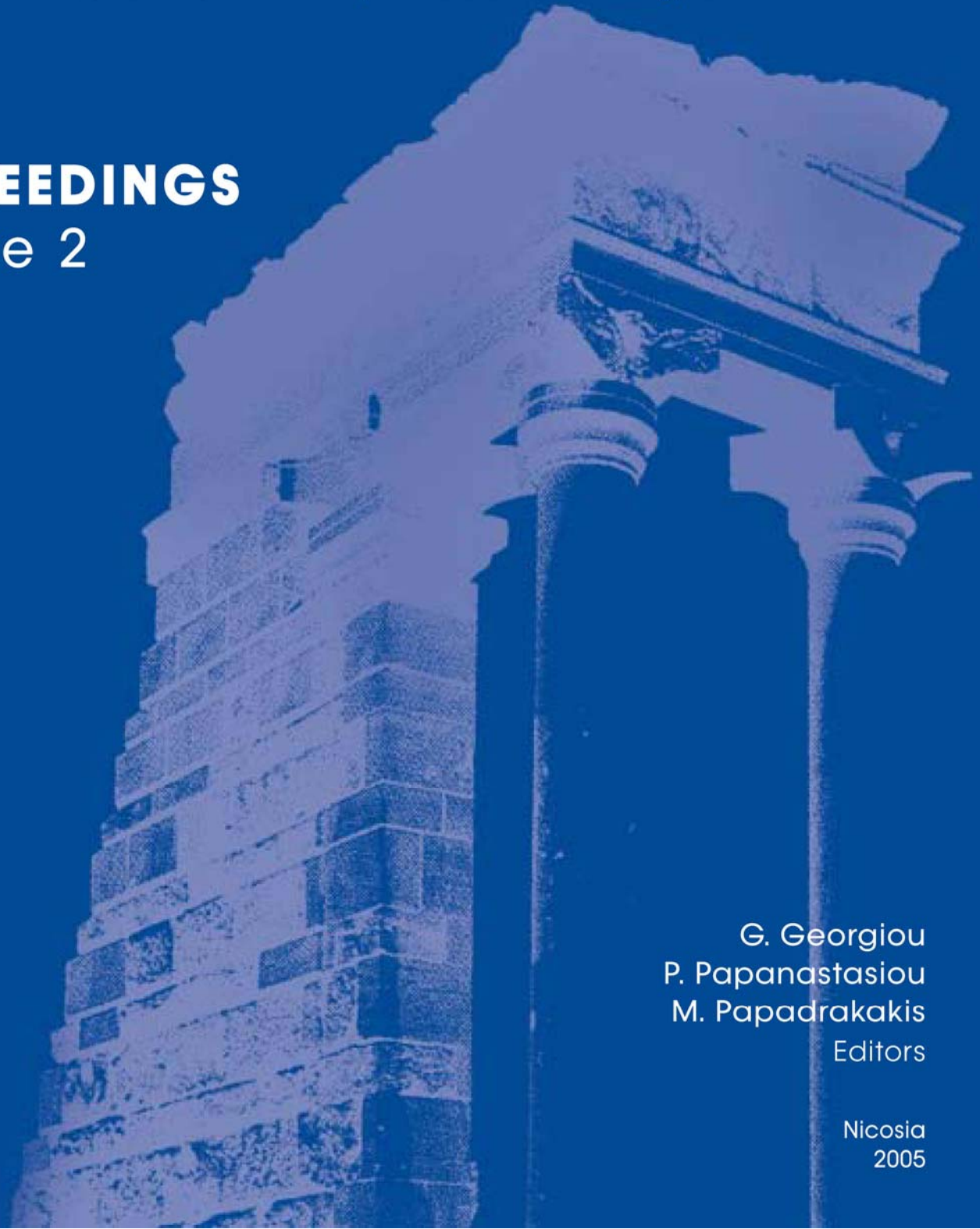
5th GRACM International Congress on Computational Mechanics

GRACM 05

29 June - 1 July 2005, Limassol, Cyprus

Dedicated to the Memory of Professor John H. Argyris

PROCEEDINGS
Volume 2



G. Georgiou
P. Papanastasiou
M. Papadrakakis
Editors

Nicosia
2005

Proceedings of GRACM 05

VOLUME 2

5th GRACM International Congress
on Computational Mechanics
Limassol, Cyprus, 29 June – 1 July 2005

Dedicated to the Memory
of Professor John H. Argyris

Edited by
G. Georgiou, P. Papanastasiou and M. Papadrakakis

Nicosia
2005

ISBN set 9963 – 644 – 50 - 3
ISBN v.2 9963 – 644 – 52 - X
Kantzilaris Publications
Nicosia
June 2005

Preface

The 5th International Congress of the Greek Association of Computational Mechanics (GRACM) will be held in Limassol, Cyprus during 29 June - 1 July 2005. The previous conferences in the series were held in Athens, 1992; Chania, 1996; Volos, 1999; and Patras, 2002.

Since their inception, the conferences have grown in size and scope covering more and wider areas of Computational Mechanics. In addition to the 8 invited plenary papers, 107 contributed papers from 12 countries have been accepted for presentation in GRACM05. These papers covered a wide range of topics: Solid and Structural Mechanics, Mechanics of Materials, Structural Dynamics and Earthquake Engineering, Stability and Chaos, Fracture Mechanics, Fluid Mechanics and Hydraulics, Aerodynamics, Transport phenomena, Electromagnetism, Biomechanics, Inverse Problems in Mechanics, Structural Control and Optimization, Artificial Intelligence and Expert Systems, System Identification, Numerical Methods and Algorithms, Finite Elements, Boundary Elements and Finite Differences.

The fifth Congress heralds new and innovative activities in a number of areas of Computational Mechanics and addresses the important issue of where the developments stand today and what the future will be in the applications of research developments for the solution of complex problems of modern technology.

The aims of the Congress are to encourage graduate student participation as well as to become a forum for critical discussion so as to lead to an assessment of past developments and future application and research needs. The outcome is expected to help researchers and engineers to shoulder important responsibilities toward the use of computational methods and mathematical models for the solution of a wide range of Engineering problems, and to pursue advanced research for the understanding and definition of the issues that remain to be addressed.

We thank the authors of the plenary and contributed papers for timely submission and participation in the Congress, the reviewers of the papers, the members of the Scientific Committee and the members of the Organizing Committee for their support and guidance. We also express our appreciation to the technical and financial cosponsors, including the Department of Civil & Environmental Engineering and the Department of Mathematics & Statistics of the University of Cyprus, the Hyperion Systems Engineering, the Cyprus Tourism Organization and the Department of Antiquities, Cyprus. Thanks are also due to the vice-president of GRACM, professor A. Boudouvis, and to all members of the Executive Board of GRACM for their continuous support and close cooperation.

An electronic color version of these Proceedings can be found at

http://www.ucy.ac.cy/~gracm05/e-Proceedings.htm

Georgios Georgiou
Panos Papanastasiou
Manolis Papadrakakis

June 2005

Author Statistics

The origin of a paper is assumed to be that of the presenting author. The numbers in parentheses include the affiliations of all co-authors.

Greece		**79** (83)
Athens	**46** (48)	
Crete	**8** (9)	
Ioannina	**3**	
Patras	**10**	
Thessaly	**3**	
Thessaloniki	**9** (10)	
Cyprus		**14**
USA		**8** (13)
UK		**5** (8)
Germany		**2** (6)
Canada		**1**
Israel		**1** (2)
Italy		**1** (2)
Japan		**1**
The Netherlands		**1** (2)
Poland		**1**
Taiwan		**1**
Bulgaria		(3)
Austria		(1)
Belgium		(1)
Hungary		(1)
Spain		(1)
Sweden		(1)
Total:		**115**

GRACM 05 Committees

Scientific Committee

M. Papadrakakis, Athens (Chair)
A. Alexandrou, Nicosia
N. Aravas, Volos
D. Beskos, Patras
A. Boudouvis, Athens
G. Georgiou, Nicosia (Co-chair)
A. Karageorghis, Nicosia
S. Karamanos, Volos
J. Katsikadelis, Athens
N. Makris, Patras
P. Papanastasiou, Nicosia (Co-chair)
G. Provatidis, Athens
K. Spiliopoulos, Athens
S. Theotokoglou, Athens

Organizing Committee

G. Georgiou (Co-Chair)
P. Papanastasiou (Co-Chair)
A. Alexandrou
C. Doumanides
S. Kassinos
P. Komodromos
Y. Smyrlis
C. Xenophontos

Invited Speakers

Nenad Bicanic, University of Glaskow (UK)
George Karniadakis, Brown University (USA)
Chris Kleijn, Delft University of Technology (The Netherlands)
Costas Pantelides, Imperial College (UK)
Glaucio Paulino, University of Illinois, Urbana (USA)
Ekkerard Ramm, University of Stuttgart (Germany)
Matasaka Tanaka, Shinshu University (Japan)
Zohar Yosibash, Ben-Gurion University (Israel)

Reviewers

A. Alexandrou
A. Boudouvis
S. Christodoulou
G. Georgiou
I. Ioannou
S. Karamanos
P. Komodromos
E. Mitsoulis
M. Papadrakakis
P. Papanastasiou
M. Phocas
G. Provatidis
K. Spiliopoulos
S. Theotokoglou
C. Xenophontos

Session Chairs

Z. Agioutantis
A. Alexandrou
D. Beskos
A. Boudouvis
J. Ekaterinaris
S. Karamanos
J. Katsikadelis
P. Komodromos
N. Makris
E. Mitsoulis
M. Neophytou
C.P. Pantelides
M. Papadrakakis
G. Provatidis
E. Sapountzakis
K. Spiliopoulos
G.E. Stavroulakis
L. Stavridis
D. Talaslidis
J. Tassoulas
E. Theotokoglou
G. Tsamasphyros
C. Xenophontos

GRACM 05 Sponsors

University of Cyprus
Department of Civil & Environmental Engineering, University of Cyprus
Department of Mathematics and Statistics, University of Cyprus
Hyperion Systems Engineering
Cyprus Tourism Organization
Department of Antiquities, Cyprus

Table of Contents

Volume 1

Biomechanics

DYNAMICS OF LARGE SCALE VEHICLE MODELS COUPLED WITH DRIVER BIODYNAMIC MODELS

C. Papalukopulos, D. Giagopulos and S. Natsiavas

Department of Mechanical Engineering
Aristotle University
54 124 Thessaloniki, Greece
(Phone: +30 2310 99 6088, Fax: +30 2310 99 6029, e-mail: natsiava@auth.gr)

Keywords: Large scale models, biodynamic models, transmissibility, random road excitation.

Abstract. *Biodynamic response of driver/seat subsystem models coupled with mechanical models of ground vehicles is investigated. When the resulting dynamical systems are complex, an appropriate component mode synthesis methodology is applied first, leading to a drastic reduction in the dimensions of the original system, without affecting significantly the accuracy of the predictions within a prespecified frequency range. The effectiveness of the methodology developed is illustrated by numerical results. In particular, frequency spectra of several response quantities related to performance of the human and vehicle models were constructed for motions resulting from periodic and random road excitation. Among other things, the results indicate that it is necessary to include the flexibility effects of the vehicle structure for improving the model accuracy throughout the frequency range of interest.*

1 INTRODUCTION

The ride comfort and safety of vehicle drivers and passengers is a diverse, challenging and multi-disciplinary subject, which has recently become an issue with large commercial and legal importance. First, human response to dynamic excitation depends on many mechanical, physiological and psychological parameters [1]. As a consequence, in addition to performing analytical studies, a good understanding of the relation between human vibration and comfort requires extensive laboratory and real environments experimental studies of response of human body to single or multi-frequency deterministic excitation or random forcing [2-4].

Most of the previous studies of ride comfort in road vehicles employed simplified models of the driver- or the passenger-seat susbsystem, without considering the effects from the coupling with the dynamics of the vehicle [3, 4]. However, preliminary results obtained with simplified quarter car models have indicated that this coupling is important, especially around the critical frequency range 3-5 Hz. In a typical situation, this simplification is mostly adopted due to the fact that the number of degrees of freedom of a vehicle structure is relatively large. The main objective of the present work is to develop and apply a systematic methodology leading to sufficiently accurate determination of dynamic response of humans riding on simplified or complex mechanical models of road vehicles. In particular, involved systems arise frequently as a result of requirements posed on the accuracy of the vehicle response. In such cases, the structural components of a vehicle are geometrically discretized by finite elements, leading to models with a quite large number of degrees of freedom, which may reach and overcome the order of a million. The basic idea is to first reduce the dimension of the systems examined by applying appropriate component mode synthesis methodologies [5, 6]. This helps the efforts towards a systematic and comprehensive study of the dynamics exhibited by large order mechanical models. Apart from increasing the computational efficiency and speed, the reduction of the system dimensions makes amenable the application of several numerical techniques for determining the dynamic response of the complex systems, which are applicable and efficient for low order dynamical systems [7]. Here, this is exploited in accelerating the determination of critical motions of the driver and the vehicle models examined resulting in response to periodic or stationary random road excitation.

The basic theoretical ingredients, leading to a reduction in the order of the vehicle mechanical models, are briefly presented in the following section. Then, typical numerical results are presented on the human and vehicle steady state dynamics due to periodic road excitation. In the fourth section, similar results are presented for the same models but when subjected to random road excitation. Finally, the most important conclusions are summarized in the last section.

2 MECHANICAL MODELS - METHOD OF ANALYSIS

In many areas of interest to the automotive industry, including the prediction of dynamic response, optimisation and control of ground vehicles (e.g., [8-12]), the most commonly employed mechanical models are two degree of freedom quarter-car models, like the one shown in Fig. 1a. This is mostly due to their simplicity and the qualitatively correct information they provide in the low frequency range, especially for ride and handling studies. On the other hand, when the accuracy of the results is of vital importance, the mechanical models employed for the vehicle structure are more complex. For instance, in Fig. 1b is shown the body of a vehicle, which has been geometrically descretized by a relatively large number of (triangular and quadrilateral) shell finite elements and a much smaller number of solid (hexahedral) finite elements, leading to a model with 1,372,989 degrees of freedom. Such a detailed discretization is necessary when the frequency range of interest is large, which is the case when investigating hand vibration or performing vibro-acoustics studies [13]. On the other hand, the corresponding wheel and suspension substructures are represented by appropriate discrete mass, stiffness and damping elements. Finally, the driver-seat subsystem is represented by either the two degree of freedom discrete mass, stiffness and damping element model shown in Fig. 2a or by the more involved five degree of freedom model presented in Fig. 2b [3, 4]. Moreover, in all cases examined, the effects of the system nonlinearities are ignored.

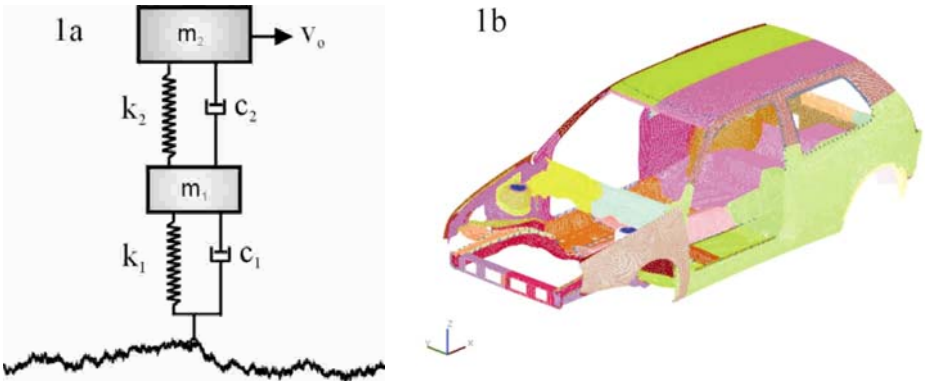


Figure 1. Vehicle models: (a) Quarter car model; (b) Finite element model.

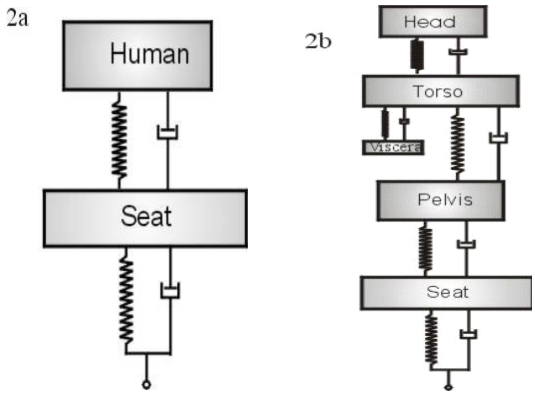


Figure 2. Human/seat subsystem models: (a) Two degree of freedom model; (b) Five degree of freedom model.

Typically, the equations of motion of a complex mechanical system include contributions from several subsystems. For instance, if for simplicity it is assumed that the system examined is composed of components I (say the vehicle body) and II (say the suspension subsystem), the equations of motion for component I alone are first derived by the finite element method in the following form

$$\hat{\mathrm{M}}_{\mathrm{I}}\,\underline{\ddot{\mathrm{x}}}_{\mathrm{I}} + \hat{\mathrm{C}}_{\mathrm{I}}\,\underline{\dot{\mathrm{x}}}_{\mathrm{I}} + \hat{\mathrm{K}}_{\mathrm{I}}\,\underline{\mathrm{x}}_{\mathrm{I}} = \underline{\hat{\mathrm{f}}}_{\mathrm{I}}(\mathrm{t})\,. \tag{1}$$

For a typical mechanical model, the order of these equations is quite large. However, for a given level of forcing frequencies it is possible to reduce significantly the number of degrees of freedom, without sacrificing the accuracy in the numerical results, by applying standard component mode synthesis methods [5]. Namely, through a coordinate transformation with form

$$\underline{x}_I = \Psi_I \underline{q}_I , \tag{2}$$

the original set of equations (1) is replaced by a considerably smaller set of equations, expressed in terms of the new generalized coordinates $\underline{q}_I$. More specifically, application of the Ritz transformation (2) into the original set of equations (1) yields the smaller set

$$M_I \ddot{\underline{q}}_I + C_I \dot{\underline{q}}_I + K_I \underline{q}_I = \underline{f}_I(t) , \tag{3}$$

where

$$M_I = \Psi_I^T \hat{M}_I \Psi_I , \quad C_I = \Psi_I^T \hat{C}_I \Psi_I , \quad K_I = \Psi_I^T \hat{K}_I \Psi_I \quad \text{and} \quad \underline{f}_I = \Psi_I^T \hat{\underline{f}}_I .$$

The most severe numerical difficulties encountered in setting up equation (3) are associated with the selection of the columns of matrix Ψ_I. In particular, this matrix includes the linear modes of the system up to a prespecified frequency plus a number of static correction modes [5], which leads to very demanding computations when large order systems are examined.

By applying an analogous treatment, a similar set of equations of motion is also obtained eventually for component II

$$M_{II} \ddot{\underline{q}}_{II} + C_{II} \dot{\underline{q}}_{II} + K_{II} \underline{q}_{II} = \underline{f}_{II}(t) . \tag{4}$$

Then, combining equations (3) and (4) leads to the equations of motion of the composite system in the classical form

$$M \ddot{\underline{x}} + C \dot{\underline{x}} + K \underline{x} = \underline{f}(t) . \tag{5}$$

The contribution of more components is treated in a similar manner, so that all the unknown coordinates are included in the vector

$$\underline{x}(t) = (x_1 \quad x_2 \quad \dots \quad x_n)^T ,$$

while M, C and K are the mass, damping and stiffness matrix of the system, respectively. These quantities include contribution from all components of the composite system examined. Finally, the elements of the vector $\underline{f}(t)$ represent the terms of external forcing, which arise from road profile geometric irregularities. In general, a typical road is characterized by the presence of large isolated irregularities, like potholes or bumps, which are superposed to smaller but continuously distributed profile irregularities. In all cases examined here, only the latter type of road irregularities are considered. Moreover, the vehicle is assumed to move on a straight path with a constant horizontal velocity component v_0 and each of the rear wheels passes over exactly the same road point as the corresponding front wheel after a constant time delay $\tau = \alpha / v_0$, where α is the vehicle wheelbase.

In a typical situation, the number of degrees of freedom of the reduced system is relatively small. Apart from increasing the computational efficiency and speed, the reduction of the system dimensions makes amenable the application of several numerical techniques, which are applicable and efficient for low order dynamical systems [6]. Here, this is exploited in accelerating the determination of steady state motions of the models examined, resulting from periodic or stationary random excitation, as explained in the following two sections.

3 NUMERICAL RESULTS FOR PERIODIC ROAD EXCITATION

This section presents a sequence of characteristic numerical results obtained by assuming deterministic road excitation. In particular, the vehicle travels over a road with harmonic profile, having amplitude s_0 and wavelength λ. Consequently, the forcing vector is proportional to s_0 and involves harmonic terms with frequency $\omega = 2\pi v_0 / \lambda$.

First, Fig. 3 presents typical frequency-response diagrams obtained at two special points of five different vehicle models, by assuming a vertical harmonic road excitation at the four wheels, with amplitude $s_0 = 1\text{cm}$. Specifically, the amplitude of the vertical acceleration obtained at the front left wheel and the driver seat is depicted as a function of the forcing frequency. In four of the models examined, the vehicle is represented by a two degree of freedom quarter car model with zero, one, two and five degrees of freedom for the driver/seat subsystem. The fifth model includes the finite element model of the vehicle body, while the driver/seat subsystem is represented by the five degree of freedom biodynamic model shown in Fig. 2b. As a general observation, the acceleration levels developed at the wheel are much higher than those developed at the seat position within the higher frequency range examined. Moreover, the presence of the human/seat subsystem makes a substantial difference in the overall dynamics, especially in the low frequency range. Finally, the results

demonstrate that the car body flexibility effects, which are taken into account more accurately by the finite element model, become important throughout the frequency range examined.

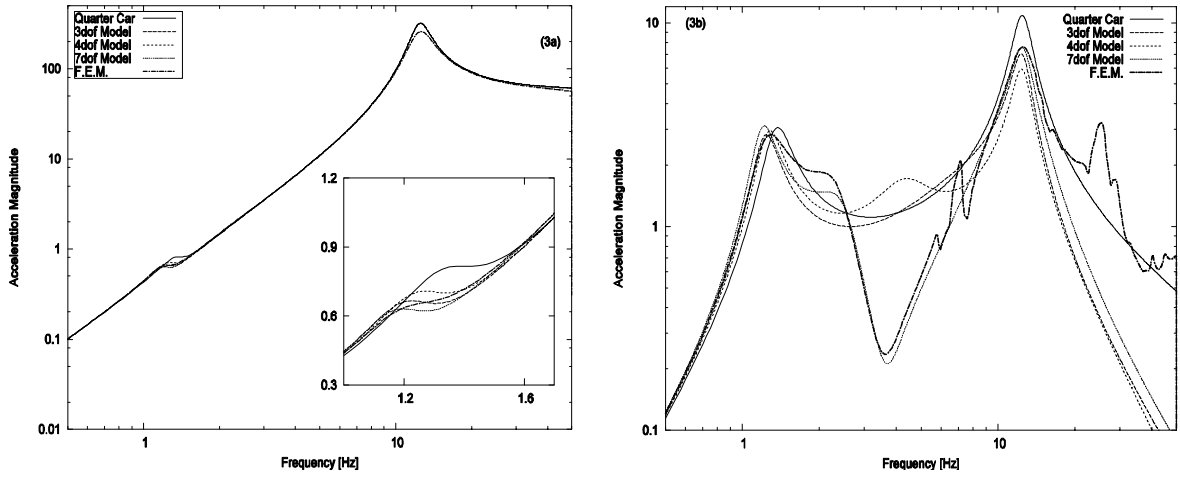


Figure 3. Frequency-response diagrams for periodic road excitation. Amplitude of vertical acceleration: (a) at the front left wheel and (b) at the driver seat.

Next, Fig. 4 shows results indicating the importance of the coupling between the dynamics of the vehicle and the biodynamic model of the human/seat subsystem. The results depicted in Fig. 4a were determined by representing the vehicle dynamics by the quarter car model of Fig. 1a, while the results shown in Fig. 4b were obtained by employing the finite element model for the vehicle structure. In both cases, the human/seat subsystem was represented by the five degree of freedom biodynamic model of Fig. 2b. In addition, the continuous lines indicate results obtained by applying the road excitation at the base of the quarter car model, while the broken lines indicate results obtained by applying the same excitation directly to the base of the biodynamic model. Specifically, Fig. 4a shows the magnitude of the transmissibility ratio of the acceleration at the pelvis to the acceleration at the ground (continuous line) or at the base of the seat (broken line), as a function of the forcing frequency. On the other hand, Fig. 4b shows the magnitude of the same transmissibility ratios for the acceleration at the driver head, as a function of the forcing frequency. The differences between the continuous and the broken lines observed in both cases illustrate the fact that ignoring the vehicle dynamics leads to significant inaccuracies when predicting human response. Similar differences were also observed for all the other degrees of freedom of the human model.

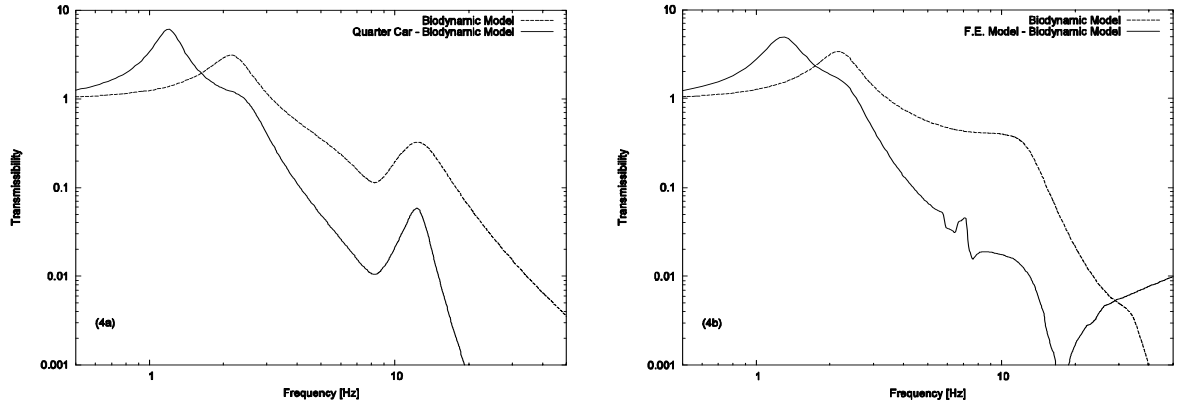


Figure 4. Magnitude of the transmissibility functions for the vertical acceleration: (a) at the driver pelvis (quarter car model) and (b) at the driver head (finite element vehicle model).

The following set of diagrams completes the picture on the response of the large scale vehicle model to periodic road excitation. In particular, Fig. 5a presents the magnitude of the transmissibility ratio for the acceleration obtained at the five degrees of freedom of the biodynamic model shown in Fig. 2a, calculated with the nominal mass of the driver ($m_d = 70kg$). Likewise, Fig. 5b depicts similar results, determined after increasing the driver mass to $m_d = 100kg$.

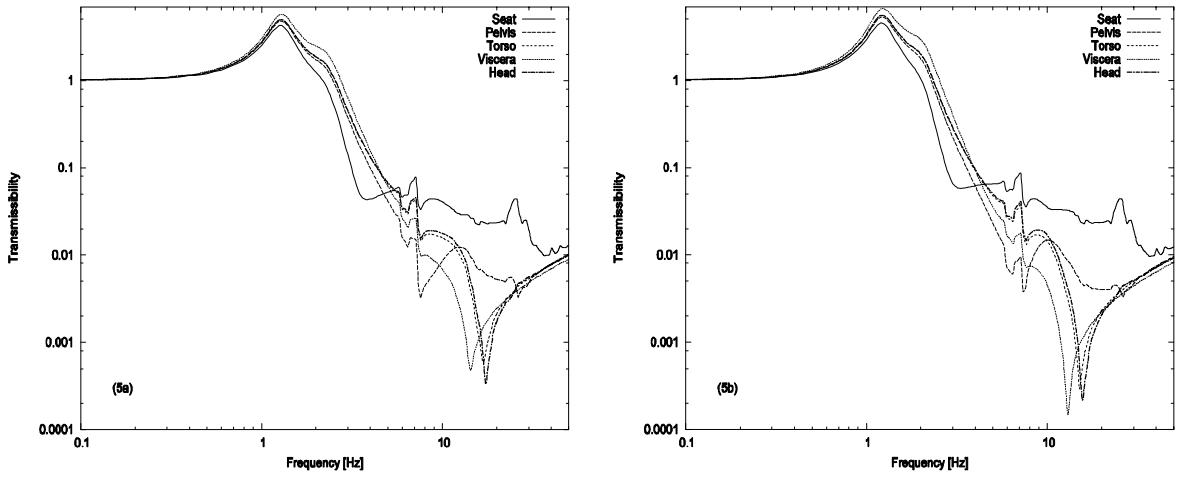


Figure 5. Magnitude of the transmissibility functions for the acceleration of the biodynamic model degrees of freedom, for a driver with mass: (a) $m_d = 70\text{kg}$ and (b) $m_d = 100\text{kg}$.

4 NUMERICAL RESULTS FOR RANDOM ROAD EXCITATION

In the sequel, more realistic road profiles were selected. Namely, the vehicle models were assumed to travel over road profiles characterized by real-valued, zero mean, stationary and Gaussian random fields. For their complete statistical description, it is sufficient to specify the power spectral density of the road irregularities, say $S_g(\Omega)$, where $\Omega = 2\pi/\lambda$ is a spatial frequency, corresponding to a harmonic irregularity with wavelength λ. According to previous investigations on the subject (e.g., [14]), the geometrical profile of typical roads fits sufficiently accurately the following simple analytical form

$$S_g(\Omega) = \begin{cases} S_g(\Omega_0)(\Omega/\Omega_0)^{-n_1}, & \text{if} \quad \Omega \le \Omega_0 \\ S_g(\Omega_0)(\Omega/\Omega_0)^{-n_2}, & \text{if} \quad \Omega \ge \Omega_0 \end{cases} \tag{6}$$

where $\Omega_0 = 1/2\pi$ is a reference spatial frequency. Then, since the mechanical models examined possess linear properties, knowledge of the road profile spectral density and the vehicle horizontal velocity permits evaluation of the spectral density of the stationary vehicle response through the well-known formula

$$S_{xx}(\omega) = H(\omega) S_{gg}(\omega) H^{T*}(\omega).$$

In the last formula, $\omega = \Omega v_0$ is the temporal frequency, $S_{xx}(\omega)$ and $S_{gg}(\omega)$ represent the spectral density matrices of the response and the forcing, respectively, while $H(\omega)$ is the matrix including the frequency response functions of the system [15].

First, Fig. 6 presents typical response diagrams obtained for the driver seat by employing the same mechanical models as those used to determine the results presented in Fig. 3. Namely, Fig. 6a shows the power spectral density of the acceleration obtained at the driver seat as a function of the forcing frequency, in response to random excitation caused at the four wheels of the vehicle. The amplitude of the road irregularities is selected to be about the same as that of the harmonic profile employed in the previous section, resulting in a good quality road, while the vehicle travels with a constant horizontal speed $v_0 = 120\,\text{km/h}$. Likewise, Fig. 6b shows similar results, obtained by assuming that the vehicle moves with a horizontal speed $v_0 = 70\,\text{km/h}$ over a bad quality road. The results depicted in Fig. 6 present a lot of similarities with the corresponding results shown in Fig. 3b. Again, the presence of the human/seat subsystem makes a substantial difference in the low frequency range examined. Moreover, the car body flexibility effects are important throughout the frequency range examined.

Next, Fig. 7 shows results indicating the importance of the coupling between the dynamics of the vehicle and the biodynamic model. Specifically, Fig. 7a depicts the power spectral density of the driver pelvis acceleration determined by using a good quality road and representing the vehicle dynamics by the quarter car model of Fig. 1a. Likewise, Fig. 7b shows similar results for the driver head, obtained for the finite element vehicle body

model of Fig. 1b. In both cases, the human/seat subsystem was represented by the five degree of freedom model, while the broken lines indicate results obtained by applying the road excitation directly to the base of the biodynamic model. Again, the differences observed illustrate the fact that ignoring the vehicle dynamics leads to significant inaccuracies in predicting the vehicle driver response.

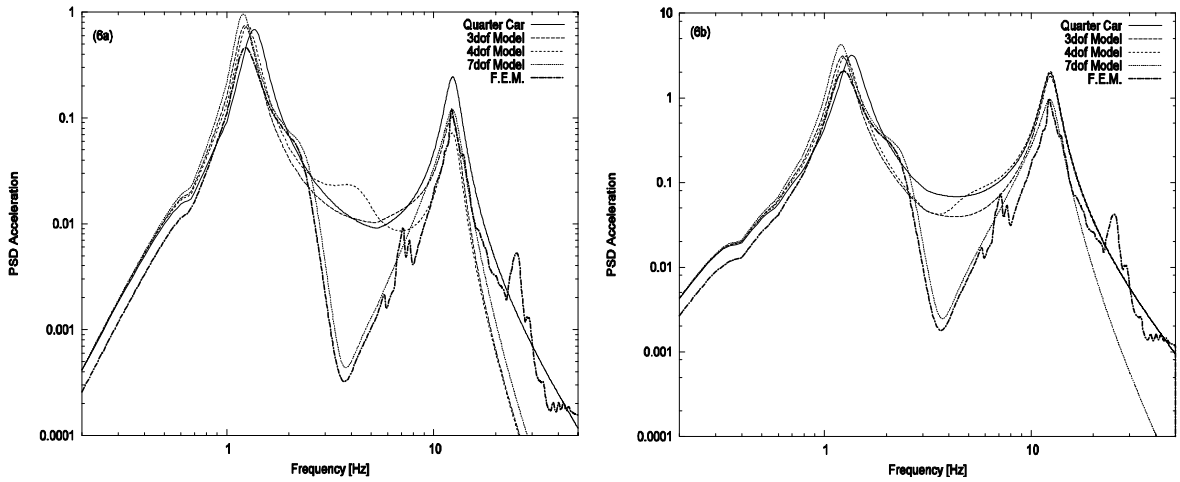


Figure 6. Power spectral density of the driver seat acceleration caused by: (a) a good quality road and $v_0 = 120\,\mathrm{km/h}$ and (b) a bad quality road and $v_0 = 70\,\mathrm{km/h}$.

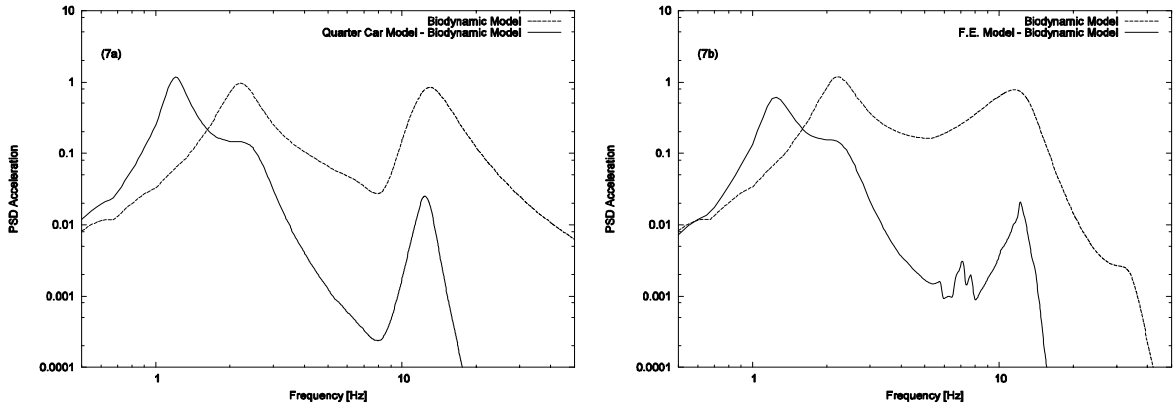


Figure 7. Power spectral density of the vertical acceleration: (a) at the driver pelvis (quarter car model) and (b) at the driver head (finite element vehicle model).

The final set of diagrams completes the picture on the response of the large scale vehicle model to random road excitation. In particular, Fig. 8a presents the power spectral density of the acceleration obtained at the front left wheel and the five degrees of freedom of the biodynamic model shown in Fig. 2a for a good quality road and $v_0 = 120\,\mathrm{km/h}$. Obviously, most of the energy in the signals referring to the degrees of freedom of the biodynamic model is distributed in a frequency range around 1 Hz, while the energy in the wheel acceleration signal is concentrated around 10 Hz. Likewise, Fig. 8b depicts similar results, determined for the off-diagonal elements of the spectral density matrix, corresponding to excitation applied to the front left wheel and measurements taken at the degrees of freedom of the biodynamic model. Most of the energy in all the signals shown is distributed and extends over the frequency range up to about 50 Hz. This is illustrated in a better way by the results of Fig. 9a, showing the cumulative rms value corresponding to the acceleration values shown in Fig. 8a. For comparison, Fig. 9b presents the same quantities but obtained by employing a quarter car model, instead.

In closing, it is important to note that the calculation of other response quantities, like apparent mass, correlation functions and vibration dose values, which are useful in assessing the ride confort of a vehicle [1-4], can also be performed in a similar fashion. For instance, the number of zero-crossings for the signals considered can be obtained as a by-product of the calculations performed. More specifically, for the forcing conditions that led to the results presented in Fig. 8, the number of zero-crossings in the vertical acceleration obtained at the left front wheel are 18, while for the driver seat are 11. Finally, Fig. 10a presents the corresponding auto- correlation

function obtained for the acceleration at the driver seat position, while Fig. 10b shows the cross-correlation function obtained for the same mechanical model between the front left wheel and the driver seat.

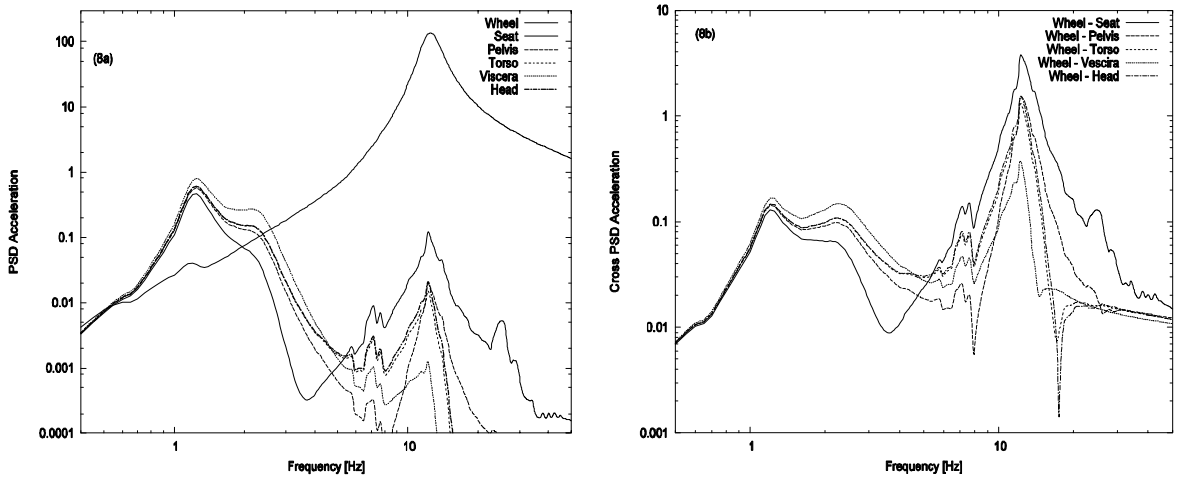


Figure 8. (a) Power spectral density of the acceleration at the front left wheel and the degrees of freedom of the biodynamic model. (b) Cross spectral density corresponding to excitation applied to the front left wheel and acceleration measured at the degrees of freedom of the biodynamic model.

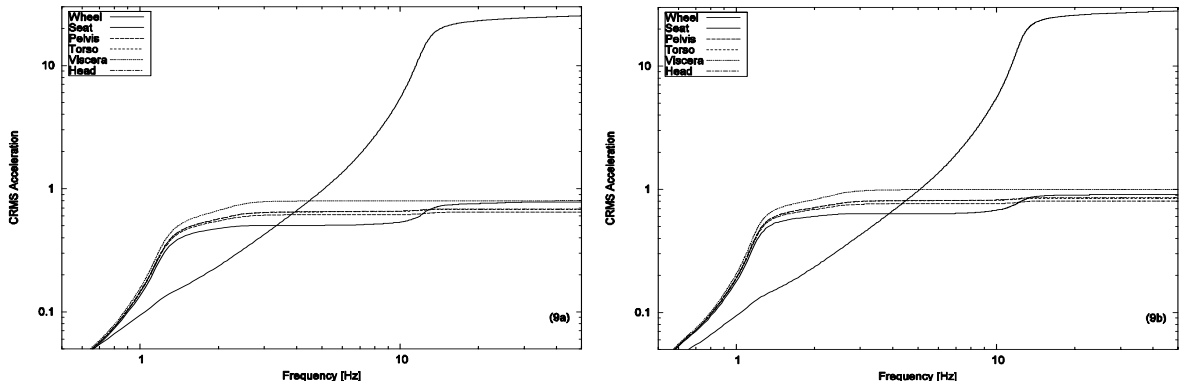


Figure 9. Cumulative rms value corresponding to the acceleration values shown in Fig. 8a. Results obtained for: (a) the finite element model and (b) the quarter car model.

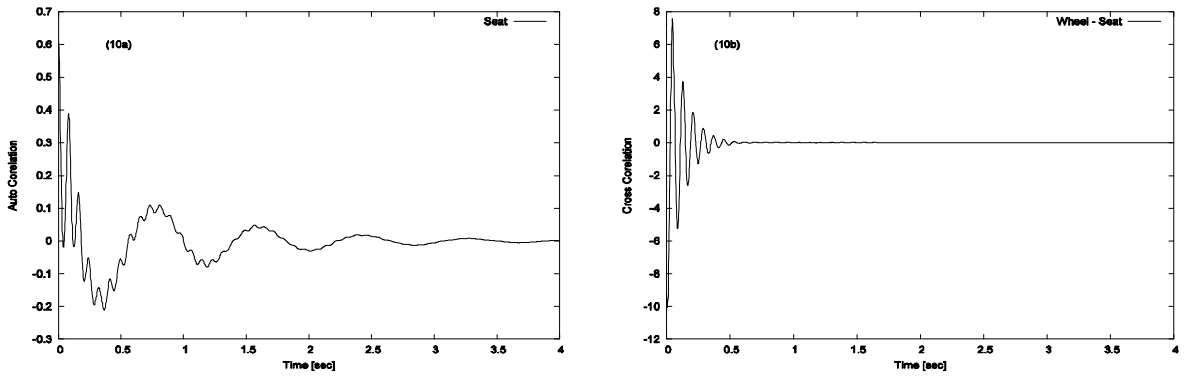


Figure 10. (a) Auto-correlation function for the acceleration at the driver seat. (b) Cross-correlation function between the front left wheel and the driver seat position.

5 CONCLUSIONS

A methodology was developed for determining response of biodynamic driver/seat models coupled with vehicle models. For large scale models, the basic idea was to first apply an appropriate reduction methodology in order to eliminate a substantial number of the original degrees of freedom, so that the reduced model is sufficiently

accurate up to a prespecified level of forcing frequencies. Then, frequency response spectra of several response quantities related to the dynamic performance of the vehicle and the driver under periodic or random road excitation were constructed. Among other things, it was shown that the car model flexibility effects are necessary for improving the model accuracy throughout the frequency range of interest. Moreover, the results demonstrated that the dynamic quantities determined form a basis for computing other response quantities, like transmissibility functions, vdv and rms values, which are valuable for assessing ride comfort in a vehicle. The methodology presented can be easily extended to include the effect of other types of excitation as well as the effect of strong nonlinearities. It will also help the efforts to select optimum values for the seat technical characteristics.

ACKNOWLEDGMENTS

This research was partially supported by a grant from the General Secretariat of Research and Technology, Greek Ministry of Development, through the PENED 2001 program (01-ΕΔ-330).

REFERENCES

[1] Griffin, M.J. (2003), *Handbook of Human Vibration,* Academic Press. San Diego, USA.

[2] Paddan, G.S. and Griffin, M.J. (1998), "A review of the transmission of translational seat vibration to the head," Journal of Sound and Vibration 215, pp. 863-882.

[3] Wu., S., Rakheja, S. and Boileau, P.-E. (1999), "Analyses of relationships between biodynamic response functions," Journal of Sound and Vibration 226, pp. 595-606.

[4] Choi, Y.T. and Wereley, N.M. (2003), "Mitigation of biodynamic response to vibratory and blast-induced shock loads using magnetorheological seat suspensions," IMECE2003, Washington, D.C.

[5] Craig, Jr., R.R. (1981), *Structural Dynamics - An Introduction to Computer Methods*, J. Wiley & Sons, New York.

[6] Verros, G. and Natsiavas, S. (2002), "Ride dynamics of nonlinear vehicle models using component mode synthesis", ASME Journal of Vibration and Acoustics, 124, pp. 427-434.

[7] Verros, G., Goudas, H., Natsiavas, S. and Hache, M. (2000), "Dynamics of large scale vehicle models," *2000 International ADAMS Conference***,** Orlando, Florida.

[8] Ellis, J.R. (1969), *Vehicle Dynamics***,** Business Books Limited, London.

[9] Gillespie, T.D. (1992), *Fundamentals of Vehicle Dynamics,* Society of Automotive Engineers, Warrendale, PA.

[10] Karnopp, D., Crosby, M.J. and Harwood, R.A. (1974), "Vibration control using semi-active generators," ASME Journal of Engineering for Industry 96, pp. 619-626.

[11] Hrovat, D. (1993), "Application of optimal control to advanced automotive suspension design," ASME Journal of Dynamic Systems, Measurement and Control 115, pp. 328-342.

[12] Metallidis, P., Verros, G., Natsiavas, S. and Papadimitriou, C. (2003), "Fault detection and optimal sensor location in vehicle suspensions," Journal of Vibration and Control 9, pp. 337-359.

[13] Kropp, A. and Heiserer, D. (2003), "Efficient broadband vibro-acoustic analysis of passenger car bodies using an FE-based component mode synthesis approach," Journal of Computational Acoustics 11, pp. 139-157.

[14] Dodds, C.J. and Robson, J.D. (1973), "The description of road surface roughness," Journal of Sound and Vibration 31, pp. 175-183.

[15] Roberts, J.B. and Spanos, P.D. (1990), *Random Vibration and Statistical Linearization,* J. Wiley and Sons, New York, NY.

5th GRACM International Congress on Computational Mechanics
Limassol, 29 June - 1 July, 2005

A PASSIVE AND AXISYMMETRIC MODEL FOR SKELETAL MUSCLE UNDER TRANSVERSE COMPRESSION AND LONGITUDINAL TENSION

A. Stoicheios, V. Kefalas and D. A. Eftaxiopoulos

School of Applied Mathematical and Physical Sciences
National Technical University of Athens, Zografou Campus, 15773 Athens, Greece

Keywords: Muscle, Experiment, Tension, Finite Elements, Compression.

Abstract. *In vitro, quasi static tension experiments have been performed on porcine femoral muscle. The nonlinear stress - strain curves obtained from these experiments, have been introduced in a finite element analysis of an ideal, axisymmetric and passive (i.e with no contractile behavior) model of skeletal muscle, under both longitudinal tension and transverse (perpendicular to the fibers) compression. The finite element computation yielded that the lowest minimum principal stresses, i.e the most compressive normal stresses in the model, develop in the vicinity of the core of the muscle belly. Thus the damage inflicted on the muscle tissue, due to a transverse compressive static load, may be more severe in the interior of the muscle than on the contact area between the external load and the muscle boundary. Analogous observations have been reported in the literature, regarding different muscle models.*

1. INTRODUCTION

The mechanical behavior of skeletal muscle under transverse compressive loads, static or dynamic, is related either to pressure sores caused to paralyzed or geriatric patients that lie in bed for long periods of time or to trauma suffered by car passengers, after a car crash. In the literature however, only quasi - static transverse compressive loads acting on skeletal muscle have been considered. Experimental data for the passive behavior of musculoskeletal tissues under compression, have been presented by Grieve and Armstrong[1]. Bosboom et al[2] have investigated experimentally the transverse force - length response of skeletal muscle under in - vivo compression and have used the Ogden[3] non - linear and viscoelastic material model, for the simulation of their experimental results. Maggana et al[4] have obtained stress - strain experimental results for transverse and longitudinal muscle compression and have used the Langevin model for their constitutive interpretation.

Linder - Ganz and Gefen[5] have done finite element analyses of human muscle in the shoulders, pelvis, head and heels, subjected to compressive loads and concerned with patients lying in bed for long periods. Oomens et al[6] and Todd and Thacker[7] have performed finite element analyses of the human buttocks under compression, related to wheelchair users. These three references[5],[6],[7] report that compressive stresses in the muscle interior are more severe than the contact compressive stresses applied on the muscle boundary. Breuls et al[8] have performed a theoretical analysis of damage evolution in skeletal muscle tissue with reference to pressure ulcer development. Spears et al[9] used finite elements to investigate the effect of different saddle shapes, on the stresses induced in the cyclist perineum during cycling.

As long as the tensile response of muscle and tendon is concerned, Hawkins and Bey[10] and Myers et al[11] performed in - vivo tensile experiments on animal muscle and tendon tissue and have obtained force - length and stress - strain results respectively. Maganaris and Paul[12] have achieved stress - strain curves for tendons under tension.

In this study we have performed in - vitro tension experiments, on porcine femoral muscle. The nominal stress - engineering strain curve for the muscle turns out to be nonlinear. This curve has been used in a finite element analysis of an ideal, simplified and axisymmetric muscle - tendon compound, subjected to combined longitudinal tension and transverse compression. The tendon stress - strain curve used in the finite element analysis was taken from Maganaris and Paul[12]. The finite element computation yielded that the minimum principal stress, i.e the most compressive normal stress, develops in the vicinity of the core of the muscle belly and not on the boundary where the external transverse compression is applied. Thus the damage inflicted on the muscle tissue due to a transverse compressive static load, may be more severe in the interior of the muscle belly than on its boundary.

2. PREPARATION FOR THE EXPERIMENTS

Tension experiments were performed on specimens taken from the femoral muscle of a Landrace male piglet. The age of the animal was 8 months and its weight was 60 kg. Immediately after the animal sacrifice, both femurs together with the surrounding tissues were removed from the animal body and kept in sodium chloride solution

0.9% w/v at normal refrigeration temperature 6^oC - 8^oC.

The experiments were performed within 1 - 4 days after the animal was sacrificed. Before the experiments, the femurs were left to thaw for 15hrs, at room temperature. Then, the muscle tissue bundle surrounding the femur was removed. Afterwards, several major femoral muscles, from tendon to tendon, together with their surrounding membrane, were removed from the bundle of muscles around the femur. Obviously, a major femoral muscle together with the tendons at its ends, has a varying cross sectional area along its length. The cross sectional area of the muscle is greater in the belly of the muscle and smaller in the neck of the muscle, where the tendon is. The specimens that were picked for the experiments were those with the most uniform cross sectional area along the belly. All specimens were tested at room temperature.

In order to measure the strain in the vicinity of the belly of the muscle, where the cross sectional area is almost uniform and is far from the tendons where the grips were applied, white strips perpendicular to the longitudinal axis of the specimens, were painted on the surface of the muscle tissue. Six strips were marked on the tissue, at approximately equal distance from each other, using a correcting fluid marker. The width of the strips was not uniform across the muscle tissue and varied from 0.5 mm to 1.5 mm. The cross sectional areas of the two specimens, namely PMT1-2 and PMT4-1, that were used in the experiments, are given in Table 1. These cross

Muscle specimen	Cross sectional area (mm^2)
PMT1-2	80.75
PMT4-1	84

Table 1: The cross sectional areas of the specimens used in the experiments

sectional areas were found by considering the muscle belly cross section as almost circular and measuring the length of its diameter using a dial extensometer. This measurement was done when the specimens were attached to the machine grips.

Prior to testing the test machine load cell was checked to have an error of less than 3%.

3. EXPERIMENTS

Tension experiments were performed on the two specimens, namely PMT1-2 and PMT4-1, of porcine muscle (see Table 1). The end tendons, together with the muscle tissue between them, were attached to the INSTRON 1121 machine grips. An instant drying glue was used on the interface between the tendons and the grips. The velocity of the grip displacement was 10mm/min. A video camera recorded the evolution of the tension experiments. The resolution of the video frames that were analyzed was 300 (height) $\times$ 500 (width) pixels and their speed was 36 frames/s. The colour contrast between the white stripes and the reddish muscle was analyzed via image processing, with the help of a code written in Mathcad. A rectangular area that contained the striped region of the muscle, was analyzed. This region was interpreted as a matrix of rows and columns of pixels. Within this area, 256 different colours were considered. Each pixel was assigned a colour, among the 256 mentioned and each colour was assigned a number between 0 and 255. The number 0 was assigned to the black colour and the number 255 was assigned to white colour, with all other colours taking the values in between. At each time and consequently at each load step that was considered, the numbers corresponding to the colours of the pixels on each row of the rectangular region, were added and the sums were plotted against the column number corresponding to each row. The peaks obviously corresponded to the white strips at the point of maximum contrast (see Figure 1). In this way, variations of white strip width due to marking irregularities, was diminished. The displacements of the peaks, measured in number of pixels, were divided by the original length between them, measured in number of pixels as well. Thus we could get a measure of the local engineering strain within the muscle. A mean value of the strains corresponding to neighboring pairs of peaks, could give an estimate of the strain in the specimen. The strains were also measured by using the INSTRON machine extensometer. The discrepancy between the measurements of the video resolution system and the extensometer was found to be 10% at strain 0.05 and 7% at strain 0.08.

4. EXPERIMENTAL RESULTS

In Figure 2 the quasi - static tension experimental results of the three muscle specimens tested (see Table 1), are presented. Fitting with polynomial curves has been performed through the experimental result points. The curve PMT4-1 was used in the finite element simulation that is presented in the following sections. It is evident that the muscle exhibits nonlinear stress - strain behavior. Such a behavior is typical for soft biological tissues and consists

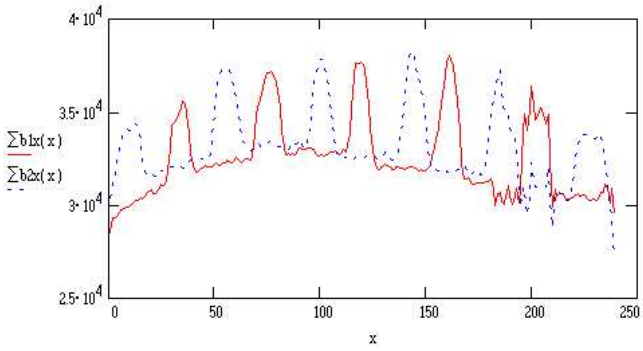


Figure 1: The white peaks on a muscle experimental specimen at time $t = 0s$ (red line) and at $t = 30s$ (blue line)

of a low stiffness region at low strains and an increased stiffness region at higher strains. The high stiffness region is due to the combined load carrying capacity of the individual muscle fibers, which when are stretched, form an overall stiff bundle. The mechanical behavior that is portrayed in the curves in Figure 2 may be more stiff than its in - vivo counterpart and the reason may be the refrigeration of the muscle prior to the experiments. Refrigeration changes the mechanical properties of the muscle[13].

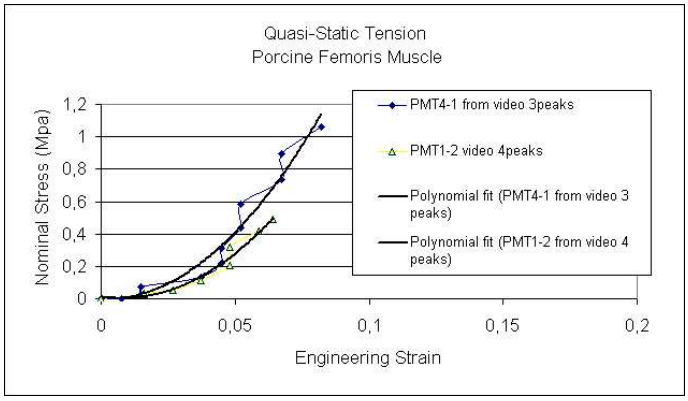


Figure 2: Experimental results from uniaxial tension tests on muscle specimens

5. FINITE ELEMENT ANALYSIS

A simplified axisymmetric model for fibrous skeletal muscle was analyzed via the finite element method. This ideal geometry, consists of a typical belly part and a neck part. A half of the generator plane surface, which produces the three dimensional muscle shape after a 360^0 revolution, appears in Figure 3. The axis of axial symmetry, lies along the left most vertical side of the model shown in Figure 3. The origin is located at the lower left most corner of the model. We consider only half of the generator plane surface, due to the mirror symmetry with respect to the horizontal plane that passes from the origin. The x axis lies along the horizontal plane of mirror symmetry and the y axis coincides with the axis of axial symmetry. Note that the lower (belly) area contains the muscle material and the upper (neck) area contains the tendon material. At the upper end of the neck the half - diameter is 1.5 cm and at the belly lower end the half - diameter is 4cm. The half length of the muscle - tendon construction, i.e the distance between the belly lower end and the neck upper end, is 12cm. Plane, two dimensional, axisymmetric triangular elements, with six nodes, have been used for the discretization of the model.

An axial tensile traction of 1MPa was applied on the end cross section of the neck (see Figure 3). A lateral, normal pressure of 1MPa was also applied on the generator curve of the muscle boundary (see Figure 3). According to the stress - strain curve used for the muscle material in this article (see Figure 2), in a simple tension experiment on a muscle bar of uniform thickness, a nominal tensile stress of 1MPa, corresponds to an engineering tensile strain of about 10%. Here the lateral load of 1MPa is compressive, but the constitutive model that we used implies an

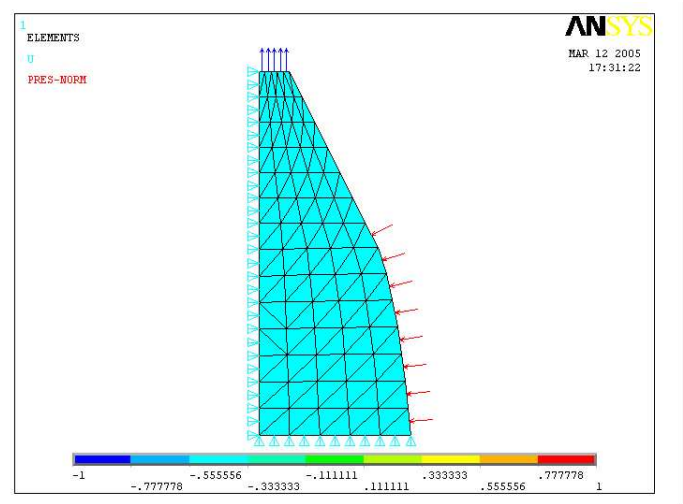


Figure 3: A coarse discretization of the muscle - tendon geometry

identical behavior under compression. We employed a static transverse compressive load in our model, while in car crashes the loads exerted on the human body are dynamic. The muscle is considered in the passive state, i.e with the contractile behavior inactive.

The displacement constraints, $u_y = 0$ on the lower horizontal boundary and $u_x = 0$ on the left vertical boundary, appear in Figure 3 too. A finer mesh in the muscle region, that was also used as a check for our results, appears in Figure 4. Note that in Figure 4 the muscle region is the lower one with the fine mesh, while the tendon region is the upper one with the coarse mesh.

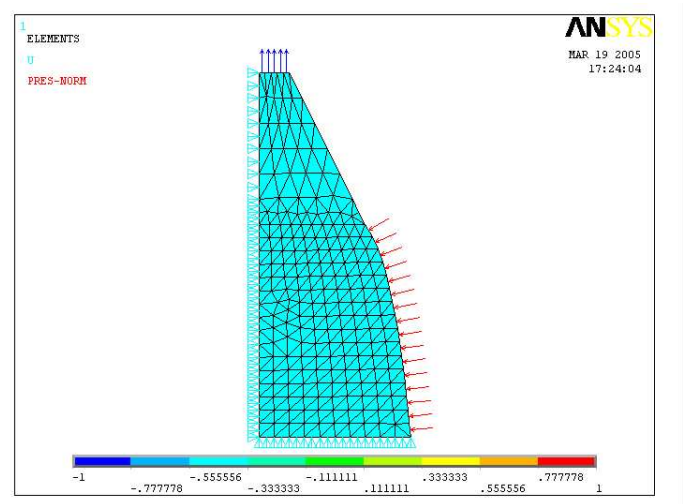


Figure 4: A fine discretization of the muscle region within the muscle - tendon geometry

A nonlinear elastic incompressible and isotropic model, for both the muscle and the tendon was used. In reality, the muscle and the tendon materials are transversely isotropic, due to their composition of fiber bundles along the axial direction. Also in reality the muscle and tendon materials may not be incompressible. Large displacements were also incorporated in the muscle and tendon deformation models.

The finite element package ANSYS 8.0 was used for the numerical simulation of the mechanical behavior of the muscle - tendon compound. The Multilinear ELAStic (MELAS) model of ANSYS was used for the introduction of the constitutive stress - strain relation of the materials (muscle and tendon). The nonlinear stress - strain curves used within the finite element simulation were taken from Figure 2 (curve PMT4-1) as long as the muscle is concerned and from Maganaris and Paul[12] (figure 4B in that paper) as long as the tendon is concerned. Via

the MELAS material model of ANSYS, these curves were converted to multilinear ones i.e to lines that consist of linear segments of varying slope. Since in the uniaxial tension experiments, the nominal stress (first Piola - Kirchhoff) and the engineering strain, both of which refer to the undeformed configuration, were recorded, we had to convert these to the Cauchy stress and to the logarithmic strain respectively, because in ANSYS an analysis with respect to the current configuration is performed. The relevant conversion relations are

$$e = \ln \lambda \tag{1}$$

where e is the logarithmic strain along the axial direction in a simple tension experiment and λ is the stretch ratio along that direction, given in turn via the relation

$$\lambda = 1 + \epsilon \tag{2}$$

where ϵ is the tensile engineering strain. The corresponding conversion relationship for the stress is

$$\sigma = \lambda \Pi \tag{3}$$

where σ is the Cauchy (true) stress and Π is the first Piola - Kirchhoff (nominal) stress. Since both the muscle and tendon materials were considered as almost incompressible, a Poisson's ratio equal to 0.45 was assigned to them.

6. FINITE ELEMENT RESULTS

Results in this paragraph are obtained by using the coarse mesh shown in Figure 3. The variation of the horizontal displacement u_x is shown in Figure 5. As expected the most negative value of u_x is at the boundary where the lateral pressure is applied. Along the axis of axial symmetry, we have $u_x = 0$. In Figure 6 the variation

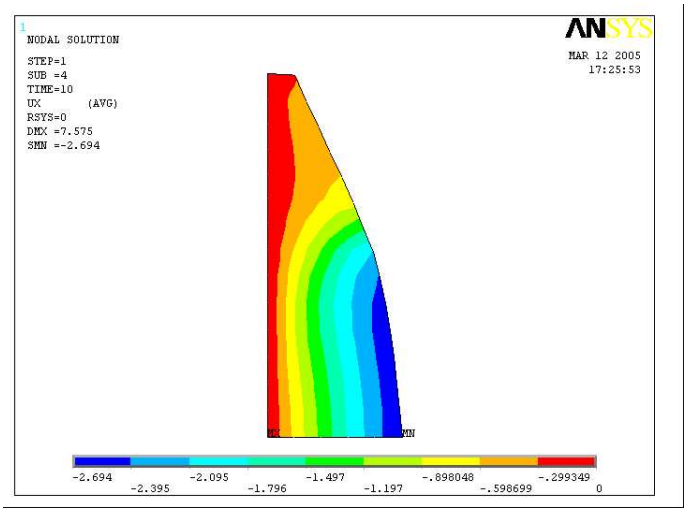


Figure 5: The variation of the horizontal displacement u_x

of the vertical displacement u_y is shown. The biggest value is on the neck upper boundary where the tensile load is applied and a zero value appears along the x axis which is an axis of symmetry. The variation of the maximum principal stress σ_1 appears in Figure 7. The greatest value develops at the neck of the compound where the tensile traction is applied. On the neck upper boundary the maximum principal stress is equal to the applied traction. The fluctuation of the minimum principal stress σ_3 is shown in Figure 8. The biggest absolute value of σ_3 develops at the central area of the belly of the muscle and not close to the line of application of the lateral pressure. Linder - Ganz and Gefen[5] have also reported that maximal principal compression and von Mises stresses, in the interior of several animal and human muscle models, exceeded the applied boundary compressive stresses. Similar conclusions are conveyed by Oomens et al[6] and Todd and Thacker[7].

In order to to check the validity of the latter observation, we reran the finite element code with the finer mesh of Figure 4 . The outcome is shown in Figure 9 and is very close to the one obtained in Figure 8.

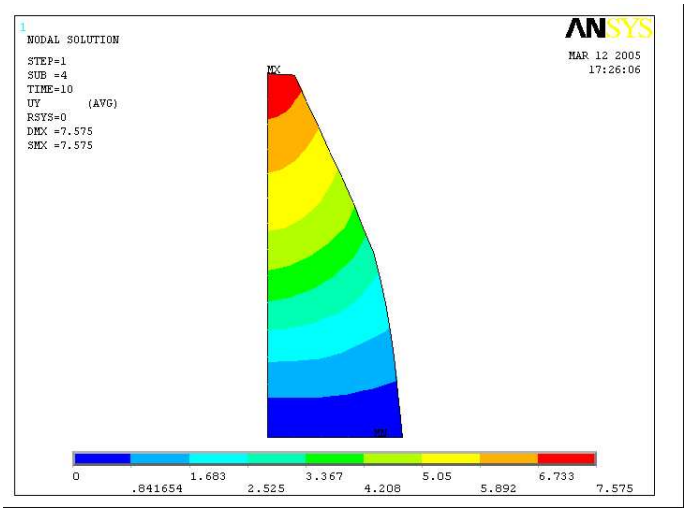


Figure 6: The variation of the vertical displacement u_y

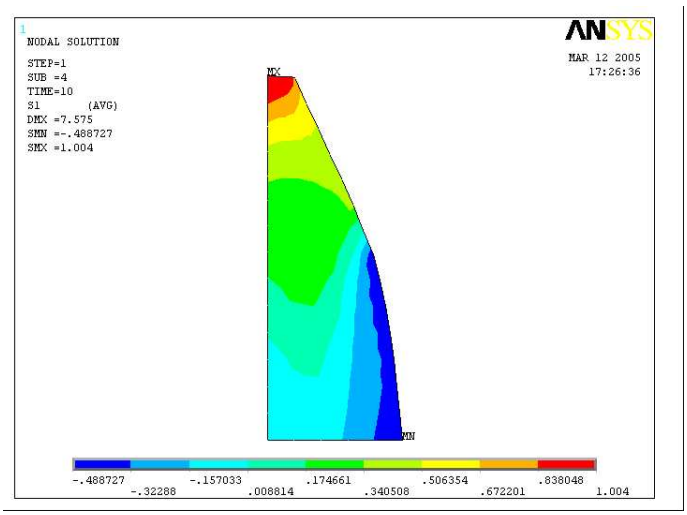


Figure 7: The variation of the maximum principal stress σ_1

7. CONCLUSIONS

Experimental and finite element analyses of skeletal muscle have been pursued in this paper. In vitro, quasi static tension experiments, on piglet femoral muscle tissue, were performed. Finite element analysis on a simplified axisymmetric muscle model, consisting of a muscle and a tendon part and subjected to an axial tensile and a transverse compressive load, has also been conducted. The main conclusions are:

- The stress - strain relation of the muscle tissue, under quasi static tension, is nonlinear. The larger the strain the stiffer the muscle becomes, since the muscle fibers work together to form a stiff bundle. Such a behavior is typical for most soft fibrous biological tissues.

- The finite element analysis of the muscle - tendon simplified axisymmetric compound, subjected to an axial tensile and a transverse compressive load, indicated that the most severe minimum principal stress, i.e the most compressive normal stress in our model, develops in the vicinity of the core of the muscle belly (see Figure 9) and not close to the boundary where the external transverse compression is applied. Analogous conclusions have been reported in the literature[5],[6],[7] for different muscle models, where maximal internal compressive stresses are greater than the applied boundary compressive stresses. Thus the damage inflicted on the muscle tissue due to a transverse compressive static load may be more serious in the muscle interior,

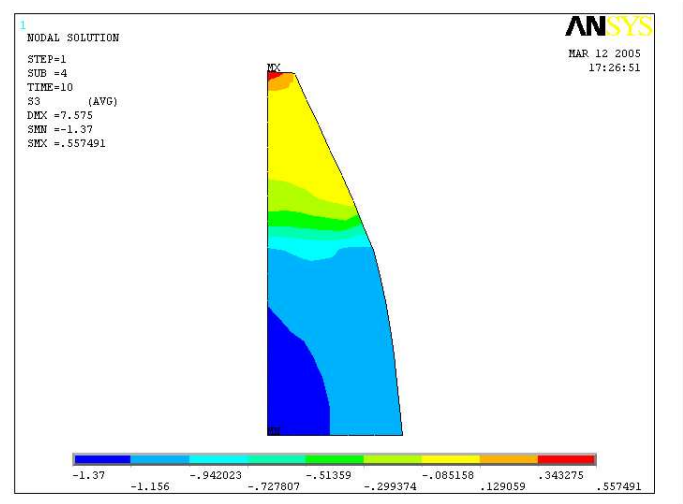


Figure 8: The variation of the minimum principal stress σ_3

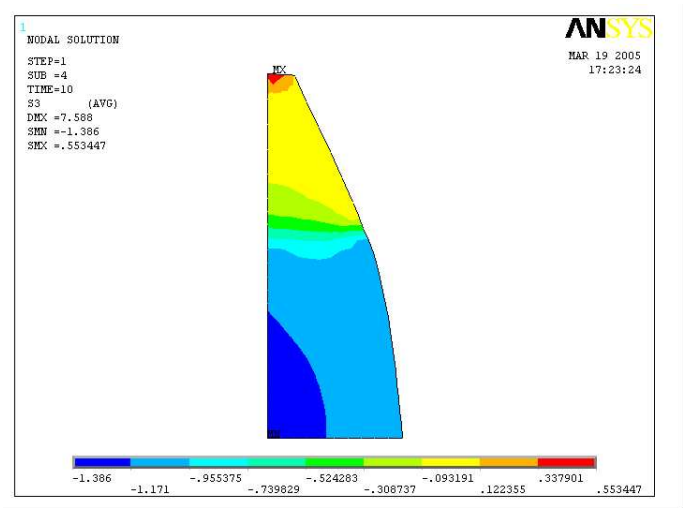


Figure 9: The variation of the minimum principal stress σ_3 under the finer mesh of Figure 4

than on the muscle boundary.

8. DISCUSSION - LIMITATIONS

The simple tension experiments, conducted on muscle tissue, were performed 1 - 4 days after the animal sacrifice. It is however known[13] that refregiration causes changes to the tissue mechanical properties. It is therefore possible, that the mechanical response concluded from these experiments, is stiffer than the one that would have occurred, if the experiments had been done immediately after the animal sacrifice.

The finite element model that we have constructed for the simulation of the muscle - tendon compound under combined longitudinal tension and transverse compression, is a simple, ideal, axisymmetric one. No interaction between the muscle and the surrounding tissues (e. g. bones) has been considered. However, axisymmetry may interprete in a crude way the action of a distributed load, along part of the muscle belly and the reaction of the underlying bone. A more realistic picture of the real geometry of such a model would have been obtained via computer tomography (CT).

The lateral compressive load that we have applied on the muscle is a static one, while in real situations of muscle injuries caused by car crashes, such loads are dynamic impact loads. Hence an impact analysis on a

nonlinear elastic material, would be more appropriate in that case.

REFERENCES

[1] Grieve, A.P., Armstrong, C.G. (1988), “Compressive properties of soft tissues”, in *Biomechanics XI-A, International Series on Biomechanics*, pp. 333-339.
[2] Bosboom, E.M.H., Hesselink, M.K.C., Oomens, C.W.J., Bouten, C.V.C., Drost, M.R. and Baaijens F.P.T. (2001), “Passive transverse mechanical properties of skeletal muscle under in vivo compression”, Journal of Biomechanics 34, pp. 1365-1368.
[3] Ogden, R.W. (1982), “Elastic deformations in rubber like solids”, in *Mechanics of Solids, The Rodney Hill 60th Anniversary Volume*, pp. 499-537.
[4] Maggana C., Kefalas V. and Eftaxiopoulos D. A. (2004), “In vitro mechanical properties of porcine skeletal muscle”, *Proceedings of the 7th National Congress on Mechanics*, Chania, Crete, June 24 - 26, Vol II, pp. 246- 251.
[5] Linder - Ganz, E. and Gefen A. (2004), “Mechanical compression-induced pressure sores in rat hindlimb: muscle stiffness, histology, and computational models”, Journal of Applied Physiology 96, pp. 2034-2049.
[6] Oomens C.W.J., Bressers O.F.J.T., Bosboom E.M.H. and Bouten C.V.C. (2001), “Deformation analysis of a supported buttock contact”, *ASME Summer Bioengineering Conference*, Snowbird, Utah, pp. 853-854.
[7] Todd B.A. and Thacker J.G. (1994), “Three-dimensional computer model of the human buttocks, in vivo”, J. Rehabil.Res. Dev. 31, pp. 111-119.
[8] Breuls, R.G.M, Bouten, C.V.C., Oomens, C.W.J., Bader, D.L. and Baaijens, F.T.P (2003), “A theoretical analysis of damage evolution in skeletal muscle tissue with reference to pressure ulcer development”, Journal of Biomechanical Engineering - Transaction of the ASME 125, pp. 902-909.
[9] Spears, I.R., Cummins, N.K., Brenchley, Z., Donohue, C., Turnbull, C., Burton, S., Mach, G.A. (2003), “The effect of saddle design on stresses in the perineum during cycling”, Medicine and Science in Sports and Exercise 35, pp. 1620-1625.
[10] Hawkins, D. and Bey, M. (1997), “Muscle and tendon force - length properties and their interaction in vivo”, Journal of Biomechanics 30, pp. 63-70.
[11] Myers, B.S., Woolley, C.T., Slotter, T.L., Garrett, W.E. (1998), “The influence of strain rate on the passive and stimulated engineering stress - large strain behavior of the rabbit tibialis anterior muscle”, Journal of Biomechanical Engineering - Transactions of the ASME 120, pp. 126-132.
[12] Maganaris, C.N. and Paul, J.P. (1999), “In vivo human tendon mechanical properties”, Journal of Physiology, 521.1, pp. 307-313.
[13] C. K. Yannakopoulos, private communication (2004).

STRESS ANALYSIS IN ABDOMINAL AORTIC ANEURYSMS APPLYING FLOW INDUCED WALL PRESSURE

Yannis Papaharilaou [*,♣], John A. Ekaterinaris[*], Eirini Manousaki[♣], Asterios N. Katsamouris[♣]

[*]Institute of Applied and Computational Mathematics
Foundation for Research and Technology- Hellas
P.O.BOX 1527, 71110 Heraklion, Crete, Greece
e-mail: yannisp@iacm.forth.gr, web page: http://www.iacm.forth.gr

[♣] Dept. of Vascular Surgery, University of Crete, Greece

Keywords: Biofluid mechanics, aneurysm rupture, computational hemodynamics, structural stress analysis.

Abstract. *Abdominal aortic aneurysm (AAA) is a localized dilatation of the aortic wall. The lack of an accurate AAA rupture risk index remains an important problem in the clinical management of the disease. To accurately estimate AAA rupture risk, detailed information on patient specific wall stress distribution and aortic wall tissue yield stress is required. A complete fluid structure interaction (FSI) study of the wall forces is impractical and thus of limited clinical value. On the other hand, isolated static structural stress analysis based on a uniform wall loading is a widely used approach for AAA rupture risk estimation that however neglects flow induced wall stress variation. Aim of this study was to assess the merit of a decoupled fluid structure analysis of AAA wall stress. Anatomically correct patient specific AAA wall models were created by 3D reconstruction of computed tomography (CT) images. Flow simulations were carried out with inflow and outflow boundary conditions obtained from patient extracted data. Static structural stress analysis was performed applying a uniform pressure wall loading and a flow induced non-uniform pressure loading. In the structural analysis a hyperelastic arterial wall model and an elastic intraluminal thrombus (ILT) model were applied. Our results show that the decoupled fluid structural analysis approach yields a more realistic AAA wall stress distribution than the isolated structural stress analysis approach thus providing a practical alternative to the more complete but computationally intensive FSI study.*

1 INTRODUCTION

Abdominal aortic aneurysm is a localized dilatation of the aortic wall. The physiological processes associated with AAA development and progression are not as yet fully understood. This pathologic condition has been found to affect 8.8 % of the population over the age of 65 [1] and if left untreated it may lead to rupture. The size of the aneurysm and its rate of expansion are parameters associated with the risk of rupture. For aneurysms with a maximum transverse diameter below 4 cm the risk of rupture is very small (but not absent). However, when the aneurysm transverse diameter is between 4 and 5 cm the risk of rupture is 0.5 % and between 5 and 6 cm it becomes 5 % rising exponentially with diameter increase [2-4]. The decision for surgical intervention for patients with AAA's is complicated by the lack of a sufficiently accurate rupture risk index. A widely used such index, based on the results from a number of clinical studies [5-7], is the maximum transverse diameter. In cases where this diameter exceeds 5-6 cm, surgical or endovascular treatment is advised. However 'small' (<5 cm) diameter aneurysms, where 'watchful waiting' requiring frequent observation is preferred to surgery, are known to rupture [8-10]. Therefore, the decision for surgical intervention, associated with a mortality rate of 4-5 % [11], should not be based exclusively on the maximum transverse diameter and a new more reliable rupture risk index should be introduced.

Recent attempts to establish a reliable AAA rupture risk index were based on the evaluation of the arterial wall stress distribution. Finite element analysis (FEA) has been used to compute the stress distribution in both simplified [12,13] and anatomically correct [14,15] AAA models. The hemodynamics of the AAA have been extensively investigated experimentally [16,17] and computationally in both idealized and anatomically correct models in steady and time varying flow [18,19]. The coupling of fluid and structure has also been studied in AAA models [20]. The role of intraluminal thrombus (ILT) on AAA wall stress still remains uncertain. Some studies support the hypothesis that the ILT introduces a cushioning effect in the transmission of the flow induced stresses to the wall that reduces the peak wall stress [21-23], while others suggest that thrombus has no effect on the progression of an AAA [24,25]. Various simplifications have been introduced to the models used in these studies with respect to the shape of the aneurysm, the inclusion and elastic properties of ILT, the thickness and elastic properties of the wall, the role of surrounding structures and, the presence of residual stresses on the

AAA wall.

The stress distribution on the aneurysmal wall is determined by the complex intra-aneurysmal hemodynamics resulting from the geometric configuration of the ILT modulated flow conduit and the effects of surrounding tissue. To date the maximum transverse dimension of the AAA is being used routinely in clinical practice as an estimate of rupture risk. However, the use of this parameter alone has lead in many cases to the underestimation of rupture risk in 'small' (<5 cm) diameter aneurysms and overestimation of the risk of rupture in 'large' (>6 cm) diameter aneurysms thus compromising the quality of patient management.

The lack of an accurate AAA rupture risk index remains an important problem in the clinical management of the disease. Accurate estimation of the patient specific AAA rupture risk requires detailed information on both the wall stress distribution and the aortic wall tissue yield stress. However, the AAA wall properties and the stress distribution cannot be measured or even derived with sufficient accuracy from non-invasive measurements *in vivo*. As an alternative, numerical approximations of the flow and wall motion equations are sought using wall constitutive models based on mean elastic properties obtained by *in vitro* mechanical testing of excised specimens of the aneurismal wall. Rhagavan et al. [26] proposed a two parameter, hyperelastic, isotropic, incompressible material model for the AAA wall utilizing uniaxial loading stress strain measurements on excised AAA specimens. The ILT solid structure has been modeled as either an elastic [27] or hyperelastic [28], isotropic, incompressible material.

The wall stress computation should ideally result from a complete FSI simulation of the wall forces. However, this approach still suffers form modeling assumptions, is very intensive computationally and thus currently impractical. Furthermore, it has been shown in idealized AAA models that the FSI approach yields peak wall stress estimates similar to those obtained by an *isolated* structural stress analysis [29]. The computational approach most widely used to estimate peak AAA wall stress is the isolated static structural analysis with a uniform peak systolic pressure wall loading. However, this approach neglects the flow induced pressure distribution on the AAA wall. Aim of this study was to assess the merit of a decoupled fluid structure approach for AAA wall stress estimation as compared to the isolated static structural stress analysis approach. Towards this end, the stress distribution computed for a uniform wall loading in an anatomically correct AAA model is compared to the stress computed for the same model but for the flow induced pressure wall loading.

2 METHODS

A 86 year old male with an intact 10 cm peak transverse diameter AAA was the subject selected for this study. The selected aneurysm geomerty exhibits significant tortouosity of the inflow conduit and the proximal segments of the iliac arteries that is expected to strongly affect the intraaneurysmal flow field. This geometric configuration is typical of large AAA's and can be attributed to the asymmetric expansion of the aneurysm sac caused primarily by the expansion constraints introduced by the proximity of the spinal column. Information on the 3D AAA geometric configuration was extracted *in vivo* by contrast enhanced high resolution spiral CT angiography. The following CT acquisition parameters were prescribed: 120 kVp, 160 mAs effective current level, 10.4 s scan time, 22.1 mm feed/rotation ratio, 380 mm in plane FOV, 2 mm slice thickness, 1.5 mm reconstruction spacing/increment, 0.5 mm slice overlap and a 512 x 512 image matrix size resulting in a 0.742 mm in plane resolution. Angiography was triggered at 120 Houndsfield units.

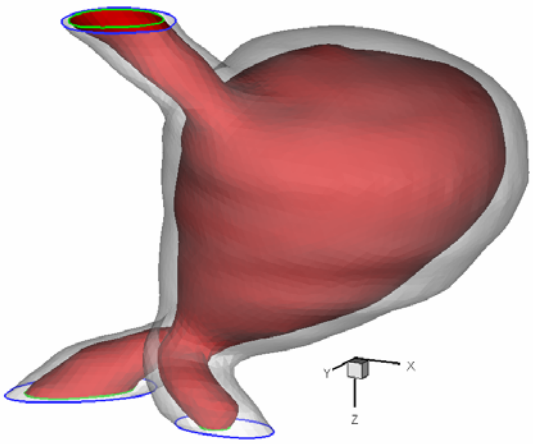


Figure 1. Internal (red) and external (grey-translucent) 3D reconstructed smoothed surfaces of the CT image extracted AAA geometry.

Segmentation and 3D surface reconstruction of the CT images was implemented using in house developed

software [30]. From the segmented CT images two 3D surfaces were reconstructed: the true vessel lumen surface and the external aortic wall surface (Figure 1). A third surface, the internal aortic wall was generated as an iso-surface 2 mm inwards from the external aortic wall. The third surface, the interface between the AAA wall endothelium and the ILT, could not be extracted from the CT images as the imaging method still lacks the level of spatial resolution and contrast required. The location of this third surface relative to the external aortic wall surface effectively determines the thickness of the AAA wall. Abnormal, small scale surface irregularities introduced during the imaging and reconstruction processes were excluded from the computational model by applying pixel width constrained smoothing of the reconstructed surfaces prior to mesh generation.

2.1 Flow Computation

The computational grid generated using Gambit had 333150 tetrahedral elements and non-uniform grid node spacing to produce higher grid density at the proximal and distal aneurysm neck regions as compared to the bulge surface. A short, native vessel blended straight tube like extension of the proximal inflow was added to the model to create a circular cross section inlet required for the application of the exact Womersley solution as the time dependent inflow boundary condition. The Navier-Stokes and continuity equations for incompressible flow neglecting body forces are expressed in vector form as:

$$\nabla \cdot \mathbf{u} = 0$$
$$\frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho}\nabla p + \nu \nabla^2 \mathbf{u} \tag{1}$$

where, $D/Dt = \partial/\partial t + \mathbf{u}\cdot\nabla$ is the substantial derivative, ρ is the fluid density, and v is the fluid kinematic viscosity.

Fluent 6.1.22 was used to solve the flow equations. For the flow field computations, the arterial wall was assumed rigid and blood was modeled as an incompressible, Newtonian fluid with a density of 1.05 gr/cm^3 and a viscosity of 4.5 cP. Blood is a suspension of red and white cells, platelets, proteins and other elements in plasma and exhibits an anomalous non Newtonian viscous behavior when exposed to low shear rates or flows in tubes of less than 1mm in diameter. However, the Newtonian fluid assumption does not affect the major flow features and is considered an acceptable approximation for modeling blood flow in the macrocirculation [31]. The AAA inflow waveform and the aortic flow split ratio in the iliac arteries were measured *in vivo* by Doppler US two hours after CT scanning of the patient. On average, the left iliac artery received 40 % of the aortic flow and the right iliac artery received 60 %. The discrete Fourier series of the measured AAA inflow waveform can be expressed as:

$$Q(t) = Q_0 + \sum_{n=1}^{N} Q_n e^{in\omega t} \tag{2}$$

where Q_0 is the steady flow component, N=16 represents the number of Fourier modes used and ω is the fundamental frequency of the measured flow waveform. From the discrete Fourier series of the volume flow rate in Eq. (2) the fully developed time varying velocity profile was computed using an expression obtained following Womersley's derivation [32]:

$$u(r,t) = \frac{2Q_0}{A}\left(1-\frac{r^2}{R^2}\right) + \sum_{n=1}^{N}\frac{Q_n}{A}\left\{\left(1-\frac{J_0(\alpha_n i^{3/2} r/R)}{J_0(\alpha_n i^{3/2})}\right)/\left(1-\frac{2J_1(\alpha_n i^{3/2})}{\alpha_n i^{3/2} J_0(\alpha_n i^{3/2})}\right)\right\} e^{in\omega t} \tag{3}$$

where J_0 and J_1 are the Bessel functions of the first kind of order zero and one respectively, A is the cross sectional area and R the inlet radius of the straight tube extension inlet and, $\alpha_n = R\sqrt{n\omega/\nu}$ is the Womersley parameter. The time averaged mean Reynolds number of the prescribed waveform was Re_m=355 and the Womersley parameter for the fundamental frequency of the measured flow waveform was α_1=16.7. The velocity profile given by Eq. (3) was applied as the time dependent inflow boundary condition. A time step size of 6×10^{-4} s was used and 10^4 time steps were required to complete one flow cycle. A time periodic solution was achieved after 7 flow cycles. A second order upwind discretization scheme was applied for the momentum equation and the SIMPLE scheme was used for pressure velocity coupling.

2.2 Finite element stress analysis

ABAQUS 6.4.1 was used to solve the momentum equations, the wall constitutive equations and the conditions of equilibrium for the static structural stress analysis. The aortic wall was modeled as an incompressible, homogenous, isotropic, hyperelastic, material with a uniform thickness of 2 mm. The finite strain constitutive model proposed by Rhagavan et al. [26] was adopted for the arterial wall with a strain energy density function given by

$$W = \alpha(I_B - 3) + \beta(I_B - 3)^2 \tag{4}$$

where, I_B is the first invariant of the left Cauchy-Green tensor **B** (I_B=tr **B**). The model parameters were set to α= 17.4 N cm^{-2} and β =188.1 N cm^{-2} that correspond to population mean values obtained from uniaxial loading tests on excised AAA wall specimens.

The ILT was modeled as an incompressible, isotropic, homogenous, linear elastic material with a Young modulus E = 0.11 MPa and a Poisson ratio ν = 0.45. These values of E and ν represent population mean values obtained from uniaxial loading tests performed on ILT specimens harvested during AAA surgery by Di Martino et al. [27]. The AAA model assembly included the ILT solid part with 41291 tetrahedral elements and the arterial wall shell part with 9690 triangular elements. Stress analysis results were obtained both for a uniform wall loading using the peak systolic arterial pressure (16 kPa or 120 mmHg) and for the non-uniform flow induced wall pressure loading computed during early systolic deceleration. Mapping of the pressure field from the finer numerical grid used for the flow computations to the coarser grid used in the stress analysis was achieved by inverse-distance interpolation. A non-slip condition was applied at the AAA wall - ILT interface. The proximal and distal ends of the model were constrained longitudinally.

3 RESULTS

The results of the time dependent flow field computation showed that most of the AAA lumen wall surface was exposed to very low wall shear stress (WSS) throughout the cardiac cycle. Regions of locally elevated WSS were located near the proximal and distal neck of the AAA bulge (Figure 2a). WSS magnitude was normalized by the straight pipe inlet Poiseuille WSS. The computed wall pressure distribution during early systolic deceleration (Figure 2b) showed a significant deviation from the peak arterial systolic pressure which has been widely used as a uniform wall loading condition in static structural stress analyses. During early systolic acceleration regions in the vicinity of the distal neck of the aneurysmal wall were exposed to a pressure loading 18 % higher than the peak systolic pressure. Furthermore, most of the aneurysm bulge wall was exposed to pressures 10 % higher than the peak systolic pressure.

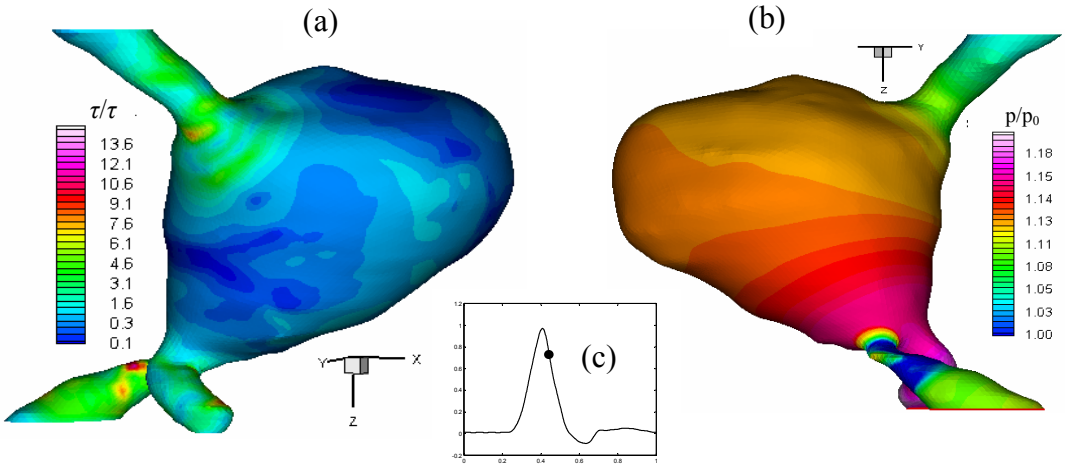


Figure 2. Computed WSS magnitude (a) and static pressure on the wall of the true lumen (b) during early systolic deceleration. WSS is normalized by the inlet equidiameter straight pipe WSS and static pressure is normalized by the systolic arterial pressure. The aortic flow waveform applied is also shown (c).

The highly complex flow field that develops in the aneurysmal sac is depicted in Figure 3 by means of stream ribbons color mapped with static pressure (a) and vorticity magnitude (b). The out of plane curvature (tortuosity) of the aortic flow conduit injecting blood into the aneurysmal expansion strongly influences the velocity distribution at the aneurismal bulge inlet. The vorticity colour mapped stream ribbon graph (Figure 3b)

clearly depicts the increase in vorticity occurring within the tortuous aortic segment leading to the aneurysmal bulge inlet.

Arterial wall stress distributions for uniform wall loading and flow induced non-uniform pressure are presented using the Von-Mises stress, a scalar measure of the stress tensor that is proportional to the strain energy density at each point expressed as

$$\sigma_{VM} = \sqrt{\frac{1}{2}\left[(\sigma_1 - \sigma_2)^2 + (\sigma_1 - \sigma_3)^2 + (\sigma_2 - \sigma_3)^2\right]} \tag{5}$$

where σ_1, σ_2, σ_3 are the principal stresses. Application of the non-uniform flow induced wall loading to the AAA model produced a 12 % increase in the computed peak wall stress as compared to the uniform wall loading result (Figures 4 and 5). Two regions of high stress were found, one located anteriorly in the distal half of the AAA bulge with a local peak of 52 N/cm^2 (Figure 4 arrow) and the other located at the proximal neck anteriorly and to the left with a local peak of 54 N/cm^2 (Figure 5 arrow). It should be noted that only the magnitude and not the locations of the peak stress regions was altered by the introduction of the non-uniform flow induced pressure wall loading.

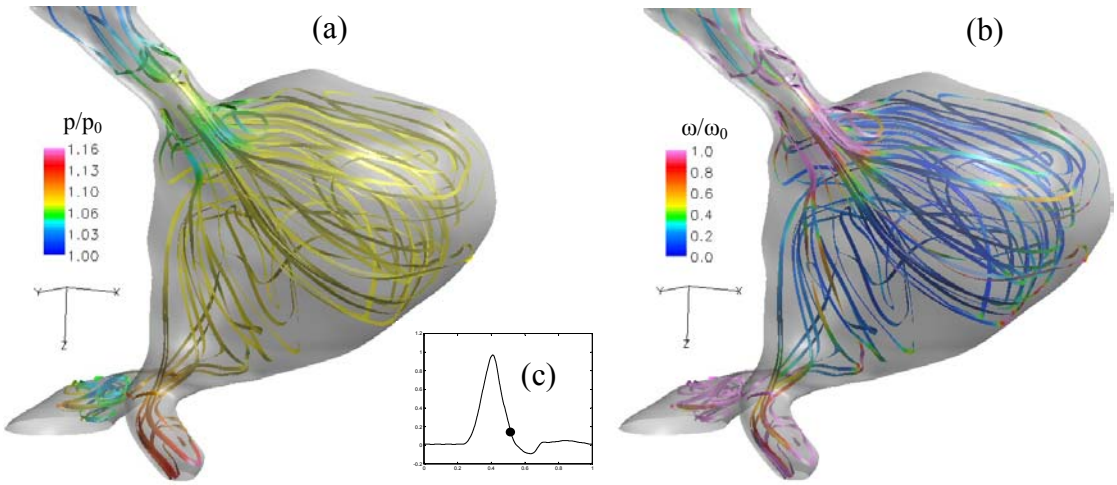


Figure 3. Stream ribbons of computed flow field during end systolic deceleration. Static pressure (a) and vorticity magnitude (b) color mapping is applied. The aortic flow waveform applied is also shown (c).

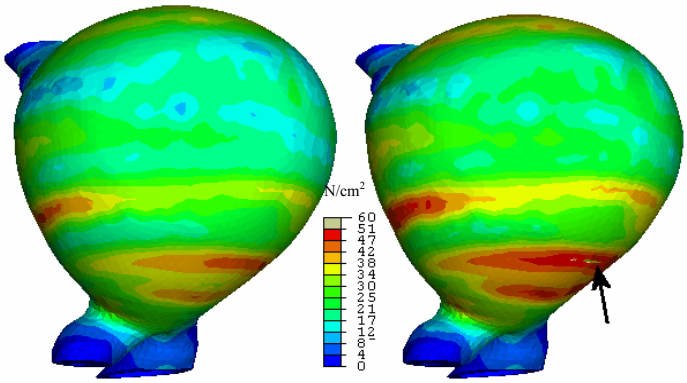


Figure 4. Computed Von Mises stress distribution on the AAA wall for uniform peak systolic pressure loading (left) and flow induced non-uniform wall loading (right). Arrow shows local maximum of wall stress.

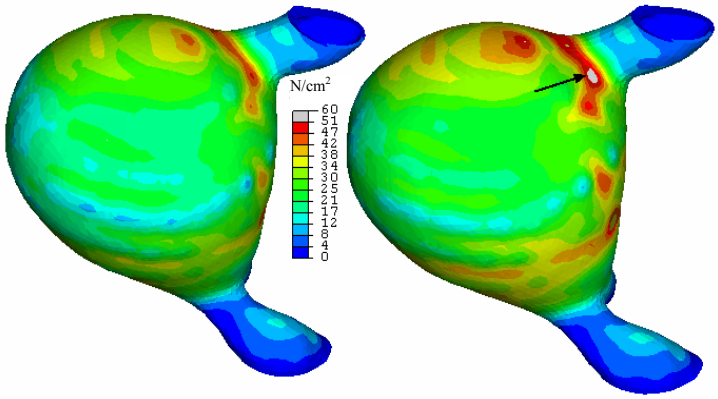


Figure 5. Computed Von Mises stress distribution on the AAA wall for uniform peak systolic pressure loading (left) and flow-induced non-uniform wall loading (right). Arrow shows local maximum of wall stress.

4 DISCUSSION

The AAA selected for this study had a peak transverse dimension of approximately 10 cm and did not rupture prior to surgery although being almost twice the size above which surgical intervention is commonly advised. This further supports the argument that the peak transverse dimension is not an absolutely reliable AAA rupture risk indicator. Model studies have shown that the law of Laplace that relates internal diameter and wall stress is not appropriate for estimating the stress field even in simplified AAA geometries. Elger et al. [12] found in models that the wall stress distribution is most strongly influenced by the shape of the aneurysm with peak stress correlated to wall curvature. This finding is in agreement with our results (see Figs. 4 and 5).

The AAA model used in the present study includes a number of simplifications and underlying assumptions. A uniform wall thickness was specified due to the inherent limitations in the imaging technique. This affects the computed stress distribution thus increasing the uncertainty of the results as compared to the exact *in vivo* conditions. However, it does not reduce the value of the comparative results presented since it will have a similar effect in the stress distribution on both computational approaches considered. The hemodynamic pressure field was computed assuming a rigid wall model. However, it has been shown both experimentally [33] and computationally [34] that the introduction of wall compliance to arterial models only has a quantitative effect on the computed wall stresses whereas the main flow features are preserved.

The computational mesh used for the structural stress analysis was based on the geometry reconstructed from the CT images obtained throughout the cardiac cycle and over multiple cycles as the acquisition was not gated to the cardiac rhythm of the subject. As a result, the mean geometric representation of the pressure pulse modulated AAA structure is reconstructed although a zero-stress state is assumed in the computation. As the zero-stress state of the AAA cannot be measured *in-vivo* one could assume that as the diastolic phase occupies most of the abdominal aorta flow cycle the reconstructed AAA geometry is an approximate representation of the diastolic pressure modulated AAA stress state. This residual stress has been neglected in this investigation although its effects are not expected to invalidate the results of this comparative study. The effects of neglecting the residual stress, which is assessed that may be important, on the computed stress distribution will be addressed in a future study.

The material properties used in this study where based on mean values and therefore the computed stress distribution is not expected to represent the exact *in vivo* wall loading conditions. It should be noted however that the aforementioned difficulties in constructing a mathematical model to simulate *in vivo* AAA wall loading conditions also apply to a further extent to the FSI approach, which is further complicated by the dynamic effects of wall motion. It is therefore very important to reduce the solution uncertainties identified in the proposed decoupled fluid structure model before introducing wall motion dynamics in a coupled fluid structure model.

Our results show that although the isolated static structural stress analysis approach captures the gross features of the stress distribution it underestimates the magnitude of the peak wall stress by as much as 12 % compared to the proposed decoupled fluid structure approach. This value may be different when other AAA cases are considered depending on the aneurysm shape and inflow conditions. However, the intra-aneurysmal flow-induced wall pressure distribution is primarily influenced by the temporal acceleration and deceleration of the flow and to a lesser extent by the size of the aneurysmal sac. Consequently, as our stress computations were based on a typical physiological AAA inflow waveform, the 12 % difference in the computed peak wall stress

should be considered as a representative result. Furthermore, the decoupled fluid structure approach yields the local AAA hemodynamic conditions thus allowing for the identification of wall regions exposed to low and oscillatory wall shear stress and high shear gradients, conditions that have been linked to the development of wall lesions. This information may then be used to further support prognosis of AAA rupture risk.

In order to establish a more reliable patient specific index of AAA rupture risk it is necessary to further improve the accuracy of the computational models used. This requires imposing realistic boundary conditions extracted from the patient *in vivo* to a computational model that couples fluid and solid dynamics. This study shows that a decoupled fluid structure approach is a practical alternative to the more complete but computationally intensive FSI study.

REFERENCES

[1] Newman, A.B., Arnold, A.M., Burke, G.L., O'Leary, D.H., and Manolio, T.A. (2001), "Cardiovascular disease and mortality in older adults with small abdominal aortic aneurysms detected by ultrasonography: the cardiovascular health study", Ann Intern Med, 134, pp. 182-90.
[2] Law, M.R., Morris, J., and Wald, N.J. (1994), "Screening for abdominal aortic aneurysms", J Med Screen, 1, pp. 110-5; discussion 115-6.
[3] Sharma, U., Ghai, S., Paul, S.B., Gulati, M.S., Bahl, V.K., Rajani, M., and Mukhopadhyay, S. (2003), "Helical CT evaluation of aortic aneurysms and dissection: a pictorial essay", Clin Imaging, 27, pp. 273-80.
[4] Wolf, Y.G., and Bernstein, E.F. (1994), "A current perspective on the natural history of abdominal aortic aneurysms", Cardiovasc Surg, 2, pp. 16-22.
[5] Galland, R.B., Whiteley, M.S., and Magee, T.R. (1998), "The fate of patients undergoing surveillance of small abdominal aortic aneurysms", Eur J Vasc Endovasc Surg, 16, pp. 104-9.
[6] Participants, T.U.S.A.T. (1998), "Mortality results for randomised controlled trial of early elective surgery or ultrasonographic surveillance for small abdominal aortic aneurysms. " Lancet, 352, pp. 1649-55.
[7] Scott, R.A., Ashton, H.A., and Lamparelli, M.J. (1999), "Vascular surgical society of great britain and ireland: fifteen years of experience using 6 cm as a criterion for abdominal aortic aneurysm surgery", Br J Surg, 86, pp. 709-10.
[8] Cronenwett, J.L., Murphy, T.F., Zelenock, G.B., Whitehouse, W.M., Jr., Lindenauer, S.M., Graham, L.M., Quint, L.E., Silver, T.M., and Stanley, J.C. (1985), "Actuarial analysis of variables associated with rupture of small abdominal aortic aneurysms", Surgery, 98, pp. 472-83.
[9] Darling, R.C., Messina, C.R., Brewster, D.C., and Ottinger, L.W. (1977), "Autopsy study of unoperated abdominal aortic aneurysms. The case for early resection", Circulation, 56, pp. II161-4.
[10]Nicholls, S.C., Gardner, J.B., Meissner, M.H., and Johansen, H.K. (1998), "Rupture in small abdominal aortic aneurysms", J Vasc Surg, 28, pp. 884-8.
[11]Katz, D.A., Littenberg, B., and Cronenwett, J.L. (1992), "Management of small abdominal aortic aneurysms. Early surgery vs watchful waiting", Jama, 268, pp. 2678-86.
[12]Elger, D.F., Blackketter, D.M., Budwig, R.S., and Johansen, K.H. (1996), "The influence of shape on the stresses in model abdominal aortic aneurysms", J Biomech Eng, 118, pp. 326-32.
[13]Mower, W.R., Baraff, L.J., and Sneyd, J. (1993), "Stress distributions in vascular aneurysms: factors affecting risk of aneurysm rupture", J Surg Res, 55, pp. 155-61.
[14]Raghavan, M.L., Vorp, D.A., Federle, M.P., Makaroun, M.S., and Webster, M.W. (2000), "Wall stress distribution on three-dimensionally reconstructed models of human abdominal aortic aneurysm", J Vasc Surg, 31, pp. 760-9.
[15]Vorp, D.A., Wang, D.H., Webster, M.W., and Federspiel, W.J. (1998), "Effect of intraluminal thrombus thickness and bulge diameter on the oxygen diffusion in abdominal aortic aneurysm", J Biomech Eng, 120, pp. 579-83.
[16]Egelhoff, C.J., Budwig, R.S., Elger, D.F., Khraishi, T.A., and Johansen, K.H. (1999), "Model studies of the flow in abdominal aortic aneurysms during resting and exercise conditions", J Biomech, 32, pp. 1319-29.
[17]Yu, S.C.M. (2000), "Steady and pulsatile flow studies in Abdominal Aortic Aneurysm models using Particle Image Velocimetry", International Journal of Heat and Fluid Flow, 21, pp. 74-83.
[18]Budwig, R., Elger, D., Hooper, H., and Slippy, J. (1993), "Steady flow in abdominal aortic aneurysm models", J Biomech Eng, 115, pp. 418-23.
[19]Taylor, T.W., and Yamaguchi, T. (1994), "Three-dimensional simulation of blood flow in an abdominal aortic aneurysm--steady and unsteady flow cases", J Biomech Eng, 116, pp. 89-97.
[20]Di Martino, E.S., Guadagni, G., Fumero A., Ballerini G., Spirito, and R., B.P., Redaelli, A. (2001), "Fluid-structure interaction within realistic three-dimensional models of the aneurysmatic aorta as a guidance to assess the risk of rapture of the aneurysm", Medical Engineering and Physics, 23, pp. 647-655.
[21]Mower, W.R., Quinones, W.J., and Gambhir, S.S. (1997), "Effect of intraluminal thrombus on abdominal

aortic aneurysm wall stress", J Vasc Surg, 26, pp. 602-8.
[22]Vorp, D.A., Mandarino, W.A., Webster, M.W., and Gorcsan, J., 3rd. (1996), "Potential influence of intraluminal thrombus on abdominal aortic aneurysm as assessed by a new non-invasive method", Cardiovasc Surg, 4, pp. 732-9.
[23]Wang, D.H., Makaroun, M.S., Webster, M.W., and Vorp, D.A. (2002), "Effect of intraluminal thrombus on wall stress in patient-specific models of abdominal aortic aneurysm", J Vasc Surg, 36, pp. 598-604.
[24]Dobrin, P.B. (1989), "Pathophysiology and pathogenesis of aortic aneurysms. Current concepts", Surg Clin North Am, 69, pp. 687-703.
[25]Schurink, G.W., van Baalen, J.M., Visser, M.J., and van Bockel, J.H. (2000), "Thrombus within an aortic aneurysm does not reduce pressure on the aneurysmal wall", J Vasc Surg, 31, pp. 501-6.
[26]Raghavan, M.L., and Vorp, D.A. (2000), "Toward a biomechanical tool to evaluate rupture potential of abdominal aortic aneurysm: identification of a finite strain constitutive model and evaluation of its applicability", J Biomech, 33, pp. 475-82.
[27]Di Martino, E., Mantero, S., Inzoli, F., Melissano, G., Astore, D., Chiesa, R., and Fumero, R. (1998), "Biomechanics of abdominal aortic aneurysm in the presence of endoluminal thrombus: experimental characterisation and structural static computational analysis", Eur J Vasc Endovasc Surg, 15, pp. 290-9.
[28]Wang, D.H., Makaroun, M., Webster, M.W., and Vorp, D.A. (2001), "Mechanical properties and microstructure of intraluminal thrombus from abdominal aortic aneurysm", J Biomech Eng, 123, pp. 536-9.
[29]Finol, E.A., Di Martino E.S., Vorp, D.A., Amon, C.H. (2003), "Fluid-structure interaction and structural analyses of an aneurysm model", *Proceedings of the Summer Bioengineering Conference*, Key Biscane, Florida, pp. 75-76.
[30]Giordana, S., Sherwin, S.J., Peiro, J., Doorly, D.J., Papaharilaou, Y., Caro, C.G., Watkins, N., Cheshire, N., Jackson, M., Bicknall, C., and Zervas, V. (2005), "Automated classification of peripheral distal by-pass geometries reconstructed from medical data", J Biomech, 38, pp. 47-62.
[31]Perktold, K., Resch, M., and Florian, H. (1991), "Pulsatile non-Newtonian flow characteristics in a three-dimensional human carotid bifurcation model", J Biomech Eng, 113, pp. 464-75.
[32]Womersley, J.R. (1955), "Method for the calculation of velocity, rate of flow and viscous drag in arteries when the pressure gradient is known", J Physiol, 127, pp. 553-63.
[33]Duncan, D.D., Bargeron, C.B., Borchardt, S.E., Deters, O.J., Gearhart, S.A., Mark, F.F., and Friedman, M.H. (1990), "The effect of compliance on wall shear in casts of a human aortic bifurcation", J Biomech Eng, 112, pp. 183-8.
[34]Perktold, K., and Rappitsch, G. (1995), "Computer simulation of local blood flow and vessel mechanics in a compliant carotid artery bifurcation model", J Biomech, 28, pp. 845-56.

ANALYSIS OF TOOTH MOVEMENT USING A BOUNDARY-TYPE SOLUTION OF REYNOLDS' EQUATIONS

Christopher G. Provatidis

School of Mechanical Engineering
National Technical University of Athens
9, Iroon Polytechniou Str, Zografou Campus, 15773 Athens, Greece
e-mail: cprovat@central.ntua.gr, web page: http://users.ntua.gr/cprovat

Keywords: Biomechanics, Hydrodynamics, Tooth movement, Centre of Resistance, Centre of rotation.

Abstract. *Biomechanics of the periodontal ligament (PDL) around human teeth is still an open problem. The assumption of nearly incompressible PDL leads to Reynolds' equations in curvilinear coordinates over the boundary of the tooth. This paper deals with the numerical solution of these equations around an axisymmetric rigid tooth that moves on its axial plane due to in-plane loading. A boundary-type symmetric stiffness matrix of order 3×3 is introduced, in order to correlate the applied force-system on the tooth with its three displacements: two translations and one rotation. The elements of this matrix are calculated using mid-point integration along the half axial boundary of the tooth being in contact with the periodontal ligament. The efficiency of the method is tested in two cases: an idealized conical and a nearly anatomical paraboloidal root. For both cases, the location of the centre of resistance, the centre of rotation and the distribution of the pressure are calculated.*

1 INTRODUCTION

The biomechanics of tooth movement is an important task that mainly concerns dentists and orthodontists but also computational mechanics and CAD/CAE scientists. Each tooth is surrounded by a periodontal ligament (shortly PDL), which protects the biological tissues by absorbing the applied mechanical loads during chewing or orthodontic treatment. Particularly in orthodontics, the loading is called "force-system" and it refers to both force and torque applied to the tooth bracket that is usually fixed in the middle of the crown. In this case, the aim of a computer simulation is to determine the centre of resistance as well as the centre of rotation for a certain force-system, or inversely, to determine the proper force-system that will produce the desirable centre of rotation so that the aesthetic malfunction is corrected through medical treatment. The centre of resistance refers to a point where the applied force F produces pure translation, while the centre of rotation corresponds to the instantaneous pole of pure rotation, well known from elementary kinematics. More details can be found in classical works[1-3].

In order to analyze tooth movements, two-dimensional elastic models using Airy's stress functions have been proposed[4-5]. Also, computational methods such as the finite element method have been proposed by many investigators[6-7]. The interested reader may also consult a state-of-the-art report[8]. The relevant analyses include linear elastic[6], viscoelastic[9] as well as isotropic and anisotropic models[10-14]. Besides the prediction of the above-mentioned centres of rotation and resistance, many investigators such as Middleton et al.[15] report that the calculation of the distribution of the pressure inside the periodontal ligament offers important information for the understanding of the induced biological effects such as *bone remodelling* during orthodontic treatment[16-17].

Since it is widely accepted that the PDL is a fluid or semi-fluid matrix material, for the purposes of this study we consider a simple mechanical representation of the PDL as an incompressible mediator between the tooth and the alveolar bone, i.e. acting force system and resulting pressure distribution. Thus, it is here assumed that the thin elastic PDL is governed by Reynolds equations. This hypothesis makes again timely the excellent work of Synge[18], 70 years ago, who proposed that tooth movement may be determined on the following assumptions:

- Tooth may be regarded as a rigid body, held in a rigid socket by a thin membrane that fills the space between;
- The periodontal ligament has a constant thickness;
- Periodontal ligament is elastic, homogeneous, isotropic and incompressible.

Using extensive tensor analysis, Synge[18] obtained two-dimensional and also axisymmetric Reynolds' equations in curvilinear coordinates along the interface between the tooth and the PDL. These formulas were solved analytically for a two-dimensional wedge and an axisymmetric cone. However, real teeth of complex geometry and variable thickness of the periodontal ligament cannot be solved in a closed form and a relevant 2-D computational method was recently proposed[19].

This paper revisits and extends the excellent theoretical work of Synge[16] so that it becomes now applicable to any arbitrary shaped axisymmetric tooth that is surrounded by an incompressible PDL of variable thickness. This is achieved by introducing a symmetric 3×3 stiffness matrix, **K**, with respect to three degrees of freedom: two translational (u,v) and one rotational (ω). The four independent elements k_{ij} of the stiffness matrix **K** are calculated using one-dimensional (such as mid-point) integration along the half of the open interface between the tooth boundary and the periodontal ligament in order to determine five regular line integrals (Is). Also, the pressure inside the PDL is calculated on the basis of the same line integrals, Is, plus one more.

The efficiency of the proposed method is presented for two test cases that approximate an upper central incisor: a typical conical-shaped model (where an analytical solution exists) and a parabolic-shaped model, respectively. For both cases, the centres of resistance, the centres of rotation and the distribution of pressure are calculated.

2 MECHANICAL THEORY

2.1 Description of the problem

From the engineering point of view, the problem is posed as follows. Let us assume a rigid axisymmetric body (T: tooth) inside another axisymmetric rigid body (S: socket). The space between these two bodies is occupied by an incompressible membrane of a prescribed variable thickness $h=h(s)$, which is a function of the arc length s along the tooth boundary, as shown in Figure 1.

All three bodies (tooth, socket and membrane) are assumed to be three-dimensional but only an axial section (shown in Figure 2) is required for the following analysis. A force-system $\mathbf{F}=\{F_x, F_y, M_z\}^{\mathrm{T}}$, is applied on the bracket of tooth and it undertakes a displacement $\mathbf{u}=\{u,v,w\}^{\mathrm{T}}$. Obviously, ($F_x$, F_y) denote forces while M_z a moment as shown in Figure 3. In general, we ask for:

(a) The displacements of the body (T)
(b) The related centres of resistance and rotation and
(c) The pressure distribution inside the incompressible membrane between the tooth apex O and the margin AB.

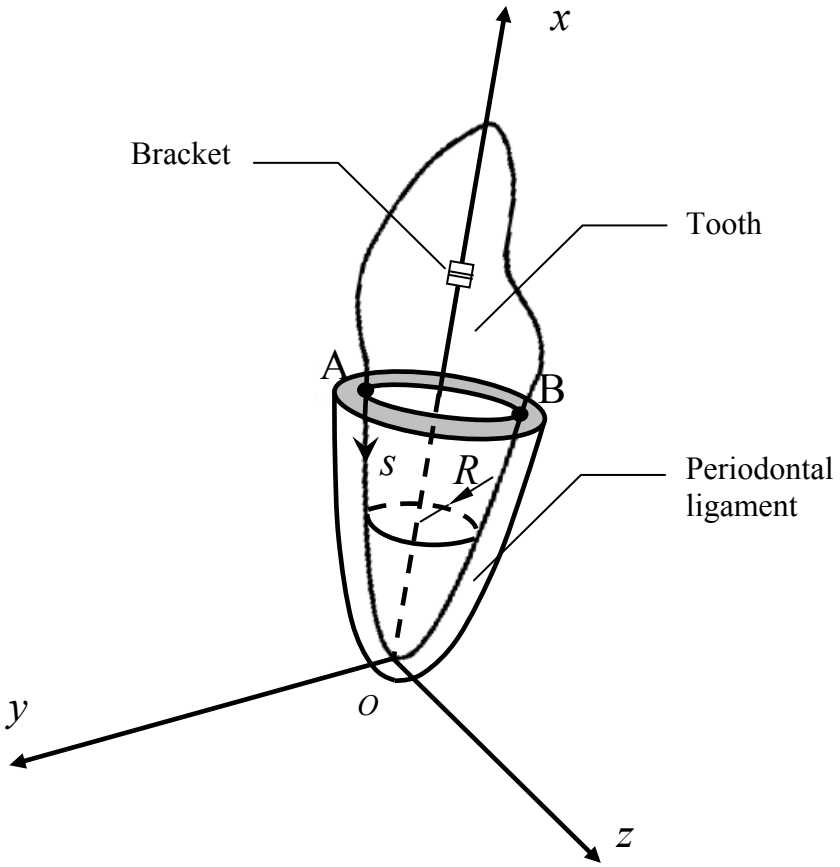


Figure 1: A three-dimensional view of an axisymmetric tooth surrounded by its periodontal ligament

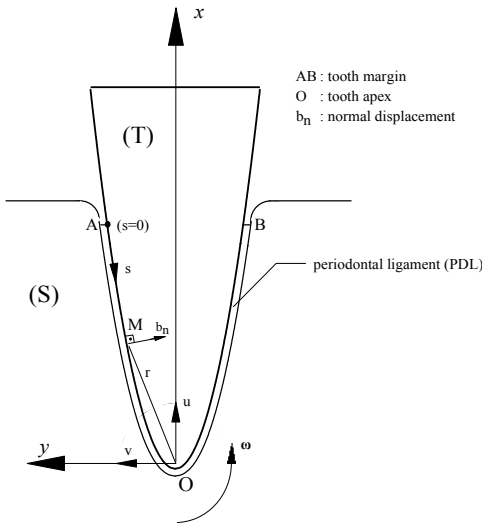


Figure 2: Geometry of the structure composed by the tooth (T), the periodontal ligament (PDL) and the socket (S). Definition of Cartesian axes and rigid-body displacements (u,v,ω).

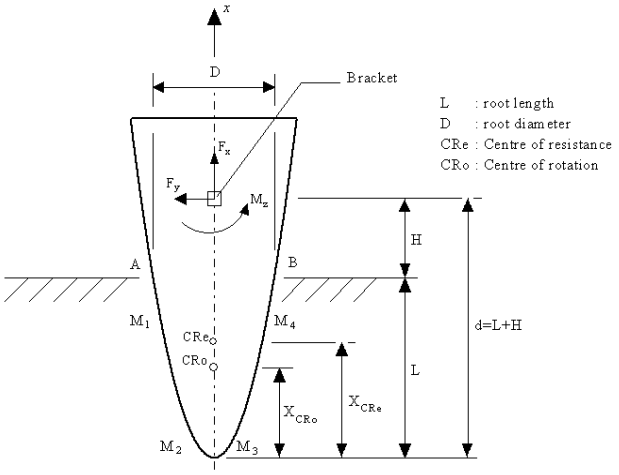


Figure 3: Definition of tooth geometry (L: root length, D: root diameter), applied force-system $\mathbf{F}=\{F_x, F_y, M_z\}^T$, centre of resistance (Cre) and centre of rotation (Cro).

2.2 Basic equations

The hydrodynamic field around a moving axisymmetric rigid tooth (T) within a rigid socket (S) is governed by[18]

$$\frac{1}{R}\frac{\partial}{\partial s}\left(Rh^3\frac{\partial p}{\partial s}\right)+\frac{h^3}{R^2}\frac{\partial^2 p}{\partial \phi^2}=12\mu\ \ b_n \tag{1}$$

with $p=p(s)$ denoting the pressure within the PDL, R the radius of the tooth, ϕ the azimuthal angle, μ the shear modulus and b_n the normal displacement component on tooth boundary. With respect to a two-dimensional coordinate system Oxy shown in Figure 2, the tooth movement is analyzed in two translations (u,v) towards x- and y-axes, as well as one rotation (ω) about the origin O that is chosen at the apex of the tooth. The displacement components of a point M are then $u-y\omega$ and $v+x\omega$, and the normal component b_n, in the sense from the socket S to the tooth T, is given by

$$b_n=-u\frac{\mathrm{d}R}{\mathrm{d}s}+\cos\phi\left[v\frac{\mathrm{d}x}{\mathrm{d}s}+\omega\left(x\frac{dx}{ds}+R\frac{\mathrm{d}R}{\mathrm{d}s}\right)\right] \tag{2}$$

The nature of the problem dictates to seek for a solution including the first harmonic in ϕ, in the form

$$p=p(s,\phi)=f(s)+g(s)\cos\phi \tag{3}$$

By substituting Eqs (2, 3) into Eq.(1) one receives

$$\frac{1}{R}\frac{\partial}{\partial s}\left[Rh^3\left(\frac{\mathrm{d}f}{\mathrm{d}s}+\frac{\mathrm{d}g}{\mathrm{d}s}\cos\phi\right)\right]+\frac{h^3}{R^2}(-g\cos\phi)=12\mu\left\{-u\frac{dR}{ds}+\cos\phi\left[v\frac{dx}{ds}+R\frac{dR}{ds}\right]\right\}$$

Then, by separating the corresponding terms, we obtain the two equations

$$\frac{1}{R}\frac{\partial}{\partial s}\left(Rh^3\frac{\mathrm{d}f}{\mathrm{d}s}\right)=-12\mu u\frac{dR}{ds} \tag{4}$$

$$\frac{1}{R}\frac{\partial}{\partial s}\left(Rh^3\frac{\mathrm{d}g}{\mathrm{d}s}\right)-\frac{h^3 g}{R^2}=12\mu v\frac{dx}{ds}+12\mu\omega\left(v\frac{dx}{ds}+R\frac{dR}{ds}\right) \tag{5}$$

The boundary conditions for the functions *f* and *g* are

$$f(0)=P_{atm},\ g(0)=0,\ g(l)=0 \tag{6}$$

the last due to the fact that *p* must be single valued at the apex (*s*=*l*).

By integrating Eq.(4) one receives

$$Rh^3\frac{\mathrm{d}f}{\mathrm{d}s}=-6\mu u R^2+C \tag{7}$$

and after a second integration, the function *f* is given by

$$f=-6\mu u\int_0^s R\,h^{-3}\,ds+C\int_0^s R^{-1}h^{-3}\,ds+C' \tag{8}$$

where the C and C' are integration constants and are calculated as follows.
For *R*=0 at the apex, Eq.(7) results in *C*=0. Then, for *s*=0 Eq.(8) implies $C'=f(0)=P_{atm}$. So, *f* is given as

$$f=P_{atm}-6\mu u\int_0^s R\,h^{-3}\,ds \tag{9}$$

and due to Eq.(3), the pressure at any point becomes

$$p(s)=P_{atm}-6\mu u\int_0^s R\,h^{-3}\,ds+g(s)\cos\phi \tag{10}$$

At this point, it is assumed that the thickness (*h*) of the membrane is *constant*. Also, a new independent variable is introduced as follows

$$\zeta=\int_0^s\frac{ds}{R} \tag{11}$$

so that ζ=0 when *s*=0 and ζ=∞ when *s*=*l*. Now, the second differential equation of our problem [Eq.(5)] becomes

$$\frac{d^2g}{d\zeta^2}-g=F(\zeta) \tag{12}$$

where

$$F(\zeta)=\frac{12\mu v}{h^3}R\frac{dx}{d\zeta}+\frac{12\mu\omega}{h^3}R\left(x\frac{dx}{d\zeta}+R\frac{dR}{d\zeta}\right) \tag{13}$$

The analytical solution of Eq.(12) is given as

$$g=C_1e^{\zeta}+C_1'e^{-\zeta}+\frac{1}{2}e^{\zeta}\int_0^{\zeta}e^{-\lambda}F(\lambda)\mathrm{d}\lambda-\frac{1}{2}e^{-\zeta}\int_0^{\zeta}e^{\lambda}F(\lambda)\mathrm{d}\lambda \tag{14}$$

For *s*=0 (ζ=0), the boundary condition $g\,\big|_{s=0}$=0 implies that the integration constants are interrelated as

$$C_1+C_1'=0 \tag{15}$$

As a result, Eq.(14) is further written as

$$g=-\sinh\zeta\int_{\zeta}^{\infty}e^{-\lambda}F(\lambda)\mathrm{d}\lambda-e^{-\zeta}\int_0^{\zeta}\sinh\lambda\,F(\lambda)\mathrm{d}\lambda \tag{16}$$

It can be easily seen that the function *g* [Eq.(16)] depends on both the geometrical factor ζ as well as the two kinematic degrees of freedom, i.e. *v* and ω (involved in *F*, cf. eq.(13)).

2.3 Pressure Distribution

2.3.1 Pressure at the apex

Since for *s*=*l* it holds that *g*(*l*)=0 [eq.(6)], Eq.(3) results in

$$(p)_0=f(l)=P_{atm}-6\mu u\int_0^s R\,h^{-3}\,\mathrm{d}s \tag{17}$$

2.3.2 Pressure at any point of the membrane

$$p(s,\phi) = P_{atm} - 6\mu u h^{-3} \int_0^s R\,ds + g(s,v,\omega)\cos\phi \tag{18}$$

2.4 Force – Displacement Relationship

The equilibrium of the tooth under the force-system system $\mathbf{F}=\{F_x,\ F_y,\ F_z\}^{\mathrm{T}}$ leads to the following relationship between this and the induced displacement $\mathbf{u}=\{u,v,\omega\}^{\mathrm{T}}$:

$$\begin{aligned} F_x &= \int_{s=0}^{s=l}\int_{\phi=0}^{\phi=2\pi 2} (p-P_{atm})\,R\frac{\mathrm{d}R}{\mathrm{d}s}\,\mathrm{d}s\,\mathrm{d}\phi \\ F_y &= -\int_{s=0}^{s=l}\int_{\phi=0}^{\phi=2\pi 2} (p-P_{atm})\,R\cos\phi\frac{\mathrm{d}x}{\mathrm{d}s}\,\mathrm{d}s\,\mathrm{d}\phi \\ M_z &= -\int_{s=0}^{s=l}\int_{\phi=0}^{\phi=2\pi 2} (p-P_{atm})\,R\cos\phi\frac{\mathrm{d}x}{\mathrm{d}s}\left(x\ \frac{\mathrm{d}x}{\mathrm{d}s}+R\ \frac{\mathrm{d}R}{\mathrm{d}s}\right)\mathrm{d}s\,\mathrm{d}\phi \end{aligned} \tag{19}$$

By substituting the pressure p from Eq.(18) in Eq.(19), we obtain

$$F_x = 6\pi\mu u h^{-3}\int_{s=0}^{s=l} R^3\,\mathrm{d}s\ ,\quad F_y = -\pi\int_{s=0}^{s=l} gR\frac{\mathrm{d}x}{\mathrm{d}s}\,\mathrm{d}s,\quad M_z = -\pi\int_{s=0}^{s=l} g\,R\left(x\frac{\mathrm{d}x}{\mathrm{d}s}+R\frac{\mathrm{d}R}{\mathrm{d}s}\right)\mathrm{d}s \tag{20}$$

3 A MODERN CONSIDERATION

3.1 Kinematic quantities

In order to generalize the above procedure to any arbitrary shaped axisymmetric tooth root, in this paper a numerical solution is proposed. For simplicity, a PDL of *constant thickness* is considered. In this case, in virtue of Eq.(16), Eq.(20) obtains its final matrix form

$$\begin{Bmatrix} F_x \\ F_y \\ M_z \end{Bmatrix} = \begin{bmatrix} k_{11} & 0 & 0 \\ 0 & k_{22} & k_{23} \\ 0 & k_{32} & k_{33} \end{bmatrix} \cdot \begin{Bmatrix} u \\ v \\ \omega \end{Bmatrix} \tag{21}$$

or, in compact form:

$$\mathbf{F} = \mathbf{K}\,\mathbf{u} \tag{22}$$

where the above 3×3 "stiffness" matrix **K** relates the applied force-system **F** on the tooth with the induced rigid-body movements **u**. The elements of **K** are given as follows:

$$k_{11} = c\ \ I_5\,,\qquad k_{22} = 2c\int_0^l \left(\sinh\zeta\ \cdot I_1 + e^{-\zeta}\cdot I_2\right)\ R\ \frac{\mathrm{d}x}{\mathrm{d}s}\ \mathrm{d}s \tag{23}$$

$$k_{23} = 2c\int_0^l \left(\sinh\zeta\cdot I_3 + e^{-\zeta}\cdot I_4\right)\ R\ \frac{\mathrm{d}x}{\mathrm{d}s}\ \mathrm{d}s\,,\qquad k_{32} = 2c\int_0^l \left(\sinh\zeta\cdot I_1 + e^{-\zeta}\cdot I_2\right)\ R\ \left(x\frac{\mathrm{d}x}{\mathrm{d}s}+R\frac{dR}{ds}\right)\mathrm{d}s \tag{24}$$

$$k_{33} = 2c\int_0^l \left(\sinh\zeta\cdot I_3 + e^{-\zeta}\cdot I_4\right)\cdot R\cdot\left(x\frac{\mathrm{d}x}{\mathrm{d}s}+R\frac{\mathrm{d}R}{\mathrm{d}s}\right)\mathrm{d}s \tag{25}$$

where the arc-length dependent functions are:

$$I_1 = I_1(s) = I_1(\zeta) = \int_\zeta^\infty e^{-\lambda}\cdot R\cdot\frac{\mathrm{d}x}{\mathrm{d}\lambda}\ \mathrm{d}\lambda\,,\qquad I_2 = I_2(s) = I_2(\zeta) = \int_0^\zeta \sinh\lambda\cdot R\cdot\frac{\mathrm{d}x}{\mathrm{d}\lambda}\ \mathrm{d}\lambda \tag{26}$$

$$I_3 = I_3(s) = I_3(\zeta) = \int_\zeta^\infty e^{-\lambda}\cdot R\cdot\left(x\frac{\mathrm{d}x}{\mathrm{d}\lambda}+R\frac{dR}{d\lambda}\right)\ \mathrm{d}\lambda\,,\qquad I_4 = I_4(s) = I_4(\zeta) = e^{-\zeta}\int_0^\zeta \sinh\lambda\cdot R\cdot\left(x\frac{\mathrm{d}x}{\mathrm{d}\lambda}+R\frac{dR}{d\lambda}\right)\ \mathrm{d}\lambda \tag{27}$$

and the constants are: $I_5 = \int_0^l R^3\ \mathrm{d}s$ =constant and $c = 6\pi\mu/h^3$ (28)

Using the above definitions, it is trivial to prove that $k_{32}=k_{23}$. In other words, the stiffness matrix **K** is symmetric.

3.2 Determination of centers

3.2.1 Centre of Resistance (Cre)

By virtue of Eq(21) in case of pure translation (ω=0) due to a force F_y, one can easily derive:

$$v = \frac{F_y}{k_{22}} \tag{29}$$

and also

$$x_{Cre} = \frac{M_z}{F_y} = \frac{k_{23}}{k_{22}} \tag{30}$$

3.2.2 Centre of rotation (Cro)

This centre corresponds to the well-known pole of motion and its location may be inside or outside the tooth, depending on the applied loads (F_x,F_y,M_z). It should become clear that the moment M_z includes not only the external applied torque M_b at the bracket B, but also the moment of the forces (F_x,F_y) with respect to the origin O. Since the motion of each point P on the tooth may be considered as a result of pure rotation around the centre of rotation Cro, it is given on the basis of the displacements of the tooth apex O (O ≡ P) as follows

$$x_{Cro} = -\frac{v}{\omega},\ y_{Cro} = \frac{u}{\omega} \tag{31}$$

For the particular case of a force F_y applied at a point on the long tooth axis Ox and being perpendicular to that (plus an external torque M_b at the bracket B), it is easily proven that u=0 and consequently y_{cro}=0. In other words, the centre of rotation lies on the axis of symmetry. Furthermore, if only a single tipping force F_y is applied at a distance d from the apex O, the solution of the equations system leads to

$$u = 0,\quad v = \frac{F_y\left(k_{33} - k_{23}d\right)}{\left(k_{22}k_{33} - k_{23}^2\right)},\quad \omega = \frac{F_y\left(-k_{23} + k_{22}d\right)}{\left(k_{22}k_{33} - k_{23}^2\right)} \tag{32}$$

whence

$$x_{Cro} = -\frac{\left(k_{33} - k_{23}\ d\right)}{\left(-k_{23} + k_{22}\ d\right)} \tag{33}$$

From the above analysis it is concluded that, generally, the centre of rotation (Cro) depends on all three induced displacements (u,v,ω) or, in other words, on the applied force-system.

3.3 Pressure distribution

After the determination of the tooth displacement $\mathbf{u} = \{u, v, \omega\}^T$, equation (18) may be applied for each point along the tooth/PDL interface. A careful inspection shows that besides the integrals I_1, I_2, I_3 and I_4 [Eqs.(26-27)], the application of eq.(18) at a certain point M of the PDL requires of one more integral I_6 that depends on this point, and it is given by

$$p(s) = P_{atm} - 6\mu u h^{-3} I_6(s) - \sinh\ \zeta\ 12\ \mu h^{-3}\left[I_1(s)v + I_3(s)\omega\right] - e^{-\zeta}\left[I_2(s)v + I_4(s)\right] \tag{34}$$

where

$$I_6 = I_6(s) = \int_0^l R\ \mathrm{d}s \tag{35}$$

4 NUMERICAL IMPLEMENTATION

4.1 General procedure

Numerical implementation consists of the accurate and efficient calculation of the five integrals, Is, explicitly given in Eqs.(26-28); theses are functions of both the tooth shape and the thickness of the PDL. Additionally, if the distribution of the pressure p_M=p(s_M) along the PDL is desired, it is necessary to additionally calculate one more integral involved in eq.(40), as it was mentioned above.

In general, the open boundary S=AOB of the tooth being in contact with the PDL, is divided into N boundary segments. These may be either straight segments (two-point definition) or curvilinear arcs such as parabolic (three-point definition). However, only straight elements were implemented in this paper. In particular, the tooth root was subdivided by N equidistant parallel planes normal to its axis of revolution. At the so produced N+1 nodal points with curvilinear coordinates s_0=0,s_1,s_2,…,s_N=AO, the coordinates (x_i, y_i), (i=1,2,…,N) are prescribed. Integrals Is and stiffness k_{ij} (i,j=1,2,3) in Eq.(21) are calculated by using one-dimensional numerical integration (such as mid-point for simplicity), the latter being equivalent to one-point Gaussian quadrature.

After the Is have been calculated, the linear system of the three equations [eq.(21)] is easily solved in (u,v,ω) and then both the centre of resistance and the centre of rotation are calculated. In the sequence, by sweeping the boundary AOB from A (s=0) to B, the pressure is successively calculated at the 2N+1 nodal points. In detail, the proposed algorithm consists of the following steps:

Step 1: Divide half of the tooth surface (boundary) into segments made by N cuts perpendicular to the axis of symmetry at equal distances along this axis.

Step 2: For each boundary segment, calculate and store the variations of the five integrals given by Eqs.(26-28), i.e. ΔI_i, i=1,…,5.

Step 3: Calculate the integral $I_5(l) = \sum_{i=1}^{N} \Delta I_i$.

Step 4: Sweep all segments along the tooth boundary and calculate the corresponding nodal values of the integrals $I_1(s_i), I_2(s_i), I_3(s_i)$ and $I_4(s_i)$, $i = 1,2,...N$.

Step 5: Calculate the line integrals in eqs.(23-25), on the basis of the nodal values that were found in Step 4.
Step 6: Apply the force-system and calculate (u,v,ω) by solving a linear system of two equations.
Step 7: Calculate the pressure distribution using Eq.(34) and the nodal values of the integrals found in Step 4.

4.2 Application

We analyze a typical upper central incisor of either conical (where an analytical solution exists) or *paraboloid* shape with the following geometrical data (Figure 3):
Root length : L = 13·0 mm; Root diameter : D = 7·8 mm; Elastic modulus (Tanne et al.[6]): E=0.68MPa ;
Poisson's ratio: v=0.50 ; Shear modulus: $\mu=E/[2(1+v)]$=0.227MPa; PDL thickness h=0·229 mm.

The atmospheric pressure was taken as P_{atm}=0·101 MPa. The centre of rotation is calculated for the case of a tipping force F_y=1N that is applied at the bracket of the tooth at a distance H=5.8mm from the line AB (tooth margin) of the PDL (Figure 3); so, the distance between the bracket and the apex becomes $d = L+H$ = 18·8 mm.

Obviously, the pressure P is a function of the arch length s. Nevertheless, for presentation purposes results are presented in Figure 4 in terms of the normalized x-coordinate, i.e. the ratio x/L. For a given large or small value of x/L, there are two corresponding points along the tooth surface, i.e. (left: M_1 and right: M_4) or (left: M_2 and right: M_3), respectively, as shown in Figure 3. It is also noted that for the conical case illustrated in Figure 4, both the analytical and the numerical technique are identical (as anticipated).

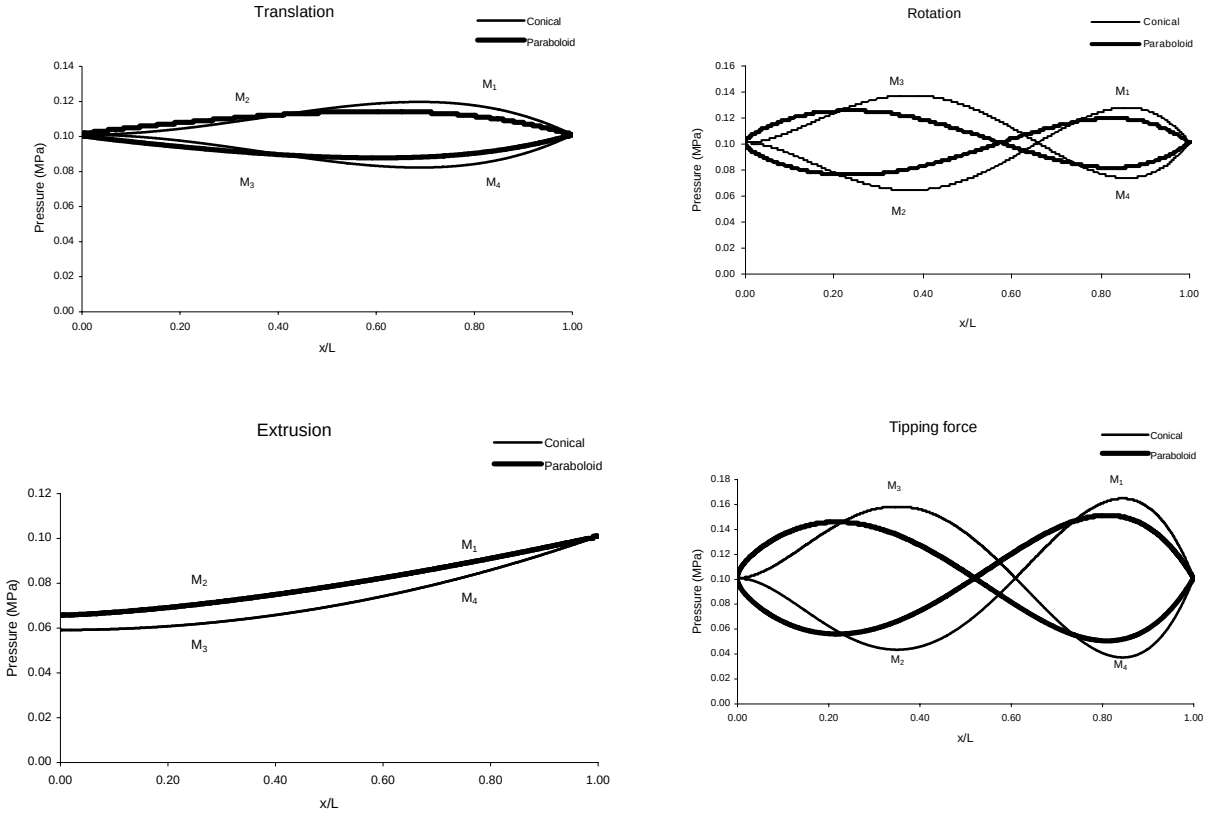


Figure 4: Pressure distribution for several orthodontic force systems: (a) Upper left: Pure translation under a unit force (F_y=1N) applied at the centre of resistance, (b)Upper right: Pure rotation under a moment (M_z=5Nmm), (c) Lower left: Extrusion under a unit force (F_x=1N) and (d) Single tipping force under a unit force (F_y=1N) applied at the tooth bracket at a distance H=5.8mm above the margin AB

Table 1 presents the results of the calculated positions of the centre of resistance (Cre) and the centre of rotation (Cro) using a variable number of boundary segments N, based on equidistant cuts by lines perpendicular to the long tooth axis. It can noticed, for example, that in case of a paraboloidal root the calculated position of the Cre converges near the value 8·253mm from the apex, that corresponds to 63.5% of the root length measured from the tooth apex (O) or 36.5% from the margin AB. Also, Table 2 presents the calculated stiffness of the PDL in pure translation, extrusion and pure moment rotation, as well as the location of the centre of resistance (Cre) as a percentage of the root length.

Table 1. Calculation of the centre of resistance (X_{cre}) and the centre of rotation (X_{cro}) for an upper central incisor (root length L=13mm, root diameter D=7.8mm) using N boundary segments along (AO). Distances are measured from the apex.

Number of boundary segments (N)	Conical root		Paraboloidal root	
	X_{cre} [mm]	X_{cro} [mm]	X_{cre} [mm]	X_{cro} [mm]
8	9·689	9·103	8·410	7·857
16	9·622	9·118	8·305	7·805
32	9·600	9·140	8·271	7·803
64	9·592	9·154	8·260	7·808
128	9·589	9·162	8·256	7·813
256	9·588	9·167	8·255	7·815
512	9·587	9·169	8·254	7·817
1024	9·587	9·170	8·254	7·818
2048	9·587	9·170	8·253	7·818
4096	9·587	9·171	8·253	7·818
Analytical	9·587	9·171	-	-

Table 2. A comparison between the stiffness and the location of the centre of resistance (Cre) of the conical and paraboloidal roots corresponding to an upper central incisor using N=4096 boundary segments.

Shape of root	STIFFNESS			Location of Cre in (%)
	Translation [mm/N]	Extrusion [mm/N]	Rotation [rad/N.mm]	
Conical	1.887×10^{-5}	1.396×10^{-5}	2.446×10^{-5}	26.3
Paraboloidal	9.380×10^{-6}	8.947×10^{-6}	2.050×10^{-6}	36.5

5 CONCLUSIONS

A boundary-type computational technique was presented for the calculation of the centres of resistance and rotation, as well as the induced pressure distribution inside the periodontal ligament of a loaded tooth. The formulation consists of a symmetric "stiffness" matrix based on Reynolds equations, which represents the resistance of the tooth-support to several force-systems. Due to the poor numerical integration scheme applied, in order to achieve convergence a significant large number of boundary segments was required.

REFERENCES

[1] Nikolai, R.J. (1985), *Bioengineering Analysis of Orthodontic Mechanics*, Lea & Febiger, Philadelphia.

[2] Burstone, C.J. (1962), Biomechanics of tooth movement. In: *Vistas in orthodontics*, Eds. B.S. Kraus and R.A. Riedel, Lea and Febiger, Philadelphia, pp. 197-213.

[3] Christiansen, R.L. and Burstone, C.J. (1969), "Centers of rotation within the periodontal ligament," Am J Orthod 55(4), pp. 353-369.

[4] Nikolai, R.J. (1974), **"**Periodontal ligament reaction and displacement of maxillary central incisor subjected to transverse crown tipping,**"** Journal of Biomechanics 7, pp. 93-99.

[5] Nikolai, R.J. and Schweiker, J.W. (1972), **"**Investigation of Root-Periodontium Interface Stresses and Displacements for Orthodontic application,**"** Experimental Mechanics 12, pp. 406-413.

[6] Tanne K., Nagataki T., Inoue Y., Sakuda M. and Burstone C.J. (1991), "Patterns of initial tooth displacements associated with various root lengths and alveolar bone heights," American Journal of Orthodontics and Dentofacial Orthopedics 100, pp. 66-71.

[7] McGuiness N.J.P, Wilson A.N., Jones M.L. and Middleton J. (1992), "A stress analysis of the periodontal ligament under various orthodontic loadings," European Journal of Orthodontics 13, pp. 115-120.

[8] Provatidis, Ch. and Toutountzakis, N. (1998), "A critical review of older and contemporary applications of biomechanical methods in orthodontics," Hellenic Orthodontic Review 1, pp.27-49.

[9] Jones M.L., Middleton J., Hickman J., Volp C., Knox J. (1998), "The development of a validated model of orthodontic tooth movement of the maxillary central incisor in the human subject," Russian Journal of Biomechanics 2, pp. 36-44.

[10] Provatidis, Ch. (1998), Bone remodelling in orthodontics, In: Computer Methods in Biomechanics and Biomedical Engineering - 2, Eds. J. Middleton, M.L. Jones and G.N. Pande, Gordon and Breach Science Publishers, The Netherlands, pp. 655-661.

[11] Provatidis C.G. (1999), "Numerical Estimation of the Centres of Rotation and Resistance in Orthodontic Tooth Movements," Computer Methods in Biomechanics and Biomedical Engineering 2, pp. 149-156.

[12] Provatidis, C.G. (2000), "A comparative FEM-study of tooth mobility using isotropic and anisotropic models of the periodontal ligament," Medical Engineering & Physics 22, pp. 359-370.

[13] Provatidis, C.G. (2001), "An analytical model for stress analysis of a tooth in translation," International Journal of Engineering Science 39, pp. 1361-1381.

[14] Provatidis, C.G. (2002), "Parametric finite element analysis and closed-form solutions in orthodontics," Computer Methods in Biomechanics and Biomedical Engineering 5(2), pp. 101-112.

[15] Middleton J., Jones M. and Wilson A. (1996), "The role of the periodontal ligament in bone modeling: The initial development of a time-dependent finite element model," American Journal of Orthodontics and Dentofacial Orthopedics 109, pp. 155-162.

[16] Provatidis, C.G. (2002), "The role of the principal strains within the periodontal ligament of a tooth during long-term intrusion," Russian Journal of Biomechanics 6 (3), pp. 38-49.

[17] Provatidis, C.G. (2003), "A bone-remodelling scheme based on principal strains applied to a tooth during translation," Computer Methods in Biomechanics & Biomedical Engineering 6 (5-6), pp. 347-352.

[18] Synge, J.L. (1933), "The tightness of the teeth, considered as a problem concerning the equilibrium of a thin elastic membrane," Philosophical Transactions of Royal Society of London, Series 231A, pp. 435-477.

[19] Provatidis, Ch. and Kanarachos, A. (2000), "Boundary-type hydrodynamic analysis of tooth movement," Engineering Analysis with Boundary Elements 24, pp. 661-669.

Fluid Mechanics and Aerodynamics

A MESHLESS METHOD FOR LIFTING-BODY FLOW SIMULATIONS

Jiahn-Horng Chen and Hong-Mau Chen

Department of Systems Engineering and Naval Architecture
National Taiwan Ocean University
2 Pei-Ning Road, Keelung, Taiwan
e-mail: B0105@mail.ntou.edu.tw

Keywords: Radial Basis Function, Lifting Body, 2-D Potential Flow

Abstract. *A meshless method with the logarithmic radial basis function is introduced to solve two-dimensional potential flow past a lifting body. The collocation approach is employed to find the numerical solution. It is found that the new radial basis function demonstrates a better convergence than the reciprocal multiquadric function that is widely employed in the literature. Furthermore, the present study successfully applies the meshless method to problems in a multi-connected domain.*

1 INTRODUCTION

The meshless method has been widely investigated in the past decade and emerged as a new category of computational methods. One of its advantages lies in that no mesh generation is required to solve differential equations numerically. It is well known that in traditional numerical methods, such as finite element methods, finite difference methods, and boundary element methods, it is usually difficult and takes much time to generate proper meshes for computational purposes. This is especially true for three-dimensional problems with complicated geometry in engineering applications.

The meshless method is currently at the stage of rampant development. Various approaches and computational procedures have been proposed and advocated in the literature. However, not every method that is claimed to be meshless is really meshless. Basically speaking, a true meshless method must possess the spirit of "meshlessness." That is, it provides a computational procedure without relating to any mesh point connectivity.

Based on this spirit, three different approaches to develop meshless methods have been successfully proposed. The first one is based on the spirit of finite element method and employs the Petrov-Galerkin weak formulation[1-2]. Detailed theories and formulations can be found in the book by Atluri and Shen[2]. In fact, this approach is the earliest one that attempted to replace the finite element method that generally needs complicated meshes in computations. However, its major disadvantage lies in that the Galerkin-type approach cannot be used for interpolation purposes. Therefore, its applications in problems with essential boundary conditions are usually difficult.

The second approach is of boundary element type. This approach is somewhat new, compared to the first one. It attempted to discretize boundary integral formulation without employing a mesh. Of course, grid points in this approach are all on the boundaries. Several procedures[1, 3-5] have been proposed with different discretization concepts.

The third approach employs radial basis functions (RBF). The essence of this approach is its employment of high-order interpolating functions to approximate solutions of differential equations. All RBFs possess the property that their values are determined only by distances and have nothing to do with directions. Therefore, they are spherically symmetric functions. According to the theories of partial differential equations, they are especially applicable to elliptic type equations.

One may find many RBFs in the literature, such as, for example, monomials, various orthogonal polynomial functions, Fourier series, and so on. These simple RBFs are only applicable to simple differential equations with simple geometries. The development of more advanced RBFs originated from interpolation of multivariate functions. Their outstanding accuracy in interpolations has been widely validated. For example, Frank[6] compared nearly 30 interpolating methods for discrete data and found that Duchon's thin-plate spline and Hardy's multiquadric function fit data best. These two functions are special cases of RBFs. In the literature, the multiquadric function, reciprocal multiquadric function, Gaussian function, and thin-plate spline are representatives for fitting discrete data. Recently, Kansa[7-8] introduced multiquadric functions to solve hyperbolic, parabolic and elliptic differential equations with collocation methods. He found that they had quite good convergence properties and achieved outstanding computational efficiency. More recently, Cheng, et al[9]

conducted a thorough study of mutiquadric RBFs applied to partial differential equations and showed their superior convergence property. In fact, the RBF is of spectral convergence order for interpolation of discrete data points. Many believe that this property is preserved in solving differential equations with collocation methods.

In addition to remarkable convergence property, there are several other advantages in using RBFs together with a collocation method to solve differential equations. For example, the solution procedure is truly meshless. Furthermore, the dimensionality of space has no effects on the convergence rate. Only the density of grid points does.

Applied to solution of partial differential equations, the RBFs have their own inherent disadvantages, too. For example, they are usually globally supported, rather than compactly supported. This leads to a full coefficient matrix. In addition, in solving simultaneous algebraic equations, they easily result in poor conditioned coefficient matrices. Several remedies for these inherent drawbacks have been proposed in the literature. Among them, the most promising one is the compactly supported RBFs[10-12].

Nevertheless, it is quite unfortunate that discussions and applications of RBFs for solving differential equations are seldom examined from the view point of differential equation theories. Even though we may approximate solutions of differential equations by superposition of RBFs which are carefully chosen, it is believed that more reliable and efficient computations can be achieved when the choice of RBFs is considered in connection with theories of differential equations.

Furthermore, one of early applications of meshless methods to potential flow computations has been due to Alturi and Zhu[1]. They illustrated capabilities of the Galerkin-based meshless method in computing the external uniform flow past a circular cylinder. Nevertheless, they were not interested in the flow physics and did not discuss in their numerical procedure how to circumvent the non-uniqueness problem encountered in this multi-connected flow setup. Following the idea of Galerkin-based meshless method, Mosqueira, et al[13] proposed "the enrichment function" to improve accuracy in computations of potential flows. In addition, Tolstykh[14] first employed simple RBFs to compute external flow problems.

2 THEORETICAL FORMULATION AND RADIAL BASIS FUNCTION

2.1 Theoretical formulation

A uniform potential flow past a lifting body in two dimensions can be described by the Laplace equation

$$\nabla^2 \phi = 0 \tag{1}$$

where $\phi = \phi(x, y)$ is the velocity potential through which we have the velocity field

$$\mathbf{u} = \nabla \phi \tag{2}$$

The boundary conditions are

$$\begin{aligned} \nabla\phi \cdot \mathbf{n} &= 0, \quad \text{on the surface of body} \\ \nabla\phi &= \mathbf{U}, \quad \text{far away from the body.} \end{aligned} \tag{3}$$

where $\mathbf{n}$ and $\mathbf{U} = U_x\mathbf{i} + U_y\mathbf{j}$ denote outward unit vector normal to the body's surface and the uniform velocity far away from the body, respectively. For the incoming flow, the angle of attack, α, is defined as

$$\alpha = \tan^{-1}\left(\frac{U_y}{U_x}\right)$$

Since the above flow setup in two dimensions establishes a multi-connected problem, an additional Kutta condition must be specified to ensure the well-posedness of the mathematical formulation.

2.2 Logarithmic Radial Basis Function

In the literature, the multiquadric function

$$f = \frac{1}{\sqrt{r^2 + c^2}} \tag{4}$$

where r represents distance between the collocation point and any point in the field and c some proper constant, is usually employed as the RBF for the solution of Eq. (1). It is interesting to find that this function is surprisingly similar in form to the analytical spherically symmetric solution of Eq. (1) in three dimensions

$$\phi = -\frac{1}{4\pi r}. \tag{5}$$

Therefore, it is somewhat expectable that the multiquadric function is a good choice in computations to approximate the solution of Laplace equation.

Nevertheless, it is well known that in two dimensions, the spherically symmetric solution to the Laplace equation is of the form

$$\phi = \frac{1}{2\pi}\ln r \tag{6}$$

Hence, one may expect that a logarithmic form of RBF

$$f = \ln(r^2 + c^2) \tag{7}$$

is a more suitable choice for computational purpose. In fact, the enrichment function proposed by Mosqueira, et al[13] is similar to Eq. (7). They employed such a function to improve numerical accuracy in solving the Laplace equation. Unfortunately, their formulation is of Galerkin type and they did not have further discussions.

2.3 The Meshless Method

Having chosen the proper RBF, we can discretize Eq. (1) and boundary conditions. First, we properly select a set of collocation points, $\mathbf{x}_i = (x_i, y_i)$, where $i = 1, 2, ..., n$. Then, we express the solution to be

$$\phi(\mathbf{x}) \approx \sum_{i=1}^{n} a_i f_i = \sum_{i=1}^{n} a_i \ln(r_i^2 + c^2) \tag{8}$$

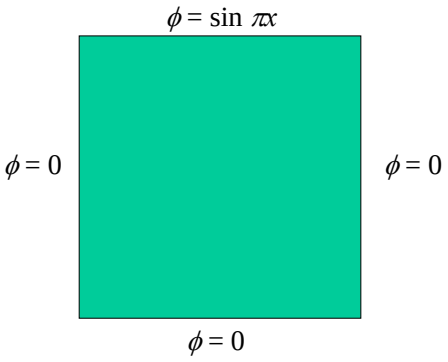
where a_i are coefficients to be determined and r_i is the distance between the field point $\mathbf{x}$ and the collocation point $\mathbf{x}_i$. The approximate solution, Eq. (8), satisfies the governing equation at all interior collocation points and the boundary conditions at boundary collocation points, from which we obtain simultaneous equation system for unknown coefficients.

3 TESTS OF CAPABILITTY OF LOGARITHMIC RADIAL BASIS FUNCTIONS

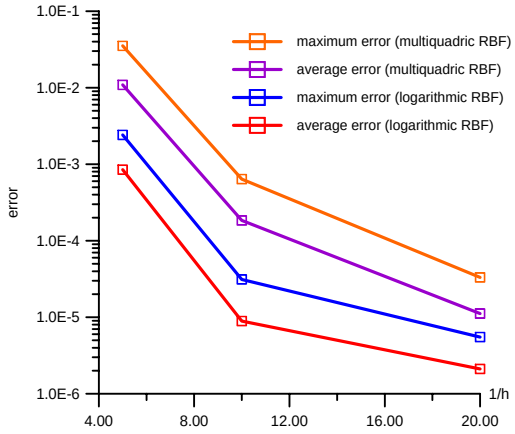
To test the capability of the logarithmic RBF in solving the two-dimensional Laplace equation, we carried out a series of case studies. The test problem is to solve the Laplace equation in a unit square domain with different types of boundary conditions.

3.1 Test 1 (Dirichlet Boundary Conditions)

For the first test, Dirichlet boundary conditions are specified on all sides of the square domain, as shown in Figure 1(a). On the right, left, and bottom sides of the domain, we prescribed $\phi = 0$. On the upper side, we



(a) Test problem with Dirichlet boundary conditions.



(b) Errors for different grid densities and RBFs.

Figure 1. Test domain and computational results.

specified a sinusoidal function, $\phi = \sin \pi x$.

Both the logarithmic and multiquadric RBFs were implemented in our computations with different densities of collocation points. In addition, to compare these two functions, we employed three sets of collocation points in computations. All the three sets are uniformly distributed. The distances in both *x*- and *y*-directions between two consecutive points are $h = 0.2,\ 0.1$, and 0.05. The total numbers of collocation points for the three sets are 25, 121, and 441, respectively.

Figure 1(b) shows the errors at the best values of *c*. The maximum error represents the maximum deviation of the computed data from the analytical solution at the collocation points. The average error denotes the L_2-norm of deviation. As we have expected, the data shown in Figure 1(b) indicate that employing the logarithmic RBF leads to more accurate computational results. Generally speaking, the results due to the logarithmic RBF are at least one-order better than those due to the multiquadric RBF if the density of collocation points keeps the same. To reach the error order of 10^{-3}, the number of collocation points for the logarithmic RBF is about 1/4 for the multiquadric RBF.

3.2 Test 2 (Neumann & Dirichlet Boundary Conditions)

For the second test, we replaced the Dirichlet boundary condition on the upper side with a Neumann type of boundary condition, $\partial\phi/\partial n = \sin \pi x$, as shown Figure 2(a).

We employed the same grid distributions stated in Sec. 3.1. The computational results are shown in Figure 2(b). Again, the results shown here were obtained with the best choices of *c*. They show that the choice of logarithmic RBF gives a better numerical solution, compared to the multiquadric RBF at the same density of collocation points. And again, to reach the error order of 10^{-3}, the number of collocation points required for the logarithmic RBF is about 1/4 for the multiquadric RBF.

3.3 Summaries from the Tests

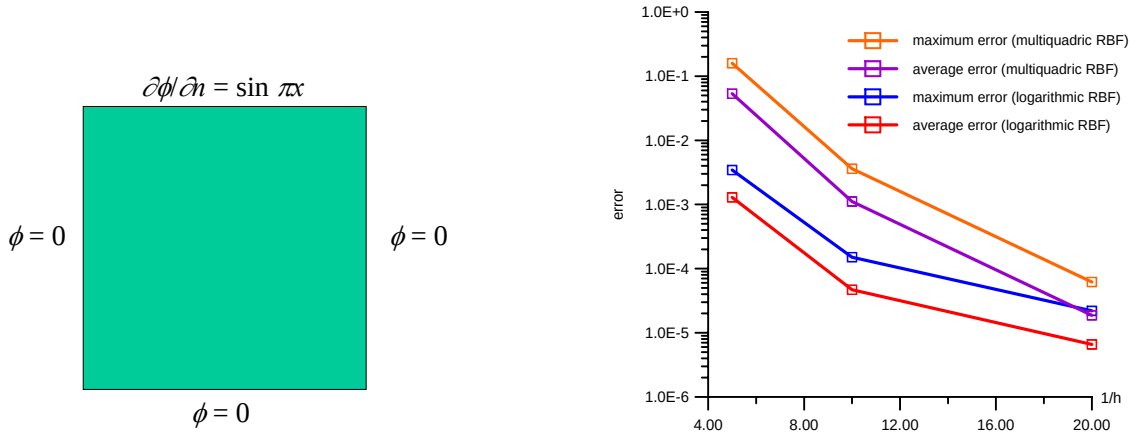
The two simple tests given above show that the logarithmic RBF seems a good choice in approximating two-dimensional problems. In fact, it exhibits a better approximation to the exact solution in either case. Of course, this is what one may expect because of its similarity to the spherically symmetric solution for the Laplace equation in two dimensions.

Here we do not discuss how the value of *c* is chosen. For detail, see the work by Chen[15]. General observations show that it depends on the density of collocation points. The value becomes smaller for a denser collocation point distribution.

4 APPLICATIONS TO LIFTING-BODY PROBLEMS

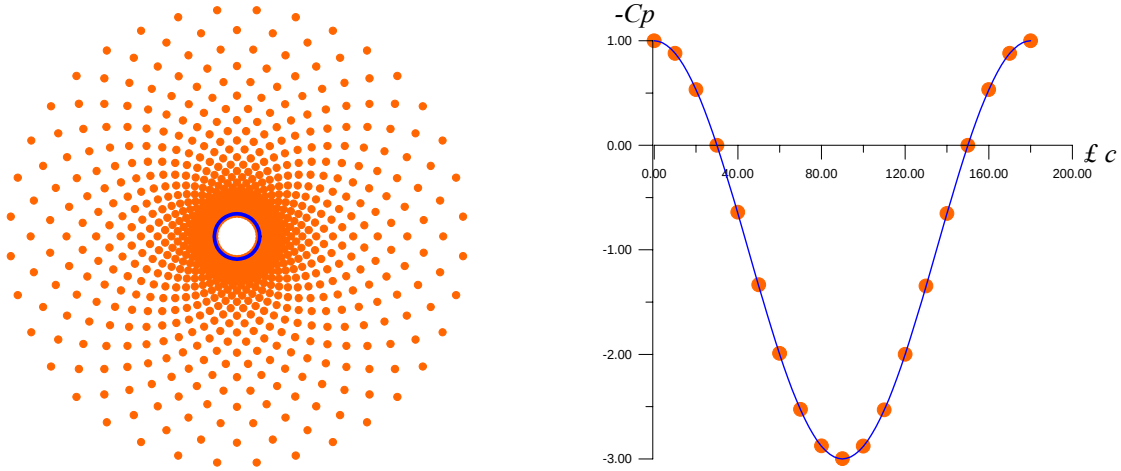
To apply the collocation method to a lifting-body flow problem, we have to address two additional features which bring forth numerical difficulties.

The first one arises because of flow physics. In the formulation, the far-field boundary condition, Eq. (3), is a



(a) Test problem with mixed boundary conditions. (b) Errors for different grid densities and RBFs.

Figure 2. Test domain and computational results.



(a) Schmatic diagram of collocation point arrangement.

(b) Pressure coefficient distribution on cylinder surface.

Figure 3. Flow past a circular cylinder without lift.

vector-valued one. Directly applying approximate expression of the velocity potential, Eq. (8), at any far-field boundary point leads to two algebraic equations. It generally results in a simultaneous equation system in which the number of equations is larger than that of unknown coefficients. To circumvent this dilemma, we devised a symmetric set of collocation points.

The second one comes from the Kutta-condition. It states that the flow leaves the trailing edge at a finite velocity. There are several practical ways to apply this condition. The way we chose here is that the flow speeds on upper and lower surfaces of the lifting body are equal at equal distances from the trailing edge. In the present study, it was usually applied at the midpoint between the trailing edge and closest collocation points on the body surface from the trailing edge.

4.1 Flow past a Circular Cylinder without Lift

The first case we studied is a uniform flow past a circular cylinder without lift. The Kutta condition was specified at the rear point of the cylinder, which gives a symmetric flow pattern.

A staggered collocation point arrangement that was used in the present study is schematically shown in Figure 3(a). In our computations, the computation domain was truncated at a radius of 10R, in which R represents the radius of cylinder.

Solving the algebraic equation system, we obtain the approximate expression of the velocity potential. Then we substituted it in the Bernoulli equation to find the pressure field. And the pressure coefficient, defined as

$$C_p = \frac{p - p_\infty}{\frac{1}{2}\rho U^2} \tag{9}$$

can be computed. Here p_∞ and ρ represent the pressure at the far field and the density of fluid, respectively. For the present study, the computed pressure coefficient on the cylinder surface is shown in Figure 3(b). The horizontal axis represents the azimuthal angle measured from the x-axis with origin at the center of cylinder. Compared to the analytical solution of pressure coefficient, the present computational results show very good agreement.

4.2 Flow past a Circular Cylinder with Lift

For the second case, we prescribed the Kutta condition at the point $(x, y) = (\sqrt{3}/2, -0.5)$. This is a somewhat unnatural specification. However, it provides a good test case in that there exists an analytical solution for such a flow. Theoretically, such a specification of the Kutta condition introduces in the flow a circulation of strength $\Gamma = 2\pi U$ in which U represent the speed of flow at the far field.

The collocation point arrangement employed in the first test case was adopted for computations in the present study. Through the same computation procedure described above, the computed pressure coefficient distribution on the cylinder is shown in Figure 4. Again, we find that the present method gives reasonable results.

4.3 Flow past a Hydrofoil

We now proceed to compute the potential flow past a two-dimensional hydrofoil. The NACA16-006 foil section, a typical section used in hydrofoils, was adopted for computations. This is a symmetric foil section without camber. The maximum thickness is 0.06*c*, in which *c* denotes the chord length of the foil section.

The collocation point arrangement is shown schematically in Figure 5. It was generated simply by the use of *H*-grid concepts in the conventional grid generation, even though we believe that not so many collocation points are required for the regions far up above and down below from the hydrofoil. Owing to its symmetry, only half of the computational domain is shown here. Furthermore, in Figure 5, the scaling in the direction normal to the incoming flow is magnified in order to show clearly the collocating points. The computation domain is 7*c* in length and *c* in half width. For computation purposes, the chord length is usually normalized to unity.

Two computational cases were carried out here. For the first one, the incident flow angle of attack is at $\alpha = 0^\circ$. Since the foil section is a symmetric one, the flow in this study case is symmetric and there is no lift. Figure 6(a) shows the pressure coefficient distribution on the upper surface of hydrofoil. Compared to the measured data[16], the results of the present method have a reasonable agreement.

For the second test, the incident angle of incoming flow is $\alpha = 3^\circ$. The results are shown in Figure 6(b). In

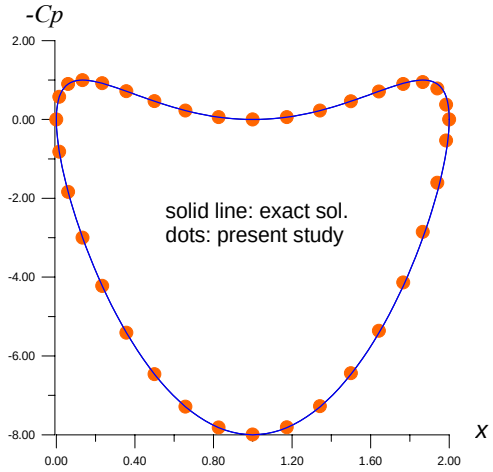


Figure 4. Pressure coefficient distribution on the cylinder for which the Kutta condition is specified at $(x, y) = (\sqrt{3}/2, -0.5)$.

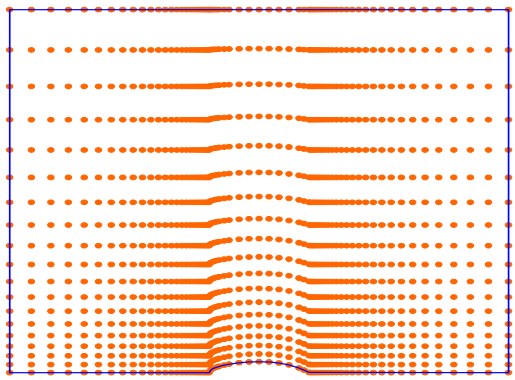
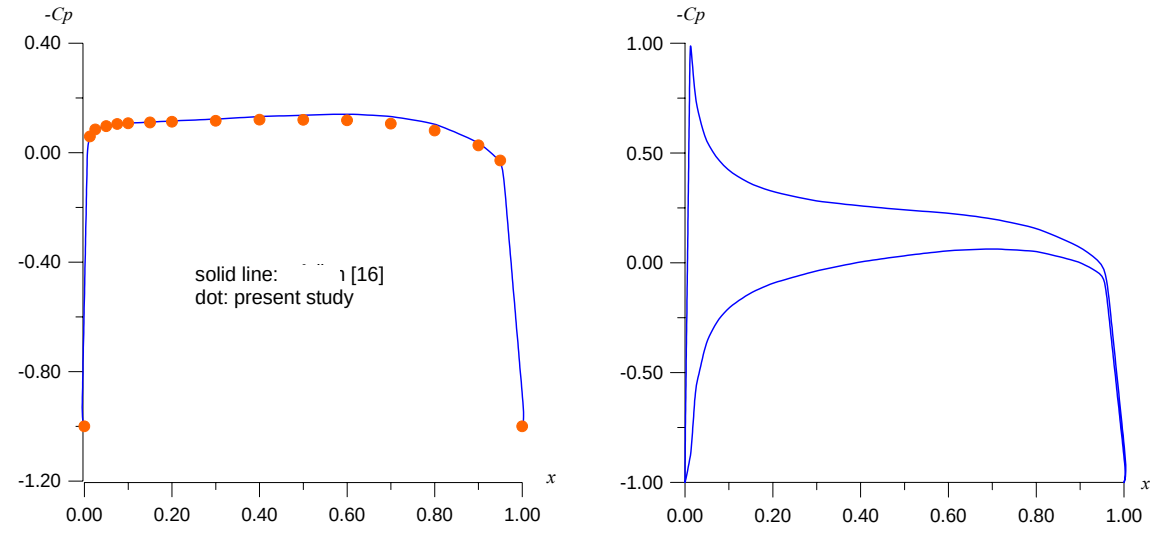


Figure 5. Schematic diagram of collocation point arrangement.



(a) α = 0°. (b) α = 3°.

Figure 6. Pressure coefficient distributions on the hydrofoil with an NACA16-006 section.

this case, we do not have data for comparisons. However, the trend shown in the plot seems reasonable. The area enclosed by the two curves of pressure coefficients distributions for the upper and lower surfaces represents the lift coefficient.

5 CONCLUSIONS

The logarithmic radial basis function, together with a collocation method, has been successfully applied to computations of lifting-body flow problems in two dimensions.

This new RBF is introduced from the theoretical point view. Results of some preliminary test cases have shown that this type of RBF gives a good approximation. In fact, the numerical solutions obtained in the test cases indicate that it gives a better approximation with fewer collocation points.

Then it was applied with a symmetric collocation point distribution to solve the potential flow around a lifting body. The Kutta condition was applied in a way suitable to collocation point method. The computed results in the test cases show that the present procedure works well.

6 ACKNOWLEDGEMENT

This work was supported by the National Science Council, Republic of China, under grant NSC92-2611-E-019-007. The authors would like to express their thanks for this support.

REFERENCES

[1] Atluri, S. N. and Zhu, T. (1998), "A new meshless local Petrov-Galerkin approach in computational mechanics," Computational Mechanics, 22, pp. 117-127.

[2] Atluri, S. N. and Shen, S. (2002), *The Meshless Local Petrov-Galerkin Method*, Tech Science Press, Encino, CA.

[3] Kulkarni, S. S., Telukunta, S. and Mukherjee, S. (2002), "Application of an accelerated boundary-based mesh-free method to two-dimensional potential theory," *IABEM 2002, International Association for Boundary Element Methods*, UT Austin, TX, USA.

[4] Chen, W. and Tanaka, M. (2002), "A meshless, integration-free, and Boundary-only RBF Technique," Computers and Mathematics with Applications, 43, pp. 379-391.

[5] Li, G. and Atluri, S. N. (2002), "Boundary cloud method: a combined scattered point/boundary integral approach for boundary-only analysis," Computer. Methods Appl. Engrg., 191, pp. 2337-2370.

[6] Frank, R. (1972), "Scattered data interpolation: tests of some methods," Math. Comput., 38, pp. 181-199.

[7] Kansa, E. J. (1990), "Multiquadrics --- A scattered data approximation scheme with applications to computational fluid dynamics ---I. Surface approximations and partial derivative estimates," Comput. Math. Appl., 19, pp. 127-145.

[8] Kansa, E. J. (1990), "Multiquadrics --- A scattered data approximation scheme with applications to computational fluid dynamics --- II. Solutions to hyperbolic, parabolic, and elliptic partial differential equations," Comput. Math. Appl., 19, pp. 147-161.

[9] Cheng, A. H.-D., Golberg, M. A., Kansa, E. J., and Zammito, G. (2003), "Exponential convergence and H-C multiquadric collocation method for partial differential equations," Numerical Methods for Partial Differential Equations, 19, pp. 571-594.

[10] Buhmann, M. D. (1998), "Radial functions on compact support," *Proc. Edingurgh Math. Soc.,* Vol. 41, pp. 33-46.

[11] Wendland, H. (1995), "Piecewise polynomial, positive definite and compactly supported radial basis functions of minimal degree," Adv. Comput. Math., 4, pp. 389-396.

[12] Wu, Z. (1995), "Compactly supported positive definite radial functions," Adv. Comput. Math., 4, pp. 283-292.

[13] Mosqueira, G., Colominas, I., Navarrina, F., and Casteleiro, M. (2001), "Meshless methods applied to potential flow problems," in Computational Fluid and Solid Mechanics (edited by K. J. Bathe), pp. 908-911, Elsevier, Oxford, UK.

[14] Tolstykh, A. I. (2001), "High-accuracy difference and meshless methods for fluid and solid mechanics," *Proceedings of the 8th International Conference of Enhancement and Promotion of Computing Methods for Engineering and Science* (EPMESC VIII), Shanghai, China.

[15] Chen, J.-H. (2005), "Logarithmic radial basis function for two-dimensional elliptic partial differential equations," in preparation.

[16] Abbott, I. H. and von Doenhoff (1959), *Theory of Wing Sections*, Dover, New York, USA.

DISCONTINUOUS GALERKIN DISCRETIZATIONS FOR VISCOUS COMPRESSIBLE FLOWS

Ioannis Toulopoulos† and John A. Ekaterinaris‡

† Department of Mathematics
University of Athens
P.O.BOX 1527, FORTH/IACM, 71110 Heraklion, Greece
toulioa@iacm.forth.gr

‡FORTH/IACM
P.O.BOX 1527, 71110 Heraklion, Crete, Greece
ekaterin@iacm.forth.gr

Keywords: High order methods, unstructured grids, shock capturing.

Abstract. *Discontinuous Galerkin discretization is applied for the numerical solution of the Euler and Navier-Stokes equations. Second- and fourth-order accurate solutions are obtained for simple inviscid and viscous flows computed with triangular meshes and polynomial bases of first and third order, respectively. For flows with discontinuities, classical slope limiters are applied for second- and third-order accurate solutions. In addition, characteristic based filters are used to make possible higher order accurate solutions of flows with discontinuities at a reduced computing cost. Discretization of the viscous terms is performed by solving an additional auxiliary equation for the gradient of the conservative variables. Inviscid and viscous flow solutions of different order of accuracy computed with triangular elements are presented.*

1 INTRODUCTION

Many CFD applications, such as rotor aerodynamics, large eddy simulations (LES) of compressible flow, and computational aeroacoustics (CAA) require high-order accuracy in space. In the last few years, efficient finite-difference methods[1-10] found widespread application in these fields. These methods were applied in simple Cartesian-type meshes [3, 7] and more complex domains through the use of generalized coordinate transformations.[5, 10] For highly accurate computations in three dimensions, grid smoothness is required to retain the designed order of accuracy of high-order methods and extreme caution is needed for the computation of metric quantities[5] that appear in the generalized coordinate transformations in order to preserve the high-order of spatial accuracy provided by centered finite-differences[5] or WENO schemes.[10] For high order methods that require wide stencils,[7, 10] specification of boundary conditions is not straight forward. In addition, domain decomposition[5] often used for complex domain grid generation (multi-block meshes) or for parallelization, requires overlap in order to ensure that the high-order accuracy of the interior scheme is retained at the interfaces of the domains.

Unstructured grid methods[11-15] overcome some problems associated with domain decomposition and discretizations of complex domains. Preservation of high order accuracy in unstructured finite volume methods[11] is computationally intensive. The recently proposed spectral volume method[15] for triangular meshes and the discontinuous Galerkin (DG) method for triangular[12, 16, 17] or quatrilateral[14] unstructured meshes overcome the shortcomings of high-order finite volume methods for unstructured grids. The DG method appears to offer certain advantages over the spectral volume method[15] because it is highly local, easy to parallelize, and possible to extend for viscous flow computations.[13, 17] Furthermore, the DG method is suitable for high-order accurate discretizations of complex domains with triangular, quadrilateral, or mixed-type elements and can preserve high order of accuracy in distorted meshes and meshes with hanging nodes that may result from local grid refinement.[14] In addition, the systematic study of Ref. 18, demonstrated that for wave propagation the DG method has very good performance even for anisotropic grids.

In this paper, applications of the DG finite element method for the numerical solution of the Euler and the Navier-Stokes (NS) equations are presented. High-order polynomial representations of the approximate solution within the elements (up to fifth order) are used. The accuracy of the DG method is demonstrated for discretizations with triangular meshes. It was found that for linear-type problems, such as weak acoustic pressure pulse spread, the achieved order of accuracy is approximately the order of the polynomial basis plus one. The achieved order of accuracy for third- and fifth-order polynomial bases is demonstrated for long time integration of a nonlinear problem, the isentropic vortex convection. Additional results for supersonic inviscid flows are presented. The accuracy of viscous flow computations is also demonstrated.

2 GOVERNING EQUATIONS

The two-dimensional Euler or Navier-Stokes equations in conservation law form are considered.

$$\frac{\partial U}{\partial t}+\frac{\partial F_{\text{inv}}}{\partial x}+\frac{\partial G_{\text{inv}}}{\partial y}=\frac{\partial F_{\text{vis}}}{\partial x}+\frac{\partial G_{\text{vis}}}{\partial y}$$

$$U=\begin{Bmatrix}\rho\\ \rho u\\ \rho \upsilon\\ e\end{Bmatrix},\quad F_{\text{inv}}=\begin{bmatrix}\rho u\\ \rho u^{2}+p\\ \rho u\upsilon\\ (e+p)u\end{bmatrix},\quad G_{\text{inv}}=\begin{bmatrix}\rho \upsilon\\ \rho u\upsilon\\ \rho \upsilon^{2}+p\\ (e+p)\upsilon\end{bmatrix} \tag{1}$$

where U is the solution vector for the conservative variables, and F_{inv}, G_{inv} are the inviscid flux vectors, and F_{vis}, G_{vis} are the viscous fluxes. The nonlinear Euler equations can be used for propagation of weak or strong acoustic type disturbances. In these cases, a source term $S(x,y,t)$, which contains disturbances of density $\rho_o=\rho_o(x,y,t)$ and pressure $p_o=p_o(x,y,t)$, is imposed as initial condition. The evolution of weak acoustic-type disturbances is governed by the linearized Euler equations for the acoustic velocities and pressure. Therefore, for validation purposes, solutions computed with the numerical solution of the full nonlinear Euler equations could be compared with available exact solutions of the linearized Euler equations.

Time marching of the viscous or inviscid equations, Eq. (1), can be performed with computationally intensive but unconditionally stable implicit methods. [16, 17] Third- or fourth-order accurate Runge-Kutta methods may also be applied for explicit time marching. In this work the total variation diminishing (TVD) third order Runge-Kutta TVD RK-3 method of Refs. 19, 20, and 21 was used. This method is TVD in the sense that the temporal operator itself does not increase the total variation of the solution. The TVD property of the time integration scheme may not be important for linear problems. It plays however, an important role for time marching of nonlinear hyperbolic problems when using DG space discretizations.

Explicit time marching of DG discretizations with the TVD RK-3 method is appropriate for demonstrating the accuracy of steady state, but more important, of time accurate solutions. For high order DG discretization the stability of explicit time marching schemes [22] degrades dramatically. As a result, numerical solutions of realistic high Reynolds number flows are not possible with explicit time marching. Recent applications of implicit schemes [16, 17] with the DG method demonstrated very good performance for both steady-state [16] and time accurate solutions[17] of time dependent problems.

3 SPACE DISCRETIZATION

The DG method was used for the first time by Reed and Hill [23] for the numerical solution of the neutron transport equation. The DG method is briefly described in this section. Further information, more details, and analysis of the DG method the can be found in the original references [20, 21] and the review articles of Refs. 12 and 24. For each time step the approximate solution, $\mathbf{u}_{\mathbf{h}}$, of the governing equations, Eq. (1), written in conservation law form as, $\partial_t \mathbf{u}+\nabla\cdot\mathbf{F}(\mathbf{u})=0$ is sought in the finite element space of discontinuous functions V_h such that $\boldsymbol{V}_h=\{\varphi_h\in \boldsymbol{L}^{\infty}(\Omega):\varphi_h|_{\boldsymbol{K}}\in \boldsymbol{V}(\boldsymbol{K}),\forall \boldsymbol{K}\in T_h\}$ where $\boldsymbol{T}_h$ is a discretization of the domain Ω using triangular or quadrilateral elements and $\boldsymbol{V}(\boldsymbol{K})$ is the local space that contains the collection of polynomials up to degree k. The weak formulation of the inviscid part of the governing equations is

$$\frac{d}{dt}\int_K u(x,t)\varphi(x)dx=-\sum_{e\in\partial K}\int_e F(u(x,t))\cdot\vec{n}_{e,K}\,\varphi(x)\,d\Gamma+\int_K F(u(x,t))\cdot\nabla(\varphi(x))dx \tag{2}$$

where $\varphi(x)$ is any sufficiently smooth function and $\vec{n}_{e,K}$ is the outward, unit normal to the edge e.

Assuming polynomial expansions of the approximate solution $\mathbf{u}_{\mathbf{h}}$ in Eq. (2), and using the same polynomials for the weighting function $\varphi(x)$ (Galerkin approximation) the weak form can be solved numerically. The mass

matrix integral at the left hand side of Eq. (2) is evaluated numerically using Gauss-Radau integration rules. The integrals on the right hand side of Eq. (2) are evaluated using quadrature rules as follows:

$$\int_e F\left(u(x,t)\right)\cdot\vec{n}_{e,K}\varphi\ (x)\ d\Gamma \approx \sum_{l=1}^{L}\psi_l\ \ F\left(u(x_{el},t)\right)\cdot\vec{n}_{e,K}\varphi\ (x_{el})\ \ |e| \tag{3}$$

$$\int_K F(u(x,t))\cdot\nabla(\varphi\ (x))\ \ dx \approx \sum_{j=1}^{J}\omega_j F(u(x_{K\ j},t))\cdot\nabla(\varphi\ (x_{K\ j}))\,|K| \tag{4}$$

It is important to compute the line integrals of Eqs. (3) and (4) using appropriate order Gaussian quadrature rules. For example, with a third-order polynomial basis a quadrature rule that integrates exactly at least sixth-order polynomial must be used in Eq. (3).

The data are assumed discontinuous across the interfaces of the continuous domain and at each interface two values are available. Therefore, the flux $F\left(u(x,t)\right)\cdot\vec{n}_{e,K}\varphi\ (x)$ is replaced by a suitable numerical flux $\tilde{F}_{e,K}(x,t)$ for the approximate solution u_h and the test function $\varphi_h\in V(K)$. Using $\tilde{F}_{e,K}(x,t)$ in Eqs. (2) and (3), and the desired expansion order for the approximate solution u_h we obtain

$$\begin{aligned}\frac{d}{dt}\int_K u_h(x,t)\varphi_h(x)dx &= \sum_{j=1}^{J}\omega_j F(u_h(x_{K\ j},t))\cdot\nabla(\varphi_h(x_{K\ j}))\,|K| \\ &-\sum_{e\in\partial K}\ \sum_{l=1}^{L}\psi_l\ \ \tilde{F}_{e,K}(x_{el},t)\varphi_h(x_{el})\ \ |e|, \qquad \forall\varphi_h\in V(K),\ \ \forall K\in T_h\end{aligned} \tag{5}$$

where time advancement of Eq. (5) is performed with the third order accurate Runge-Kutta method.

The major difference of the DG formulation with a standard node-based Galerkin finite element method is that the expansion in each element is local without any continuity across the element boundaries. The value of the numerical flux $\tilde{F}_{e,K}(x,t)$ at the edge of the boundary of the element K depends on two values of the approximate solution, one from the interior (right) of the element K, $u^R = u_h(x^{\mathrm{int}(K)},t)$, and the other from the exterior (left) of the element K, $u^L = u_h(x^{ext(K)},t)$. Any consistent, conservative exact or approximate Riemann solver can be used to obtain the numerical flux $\tilde{F}_{e,K}\left(u(x^{\mathrm{int}(K)},t),u(x^{ext(K)},t)\right)$ as follows

$$\tilde{F}_{e,K}(u^L,u^R)=\frac{1}{2}\left[F(u^L)\cdot\vec{n}_{e,K}+F(u^R)\cdot\vec{n}_{e,K}\right]-F^*(u^L,u^R) \tag{6}$$

where $F^*(u^L,u^R)$ is the dissipative part of the numerical flux. The computationally efficient local Lax-Friedrichs flux and the Roe's flux were used in our implementation.

It was shown (see Ref. 12 and references therein) that the formal order of accuracy of the DG method is at least $n+1/2$ if polynomials of degree at most n are used as basis functions. Furthermore, it was shown that for linear problems and Cartesian-type or semi-uniform triangular grids the order of accuracy is $(n+1)$. For simplicity in the rest of this paper, the method is called $(n+1)$ th order accurate if the basis functions are polynomials of degree at most n.

The approximate solution within each element is expanded in a series of local bases functions (polynomials) as follows:

$$u_h(x,y,t)=\sum_{j=1}^{J}c_j(t)P_j^k(x,y) \tag{7}$$

where $c_j(t)$, $j=1,2,...,J$ are expansion coefficients or degrees of freedom for each element, to be computed at each time step, and $P_j^k(x,y)$ are polynomial nodal or modal bases of degree k the most. In this work, first-

third- and fifth-order nodal polynomials for the master element were used. First order polynomial expansion, $k=1$, can be obtained with a hierarchical basis that contains $(k+1)(k+2)/2=3$ first order polynomials. The following non-hierarchical nodal basis of orthogonal, first-order polynomials was used: $P_1^1=1-2y$, $P_2^1=2x+2y-1$, $P_3^1=1-2x$. Each of these polynomials takes unit value at one node, located in the middle of an edge, and zero value at the other nodes located at the middle of the other edges. Orthogonality of the P_1^1, P_2^1, P_3^1 polynomials implies that the mass matrix resulting from the integration at the left hand side of Eq. (4) is diagonal. Second and higher-order nodal polynomial bases are not orthogonal. Nonorthogonality of the polynomials P_j^3, $j=1,...,10$, for example, implies that the mass matrix of Eq. (5) must be computed using high-order accurate Gauss-Radau integration. The fifth-order polynomial basis ($k=5$) requires $(k+1)(k+2)/2=21$ fifth degree polynomials. This expansion basis yields sixth-order accuracy but is very computationally intensive for the solution of the Navier-Stokes equations.

4 LIMITING OF DG DISCRETIZATIONS

Limiting operators[12, 20] $\Lambda\Pi_h$ on piecewise linear DG expansions u_h are constructed in such a way that they satisfy accuracy (if u_h is linear, then $\Lambda\Pi_h u_h = u_h$) conservation of mass for every element K, and slope limiting (the gradient of $\Lambda\Pi_h u_h$ is not bigger than that of u_h for each element K). Theoretical analysis of the slope-limiting operators can be found in Cockburn et al. [20] Cockburn and Shu.[25]

For solutions with P_j^1 base functions inside the triangle K the expansion of Eq. (7) is used for the approximate solutions $u_h(x,y,t)$, where the degrees of freedom or expansion coefficients $c_i(t)$ are the values of the numerical solution at the midpoints of the edges. The basis function is a linear function that takes unit value at the midpoints m_i of the $i-th$ edge and zero value at the midpoints of the other two edges. For more details see Refs. 12 and 20. Limiting of higher order accurate DG discretizations is more complex and computationally intensive. Therefore characteristic based filters in the spirit of Ref. 3 were used to make possible high order accurate solutions with discontinuities.

5 VISCOUS TERM DG DISCRETIZATION

The compressible Navier-Stokes (NS) equations can be written in compact vector form as follows:

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot \mathbf{f_i}(\mathbf{u}) + \nabla \mathbf{f_v}(\mathbf{u}, \nabla \mathbf{u}) = 0 \tag{8}$$

where $\mathbf{u}$ is the vector of the conservative variables and $\mathbf{f_i}$, $\mathbf{f_v}$ denote the inviscid and viscous flux functions. The discretization of the viscous, diffusive part of the NS equations with the DG method is less well known and different than the method described previously for the convective, inviscid part. The straightforward way to extend the scheme of the DG discretization of Eq. (2), which was developed for the convective part of the NS equations, using a centered flux leads to numerically stable but inconsistent solutions.[13, 26] The numerical solutions seem to converge with mesh refinement but have $O(1)$ errors with exact solutions.

Bassi and Rebay[27] introduced a formulation for the discretization of the compressible Navier-Stokes equations with the DG method that is convergent and consistent. Baumann and Oden[28] presented another successful method that avoids the inconsistencies of the simple formulation by adding extra penalty terms to the inner boundaries. The consistent formulation of Bassi and Rebay[27] for the spatial discretization of the viscous term in the NS equations was constructed by resorting to a mixed finite element formulation. The second-order derivatives of the conservative variables required for the viscous terms were obtained by using the gradient of the conservative variables, $\nabla \mathbf{u} = \mathbf{S}(u)$, as auxiliary unknowns of the NS equations. The NS equations were therefore reformulated as the following coupled system to the unknowns $\mathbf{S}$ and $\mathbf{u}$.

$$\begin{aligned} \mathbf{S} - \nabla \mathbf{u} &= 0 \\ \partial_t \mathbf{u} + \nabla \cdot \mathbf{f}_i(\mathbf{u}) + \nabla \cdot \mathbf{f}_v(\mathbf{u}, \mathbf{S}) &= 0 \end{aligned} \tag{9}$$

The weak formulation of the first equation of the system of Eq. (9) is

$$\int_K S_h \phi \, dx - \oint_e \mathbf{u_h} \cdot \mathbf{n} \phi \, d\Gamma + \int_K \mathbf{u_h} \nabla \phi \, dx = 0 \tag{10}$$

where the term $\mathbf{u}_h \cdot \mathbf{n}$ in the second (contour) integral of Eq. (10) is replaced by a numerical flux $H_s(\mathbf{u}^-, \mathbf{u}^+, \mathbf{n})$. This numerical flux is the average between the two interface states

$$H_s(\mathbf{u}^-, \mathbf{u}^+, \mathbf{n}) = \frac{1}{2}(\mathbf{u}^- + \mathbf{u}^+)\mathbf{n} \tag{11}$$

The computed auxiliary variables $\mathbf{S}_h$ are used to form the second equation of the system in Eq. (11) as follows:

$$\begin{aligned} \int_K \mathbf{u_h} \phi \, dx - \oint_e \mathbf{f}_{\text{inv}}(\mathbf{u_h}) \cdot \mathbf{n} \phi \, d\Gamma \quad + \int_K \mathbf{f}_{\text{inv}}(\mathbf{u_h}) \nabla \phi \, dx \\ + \oint_e \mathbf{f}_{\text{vis}}(\mathbf{u_h}, \mathbf{S_h}) \cdot \mathbf{n} \phi \, d\Gamma - \int_K \mathbf{f}_{\text{vis}}(\mathbf{u_h}, \mathbf{S_h}) \nabla \phi \, dx = 0 \end{aligned} \tag{12}$$

In Eq. (12), the term $\mathbf{f}_{\text{vis}}(\mathbf{u}_h, \mathbf{S}_h) \cdot \mathbf{n}$ is replaced with the following centered numerical flux

$$\mathbf{h}_v(\mathbf{u}^-, \mathbf{S}^-, \mathbf{u}^+, \mathbf{S}^+, \mathbf{n}) = \frac{1}{2}[\mathbf{f}_v(\mathbf{u}^-, \mathbf{S}^-) + \mathbf{f}_v(\mathbf{u}^+, \mathbf{S}^+)] \cdot \mathbf{n} \tag{13}$$

6 RESULTS

The accuracy of the numerical method presented in the previous sections was first demonstrated for simple problems in Cartesian domains. Next inviscid and viscous flow solutions in more complex domains were obtained. Time accurate numerical solutions were computed for the propagation of weak disturbances. The accuracy of the numerical solutions was evaluated by comparing the computed results with the exact solution of the linearized Euler equations. A test problem, which has an exact solution, the propagation and reflection of a Gaussian pressure pulse from a solid wall, was considered. A small disturbance initial condition for the full Euler equations is specified by $p(x, y) = \exp(-\ln 2[x^2 - (y - y_o)^2]/w)$ where w is the width of the pulse and y_o is the distance from the wall. The computed pressure disturbances from the numerical solution of the full nonlinear Euler equations were found in very good agreement with the exact solution of the linearized Euler equations.

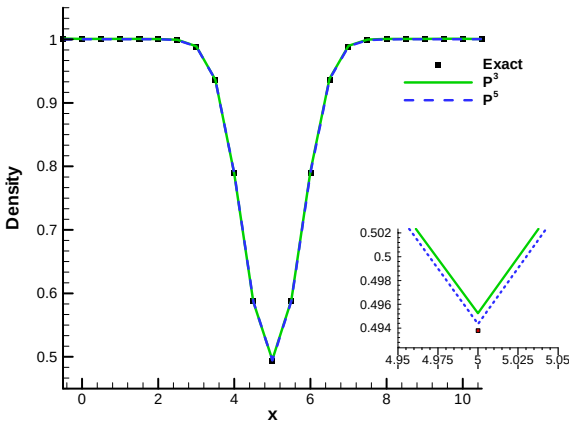


Figure 1. Comparison of the computed solution ($\Delta x = 0.5$) with exact result at time T=10.

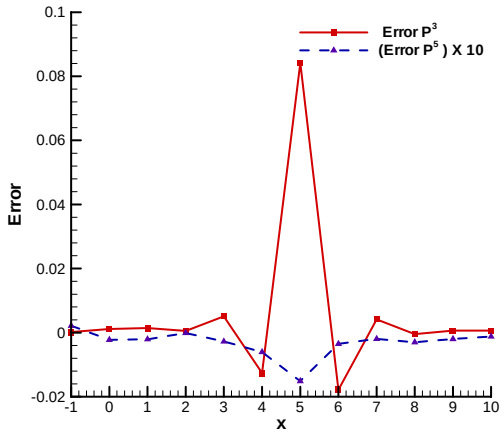


Figure 2. Error of numerical solutions computed with $\Delta x = 1.0$, P^3, and P^5 polynomial bases.

Further validation of time accurate solutions was carried out for the convection of an isentropic vortex.

Numerical solutions of the Euler equations were computed until nondimensional time T=10 (100 unit lengths) with third- and fifth-order polynomial bases on a series of meshes with $\Delta x = 0.25$, $\Delta x = 0.50$, and $\Delta x = 1.0$. A comparison of the computed solutions with the exact result is shown in Fig. 1. The error of the computed solutions with $\Delta x = 1.0$ is shown in Fig. 2. Grid convergence of the computed solutions is shown in Fig. 3. The plot of Fig. 3 that includes the 4th and 6th order theoretical slopes, indicates that the numerical solutions achieves the designed order of accuracy.

Supersonic flow computations with the total variation bounded (TVB) limiter are shown in Figs. 4 and 5. Fig. 4 shows the computed pressure field over a cylinder at $M_\infty = 2.0$. The oblique shock reflection problem at $M_\infty = 3.9$ was also solved numerically. The computed solutions of Figs. 4 and 5 indicate that the TVB limiter with the parameter values of Ref. 12, is appropriate for capturing of strong shocks. Further investigation is required for accurate and sharp capturing of transonic shocks. For transonic flow computations, in Ref. 29 instead of the TVB limiter a stabilization operator was applied. It was shown in Ref. 29 that computed solutions with the stabilization operator and sufficient grid resolution preserve monotonicity and produce sharp capturing of transonic shocks. Evaluation of our approach for stabilization of DG discretizations for transonic flow using characteristic based filters[3] is currently in progress.

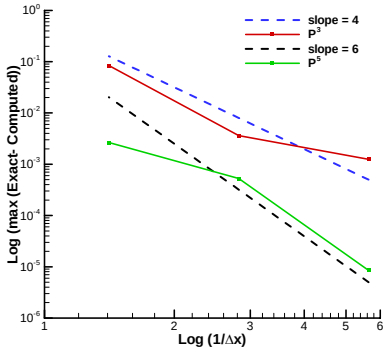


Figure 3. Grid convergence of the numerical solutions computed with P^3 and P^5 polynomials.

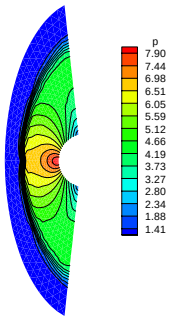


Figure 4. Computed pressure for inviscid $M_\infty = 2.0$ supersonic flow over a cylinder.

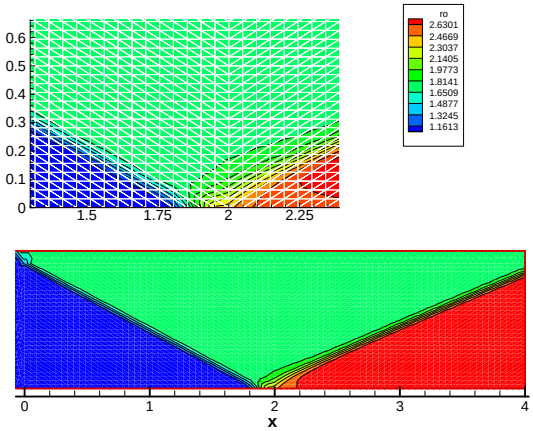


Figure 5. Computed Mach number for oblique shock reflection; $M_\infty = 3.9$.

Computations of viscous flows with third order polynomial expansions (fourth-order accurate solutions) were carried out for low Reynolds number, laminar, compressible viscous flows. It was pointed out in section 2 (see also remarks by Atkins and Shu[22]) that with the increase of the order of polynomial expansion basis of the DG discretizations the CFL stability reduces dramatically. As a result, with the available computing resources it was possible to demonstrate and validate the approach for the DG viscous term discretization only for solutions of simple, low Reynolds number viscous flows, which do not require very small grid spacing to accurately

capture the near wall steep gradients. Two representative examples are shown; (1) channel flow, (2) low Mach number laminar boundary layer that is compared with the Blasius incompressible flow solution.

The computed flowfield in the channel is shown in Fig. 6. A comparison of the computed velocity distribution with the analytic parabolic velocity distribution is shown in Fig. 7. In spite of the coarse mesh very good agreement of the solution computed with P^3 polynomial bases (for both inviscid and viscous fluxes) is obtained with the exact result. The laminar boundary layer flow at $\mathrm{Re} = 500$ was computed using three meshes. The coarse mesh contains 1350 triangular elements and the minimum length of the orthogonal triangle side (distance from the wall) is $\Delta y = h = 0.04$. The minimum distance from the wall of the medium and fine mesh is $\Delta y = h = 0.02$ and $\Delta y = h = 0.01$, respectively. The computed solution is shown in Fig. 8. A comparison of the coarse and fine grid solutions with the Blasius solution is shown in Fig. 9. Good agreement is obtained for the fine grid solution.

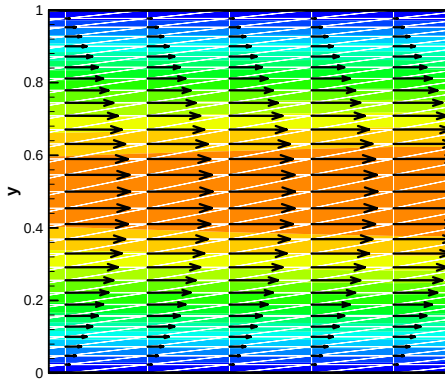


Figure 6. Computed velocity field for channel flow at $M_\infty = 0.3$.

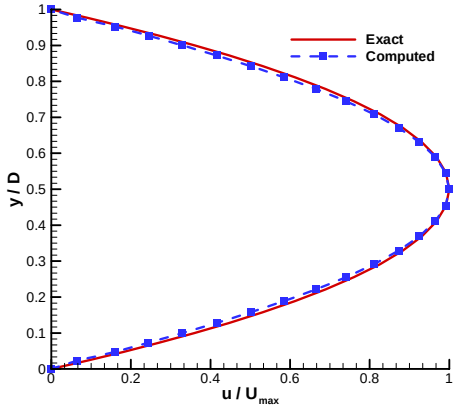


Figure 7. Comparison of the computed solution with the incompressible flow exact result.

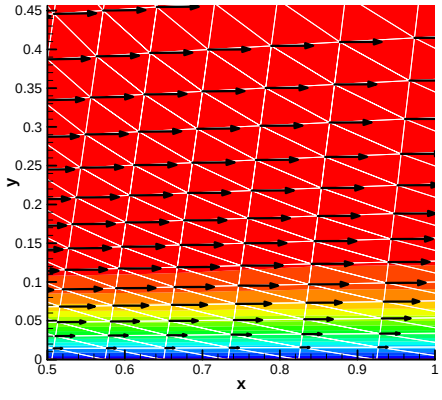


Figure 8. Computed velocity for zero pressure gradient flat plate boundary layer flow at $M_\infty = 0.3$.

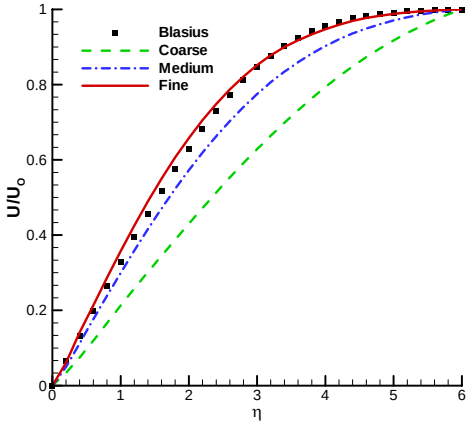


Figure 9. Comparison of the computed velocity with the Blasius exact result.

7 CONCLUSIONS

The DG finite element method was used for high order accurate discretization of the Euler and Navier-Stokes equations. Time marching was performed with the explicit, third order accurate, TVD-RK3 method. Supersonic shocks were captured using a TVB limiter with first-order polynomial expansions. Application of this limiter with higher order polynomial expansions is computationally intensive and alternative ways to suppress oscillations are currently investigated. The viscous terms in the Navier-Stokes equations were discretized in the DG framework using an auxiliary equation for the gradient of the conservative variables. Computed solutions for viscous and invicid flows were in good agreement with exact results. Severe time step limitations for high-order polynomial expansions with explicit time integration make high-order accurate

numerical solutions very intensive computationally and incorporation of implicit time marching is required for solutions of realistic viscous flow problems with the DG method.

References

[1] Lele, S.K., "Compact Finite Difference Schemes with Spectral-like Resolution," J. of Computational Physics, Vol. 103, No. 1, 1992, pp. 16-42.

[2] Tam, C.K. and Webb, J.C., "Dispersion-Relation Preserving Finite Difference Schemes for Computational Acoustics," J. of Computational Physics, Vol. 83, No. 1, 1993, pp. 262-281.

[3] Yee, H.C., Sandham, N.D., and Djomehri M. J., "Low-dissipative high-order shock capturing methods using characteristic-based filters," J. of Computational Physics, Vol. 150, No. 1, 1999, pp. 199-238.

[4] Gaitonde, D.V. and Visbal, M.R., "Pade-Type High-Order Boundary Filters for the Navier-Stokes Equations," AIAA Journal, Vol. 38, No. 11, 2000, pp. 2103-2112.

[5] Visbal, M.R. and Gaitonde, D.V., "On the Use of High-Order Finite-Difference Schemes on Curvilinear and Deforming Meshes," J. of Computational Physics, Vol. 181, 2002, pp. 155-185.

[6] Jiang, G.-S. and Shu, C.-W., "Efficient Implementation of Weighted ENO Schemes," J. of Computational Physics, Vol. 126, No. 1, 1996, pp. 202-228.

[7] Balsara, D.S. and Shu, C.-W., "Monotonicity Preserving Weighted Essentially Non-oscillatory Schemes with Increasingly High Order of Accuracy," J. of Computational Physics. Vol. 160, 2000, p. 405-452.

[8] Wang, Z.J. and Chen, R.F., "Optimized Weighted Essentially Non-oscillatory Schemes for Linear Waves with Discontinuity," J. of Computational Physics, Vol. 174, No. 1, 2001, pp. 381-404.

[9] Pirozzoli, S., "Conservative Hybrid Compact-WENO Schemes for Shock Turbulence Interaction," J. of Computational Physics, Vol. 178, No. 1, 2002, pp. 81-117.

[10] Ekaterinaris, J.A., "Performance of High-Order-Accurate, Low-Diffusion Numerical Schemes for Compressible Flow," AIAA Journal, Vol. 42, No. 3, 2004, pp. 483-500.

[11] Abgral, R., "On Essentially Non-Oscillatory Schemes on Unstructured Meshes. Analysis and Implementation," J. of Computational Physics. Vol. 114, No. 1, 1994, pp. 45-58.

[12] Cockburn, B, Karniadakis, G. E., and Shu, C.-W. (Editors) "Discontinuous Galerkin Methods," Springer, 1999.

[13] Yan, J. and Shu, C-W., "Local Discontinuous Galerkin Methods for Partial Differential Equations with Higher Order Derivatives," NASA/CR-2002-211959 and ICASE Report No. 2002-42, Nov. 2002.

[14] van der Veght, J.J.W. and van der Ven, H., Discontinuous Galerkin Finite Element Method with Anisotropic Local Grid Refinement for Inviscid Compressible Flow," J. of Computational Physics, Vol. 141, No. 1, 1998, pp. 46-77.

[15] Wang, Z.J. "Spectral (Finite) Volume Method for Conservation Laws on Unstructured Grids. Basic Formulation: Basic Formulation**,"** J. of Computational Physics, Vol. 178, No. 1, 1 2002, pp. 210-251, and "Spectral (Finite) Volume Method for Conservation Laws on Unstructured Grids: II. Extension to Two-Dimensional Scalar Equation," J. of Computational Physics, Vol. 179, No. 2, 2002, pp. 665-697.

[16] Rasetarinera, P. and Hussaini, M.Y., "An Efficient Implicit Discontinuous Spectral Galerkin Method," J. of Computational Physics, Vol. 172, No. 2, pp. 718-738.

[17] Bassi, F., Crivellini, A., Rebay, S., and Savini, M., "Discontinuous Galerkin Solutions of the Reynolds-Averaged Navier-Stokes and k-ω Turbulence model Equations," Computers and Fluids, 2004,to appear.

[18] Hu, F.Q., Hussaini, M.Y., and Rasetarinera, P., "An Analysis of the Discontinuous Galerkin Method for Wave Propagation Problems," J. of Computational Physics, Vol. 151, No. 2, 1999, pp. 921-946.

[19] Gottlieb, S. and Shu, C.W., "Total Variation Diminishing Runge-Kutta Schemes," Mathematics of Computation, Vol. 67, 1998, pp. 73-85.

[20] Cockburn, B., Hou, S., and Shu, C.W., "TVB Runge-Kutta Local Projection Discontinuous Galerkin Finite Element Method for Conservation Laws IV: The Multidimensional Case," Mathematics of Computation, Vol. 54, 1990, pp. 545-581.

[21] Cockburn, B., and Shu, C.W., "The Runge-Kutta Discontinuous Galerkin Method for Conservation Laws," J. of Computational Physics, Vol. 141, 1998, p. 199.

[22] Atkins H. and Shu, C.W., "Quadrature-free Implementation of the Discontinuous Galerkin Method for Hyperbolic Equations," AIAA Journal, Vol. 36, No. 6, 1998, pp. 775-782.

[23] Reed, W.H. and Hill, T.R., "Triangular Mesh Methods for the Neutron Transport Equation," Technical Report LA-UR-73-479, Los Alamos Scientific laboratory, 1973.

[24] Cockburn, B., and Shu, C.W., "Runge-Kutta Discontinuous Galerkin Method for Convection Dominated Problems," J. of Scientific Computing, Vol. 16, No. 3, 2001, pp. 173-261.

[25] Cockburn, B., Shu, C.W., "TVB Runge-Kutta Local Projection Discontinuous Galerkin Finite Element Method for Conservation Laws II: General Framework," Mathematics of Computation, Vol. 52, 1989, pp. 411-435.

[26] Zhang, M., and Shu, C.W., "An Analysis of Three Different Formulations of the Discontinuous Galerkin Method for Diffusion Equations," Mathematical Models and Methods in Applied Sciences, Vol. 13(3), 2003, 395-413.

[27] Bassi, F., Rebay, S., "A High-Order Accurate Discontinuous Finite Element Method for the Numerical Solution of the Compressible Navier-Stokes Equations," J. of Computational Physics, Vol. 131(1), 1997; pp. 267-279.

[28] Baumann, C.E., Oden J.T. "A Discontinuous hp Finite Element Method for Convection-Diffusion Problems," Computational Methods in Applied Mech. Eng. Vol. 175, 1999; pp. 311-341.

[29] van der Vegt, J.J.W., Van der Ven, H., "Space Time Discontinuous Galerkin Finite Element Method with Dynamic Grid Motion for Inviscid Compressible Flows: I. General Formulation," J. of Computational Physics, Vol. 182, No. 2, 2002, pp. 546-585.

A NEW ALGEBRAIC STRUCTURE-BASED TURBULENCE MODEL WITH PROPER HANDLING OF STRONG ROTATION

Stavros C. Kassinos*,† and Carlos A. Langer*

* Department of Mechanical and Manufacturing Engineering
University of Cyprus
P.O.BOX 20537, 1678 Nicosia, Cyprus
E-mail: kassinos@ucy.ac.cy

†Center for Turbulence Research
Stanford University, Stanford CA 94305-3030, USA

Keywords: Rotation, turbulence modeling, structure based, algebraic model, V2F, near wall.

Abstract. *A primary goal of turbulence modeling based on the Reynolds-Averaged Navier-Stokes equations is to determine the Reynolds stress tensor in order to close the turbulence problem at the mean velocity level. However, the Reynolds stresses alone do not characterize adequately the turbulence, especially in presence of rotation; the structure of the turbulence is also important. Hypothetical turbulent eddies are used to bring awareness of turbulence structure into the turbulence model. Averaging over an ensemble of eddies produces a set of one-point statistics, representative of the eddy field, and a set of equations of state relating the Reynolds stresses to these statistics. An algebraic model for the eddy statistics is constructed in terms of the local mean deformation normalized by a turbulence time scale. The algebraic model is closed with evolution equations for turbulent scales borrowed from the* v^2-f *[3] turbulence model (V2F), providing the turbulence time scale, the turbulent kinetic energy, and a velocity scale for the turbulent transport of the turbulent scales. It is shown that the algebraic structure-based evaluation of the Reynolds stresses can be used directly with a* conventional *model for the turbulence scales. The model is evaluated in spanwise rotating channel flow and in flat plate boundary layers.*

1 INTRODUCTION

Flow predictions have become a standard feature of modern flow system design. Where turbulence is important there is need to have a good model for the turbulent transport and for the turbulent stress, in order to produce adequate predictions of skin friction, flow separation, heat and mass transfer, and other flow features. As a result of the efforts of many contributors, turbulence models are now quite adequate for simple flows, but there remain important engineering problems where improved models are needed. For example, improved models for turbulence in rotating systems would enable better turbomachinery design. Linear eddy-viscosity models are known to be inaccurate in predicting the effect of strong streamline curvature and frame rotation. There is no shortage of modifications and adjustments proposed in the literature to correct their behavior. For example Shih et al.[1] modify the k–ε model by introducing coefficients in the ε-equation that depend on the shear rate and frame rotation. A more consistent redesigning of the ε-equation for flows with rotational effects has been proposed by Haire and Reynolds[2]. Another recent attempt by Durbin and Pettersson-Reif[3] consists in the modification of the eddy-viscosity coefficient (again by introducing dependency on the shear rate and frame rotation). In the latter case the justification for the choice of the selected functional dependency comes from the study of solutions of second-moment models in the case of homogeneous rotating shear. Although these modifications are shown to provide encouraging predictions for simple flows with rotation, their accuracy for more complex situations remains unclear. Differential Reynolds stress models, on the other hand, possess the obvious advantage that the turbulence production terms and the stress anisotropy are automatically accounted for. Unfortunately, the difficulties in modeling the stress redistribution terms and their inherent numerical stiffness make them not amenable to mainstream use in engineering calculations. Algebraic Reynolds stress models have received a substantial amount of attention given the potential benefit of introducing stress anisotropy in the *controlled* environment of an eddy-viscosity closure. Several models[4,5] have been devised with various degree of success. The basic idea behind these models is to express the Reynolds stress tensor as a function of one or more (up to ten) different tensors. This is not different from what is used to derive the so-called non-linear eddy-viscosity models where additional (high-order) terms are added to the Boussinesq relationship between mean strain and Reynolds stresses.

Reynolds and coworkers[6,7,8] have repeatedly argued that for adequate modeling and description of rotating turbulence, information about the turbulence structure is crucial. The Reynolds stresses only characterize the componentality of turbulence, that is, which velocity components are more energetic. The turbulent field has much more information than that contained in the Reynolds stresses, which is important in presence of rotation, and which is described by the turbulence structure. For instance, the dimensionality of the flow is important. This carries information about which directions are favored by the more energetic turbulent eddies: if the turbulent eddies are preferentially aligned with a given direction, then the dimensionality is smaller along that direction. In the Algebraic Structure-Based Model (ASBM), hypothetical turbulent eddies are used to bring awareness of turbulence structure into the turbulence model. Averaging over an ensemble of eddies produces a set of one-point statistics, representative of the eddy field, and a set of equations of state relating the Reynolds stresses to these statistics. The structure-based approach to build the Reynolds stress closure has lead to the development[9,10] of an ASBM closed with a *novel* two-equation model based on transport equations for the turbulent kinetic energy, k, and for the large-scale turbulent enstrophy $\tilde{\omega}^2$. The model has been calibrated for channel flow simulations and the results have shown excellent agreement with available direct numerical simulation (DNS) data.

The primary objective of this work was to explore the ASBM with transport equations for conventional turbulence scalars, namely the V2F model. Advantages of this set of scale equations are its availability in computational fluid dynamics (CFD) codes and the fact that it relies on a scalar diffusivity, v^2, for turbulent transport of the scalars, which eases the model's numerical implementation in complex geometries/flows. Results are presented for boundary layer flows, with and without pressure gradients, and for channel flow, with and without spanwise rotation.

2 THE STRUCTURE-BASED ALGEBRAIC STRESS MODEL

The eddy-axis concept[7] is used to relate the Reynolds stress and the structure tensors to parameters of a hypothetical turbulent eddy field. Each eddy represents a two-dimensional turbulence field, and is characterized by an eddy-axis vector, a_i. The turbulent motion associated with this eddy is decomposed in a component along the eddy axis, the jetal component, and a component perpendicular to the eddy axis, the vortical component. This motion can be further allowed to be flattened in a direction normal to the eddy axis (a round eddy being characterized by a random distribution of kinetic energy around its axis). Averaging over an ensemble of turbulent eddies gives statistical quantities representative of the eddy field, along with constitutive equations relating the normalized Reynolds stresses and turbulence structure to the statistics of the eddy ensemble:

$$\begin{aligned} r_{ij} = \frac{\overline{u_i' u_j'}}{2k} &= (1-\phi)\tfrac{1}{2}(\delta_{ij} - a_{ij}) + \phi a_{ij} \\ &+ (1-\phi)\chi[\tfrac{1}{2}(1-a_{nm}b_{nm})\delta_{ij} - \tfrac{1}{2}(1-a_{nm}b_{nm})a_{ij} - b_{ij} + a_{in}b_{nj} + a_{jn}b_{ni}] \\ &+ (-\gamma \Omega_k^T / \Omega^T)(\varepsilon_{ipr}a_{pj} + \varepsilon_{jpr}a_{pi})\{\tfrac{1}{2}[1-\chi(1-a_{nm}b_{nm})]\delta_{kr} + \chi b_{kr} - \chi a_{kn}b_{nr}\}. \end{aligned} \tag{1}$$

The eddy-axis tensor, $a_{ij} = \langle V^2 a_i a_j \rangle$, is the energy-weighted average direction cosine tensor of the eddy axes. The eddy-axis tensor is determined by the kinematics of the mean deformation. Eddies tend to become aligned with the direction of positive strain rate, and they are rotated kinematically by mean or frame rotation.

Motion around the eddy is called vortical, and motion along the axis is called jetal. The eddy jetting parameter ϕ is the fraction of the eddy energy in the jetal mode, and $(1-\phi)$ is the fraction in the vortical mode. Under irrotational mean deformation, eddies remain purely vortical $(\phi = 0)$. Shear produces jetal eddies, and in the limit of infinite rapid distortion $(\phi \to 1)$ for shear in a non-rotating frame. For shear in a rotating frame, ϕ ranges from 1 for zero frame rotation to 0 for frame rotation that exactly cancels the mean rotation in the frame, for which the mean deformation in an inertial frame is irrotational.

The eddy helix vector γ_k arises from the correlation between the vortical and jetal components. Hence $\gamma_k = 0$ for purely vortical turbulence $(\phi = 0)$ or for purely jetal turbulence $(\phi = 1)$. Typically γ_k is aligned with the total rotation vector Ω_k^T. The eddy-helix vector is the key factor in setting the shear stress in turbulent fields.

Flattening is used to describe the degree of asymmetry in the turbulent kinetic energy distribution around an eddy. A round eddy has no preferential distribution. If the motion is not axisymmetric around the eddy axis, the eddy is called flattened. The eddy-flattening tensor, b_{ij}, is the energy-weighted average direction cosine tensor

of the flattening vector. The intensity of the flattening is given by the flattening parameter, χ. Under rapid irrotational deformation in a fixed frame eddies remain axisymmetric. Rotation tends to flatten the eddies in planes perpendicular to the rotation direction.

Following Reynolds et al.[11], the eddy-axis tensor, a_{ij}, is computed on the analysis frame, where the turbulence might be at equilibrium or very close to it. The eddy-axis tensor is computed with no reference to the frame rotation, as it is only kinematically rotated by it[7,2]. The evaluation is divided in two parts. Initially a strained eddy-axis tensor, a_{ij}^s, is evaluated based on the irrotational part of the mean deformation. Next a rotation operation is applied, sensitizing the eddy-axis tensor to mean rotation. This procedure produces eddy-axis tensor states that mimic the limiting states produced under rapid distortion theory (RDT) for different combinations of mean strain with on-plane mean rotation, while guaranteeing realizability of the eddy-axis tensor.

The strained a_{ij}^s is given by

$$a_{ij}^s = \frac{1}{3}\delta_{ij} + \frac{(S_{ik}^* a_{kj}^s + S_{jk}^* a_{ki}^s - \frac{2}{3} S_{mn}^* a_{nm}^s \delta_{ij})\tau}{a_0 + 2\sqrt{\tau^2 S_{kp}^* S_{kq}^* a_{pq}^s}} \tag{2}$$

where $S_{ij}^* = S_{ij} - S_{kk}\delta_{ij}/3$ is the traceless strain-rate tensor with $S_{ij} = \left(\partial u_i/\partial x_j + \partial u_j/\partial x_i\right)/2$, τ is a time scale of the turbulence, Eq. (14), and $a_0 = 1.6$ is a 'slow' constant. This gives realizable states for the eddy-axis tensor under irrotational deformations.

The final expression for the homogeneous eddy-axis tensor, a_{ij} (for near-wall regions see Eq. (9)), is obtained by applying a rotation transformation to the strained eddy-axis tensor, a_{ij}^s,

$$a_{ij} = H_{ik} H_{jl} a_{kl}^s, \qquad H_{ij} = \delta_{ij} + h_1 \frac{\Omega_{ij}}{\sqrt{\Omega_{pp}^2}} + h_2 \frac{\Omega_{ik}\Omega_{kj}}{\Omega_{pp}^2}, \tag{3}$$

where $\Omega_{pp}^2 = \Omega_{pq}\Omega_{pq}$. The orthonormality conditions $H_{ik}H_{jk} = \delta_{ij}$ and $H_{ki}H_{kj} = \delta_{ij}$ require

$$h_1 = \sqrt{2h_2 - h_2^2/2}. \tag{4}$$

h_2 is determined with reference to RDT for combined homogeneous plane strain and rotation[11,2],

$$h_2 = \begin{cases} 2 - 2\sqrt{\frac{1}{2}(1+\sqrt{1-r})} & \text{if } r \le 1 \\ 2 - 2\sqrt{\frac{1}{2}(1-\sqrt{1-1/r})} & \text{if } r > 1 \end{cases}, \tag{5}$$

where $r = (a_{pq}\Omega_{qr}S_{rp}^*)/(S_{kn}^* S_{nm}^* a_{mk})$.

The flattening tensor b_{ij} is modeled in terms of the mean rotation rate vector, Ω_i, and the frame rotation rate vector, Ω_i^f,

$$b_{ij} = \frac{(\Omega_i + C_b\Omega_i^f)(\Omega_j + C_b\Omega_j^f)}{(\Omega_k + C_b\Omega_k^f)(\Omega_k + C_b\Omega_k^f)}, \quad C_b = -1.0 \tag{6}$$

The helix vector γ_k is taken as aligned with the total rotation vector,

$$\gamma_k = \gamma \frac{\Omega_k^T}{\sqrt{\Omega_p^T \Omega_p^T}}, \quad \gamma = \beta\sqrt{\frac{2\phi(1-\phi)}{1+\chi}}. \tag{7}$$

Modeling ϕ, β (see Eq. 7), and χ is a crucial part in the construction of the model. The equations for these scalars are found by analyzing target turbulent states corresponding to a mean deformation. Throughout the

model development there is a strong effort to make it consistent with RDT solutions, aiming to improve model dependability and realizability for a wide range of mean deformations, as well as to obtain guidance in the functional shape chosen for the structure parameters. Tentative functional forms for the structure parameters are thus chosen with reference to RDT. A set of parameter values is chosen to mimic the isotropic turbulent state (the eddy structure is expected to consist of axisymmetric $(\chi = 0)$, vortical $(\phi = 0)$ eddies). Finally interpolation functions (along with model constants) are chosen to bridge these limiting states (isotropy and RDT). They are selected specially to match a canonical state of sheared turbulence, observed in the log region of a boundary layer.

The structure scalars are parameterized in terms of η_m, η_f, and a^2, representatives of the ratio of the mean flow rotation to the mean flow strain, frame rotation to mean flow strain, and a measure of anisotropy respectively,

$$\eta_m \equiv \sqrt{\frac{a_{ij}\Omega_{ik}\Omega_{jk}}{a_{pq}S_{pr}S_{qr}}}, \qquad \eta_f \equiv \eta_m - \mathrm{sign}\left(a_{mn}\Omega^T_{ms}S_{sn}\right)\sqrt{\frac{a_{ij}\Omega^T_{ik}\Omega^T_{jk}}{a_{pq}S_{pr}S_{qr}}}, \qquad a^2 \equiv a_{pq}a_{pq}, \tag{8}$$

η_m, η_f are weighted by the eddy-axis tensor, allowing the model to satisfy the principle of material frame indifference to rotation, when appropriate.

As a no-slip wall is approached, the velocity is driven to zero through the action of viscous forces. Furthermore, the velocity vector is reoriented into planes parallel to the wall through an inviscid mechanism (wall blocking) which acts over distances far larger than the viscous length scale. Thus the velocity component normal to the wall is driven to zero faster than the tangential components. In the structure-based model it is postulated that the eddy orientation shall also be parallel to the wall. A wall-blocking procedure is then introduced to reorient the eddies into planes parallel to the wall. The structure parameters are also sensitized to wall blocking, such that the modeled Reynolds stresses are consistent with the expected near wall asymptotic behavior.

As in Reynolds et al.[11], the homogeneous eddy-axis tensor, a^h_{ij}, is computed based on the homogeneous algebraic procedure, Eqs. (2, 3) (note that the superscript 'h' has been added in the current section). It is then partially projected onto planes parallel to the wall,

$$a_{ij} = H_{ik}H_{jl}a^h_{kl}, \qquad H_{ik} = \frac{1}{D_a}(\delta_{ik} - B_{ik}), \qquad D_a^2 = 1 - (2 - B_{kk})a^h_{kl}B_{nm} \tag{9}$$

where H_{ik} is the partial-projection operator, and D_a^2 is such that the trace of a_{ij} remains unity. The blockage tensor H_{ik} gives the strength and the direction of the projection. If the wall-normal direction is x_2, then B_{22} is the sole non-zero component, and varies between 0 (no blocking) far enough from the wall, to 1 (full blocking) at the wall. B_{ij} is computed by

$$B_{ij} = \frac{\Phi_{,i}\Phi_{,j}}{\Phi_{,k}\Phi_{,k}}\Phi, \quad \text{if} \quad \Phi_{,k}\Phi_{,k} > 0\,. \tag{10}$$

If all gradients of Φ vanish, B_{ij} is computed from an average over surrounding points. The blocking parameter, Φ, is computed by an elliptic relaxation equation

$$L^2\frac{\partial^2\Phi}{\partial x_k \partial x_k} = \Phi, \quad L = C_L\,\mathrm{Max}\left(\frac{k^{3/2}}{\varepsilon}, C_\nu\sqrt[4]{\frac{\nu^3}{\varepsilon}}\right). \tag{11}$$

with $\Phi = 1$ at solid boundaries, and $\Phi_{,n} \equiv \partial\Phi/\partial x_n = 0$ at open boundaries, where x_n is the direction normal to the boundary. The definition of L is inspired by Durbin and Pettersson-Reif[3]. Here $C_\nu = 50$, and

$$C_L = 1.0\frac{S\tau}{S^2\tau^2 + 15.} \tag{12}$$

with $S^2 = 2S_{ij}S_{ji}$. This form is chosen so as to limit the growth of L in rotating flows, when ε decreases substantially. An overgrown L would enforce too much blocking on the turbulence structure over too much of

the flow. An alternative solution[12] would be to add to ε the viscous dissipation. This would in fact again limit the decay of ε near a stable wall in rotating flows.

To recover proper asymptotic behavior of the Reynolds stresses, $r_{12} \propto O(x_2)$ and $r_{22} \propto O(x_2^2)$, as the wall at $x_2 = 0$ is approached, the homogeneous jetal, ϕ^h, and helix, γ^h, parameters are modified using

$$\phi = 1 + (\phi^h - 1)(1 - B_{kk})^2, \quad \gamma = \gamma^h (1 - B_{kk}). \tag{13}$$

A consequence of this approach is that realizability is automatically satisfied for r_{ij}.

3 SCALES AND NUMERICAL IMPLEMENTATION

In Langer and Reynolds[9] the model was closed with two scalar transport equations, for the turbulent kinetic energy, k, and the large-scale turbulent enstrophy, $\widetilde{\omega}^2$. These make use of the structure information made available by the ASBM. They provide the energy to scale the normalized stresses that come out of the ASBM, a length scale needed to determine the blocking of the stresses, and also a time scale, τ, used to normalize the mean flow strain and rotation rates, as well as the frame rotation rate, which are fed to the ASBM. Here these scale equations are substituted by a set of V2F scale equations[13], not reproduced here. Besides the turbulent kinetic energy and a blocking length scale, these scalars also provide a velocity scale, v^2, for the turbulent transport term present in the scale equations, and again a time scale, used to normalize strain and rotation rates fed to the ASBM. The time scale is given by:

$$\tau^2 = \left(\frac{k}{\varepsilon}\right)^2 + \left(2.0\sqrt{\frac{\nu}{\varepsilon}}\right)^2. \tag{14}$$

We consider the V2F scale equations here as an early simplification in the model development. The V2F equations are more readily available in existing 3D multi-purpose codes, have been well tested and developed, and provide a simple turbulent transport model. In the original model, the turbulent transport followed a tensorial model, which could be challenging to implement in an implicit solver.

The mean momentum equations receive the gradient of Reynolds stresses returned by the model, $R_{ij} = 2kr_{ij}$. In eddy-viscosity models it is trivial to treat the Reynolds stress gradient term implicitly, which is important for stability. With the ASBM the Reynolds stresses are computed with an algebraic procedure and added explicitly to the momentum equations. For improved numerical stability, the Reynolds stress gradient is treated in a manner akin to the deferred correction approach[14], where n represents the iteration level

$$\frac{\partial}{\partial x_j}(R_{ij}) = \frac{\partial}{\partial x_j}\left(\nu_T \frac{\partial U_i^{n+1}}{\partial x_j}\right) - \frac{\partial}{\partial x_j}\left(\nu_T \frac{\partial U_i^{n}}{\partial x_j} - R_{ij}^{n}\right), \tag{15}$$

and ν_T is the turbulent viscosity as provided by the V2F model. Upon convergence the terms involving the turbulent viscosity cancel out and only the Reynolds stress gradient remains.

4 NUMERICAL RESULTS

We first compare the model with two flat-plate boundary layers: a zero pressure gradient (ZPG) boundary layer at $\mathrm{Re}_\theta = 1410$ [15], and an adverse pressure gradient (APG) boundary layer[16]. Figures (1) and (2) show comparisons of the ASBM combined with the V2F scale equations for the ZPG and for the APG boundary layers, respectively. Comparisons are made using wall units, and in the APG case, only one station is reported, 2/3 of the way through what Spalart and Watmuff[16] call the 'comparison region', $x = 0.80$. The mean velocity is very well captured in the ZPG case, while the agreement is not so good in APG case. The discrepancy in the value of the free stream velocity indicates an underprediction of the skin friction coefficient. The anisotropies of the turbulence intensities predicted by the model are in good agreement with the DNS.

For a fully developed channel flow undergoing orthogonal mode rotation, the mean continuity equation is automatically satisfied $(V = \partial U / \partial x = 0)$. The momentum equation in the streamwise direction becomes a balance between the (modified) pressure gradient and the viscous and turbulent fluxes. The presence of frame rotation in this particular configuration allows the investigation of the model's response to rotation in perhaps the most critical fashion, when all secondary effects are solely due to the turbulence. In more complex flows, other terms representative of the mean motion (convective, Coriolis, body force, and especially pressure gradient

terms) can also be active in the momentum equations, sharing some of the job of representing distinct physical phenomena. Then inaccuracies in the turbulence model itself may be overshadowed due to smaller importance of the Reynolds stress in the momentum equations.

Results shown here correspond to a family of pressure-driven fully-developed channel flows, in presence of spanwise frame rotation, aligned with the mean flow vorticity. The mean flow is given by $U_i = \{U(y),0,0\}$ in a coordinate system $x_i = \{x, y, z\}$ where y is the wall normal, x is the streamwise direction with streamwise

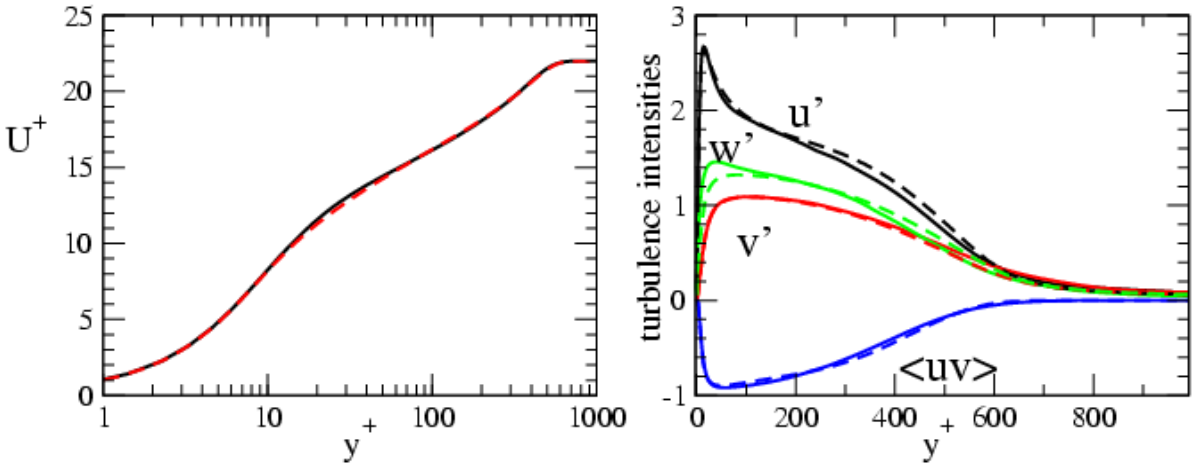


Figure 1. Mean velocity and turbulence intensities for ZPG boundary layer. Solid lines: DNS[15], Dashed lines: model

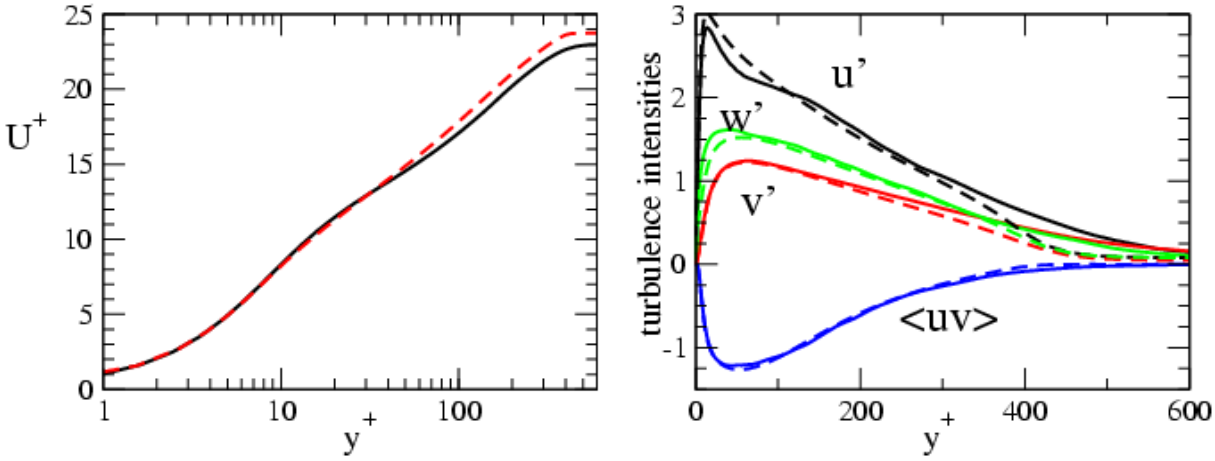


Figure 2. Mean velocity and turbulence intensities for APG boundary layer, at x=0.80. Solid lines: DNS[16], Dashed lines: model.

velocity U, and z is the spanwise direction. The frame rotation rate vector is given by $\Omega_i^f = \{0,0,\Omega_3^f\}$. The wall-normal mean velocity necessarily vanishes by continuity for a fully developed channel flow with zero velocity at the walls. The solutions depend on two parameters; the friction Reynolds number, $\mathrm{Re}_\tau = u_\tau h/\nu$, and the rotation number, $Ro = \Omega^f 2h/U_b$, where Ω^f is the magnitude of the frame rotation rate, h is the half height of the channel and U_b is the bulk velocity in the channel. For the fully developed rotating channel flow the friction velocity can be defined in terms of the streamwise pressure gradient, $u_\tau^2 = -hd(P/\rho)dx$.

Figure 3 corresponds to a set of fixed frame channel flows. On the left are mean velocity profiles in wall coordinates (normalized by the wall shear stress and viscosity) for a series of friction Reynolds numbers. Comparisons are made with DNS[17] at $\mathrm{Re}_\tau = \{180,395,590\}$. Two distinct log laws[18,19] are also shown. On the center are the turbulence intensities for $\mathrm{Re}_\tau = 590$. The anisotropy predicted in the log region is a testament to the accuracy of the ASBM in this case. On the right are the structure-dimensionality components for this case.

Of note is the d_{11} component. It is the smaller component indicating structures preferentially aligned with the x-direction. Furthermore it shows a minimum near the wall, where near-wall streaks aligned with the flow direction have come to be expected.

Figure 4 corresponds to a set of channel flows in a rotating frame, where Ω_3^f is aligned with the mean flow vorticity. The ASBM is compared against DNS[20]. On the left are mean velocity profiles normalized by the bulk velocity at $\mathrm{Re}_\tau = 360$. With frame rotation, $Ro = 0.22$, the velocity profile becomes asymmetric about the centerline of the channel. In spanwise frame rotation, the Coriolis terms drop out of the mean flow equations making them insensitive to direct effects of frame rotation. The mean flow asymmetry is then a secondary effect due to the effect of the frame rotation on the Reynolds stresses and turbulence structure. In the center the ASBM

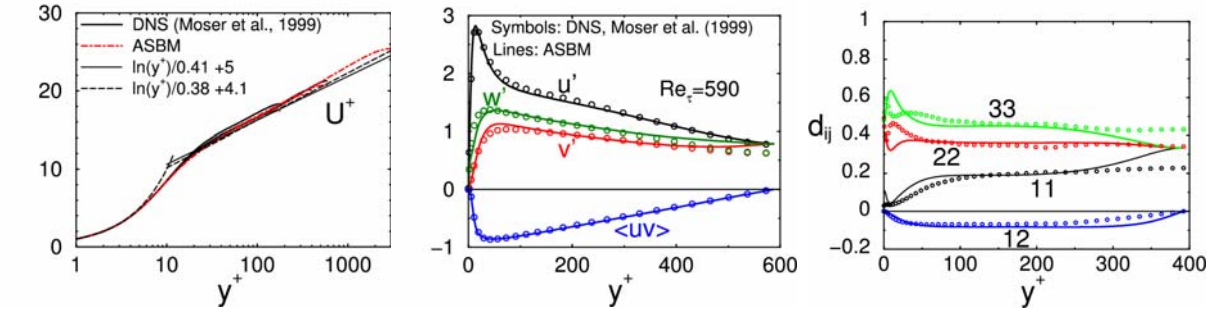


Figure 3. From left to right. Mean velocity, turbulence intensities and dimensionality tensor components for fixed-frame fully-developed channel flow. Symbols: DNS[17], Lines: model, correlations[18,19].

turbulence intensities are compared to the DNS results. The ASBM captures the anisotropy of the Reynolds stresses and its dependence on frame rotation. Notice in particular the fact that the wall-normal intensity, v', outgrows the streamwise intensity u' in the core region of the channel. In the right are profiles of the structure dimensionality tensor. Comparing with Figure (3) it is clear that the dimensionality is little affected by the frame rotation. It does display an asymmetry, but this results directly from the asymmetry in the mean velocity gradient. There are no dramatic changes as in the Reynolds stresses.

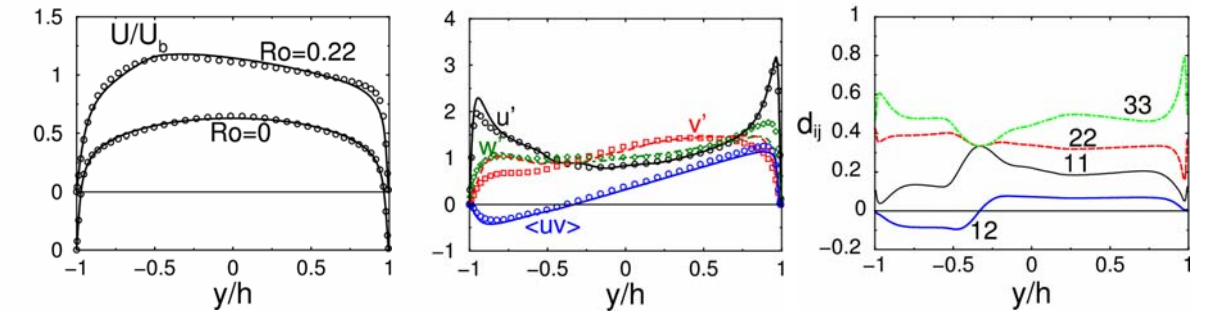


Figure 4. From left to right. Mean velocity, turbulence intensities and dimensionality tensor components for spanwise-rotating fully-developed channel flow. Symbols: DNS[20], Lines: model.

4 CONCLUSIONS

A new algebraic structure-based model has been presented as an alternative for the engineering analysis of complex flows. The results presented here demonstrate (i) the capability of the ASBM, coupled to the V2F equations, to be integrated directly to the wall, (ii) the response of the model combination to boundary layer flows and to situations where strong rotation is present, and where the turbulence is the sole responsible for the secondary effects observed, and (iii) the possibility of coupling the ASBM with the conventional scale equations, readily available in CFD packages.

REFERENCES

[1] Shih, T. H., Liou, W. W., Shabbir, A. and Zhu, J., (1995), A new k–ε eddy-viscosity model for high Reynolds number turbulent flows, *Comp. Fluids*, vol. 24, pp. 227-238.

[2] Haire, S. L. and Reynolds, W. C., (2003), Toward an Affordable Two-Equation Structure-Based Turbulence Model, *Technical Report No. TF-84*, Mech. Eng. Dept., Stanford Univ.

[3] Durbin, P. A. and Pettersson Reif, B. A. (2001), *Statistical Theory and Modeling for Turbulent Flows*, John Wiley and Sons.

[4] Gatski, T. B. and Speziale, C. G. (1993), On explicit algebraic stress models for complex turbulent flows, *J. Fluid Mech.*, vol. 254, pp. 59-78.

[5] Wallin, S. and Johansson, A. V., (2002), Modelling streamline curvature effects in explicit algebraic Reynolds stress turbulence models, *Int. J. Heat Fluid Flow*, vol. 23, no. 5, pp. 721-730.

[6] Reynolds, W. C., (1991), Towards a Structure-Based Turbulence Model, In: T. B. Gatski, S. Sarkar, and C. G. Speziale (Eds.), *Studies in Turbulence*, Lumley 60th birthday Symposium, pp. 76-80, Springer-Verlag, New York.

[7] Kassinos, S. C. and Reynolds, W. C., (1994), A Structure-Based Model for the Rapid Distortion of Homogeneous Turbulence, *Technical Report TF-61*, Mech. Eng. Dept., Stanford Univ.

[8] Kassinos, S. C., Reynolds, W. C. and Rogers, M. M., (2001), One-Point Turbulence Structure Tensors, *J. Fluid Mech.*, vol. 249, pp. 337-71.

[9] Langer, C. A. and Reynolds, W. C., (2003), A New Algebraic Structure-Based Turbulence Model for Rotating Wall-Bounded Flows, *Technical Report No. TF-85*, Mech. Eng. Dept., Stanford Univ.

[10] Reynolds, W. C., Langer, C. A. and Kassinos, S. C., Rogers, M. M., (2002), Structures and Scales in Turbulence Modeling, *Phys. Fluids*, vol. 14, no. 7, pp. 2485-2492.

[11] Reynolds, W. C., Kassinos, S. C., Langer, C. A. and Haire, S. L., (2000), New Directions in Turbulence Modeling, Presented at the *Third International Symposium on Turbulence, Heat, and Mass Transfer*, Nagoya, Japan, April 3-6.

[12] Pettersson, B. A. and Andersson, H. I., (1997), Near-wall Reynolds-stress modelling in noninertial frames of reference, *Fluid Dynamics Research*, vol. 19, pp. 251-276.

[13] Lien, F. S. and Durbin, P. A., (1996), Non-linear k - ε - v^2 modeling with application to high-lift, *CTR summer proceedings*.

[14] Ferziger, J. H. and Perić, M. (2002), *Computational Methods for Fluid Dynamics*, 3rd ed., Springer-Verlag.

[15] Spalart, P. R., (1988), Direct numerical study of a turbulent boundary layer up to $\mathrm{Re}_\theta = 1410$, *J. Fluid Mech.*, vol. 187, pp. 61-98.

[16] Spalart, P. R. and Watmuff, J. H., (1993), Experimental and numerical investigation of a turbulent boundary layer with pressure gradients, *J. Fluid Mech.*, vol. 249, pp. 337-371.

[17] Moser, R., Kim, J. and Mansour, N. N., (1999), Direct Numerical simulation of turbulent channel flow up to $\mathrm{Re}_\tau = 590$, *Phys. Fluids*, vol. 11, no. 4, pp. 943-945.

[18] Coles, D. E. and Hirst, E. A., (1969), Memorandum on Data Selection, *Proceedings 1968 AFOSR-IFP-Stanford Conference on Computation of Boundary Layers*.

[19] Osterlund, J. M., Johansson, A. V., Nagib, H. M. and Hites, M. H., (2000), A Note on the Overlap Region in Turbulent Boundary Layers, *Phys. Fluids*, vol. 12, no. 1, pp. 1-4.

[20] Alvelius, K., (1999), Studies of Turbulence and its Modeling through Large Eddy- and Direct Numerical Simulation, Ph.D. Thesis, Dept. Mech. Eng., KTH, Stockholm, Sweden.

A NUMERICAL SIMULATION METHODOLOGY FOR HYDRAULIC TURBOMACHINES

John S. Anagnostopoulos

Department of Mechanical Engineering / Fluids Section
National Technical University of Athens
9 Heroon Polytechniou ave., Zografou, 15780 Athens, Greece
e-mail: j.anagno@fluid.mech.ntua.gr

Keywords: hydraulic turbomachines, numerical simulation, Cartesian grid, centrifugal pump impeller, design optimization.

Abstract. *A numerical algorithm based on the control volume approach is developed for the simulation of the flow in a 2-dimensional centrifugal pump impeller. The flow domain is discretized with a polar, Cartesian mesh that covers a periodically symmetric section of the impeller. Advanced numerical techniques for adaptive grid refinement and for the treatment of cells that contain both fluid and solid regions are implemented in order to achieve a fully automated grid construction for any number of blades or blade geometry. The process is verified by a precision check of the numerical results. Then, a methodology is developed to extract the characteristics curves of a pump using the numerical results for the impeller section only. This also needs an estimation of the additional losses through the casing and the inlet and outlet sections of the pump. The regulation of the corresponding loss coefficients is done for a commercial pump, for which there are available measurements with both the original and a modified design impeller (different blade number and geometry) in place. Finally, the completed evaluation algorithm is used to find the optimum impeller geometry that maximizes the pump best efficiency point, using as design variables the blade angles at the leading and the trailing edge. The results show that the developed methodology, which is easily extensible to 3-dimensions, can be used for performance prediction and design optimization in hydraulic turbomachines.*

1 INTRODUCTION

The numerical simulation of the fluid flow for the design of hydraulic turbomachinery has become a requisite tool in order to increase efficiency and reduce cavitation. However, in spite of the continuous increase in computing power, the inverse design numerical optimization is still a laborious task, because it needs a large number of flow field evaluations. Such computations may be very costly, especially when the entire 3D domain in both the impeller/runner and the casing of the machine are simulated, and many design variables are incorporated. For this reason, few real 3D inverse design methods have been developed, as the inverse time marching method,[1] the pseudo-stream function method,[2] and the Fourier expansion singularity method.[3] These methods are very time consuming and exhibit some difficulties in correlating the design parameters with the blade geometry (the first two) or convergence problems (the latter). A quasi-3D method is recently proposed,[4, 5] which performs a blade-to-blade solution and saves computer time by using only one representative hub-to-shroud surface. All the above models are based on the inviscid simplified assumption. The application of fully 3-dimensional turbulent flow analysis tools for the impeller/runner design increases in the last years,[6–9] whereas the use of Navier-Stokes validation in inverse design optimization methods is still not a common practice, since in addition to the time-consuming calculations, there is a need for automated mesh generation in complex geometries. Some recently developed models are based on this advanced approach.[10,11]

The computer time needed by these models depends strongly on the generation cost of the body-fitted grid, as well as on the grid quality. An alternative practice in complex domains is the use of Cartesian grids that need much reduced construction effort. The main drawback of these grids is the inability to fit the grid lines to a sloped boundary. Several numerical techniques have been developed to improve the accuracy in such regions, globally classified in cell-cut and immersed boundary methods.[12,13]

In the present work such advanced numerical techniques developed for the automatic generation of Cartesian grids in irregular geometries are incorporated in a computer algorithm for the simulation of a centrifugal pump operation. The optimization algorithm has been recently tested for laminar flows in simple geometries.[14] In order to accelerate the shape optimization process, the computational domain contains only the pump impeller in a 2-dimensional approach, while special modeling is applied to produce the characteristic curves of the whole pump. With this methodology the cost per evaluation is much reduced compared to a fully 3D simulation of the

pump, giving thus a quick and reasonable estimation of the optimum values of the design parameters.

2 THE NUMERICAL METHODOLOGY

For the simulation of the flow in a 2-dimensional pump impeller the incompressible Navier-Stokes equations are expressed in polar coordinates and for a rotating with the impeller system. The latter, in conjunction with a circular inner and outer boundaries used here, results in a steady flow field throughout the domain, hence the governing equations for a horizontal space (no gravity) become:

Continuity: $$\nabla \cdot \vec{w} = 0 \tag{1}$$

Momentum: $$\vec{w} \cdot \nabla \vec{w} = -2\vec{\omega} \times \vec{w} - \vec{\omega} \times (\vec{\omega} \times \vec{w}) - \frac{\nabla p}{\rho} + \nu \nabla^2 \vec{w} \tag{2}$$

where $\vec{w}$ is the fluid velocity in the rotating system (relative fluid velocity), ω is the angular rotation speed of the impeller, and p, ρ, ν are the fluid pressure, density and kinematic viscosity, respectively. The system of the averaged form of the above equations, along with the k-ε turbulence model, is numerically solved with the finite volume approach and a collocated grid arrangement, using a preconditioned bi-conjugate gradient (Bi-CG) solver. Setting cyclic boundary conditions the solution can be restricted to the $1/z$ part of the impeller, where z is the number of blades.

2.1 Numerical grid

As stated in the Introduction, the use of Cartesian grids for numerical design optimization provides the significant advantage of a fast and automated grid generation process. Other desirable features include the easiness in the construction and control of locally or adaptively refined regions, as well as the capability to use discretization schemes of higher, in general, accuracy, compared to other grid types. All the above features are recently developed and incorporated in the numerical algorithm, which is used for the present study after transporting to polar, rotating coordinates. The numerical technique, more details of which can be found in Anagnostopoulos,[15] introduces a multiple stencil that allows the application of second order discretization schemes to any grid cell, regardless of its refinement ratio or local grid topology, therefore it is applicable not only to refined but also to completely unstructured Cartesian grids. Moreover, in spite of the gain in accuracy, the resulting expressions remain simple and robust, and for some schemes become even simpler than the original ones.[15] Finally, this method can be easily extended to 3-dimensions. An indicative picture of such a 2-dimensional grid, adaptively refined near the blades of a centrifugal impeller, is shown in Figure 1. For a given accuracy of the results, it was found that the above technique achieves considerably greater savings in computer cost than each of the local grid refinement and the higher-order discretization methods alone.[15]

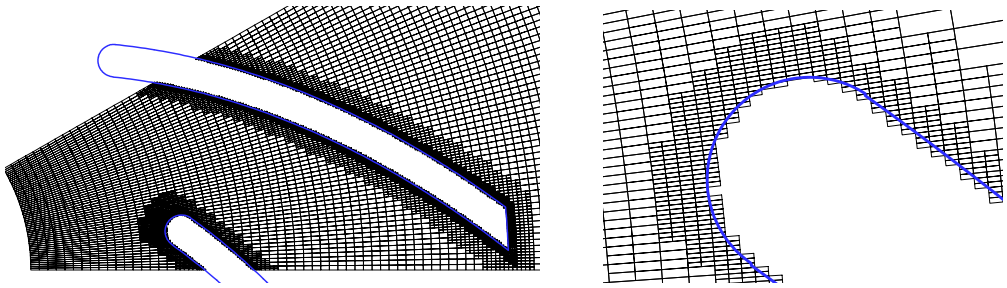


Figure 1. Indicative grid for a 2D pump impeller and detailed view at the blade leading edge.

2.2 Geometry representation

A cell-cut, sharp-interface grid construction method, developed and tested with success in various applied studies in the past,[16] is now modified and further improved in order to increase its accuracy from first order to almost second order near irregular boundaries. With this new method no cell-merging is performed, but all the grid cells that are totally or even partly filled with fluid are solved using the same general equation of the following linearized form:

$$(A_P - \gamma_V S_P)\Phi_P = \sum_i \gamma_i A_i \Phi_i + \gamma_V S_U, \quad A_P = \sum_i \gamma_i A_i, \quad i = E, W, N, S, U, D \tag{3}$$

where A_i are the coefficients linking the dependent variable Φ_P with its neighbors on the adjacent grid volumes,

and S_U, S_P are additional source terms. The geometric coefficients γ_i and γ_v represent the free portion (not blocked by the solid boundary) of the cell faces and volume, respectively.[16]

The fluid variables (velocities, pressures, etc) are computed at the centroid of the Cartesian cells, which for a partly filled cell does not coincide with its geometric centre, as shown in Figure 2. For this reason, special stencils are introduced to compute the cell-face values and the gradients of the flow variables, making a compromise between simplicity and accuracy, which results in cost-effective relations with almost second-order accuracy. The additional terms are included in the coefficients of the general equation (3), whereas all the needed geometric quantities are computed by a preprocessing algorithm. As a result, after defining the geometry of the computational domain, the grid construction process can be performed in a fast and fully automated way. Wall boundary conditions are also set automatically to every boundary cell (e.g. cells P_1 to P_4 in Fig. 2).

The above partly-filled-cells (PFC) method preserves the accuracy of the boundary representation and retains the conservation property. Moreover, it was not found to affect the stability of the solution algorithm, while its simplicity makes it easily applicable to both 2D and 3D complex geometries. An indicative picture of such a grid can be observed in Figure 1, where only the solvable cells are drawn.

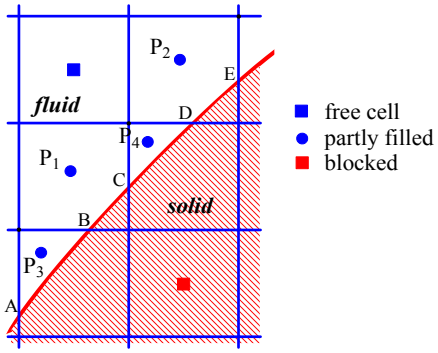


Figure 2. Treatment of partly filled grid cells.

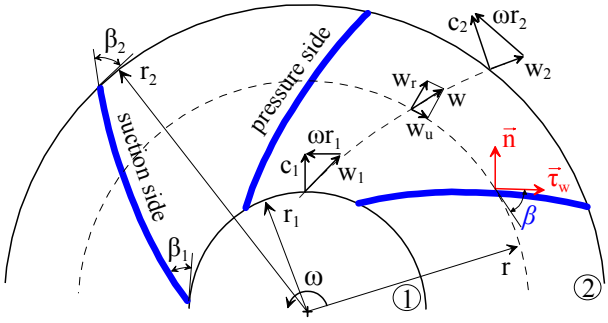


Figure 3. Sketch of a centrifugal pump impeller.

2.3 Impeller head and power calculations

Using a convergent flow field, the net energy added to a unit mass of fluid by the impeller can be calculated after computing the total energy of the fluid at the impeller inlet and exit, as shown in Figure 3. Hence the fluid head H_{12} is obtained from the flux-weighted relation:

$$H_{12} = H_2 - H_1 = \frac{1}{Q_u} \cdot \int \left(\frac{p_2 - p_1}{\rho g} + \frac{c_2^2 - c_1^2}{2g} \right) dq \tag{4}$$

where c is the absolute velocity of the fluid, Q_u the flow rate through the impeller and g the gravity acceleration, while the subscripts 1 and 2 denote impeller inlet and exit conditions, respectively. The integration is approximated by a summation over the radial flow rates δq at all the grid cells facing the inlet or the exit circumference of the impeller. On the other hand, the power absorbed from the impeller, N_u, can be calculated from the torque M_u on the blades:

$$N_u = \rho \cdot g \cdot Q_u \cdot H_u = \omega \cdot M_u = \omega \cdot \int_{r_1}^{r_2} \left[\left(\vec{r} \times \vec{n} \right) \cdot p + \left(\vec{r} \times \vec{\tau}_w \right) \cdot cot\, \beta \right] \cdot b \cdot dr \tag{5}$$

where $\vec{n}$ the unit vector normal to the blade surface, $\vec{\tau}_w$ the wall shear stress, β the blade angle and b the impeller width, whereas the integration covers both the pressure side and the suction side of the blade (Fig. 3). Although the simulation is 2-dimensional, the impeller width b is a function of radius r, as in the real impeller.

2.4 Pump characteristic curves

Although the simulation is restricted to the impeller geometry, which furthermore is considered 2-dimensional, the numerical results can be used as a basis to estimate the performance of the entire pump. To do this, the additional hydraulic losses are properly expressed and abstracted from the head results, according to the following analysis (Figure 3). The effect of the impeller shroud and hub surfaces can be computed using Darcy's law, from the relation:

$$dh_1 = \int_{r_1}^{r_2} \lambda \frac{dr}{D_h} \cdot \frac{w_r^2 + w_u^2}{2g} \tag{6}$$

where λ is the friction coefficient, w_r and w_u are the flow radial and (relative) tangential velocity components, and $D_h \approx 2b$ is the hydraulic diameter, with b the impeller width.

At the exit of the real pump impeller the flow enters into the spiral volute and decelerates. The sudden expansion losses are proportional to the absolute exit velocity of the fluid, namely:

$$dh_2 = k_2 \cdot c_2^2 \tag{7}$$

The above Eqs. (6) and (7) are flux weighted along the impeller periphery in order to compute the corresponding average values.

Away from the design flow rate conditions additional losses appear at the suction side, due to the impeller incidence and the inlet pipe recirculation, as well as at the volute tongue. All these losses can be included in an approximate expression of the type:

$$dh_3 = k_3 \cdot \left(1 - Q/Q_0\right)^2 \tag{8}$$

where Q_0 is the design flow rate. The rest hydraulic losses in the inner and outer pump section, as well as in the spiral volute, can be taken proportional to the pump flow rate Q:

$$dh_4 = k_4 \cdot Q^2 \tag{9}$$

Using the above relations, the pump head H can be estimated by abstracting from the computed impeller head H_u (Eq. 4) all the above losses:

$$H = H_{12} - dh_1 - dh_2 - dh_3 - dh_4 \tag{10}$$

The mechanical losses at the shaft bearings, along with disk friction power, can be taken into account by increasing the computed impeller power N_u (Eq. 5):

$$N = N_u / \eta_m \tag{11}$$

and the volumetric losses due to leakage flow are expressed as:

$$\eta_Q = Q / Q_u \tag{12}$$

Hence, the pump efficiency can be finally computed from the relation:

$$\eta = \frac{\rho \cdot g \cdot H \cdot Q}{N} = \frac{H}{H_u} \cdot \eta_Q \cdot \eta_m \tag{13}$$

In order to use the above relations, the values of the coefficients k_2 to k_4 and the efficiency factors η_Q and η_m must be known, otherwise they can be regulated using experimental or statistical data.

2.5 Numerical optimization

In order to find the combination of the impeller design variables that maximizes the target value, an optimization algorithm is developed based on the unconstrained gradient approach. This selection is made after some preliminary numerical tests showed that the cost function (here the pump efficiency) does not exhibit local maxima outside a global maximum region. However, this is not an analytic function, and hence a problem of non-continuity and scattering arises. The algorithm is specially designed to operate even for such discrete data, using a varied with trial-and-error step along the gradient direction. Also, the gradients are computed using forward finite differences at the beginning, and central differences when the cost function approaches maximum, and with a variable step size. The algorithm converges very fast within the region where the cost function maximizes, although due to scattering it cannot always find the absolute maximum. However, the fast performance allows repeating the calculations from different starting values, to verify or correct the resulting optimum.

3 RESULTS

3.1 Accuracy and precision checks

The accuracy of the representation of the blade geometry with the PFC method was tested first, along with

the grid dependency of the flow field results. This was performed using the base-case geometry of Figure 4, that is a periodically symmetric section of the centrifugal impeller (z = 9 blades). Then, the impeller is turned step-by-step at a small fraction (1/20) of the tangential width of the boundary cells, resulting in a number of different grid configurations at the blade boundary line. The scattering in the corresponding flow field results is due to the numerical error introduced by the PFC method.

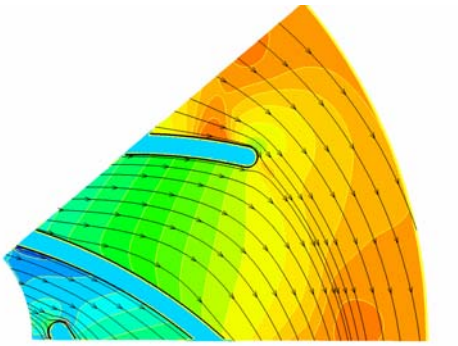


Figure 4. Pressure contours and flow streamlines in the computational domain.

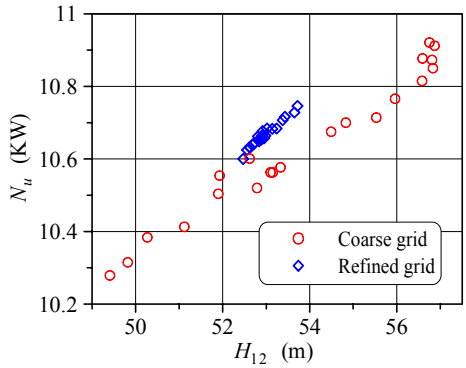


Figure 5. Predicted values of the fluid head and the impeller power.

The computed values for the fluid head and impeller power (Eqs. 4 and 5) are concentrated in Figure 5, for two grids: a coarse (8000 nodes) and a finer (27000 nodes) that has two refined layers around the blade, as in the example of Figure 1. The precision of the refined grid results is satisfactory, since the mean and maximum deviation from the mean values are of the order of 0.5% and 1% respectively. These are about one fourth of the coarse grid corresponding ones, confirming that the PFC method preserves the accuracy of the discretization scheme.

On the other hand, the differences in the mean values between the two grids represent the grid-dependency of the results, and they are again of the order of 1%, which is an adequate accuracy. Consequently, the refined grid is selected to be used for the rest calculations.

3.2 Regulation of the model

The adjustable coefficients involved in the model Eqs. (6) to (13) are regulated using the characteristic curves of a commercial centrifugal pump operating with a new impeller, constructed in the Laboratory. The impeller has 9 two-dimensional (non-twisted) blades with inlet and exit diameter D_1=70 mm, D_2=190 mm, exit width b_2=9 mm, and inlet and exit angle β_1=26 deg, β_2=49 deg (Fig. 4). The blade shape is a simple circular arc of constant thickness 5 mm, and with both ends rounded.

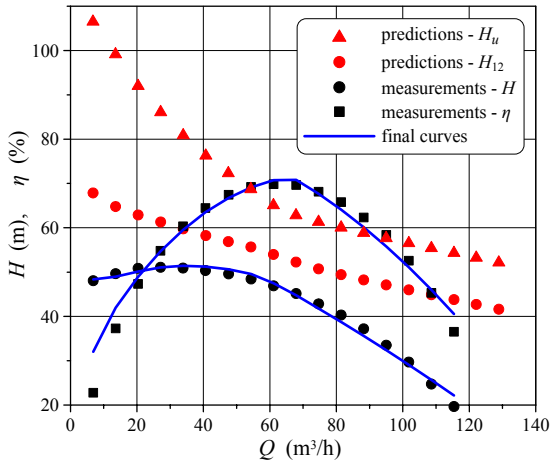


Figure 6. Measured and computed characteristic curves of the pump.

The measured characteristic curves H–Q and η–Q for this pump are shown in Figure 6, along with the numerical results from the 2-dimensional impeller simulation, obtained by Eqs. (4) and (5). The head H_{12} that

the water acquires in the impeller is as expected quite higher than the pump head H. Next, the efficiency coefficients η_Q and η_m were defined from statistical data, whereas using least-squares regression analysis, the values of the adjustable coefficients k_2, k_3 and k_4 were computed so as the pump head H from Eq. (10) approximates well the corresponding experimental curve.

Finally, the efficiency of the pump is calculated from Eq. (13) and the resulting curve is also plotted in Figure 6. Although this curve is not produced by fitting, the agreement with the corresponding measured curve is very good, and this verifies the consistency of the followed modeling strategy and the validity of the adjusted coefficients for this particular pump.

3.3 Optimal blade design

The objective here is to maximize the best efficiency value of the pump, using as design variables the inlet and the exit blade angles. However, the exact location of the best efficiency point (BEP) of the pump depends on the blade design, therefore it must be found for every set of the blade angles. This would need the construction of the η-Q characteristic curve of the pump, by computing several points on it with corresponding runs of the evaluation algorithm. An alternative and much faster method is tested here, according to which the unknown flow rate at the BEP is treated as an additional design variable, along with the blade angles, and its value is obtained when the optimizer converges.

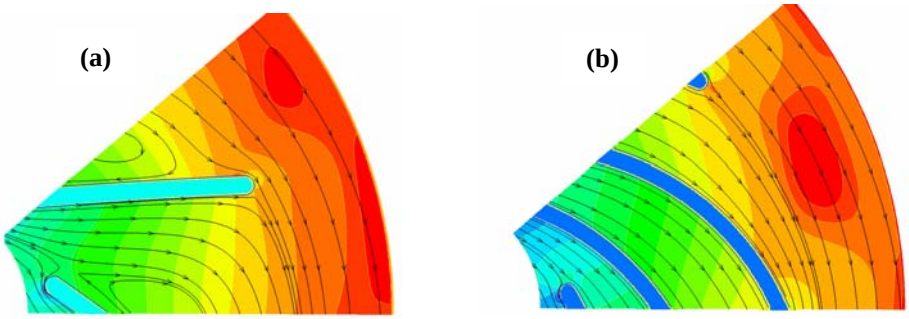


Figure 7. Examples of blade shape and the computed flow field: a) β_1=35, β_2=72 deg; b) β_1=21, β_2=22 deg.

During the optimization process the evaluation algorithm is capable to generate the grid and solve the flow equations for a wide range of different blade configurations, two extreme examples of which are plotted in Figure 7. The convergence rate of the optimization algorithm, which is described previously in chapter 2.5, is shown in Figure 8, for two different starting values. Although the starting values are selected far from the optimal region, the algorithm reaches there in less than 20 evaluations, whereas final convergence occurs in about 60 to 80 evaluations, which can be performed in about 15 to 20 hours by a P4 PC.

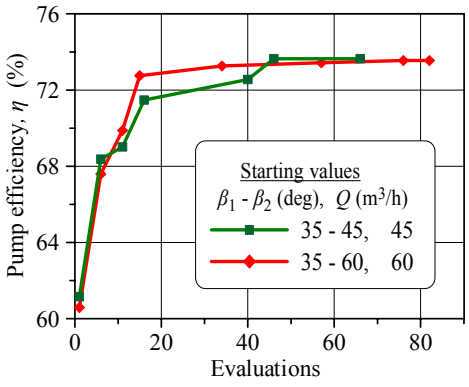


Figure 8. Convergence history of the optimization algorithm.

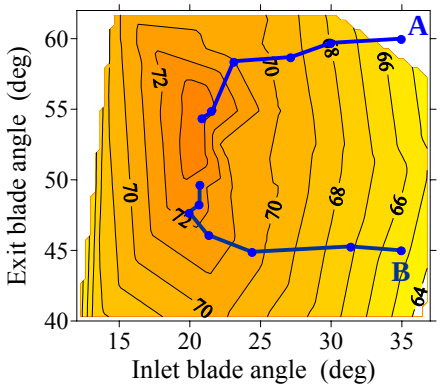


Figure 9. Variation of the design variables during optimization.

The corresponding variation of the blade angles during the optimization process is drawn in Figure 9, where contour lines of constant efficiency are also plotted, as they computed for the optimal flow rate. The two paths converge to the same region, although not to exactly the same point, due to non-continuity effects. The optimal inlet blade angle is about 20-21 deg, which is more consistent with the present rotation speed (3000 rpm) and the

impeller inlet diameter (the real laboratory blade had been constructed with an inlet angle of 26 deg, in order to operate the pump effectively as a turbine too). On the other hand, the exit blade angle exhibits a wider optimal region, ranging between 50 and 58 deg, which is in agreement with theoretical and statistical data. For example, the classic blade number selection relation of Pfleiderer[17]

$$z = 6{,}5 \cdot \frac{D_2 + D_1}{D_2 - D_1} \cdot sin\left(\frac{\beta_1 + \beta_2}{2}\right) \tag{14}$$

for $\beta_2 = 54$ deg. gives $z = 8.6 \rightarrow 9$ blades.

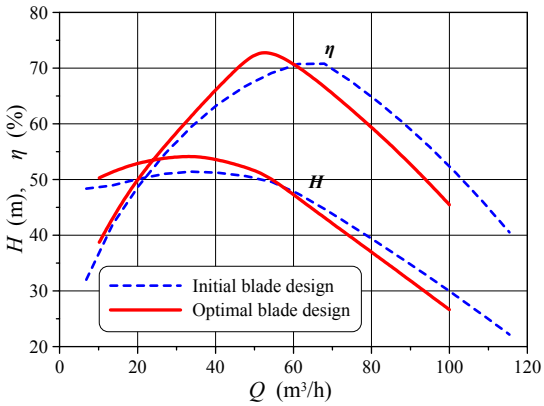


Figure 10. Pump characteristic curves with the initial and the optimal impeller.

The calculated characteristic curves of the pump with the optimized impeller are drawn in Figure 10, along with the corresponding initial blade curves (shown also in Fig. 6). The maximum efficiency of the pump with the optimal blade is about 3% higher (from 70.5% to about 73.5%). However, the BEP is displaced to smaller flow rates comparing to the initial one (52 m^3/h from 63 m^3/h), and the same is valid for the maximum head. The latter is now reasonably higher because of the larger exit angle (initially $\beta_2 = 49$ deg).

4 CONCLUSIONS

A numerical methodology for the calculation of the flow field in a pump impeller and the prediction of the pump characteristics curves is developed, regulated, and tested against experimental and statistical data, with encouraging results. The computer algorithm is found suitable to be used for design optimization purposes in hydraulic pumps and turbines, thanks to the automatic grid generation and the increased precision of representing irregular boundaries. The main advantage validated in the present study is that the methodology provides the ability to localize the optimal region of the design variables at low computer cost.

A more elaborate design would require the detailed simulation of the fluid passage through the 3D impeller/ runner geometry and even the casing of the machine, by performing costly Navier-Stokes solutions. Consequently, the less accurate but much faster modeling approach proposed here can be used as a starting optimization strategy in order to locate the region of maxima, and thus to reduce the number of the subsequent 3D evaluations.

REFERENCES

[1] Zangeneh, M., Goto, A. and Takemura, T. (1996), "Suppression of secondary flows in a mixed-flow pump impeller by application of three-dimensional inverse design method: Part 1 – design and numerical validation", ASME Trans., J. of Turbomachinery 118, pp. 536-543.

[2] Xu, J.Z. and Gu C.W. (1992), "A numerical procedure of three-dimensional design problem in turbomachinery", ASME Trans., J. of Turbomachinery, 114, pp. 548-582.

[3] Borges, J.E. (1994), "A three-dimensional inverse method for turbomachinery: Part-1 Theory", ASME Trans., J. of Turbomachinery, 112, pp. 346-354.

[4] Peng, G., Cao, S., Ishizuka, M. and Hayama, S. (2002), "Design optimization of axial flow hydraulic turbine runner: Part I – an improved Q3D inverse method", Intl. J. Numerical Methods in Fluids, 39, pp. 517-531.

[5] Peng, G., Cao, S., Ishizuka, M. and Hayama, S. (2002), "Design optimization of axial flow hydraulic turbine runner: Part II – multiobjective constrained optimization method", Intl. J. Numerical Methods in Fluids, 39,

pp. 533-548.
[6] Benra, F.K. (2001), "Economic development of efficient centrifugal pump impellers by numerical methods", World Pumps, May 2001, pp. 48-53.
[7] Yulin, W., Jianming, Y. and Shuliang, C (1999), "Advanced design for large hydraulic turbine runner", *Proceedings ASME/JSME FEDSM' 99,* San Francisco, California, July 18-23, S-287.
[8] Garrison, L.A., Richard, K.F., Sale, M.J. and Cada, G. (2002), "Application of biological design criteria and computational fluid dynamics to investigate fish survival in Kaplan turbines", *Proceedings HydroVision 2002,* Portland, Oregon, July 29 – August 2.
[9] Swiderski, J., Martin, J.N., Norrena, R. (2001), "Automated runner blade design optimization process based on CFD verification", *Proceedings Waterpower XII*, Salt lake City.
[10] Mazzouji, F., Couston, M., Ferrando, L., Garcia, F. and Debeissat F. (2004), "Multicriteria optimization: viscous fluid analysis – mechanical analysis", *Proceedings 22nd IAHR Symposium on Hydraulic Machinery and Systems,* Stockholm, Sweden, June 29 – July 2, A04-1.
[11] Kueny, J.L., Lestriez, R., Helali, A., Dmeulenaere, A. and Hirsch, C. (2004), "Optimal design of a small hydraulic turbine", *Proceedings 22nd IAHR Symposium on Hydraulic Machinery and Systems,* Stockholm, Sweden, June 29 – July 2, A02-2.
[12] Ye, T., Mittal, R., Udaykumar, H.S. and Shyy, W. (1999), "An accurate Cartesian grid method for viscous incompressible flows with complex immersed boundaries", J. of Computational Physics, 156, pp. 209-240.
[13] Fadlun, E.A., Verzicco, R., Orlandi, P. and Mohd-Yusof, J. (2000), "Combined immersed-boundary finite-difference methods for three-dimensional complex flow simulations", J. Computational Physics, 161, pp. 35-60.
[14] Grapsas, V. and Anagnostopoulos, J. (2004), "Numerical optimization of the hydrodynamic shape of fluid flow systems", *Proceedings 1st IC-SCCE,* Athens, Greece, 8-10 September.
[15] Anagnostopoulos, J. (2003), "Discretization of transport equations on 2D Cartesian unstructured grids using data from remote cells for the convection terms", Intl. J. for Numerical Methods in Fluids, 42, pp. 297-321.
[16] Anagnostopoulos, J., Bergeles. G., Epple, B. and Stegelitz, P. (2001), "Simulation of grinding and drying performance of a fluid-energy lignite mill", ASME Trans., J. Fluids Engineering, 123(2), pp. 303-310.
[17] Pfleiderer, C. (1961), *Die Kreiselpumpen,* 5th edition, Springer-Verlag, Berlin.

OPTIMAL GRID ADAPTATION THROUGH A POSTERIORI ERROR ANALYSIS

K.N. Gkagkas, D.I. Papadimitriou, K.C. Giannakoglou

Lab. of Thermal Turbomachines,
National Technical University of Athens,
P.O. Box 64069, 15710, Athens, Greece,
e-mail: kgianna@central.ntua.gr

Keywords: A Posteriori Error Analysis, Adjoint Method, Grid Adaptation, Aerodynamics, Turbomachinery

Abstract. *A discrete adjoint approach to grid adaptation is presented. In particular, this paper is concerned with the prediction of integral flow quantities, such as the forces acting upon isolated or cascade airfoils, with user–defined accuracy. The aim is to achieve this accuracy through a small number of computations on successively adapted coarse grids. On each grid the flow and adjoint equations are solved. The adaptation sensor on each grid is computed in terms of flow and adjoint variables and residuals. The method application is considered to be succesful if the overall computational cost is less than that required to solve the problem on a very fine grid, safely adequate to reach the same accuracy but, unfortunately, not known in advance. An a posteriori error analysis formulation, that is the tool guiding the grid adaptation, is adjusted to an upwind compressible flow solver, investigated with respect to its parameters and extended to cascade flows.*

1 INTRODUCTION

Often, the conclusive output of a flow analysis is one (or more) integral quantities which need to be computed with acceptable accuracy for engineering applications. Typical examples of integral outputs in aerodynamics or turbomachinery are the lift and drag of an isolated airfoil or the peripheral force acting on a cascade blade. This occurs frequently in design optimization problems, where the aerodynamic shape with the minimum or maximum value of an integral quantity (minimum drag, maximum lift, maximum loading in a peripheral cascade, etc) is sought. Since search methods (in particular, those based on evolutionary algorithms) require a great amount of evaluations to reach the optimal solution, the CPU cost per evaluation needs to be as low as possible. One way to minimize the cost is by reducing the grid size, without however damaging the prediction accuracy, at least for the integral output of interest.

In view of the above, the accuracy with which the entire flow field is calculated is of importance only so far as this affects this integral output. It is known that flow equation models, discretization schemes and grid resolution are the main issues which determine the accuracy in CFD computations. In what follows the inviscid flow equations are solved and the discretization scheme is a vertex–centered finite–volume method for unstructured grids, [1]. The inviscid fluxes are computed by means of the Roe's approximate Riemann solver [2] with second–order spatial accuracy [3]. In this framework, the integral output needs to be computed with user–defined accuracy through computations made on the coarser possible grid, i.e. with the minimum CPU cost.

This can be done through formulating and solving an appropriate adjoint (dual) problem. Using the flow and adjoint variables and residuals, a measure of the expected contribution of each grid node to the error in the integral output is estimated, [4-8]. This measure, in the form of a scalar sensor field over the grid edges, is used to selectively enrich the coarse grid in error inducing areas. The adaptation leads to a new grid on which the flow and adjoint equations are solved again and this procedure goes up as long as high error inducing grid subsets are identified. The computation on the finally adapted grid yields the

integral output with the desired accuracy. In the expense of this iterative algorithm, which relies upon the numerical solution of flow and adjoint equations on sequentially adapted grids, engineers overcome the necessity to generate extremely fine grids and the CPU cost reduces. We will show that the cost of solving the direct and adjoint equations on successive grids is lower than that of solving the flow equations on a very fine mesh.

The structure of this paper is as follows: The formulation of the adjoint problem and the definition of grid adaptation sensors are presented first. More about the method can be found in any of the aforementioned works on a posteriori error analysis. In the results section, the method is applied to flow problems concerned with the accurate prediction of lift and drag in isolated airfoils and the peripheral force in a compressor cascade. Through these cases, particular features of this method are highlighted.

2 A POSTERIORI ERROR ANALYSIS

Let $f(U)$ be an integral flow quantity, resulting from the integration of the flow variables U over a part of the domain boundary. A specific accuracy level is required for $f(U)$. Let us also consider two computational grids, namely the coarse (index H) and the fine (index h) ones, over the flow domain. U_H and U_h can be calculated through satisfying the flow equations, $R_H(U_H) = 0$ or $R_h(U_h) = 0$, on each grid. Then, $f_H(U_H)$ and $f_h(U_h)$ can be computed through the same integration scheme. Note that U_H and $f_H(U_H)$ are obtained using low–cost computations and are not so accurate since both the solution of the flow equations and integration are carried out using the coarse grid. In contrast, the solution of $R_h(U_h) = 0$ is computationally expensive and, practically, undesirable. Finally, starting from U_H, one may interpolate it onto the fine grid by means of a prolongation operator I_h^H to get

$$U_h^H = I_h^H U_H \tag{1}$$

and, $f_h(U_h^H)$, through integration over the fine grid.

By expanding the first–order Taylor series about $f_h(U_h^H)$ and $R_h(U_h^H)$, we get (= is used instead of $\approx$)

$$f_h(U_h) = f_h(U_h^H) + \frac{\partial f_h}{\partial U_h}|_{U_h^H}(U_h - U_h^H) \tag{2}$$

$$R_h(U_h) = R_h(U_h^H) + \frac{\partial R_h}{\partial U_h}|_{U_h^H}(U_h - U_h^H) = 0 \tag{3}$$

where $\frac{\partial f_h}{\partial U_h}|_{U_h^H}$ and $\frac{\partial R_h}{\partial U_h}|_{U_h^H}$ are computed using the prolongated field U_h^H. According to eq. 3, $U_h - U_h^H$ is given by

$$U_h - U_h^H = -\left[\frac{\partial R_h}{\partial U_h}|_{U_h^H}\right]^{-1} R_h(U_h^H) \tag{4}$$

which, upon substitution into eq. 2, provides an estimate of the integral functional as follows

$$f_h(U_h) = f_h(U_h^H) - \frac{\partial f_h}{\partial U_h}|_{U_h^H}\left[\frac{\partial R_h}{\partial U_h}|_{U_h^H}\right]^{-1} R_h(U_h^H) \tag{5}$$

The matrix inversion in eqs. 4 and 5 can be handled by introducing the adjoint variables Ψ, satisfying the so–called adjoint equations

$$\left[\frac{\partial R_h}{\partial U_h}|_{U_h^H}\right]^T \Psi_h|_{U_h^H} = \left(\frac{\partial f_h}{\partial U_h}|_{U_h^H}\right)^T \tag{6}$$

In terms of the adjoint variables, the functional $f_h(U_h)$ is merely expressed as

$$f_h(U_h) = f_h(U_h^H) - \left(\Psi_h|_{U_h^H}\right)^T R_h(U_h^H) \tag{7}$$

Eq. 7 can be considered as a better approximation to $f_h(U_h)$, compared to $f_h(U_h^H)$. However, solving for Ψ_h should be avoided, as we did for any other computation on the fine grid. So, instead of solving eq. 6, the adjoint equations are written and solved on the coarse grid, i.e.

$$\left[\frac{\partial R_H}{\partial U_H}\right]^T \Psi_H = \left(\frac{\partial f_H}{\partial U_H}\right)^T \tag{8}$$

Ψ_H is then interpolated over the fine grid nodes, through the prolongation operator J_h^H,

$$\Psi_h^H = J_h^H \Psi_H \tag{9}$$

So, instead of eq. 6, the following equation can be used

$$\tilde{f}_h(U_h) = f_h(U_h^H) - (\Psi_h^H)^T R_h(U_h^H) \tag{10}$$

3 GRID ADAPTATION

The last term in eq. 10 stands for a correction term (error) through which a better (than $f_h(U_h^H)$) estimate of f can be obtained, without however solving any partial differential equation on the fine grid. This is possible in the expense of additionally solving the adjoint equations on the coarse grid, i.e. with as much as twice the cost of computing U_H. However, if higher accuracy is needed, the coarse grid can be adapted, particularly in areas which induce the maximum error in the integral output and the same procedure is repeated. The grid adaptation must be driven by a sensor which is proportional to the aforementioned error. In [?], it is demonstrated that this error can be written as either

$$f_h(U_h) - f_h(U_h^H) = \left(\Psi_h^H\right)^T R_h(U_h^H) + \left(\Psi_h|_{U_h^H} - \Psi_h^H\right)^T R_h(U_h^H) \tag{11}$$

or

$$f_h(U_h) - f_h(U_h^H) = \left(\Psi_h^H\right)^T R_h(U_h^H) + \left\{R_h^\Psi\left(\Psi_h^H\right)\right\}^T (U_h - U_h^H) \tag{12}$$

where the residual of the adjoint equation on the fine grid is

$$R_h^\Psi(\Psi) \equiv \left[\frac{\partial R_h}{\partial U_h}|_{U_h^H}\right]^T \Psi - \left(\frac{\partial f_h}{\partial U_h}|_{U_h^H}\right)^T \tag{13}$$

From the above relations, an adaptation criterion, which takes into consideration the errors associated with both the flow and adjoint equations, is, [?]

$$\epsilon_k = \frac{1}{2}\sum_{l(k)} |\left[Q_h^H \Psi_H - L_h^H \Psi_H\right]_{l(k)}^T \left[R_h(L_h^H U_H)\right]_{l(k)}| + |\left[Q_h^H U_H - L_h^H U_H\right]_{l(k)}^T \left[R_h^\Psi(L_h^H \Psi_H)\right]_{l(k)}| \tag{14}$$

where L and Q are linear and quadratic interpolation schemes. Considering that the fine grid h is defined based on the existing coarse grid, ϵ_k is assigned to each coarse grid element (here, edge) and the summation over $l(k)$ takes into account any fine grid element associated with the coarse grid element k. Finally, the local adaptation parameter is defined by

$$\eta_k = \frac{\epsilon_k}{e_o} \tag{15}$$

where e_o is a user–defined allowed error for the integral output. Grid edges marked with $\eta_k > 1$ should be refined. The grid adaptation cycles terminate when there are no more edges marked for refinement.

4 THE OVERALL ALGORITHM – PRACTICALITIES

The repetitive grid algorithm which leads to the computation of $f(U)$ with prescribed accuracy includes the following steps.

1. Solve the flow and adjoint equations using the same coarse unstructured grid and compute $f(U_H)$.
2. Generate the fine grid (usually the "quadruple" grid, created by subdividing each coarse grid triangle into four triangles) and compute $L_h^H U_H$, $Q_h^H U_H$, $L_h^H \Psi_H$ and $Q_h^H \Psi_H$ over its nodes.
3. Compute the flow and adjoint equation residuals at the fine grid nodes, using $L_h^H U_H$ and $L_h^H \Psi_H$

4. Calculate the adaptation sensor η_k for each coarse grid edge and refine accordingly the coarse grid.

5. Update the coarse grid and return to step (1).

The numerical solution of the Euler equations is carried out through the aforementioned time–marching solver. The numerical inviscid fluxes crossing the interface between any pair of adjacent node–centered control volumes, are computed through the Roe's approximate Riemann solver, [**?**]. They become second–order accurate through variable extrapolation, [**?**], which requires the primitive variable gradient at the grid nodes. Gradients are computed over the triangular elements and, then, scatter–added to the nodes. The discretized system of equations is solved using the pointwise implicit Jacobi method. The matrix coefficient is filled in by considering only first–order convection terms; thus, the non–zero pattern of the coefficient matrix, in graph theory terms. coincides with that of the adjacency matrix in graph theory.

The previous assumption concerning the formation of the matrix coefficient is important. During the solution of eq. 8, $\left[\frac{\partial R_H}{\partial U_H}\right]^T$ is set equal to the transpose of the aforementioned coefficient matrix. Thus, in the sake of computational convenience, only the first–order terms are taken into account in the discrete adjoint equation. The r.h.s. term in the same equation, $\left(\frac{\partial f_H}{\partial U_H}\right)^T$ is expressed according to the trapezoidal integration law. In the present problems, the trapezoidal rule is exclusively used to compute integral outputs. Switching to a more accurate integration formula is possible and can be employed in a straightforward manner; however, this does not affect the conclusions drawn below.

5 RESULTS–DISCUSSION

The first case is concerned with the study of the flow developed around the isolated RAE2822 profile. The flow is inviscid with $M_\infty = 0.50$ and $\alpha_\infty = 3^o$. The maximum Mach number over the airfoil is about 0.92 and drag should approach zero. The initial grid is generated through the advancing front method after defining 112 nodes along the airfoil contour, clustered close to the leading and trailing edges. The initial grid (URG1) is fully unstructured, with as many as 1448 nodes and 2763 triangles. No particular care concerning the location of nodes is taken.

A couple of computations on successively refined grids are first made to obtain some reference results, concerning lift and drag coefficients. Starting from URG1, a grid (URG2) with as many as four times its triangles is generated by splitting each one of its triangles into four. The generation of the so–called *uniformly refined grids* (URG) is repeated three times. Over these four grids, the flow solver is used to predict the flow field and, through integration, to compute lift (C_l) and drag (C_d) coefficients. These values are tabulated in table 1 which also shows the CPU cost of each computation. All four computations are made with the same initialization; even if the computation on any fine grid could start from the converged solution on the previous (coarser) grid, this is avoided in the sake of fairness in the comparison of CPU costs. A solution is considered to be converged if the maximum residual becomes lower than 10^{-20} and this criterion determines the CPU cost of each computation. All computations are made on an Intel Pentium M processor at $1.80GHz$. Note that C_d approaches zero without reaching it, due to numerical diffusion.

Grid	Nodes	Triangles	C_l	C_d	CPU secs.
URG1	1448	2763	0.6674	0.008744	13.3
URG2	5659	11052	0.6907	0.004243	111.0
URG3	22370	44208	0.6975	0.003551	1124.1
URG4	88948	176832	0.6998	0.003463	11331.2

Table 1: Flow around the RAE2822 airfoil. Computed (reference) C_l and C_d values using four uniformly refined grids.

The grid adaptation with a posteriori error analysis is employed four times, with the same starting grid (URG1). Each time, a different e_o value is used; the e_o values used are listed in fig. 1 which shows the changes of C_l during the successive grid refinements in terms of grid size (left) and CPU cost (right). In all cases, a small number of adaptation cycles (around five) is needed. After the first or second cycle, the grid size increases slightly. On the other hand, the CPU cost of the flow analysis on each refined grid is, more or less, the same, since the starting flow field is interpolated from the converged solution on the coarser grid. All four runs show the correct trend of the C_l value computed over the finally refined grids.

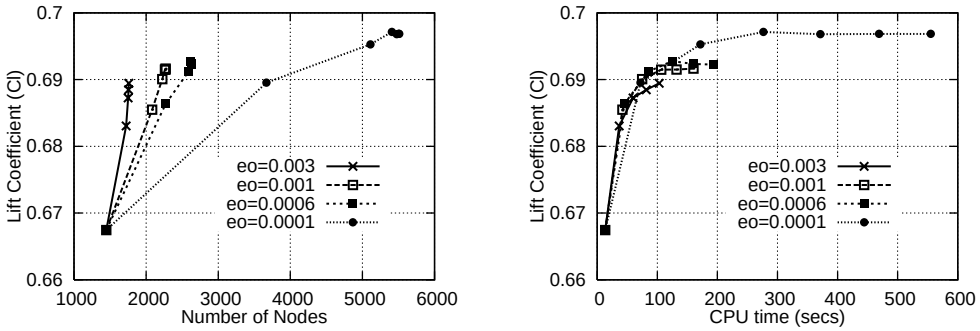


Figure 1: Flow around the RAE2822 airfoil, with user–defined accuracy in C_l. Computed C_l values during the grid adaptation driven by η_k, plotted in terms of the number of nodes (left) and CPU cost (right). The curves shown correspond to four different e_o values.

Fig. 2 compares the cost for obtaining the final C_l value, with user–defined accuracy, with that of using the URGs of table 1. Using the two runs with the lower e_o values, useful conclusions can be drawn. The $e_o = 0.0006$ run computes a better (slightly higher, i.e. closer to that of URG4) C_l value than that computed using URG2. This computation leads to an adapted grid with 2634 nodes (URG2 possesses 5659 nodes which, despite that, yields a slightly worse C_l value). The cost for the computation using $e_o = 0.0006$ is about 193 secs. Using the more strict C_l criterion ($e_o = 0.0001$), the C_l value computed using URG3 is obtained. However, fig. 2, the cost of this run (555 secs) is about half of the cost for URG3 (1124.1 secs). The economy in CPU cost is, in fact, noticeable.

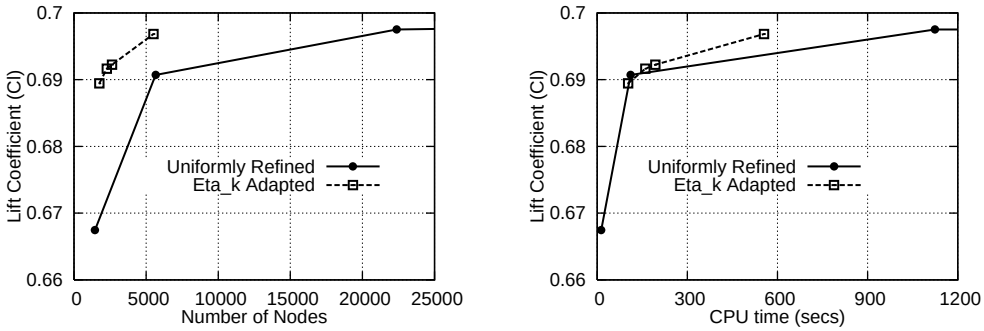


Figure 2: Flow around the RAE2822 airfoil, with user–defined accuracy in C_l. Comparison of C_l values computed using (a) a posteriori error analysis with grid adaptation and (b) the URGs, table 1.

Two more comparisons for the same case can be made using the plots shown in fig. 3. First, eq. 10 is used to correct the C_l computed on each grid during the successive adaptation cycles. Compared to fig. 1 (left) a slight additional improvement is shown in fig. 3 (left). Also, the C_l values on the successively refined grids (for $e_o = 0.0001$) are extrapolated according the Richardson's formula and this offers an additional way of exploiting the a posteriori error analysis method, fig. 3 (right).

The same airfoil, with the same infinite flow conditions and starting grid (URG1) is used for four new computations. This time, the target is the computation of C_d with prescribed accuracy. Four e_o values are used. The results obtained are shown in figs. 4, 5, according to the previously used presentation mode. The conclusions that can be drawn are similar. For instance, fig. 5 (left) shows that the same accuracy level can be obtained using a much coarser grid. From fig. 5 (left) and the CPU costs listed in table 1, it is also obvious that the lower e_o value ($e_o = 0.0001$) reaches the best value for C_d in about 3000 secs compared to the four times more expensive computation using URG4. As in the case of C_l correction, the correction of C_d through eq. 10 or Richardson extrapolation, (not shown here in the interest of space) can be used to further improve the C_d predictions. The starting grid as well as two finally adapted grids for $e_o = 0.0001$ used as threshold for C_l and C_d are shown in fig. 6.

The last case is concerned with the accurate prediction of the peripheral force coefficient in a compressor cascade. Here, a 2D controlled diffusion airfoil cascade is utilized and the integral quantity that needs to be computed with desired accuracy is the non-dimensional force component in the pitchwise direction.

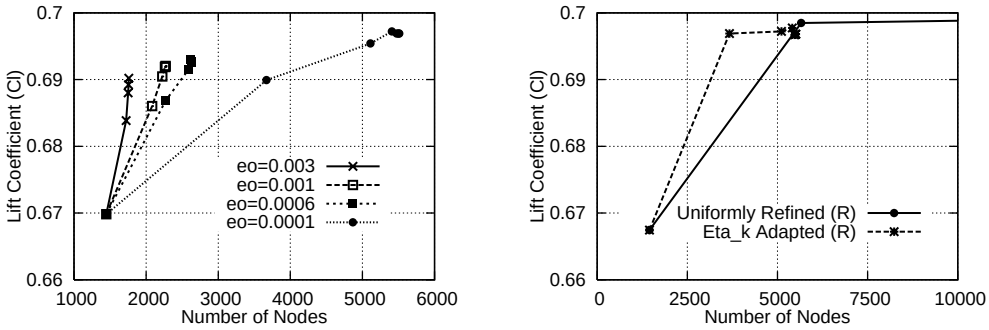


Figure 3: Flow around the RAE2822 airfoil, with user–defined accuracy in C_l. Corrected through eq. 10 (left) and Richardson extrapolated (right, for $e_o = 0.0001$) C_l values computed using (a) a posteriori error analysis with grid adaptation and (b) the URGs, table 1.

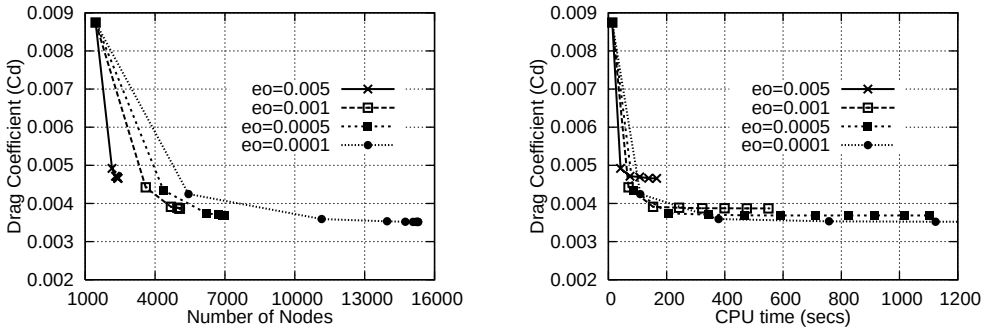


Figure 4: Flow around the RAE2822 airfoil, with user–defined accuracy in C_d. Computed C_d values during the grid adaptation driven by η_k, plotted in terms of the number of nodes (left) and CPU time (right).

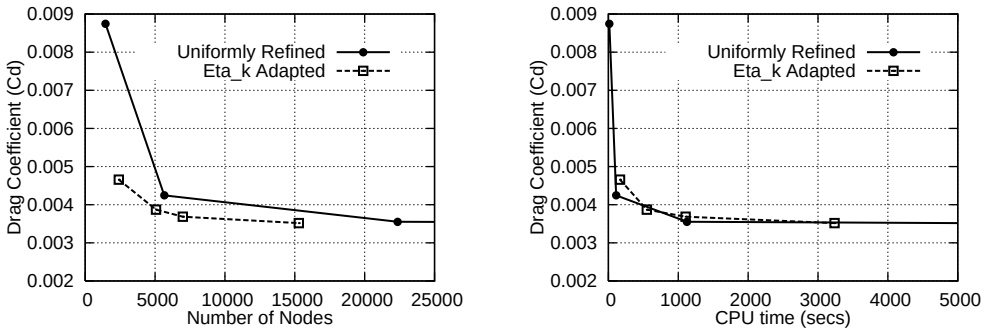


Figure 5: Flow around the isolated RAE2822 airfoil, with user–defined accuracy in C_d. Comparison of C_d values computed using (a) a posteriori error analysis with grid adaptation and (b) the URGs, table 1.

Since the flow is inviscid, this is derived through the integration of the pressure distribution around the airfoil and projection in the pitchwise direction. The coefficient of interest is denoted by C_{Fy}. The flow conditions are: $M_{2,is} = 0.37$ and $\alpha_1 = 47^o$. In all computations, the Jacobi method is used with CFL number equal to 50. The convergence criteria are the same as previously. Of course, modifications in the adjoint method are necessary, in order to account for periodicity.

Table 2 shows the three URGs used to get reference C_{Fy} values so as to compare the expected gain from the use of grid adaptation based on the a posteriori error analysis with uniformly (and, thus, uncontrollably expensive) grid refinement. URG1 is the starting grid in every subsequent calculation.

Fig. 7 shows the results of five computations with a wide span of e_o values. In contrast to the monotonic curves previously shown, in all cases, an overshooting in C_{Fy} is observed before reaching its final value, on the finally adapted grid. According to table 2, we may assume that the desired C_{Fy} value is about 0.033. We can see from fig. 7 that the lower the desired accuracy threshold e_o, the closer to 0.033 the terminal C_{Fy} value is captured.

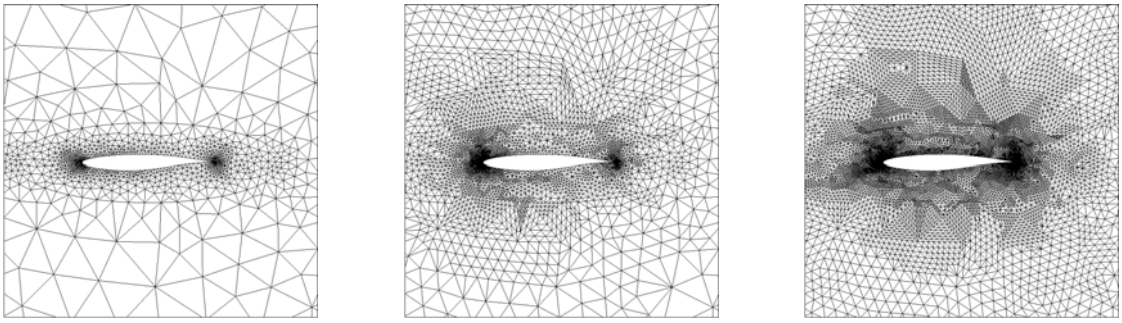


Figure 6: Flow around the RAE2822 airfoil. Left: starting grid (URG1), Middle: finally adapted grid (5571 nodes, 10777 triangles), computed using $e_o = 0.0001$ for C_l. Right: finally adapted grid (15290 nodes, 30170 triangles), computed using $e_o = 0.0001$ for C_d.

Grid	Nodes	Triangles	C_{Fy}	CPU secs.
URG1	1809	3327	0.032880	24.6
URG2	6945	13308	0.032988	189.8
URG3	27198	53232	0.033014	1472.6

Table 2: Flow in a 2D compressor cascade. Computed (reference) peripheral force coefficient values using three uniformly refined grids.

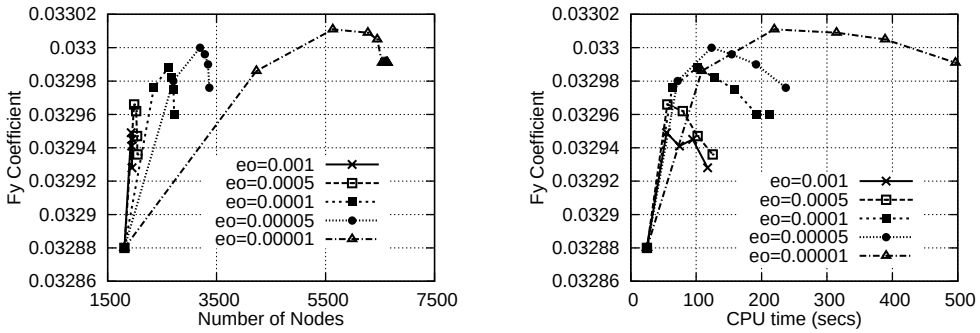


Figure 7: Flow in a 2D compressor cascade, with user–defined accuracy in C_{Fy}. Computed C_{Fy} values during the grid adaptation driven by η_k, plotted in terms of the number of nodes (left) and CPU cost (right), for five different e_o values.

Fig. 8 compares the results of grid adaptation through a posteriori error analysis with those obtained using the three URGs. In terms of grid size, the a posteriori error analysis based adaptation leads to much smaller numbers of grid nodes. The difference is not that important in terms of CPU cost; however, this depends on the convergence criteria used during the repetitive solution of the flow and adjoint equations and could be improved through the selection of different values. Fig. 9 presents the finally adapted grids for $e_o = 0.0005$ and $e_o = 0.00005$. It is obvious that the latter leads to a considerably finer grid close to the blade airfoil.

6 CONCLUSIONS

In this paper, the a posteriori error analysis method combined with grid adaptation techniques was presented, as a tool that ensures the computation of integral quantities with user–defined accuracy. The expected gain, which was confirmed by the examined cases, is that our goal can be achieved without using uncontrollably fine grids; local criteria can be used to drive the grid refinement in areas which induce the higher errors in the computation of the integral output of interest. The repetitive solution of the flow and adjoint equations are carried out on comparatively coarse grids, so the overall CPU cost reduces. Correction formulaes, based on either the adjoint formulation or Richardson extrapolations can be used to improve the prediction accuracy, with the same computing cost. In this paper, the a posteriori error analysis method is extended to periodic turbomachinery flows.

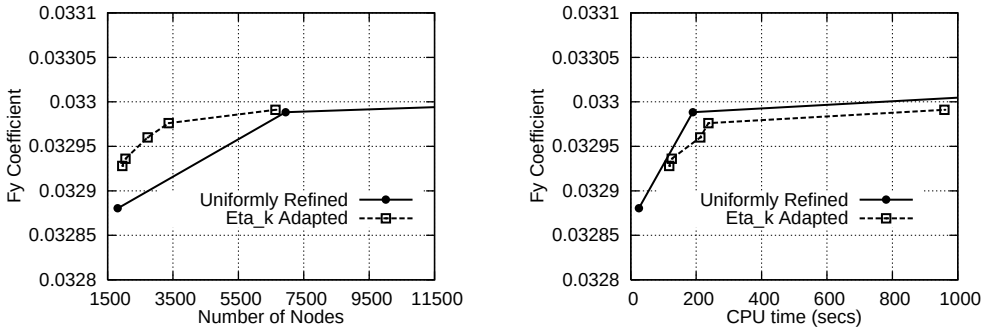


Figure 8: Flow in a 2D compressor cascade, with user–defined accuracy in C_{Fy}. Comparison of C_{Fy} values computed using (a) a posteriori error analysis with grid adaptation and (b) the URGs, table 2.

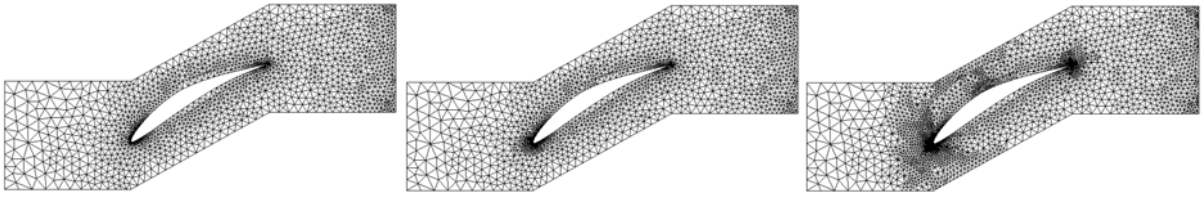


Figure 9: Flow in a 2D compressor cascade, with user–defined accuracy in C_{Fy}. Left: Initial coarse grid (URG1, 1809 nodes, 5136 triangles). Middle: finally adapted grid (2048 nodes, 3786 triangles), computed using $e_o = 0.0005$. Right: finally adapted grid (3368 nodes, 6354 triangles), computed using $e_o = 0.00005$.

ACKNOWLEDGMENT

This work was supported by the Iraklitos Basic Research Project, funded by the Hellenic Ministry of Education.

REFERENCES

[1] Koubogiannis, D.G., Poussoulidis, L.C., Rovas, D.V. and Giannakoglou, K.C. (1998), "Solution of Flow Problems Using Unstructured Grids on Distributed Memory Platforms", Computer Methods in Applied Mechanics and Engineering 160, pp. 89-100.

[2] Roe, P. (1981), "Approximate Riemann Solvers, Parameter Vectors and Difference Schemes", Journal of Computational Physics 43, pp. 357-372.

[3] van Leer, B. (1982), "Flux vector Splitting for the Euler Equations", Lecture Notes in Physics 170, pp. 405-512.

[4] Giles, M.B. (1997), "On adjoint equations for error analysis and optimal grid adaptation in cfd", tech. rep., Oxford University Computing Laboratory, Numerical Analysis Group, Wolfson Building, Parks Road, Oxford, England, Report no. 97/11, September 1997.

[5] Venditti, D.A. and Darmofal, D.L. (1999), "A multilevel error estimation and grid adaptive strategy for improving the accuracy of integral outputs", AIAA 99-3292.

[6] Venditti, D.A. and Darmofal, D.L. (2000), "Adjoint error estimation and grid adaptation for functional outputs: Application to quasi-one-dimension flow", Journal of Computational Physics 164, pp. 204-227.

[7] Venditti, D.A. and Darmofal, D.L. (2001), "A grid adaptive technology for functional outputs of compressible flow simulations", AIAA 2001-2659.

[8] Venditti, D.A. and Darmofal, D.L. (2002), "Grid adaptation for functional outputs: Application to two-dimensional inviscid flows", Journal of Computational Physics 176, pp. 40-69.

5^{th} GRACM International Congress on Computational Mechanics
Limassol, 29 June - 1 July, 2005

LOW MACH NUMBER PRECONDITIONING FOR 2D AND 3D UPWIND FLOW SOLUTION SCHEMES ON UNSTRUCTURED GRIDS

V.G. Asouti, D.I. Papadimitriou, D.G. Koubogiannis, K.C. Giannakoglou

Lab. of Thermal Turbomachines,
National Technical University of Athens,
P.O. Box 64069, 15710, Athens, Greece,
e-mail: vasouti@mail.ntua.gr

Keywords: Preconditioning

Abstract. *An existing method and the corresponding software for the numerical solution of the Euler and Navier–Stokes equations in high–speed flows is extended to account for low–speed flows as well. This is achieved through the multiplication of the governing equations by a precondition matrix which is defined at each grid node in terms of the local Mach number and ensures adequately clustered eigenvalues and, thus, optimal convergence characteristics at all flow speeds. A second–order upwind scheme is adapted to the preconditioned system of equations through appropriate assumptions, which are clearly presented in this paper. 2D or 3D, inviscid or turbulent flow problems are analyzed, in external aerodynamics and turbomachinery.*

1 INTRODUCTION

It is well known that the numerical solution of the compressible fluid flow equations for the low Mach regime suffers from slow convergence and increased computing cost. It is also known that the main reason for the performance degradation of the relevant software, based on time–marching schemes and the theory of hyperbolic system of equations, is the large disparity between acoustic waves and fluid speeds. To overcome this problem, completely different formulations for the prediction of low–speed or incompressible flows such as pressure correction and pseudo–compressibility methods, have been developed. However, maintaining and extending two different CFD tools by the same research group is, in fact, cumbersome. The only way to use the same time–marching solution method regardless of the flow speed, is through preconditioning.

Conceptually, preconditioning is based on the multiplication of the pseudo–time derivative by an appropriate precondition matrix without affecting the steady state solution. The precondition matrix is defined in terms of the local Mach number [1], [2], [3] and degenerates to the unit matrix at sonic speed. According to the hyperbolic system theory, the Jacobians of the convection terms are multiplied by the inverse of the precondition matrix and this gives rise to much more clustered eigenvalues compared to those of standard Jacobians. The selection of the precondition matrix depends on the vector of solution variables. In the literature, different precondition matrices have been proposed depending on whether the flow equations are solved in terms of $\overrightarrow{Q} = [p\ u\ v\ T]^T$ or $\overrightarrow{W} = [\varrho\ \varrho u\ \varrho v\ E]^T$ (see [4] or [5] and [6], respectively).

The implementation of preconditioning in a numerical flow solver depends practically on the discretization scheme used. The present method [7], [8] is based on the finite volume technique for unstrucured grids with an upwind scheme for the discretization of convection terms. The latter are computed by sweeping the grid edges and employing a 1D Riemann flow solver between the two edge nodes. To maintain the existing formulation, a couple of assumptions concerning the management of the precondition matrix are made. These assumptions concern even the residual of the iteratively solved equations and contribute to the elimination of pressure oscillations that the conventional system of equations produce at low Mach numbers, particularly close to the leading and trailing edges of airfoils.

2 GOVERNING EQUATIONS AND LOW–MACH PRECONDITIONING

Though this paper is concerned with both 2D and 3D all–speed flows, for the sake of simplicity, the analysis of the method will be restricted to 2D flows; any extension to 3D flows is straightforward and, thus, omitted. The 2D Euler equations for compressible flows are written, in conservative form, as follows

$$\frac{\partial \overrightarrow{W}}{\partial t} + \frac{\partial \overrightarrow{F}}{\partial x} + \frac{\partial \overrightarrow{G}}{\partial y} = 0 \tag{1}$$

where

$$\overrightarrow{W} = \left\{ \begin{array}{c} \varrho \\ \varrho u \\ \varrho v \\ E \end{array} \right\} , \qquad \overrightarrow{F} = \left\{ \begin{array}{c} \varrho u \\ \varrho u^2 + p \\ \varrho uv \\ (E+p)u \end{array} \right\} , \qquad \overrightarrow{G} = \left\{ \begin{array}{c} \varrho v \\ \varrho uv \\ \varrho u^2 + p \\ (E+p)v \end{array} \right\} \tag{2}$$

Here ϱ is the density, u and v the velocity components, E the total energy per unit volume and p the pressure. Note that only steady flow simulations are of interest. Starting from eq. 1, the preconditioned equations are obtained by multipliying the pseudo–time derivative term by the inverse of an appropriate precondition matrix Γ, namely

$$\Gamma^{-1} \frac{\partial \overrightarrow{W}}{\partial t} + \frac{\partial \overrightarrow{F}}{\partial x} + \frac{\partial \overrightarrow{G}}{\partial y} = 0 \tag{3}$$

or

$$\frac{\partial \overrightarrow{W}}{\partial t} + \Gamma A_x \frac{\partial \overrightarrow{W}}{\partial x} + \Gamma A_y \frac{\partial \overrightarrow{W}}{\partial y} = 0 \tag{4}$$

where $A_x = \frac{\partial \vec{F}}{\partial \vec{W}}$ and $A_y = \frac{\partial \vec{G}}{\partial \vec{W}}$ are the Jacobian matrices for the conservative variables. Eq. 4 can also be written in terms of the primitive variable array $\overrightarrow{V} = [\varrho \ u \ v \ p]^T$ as follows

$$\frac{\partial \overrightarrow{V}}{\partial t} + \overline{\Gamma}\,\overline{A}_x \frac{\partial \overrightarrow{V}}{\partial x} + \overline{\Gamma}\,\overline{A}_y \frac{\partial \overrightarrow{V}}{\partial y} = 0 \tag{5}$$

where $\overline{A}_x$ and $\overline{A}_y$ are the corresponding Jacobian matrices, $\overline{A}_x = M^{-1}A_xM$, $\overline{A}_y = M^{-1}A_yM$, $\overline{\Gamma} = M^{-1}\Gamma M$ and $M = \frac{\partial \vec{W}}{\partial \vec{V}}$.

As already mentioned in the introduction, the role of the precondition matrix Γ (and, subsequently, the role of $\overline{\Gamma}$) is to alleviate the disparity between acoustic waves and the fluid speeds which characterize the non–preconditioned (regular) system of flow equations and is the main reason for the slow convergence of any numerical solution method applied to low Mach number flows. This paper adopts the precondition matrix originally proposed by [6], namely

$$\overline{\Gamma} = \left[\begin{array}{cccc} 1 & 0 & 0 & -\frac{1-a}{c^2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & a \end{array} \right] \tag{6}$$

where $a = \min[1, M^2]$ and M is the local Mach number; Γ can be obtained from $\overline{\Gamma}$, since $\Gamma = M\overline{\Gamma}M^{-1}$.

Through the application of the Gauss' divergence theorem, the integration of eqs. 5 over any finite volume cell leads to the integral of numerical fluxes crossing its boundary. Let $\vec{n} = (n_x, n_y)$ denote the normal vector to the boundary; the application of any upwind scheme to numerically compute these fluxes as well as the convergence characteristics of the resulting solution method is determined by the eigenvalues of the preconditioned directional Jacobian matrix

$$\overline{A}_\Gamma = \overline{\Gamma}\,\overline{A} = \overline{\Gamma}(\overline{A}_x n_x + \overline{A}_y n_y) \tag{7}$$

which are

$$\begin{aligned} \lambda_1 &= \lambda_2 = \vec{v} \cdot \vec{n} \\ \lambda_{3,4} &= \frac{1}{2}\{(1+a)\vec{v} \cdot \vec{n} \pm \sqrt{[(1-a)\vec{v} \cdot \vec{n}]^2 + 4ac^2|\vec{n}|^2}\} \end{aligned} \tag{8}$$

It can be shown that, for low Mach number flows, the eigenvalues, (eq. 8) are much more clustered than those of the non–preconditioned system $\lambda_1^* = \lambda_2^* = \vec{v} \cdot \vec{n}$, $\lambda_3^* = \lambda_4^* = \vec{v} \cdot \vec{n} + c\,|\vec{n}|$, giving thus rise to better convergence properties. Fig. 1 compares the ratio λ_3/λ_1 of the preconditioned and the non–preconditioned system λ_3^*/λ_1^* for two different ranges of the Mach number. It can clearly be observed that, through preconditioning, eigenvalues become much more clustered as Mach number approaches zero.

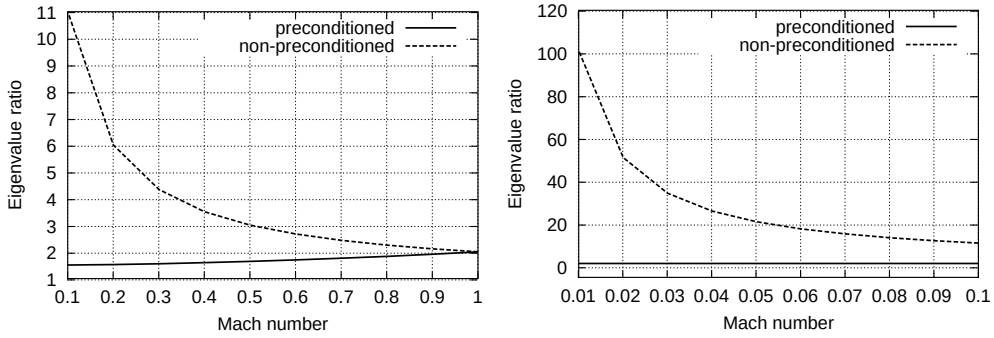


Figure 1: Comparison of the λ_3/λ_1 ratio for the preconditioned and the non–preconditioned system for $0.1 \leq M \leq 1$ (left) and $0.01 \leq M \leq 0.1$ (right).

3 DISCRETIZATION AND NUMERICAL SOLUTION

The preconditioned equations are integrated over vertex–centered finite volumes $\mho$ defined in fig. 2 for 2D problems. To carry out the integration, the assumption that Γ stays out of the integral, which facilitates considerably the subsequent development of equations, is made. The meaning of this assumption is that, in any vertex–centered cell, Γ remains fixed and equal to that defined at the enclosed node P. So

$$\iint_{\mho} \frac{\partial \overrightarrow{W}}{\partial t}\, d\mho + \Gamma \int_{\partial \mho} (\overrightarrow{F} n_x + \overrightarrow{G} n_y)\, d\partial\mho = 0 \tag{9}$$

which, through further analysis of terms, yields

$$\frac{\mho_P}{\Delta t_P}\, \delta \overrightarrow{W}_P + \Gamma_P \sum_{Q \in nei(P)} \overrightarrow{\Phi}_{PQ} = 0 \tag{10}$$

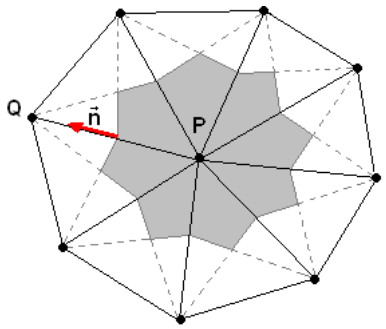


Figure 2: Control volume surrounding a node.

or, equivalently,

$$\frac{\mho_P}{\Delta t_P}\, \delta \overrightarrow{W}_P + \Gamma_P \sum_{Q \in nei(P)} (\Gamma^{-1} \Gamma \overrightarrow{\Phi}_{PQ}) = 0 \tag{11}$$

where $nei(P)$ is the set of grid nodes that are linked with P through grid segments and $\overrightarrow{\Phi}_{PQ}$ is the numerical flux crossing the interface of the finite volumes defined around adjacent nodes P and Q. Between P and Q, the 1D Roe approximate Riemann solver [9] is employed, according to which

$$\overrightarrow{\Phi}_{PQ} = \frac{1}{2}[H(\overrightarrow{W}_P, \vec{n}_{PQ}) + H(\overrightarrow{W}_Q, \vec{n}_{PQ})] - \frac{1}{2}|\tilde{A}_{PQ}|\Delta\overrightarrow{W}_{PQ} \tag{12}$$

where $\tilde{A}_{PQ}$ is the Roe–averaged Jacobian at the midnode between P and Q and $\Delta\overrightarrow{W}_{PQ} = \overrightarrow{W}_P - \overrightarrow{W}_Q$ For second order accuracy, $\overrightarrow{W}_L = \overrightarrow{W}_P + \frac{1}{2}\overrightarrow{PQ}\,\nabla\overrightarrow{W}_P$ and $\overrightarrow{W}_R = \overrightarrow{W}_Q + \frac{1}{2}\overrightarrow{PQ}\,\nabla\overrightarrow{W}_Q$ should substitute for $\overrightarrow{W}_P$ and $\overrightarrow{W}_Q$, respectively. The last term in eq. 12 is rewriten as follows:

$$\begin{aligned} |\tilde{A}_{PQ}|\Delta\overrightarrow{W}_{PQ} &= |\tilde{\Gamma}^{-1}_{PQ}\tilde{\Gamma}_{PQ}\tilde{A}_{PQ}|\Delta\overrightarrow{W}_{PQ} \\ &\simeq \tilde{\Gamma}^{-1}_{PQ}|\tilde{\Gamma}_{PQ}\tilde{A}_{PQ}|\Delta\overrightarrow{W}_{PQ} \\ &= \tilde{\Gamma}^{-1}_{PQ}|\tilde{A}_{\Gamma_{PQ}}|\Delta\overrightarrow{W}_{PQ} \end{aligned} \tag{13}$$

Here $|\tilde{A}_{\Gamma_{PQ}}|$ is defined by

$$|\tilde{A}_{\Gamma_{PQ}}| = \tilde{P}_{\Gamma_{PQ}}|\tilde{\Lambda}_{\Gamma_{PQ}}|\tilde{P}^{-1}_{\Gamma_{PQ}} \tag{14}$$

where $\tilde{\Lambda}_\Gamma$ is the diagonal eigenvalue matrix of $\tilde{A}_{\Gamma_{PQ}}$ whereas $\tilde{P}_\Gamma$ and $\tilde{P}^{-1}_\Gamma$ are the diagonalization matrices composed of the right and left eigenvectors, respectively. Subscript Γ denotes that the matrices are derived from the preconditioned system and any quantity marked with $\tilde{}$ is Roe–averaged. Through eqs. 13 and 14, eq. 12, becomes

$$\overrightarrow{\Phi}_{PQ} = \frac{1}{2}[H(\overrightarrow{W}_P, \vec{n}_{PQ}) + H(\overrightarrow{W}_Q, \vec{n}_{PQ})] - \frac{1}{2}\tilde{\Gamma}^{-1}_{PQ}|\tilde{A}_{\Gamma_{PQ}}|\Delta\overrightarrow{W}_{PQ} \tag{15}$$

According the stability criteria applied for the preconditioned system, the local time step is

$$\Delta t = \frac{CFL\ h_T}{\frac{1}{2}\{(1+a)|\vec{v}|\ +\ \sqrt{[(1-a)|\vec{v}|]^2\ +\ 4ac^2|\vec{n}|^2}\}} \tag{16}$$

which is simplified to $\Delta t = \frac{CFL\ h_T}{|\vec{v}|+c|\vec{n}|}$ for the non preconditioned case ($a = 1$). In the 2D case, eq. 16 is used to compute Δt at each triangle using its minimum height h_T and the CFL number defined by the user. Time steps are then scatter–added to nodes.

4 RESULTS AND DISCUSSION

A number of test problems has been selected to demonstrate the capability of the programmed software to cope with all–speed flows. Our intention is not to demonstrate how accurate the computed results are, but to convince the reader that, through preconditioning, one can exploit software based on the theory of hyperbolic equations even in low Mach number flows where, by nature, time–marching methods are slow. No comparison with experimental or other reference computational results is shown, since the non–preconditioned method was adequately validated in the past at high subsonic and transonic flows. The demonstration that follows is concerned with inviscid and viscous flows, around an isolated airfoil, in a compressor cascade and around a complete aircraft. Any comparison concerning convergence speed is presented as a function of iterations. The CPU cost per iteration of the preconditioned system of equations is slightly higher than that of the non–preconditioned one, due to the excess number of floating point operations it involves. However, the difference in CPU cost is almost negligible and, thus, the comparison in terms of iterations can be interpreted as a comparison in terms of cost.

The first problem is concerned with the computation of the inviscid flow around the isolated NACA12 airfoil. The same unstructured grid (950 nodes, 1800 triangles) was used to predict the flow field at three different infinite Mach numbers, namely $M_\infty = 0.1$, 0.01 and 0.005, with the same infinite flow angle $\alpha_\infty = 5^o$. Fig. 3 compares the residual drop in terms of iterations. Note that no stopping criterion was used, so any comparison between the performance of the non-preconditioned and preconditioned equations can be objectively quantified, depending on the desired maximum allowed residual. In all cases, preconditioning leads to better convergence characteristics; it is clear that the lower the Mach number the lower the computing cost.

In the $M_\infty = 0.1$ case, the gain achieved through preconditioning is not that important, though it does exist. For higher M_∞ values, both systems give very similar convergence characteristics and this can

be explained by the precondition matrix form, eq. 6, which tends to the unit matrix. By comparing the residual curves of the preconditioned equations in all three cases, it can be seen that they remain close to each other whereas the convergence of the non-preconditioned equations drifts much more slowly as the Mach number decreases. In the cases $M_\infty = 0.01$ and 0.005, assuming a stopping criterion for the residual equal to 10^{-10}, preconditioning leads to convergence of about five to eight times lower CPU cost.

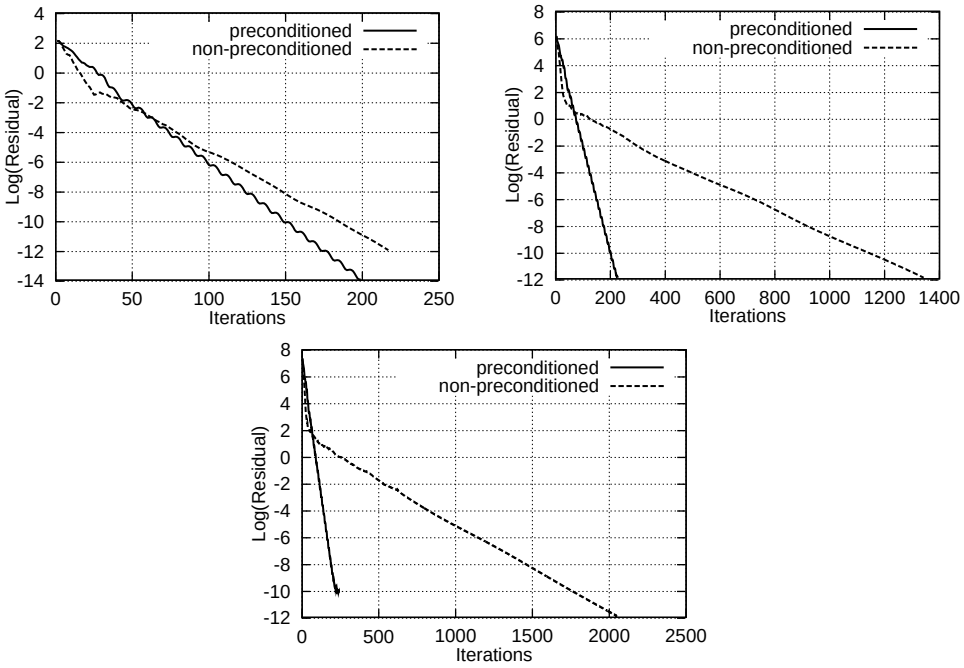


Figure 3: Inviscid flow around the NACA12 airfoil. Convergence diagrams for $M_\infty = 0.1$ (top–left), $M_\infty = 0.01$ (top–right) and $M_\infty = 0.005$ (bottom).

Fig. 4 shows a close up view of the unstructured grid used and the Mach number contours computed through the preconditioned equations around the airfoil, for $M_\infty = 0.005$.

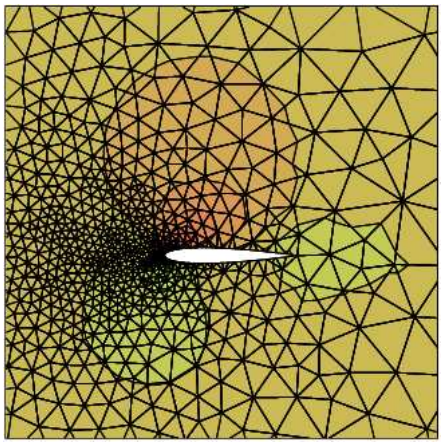


Figure 4: Inviscid flow analysis for NACA12 airfoil. Mach number contours for $M_\infty = 0.005$, and Mach number increment $\Delta M = 0.0002$.

Fig. 5 compares the pressure distribution around the airfoil, produced by the preconditioned and non–preconditioned solver with the same computing cost. Close to the leading and trailing edges, the preconditioned equations eliminate non–physically accepted pressure kinks, thanks to the modified last term in eq. 12 which acts as smoothing term.

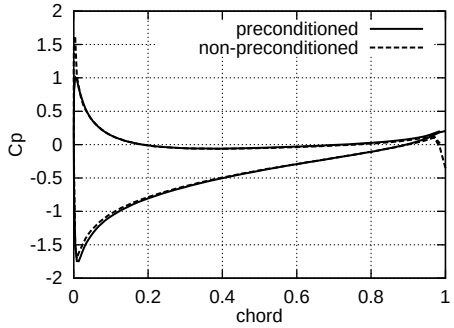


Figure 5: Inviscid flow analysis for NACA12 airfoil. Pressure coefficient distribution ($M_\infty = 0.1$).

The second problem examined is that of the inviscid flow analysis in a 3D compresor cascade. The exit isentropic Mach number equals to 0.1 or 0.3 and $a_1 = 47^o$. A 2D unstructured grid (1800 nodes, 3300 triangles) was generated at first which was then stacked in the spanwise direction to create the 3D unstructured grid with 40000 tetrhedra and 9000 nodes. Symmetry conditions were employed over the upper and the lower plane in the spanwise direction. Fig. 6 shows the residual convergence history for $M_{2,is} = 0.3$ and 0.1. In the high Mach case, both solvers, either with or without preconditioning, converge easily ; however, the preconditioned equations converge faster. In the low Mach number case, the preconditioned equations solver converges within 400 iterations (the stopping criterion for the residual is the same as in the previous case); on the other hand, without preconditioning, even 4000 iterations do not suffice to get an adequately converged solution.

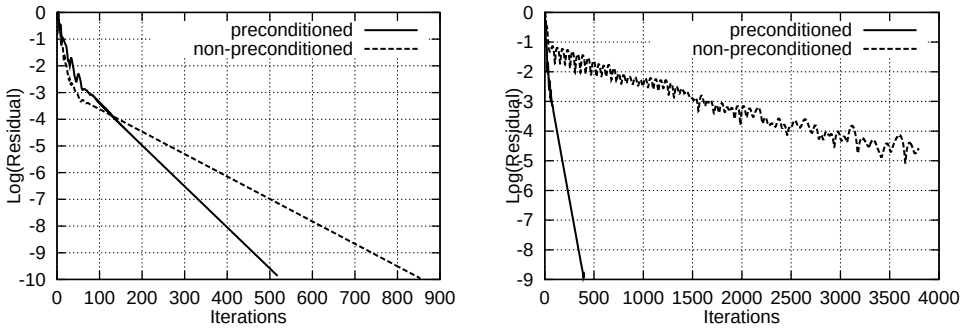


Figure 6: Inviscid flow in a 3D compressor cascade. Convergence diagrams for $M_{2,is} = 0.3$ (left) and $M_{2,is} = 0.1$ (right).

The turbulent flow over the same linear cascade was also analysed. The 2D grid consists of 5000 nodes and 9400 triangles, which results to a 3D grid with 170000 tetrahedra and 34700 nodes. The Spalart–Allmaras one–equation turbulence model [10] was used along with the wall function technique; in particular, a non zero (slip) velocity was allowed to occur over the blade nodes through assuming that the real solid wall is located at distance δ from the boundary node; δ is a user-defined parameter so that boundary nodes be in the logarithimic region of the boundary layer. Fig. 7 shows the calculated Mach number contours and compares the convergence of the preconditioned and non–preconditioned equations for $Re = 100000$ and $M_{2,is} = 0.1$. With the maximum allowed residual value be equal to 10^{-5}, the preconditioned equation converge at half the CPU cost of the conventional solver. The lower the stopping residual threshold, the more important the CPU gain.

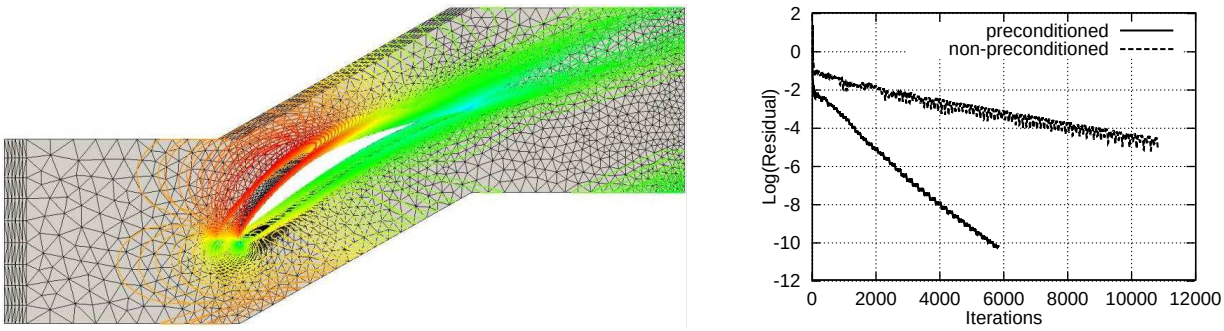


Figure 7: Turbulent flow in a 3D compressor cascade ($Re = 100000$). Mach number contours of the preconditioned system (left) and convergence diagram (right).

Last case is that of the computation of the inviscid flow around a complete aircraft. The computational grid generated around half of the aircraft (due to symmetric flow conditions) consists of 256000 tertahedra and 45000 nodes.

This case was studied for $M_\infty = 0.1$ and $\alpha_\infty = 0^o$. Fig. 8 shows the convergence curves for the preconditioned and non–preconditioned equations and the iso–Mach contours over the aircraft surface. The preconditioned equations converge faster and the gain in CPU is expected to increase at lower Mach numbers.

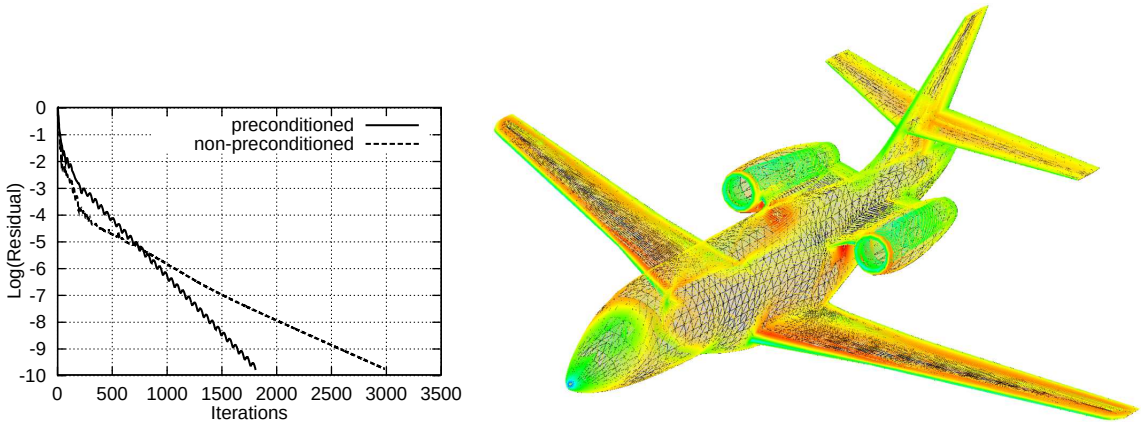


Figure 8: Inviscid flow around an aircraft. Convergence diagram (left) and iso–Mach contours for $M_\infty = 0.1$ (right).

5 CONCLUSIONS

The implementation of low–Mach preconditioning in a time–marching, primitive variable flow solver can increase its robustness, yielding equally satisfactory convergence at all flow speeds. A couple of assumptions is made during this implementation in the contest of a second–order upwind scheme, as demonstrated in this paper. These assumptions often affect positively the accuracy of the predictions by improving the quality of the solution in areas close to leading and trailing edge of an airfoil. The proposed method leads to a considerable economy in CPU cost which becomes more important as the Mach number decreases.

AKNOWLEDGEMENTS

The first two authors were supported by grants from the Research Committee of the National Technical University of Athens and the Beneficial Foundation Alexandros S. Onasis.

References

[1] Turkel E., Radespiel R., Kroll N. " Assesment of Preconditioned Methods for Multidimensional Aerodynamics", Computers & Fluids, Vol. 26, No.6, pp. 613-634, 1997.

[2] van Leer B., Lee W.T., and Roe P., "Characteristic Time-Stepping or Local Preconditioning of the Euler Equations", AIAA Paper 91-1552, 1991.

[3] Choi Y.H., and Merkle C.L., "The Application of Preconditioning in Viscous Flows", Journal of Computation Physics, Vol. 105, No. 2, pp. 207-233, 1993.

[4] Turkel E., "Preconditioned Methods for Solving the Incompressible and Low Speed Compressible Equations", Journal of Computation Physics, Vol. 72, No.2, pp. 277-298, 1987.

[5] Weiss J.M., Smith W.A., "Preconditioned Applied to Variable and Constant Density Flows", AIAA, Vol. 33, No.11, Paper 94-2209, 1995.

[6] Eriksson Lars-Erik, "A preconditioned Navier-Stokes Solver for Low Mach Number Flows", Computational Fluid Dynamics, 1996.

[7] Giannakoglou K.C., Koubogiannis D.G., Giotis A.P. and Rovas D.V., "A 3-D Solver Using Unstructured Grids with Tetrahedral Elements", Proc. of the 5th National Congress on Mechanics, Ioannina, pp. 973-980, August 1998.

[8] Koubogiannis D.G., Athanassiadis A.N. and Giannakoglou K.C., "One- and Two-Equation Turbulence Models for the Prediction of Complex Cascade Flows Using Unstructured Grids", Computers & Fluids, 32, pp. 403-430, 2003.

[9] Roe P. L. "Approximate Riemann Solvers, Parameter Vectors, and Difference Schemes", Journal of Computation Physics, Vol. 43, pp. 357-372, 1981.

[10] Spalart P. R. and Allmaras S. R., "A One-Equation Turbulence Model For Aerodynamic Flows", AIAA, Paper 92-0439, 1992.

Non-Newtonian Fluid Mechanics and Transport Phenomena

SIMULATION OF LAMINAR POLYMER MELT FLOW AND TURBULENT AIR-JET COOLING IN BLOWN FILM PRODUCTION

John Vlachopoulos and Vasiles Sidiropoulos
CAPPA-D, Chemical Engineering, McMaster University
Hamilton, Ontario, Canada L8S 4L7
Email: vlachopj@mcmaster.ca

Keywords: polymer extrusion, die flow, bubble inflation, turbulent jets

Abstract. Blown film extrusion involves melting and pumping of a polymer melt through a die (usually of the spiral variety) for the formation of a tubular film, which is stretched in the machine direction and simultaneouysly radially inflated. Cooling is accomplished by tangentially impinging external air jets and sometimes by internal vertically impinging air jets. Simulations of laminar very low Re flow through the spiral die geometry are performed for the determination of flow channel geometry that would result in even melt outlflow from the die. The bubble inflation process is described on the basis of thin membrane theory using ordinary differential equations. Bubble cooling by external and internal air jets is simulated using k-epsilon theory. The importance of the Venturi and Coanda effects in bubble stability and film production is discussed.

1. INTRODUCTION

Blown film extrusion is the most important process for the production of plastic films, ranging in thickness from 0.5 mm to as thin as 5 μm. The polymer is melted in an extruder and the hot melt is pumped through a die to form a thin walled tube, which is simultaneously axially drawn and radially expanded. In most installations the extruder(s) are horizontal and the blown film bubble is formed vertically upward as shown in figure 1.

Cooling is accomplished by air jets (frequently issuing from dual air lips) externally and vertical jets impinging on the bubble internally. The rate of cooling determines the film production rate. The air-jet velocity through the Venturi effect in narrow passages helps stabilize the thin bubble of molten polymer, which eventually solidifies at the so-called freeze-line.

2. DIE FLOW SIMULATION AND DESIGN

The purpose of an extrusion die is to impart the desired shape to the polymer melt stream produced continuously by the extruder. In blown film extrusion a thin tubular film is formed as the melt flows through the die lips. The die lip gap width usually ranges from 0.76mm to 3mm and die diameter from a few centimeters for laboratory lines to more than one meter for industrial installations producing more than one ton of film per hour. The annular flow is formed in the gap between the inner mandrel and the outer die body. By far the most common die geometry for blown film production is the spiral one, which is shown schematically in figure 2. The polymer is fed by a number of melt tubes ending with a "port" at the start of each spiral. It flows both along the spirals and in the gap between the mandrel and the die body. The flow rate becomes progressively uniform around the circumference towards the die exit.

Above the die lips the outgoing polymer stream must have uniform thickness. Uneven thickness tends to be exaggerated by the subsequent inflation and stretching operations resulting in unacceptable film production. Polymer melts tend to flow preferentially directly above the ports resulting in periodic thick and thin sections as shown schematically in figure 3. The tendency for uneven thickness distribution becomes greater as the power-law exponent in the viscosity function decreases, i.e. shear thinning increases. Thickness variation of more than 5% above the die lips is usually unacceptable. Another problem is the poor flow distribution along the spirals and between mandrel and die body. Many dies are currently in operation in industrial installations capable of producing a film of acceptable uniformity but having part of the spiral channel with stagnant flow regions. Stagnant flow regions may result in polymer degradation and other production problems.

Computer modeling is necessary for die designs producing uniform thickness without any stagnant flow regions. This can be achieved by 2-D approximations of the Hele-Shaw variety using control volumes[1,2,3] relating local pressure to local flow rates or fully 3-D finite element simulations[4] for very slow viscous (creeping) non–Newtonian flow.

3. BUBBLE FORMATION AND MODELING

As the molten polymer emerges from the annular die lips, the thin tube is undergoing non-uniform biaxial stretching. The deformation effectively ceases at the freeze line, because the stresses are not large enough to deform the solid film. For typical film production rate and thinness, it is reasonable to neglect inertia forces, surface tension, air drag and gravity on the highly viscous melt. Perhaps the neglect of gravity is the weakest assumption when high production rates are involved.

Figure 4 shows a sketch of an axisymmetric thin tubular film (of thickness h) with an embedded Cartesian system having unit vectors in the direction of flow **m** (meridional), in the tangential **t** and normal **n** directions. Mass conservation gives

$$Q = 2\pi R h V_m = \text{constant} \tag{1}$$

where Q is the volumetric flow rate, R is the bubble radius, h the bubble thickness and V_m the velocity in the meridional direction.

The equilibrium of forces in the normal direction for the thin membrane assumption is

$$\frac{\Delta P}{h} = \frac{\sigma_m}{\rho_m} + \frac{\sigma_t}{\rho_t} \tag{2}$$

where ρ_m, ρ_t are the curvatures in two directions **m** and **t** , σ_m and σ_t are the corresponding stresses and ΔP is the inflation pressure.

The force balance on any plane z gives,

$$-\pi R^2 \Delta P + 2\pi R h \sigma_m \cos\theta = F_T = \text{constant} \tag{3}$$

where F_T is the total force exerted on the thin membrane.

By determining the two radii of curvature and combining with the force balance equations we end up with a single ordinary differential equation[5,6,7] which, when solved, provides the bubble shape for a Newtonian viscosity model. The role of viscoelasticity is to decrease the ultimate bubble radius[7,8].

4. BUBBLE COOLING

The air-cooling system is an integral part of any blown film line. It greatly affects not only the heat transfer from the molten polymer film but also the stability and the shaping of the bubble. The bubble shape is primarily determined by mechanical manipulations and aerodynamics[9]. Film cooling ultimately affects both production rate and final film properties.

Two important aerodynamic phenomena may be associated with the cooling airflow, namely the *Venturi* and *Coanda* effects. The well-known Venturi effect is caused when a fluid flows through a constricted area: its speed increases and the pressure drops. In film blowing, the lower orifice air is flowing through the narrow gap between the bubble surface and the air ring cone. The resulting Venturi effect causes the bubble to be pulled towards the air ring cone, visibly deforming the bubble.

The Coanda effect occurs when a free jet emerges close to a surface: the jet tends to bend, "attach" itself and flow along the surface. The surface may be flat or curved and located inclined or offset to the jet. The Coanda effect is more pronounced near curved surfaces. Blown film bubble surfaces with the cooling air impinging on them at an angle, offer the possibility of Coanda type jet attachments and detachments.

Sidiropoulos and Vlachopoulos[10,11] studied the aerodynamics of dual orifice air-rings using a commercially available finite volume software package to solve the Reynolds averaged Navier-Stokes equations.

$$\frac{\partial(\rho U_i)}{\partial t} + \frac{\partial(\rho U_i U_j)}{\partial x_j} = \frac{\partial}{\partial x_j}\left[\mu\left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i}\right) - \left(\frac{2}{3}\mu\frac{\partial U_k}{\partial x_k}\right)\right] - \frac{\partial p}{\partial x_i} + \rho g_i + \frac{\partial}{\partial x_j}\left(-\rho\overline{u_i' u_j'}\right) \tag{4}$$

In these equations the effect of turbulence is included through the Reynolds stresses ($-\rho\overline{u_i' u_j'}$), which involve products of the velocity fluctuations. A variation of the so-called k-ε turbulent model was used to achieve closure of the system of differential equations. In this approach the contribution of Reynolds stresses to the momentum balance is introduced through the concept of the "effective" viscosity. Although there are more accurate models to simulate the airflow, the results are often marginally different to justify the increased complexity of the simulation .

Figure 5 shows a simulation example depicting the calculated airflow streamlines around a typical blown film bubble. The two simulated airflows (presented on the left and right side of the same figure) correspond to slightly different setups for the adjustable part of the air ring. Although the geometrical differences

are minute, the simulation predicts significant differences in the local airflow pattern, which consequently result in differences in local heat fluxes.

The numerical simulations explained the importance of under-pressure generated by the Venturi effect, which is large enough to force the bubble to take the shape of the air ring. It appears that the Venturi effect overwhelms any tendency for different bubble shapes dictated by viscoelasticity. Generally, the Venturi effect is used to stabilize the process. However, it is known from industrial practice that sometimes even small adjustments to the airflow may cause bubble instability.

Internal Bubble Cooling (frequently called IBC) of blown film bubbles employs various mechanisms to exchange the internal air of the bubble. Naturally, this helps to increase the cooling of the film, which would ultimately lead to increased production rates. IBC typically involves specifically designed equipment, engaged in exchanging the warm internal air with colder external air and also constantly circulating and mixing the internal air. In some cases the external air is chilled before injected inside the bubble to maximize the cooling benefit. With IBC, the expected production rate improvement becomes increasingly important as die size and film layflat width increase. Production rate increases range from 20% for small bubbles (up to 8´´ in die diameter) to as much as 80% for very large bubbles. A simulation of IBC has been published[12].

5. CONCLUDING REMARKS

The use computer simulation for the determination of flow channel geometry to produce uniform outflow of a thin polymer tube and stagnation-free regions is absolutely necessary for successful die designs. The bubble inflation modeling provides insights into the polymer deformation process, but the bubble shape is primarily determined by mechanical manipulations and the Venturi effect. Turbulent air jet simulations provide significant information on air ring and internal bubble cooling system design.

6. REFERENCES

1. O'Brien, K.T. (Ed) (1992), "Applications of CAE in Extrusion and Other Continuous Processes", Hanser Publishers, Munich
2. Perdikoulias, J., Vlachopoulos, J., and Vlcek, J., "Spiral Die Analysis", Chapter 2.2, pp. 39-57 in Kanai, T., and Campbell, G.A. [eds] (1999), "Film Processing", Hanser Publishers, Munich
3. Perdikoulias, J., Vlcek, J., and Vlachopoulos, J. (1987),`` Polymer Flow through Spiral Mandrel Dies" Adv. Polym. Technol., 7, p.333
4. Coyle, D.J., and Perdikoulias, J. (1993).,``Flow simulation and Visualization in Spiral Mandrel Dies'' SPE Antec Tech. Papers, p.2445
5. Pearson, J. R. A., and Petrie, C. J. S. (1970), "The Flow Of A Tubular Film. Part 1. Formal Mathematical Representation", J. Fluid Mech., 40, p.1
6. Pearson, J. R. A., and Petrie, C. J. S. (1970), "The flow of a tubular film. Part 2. Interpretation of the model and discussion of solutions", J. Fluid Mech., 42, p. 609
7. Tanner, R. I. (2000), "Engineering Rheology", Oxford University Press, Oxford
8. Alaie, S.M. and Papanastasiou,T.C. (1993), "Modeling of Non-Isothermal Film Blowing with Integral Constitutive Equations", Intern.Polym.Proc.,7, p.51
9. Kurtz, S. J. (1995), "Relationship of stresses in blown film processes", Int.. Polym. Proc., 10, p.148
10. Sidiropoulos, V., and Vlachopoulos, J. (2000), "An Investigation of Venturi and Coanda Effects in Blown Film Cooling", Int. Polym. Proc., 15, p.40
11. Sidiropoulos, V., and Vlachopoulos, J. (2000), "The Effects Of Dual-Orifice Air-Ring Design On Blown Film Cooling", Polym.Eng.Sci., 40,p.1611
12. Sidiropoulos, V., and Vlachopoulos, J. (2001), "Numerical Study Of Internal Bubble Cooling (IBC) In Film Blowing", Int. Polym.Proc.,16, p.48
13. Sidiropoulos, V., Tian, J. J., and Vlachopoulos, J. (1996), "Computer Simulation of Film Blowing", J. of Plast. Film & Sheet., 12,p.107

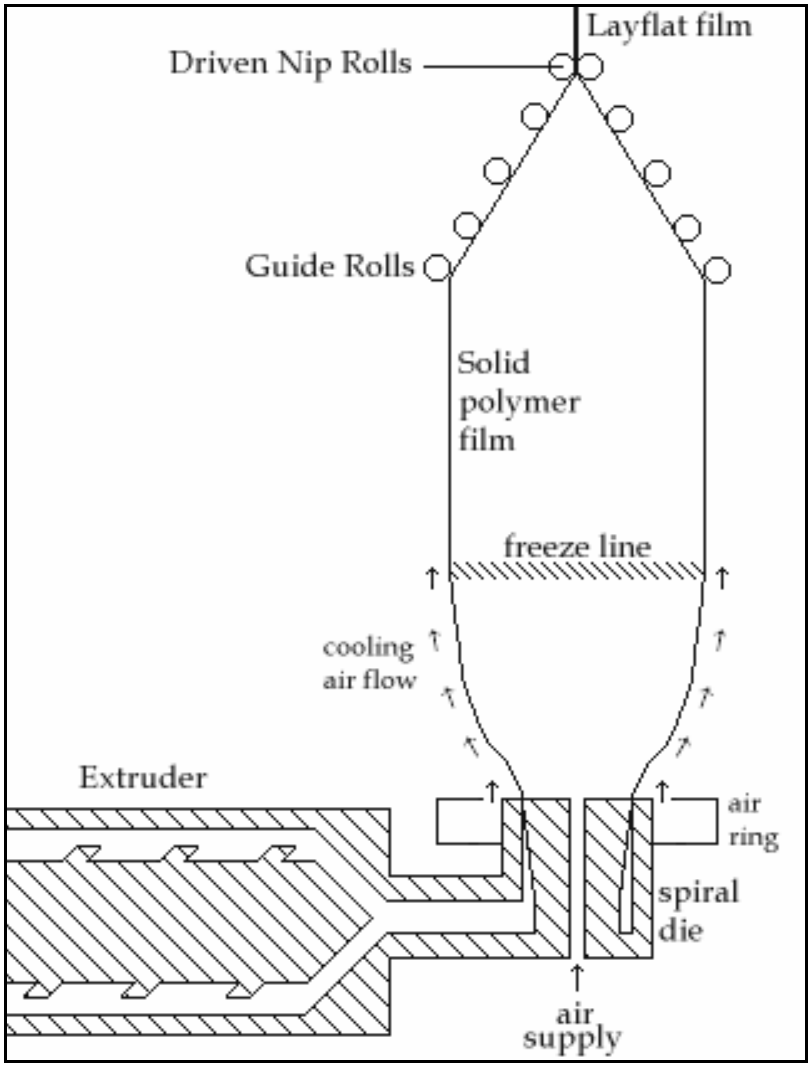


Figure 1. Schematic of the film blowing process

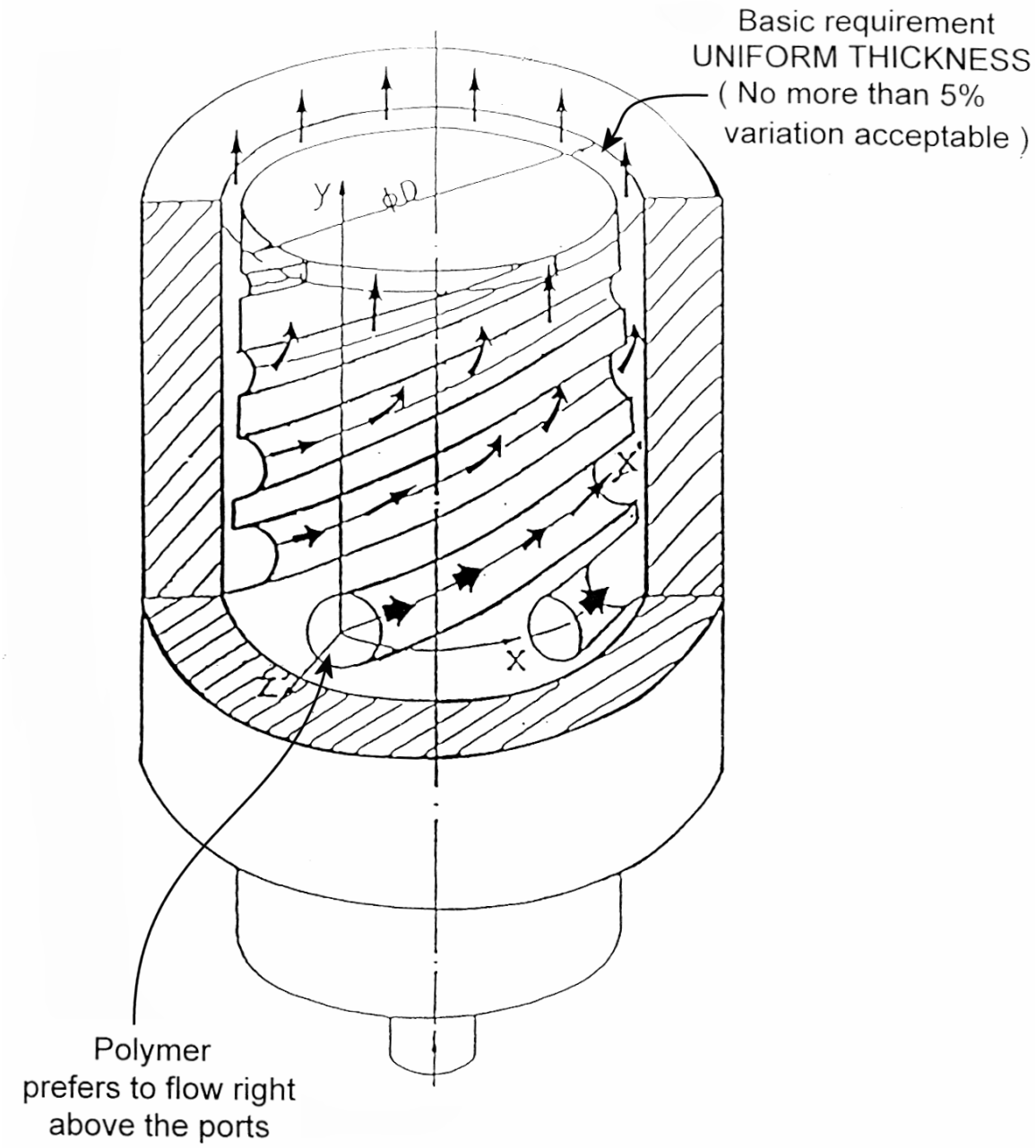


Figure 2. Schematic of a spiral mamdrel die

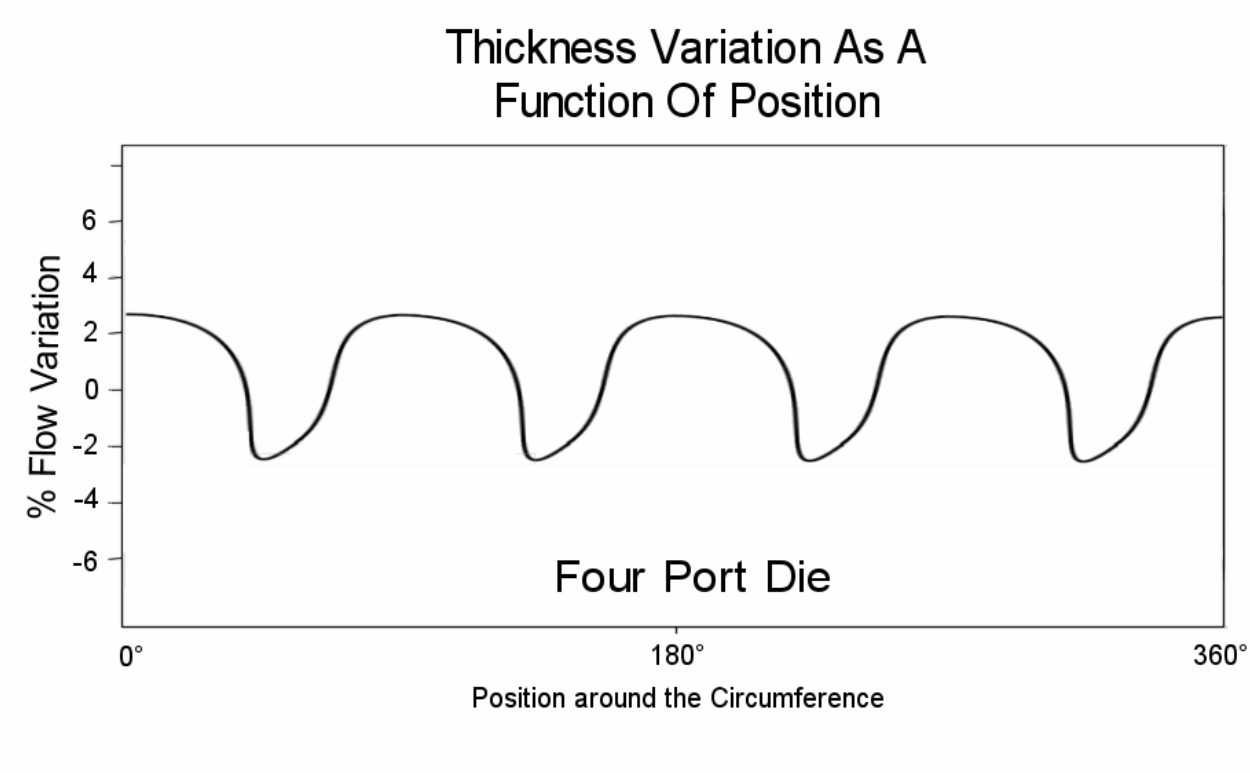


Figure 3. Thickness variation as a function of position for a four port die

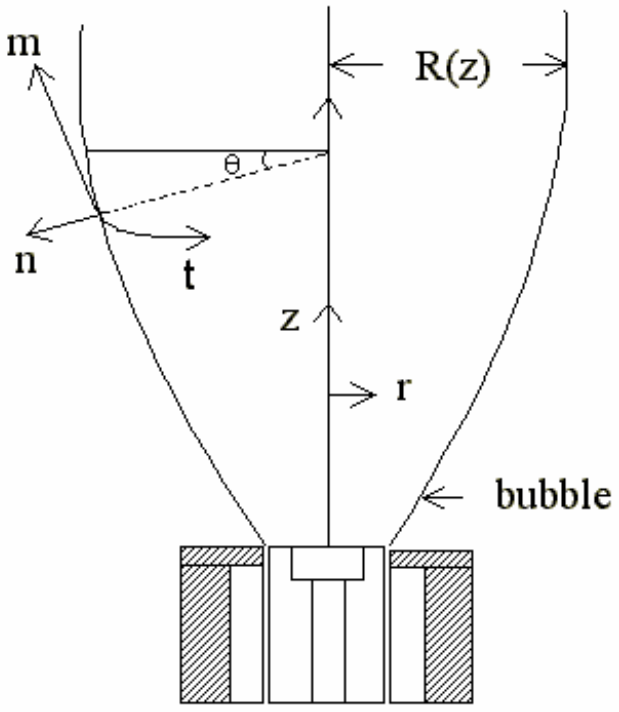


Figure 4. Blown film bubble and intrinsic coordinate system (**n, m, t)**

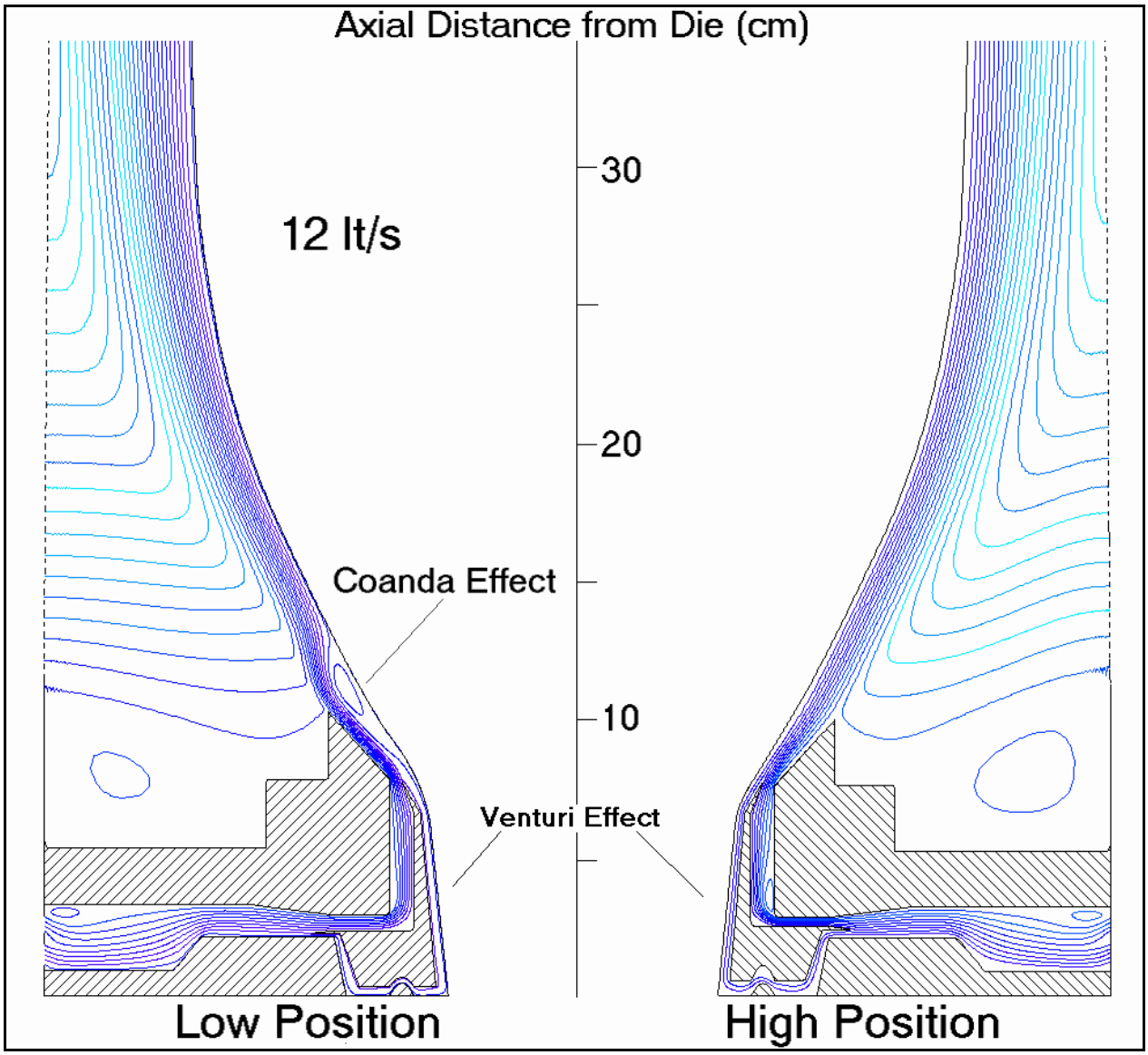


Figure 5. Cooling air streamlines around an LLDPE blown film bubble
for different setups of the adjustable ring (moderate air-flow)
left side: low position, right side: high position

SQUEEZE FLOW OF VISCOPLASTIC MATERIALS FULLY ACCOUNTING FOR THE HIGHLY DEFORMING DOMAIN OF THE MATERIAL

George Karapetsas and John Tsamopoulos

Laboratory of Computational Fluid Dynamics,
Department of Chemical Engineering,
University of Patras, Patras 26500, Greece

Keywords: Squeeze flow, viscoplastic material, yield surface, free surface

Abstract. *We examine the transient, axisymmetric, squeeze flow of viscoplastic materials, under creeping flow. Both cases of the disks moving with constant velocity and under constant force are studied. This is a transient simulation, performed for the first time for such materials, in order to determine the transient force or velocity, respectively, the shape of the liquid/air interface and the effect of gravity very accurately. All these are impossible with the quasi steady state models used up to now. For the simulation of the viscoplastic behaviour we employ the approximate model suggested by Papanastasiou. As a numerical scheme we use the mixed finite element method coupled with a quasi-elliptic mesh generation scheme in order to follow the large deformations of the flow domain. We show that unyielded material arises around the two stagnation points of flow at the center of the disks and that the size of these domains increases with the Bingham number. When the disks move with constant velocity the unyielded region decreases, as time passes and the distance between the two disks decreases. On the other hand, when a constant force is applied on the disks, they slow down until they finally stop, because eventually all the material between them becomes unyielded.*

1. INTRODUCTION

Squeeze flow is widely used as a typical experiment for the rheological characterization of Non-Newtonian fluids [1]. The rheological experiment is usually conducted in two different ways, either by measuring the force required to push the disks at a constant velocity [2] or by measuring the velocity of the disks towards each other when a constant force is applied [3]. Our efforts are aimed at developing an accurate and efficient numerical method in order to simulate both versions of such a rheological experiment for the special case of a viscoplastic material. The deformation and flow of such materials are very important since many multicomponent fluids such as suspensions, pastes, paints, foodstuffs, foams and slurries, which are very frequently encountered in industrial processes, are viscoplastic [4]. When the applied stress is below a particular value, which is called yield stress, these materials exhibit small or no deformation at all (solid like behavior). Above the yield stress these materials flow with viscosity that depends on the local shear rate.

The behavior of viscoplastic materials in squeeze flow has attracted the attention of several researchers in the past. In early theoretical studies, the most usual approach was the lubrication approximation [5]. However, although it was noticed since then that this kind of analysis led to profound kinematic inconsistencies in the calculated velocity fields, it is still used extensively to evaluate experimental results. O'Donovan & Tanner [6] were the first who recognized the need to solve the squeeze film problem without assuming the lubrication approximation. They solved numerically the constant velocity problem employing the biviscosity model for modeling the viscoplastic behavior of the material and they finally concluded that unyielded material arise only adjacent to the centre of the plates. Quite recently Smyrnaios & Tsamopoulos [7] provided a qualitative analysis and numerical simulations for this problem assuming a quasi steady state model. They employed both the original Bingham model [8] and the exponential one, which was proposed by Papanastasiou [9], clarifying all the misleading speculations on the position and the existence of yield surfaces in this flow field. They showed that unyielded material could only exist around the two stagnation points of flow confirming and extending the work by O'Donovan & Tanner [6]. Matsoukas & Mitsoulis [10] also solved numerically the squeeze flow of viscoplastic materials, assuming steady state, for both planar and axisymmetric flow confirming the earlier results by Smyrnaios & Tsamopoulos [7] for the axisymmetric case. They have also provided a more exact relation for the force that must be applied on the disks' surface in order to maintain their constant velocity.

In the present work we solve the transient squeeze flow of a viscoplastic material for both cases where the disks are moving with constant velocity and under constant force. The transient simulation which is done for the first time for such fluids enables the determination of the transient force or velocity, respectively, the shape of the liquid/air interface, and the location of the yield surface. Moreover we are able to study the effect of possible slip of the fluid on the disk surface and the effect of gravity on the flow field.

2. PROBLEM FORMULATION

We consider the axisymmetric squeeze flow of a viscoplastic material with a constant yield stress, τ_y, and upon yielding a constant dynamic viscosity, η_o. We assume that the fluid has a surface tension, σ and that it is incompressible with constant density, ρ. Fig. 1 illustrates a schematic of the flow geometry examined herein: the space between two parallel coaxial disks, whose radial dimension is represented with a, is filled with a cylindrical sample of a viscoplastic material. The radius of this sample is denoted with b. Initially the disks are not moving and so the free surface of the fluid is assumed to be straight. At the same time, the pressure inside the fluid is considered to be uniform, assuming that the ambient pressure is zero.

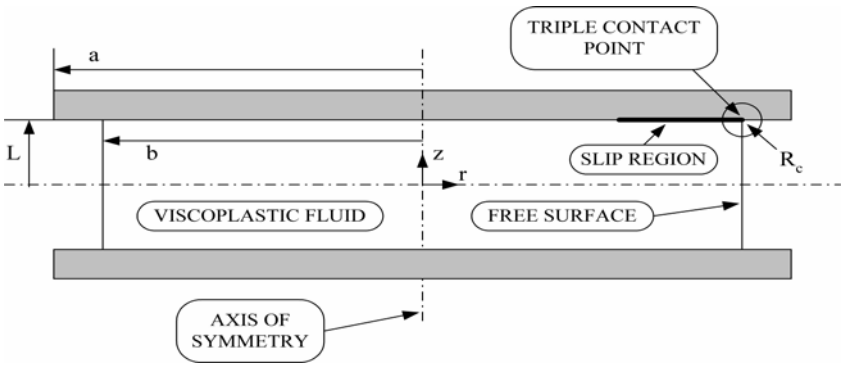


Figure 1. Schematic of the squeeze flow geometry between two parallel coaxial disks.

We scale all lengths with half the initial distance of the two disks, L, and time with L/V^*, where V^* is the characteristic velocity. For the constant velocity problem the choice of the characteristic velocity V^* is rather obvious and it is the constant velocity of the disks V. On the other hand although no characteristic velocity exists for the constant force problem, due to the decelerating nature of the flow, we have chosen as such the initial velocity of the disks $V(t=0^+)$. In addition, both pressure and stress components are scaled with viscous scale $\frac{\eta_o V^*}{L}$. Thus the dimensionless groups that arise are the Reynolds number $Re=\frac{\rho VL}{\eta_o}$, the Bingham number $Bn=\frac{\tau_y L}{\eta_o V}$, the Capillary number $Ca=\frac{\eta_o V}{\sigma}$, the Stokes number $St=\frac{\rho g L^2}{\eta_o V}$, when gravity is taken under consideration, and finally the aspect ratios $\varepsilon=\frac{L}{b}$ and $w=\frac{a}{L}$.

The flow of an incompressible fluid is governed by the momentum and mass conservation equations which in their dimensionless form are:

$$Re\frac{D\underline{v}}{Dt}+\underline{\nabla}P+\underline{\nabla}\cdot\underline{\underline{\tau}}-St\,\underline{e}_z=0 \tag{1}$$

$$\underline{\nabla}\cdot\underline{v}=0 \tag{2}$$

where $\underline{\underline{\tau}}$ is the viscous part of the total stress tensor $\underline{\underline{\sigma}}$,

$$\underline{\underline{\sigma}}=P\underline{\underline{I}}+\underline{\underline{\tau}}, \tag{3}$$

while $\underline{v}$, P are the axisymmetric velocity vector and the pressure respectively.

To complete the description of the flow problem specification of a constitutive equation which describes the rheology of the fluid is required. In the present study we employ the continuous constitutive equation that has been proposed by Papanastasiou [9] which relates the stress tensor, $\underline{\underline{\tau}}$, to the rate of strain tensor, $\underline{\underline{\dot{\gamma}}}$, by a simple exponential relation. The dimensionless form of this constitutive equation is

$$\underline{\underline{\tau}}=-\left[1+Bn\frac{1-e^{-N\dot{\gamma}}}{\dot{\gamma}}\right]\underline{\underline{\dot{\gamma}}} \tag{4}$$

where $\underline{\underline{\dot{\gamma}}}$ is the rate of strain tensor defined as $\underline{\underline{\dot{\gamma}}}=\underline{\nabla}\underline{v}+\underline{\nabla}\underline{v}^T$, $\dot{\gamma}$ is its second invariant, $\dot{\gamma}=[\frac{1}{2}\underline{\underline{\dot{\gamma}}}:\underline{\underline{\dot{\gamma}}}]^{1/2}$ and N is the stress growth exponent $N=\frac{mV}{L}$.

Smyrnaios & Tsamopoulos [7] have shown that for relatively large values of the exponent coefficient, N, this model closely approximates the discontinuous Bingham behavior. However Burgos et al. [11] have suggested that extremely large values of that coefficient should be avoided for reasons that have to do with numerical stability and the stiffness of the resulting discrete system. The main advantage of this constitutive equation is the fact that it is continuous and it holds uniformly in yielded and unyielded regions. Thus the determination of the shape and the location of the yield surface can be done a posteriori in contradiction to the discontinuous Bingham model. The criterion for this determination has the following form

Yielded material: $\tau > Bn$ (5)

Unyielded material: $\tau \le Bn$ (6)

Along the free surface the velocity field should satisfy a local force balance between the surface tension and the viscous stresses in the liquid

$$\underline{n} \cdot \underline{\underline{\sigma}} = \frac{2H}{Ca}\underline{n} \tag{7}$$

where $\underline{n}$ is the outward unit normal to the free surface and $2H$ is its mean curvature which is defined as

$$2H = -\underline{\nabla}_s \cdot \underline{n}\,, \quad \underline{\nabla}_s = (\underline{\underline{I}} - \underline{n}\underline{n}) \cdot \underline{\nabla} \tag{8}$$

Taking the tangential and normal to the free surface components of the force balance we get

$$\underline{t}\underline{n} : \underline{\underline{\sigma}} = 0 \tag{9}$$

$$\underline{n}\underline{n} : \underline{\underline{\sigma}} = \frac{2H}{Ca} \tag{10}$$

In addition, the boundary conditions that must be imposed on the axis of symmetry (r = 0) and on the plane of symmetry (z = 0) are

$$\underline{n} \cdot \underline{v} = 0 \tag{11}$$

$$\underline{t}\underline{n} : \underline{\underline{\sigma}} = 0 \tag{12}$$

On the surface of the disk, two boundary conditions are imposed. On the axial direction the boundary condition depends actually on the problem that is examined each time. That is for the constant velocity problem the fluid has the same axial velocity with the disks, and thus

$$\underline{n} \cdot \underline{v} = -1 \tag{13}$$

while for the constant force problem the condition that must be imposed is

$$F - \int_0^{R_c} 2\pi\sigma_{zz} r dr = 0 \tag{14}$$

In the present study we also examine the effect of slip of the fluid on the surface of the disks. To this end we adopt a modification of the slip model that was originally proposed by Kamal et al. [12]. The slip coefficient is an exponential function of the radial distance from the triple contact point and thus the slip model has the following form

$$\underline{t}\underline{n} : \underline{\underline{\sigma}} = \beta_{sl} e^{(-\alpha_{sl}(r-R_c))}\, \underline{t} \cdot \underline{v} \tag{15}$$

where R_c is the radial distance of the triple contact point from the axis of symmetry, β_{sl} is a parameter which is used to adjust the level of slip velocity in comparison to the wall shear and α_{sl} is another parameter which is used to adjust the length of the slip region. The model reduces to the no-slip condition as these two parameters increase whereas when α_{sl} becomes zero, slip occurs over the entire wall boundary.

The model is completed by assuming that the fluid initially is at rest $\underline{v}(r, z, t = 0) = 0$, the free surface is flat with radius b and finally that the fluid is under constant pressure $P(r, z, t = 0) = \frac{\varepsilon}{Ca}$.

3. NUMERICAL IMPLEMENTATION

In order to numerically solve the above set of equations we have chosen the mixed finite element method combined with an elliptic grid generation scheme for the discretization of the transient physical domain.

3.1 Elliptic grid generation

In the present study we employ a quasi elliptic mesh generation scheme that has been recently developed and applied in several cases by Dimakopoulos & Tsamopoulos [13]. With this scheme the time dependent physical domain (r, z) is mapped onto a fixed with time computational one (η, ξ). As such computational domain we

choose the volume that is initially occupied by the fluid. This mapping is based on the solution of the following system of quasi elliptic partial differential equations

$$\underline{\nabla}\cdot\left(\varepsilon_1\sqrt{\frac{r_\xi^2+z_\xi^2}{r_\eta^2+z_\eta^2}}+(1-\varepsilon_1)\right)\underline{\nabla}\xi=0 \tag{16}$$

$$\underline{\nabla}\cdot\underline{\nabla}\eta=0 \tag{17}$$

where the subscripts denote the differentiation with respect to the variable indicated and ε_1 is a parameter that is adjusted by trial and error. In order to solve the above system of differential equations appropriate boundary conditions must be imposed. On the fixed boundaries we impose the equations that define their position and the remaining degrees of freedom are used for the equidistribution of the nodes along the boundaries. In addition, along the moving interface we impose the kinematic equation

$$\frac{D\underline{F}}{Dt}=\underline{v} \tag{18}$$

where $\underline{F}=r\underline{e_r}+z\underline{e_z}$ is the position vector of the free surface, together with a condition that requires the uniform distribution of the nodes along the free surface.

3.2 Mixed finite element method

The computational domain is discretized using triangular elements because they can conform better to the large deformations of the transient physical domain. We approximate the velocity vector with quadratic basis functions, ϕ^i, and the pressure as well as the position vector with linear basis functions, ψ^i. We employ the finite element/Galerkin method which results into the following weak forms of the momentum and mass balances

$$\int_\Omega\left[\mathrm{Re}\frac{D\underline{v}}{Dt}\phi^i+\nabla\phi^i\cdot\underline{\underline{\sigma}}\right]d\Omega+\int_\Gamma\left[\underline{n}\cdot\underline{\underline{\sigma}}\right]\phi^i d\Gamma=0 \tag{19}$$

$$\int_\Omega\psi^i\nabla\cdot\underline{v}\,d\Omega=0 \tag{20}$$

where $d\Omega$ and $d\Gamma$ are the differential volume and surface area respectively. The surface integral that appears in the momentum equation is split into four parts, each one corresponding to a boundary of the physical domain.
We must also derive the weak forms of the mesh generation equations. Thus after applying the divergence theorem we get

$$\int_\Omega\left(\varepsilon_1\sqrt{\frac{r_\xi^2+z_\xi^2}{r_\eta^2+z_\eta^2}}+(1-\varepsilon_1)\right)\underline{\nabla}\xi\cdot\underline{\nabla}\psi^i d\Omega=0 \tag{21}$$

$$\int_\Omega\underline{\nabla}\eta\cdot\underline{\nabla}\psi^i d\Omega=0 \tag{22}$$

The resulting set of algebraic equations is solved with the following two-step Newton-Raphson/Non-linear Gauss Seidel iteration scheme. At each time step the momentum and mass balances are solved until convergence assuming the physical domain from the previous time step. Thus once the velocities and pressure are known, the new locations of the mesh points can be determined from the mesh generation equations and their boundary conditions. This procedure continues until convergence is achieved for both the flow and mesh equations. Finally the set of algebraic equations is integrated in time with the Predictor-Corrector Euler method using an automatic adaptation of the time step for optimizing the code performance. The iterations are terminated using a tolerance for the absolute error of the residual vector which is set to 10^{-9}.

4. RESULTS AND DISCUSSION

4.1 Constant velocity problem

Fig. 2 illustrates the flow field in a volume of a viscoplastic material that initially fills only partially the space between the two disks, that is when $a>b$. The snapshots are taken at times $t=0.2$, $t=0.5$, $t=0.6$ and each one of them shows the contour plots of the axial velocity, on the upper half, and the radial velocity, on the lower half, for $Bn=10$, $\mathrm{Re}=0$, $Ca=10^3$, $St=0$, $w=15$, $\varepsilon=0.1$, $\alpha_{sl}=5$, $\beta_{sl}=1$, $N=500$. At early times the fluid moves inside the disks while near the end of the simulation it exits from them. On the left hand side we can

see the axis of symmetry while on the right hand side we see the free surface of the fluid. We observe that the free surface, which initially was straight, deforms everywhere even at small times. The axial velocity has its smallest (negative) values near the disk walls because the fluid follows the motion of the disks in the axial direction and its values monotonically increase towards zero at the mid-plane.

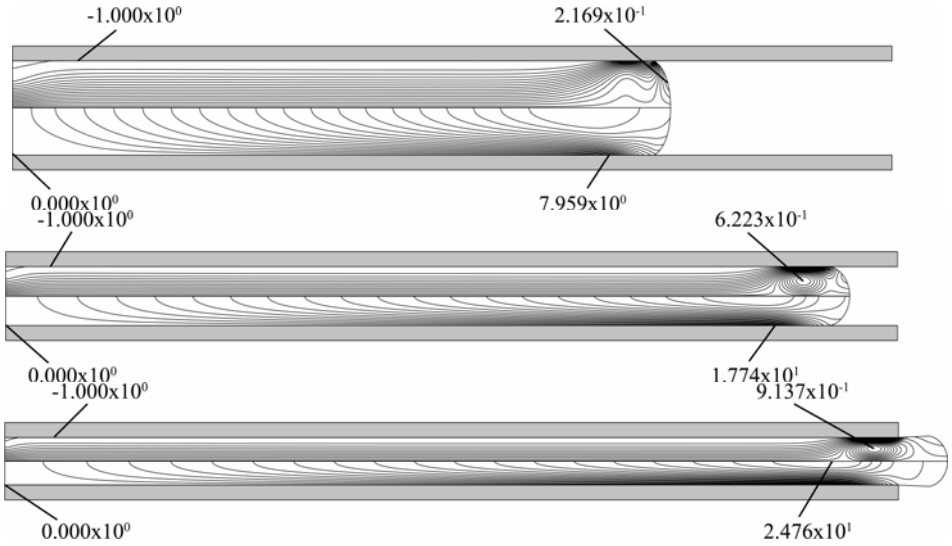


Figure 2. Contour plots of the axial, upper half, and the radial, lower half, velocity component at $t = 0.2$, $t = 0.5$, $t = 0.6$ and for $Bn = 10$, $\mathrm{Re} = 0$, $Ca = 10^3$, $St = 0$, $w = 15$, $\varepsilon = 0.1$, $\alpha_{sl} = 5$, $\beta_{sl} = 1$, $N = 500$

As for the radial velocity component on the axis of symmetry it becomes zero because of the symmetry. Moreover, the slip condition which was introduced in this simulation divides the surface of the disks into a noslip region where the radial velocity is zero and a slip region near the triple contact point. In the slip region, although the disks force the fluid to move in the axial direction, the fluid near the disks follows only partly that motion because it is also allowed to slip in the radial direction. Another significant difference is observed on the axial velocity field near the axis of symmetry. The axial velocity of the fluid remains equal to the velocity of the disk up to almost half the distance of the disk from the mid-plane and this happens because unyielded material arises in that region. Smyrnaios & Tsamopoulos [7] have also noticed a similar dependence of the size of these domains on the aspect ratio. It must be noted though that the different aspect ratios in their quasi steady state analysis correspond to different time moments in our transient simulation.

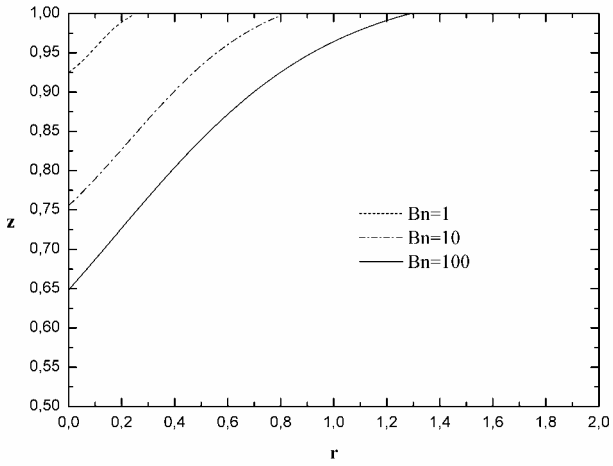


Figure 3. Yield surfaces for $\mathrm{Re} = 0$, $Ca = 10^3$, $St = 0$, $w = 10$, $\varepsilon = 0.1$, $\alpha_{sl} = 10$, $\beta_{sl} = 1$, $N = 500$ and various Bn numbers at time $t = 10^{-5}$

Fig. 3 shows the shape of the yield surface as a function of the Bn number for $\mathrm{Re} = 0$, $Ca = 10^3$, $St = 0$, $w = 10$, $\varepsilon = 0.1$, $\beta_{sl} = 1$, $\alpha_{sl} = 10$ and $N = 300$ at time $t = 10^{-5}$. As we can see in the figure the

viscoplasticity of the material affects significantly the size of the unyielded regions that arise in the flow domain. More specifically their size increases substantially with the increase of the Bn number.

We also examined the effect of the slip coefficients α_{sl} and β_{sl} on the flow field. When the value of β_{sl} increases the level of the slip velocity, near the triple contact point, decreases which has an effect on the shape of the fluid/air interface. However the velocity field away from the triple contact point is not affected significantly. The other parameter, α_{sl}, controls the rate of decrease of the slip velocity and thus it can be used for adjusting the length of the slip region. The smaller the value of α_{sl} the larger the slip region becomes. Simulations have shown that α_{sl} has a very small effect on the shape of the fluid/air interface. However its effect on the domain of the unyielded region is rather important. Fig. 4 shows the shape of the yield surface for two different parameters of α_{sl}.

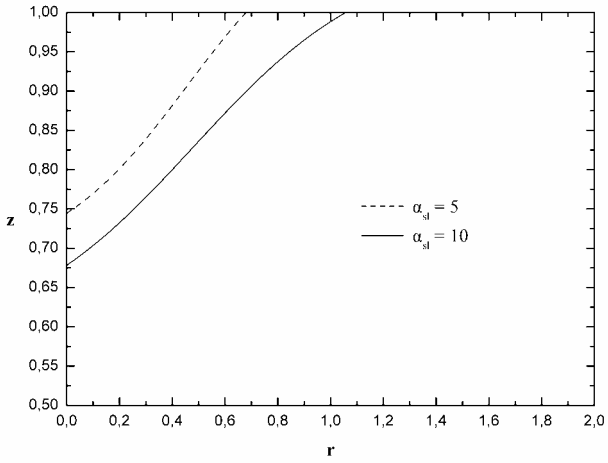


Figure 4. Yield surfaces at time $t = 10^{-5}$ for $Bn = 30$, $\mathrm{Re} = 0$, $Ca = 10^3$, $St = 0$, $w = 5$, $\varepsilon = 0.5$, $\beta_{sl} = 10$, $N = 300$ and various α_{sl}

We observe that when the slip region becomes longer the domain of the unyielded material becomes smaller or for even smaller value of α_{sl}, that is when $\alpha_{sl} = 2.5$ which is not presented here, the unyielded area totally disappears. This happens because as the slip length increases and slip is allowed at a larger portion of the disk surface, the fluid follows only partly the axial motion of the disks towards the mid-plane and flows also in the radial direction. Thus the flow becomes extensional near the disk surface and the axis of symmetry which prohibits the formation of an unyielded domain.

It is very interesting to calculate the squeeze force that must be applied on the disks in order to maintain their constant velocity because usually in such a rheological experiment the force is the only measured quantity. We calculate the force acting on the disks by integrating the normal to the disks component of the total stress tensor, σ_{zz}, over the portion of the disk surface which comes into contact with the fluid

$$F = \int_0^{R_c} 2\pi\sigma_{zz} r dr$$

Fig. 15 shows the dependence of the squeeze force on time and for various Bn numbers. The rest of the parameters are $\mathrm{Re} = 0$, $Ca = 10^3$, $St = 0$, $w = 10$, $\varepsilon = 0.1$, $\beta_{sl} = 1$, $\alpha_{sl} = 5$ and $N = 300$. We observe that the squeeze force increases significantly as time passes because of the decreased distance of the disks. Furthermore we notice that there is a substantial increase in the force that must be applied on the disks as the Bn number increases especially at early times. This dependence of the force on the viscoplasticity of the material was also noted by previous researchers in the past [7],[10]. The variation of the force is almost of one order of magnitude as the Bn number increases from 1 to 50. Thus we can say that measuring the force that must be applied on the disks to maintain their constant velocity can be an efficient way of determining the value of the yield stress for a viscoplastic fluid.

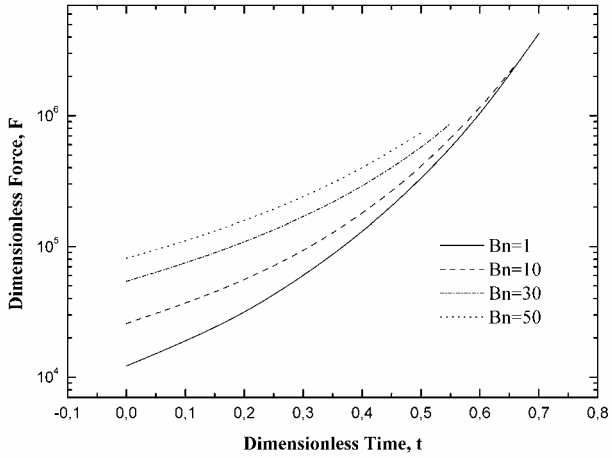


Figure 5. Evolution of the squeeze force with time for $\mathrm{Re}=0$, $Ca=10^3$, $w=20$, $\varepsilon=0.1$, $\alpha_{sl}=5$, $\beta_{sl}=1$, $N=300$ and various Bn numbers

4.2 Constant force problem

The measured quantities in a squeeze flow experiment that is conducted under constant force are the axial velocity of the disks and their position with time. The motion of the disks in such an experiment is decelerating because as they approach each other the resistance of the fluid continuously increases which makes them to slow down and eventually to stop. Fig.16 shows the evolution of the axial velocity and position of the upper disk with time for the case of a viscoplastic fluid with $Bn=10$, $\mathrm{Re}=0$, $Ca=10^3$, $St=0$, $w=5$, $\varepsilon=0.5$, $\beta_{sl}=100$, $\alpha_{sl}=5$ and $N=300$.

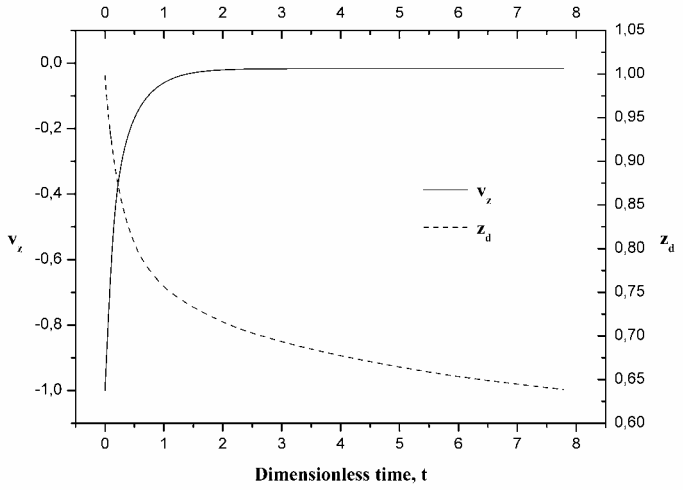


Figure 16. Evolution of the squeeze velocity and the distance of the upper disk from the mid-plane z_d with time for $Bn=10$, $\mathrm{Re}=0$, $Ca=10^3$, $w=5$, $\varepsilon=0.5$, $\alpha_{sl}=5$, $\beta_{sl}=100$, $N=300$

Indeed, one can see in the figure that the velocity of the disks reduces very rapidly, especially at early times. It is characteristic that the disk has already lost the 90% of its axial velocity already by the time $t=0.70$ while the distance of the upper disk from the mid-plane is $z_d=0.78$. After that point the disk velocity reduces with much smaller rate and it finally becomes practically zero at time $t=7.78$. The position of the disk at that time moment is $z_d=0.64$. The motion of the disks eventually stops because all the material between the two disks behaves as a rigid solid.

5. CONCLUSIONS

We studied the transient squeeze flow of a viscoplastic material between two parallel coaxial disks. Both cases of the disks moving with constant velocity and under constant force, were studied. This is the first truly transient simulation of squeeze flow for viscoplastic material and allows us to determine the shape of the liquid/air interface which was impossible with the quasi-steady state models. The simulation of the process is based on the mixed finite element method for the discretization of the governing equations coupled with a quasi-elliptic mesh generation scheme in order to follow the large deformations of the physical domain.
We examined the effect of the yield stress, the slip coefficients and the gravity forces. In the constant velocity problem is shown that unyielded material arises around the two stagnation points of flow at the disk centers, verifying previous steady state calculations. The size of these domains increases with the Bingham number but reduces with time because of the decreased distance of the disks. The force that must be applied on the surface of the disks in order to maintain their constant velocity increases substantially with Bingham number and with time. The slip condition on the surface of the disks is found to affect significantly the flow field. More specifically the level of the slip velocity near the triple contact point alters the contact angle of the moving interface with the surface of the disks, which decreases from 180^o. Moreover as the length of the slip region increases, the size of the unyielded areas reduces significantly. Under typical gravity conditions the flow inside the two disks is not affected significantly whereas when enough fluid exits them, the effect of gravity on the flow field as well as on the shape of the free surface becomes rather important.
Finally, when constant force is applied on the disks their motion is decelerating until they finally stop since at that time all the material between them behaves as rigid solid. The time that is needed for the disks to stop moving as well as their final position depends strongly on the viscoplastic properties of the material. As the Bingham number increases the time reduces while the final distance of the disks increases.

AKNOWLEDGEMENT

The Project is co-funded by the European Social Fund and National Resources of Greece under the Pythagoras II program.

REFERENCES

1. Bird R.B., Hassager O., Armstrong R.C., and Curtiss F.C., “Dynamics of Polymeric Liquids, Vol 1: Fluid Mechanics”, 2nd Ed., John Wiley & Sons, New York (1987)
2. Sherwood J.D., Meeten G.H., Farrow C.A., and Alderman N.J., “Squeeze-film rheometry of non-uniform mudcakes”, J. Non Newtonian Fluid Mech., 39 (1991) 311-334
3. Chan T.W., Baird D.G., “An evaluation of a squeeze flow rheometer for the rheological characterization of a filled polymer with a yield stress”, Rheol. Acta, 41 (2002) 245-256
4. Bird R.B., Dai G.C., Yarusso B.J., “The rheology and flow of viscoplastic materials”, Rev. Chem. Eng., 1 (1983) 1-70
5. Lipscomb G.G., Denn M.M., “Flow of Bingham fluids in complex geometries”, J. Non Newtonian Fluid Mech., 14 (1984) 337-346
6. O’Donovan E.J., Tanner R.I., “Numerical study of the Bingham squeeze film problem”, J. Non-Newt. Fluid Mech., 15 (1984) 75-83
7. Smyrnaios D. N., Tsamopoulos J. A., “Squeeze flow of Bingham plastics”, J. Non Newtonian Fluid Mech., 100 (2001) 165-190
8. Bingham E.C., “Fluidity and Plasticity”, McGraw-Hill (1922)
9. Papanastasiou T.C., “Flows of materials with yield”, J. Rheol., 36 (1992) 389-407
10. Matsoukas A., Mitsoulis E., “Geometry effects in squeeze flow of Bingham plastics”, J. Non Newtonian Fluid Mech., 109 (2003) 231-240
11. Burgos G.R. & Alexandrou A.N., “On the determination of yield surfaces in Herschel-Bulkley fluids”, J. Rheol., 43(3) (1999) 463-483
12. Kamal M. R., Goyal S. K., and Chu E., “Simulation of injection mold filling of viscoelastic polymer with fountain flow”, AICHE J., 34 (1), 94-105 (1988)
13. Dimakopoulos Y. & Tsamopoulos J., “A quasi-elliptic transformation for moving boundary problems with large anisotropic deformations”, J. Comp. Phys., 192 (2003) 494-522

COMPRESSION OF MATERIALS WITH YIELD STRESS

Andreas N. Alexandrou[1], George Florides[1], and Georgios Georgiou[2]

[1]Department of Mechanical and Manufacturing Engineering
University of Cyprus CY 1618, Nicosia, Cyprus
e-mail: andalexa@ucy.ac.cy

[2] Department of Mathematics and Statistics
University of Cyprus
P.O.BOX 20537, 1678 Nicosia, Cyprus
e-mail: georgios@ucy.ac.cy , web page: http://www.ucy.ac.cy/~georgios

Keywords: Computational rheology; Finite elements; Herschel-Bulkley fluids; Bingham fluids; Compression test; Squeeze flow.

Abstract. *Semisolid slurries used in the processing of alloys are materials with thixotropic behavior and yield stress. All existing and future semisolid technologies are based upon the unique combination of their solid-like and fluid-like behavior. However, an intrinsic difficulty in utilizing the potential of the process is the rather complex flow behavior of the slurries. As a two-phase mixture of liquid and solid particles the bulk flow behavior is determined by the way the two phases interact and evolve during processing. This suggests that process variables change continuously in a way that is very different from liquid casting.*

The objective of rheological studies of the slurries is to describe qualitatively the solid phase structure, the contribution from the liquid phase, and to express the apparent mechanical response in terms of structural parameters, such as grain size, degree of agglomeration, etc., and their kinetics. Presently, this ultimate objective is rather difficult to achieve. Instead, the alternative is to introduce a phenomenological approach that captures and generalizes the salient features of semisolid behavior. These features are: (a) the existence of a finite yield stress, (b) the apparent shear-thinning behavior in steady-state shear rate, and (c) the apparent shear-thickening behavior during rapid transients where the shear rate is variable. Broadly speaking, the mechanical response of semisolid slurries corresponds to that of a thixotropic, non-linear visco-plastic material with history-dependent material parameters.

However, the determination of material constants for the assumed behavior is difficult to achieve due to the high temperature of the suspensions and the short duration of the material response. In this paper we demonstrate the use of computational rheology as a way to determine the material constants. We show that by modeling exactly actual classical tests using computational methods we can determine the material constants that fit the assumed bulk rheological behavior. We show also that the same approach can be used to test various hypotheses about structure breakdown and the associated time scales.

1 INTRODUCTION

Semisolid slurries are a mixture of rounded, rosette-like solid particles and liquid at a temperature between the solidus and liquidus isotherms. The average solid volume fraction is a function of the bulk temperature of the suspension and depending on the local temperature it varies from zero to unity. These materials exhibit a yield stress, which depends on the solid volume fraction. During processing the viscosity changes significantly due to segregation effects of the internal microstructure. In general, semisolid materials behave like visco-plastic fluids characterized by a finite yield stress τ_0 and by material properties that are time and shear rate depended. Therefore, Herschel-Bulkley fluid model with time dependent parameters could be applied to characterize the flow and time depended phenomena of semisolid slurries [1].

The Herschel-Bulkley model is based on the combination of the Bingham and power-law models. Thus, Herschel-Bulkley fluids exhibit a yield stress, τ_o, which is the minimum stress required for the material to deform. The existence of yield stress in semisolid materials is due to bonding and dry friction between particles. Once the yield stress is exceeded the material behaves either as a shear-thickening or as shear-thinning fluid with a non-linear stress strain relationship [1]:

$$\left.\begin{array}{ll} \underline{\underline{\dot{\gamma}}} = \underline{\underline{0}} & \tau \le \tau_o \\ \underline{\underline{\tau}} = \left(\dfrac{\tau_0}{\dot{\gamma}} + K\, \dot{\gamma}^{n-1} \right) \underline{\underline{\dot{\gamma}}} & \tau \ge \tau_o \end{array}\right\} \quad (1)$$

where K is the consistency index and n is the power-law index, which determines the nonlinear behaviour of the fluid upon yielding; the fluid is shear-thinning for $n<1$ and shear-thickening for $n>1$ [2]. For $n=1$ the Bingham model is recovered, in which case K represents the constant viscosity. The three material parameters, K, τ_o and n are determined from experimental data. The Herschel-Bulkley model cannot be used in practical problems because it is singular when the shear rate approaches zero. With the exception of ideal unidirectional problems, viscoplastic behavior is usually described using non-singular approximations of Eq. (1). A popular regularization is the one proposed by Papanastasiou, which is valid uniformly at all levels of $\underline{\underline{\dot{\gamma}}}$ [2,3,4,5]:

$$\underline{\underline{\tau}} = \left\{ \frac{\tau_o \left[1 - \exp\left(-m\dot{\gamma}\right) \right]}{\dot{\gamma}} + K\, \dot{\gamma}^{n-1} \right\} \underline{\underline{\dot{\gamma}}} \qquad \text{with} \quad \dot{\gamma} = \sqrt{\frac{1}{2} II_\gamma} \quad (2)$$

In the above constitutive relation *II* stands for the second invariant of a tensor and *m* is a stress growth exponent. The above regularized constitutive equation provides a satisfactory approximation of the Herschel-Bulkley model for sufficiently large values of *m* and is valid uniformly at all levels of $\dot{\gamma}$. The need to determine the yielded and unyielded regions of the flow is thus eliminated. It should be noted that large values of *m* might lead to convergence difficulties while small values can lead to wrong results. Hence, the value of *m* should be chosen very carefully [3, 4, 6].

2 COMPRESSION EXPERIMENT

2.1 Description of the Experiment

In this study, we investigated numerically the compression of a finite amount of Herschel-Bulkley fluid by considering constant, time-independent properties. The numerical simulations were performed under constant load as well as under constant velocity applied only on the upper side of the sample. This is an important difference from other experimental studies, in which the compression is performed simultaneously from both sides of the sample. The objective here is to study the influence of the controlling parameters such as the Bingham (Bn) number on the compression experiment and other derived material quantities. The topology of the yielded ($\tau > \tau_o$) and unyielded ($\tau \leqslant \tau_o$) zones and their evolution during processing is examined here as a key derived quantity. This aspect is important in understanding the structural changes and rheological attributes of the material during flow. Alexandrou and his collaborators [3, 7] demonstrated that regularized models such as the Papanastasiou, the biviscosity and other viscoplastic models could predict the flow and represent the topography of the yield surfaces reasonably well provided that the regularizing parameters are properly selected.

Figure 1 shows the axisymmetric sample of initial radius R and height H=2R used in the simulations. Due to the non-dimensionalization the height is equal to one (H=1) and the width is D/2=R=0.5. The material is placed on a plate and compressed from the top by applying either constant force or constant velocity.
Due to symmetry, only one-half of the sample is used in the simulations. A related flow to this problem is Stefan's squeezing flow, in which both sides are set under external pressure. Symmetry boundary conditions are imposed along the axis of symmetry (r=0) and the velocity vector is zero along the bottom.

In the experiment performed under constant velocity, the specimen is compressed (from the top), and the transverse velocity (pointing downwards) is set to unity. While in the case of experiment performed under constant load, the boundary condition at the top of the sample is given by the relation:

$$F = \int_s \underline{\underline{\sigma}} \cdot e_z \, dS \quad (3)$$

where S refers to the area at the top. The velocity in this case is everywhere set to zero at t=0.

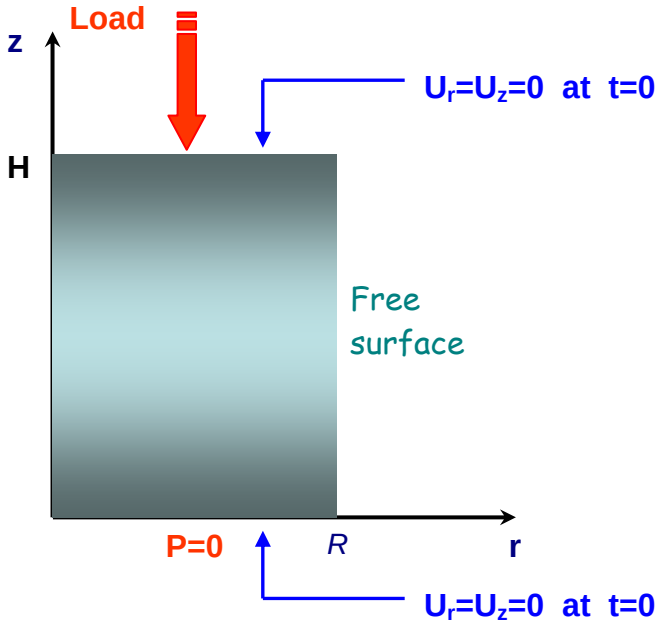


Figure 1: Geometry of the compression experiment.

2.2 Governing Equations

The flow is governed by the continuity and momentum equations for an incompressible fluid:

$$\nabla \cdot \mathbf{u} = 0 \tag{4}$$

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) = \nabla \cdot \underline{\underline{\sigma}} \tag{5}$$

where ρ is the density of the fluid, and σ is the total stress tensor defined as σ = -pI + τ, p being the pressure and I the unit tensor. The viscous stress tensor τ is assumed to obey the regularized Herschel-Bulkley equation (2) which closes the system of equations (4) and (5).

The governing equations in Lagrangian coordinates together with the constitutive relation, boundary conditions and the free surface are descretized using the mixed-Galerkin finite element method with the standard nine-node quadrilateral elements. The resulting non-linear system of equations was linearized using a Newton-Raphson iteration procedure and a solution is considered converged when the maximum error is less than 10^{-5}.

2.3 Non-dimensionalization

The non-dimensional form of the equations 4 & 5 for both experimental cases (constant load and constant velocity) becomes:

$$\nabla \cdot \mathbf{u}^* = 0 \tag{5}$$

$$\left(\frac{\partial \mathbf{u}^*}{\partial t^*} + \mathbf{u}^* \cdot \nabla \mathbf{u}^* \right) = \nabla \cdot \sigma^* \tag{6}$$

The dimensionless form of the regularized Herschel-Bulkley constitutive relation (2) takes the form:

$$\underline{\underline{\tau}}^* = \left[\mathrm{Bn} \frac{1 - \exp(-M\dot{\gamma}^*)}{\dot{\gamma}^*} + \dot{\gamma}^{*(n-1)} \right] \underline{\underline{\dot{\gamma}}}^* \tag{7}$$

where $\underline{\underline{\dot{\gamma}}}^*$ is the dimensionless rate of strain tensor, Bn is the Bingham number and M the dimensionless

stress growth exponent.

In the case of constant load there is no characteristic velocity and the Bingham number Bn and the dimensionless growth exponent M are defined in terms of the force F:

$$Bn = \frac{\tau_o H^2}{F} \tag{8}$$

$$M = m\left(\frac{F}{H^2 K}\right)^{1/n} \tag{9}$$

where H is the initial height of the sample.

In the case of constant velocity the Bingham number Bn and the dimensionless growth exponent M are defind by:

$$Bn = \frac{\tau_o H^n}{K U^n} \tag{10}$$

$$M = \frac{mU}{H} \tag{11}$$

where U is the compression velocity.

3 NUMERICAL RESULTS

For the selection of the most suitable mesh refinement we employed five meshes (10x10 to 20x20) with number of elements in the range from 100 to 400 (10x10 to 20x20) and we examined the convergence of the numerical results. By taking into consideration also computer's processor speed, we selected the mesh with 256 (16X16) elements, which gives practically converged results.

In Figure 2 is shown typical mesh shapes with 256 (16X16) elements, obtained during an experiment under constant load with Bn=0.1 and M=300. Note that the mesh appears to be more refined due to the fact that the graphics package divides each nine-node element into four quadrilaterals.

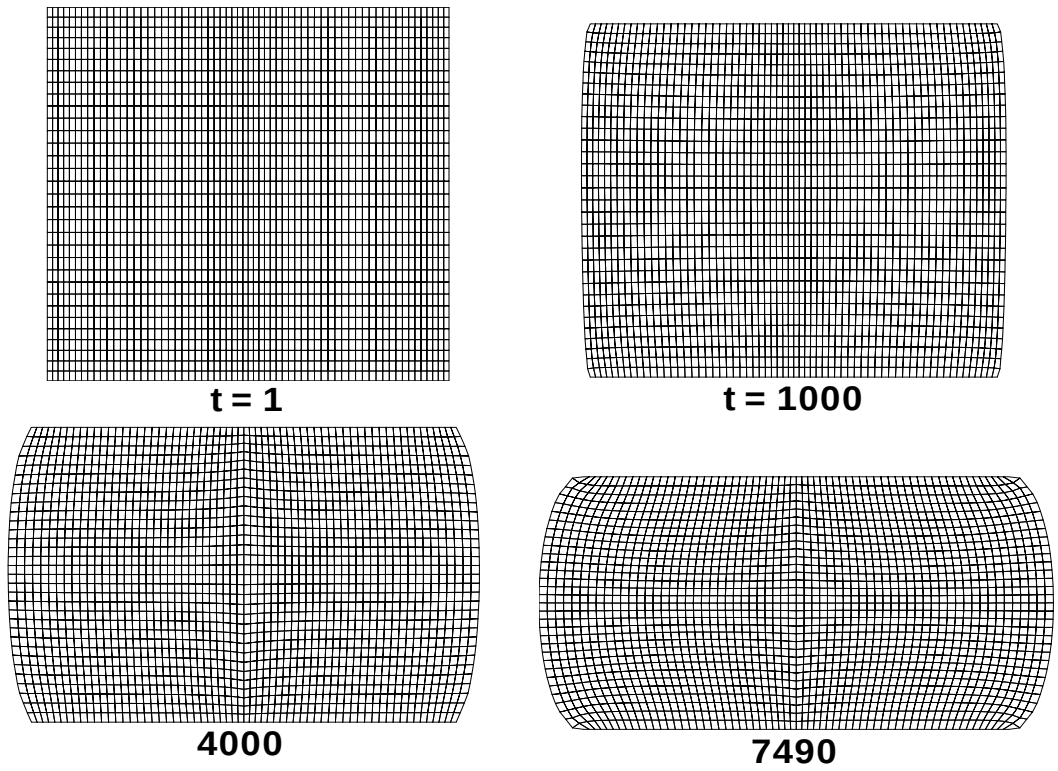


Figure 2: Typical meshes obtained during a constant force experiment; Bn=0.1 and M=300

The effects of the Bingham number, Bn, and the power-law exponent, n, on the evolution of the sample height, i.e. on the rate of compression, under constant load are illustrated in Figures 3 and 4 respectively. Figure 3 shows the evolution of the sample's height for five different Bingham numbers (Bn=0.05, 0.1, 0.3, 1 and 2), which increases by increasing the Bingham number (Bn) and reaches earlier the steady state. In experiments under constant load, compression is in general very rapid initially and then decelerates slowly approaching a "steady" state. As for the influence of the power-law exponent illustrated in Figure 4 the rate of compression, as expected, increases for higher values of *n*.

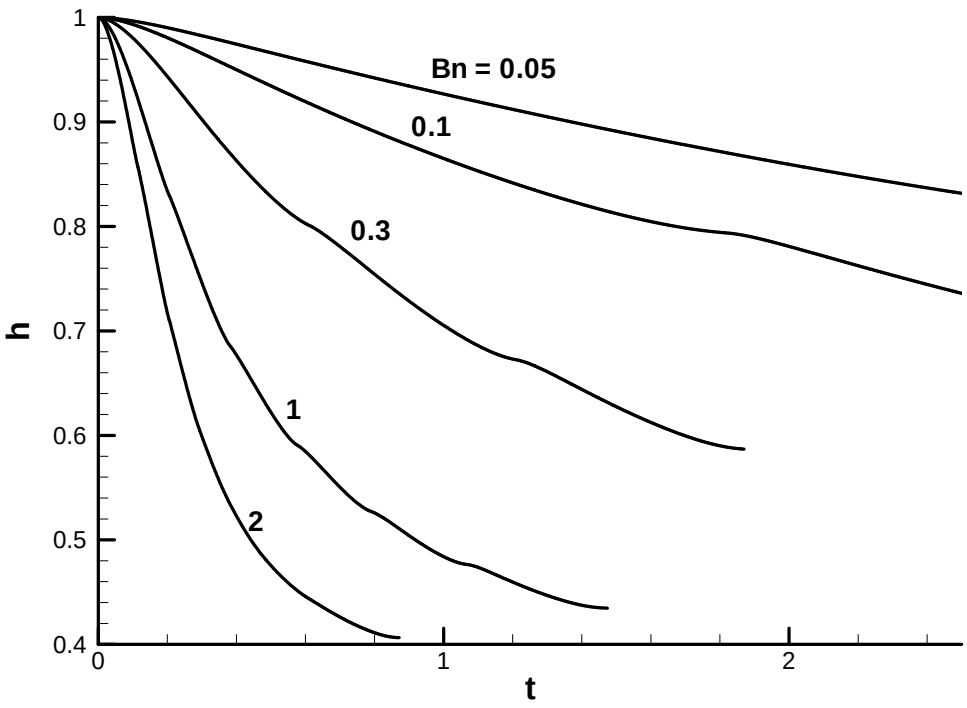


Figure 3: Effect of the Bingham number on the evolution of the sample's height;

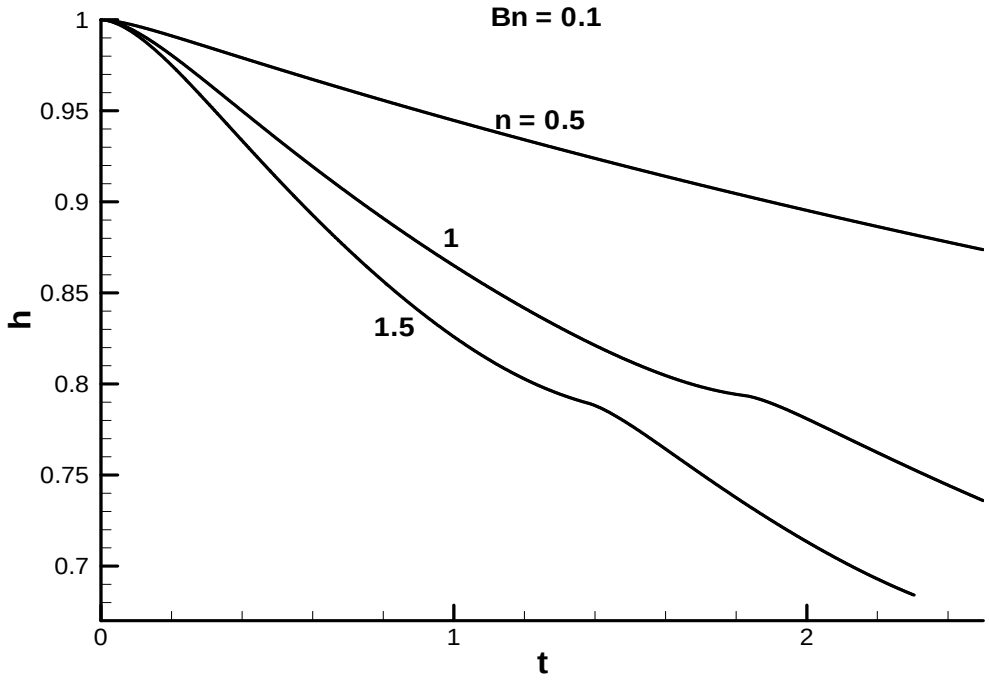


Figure 4: Effect of the power-law exponent n on the evolution of the sample's height in a constant force experiment; Bn=0.1

Figure 5 shows snapshots of the sample in an experiment under constant load and Bn = 0.1, taken at different times. Note that for presentation purposes the snapshots are not in scale. The evolution of the yielded (grey colour) and the unyielded (black colour) areas is also shown. Initially the material is yielded at the edges of the sample's upper side. As expected, the size of the yielded regions increases as the experiment proceeds. The compression of the sample begins soon after the whole material is yielded, with the exception of the unyielded spot located just above the bottom of the sample and around the axis of symmetry, which persists till the end of the compression experiment. Although the load remains constant on the top side decreases, due to the increase of the area of the compressed side, the total applied stress. Interestingly, in these final stages the unyieded area grows again instantaneously because the effective stress field in the sample decreases. This phenomenon is due to the fact that the sample is compressed with constant load, which is applied from its top side.

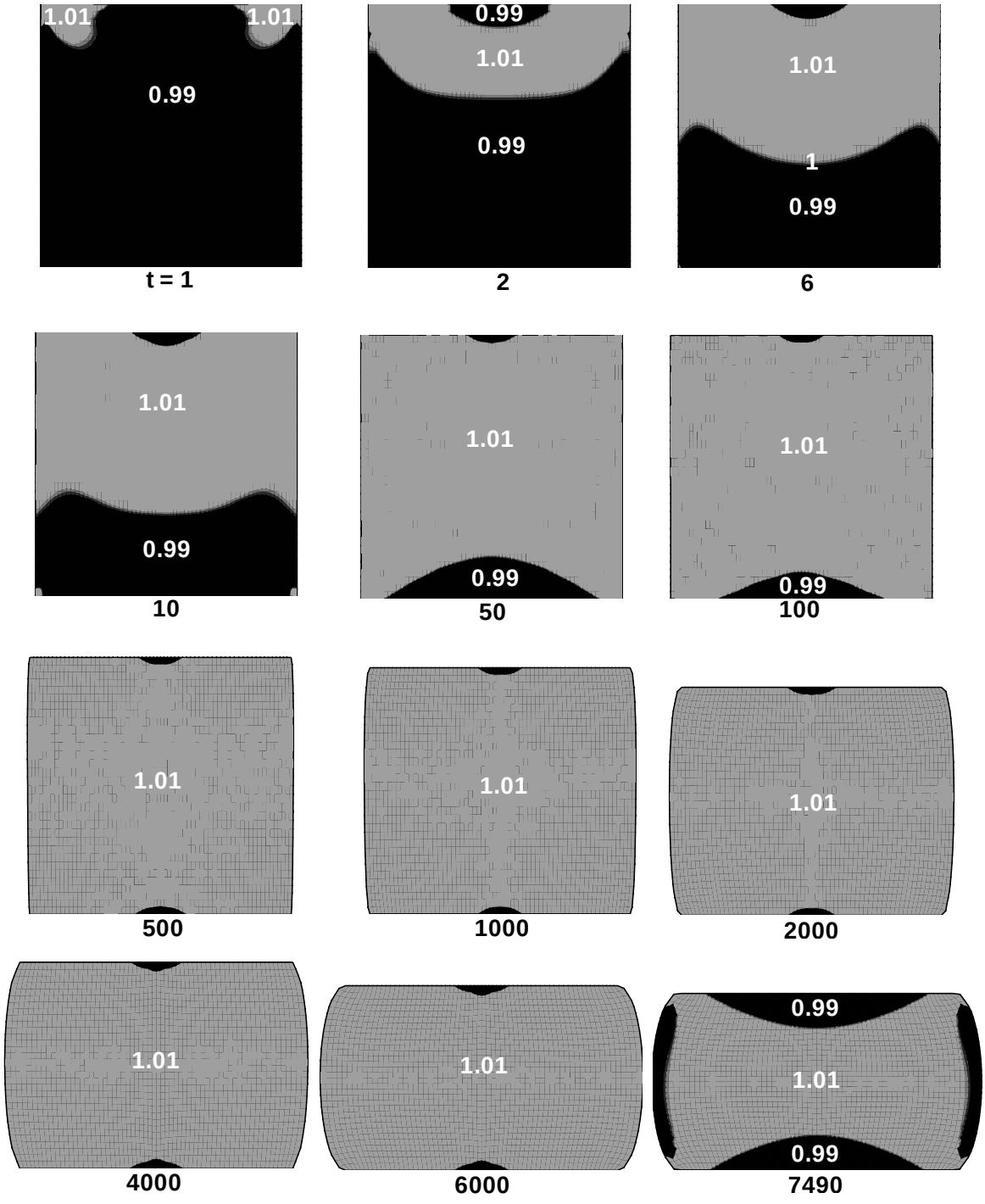


Figure 5: The evolution of the yielded (grey colour) and unyielded (black colour) areas in a constant force experiment; Bn = 0.1.

REFERENCES

[1] A.N. Alexandrou, G.C.Florides, G.C. Georgiou and D. Apelian, Rheological effects of structure breakdown on semisolid slurries, Proceeding of the 8th International Conf. on Semi-solid Processing of Alloys and Composites, Limassol (Cyprus), September 21th-23th 2004.

[2] A.N. Alexandrou, Y. Pan, D. Apelian and G. Georgiou, Semisolid material characterization using computational rheology, Proceedings of the 7th International Conference on Semi-Solid Processing of Alloys and Composites, Y.Tsutsui, M. Kiuchi and K. Ichikawa (Eds.), Tsukuba, Japan, 417-422 (2002).

[3] A.N. Alexandrou, T.M. McGilvreay and G. Burgos, Steady Herschel-Bulkley fluid flow in three-dimensional expansions, J. Non-Newtonian Fluid Mech. 100, 77-96 (2001).

[4] G.R. Burgos, Rheology of semisolid metal suspensions, PhD Thesis, Worcester Polytechnic Institute, 1999.

[5] G.R. Burgos, A.N. Alexandrou and V.M. Entov, Two-Phase Model of Flow of Semi-Solid Materials, Proceedings of 5th International Conf. On Semi-solid Processing of Alloys and Composites, Colorado, USA, 217-224 (1989).

[6] T.C. Papanastasiou, Flows of materials with yield, J. Rheology, 31 (1987) 385-404.

[7] G.R. Burgos, A.N. Alexandrou and V. Entov, On the determination of the yield surfaces in Herschel-Bulkley fluids, J. Rheology, 43 (1999) 463-483.

TWO-DIMENSIONAL SIMULATIONS OF THE EFFECT OF THE RESERVOIR REGION ON THE PRESSURE OSCILLATIONS OBSERVED IN THE STICK-SLIP INSTABILITY REGIME

Eleni Taliadorou[1], Georgios Georgiou[1,†], and Andreas Alexandrou[2,†]

[1]Department of Mathematics and Statistics
University of Cyprus, P.O.BOX 20537, 1678 Nicosia, Cyprus
e-mail: taliadorou@cytanet.com.cy, georgios@ucy.ac.cy
[2]Department of Mechanical and Manufacturing Engineering
University of Cyprus, P.O.BOX 20537, 1678 Nicosia, Cyprus
e-mail: andalexa@ucy.ac.cy

[†]Greek Association of Computational Mechanics

Keywords: Stick-slip instability, Extrusion, Compressibility, Slip, Pressure oscillations.

Abstract. *The time-dependent, two-dimensional compressible Newtonian flow over the reservoir-die region is solved assuming that slip occurs along the die wall following a non-monotonic slip law. The combination of compressibility and nonlinear slip leads to self-sustained oscillations of the pressure drop and of the mass flow rate at constant piston speed, when the latter falls into the unstable negative slope regime of the flow curve. The effect of the reservoir volume on the amplitude, the frequency and the waveform of the pressure oscillations is studied and comparisons are made with experimental observations concerning the stick-slip polymer extrusion instability.*

1 INTRODUCTION

Slip at the wall is considered to be a key factor in polymer extrusion instabilities, such as the stick-slip instability[1,2]. A characteristic of the stick-slip instability not encountered with other types of extrusion instability, such as sharkskin and gross melt fracture, is that this is accompanied by pressure and mass flow rate oscillations which result in extrudate shapes characterized by alternating rough and smooth regions[1,2]. Recent work concerning numerical modeling of the stick-slip instability has been reviewed by Achilleos et al.[3] who discuss three different instability mechanisms: (a) combination of nonlinear slip with compressibility; (b) combination of nonlinear slip with elasticity; and (c) constitutive instabilities. In the present work, we investigate further the compressibility-slip instability by means of numerical simulations.

The compressibility-slip mechanism has been tested by Georgiou and Crochet[4,5] in the Newtonian case, with the use of an arbitrary non-monotonic slip equation relating the wall shear stress to the slip velocity. These authors numerically solved the time-dependent compressible Newtonian Poiseuille and extrudate-swell flows with non-linear slip at the wall, showing that steady-state solutions in the negative-slope regime of the flow curve (i.e. the plot of the wall shear stress versus the apparent shear rate or the plot of the pressure drop versus the volumetric flow rate) are unstable, in agreement with linear stability analysis. Self-sustained oscillations of the pressure drop and of the mass flow rate at the exit are obtained, when an unstable steady-state solution is perturbed, while the volumetric flow rate at the inlet is kept constant. These oscillations are similar to those observed experimentally with the stick-slip extrusion instability. In a recent work, Georgiou[6] obtained similar results for the compressible, axisymmetric Poiseuille and extrudate-swell flows of a Carreau fluid with slip at the wall, using an empirical slip equation that is based on the experimental measurements of Hatzikiriakos and Dealy[7,8] on a HDPE melt. Unlike the experimental observations[8,9,10], however, the limit cycles of the periodic solution obtained in all these numerical studies do not follow the steady-state branches of the flow curve.

As pointed out by Georgiou[6], including the reservoir is necessary in order to account for the compression and decompression of most part of the fluid, and obtain limit cycles following the steady-state branches of the flow curve, i.e. for obtaining pressure and extrudate flow rate oscillations characterized by abrupt changes, as is the experiments. Only such abrupt changes can lead to extrudates with alternating relatively smooth and sharkskin regions, which is the basic characteristic of the stick-slip instability. Note that the reservoir region is taken into account in various one-dimensional phenomenological models, which are also based on the compressibility/slip mechanism[11]. These describe very well the pressure oscillations but they are not predictive, because they require as input certain experimental parameters.

The objective of the present work is to extend the simulations of Georgiou[6] by including the reservoir

region and study the effect of the reservoir length on the pressure oscillations. According to experiments[8,9,12], the period of the oscillations scales roughly with the volume of the polymer melt in the reservoir. Weill[9] and Durand et al.[10] also studied experimentally the effect of the reservoir length on the durations of compression and relaxation and found that both times increase linearly with the reservoir length, which indicates that the latter does not affect the waveform of the oscillations.

In Section 2, the governing equations, the slip equation and the boundary and initial conditions are discussed. In Section 3, we describe briefly the numerical method, present the numerical results, and make comparisons with experimental observations. Finally, in Section 4, we summarize the conclusions.

2 GOVERNING EQUATIONS AND BOUNDARY CONDITIONS

The geometry of the flow corresponds to the actual setup used in the experiments of Hatzikiriakos and Dealy[8]. There is a contraction region at 45 degrees between the barrel and the die as shown in Fig. 1. The actual values of the radii of the barrel and the die, denoted respectively by R_b and R, and the length of the die, L, are tabulated in Table 1.In the simulations, the length of the barrel, L_b, varied from 20R to 200R.

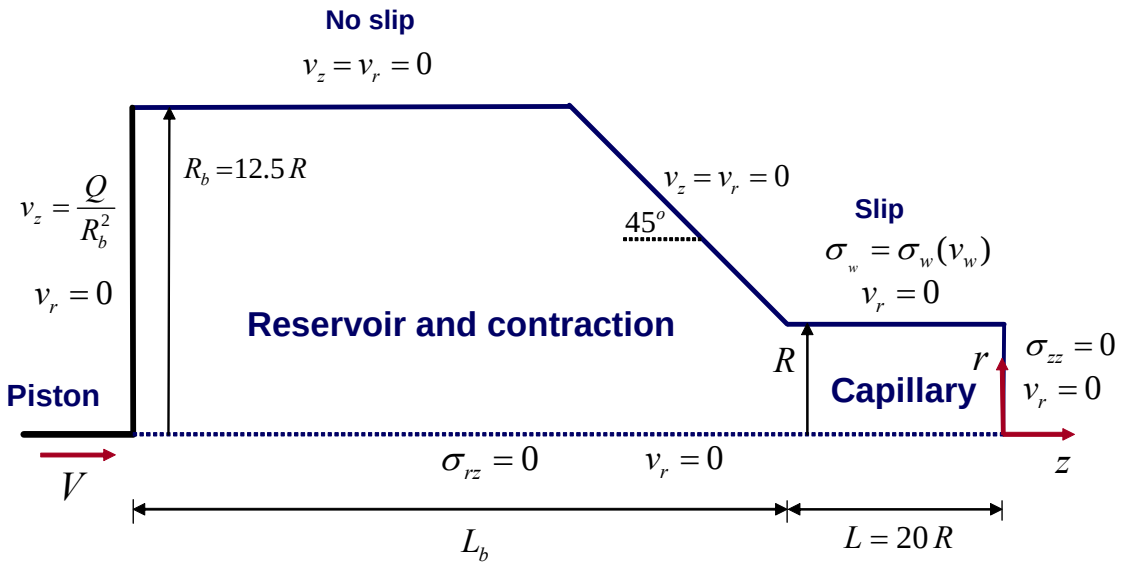


Figure 1. Geometry and boundary conditions for the time-dependent. compressible, axisymmetric flow over the reservoir-capillary region, with slip along the capillary wall

Symbol	Parameter	Value
R_b	Radius of the barrel	0.9525 cm
L_b	Length of the barrel	
	Contraction angle	45 degrees
R	Radius of the die	0.0381 cm
L	Length of the die	0.762 cm

Table 1 : Symbols and values of various lengths concerning the flow geometry

To non-dimensionalize the governing equations, we scale the lengths by the capillary radius, R, the velocity vector, **v**, by the mean velocity V in the capillary, the pressure, p, by $\eta V/R$, η denoting the constant viscosity, the density, ρ, by a reference density, ρ_0, and the time by R/V. With these scalings, the dimensionless continuity and momentum equations for time-dependent, compressible, isothermal viscous flow in the absence of body forces become:

$$\frac{\partial \rho}{\partial t} + \nabla \bullet \rho \mathbf{v} = 0 \qquad (1)$$

and

$$Re\,\rho\left(\frac{\partial \mathbf{v}}{\partial t}+\mathbf{v}\bullet\nabla\mathbf{v}\right)=-\nabla p+\nabla^2\mathbf{v} \tag{2}$$

where *Re* is the Reynolds number, defined by

$$Re\equiv\frac{\rho_0 RV}{\eta} \tag{3}$$

The above equations are completed by an equation of state relating the pressure to the density. We used the first-order expansion:

$$\rho=1+B\,p \tag{4}$$

where *B* is the compressibility number,

$$B\equiv\frac{\beta\eta V}{R} \tag{5}$$

β being the isothermal compressibility.

Along the capillary wall, slip is assumed to occur following the three-branch multi-valued slip model:

$$v_w=\begin{cases}A_1\sigma_w^{m_1}, & 0\le v_w\le v_{c2}\\ A_2\sigma_w^{m_2}, & v_{c2}\le v_w\le v_{\min}\\ A_3\sigma_w^{m_3}, & v_w\ge v_{\min}\end{cases} \tag{6}$$

where v_w is the relative dimensionless velocity of the fluid with respect to the wall, σ_w is the dimensionless shear stress on the wall, v_{c2} is the maximum slip velocity at σ_{c2}, and v_{min} is the minimum slip velocity at σ_{min}. The third branch is the power-law slip equation suggested by Hatzikiriakos and Dealy[8] for the right branch of their flow curve. The first branch results from the slip equation they propose for the left branch of their slope curve after substituting all parameters for resin A at 180°C and taking the normal stress as infinite. Finally, the second negative-slope branch, which corresponds to the unstable region of the flow curve for which no measurements have been possible, is just the line connecting the other two branches. The values of all the slip equation parameters and the definitions of the dimensionless numbers A_i can be found in Ref. 6.

The other boundary conditions of the flow are shown in Fig. 1. Along the axis of symmetry, we have the usual symmetry conditions. Along the barrel and the contraction walls both velocity components are zero (no slip). Along the capillary wall, only the radial velocity is zero, whereas the axial velocity satisfies the slip equation (6). At the inlet plane, it is assumed that the radial velocity component is zero while the axial velocity is uniform, corresponding to the motion of the piston at constant speed. Note that the imposed volumetric flow rate, *Q*, is scaled by $\pi R^2 V$. The simulations are carried out on a fixed domain, i.e. the motion of the piston is not taken into account. This is a reasonable assumption provided that the piston speed is low. At the capillary exit, the radial velocity component and the total normal stress are assumed to be zero.

Finally, as initial condition, we use the steady-state solution corresponding to a given volumetric flow rate Q_{old} that we perturb to Q at t=0.

3 NUMERICAL RESULTS

We use the finite element formulation for solving this Newtonian flow problem, employing biquadratic-velocity and bilinear-pressure elements. For the spatial discretization of the problem, we use the Galerkin forms of the continuity and momentum equations. For the time discretization, the standard fully-implicit (Euler backward-difference) scheme is used. Various finite element meshes have been used in the simulations with the reservoir length, L_b, ranging from 20 to 200. These were refined near the walls, and around the entrance of the capillary. The longest mesh (L_b=200) consisted of 4511 elements corresponding to 42403 unknowns. In all results presented below the following values for the slip equation parameters and the compressibility number

have been used: m_1=3.23, A_1=0.0583, m_2=2.86, A_2=0.929, m_3=-4.43, A_3=4.04, and B=1.54 10^{-4} .

We first constructed the steady-state flow curves for the reservoir-capillary region. In Fig. 2, we show the log-log plot of the pressure drop, measured along the centerline from the piston to the die exit, versus the volumetric flow rate obtained with Re=0.01 and L_b=80. Due the non-monotonicity of the slip equation, the flow curve exhibits a maximum and a minimum, which define the limits of the unstable regime, i.e. only the steady-state solutions corresponding to the two positive-slope branches are stable. As already mentioned, the steady-state solutions are perturbed by changing the volumetric flow rate from an old value to the desired one Q. Given that the flow is compressible, the behavior of the time-dependent solution depends on whether the new value of Q corresponds to a positive-slope branch, or to the negative-slope branch which is unstable. In the first case, the new steady-state is obtained without any oscillations, whereas, in the second case, the solution is oscillatory and, after a transition period, becomes periodic. Self-sustained oscillations of the pressure drop and the mass flow rate are obtained which are similar to those observed experimentally in the stick-slip extrusion instability regime. All the results presented below have been obtained in the unstable regime.

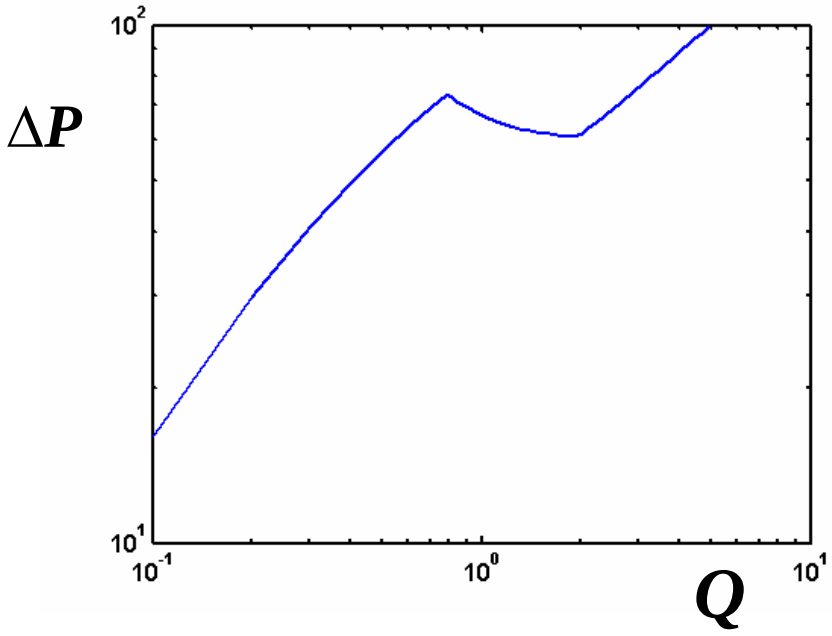


Figure 2. Flow curve for Re=0.01 and L_b=80

In Fig. 3, we show the oscillations of the pressure drop (Fig. 3a) and the volumetric flow rate (Fig. 3b) obtained by perturbing the steady-state solution for Re=0.01, L_b=80 and Q=1.35. In Fig. 3a, we show two different possibilities when the pressure drop is measured across the entire flow domain, $(\Delta P)_{tot}$, and across the capillary, $(\Delta P)_{cap}$. Sudden jumps of the pressure drop are observed in the latter case. The volumetric flow rate at the capillary exit is also characterized by sudden jumps which is consistent with experimental observations. Plotting the trajectory of the solution on the flow curve plane (Fig. 4) shows that, after a transition period, a limit cycle is reached which follows exactly the positive-slope branches of the steady-state flow curve. The volumetric flow rate increases together with the pressure following exactly the left positive-slope branch of the flow curve and, when the pressure reaches its maximum value, Q jumps to the right positive slope branch. The volumetric flow rate then starts decreasing together with the pressure following this branch till the pressure reaches its minimum and then jumps to the left positive-slope branch and starts the next oscillation cycle. This behavior agrees well with experimental observations[8,10]. Note also that in our previous study[6], the limit cycles did not follow the steady-state flow curve due to the omission of the reservoir region. This drawback was also exhibited by the one-dimensional model of Greenberg and Demay[13] , which does not include the barrel region. Note that one-dimensional phenomenological relaxation/oscillation describe the oscillations of the pressure and the volumetric flow rate in the stick-slip instability regime under the assumption that these follow the experimental flow curve[14,15]. The present simulations are the first to show that the limit cycle follows the steady-state flow curve.

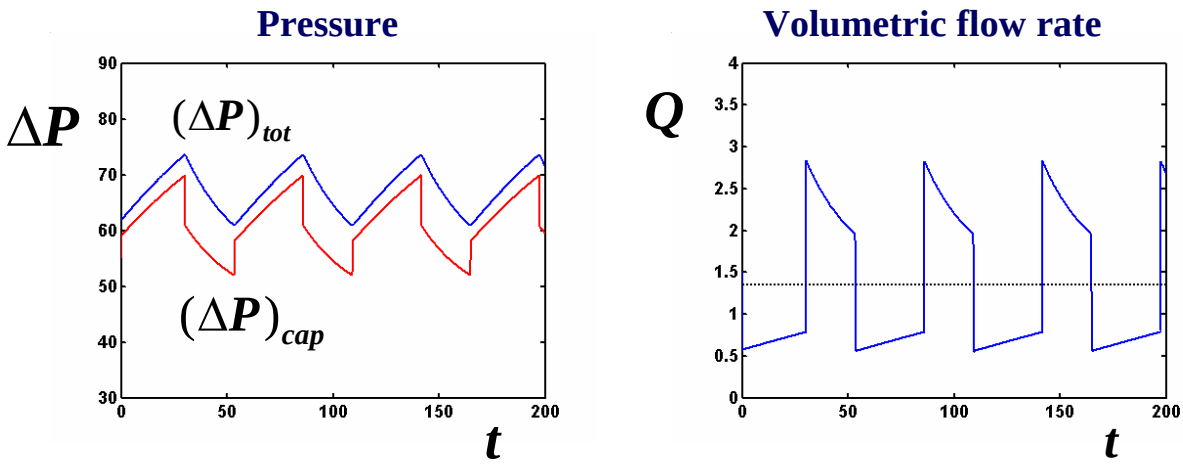


Figure 3. Pressure and flow rate oscillations for Q=1.35, Re=0.01 and L_b=80

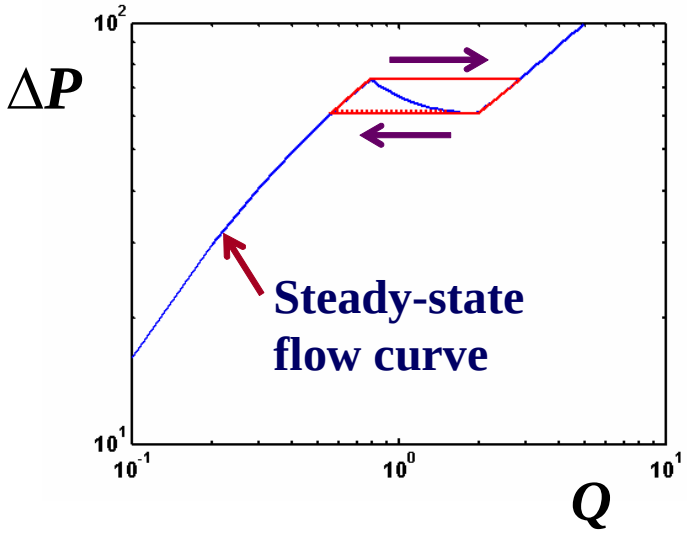


Figure 4. Trajectory of the solution on the flow curve plane; Q=1.35, Re=0.01 and L_b=80

We then reduced the value of the Re from 0.01 to 0.001 in an attempt to approach the experimental value ($1.43\ 10^{-5}$). As shown in Fig. 5, where we compare the oscillations of ΔP during one cycle for Re=0.01 and 0.001, L_b=80 and Q=1.35, decreasing the Reynolds number has no practical effect on the oscillations. However, the artificial overshoots are observed in the flow rate. Thus instead of trying to eliminate the overshoots by reducing the time step (which would have resulted into much longer runs), we decided to continue the runs with Re=0.01. Note that in our previous study[6] for the extrudate-swell flow, in which the reservoir region has been excluded, we observed that as the Reynolds number is reduced the amplitude of the pressure-drop oscillations is reduced, the amplitude of the mass-flow-rate oscillations is increased and the frequency of the oscillations is considerably increased. This shows once again the importance of including the reservoir region.

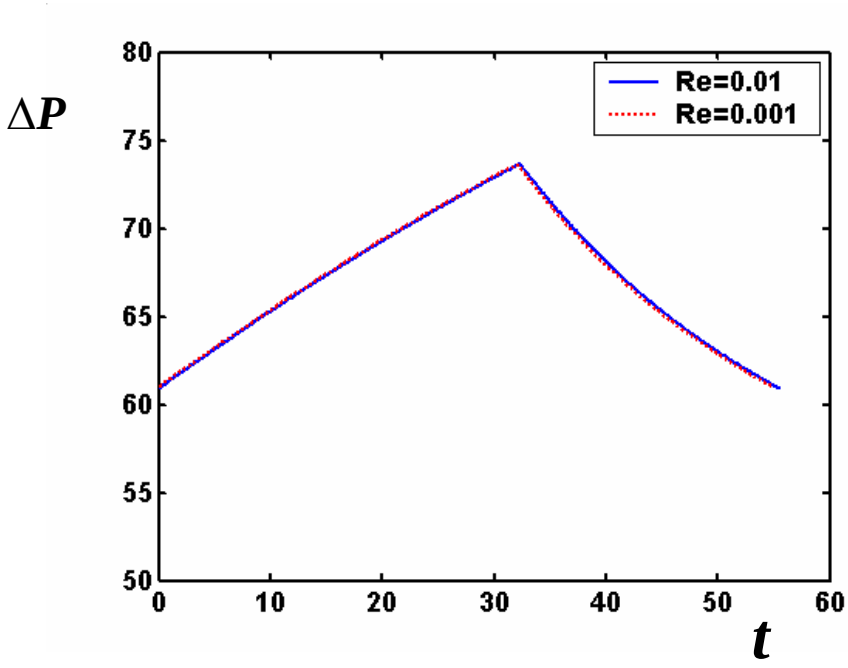


Figure 5. Comparison of the pressure oscillations for Re=0.01 and 0.001; L_b=80 and Q=1.35

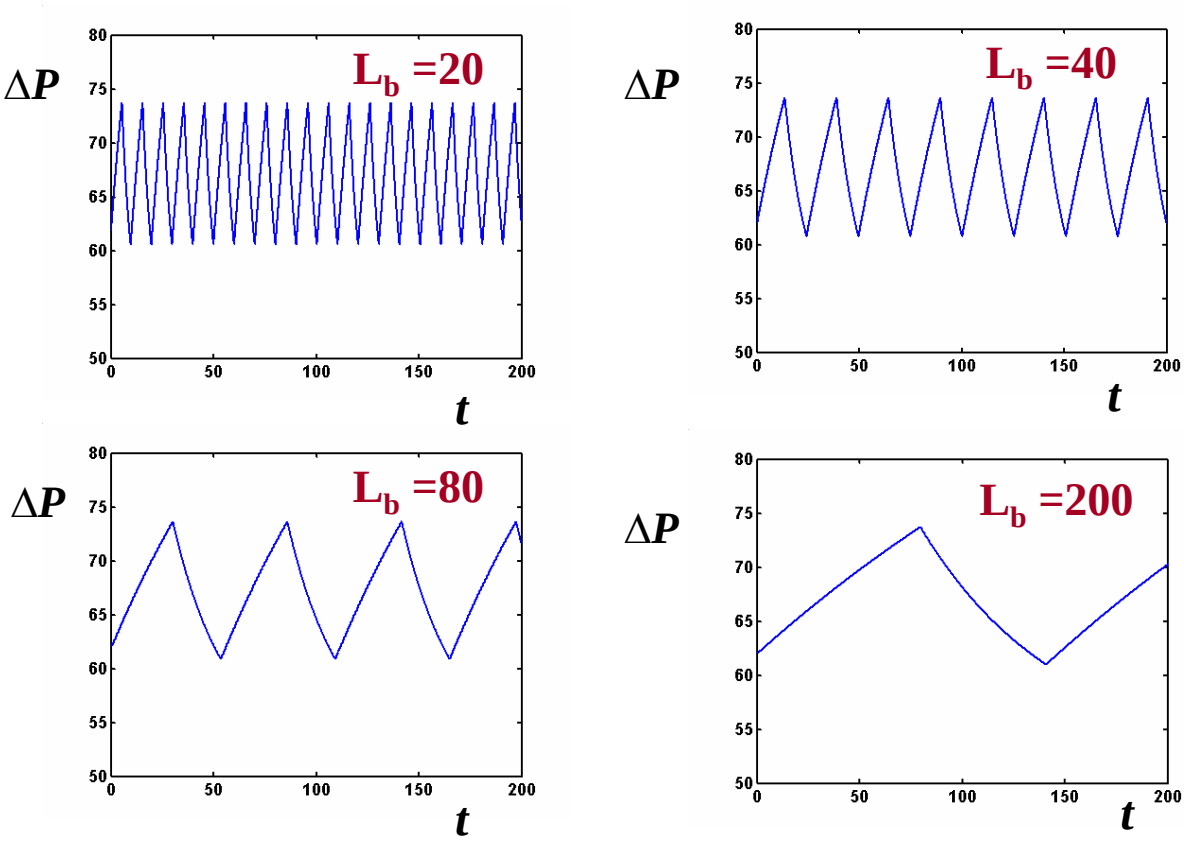


Figure 6. Effect of the reservoir length on the pressure oscillations; Q=1.35 and Re=0.01

In order to study the effect of the reservoir length on the pressure oscillations we obtained results for various values of L_b. In Fig. 6, we show the pressure oscillations for different values of L_b, Re=0.01 and Q=1.35. We observe that the period of the pressure oscillations increases with L_b while their amplitude seems to be less sensitive. This is more clearly shown in Fig. 7, where the period and the amplitude of the pressure oscillations are plotted versus the reservoir volume. In agreement with experiments with different polymer melts[8,10,16,17] , the period T increases linearly with the reservoir volume while the amplitude is essentially constant. In Fig. 7a, the period appears to pass through the origin which is not the case with the experiments. Finally, in order to show the effect of the reservoir on the waveform of the pressure oscillations we plotted the normalized pressure oscillations during one cycle for L_b =20 and 200 (Fig. 8). The waveform is independent of the reservoir length, i.e., the durations of the compression and relaxation increase linearly with the reservoir length. This agrees well with the experiments of Weill[9], Hatzikiriakos and Dealy[8] and Durand et al.[10].

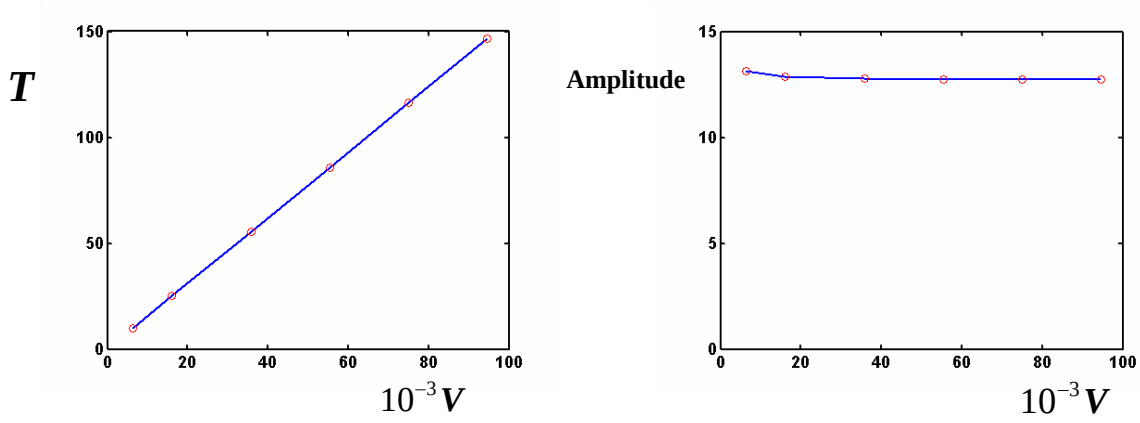


Figure 7. The period and the amplitude of the pressure oscillations versus the reservoir volume; Q=1.35 and Re=0.01

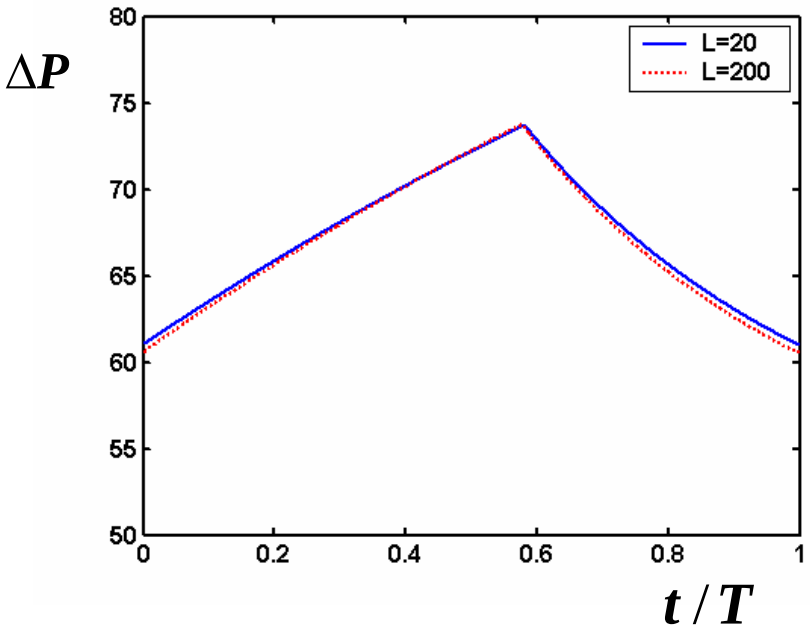


Figure 8. Effect of the reservoir length on the waveform of the pressure oscillations; Q=1.35 and Re=0.01

4 CONCLUSIONS

We solved numerically the time-dependent, compressible flow of a Newtonian fluid over the reservoir-capillary region, assuming that slip occurs along the capillary wall following a non-monotonic slip law based on the experimental findings of Hatzikiriakos and Dealy[7,8] for certain polyethylene melts. By using meshes of different length, we have studied the effect of the reservoir length on the pressure oscillations occurring when the imposed flow rate falls in the unstable negative-slope regime of the flow curve. Our calculations showed that the pressure oscillations follow the steady-state flow curve and that their period increases linearly with the reservoir length, while their amplitude and waveform remain unaffected. These results are in good agreement with the experiments of Weill[9] , Hatzikiriakos and Dealy[8], Durand et al.[10], and others, which have also shown that the period and the shape of the pressure oscillations vary also with the imposed flow rate, where their amplitude remains unaffected. The effect of Q on the pressure oscillations is currently under study.

REFERENCES

[1] Denn, M.M. (2001), "Extrusion instabilities and wall slip," Annu. Rev. Fluid Mech. 33, pp. 265-287.
[2] Hatzikiriakos, S.G. and Migler, K. (2004), *Polymer Processing Instabilities: Control and Understanding*. New York: Mercel Dekker.
[3] Achilleos, E., Georgiou, G. and Hatzikiriakos, S.G. (2002), "On numerical simulations of polymer extrusion instabilities," Applied Rheology 12, pp. 88-104.
[4] Georgiou, G.C. and Crochet, M.J. (1994), "Compressible viscous flow in slits with slip at the wall," J. Rheology 38, pp. 639-654.
[5] Georgiou, G.C. and Crochet, M.J. (1994), "Time-dependent compressible extrudate-swell problem with slip at the wall," J. Rheology 38, pp. 1745-1755.
[6] Georgiou, G. (2003), "The time-dependent compressible Poiseuille and extrudate-swell flows of a Carreau fluid with slip at the wall," J. Non-Newtonian Fluid Mech. 109, pp. 93-114.
[7] Hatzikiriakos, S.G. and Dealy, J.M. (1992), "Wall slip of molten high density polyethylenes. II. Capillary rheometer studies," J. Rheology 36, pp. 703-741.
[8] Hatzikiriakos, S.G. and Dealy, J.M. (1992), "Role of slip and fracture in the oscillating flow of HDPE in a capillary," J. Rheology 36, pp. 845-884.
[9] Weill, A. (1980), Capillary flow of linear polyethylene melt: Sudden increase of flow rate. J. Non-Newtonian Fluid Mech., 7, pp. 303-314.
[10] Durand, V., Vergnes, B., Agassant, J.F., Benoit, E. and Koopmans, R.J. (1996), "Experimental study and modeling of oscillating flow of high density polyethylenes," J. Rheology, 40, pp. 383-394.
[11] Georgiou, G. (2004), "Stick-slip instability," Chapter 7 in *Polymer Processing Instabilities: Control and Understanding*, Hatzikiriakos, S.G. and Migler, K. (Eds.), New York: Mercel Dekker.
[12] Myerholtz, R.W. (1967), "Oscillating flow behavior of high-density polyethylene melts," J. Appl. Polym. Sci. 11, pp. 687-698.
[13] Greenberg, J.M. and Demay, Y. (1994), "A simple model of the melt fracture instability," Eur. J. Appl. Maths., 5, pp. 337-357.
[14] Adewale, K.P. and Leonov, A.I. (1997), "Modelling spurt and stress oscillations in flows of molten polymers," Rheol. Acta 36, pp. 110-127.
[15] Den Doelder, C.F.J., Koopmans, R.J., Molenaar, J. and Van de Ven, A.A.F. (1998), "Comparing the wall slip and the constitutive approach for modelling spurt instabilities in polymer melt flows," J. Non-Newtonian Fluid Mech., 75, pp. 25-41.
[16] Robert, L., Vergnes, B. and Demay, Y. (2001), "Complex transients in the capillary flow of linear polyethylene," J. Rheology 44, pp. 1183-1187.
[17] Sato, K. and Toda, A. (2001), "Physical mechanism of stick-slip behavior in polymer melt extrusion: Temperature dependence of flow curve," J. Phys. Soc. Japan 70, pp. 3268-3273.

EDGE EFFECTS ON FLOW AND STRESS IN DRYING COATINGS

H. Radhakrishnan and L. E. Scriven

Department of Chemical Engineering & Materials Science
and Coating Process Fundamentals Program
Industrial Partnership for Research in Interfacial and Materials Engineering
University of Minnesota
Minneapolis, MN 55455 USA
Email: hari@cems.umn.edu, pjensen@cems.umn.edu

Keywords: Coatings, Drying, Elastic Stresses, Viscous Flow, Finite Element Method

Abstract: *Applied liquid coatings are often dried or cured to their final form. Under uniform conditions, drying can be taken to be one-dimensional, i.e. top-down, in most coatings except at the edges, where it is both top-down and edge-in. However, uniform drying conditions are difficult to achieve, and in-plane gradients in concentration, shrinkage, and stress develop. Drying at the edge is the archetype of such non-uniform drying; modeling it can help in assessing the effects of in-plane gradients. A two-dimensional theory of drying is formulated and used to model non-uniformly drying coatings. The equation system is solved by a method of lines. The solutions illustrate effects of non-uniformities on concentration distribution, and consequently on shrinkage and stress gradients. The resulting stresses can produce curling, cupping, cockling, cracking, delaminating from the substrate, and other defects in the final coating.*

1 INTRODUCTION

Coatings are generally applied as liquid, then dried or cured to their final solid form. Whether by gelation or vitrification, the transition from liquid to solid can be approximated as occurring at a specific concentration[1,2]. Solvent evaporation during drying reduces the stress-free state of the coating causing its current state to shrink. If the coating is liquid, its current state shrinks isotropically and with no slip at the substrate. If differential shrinkage produces deviatoric strain, deviatoric viscous stress appears in proportion to the local strain rate; the proportionality is the viscosity. If the coating is an elastic solid, in-plane shrinkage of its current state is frustrated by adhesion to the substrate. If differential shrinkage produces elastic strain, elastic stresses appear in proportion to the local strain; the proportionality is the elastic modulus. Only solvent loss and shrinkage after solidification produces stress in the drying coating[1,2].

Far from the edge, drying is top-down and, under uniform conditions, gradients in concentration, stress and shrinkage are one-dimensional: normal to the substrate. However, non-uniformities in mass transfer, coating thickness, and substrate profile can lead to in-plane gradients in concentration, shrinkage and stress; so can other factors. Near the edge, drying is both top-down and edge-in; and even under uniform conditions, gradients in concentration, stress and shrinkage are two- or three-dimensional: normal and parallel to the substrate. One-dimensional models of drying and stress development are inadequate for understanding these gradients and their consequences.

2 THEORY AND COMPUTATION

Theory of two-dimensional drying is brought to bear on coatings applied as liquid that solidify at a moving front to their final dried solid form. The governing equations are those describing solvent mass transfer by diffusion and convection; viscous flow in the liquid coating; appearance and subsequent migration of the solidification front; shrinkage and stress development in the solid coating; and the effect of falling solvent concentration on the coating's diffusivity, viscosity, elastic modulus, yield stress and post-yield viscosity.

The governing equations are highly non-linear, and difficult if not impossible to solve in terms of standard domain-spanning functions. So the equations are solved by a "method of lines:" Galerkin's method with finite element basis functions in space, and finite-differencing of the time-dependent basis function coefficients. DASSL package[3] is used to solve the equations by a Newton's method with secant-approximated Jacobian[4]; and Hood's frontal solver[5] is used to solve the linear matrix problem at each iteration.

The model system chosen resembles closely a polystyrene-toluene solution. The coating behaves like a Newtonian liquid before solidification, a neo-Hookean elastic solid after. Wherever the local stress in the solid coating exceeds the yield stress, the coating relaxes stress to the yield value by plastic yielding of the stress-free state. Von Mises yield criterion[6] and its associated flow rule are used to model stress relaxation by yielding.

Available experimental data on the concentration dependence of diffusivity, viscosity, and elastic modulus[7,8,9,10] were fitted empirically and the curves were extended to regions where data were not available. There appears to be neither experimental data nor theoretical framework about the concentration dependence of yield stress and post-yield viscosity. So the yield stress was taken to be a constant fraction of the elastic modulus, and post-yield viscosity was assumed to vary with composition as does the Newtonian viscosity.

3 RESULTS AND DISCUSSION

3.1 Drying near the edge after solidification

Solidified drying coatings that adhere to the substrate cannot shrink freely in the in-plane direction i.e., parallel to the substrate. The difference between the current state and the stress-free state of the coating is elastic strain, to which stress is proportional; the proportionality factor is elastic modulus. At the edge, drying is both top-down and edge-in, and inherently two-dimensional. Non-uniform solvent removal there causes non-uniform shrinkage, and produces in-plane gradients in stress, as shown in Figure 1. Stress varies close to the edge but only imperceptibly more than four to six thicknesses away. Far away from the edge, the cross- and down-web tensile stresses fall to their transversely isotropic value, as seen in Figures 1a and 1b; the normal peeling stress and shear stress vanish, as seen in Figures 1c and 1d. Near the edge, the cross- and down-web tensile stresses vanish at the free corner; and the peeling and shear stresses are theoretically "infinite" — an elastic singularity — at the pinned corner. In-plane stress gradients can produce defects such as cockle. The high tensile stress concentration at the pinned edge can lead to delamination from the substrate. Even uniform in-plane tensile stress away from the edge, when excessive, can produce defects like curling and cupping, and failures like cracking and crazing. If the coating yields, i.e. relaxes stresses plastically, the local stress in the coating falls to the yield value everywhere that value has been exceeded. Thus, plastic yielding reduces the level of stress overall and variation of the concentrated stress near edges, and therefore the danger of defects and failures. Predictions from the model show that high elastic modulus and high yield stress raise the level of stress and in-plane stress gradients; high post-yield viscosity prevents the stresses from relaxing rapidly, producing a stress peak.

The high elastic modulus and yield stress of hard coatings make them susceptible to cracking and delamination. A method sometimes advocated to lower the overall stress is to apply a thin sub-layer of softer material between the hard layer and the substrate. The idea is that the softer sub-layer would allow the upper layer to retract more from the edge without significantly affecting the coating's functionality. Retraction of the upper layer's edge would allow its current state to be closer to its stress-free state, thereby lowering its strain and stress.

Figure 2 shows the effect of sub-layer thickness on the upper layer's edge retraction and stress concentration. The upper layer's elastic modulus, yield strength, and post-yield viscosity are five times greater than those of the sub-layer. The upper layer's initial thickness is 50 microns in all cases; whereas the sub-layer's thickness was varied from 0 to 50 microns. With rising sub-layer thickness, the edge of the upper layer retracts more and its stress falls. However, edge retraction is limited and soon asymptotes. The retraction's effect is felt only near the edge, and falls quickly about four to six thicknesses away. Figure 3 shows the effect of the sub-layer's thickness on stress concentration in the sub-layer. Because edge retraction lowers stress in the top-layer, it exerts less traction at the interface between the two layers producing less stress in the sub-layer. The stress concentration at the pinned edge is significantly lower than that in the single layer coating. Predictions from the model indicate that the amount of edge retraction depends on the ratio of the moduli of the two layers and the ratio of their thicknesses. Weaker, thicker sub-layers allow the upper layer's edge to retract more but only up to a limit. Therefore the method advocated cannot lower stress overall, yet it can at an edge, and thereby reduce the danger of delaminating from the substrate.

3.2 Drying near the edge before solidification

Whereas the concentration of stress in a solidified coating reaches no more than six thicknesses from the edge (as noted earlier by Tam[11] and Lei[12]), instances are numerous[13] where the edge effects have intruded much farther. Such defects can develop or begin developing when the coating is still liquid.

To examine this aspect, a flat liquid coating with an initially rounded edge, as shown in Figure 4a was modeled. If drying accompanies flow, the evolution of edge shape can be split into two stages: in the first, drying is insignificant and the volume of liquid is constant; in the second, drying becomes appreciable and the liquid volume shrinks continuously.

In the first stage, flow of the liquid coating is driven by capillarity, i.e. the gradient in curvature of the free surface. Liquid is driven away from the curved edge toward the middle, as shown in Figures 4b to 4e, until ultimately the coating reaches its static equilibrium shape. The profile is an arc of a circle if the effect of gravity

is negligible, as shown in Figure 4f.

In the second stage, the loss in solvent volume causes the stress-free state of the liquid to shrink. The difference between the current state and stress-free state is deviatoric strain to which deviatoric viscous stress is proportional; the proportionality is viscosity. If the viscous stress exerts a net local force (i.e. non-vanishing divergence of the stress tensor — or stress dyadic), the liquid coating flows. The capillary pressure force is much larger than the viscous force from the rate of shrinkage. The liquid profile remains an arc of a circle. Because solvent evaporates, the arc's radius changes with time, as shown in Figures 4g to 4i.

Only in narrow stripe coatings can the liquid attain the arcuate shape. Otherwise, the coating solidifies before the capillary flow from the edge reaches the middle of the coating. The shape of the dried coating depends on the liquid shape at the time of solidification. Initial solvent concentration and drying conditions determine the time to solidify; resistance to flow determines the liquid edge shape. Flow resistance depends on the coating's viscosity, surface tension, and thickness. Solutions of the governing equations illustrate that higher viscosity, lower surface tension, and thinner coatings increase the flow resistance and in that way prevent the edge effect from extending far inward.

3.3 Drying near the edge during solidification

Edge effects that develop when the coating is still liquid would continue to move inward unless they were arrested by solidification. Solidification starts where the solvent concentration is the lowest, usually at the edge where the coating is the thinnest, and propagates to the interior of the coating. To study the effect of solidification on edge effects, a drying coating that solidifies at a moving front was modeled. Solidification was taken to occur at a specified concentration, and the solidification front was tracked explicitly with separate material ("Lagrangian") meshes for the liquid and for the solid. The boundary between the meshes was the solidification front.

Initially, the coating was taken to be fully liquid, and drying, shrinkage and flow are modeled as described above. Onset of solidification was declared when the solvent concentration in a small number of finite elements, about five or six, dropped below the specified concentration. The solver was halted; the coating was remeshed into solid and liquid parts; the old solution was transferred onto the new mesh; and the solver was restarted. Figure 5a shows the liquid and solid regions near the edge at the onset of solidification. The meshes deformed so that the interface always corresponded to the isopleth of the solidification concentration, as shown in Figures 5b to 5h. The two meshes were remeshed when necessary to maintain numerical accuracy. End of solidification was declared when the maximum solvent concentration in the coating was no more than 0.1% of the solidification concentration. The coating continued to dry as a solid until no more solvent remains in the coating.

The in-plane stress gradients in a fully dried coating are shown in Figure 6. Predictions from the model show that solidification alleviates some of the edge effects that developed when the coating was still liquid. As in the fully solid case, stress varies close to the edge but falls off quickly away from the edge. Perceptible edge effects extend up to ten thicknesses from the edge, compared to six thicknesses in the fully solid case and twenty thicknesses in the fully liquid case.

4 SUMMARY AND CONCLUSIONS

Elastic stresses develop in a solidified drying coating in response to its frustrated in-plane shrinkage. The effects of the coating's elastic modulus, yield strength, and post-yield viscosity on stress development and distribution can be analyzed and predicted, as is done here. Edge retraction of the upper layer in a drying two-layer coating can reduce stress near the edge of a coating and so reduce the danger of delaminating from the substrate but cannot significantly affect stress in the rest of the coating.

Edge effects that arise from flow in the liquid phase intrude farther from the edge than those from post-solidification shrinkage and elastic stress. The final edge shape of dried coating depends on the edge shape of the liquid coating at the start of solidification, which in turn depends on initial shape, surface tension, and the liquid coating's resistance to flow. The thinner and more viscous a coating, and the lower its surface tension, the narrower the edge region where thickness varies in an otherwise uniform coating.

Edge effects that arise in the liquid phase tend to be alleviated by flow until a coating solidifies. When solidification begins at the edge, flow in the liquid coating away from the edge relative to the solidified coating reduces the thickness variation making the coating more uniform. Elastic stresses can appear only after solidification, and they develop in the same way as stresses developed in drying solid coatings.

REFERENCES

[1] S. G. Croll. "Internal stress in a solvent-cast thermoplastic coating". Journal of Coatings Technology, 50(636), 1978.

[2] S. G. Croll. “The origin of residual internal stress in solvent-cast thermoplastic coatings”. Journal of Applied Polymer Science, 23:847–858, 1979.

[3] K. E. Brenan, S. L. Campbell, and L. R. Petzold. *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*. North-Holland, 1989.

[4] D. E. Salane. “Adaptive routines for forming jacobians numerically.” Technical Report SAND86-1319, Sandia National Laboratories, Albuquerque, NM, 1986.

[5] I. S. Duff, A. M. Erisman, and J. K. Reid. *Direct Methods for Sparse Matrices*. Oxford University Press, 1986.

[6] L. E. Malvern. *Introduction to the Mechanics of a Continuous Medium*. Prentice-Hall, Englewood Cliffs, New Jersey, 1969.

[7] J. Rauch and W. Köhler. “Collective and thermal diffusion in dilute, semidilute, and concentrated solutions of polystyrene in toluene”. Journal of Chemical Physics, 119(22):11977–11988, 2003.

[8] W. D. Richards and R. K. Prudhomme. “The viscosity of concentrated polymer solutions containing low molecular weight solvents”. Journal of Applied Polymer Science, 31:763–776, 1986.

[9] D.-M. Kim and E. B. Nauman. “Solution viscosity of polystyrene at conditions applicable to commercial manufacturing processes”. Journal of Chemical Engineering Data, 37:427–432, 1992.

[10] Y. Narukawa. Personal communication, 2002.

[11] S.-Y. Tam. *Stress Effects in Drying Coatings*. PhD thesis, University of Minnesota, Published by University Microfilms International, Ann Arbor, MI, 1997.

[12] H. Lei. *Flow, Deformation, Stress, and Failure in Solidifying Coatings*. PhD thesis, University of Minnesota, Published by University Microfilms International, Ann Arbor, MI, 1999.

[13] E. B. Gutoff and E. D. Cohen. *Coating and Drying Defects: Troubleshooting Operating Problems*. Wiley Interscience, 1995.

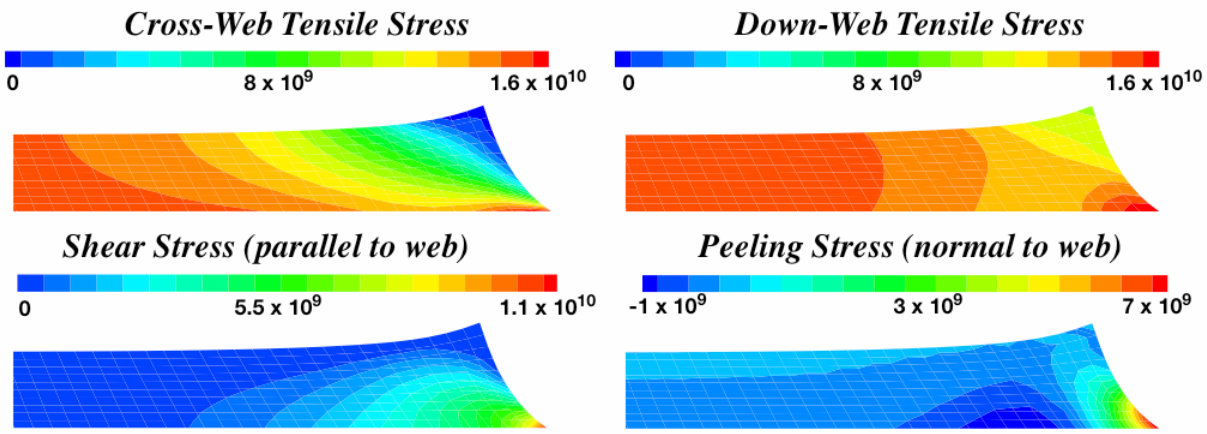


Figure 1: Stress distribution near the edge of a fully dried solid coating.

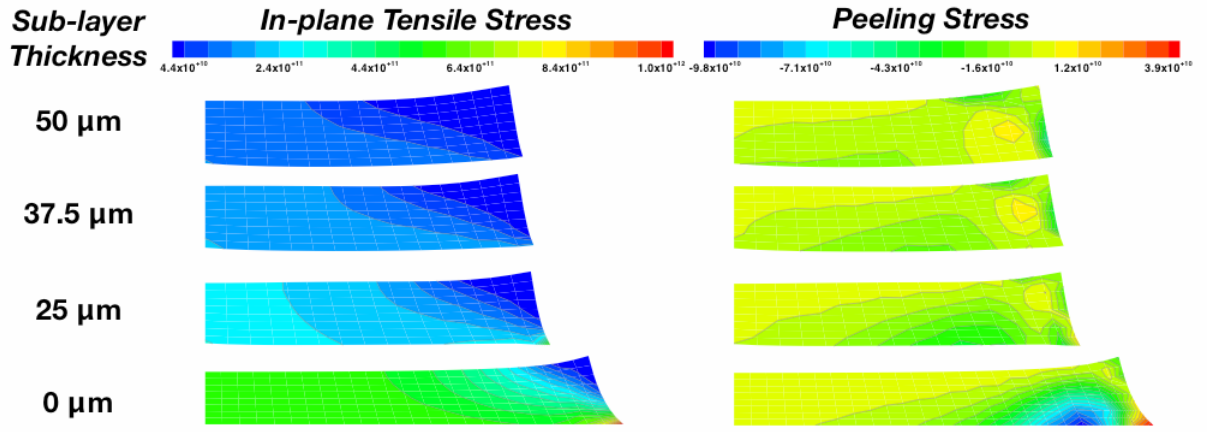


Figure 2: Cross-web tensile and peeling stresses in the top layer of a two-layer solid coating.

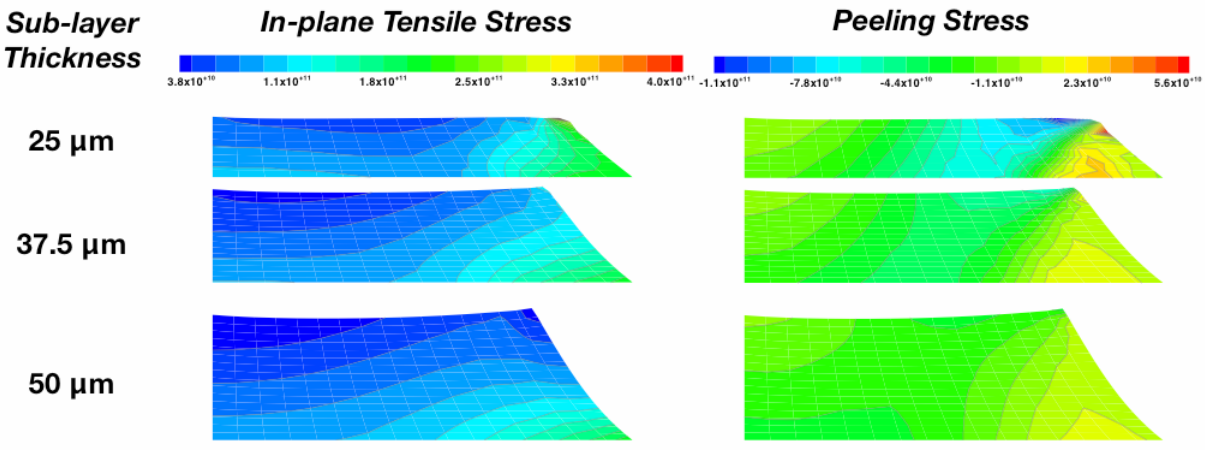


Figure 3: Cross-web tensile and peeling stresses in the bottom layer of a two-layer solid coating.

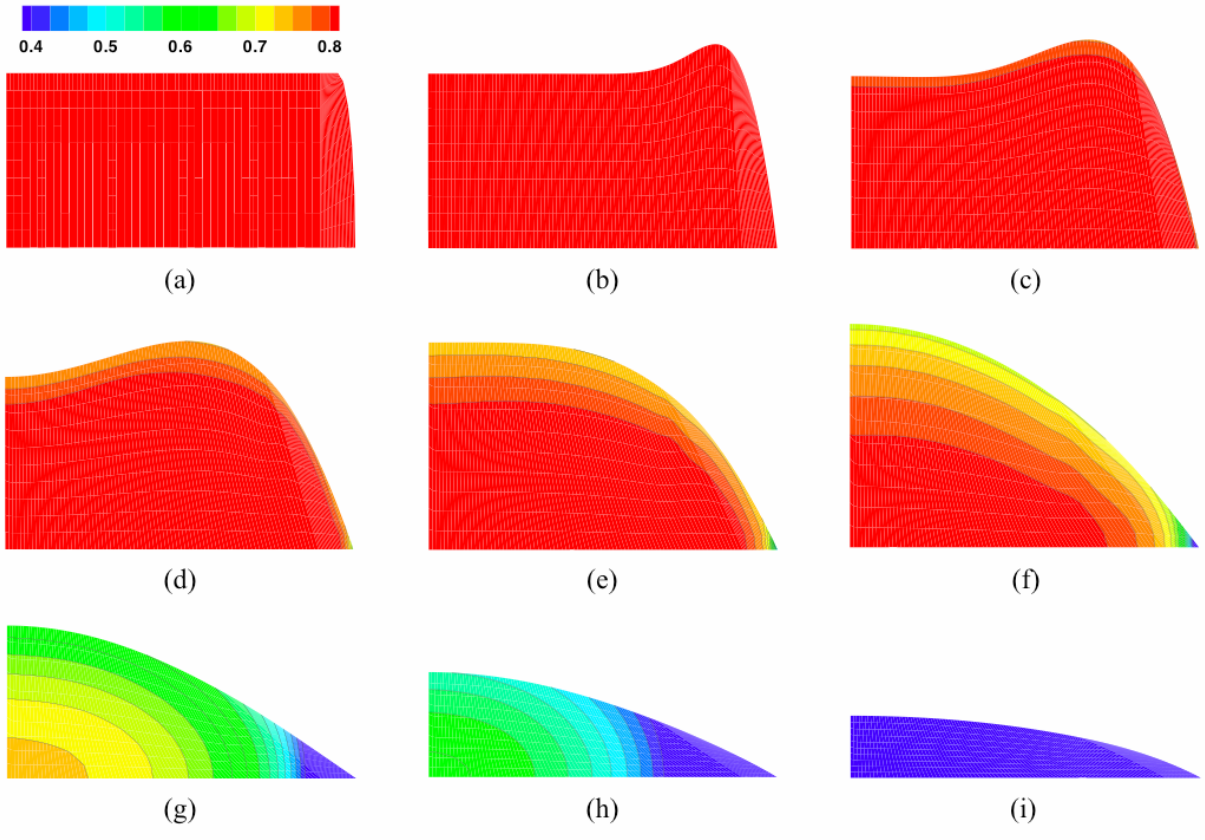


Figure 4: Edge shape of a still liquid coating at different times of drying.

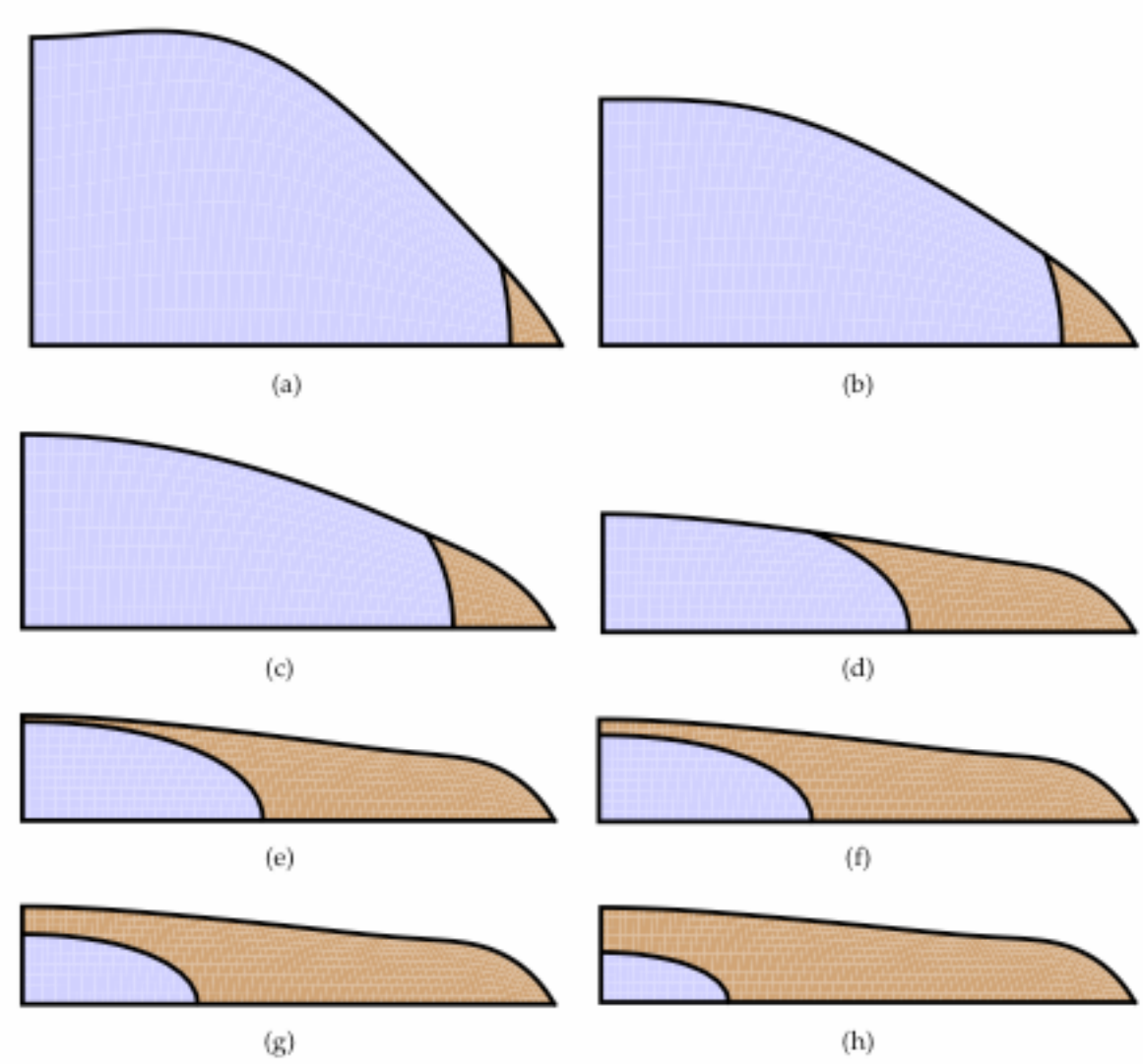


Figure 5: Solidification front movement near the edge of a drying coating.

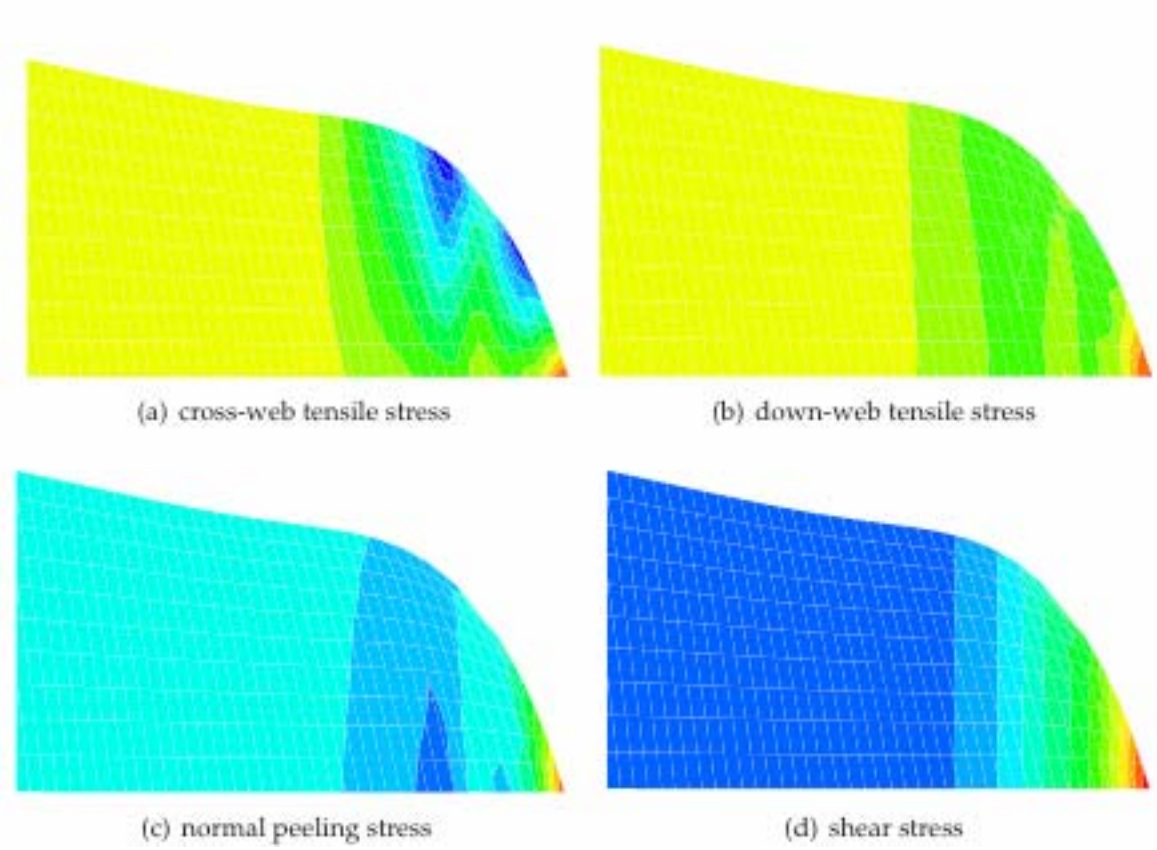


Figure 6: Final stresses near the edge of an applied liquid coating that solidified at a moving front.

FOUNTAIN FLOW REVISITED: THE EFFECT OF VISCOELASTICITY ON THE FREE SURFACE

Evan Mitsoulis

Computer-Aided Materials Processing – Rheology & Design (CAMP-R&D)
School of Mining Engineering and Metallurgy
National Technical University of Athens
Zografou 157 80, Athens, Greece
e-mail: mitsouli@metal.ntua.gr

Keywords: Viscoelasticity, Fountain Flow, Computational Rheology, Finite Element Method.

Abstract. *Fountain flow is the phenomenon of deceleration and outward motion of fluid particles as they approach a slower moving interface. Numerical simulations have been undertaken for the flow of Newtonian and viscoelastic fluids, obeying an integral constitutive equation of the K-BKZ type, capable of describing the behaviour of polymer melts. The flow simulations are performed for planar and axisymmetric geometries and show the shape and extent of the free surface, as well as the stresses and pressures in the system. The semicircle is a good approximation for the free surface of fountain flow, but detailed computations show the effect of elasticity on the free surface, which is non-monotonic as the elasticity level (or Weissenberg number) increases. The Newtonian values are faithfully reproduced and the new viscoelastic results show subtle differences, which influence the stress levels in the flow field.*

1 INTRODUCTION

Fountain flow is a well-known fluid-mechanics phenomenon present in all cases where a moving fluid displaces another immiscible fluid[1]. Within the context of non-Newtonian fluid mechanics, this type of flow is of interest in polymer processing, and in particular in the flow of polymer melts in injection-mold filling[2]. The theoretical problem has been adequately addressed in the mid-80's by a number of researchers[3-5], who dealt primarily with Newtonian fluids and showed intricate patterns developing when tracing particles, in agreement with experiments[3-6]. Mavridis et al.[3] showed in a sample run that power-law fluids only slightly affect the position of the flow front. Beris[7] argued that fountain flow patterns (especially the observed V-shapes[8]) are independent of the model and form as a consequence of fluid-mechanics principles that the fluid obeys at the front. Quite recently, Bogaerds et al.[9] performed calculations with a viscoelastic model (the extended pom-pom[10]) and showed that the flow front is not affected appreciably by the elasticity of the polymer melt, and it does not deviate much from a semicircle for a wide range of flow rates in a planar domain. All these works have shown that the results range from 0.8 to 1.05 of the radius of the semicircle, with the lower values corresponding to the axisymmetric case of a Newtonian fluid, while shear-thinning of power-law fluids brings the results closer to a semicircle front.

Although the problem is well understood from the physics and fluid mechanics points of view, it has become evident that numerically subtle differences exist for different fluid models and types of domain geometry (planar vs. axisymmetric). It is, therefore, the purpose of the present paper to address these issues and provide detailed results both for the free surface location and the other flow variables in the system for Newtonian and viscoelastic fluids.

In the present work, a standardized commercial low-density polyethylene melt (IUPAC-LDPE)[11] with long-chain branching is studied with the purpose of finding how rheology affects the flow behaviour in fountain flow. First the polymer is rheologically well-characterized with a modified K-BKZ integral model. Then it is simulated in fountain flow between parallel plates with the purpose of determining the free surface and finding the amount of centerline front movement. The results from the simulations are compared with the corresponding Newtonian results. Conclusions are drawn regarding the behaviour of different types of fluids in injection molding, where fountain flow is prominent.

2 MATHEMATICAL MODELLING

The problem at hand is that of a polymer melt flowing under a pressure gradient in a tube (or between parallel

plates) and acquiring a free surface at the moving front, called the "*fountain*". The tube has a radius R (or the plates have a half gap of H) and a length of L as shown here schematically in Figure 1. The polymer flows under shear flow in most of the domain except near the front, where due to the presence of the free surface and the stagnation point at B, rearranges itself to accommodate the free surface. The centerline front location z_{cl} reduced by the radius R gives rise to the definition of $\chi = \Delta z_{cl}/R$, in equivalence to the definition of *extrudate swell*[12]. In what follows, we consider that the channel length L is long enough ($L/R = L/H$ =10) so that it is adequate to impose at entry a fully developed velocity profile even for highly viscoelastic polymer melts.

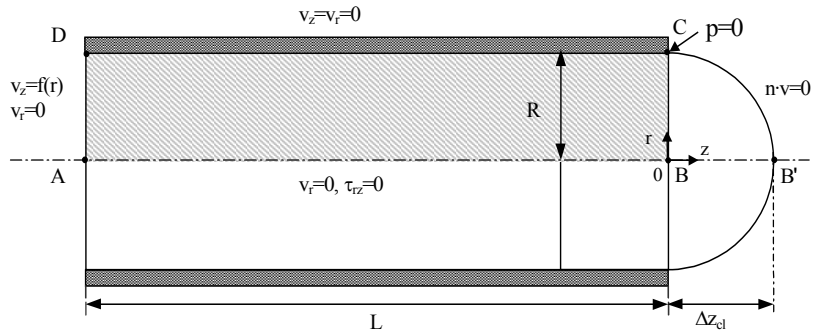


Figure 1. Schematic representation of fountain flow in a tube together with the boundary conditions for determination of the free surface at the moving front.

The flow is governed by the usual conservation equations of mass and momentum for an incompressible fluid under isothermal conditions:

$$\nabla \cdot \overline{\mathbf{v}} = 0 \tag{1}$$

$$0 = -\nabla p + \nabla \cdot \overline{\overline{\boldsymbol{\tau}}} \tag{2}$$

where $\overline{\mathbf{v}}$ is the velocity vector, $\overline{\overline{\boldsymbol{\tau}}}$ is the extra-stress tensor, and p is the scalar pressure.

The constitutive equation that relates the stress to the deformation history is a K-BKZ integral equation. In its more general form, it is written as[12]

$$\overline{\overline{\tau}} = \frac{1}{1-\theta}\int_{-\infty}^{t}\sum_{k=1}^{N}\frac{a_k}{\lambda_k}\exp\left(-\frac{t-t'}{\lambda_k}\right)H(I_C, I_{C^{-1}})\left(\overline{\overline{C}}_t^{-1}(t') + \theta\overline{\overline{C}}_t(t')\right)dt' \tag{3}$$

where λ_k and a_k are the relaxation times and relaxation modulus coefficients at a reference temperature T_0, respectively, θ is a material constant, and $H(I_C, I_{C^{-1}})$ is a strain-memory function depending on I_C, $I_{C^{-1}}$, the first invariants of the Cauchy-Green tensor C_t and its inverse C_t^{-1}, the Finger strain tensor. The material constant θ is given by

$$\frac{N_2}{N_1} = \frac{\theta}{1-\theta} \tag{4}$$

where N_1 and N_2 are the first and second normal stress differences, respectively.

The strain-memory (or damping) function H proposed by Papanastasiou et al.[13] has the following form

$$H(I_{C^{-1}}) = \frac{\alpha}{(\alpha-3)+\beta I_{C^{-1}}+(1-\beta)I_C} \tag{5}$$

where α and β are material constants. This damping function, called *PSM*, has been used extensively for predicting the viscoelastic character of polymer melts with good success[13,14]. Note that in principle, multiple H_k can be used by having different α_k's and β_k's for every relaxation mode.

3 DIMENSIONLESS PARAMETERS

For viscoelastic materials a generalized Weissenberg number *Ws* is defined as[12]:

$$Ws = < \lambda > \frac{V}{H} \tag{6}$$

where $<\lambda>$ is the average relaxation time of the polymer and V is a characteristic velocity, usually taken as the average velocity of the melt, V_{avg}.

In the case of fountain flow of polymer melts and with respect to Figure 1, it has been customary to define the quantity $\dot{\gamma}_a$, called the *apparent shear rate*, which is related to the volumetric flow rate Q and the width W by

(axisymmetric geometry) $\dot{\gamma}_a = 4Q / \pi R^3$ (7a)

(planar geometry) $\dot{\gamma}_a = 6Q / W(2H)^2$ (7b)

The apparent shear rate can be used as a convenient representation of flow rate in the absence of a single relaxation time. This measure has been used in the experiments by Meissner[11] and will be used in the present work as well.

4 RHEOLOGICAL CHARACTERIZATION

This work is concerned with a low-density polyethylene (LDPE), under the code name IUPAC-LDPE melt A, and used previously as a benchmark material for rheological characterization and flow behaviour[11,14]. Some of its specific data is given in Table 1[15].

Table 1: Material parameter values for the IUPAC-LDPE melt A[15].

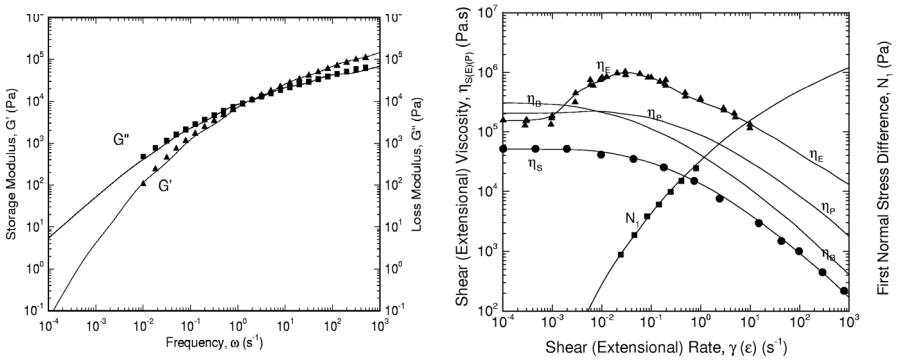
Property	LDPE
Density, ρ (25°C) (g/cm^3)	0.918
Mass-Average Molar Mass, M_w (g/mole)	258,000
Polydispersity, M_w/M_n	16.6
Melt Temperature, T_m (°C)	108
Zero-Shear-Rate Viscosity, η_0 at 150°C (Pa·s)	51,500
Activation Energy, E_a (J/mole)	58,000

This polyethylene has a random-like long-chain branched structure due to radical polymerization at high pressure. The degree of long-chain branching cannot be determined quantitatively. The rheological material functions of polymer melts are dynamic properties, such as the storage and loss moduli, and steady properties, such as their viscosity in shear and elongation, and the normal stresses in shear. The viscosity in shear was measured by mechanical-dynamic experiments using a rotational rheometer. In Figure 2a, storage and loss moduli at 150°C are shown as a function of frequency (symbols on the graphs). These data can be used for the determination of the relaxation spectra, which are required for modelling the flow behaviour (lines on the graphs). Figure 2b shows the viscosity function calculated from G' and G''.

The above material data have been modelled with the K-BKZ integral eqn. (3) and the damping function *PSM* (eqn. 5) in order to determine the parameters of the model. These are reproduced in Table 2 and show a spectrum with 8 relaxation times. From the relaxation spectrum it follows that the zero-shear viscosity $\eta_0 = \Sigma(a_k\lambda_k)$ = 51,064 Pa·s and the average relaxation time $<\lambda> = \Sigma(a_k\lambda_k^2)/\Sigma(a_k\lambda_k)$ = 58.7 s, thus giving evidence of the pronounced elasticity of LDPE.

Table 2: Material parameter values used in eqn. (3) for fitting data of the IUPAC-LDPE melt at 150°C (θ = –0.25, ρ = 0.92 g/cm^3, $<\lambda>$ = 58.7 s, η_0 = 51,064 Pa·s).

k	λ_k (s)	a_k (Pa)	α_k	β_k
1	10^{-4}	129,000	14.38	0.018
2	10^{-3}	94,800	14.38	0.018
3	10^{-2}	58,600	14.38	0.08
4	10^{-1}	26,700	14.38	0.12
5	10^{0}	9,800	14.38	0.12
6	10^{1}	1,890	14.38	0.16
7	10^{2}	180	14.38	0.03
8	10^{3}	1.0	14.38	0.002



(a) (b)

Figure 2. Rheological data for the IUPAC-LDPE melt. Symbols represent experimental data. Lines correspond to model predictions according to the K-BKZ model (eqn. 3) with the parameters of Table 2.

With the parameters of Table 2, it is possible to fit the experimental data for different shear rates and obtain the shear-thinning behaviour for the shear viscosity, as well as the quadratic behaviour at very low shear rates for the first normal stress difference, as shown in Figure 2.

For the uniaxial elongational viscosity η_E, multiple values of the β–parameter have been used to reduce the amount of strain-hardening at higher elongational rates. We also note in Figure 2b that the planar extensional viscosity η_P starts from the Newtonian value of $\eta_P = 4\eta_0$ and follows the shear viscosity η_S in its thinning behaviour. The biaxial extensional viscosity η_B starts from the Newtonian value of $\eta_B = 6\eta_0$ and also follows the shear viscosity in its thinning behaviour. Previous experience with simulations for polyethylene melts[8] has shown that the extensional behaviour is crucial in predicting correctly their flow behaviour in both axisymmetric and planar flows.

The data for the second normal stress difference N_2 have shown a non-zero value, $\theta = -0.25$[12].

5 METHOD OF SOLUTION

The special numerical scheme developed by Luo and Mitsoulis[16] for the implementation of integral-type constitutive equations with the finite element method (FEM) has been used. This scheme is effectively an EVSS-G/SI scheme in the jargon of viscoelastic simulations. This means that there is a splitting of elastic and viscous stresses (Elastic-Viscous-Stress-Splitting), that the viscous stresses enter the stiffness matrix using a reference viscosity ($\eta_{ref} = 1$), while the elastic stresses enter the load vector and are incremented to reach higher elasticity levels. The elastic stresses are calculated according to the constitutive eqn. (3) using streamline integration (SI). Galerkin averaging is used for the velocity-gradient field (the G in EVSS-G) to obtain a smooth field. However, our method uses as primary variables only the two velocities and pressure (*u-v-p formulation*) instead of also using the stresses and the rates-of-strain, as done in the *mixed formulation (u-v-p-*τ_{ij}*-*$\dot{\gamma}_{ij}$*)* for differential models. This renders the simulations extremely fast, even for multiple relaxation modes.

The numerical algorithm for convergence is Picard iteration, i.e., direct substitution. Convergent solutions have been obtained for a wide range of apparent shear rates for a fixed domain[16]. However, for flows *with free surfaces,* such as the ones considered here, and due to the highly viscoelastic nature of the melts manifested by very high deformations in the order of 100%, it was found necessary to proceed carefully and use severe under-relaxation for the free surface movement in the order of 10%, sometimes allowing only as much as 1% movement of the newly updated position.

The solution procedure advances slowly from low flow rates (Newtonian behaviour) to higher ones by using a *flow rate increment scheme*. On average 13 CPU s per iteration were needed with a mesh having 600 elements on an Athlon^{TM} 64 processor at 3400+ MHz for a total of 1150 iterations up to $\dot{\gamma}_a$ = 300 s^{-1}. The criterion for convergence was 10^{-3} for the maximum changes in the velocities, 10^{-2} for the pressure, and 10^{-3} for the free surface.

The initial meshes form a rectangle, with the origin at the RHS centerline. During the solution process the mesh deforms to accommodate the fountain flow with the free surface at the front. The deformed finite element meshes used in this study at the end of the computations are shown in Figure 4 for Newtonian fluids (both geometries) and for the LDPE melt in planar flow. The domain in its entirety extends to $-10H$ upstream of the front to ensure the correct imposition of a fully developed velocity profile upstream. Because of symmetry only half the domain is used. The grid consists of 600 quadrilateral 8-node serendipity elements, 1901 nodes, and 4453 unknown u-v-p degrees of freedom (DOF).

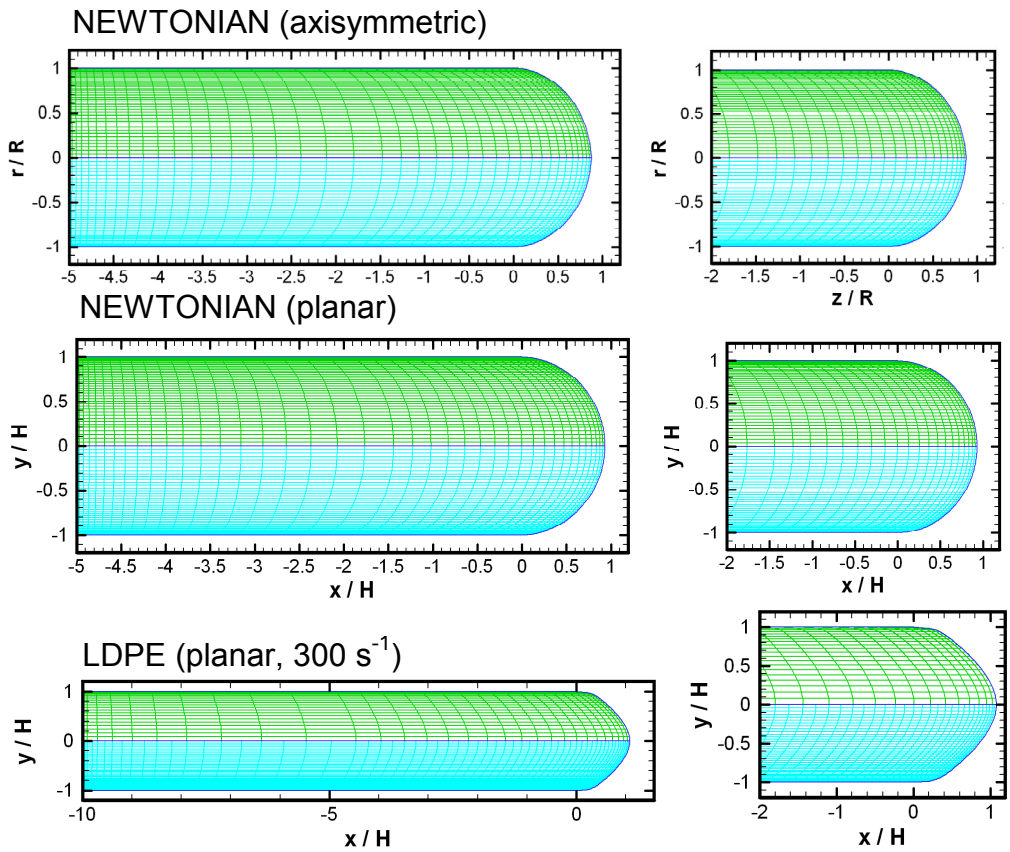


Figure 4. Deformed finite element meshes used in the simulations: (a) full view (upper half is the original mesh, lower half has quadruple number of elements for the calculation of the stresses and the stream function), (b) blown-up view of the computational domain near the front.

6 RESULTS AND DISCUSSION

6.1 Newtonian results – Flow field

First the numerical simulations have been carried out for Newtonian fluids in both planar and axisymmetric geometries. The Newtonian values are: $\chi = 0.876$ (axisymmetric) and $\chi = 0.931$ (planar). We note a difference of 5.5% in favour of planar geometry (cf. the extrudate swell values of 13% for axisymmetric and 18.5% for planar geometries[12,17]). Thus, the maximum thrust forward of the advancing front at the centerline can be at most 87.6% (tube) and 93.1% (flat channel) for Newtonian fluids.

Behrens et al.[5] have produced a table indicating such values, which is updated here as Table 3. The reasons for the discrepancies are due to the mesh construction and its density, especially around the singular exit point C.

Table 3. Dimensionless centerline front location χ for Newtonian fluids.

Investigator	Geometry	Frame of Reference	χ
Mavridis et al.[3]	Planar	Moving-steady	0.90
Coyle et al.[4]	Planar	Moving-steady	0.93
Behrens et al.[5]	Planar	Moving-steady	0.94
Behrens et al.[5]	Planar	Fixed-transient	0.91
Bogaerds et al.[8]	Planar	Moving-steady	0.92
This work	Planar	Moving-steady	0.93
Mavridis et al.[3]	Axisym.	Moving-steady	0.83
Behrens et al.[5]	Axisym.	Moving-steady	0.82
Behrens et al.[5]	Axisym.	Fixed-transient	0.86
Behrens et al.[5]	Axisym.	Exp. result	0.83±0.04
This work	Axisym.	Moving-steady	0.88

It must be pointed out that the fountain flow problem is more difficult than the extrudate swell problem, because the free surface is perpendicular to the main flow direction, whereas in extrudate swell the free surface follows the main flow direction. Also, in the Newtonian extrudate swell the domain gets deformed in the order of 13-19%, while in fountain flow the domain is deformed in the order of 88-93%.

The numerical solutions give a wealth of information regarding all kinematic and dynamic variables of the flow field. The results for axisymmetric and planar geometries are shown in Figure 5, where contours are given for several variables. The kinematic variables are the stream function (PSI), the two velocities v_x or v_z (U) and v_y or v_r (V), and the shear rate $\dot{\gamma}_{xy}$ or $\dot{\gamma}_{rz}$ (GXY), while the dynamic variables are the pressure, p (P), and the 3 (planar) or 4 (axisymmetric) extra stresses, τ_{xx} or τ_{zz}, τ_{yy} or τ_{rr}, τ_{xy} or τ_{rz} and $\tau_{\theta\theta}$ (or TXX, TYY, TXY, T33). Interesting points to notice is the fully developed flow upstream, which extends down to about one radius R or half gap H from the exit. The streamlines, u-contours and shear rate contours are parallel to the walls and the isobars are vertical lines. The maximum value for the pressure occurs at entry (not shown), while the reference zero value has been set at the exit point C. Small negative pressures also occur at the flow front and are a consequence of the singular point at the exit wall. The strain rates are equal to the stresses due to assuming a unit viscosity. Most of the rearrangement occurs right at the exit where the minima and maxima occur because of the exit singularity. Eleven (11) contours have been drawn between the minimum and the maximum values.

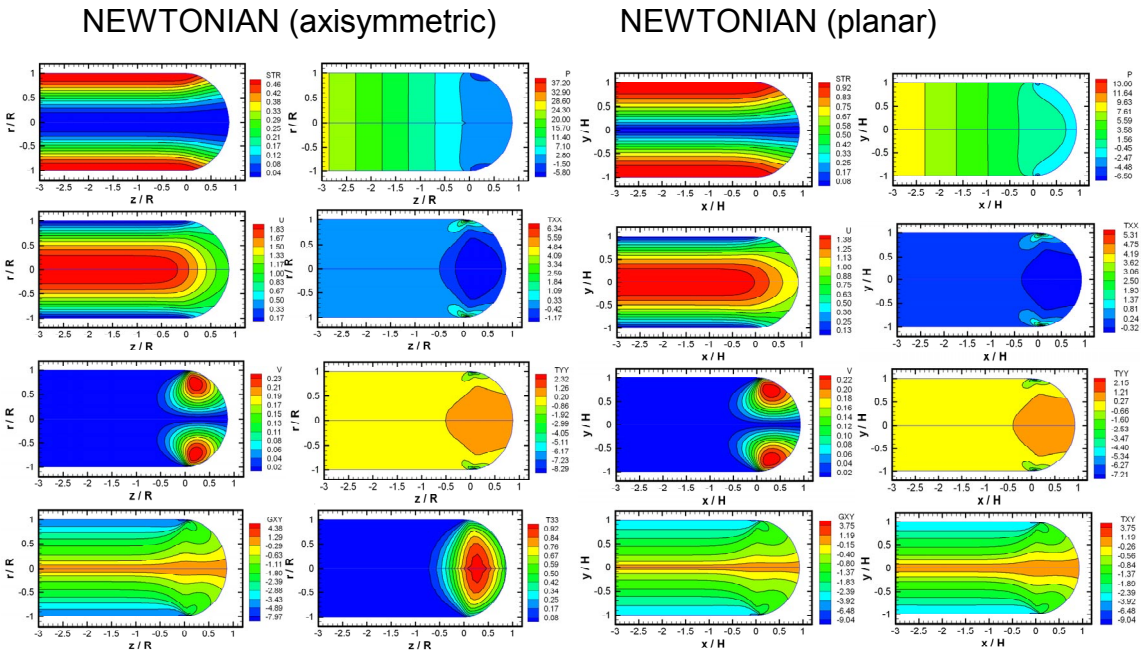


Figure 5. Contours of kinematic and dynamic variables obtained from the simulations of fountain flow for Newtonian fluids. Eleven (11) contour values are shown between maximum and minimum.

6.2 Viscoelastic results – Flow field

The viscoelastic numerical simulations have been carried out for a wide range of apparent shear rates $\dot{\gamma}_a$ from 0.001 s^{-1} and reaching as high as 300 s^{-1}. Typical results from one such run are shown in Figure 6, where contours are given for the highest apparent shear rate $\dot{\gamma}_a$ = 300 s^{-1} (Ws=5870, from eq. 6) for several variables.

The kinematic variables are the stream function (PSI), the two velocities v_x (U) and v_y (V), and the shear rate $\dot{\gamma}_{xy}$ (GXY), while the dynamic variables are the pressure, p (P), and the 3 extra stresses, τ_{xx}, τ_{yy}, τ_{xy} (or TXX, TYY, TXY). The pressure and the stresses are given in dimensionless form, being divided by $\eta_0 V_{avg}/H$. The velocities are divided by V_{avg}. Interesting points to notice and in contrast with the Newtonian flow fields are the squeezing and bending of the streamlines near the exit due to the viscoelastic nature of the melt. Also the fully developed shear flow is reduced compared to the Newtonian. The melt now takes a full gap ($2H$) to rearrange itself near the exit, and produces a bullet-like profile with less roundness than in the Newtonian case. Because of the assumed non-zero second normal stress difference ($\theta \neq 0$), the isobars are not vertical lines but exhibit a parabolic profile between the walls. The maximum value of the pressure is at entry (not shown), while high negative pressures occur near the singularity, which is now much more pronounced. The shear rates for this case are indeed high, reaching values at the die wall > 500 s^{-1}. The stress contours show that the flow rearranges itself

more than a gap from the exit, with most of the rearrangement occurring right at exit where the minima and maxima occur because of the exit singularity.

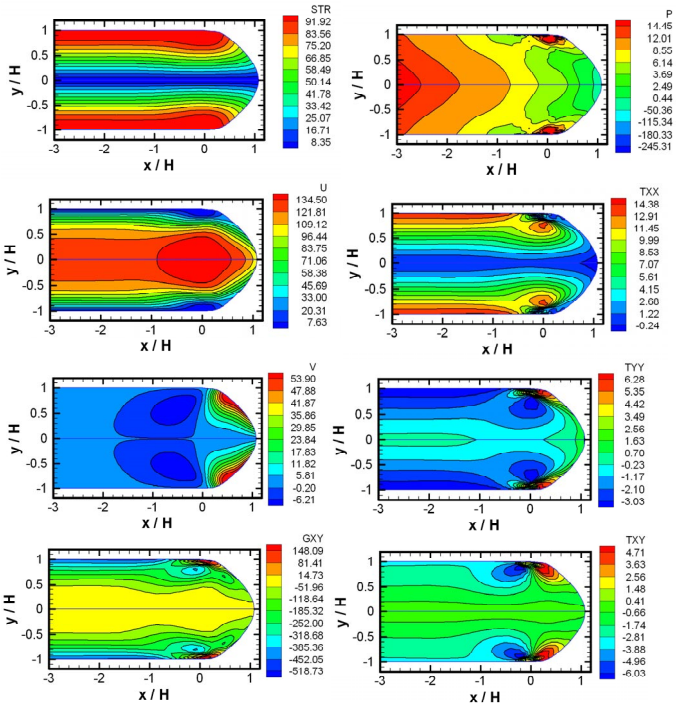


Figure 6. Contours of kinematic and dynamic variables obtained from the simulations of fountain flow for the IUPAC-LDPE melt A at $\dot{\gamma}_a$ = 300 s^{-1}. Eleven (11) contour values are shown between maximum and minimum.

6.3 Viscoelastic results – Centerline front location

The dependence of the centerline front location on the apparent shear rate $\dot{\gamma}_a$ is presented in Figure 7 for the IUPAC-LDPE melt A. The behaviour of the front location is typical of polymer melts, i.e., starting from the Newtonian value of 93% of the half gap *H* at extremely low shear rates, it first decreases reaching a minimum of 79% around 0.3 s^{-1}, and then increases substantially reaching values of 108% at apparent shear rates of 300 s^{-1}. An important observation that can be made is that the phenomenon is nonlinear. In the past, it has been assumed that fountain flow profiles are approximately those of a semicircle[1-2], and they do not change much either for different fluid models or different types of geometry. The current results show that those observations are generally true in a rough way. However, different fluid models do play a role, the χ-variable depends for polymer melts on the flow rate, and the type of geometry is also an important parameter.

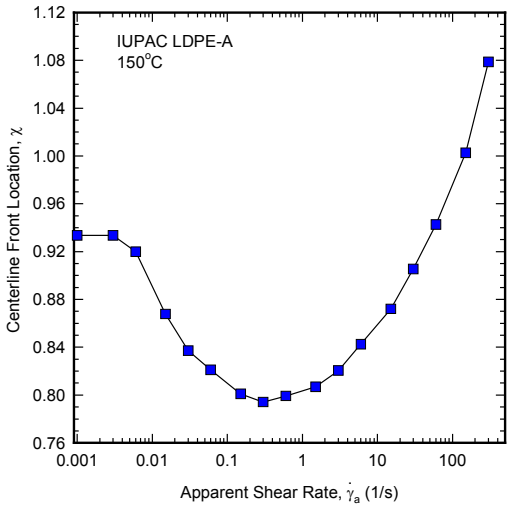


Figure 7. Centerline front location as a function of apparent shear rate for the IUPAC-LDPE melt A at 150ºC.

7 CONCLUSIONS

The current work has addressed the issue of simulating Newtonian and viscoelastic polymer melts flowing under a pressure gradient between parallel plates or tubes for the determination of fountain flow. The simulations were undertaken for the flow of a standardized branched low-density polyethylene melt (IUPAC-LDPE). For the viscoelasticity of the polymers, rheological data of dynamic and steady properties were used over a wide-range of shear and elongational rates. The rheological data has been fitted with the K-BKZ/PSM model with a spectrum of 8 relaxation times and associated parameters for the melt. This well-known model has been used for the first time in fountain flow simulations of polymer melts.

The simulations have addressed the issue of finding the moving fountain front by constructing a streamline, which is the free surface of the fluid. This is done after a solution for the conservation and constitutive equations has been achieved via the Finite Element Method. FEM uses as primary variables the two velocities and pressure (primitive variables approach). Then the stresses are calculated via streamline integration and then the free surface coordinates are found by using the updated velocity field and integrating the velocity profiles there. The simulations revealed a distinct behaviour in the LDPE melt compared with the Newtonian simulations. LDPE shows a bullet-like fountain flow at high shear rates with less roundness of the front than the Newtonian fluids. These results are a prelude for a full study of non-isothermal effects in fountain flow of polymer melts.

ACKNOWLEDGMENTS

Financial support from the European Union's Centres of Excellence consortium under the title of "Polymer Injection Advanced Molding" (PIAM) project (Contract # NMP2-CT-2003-505878) is gratefully acknowledged. Also, financial assistance from the SOCRATES program (a subprogram of ERASMUS) of the EU for bilateral collaboration between Greece and Cyprus in the form of a travel grant, is gratefully acknowledged.

REFERENCES

[1] Rose, W. (1961), "Fluid-Fluid Interfaces in Steady Motion", Nature, Vol. 191, pp. 242-247.

[2] Tadmor, Z., Gogos, C.G. (1979), *Principles of Polymer Processing*, SPE Monograph Series, Wiley, New York.

[3] Mavridis, H., Hrymak, A.N., Vlachopoulos, J. (1986), "Finite Element Simulation of Fountain Flow in Injection Molding", Polym. Eng. Sci., Vol. 26, pp. 449-454.

[4] Coyle, D.J., Blake, J.W., Macosko, C.W. (1987), "The Kinematics of Fountain Flow in Mold-Filling", AIChE J., Vol. 33, pp. 1168-1177.

[5] Behrens, R.A., Crochet, M.J., Denson, C.D., Metzner, A.B. (1987), "Transient Free-Surface Flows: Motion of a Fluid Advancing in a Tube", AIChE J., Vol. 33, pp. 1178-1186.

[6] Mavridis, H., Hrymak, A.N., Vlachopoulos, J. (1988), "Transient Free-Surface Flows in Injection Mold Filling", AIChE J., Vol. 34, pp. 403-410.

[7] Beris, A.N. (1987), "Fluid Elements Deformation behind an Advancing Flow Front", J. Rheol., Vol. 31, pp. 121-124.

[8] Schmidt, R.L. (1974), "A Special Mold and Tracer Technique for Studying Shear and Extensional Flows in a Mold Cavity during Injection Molding", Polym. Eng. Sci., Vol. 14, pp. 797-801.

[9] Bogaerds, A.C.B., Hulsen, M.A., Peters, G.W.M., Baaijens, F.P.T. (2004), "Stability Analysis of Injection Molding Flows", J. Rheol., Vol. 48, pp. 765-785.

[10] Verbeeten, W.M.H., Peters, G.W.M., Baaijens, F.P.T. (2001), "Differential Constitutive Equations for Polymer Melts: The Extended Pom-Pom Model", J. Rheol., Vol. 45, pp. 823-844.

[11] Meissner, J. (1975), "Basic Parameters, Melt Rheology, Processing and End-Use Properties of Three Similar Low Density Polyethylene Samples", Pure Appl. Chem., Vol. 42, pp. 551-612.

[12] Tanner, R.I. (2000), *Engineering Rheology*, 2nd Ed., Oxford University Press, Oxford.

[13] Papanastasiou, A.C., Scriven, L.E., and Macosko, C.W. (1983), "An Integral Constitutive Equation for Mixed Flows: Viscoelastic Characterization", J. Rheol., Vol. 27, pp. 387-410.

[14] Mitsoulis, E. (2001), "Numerical Simulation of Entry Flow of the IUPAC-LDPE Melt", J. Non-Newtonian Fluid Mech., Vol. 97, pp. 13-30.

[15] Mitsoulis, E., Schwetz, M., Münstedt, H. (2003), "Entry Flow of LDPE Melts in a Planar Contraction", J. Non-Newtonian Fluid Mech., Vol. 111, pp. 41-61.

[16] Luo, X.-L., Mitsoulis, E. (1990), "An Efficient Algorithm for Strain History Tracking in Finite Element Computations of Non-Newtonian Fluids with Integral Constitutive Equations", Int. J. Num. Meth. Fluids, Vol. 11, pp. 1015-1031.

[17] Mitsoulis, E., Vlachopoulos, J., Mirza, F.A. (1984), "Simulation of Extrudate Swell from Long Slit and Capillary Dies", Polym. Proc. Eng., Vol. 2, pp. 153-177.

SIMULATING THE SATURATION FRONT USING A FRACTIONAL DIFFUSION MODEL

E. Gerolymatou[1], I. Vardoulakis[1] and R. Hilfer[2]
[1] Department of Mechanics,
Faculty of Applied Mathematics and Physics,
N.T.U. Athens ,
9 Iroon Politechneiou street,
15780 Athens, Greece
corresponding author: eleni@mechan.ntua.gr.
[2] Institute for Computational Physics,
Physics on Supercomputers,
University of Stuttgart,
Pfaffenwaldring,
2770569 Stuttgart, Deutschland

Keywords: Fractional Calculus, Anomalous Diffusion, Fractional Time Evolutions.

Abstract. *In this paper the possibility of making use of fractional derivatives for the simulation of the flow of water through porous media and in particular through soils is considered. The Richards equation, which is a non-linear diffusion equation, will be taken as a basis and is used for the comparison of results. Fractional derivatives differ from derivatives of integer order in that they entail the whole history of the function in a weighted form and not only its local behavior, meaning that a different numerical approach is required. Previous work on the topic will be examined and a consistent approach based on fractional time evolutions will be presented.*

1 INTRODUCTION

Infiltration is defined as the flow of water through porous media and in particular through soils. It follows the ordinary laws of hydrodynamics. The Richards equation, which is a non-linear diffusion equation, is usually used for its description [1, 2, 3], even though in several cases it fails to predict variations in the behavior of different types of soil. The present work is an attempt to ascertain whether fractional calculus is suitable as a tool for the simulation of the saturation front in partially saturated porous media. Therefore the substitution of the derivative with respect to time with a fractional derivative of order smaller than unity is considered [4, 5]. Fractional calculus is a branch of mathematics related to integrals and derivatives of arbitrary order and dates back to the 17th century [6]. Fractional derivatives differ from derivatives of integer order in that they entail the whole history of the function in a weighted form and not only its local behavior. Lately it was found to have many applications in physics and mechanics, especially concerning the description of anomalous diffusion [7, 8]. It has been suggested[5] to replace the time derivative in the Richards equation by a fractional derivative as a way to describe experimental observations that show deviations from normal diffusive scaling. We found that the referenced paper [5] contains several theoretical errors (see also [9]), and we discuss ways how these can be eliminated. An improved and consistent approach based on fractional time evolutions [7] will be presented.

2 ABSORPTION IN POROUS MEDIA

2.1 Derivation of the Richards equation

A soil mass generally consists of a network of partially or totally interconnected interspaces of various sizes and shapes. These interspaces may be filled with air or water or both. The volumetric moisture content $\theta(t,\mathbf{x})$, also called local volume fraction of water, is defined as the ratio of the volume of water to the volume of a representative elementary soil volume located at position $\mathbf{x}$. Under the assumption that the porosity (defined as the volume fraction of pores) is constant, and that the speed of the solid phase vanishes, the mass balance for the liquid phase yields:

$$\frac{\partial\theta}{\partial t} = -\mathrm{div}(\mathbf{q}) \qquad (1)$$

where $\mathbf{q}$ is the specific discharge of fluid through the interstices of the solid matrix. The flow of liquids in unsaturated media is determined by the pressure, gravity and capillary forces acting on the liquid. We consider the capillary potential (per unit weight of water):

$$\Psi = \left(p / \rho_w g + z \right) \tag{2}$$

where the pressure is determined by the surface tension and curvature of the air-liquid interface. We will consider horizontal absorption and therefore neglect the effect of gravity. Thus the moisture discharge vector is related to the total potential by means of the following equation:

$$\mathbf{q} = -K \cdot \mathrm{grad}\left(\Psi\right) \tag{3}$$

which is known as the moisture conduction equation. Combining equations (1) and (3) we acquire the following equation, known as the Richards equation:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x}\left(D(\theta) \frac{\partial}{\partial x}\theta \right) + \frac{\partial}{\partial y}\left(D(\theta) \frac{\partial}{\partial y}\theta \right) + \frac{\partial}{\partial z}\left(D(\theta) \frac{\partial}{\partial z}\theta \right) \tag{4}$$

where $D(\theta) = K\left(d\Psi / d\theta\right)$ is known as the moisture diffusivity. For example from Philip [10] we get the following empirical moisture diffusivity function

$$D \approx D_0 \exp(c \cdot \theta) \tag{5}$$

where $D_0 = 8.3 \cdot 10^{-7} cm^2 / min$ and $c = 19$. In Fig. 1 the experimental data and the fitted function of eq. (5) are presented. In what follows we will only consider one - dimensional moisture diffusion along horizontal soil columns. For this case the Richards equation will take the form

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x}\left(D(\theta) \frac{\partial}{\partial x}\theta \right). \tag{6}$$

We choose the coordinate system such that $x = 0$ corresponds to the left end of the horizontal soil column.

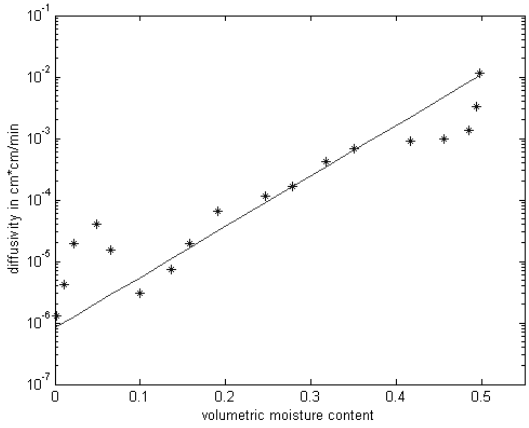


Fig. 1: Empirical diffusivity function from reference [10].
The fitted line corresponds to eq. (5).

2.2 Discussion

In the former analysis, the soil skeleton has been assumed to be rigid and the inertial effects have been assumed to be negligible, as the progress of the phenomenon is slow. The thermal effects and the effects of condensation and evaporation have been neglected, as well as the complications arising from the interplay of air and water [11,12]. With the introduction of the similarity variable

$$\xi = \frac{x}{\sqrt{t}} \tag{7}$$

Richard's equation (6) transforms into an ordinary differential equation, which has been used to find analytical solutions for soil water flow problems and also to find the dependence of the conductivity on the degree of saturation [13]. However, as summarized in a recent publication by Pachepsky et al. [5], significant deviations from the scaling law eq. (7) have been observed in many published experiments. Relationships between positions and times at which a particular value of the volumetric moisture content is observed suggest a similarity transformation of the form

$$\xi = x \cdot t^{-\alpha/2} \tag{8}$$

The case of $\alpha < 1$ could be interpreted as non-Brownian transport of particles that remain motionless for extended periods of time, for example, when waiting periods have a power law distribution.

In the aforementioned paper of Pachepsky et al. [5] the defect of Richards law is addressed by resorting into non-standard diffusion mathematical models that involve fractional derivatives with respect to time. It is worth noticing that one of us[4] has shown that the analysis of Pachepsky et al. (2003) et al. is mathematically flawed. An attempt to correct this analysis and the corresponding numerical integration of the fractional diffusion equation is shown below.

3 FRACTIONAL CALCULUS

Fractional calculus is the field of mathematical analysis which deals with the investigation and applications of integrals and derivatives of arbitrary order [6,7]. Although the term 'fractional calculus' is actually a misnomer, the designation 'integration and differentiation of arbitrary order' being more appropriate, it is well established due to prevailing use. In contrast to integration and differentiation of integer order, for integration and differentiation of arbitrary order a great variety of definitions exists. That is both one of the advantages and one of the disadvantages of fractional calculus.

If $f(x)$ is locally integrable on (γ, ∞), then the right hand fractional Riemann-Liouville integral of $f(x)$ of order $\alpha > 0$ is defined as

$$I_{\gamma+}^{\alpha} f(x) = \frac{1}{\Gamma(\alpha)} \int_{\gamma}^{x} (x-u)^{\alpha-1} f(u) du \tag{9}$$

for almost all $-\infty < \gamma < x < \infty$ and for suitable f. The subscripts in I denote the terminals of integration in the given order.

The following general definition of fractional derivatives was introduced in [7]: The right sided fractional derivative of order $0 < \alpha < 1$ and type $0 \le \beta \le 1$ with respect to t is defined by

$$D_{\gamma+}^{\alpha,\beta} f(t) = \left(I_{\gamma+}^{\beta(1-\alpha)} \frac{d}{dt} \left(I_{\gamma+}^{(1-\beta)(1-\alpha)} f \right) \right)(t) \tag{10}$$

for functions for which the expression on the right hand side exists.

The fractional Riemann – Liouville derivative is a special case of eq. (10) corresponding to $\beta = 0$, namely

$$D_{\gamma+}^{\alpha,0} f(x) = \frac{d}{dx} I_{\gamma+}^{1-\alpha} f(x) \tag{11}$$

where $0 < \alpha < 1$.

Another definition introduced by Liouville [14] in 1832, but often referred to as "the Caputo approach", corresponds to the fractional derivative of order α and type 1 as defined in eq. (10) and has proved to be very popular among engineers, especially as far as the field of viscoelasticity is concerned. It reads

$$D_{\gamma+}^{\alpha,1} f(x) = I_{\gamma+}^{1-\alpha} \frac{d}{dx} f(x) = \frac{1}{\Gamma(1-\alpha)} \int_{\gamma}^{x} \frac{f'(u)}{(x-u)^{\alpha}} du \tag{12}$$

where $0 < \alpha < 1$. This is a far more restrictive definition than the previous one, in that it demands that the derivative of $f(x)$ be absolutely integrable.

It is crucial to note that the different definitions of the fractional derivatives and integrals have a different

physical meaning. It is therefore of great significance that care is taken, when entering the field of applications. One example of this fact is the connection of the fractional derivatives to continuous time random walks. Eq.(13) has a rigorous relationship with continuous time random walks, whereas the solution of eq.(14) does not admit a probabilistic interpretation [15].

$$D_{0+}^{\alpha,1} f(x,t) = C_\alpha \frac{\partial^2 f(x,t)}{\partial x^2} \tag{13}$$

$$D_{0+}^{\alpha,0} f(x,t) = C_\alpha \frac{\partial^2 f(x,t)}{\partial x^2} \tag{14}$$

where C_α is a fractional diffusion constant.

4 PREVIOUS WORK

Pachepsky et al. [5], following the scaling deviations observed in experiments from the scaling resulting from the Richards equation (cf. Table 1 of that reference) considered the equation

$$D_{0+}^{\alpha,0}\theta = \frac{\partial}{\partial x}\left(D_\alpha(\theta)\frac{\partial \theta}{\partial x}\right) \tag{15}$$

replacing the derivative with respect to time with a fractional one. Consequently they attempted a solution of the resulting time-fractional absorption equation by inserting the similarity variable from eq. (8) and transforming the equation into an ordinary differential equation, as was done by Philip, who introduced the similarity transform in the Richards equation. In an attempt to reproduce the results, we found that the transformation of eq. (15) into an ordinary fractional differential equation is not possible in the same way as in the case $\alpha = 1$. This is because the authors assumed the following relationship to hold:

$$D_{0+}^{\alpha,0}\theta = \frac{d\theta}{d\xi} D_{0+}^{\alpha,0}\xi \tag{16}$$

which leads to the ordinary fractional differential equation:

$$\frac{d}{d\xi}\left(D(\theta)\frac{d\theta}{d\xi}\right) - \frac{\Gamma(1-\alpha/2)}{\Gamma(1-3\alpha/2)}\xi\frac{d\theta}{d\xi} = 0 \tag{17}$$

Let as assume as a counterexample that $\theta = t^a$, $\xi = t^b$ .Then, for the eq. (16) the following relationship should hold for all values of b:

$$z(b) = \frac{1}{b}\frac{\Gamma(1+b)}{\Gamma(1+b-a)} = const \tag{18}$$

However, as can be seen in Fig. 2 this is not true. This finding gives rise to serious doubts concerning the validity of the numerical solution presented in [5]. An additional implication is the fact that the equation considered would in fact require an initial condition of integral type, which is not provided and in experimental situations is hard to obtain.

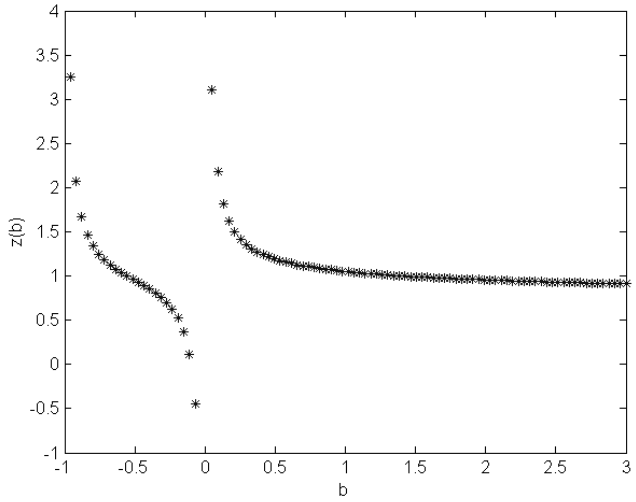


Figure 2: Counterexample demonstrating the inapplicability of the "chain rule", used in eq. (16)

5 FRACTIONAL RICHARDS EQUATION

The derivative of order equal to unity may be defined as follows

$$\frac{d}{ds} f(s) = \lim_{t\to 0} \frac{f(s) - f(s-t)}{t} = -\lim_{t\to 0} \frac{\mathsf{T}(t) f(s) - f(s)}{t} \tag{19}$$

which identifies $-d/dt$ as the infinitesimal generator of time translations defined as

$$\mathsf{T}(t) f(s) = f(s-t) \tag{20}$$

As shown by one of us [7], fractional derivatives arise respectively as the infinitesimal generators of coarse grained time evolutions

$$T_\alpha(t) f(s) = \int_0^\infty \mathsf{T}(u) f(s) h_\alpha\left(\frac{u}{t}\right)\frac{du}{t} \tag{21}$$

where t is considered as a duration of time and therefore it is always positive, and h_α is a one-sided stable law [7]. The order α of the derivative lies between zero and unity, and gives a quantitative measure for the decay of the averaging kernel h_α. The case $\alpha \neq 1$ indicates that memory effects and history dependence may become important.

Taking into consideration the nature of the problem and the initial conditions provided, it is obvious that we need to consider a fractional derivative of type $\beta = 1$, which would result in the following equation

$$D_{0+}^{\alpha,1}\theta(t,x) = \frac{\partial}{\partial x}\left[D_\alpha(\theta)\frac{\partial\theta(t,x)}{\partial x}\right] \tag{22}$$

with initial condition

$$\theta(0,x) = \tilde{\theta}(x) \tag{23}$$

where D_α is the fractional diffusivity and is in general dependent on θ. From this point on we will refer to eq. (22) as the fractional Richards equation.

6 NUMERICAL METHOD

For the solution of eq. (22) we will make use of an Adams-Bashforth-Moulton algorithm introduced by

Diethelm and Freed[16]. Eq. (22) is rewritten as a weakly singular Volterra equation of the second type

$$\theta(t,x)=\theta(0,x)+\frac{1}{\Gamma(\alpha)}\int_0^t (t-u)^{\alpha-1}\frac{\partial}{\partial x}\left(D_\alpha\frac{\partial\theta}{\partial x}\right)du \tag{24}$$

Considering an equidistant mesh

$$\theta_{n+1}(x)=\tilde{\theta}(x)+\frac{1}{\Gamma(\alpha)}\left(\sum_{j=0}^{n}a_{j,n+1}f\left(t_j,x,\theta_j(x)\right)+a_{n+1,n+1}f\left(t_{n+1},x,\theta_{n+1}^P(x)\right)\right) \tag{25}$$

where the predictor $\theta_{n+1}^P(x)$ is evaluated by the relationship

$$\theta_{n+1}^P(x)=\tilde{\theta}(x)+\frac{1}{\Gamma(\alpha)}\sum_{j=0}^{n}b_{j,n+1}f\left(t_j,x,\theta_j(x)\right) \tag{26}$$

and the constants are evaluated as follows

$$a_{j,n+1}=\begin{cases}\dfrac{h^\alpha}{\alpha(\alpha+1)}\left(n^{\alpha+1}-(n-\alpha)(n+1)^\alpha\right) & \text{if } j=0\\ \dfrac{h^\alpha}{\alpha(\alpha+1)}\left((n-j+2)^{\alpha+1}-2(n-j+1)^{\alpha+1}+(n-j)^{\alpha+1}\right) & \text{if } 1\le j\le n\\ \dfrac{h^\alpha}{\alpha(\alpha+1)} & \text{if } j=n+1\end{cases} \tag{27}$$

$$b_{j,n+1}=\frac{h^\alpha}{\alpha}\left((n+1-j)^\alpha-(n-j)^\alpha\right).$$

$f(t,x,\theta)$ signifies $\dfrac{\partial}{\partial x}\left(D_\alpha\dfrac{\partial\theta}{\partial x}\right)$, and for its evaluation finite elements will be used.

7 RESULTS

The fractional Richards equation, namely eq. (22), was solved by means of the numerical method presented in section 6. For the fractional diffusivity function we assumed

$$D_\alpha(\theta)=D_\alpha \tag{28}$$

where $D_\alpha=0.1\,cm^2/min^\alpha$. The initial condition was assumed to be a step function given as

$$\tilde{\theta}(x)=\begin{cases}c_1, & x=0\\ c_2, & x>0\end{cases} \tag{29}$$

where $c_1=0.6$ and $c_2=0.2$.

In Figs. (3a) and (3b) the volumetric moisture content as a function of the distance from the beginning of the soil column is shown for the fractional Richards equation and Richards equation respectively. The curves displayed are isochrones corresponding to times equal to 0, 100, 200, 300, 400, 500, 600, 700, 800, 900 and 1000 min. The lowermost curve corresponds to the initial conditions, whereas the uppermost to time equal to 1000 min. For the case of the fractional Richards equation these results were achieved for order of the fractional derivative equal to 0.9. As can be seen in this case the process appears indeed to be slower than evaluated by the Richards equation and could therefore be termed as sub-diffusive.

In Fig 3(a) we plot x versus t, where x and t solve the equation

$$\theta(t,x)=0.3. \tag{30}$$

Here $\theta(t,x)$ is the solution of eq. (22) for $\alpha = 0.9$, $D_\alpha(\theta)$ given in eq. (28) and initial conditions as in eq. (29).

In Fig. 4(b) the isochrones of volumetric moisture content as a function of x are exhibited for different values of the fractional derivative and time equal to 1000 min. The curve closest to the axis corresponds to α=0.70, whereas the uppermost curve corresponds to the classical Richards equation.

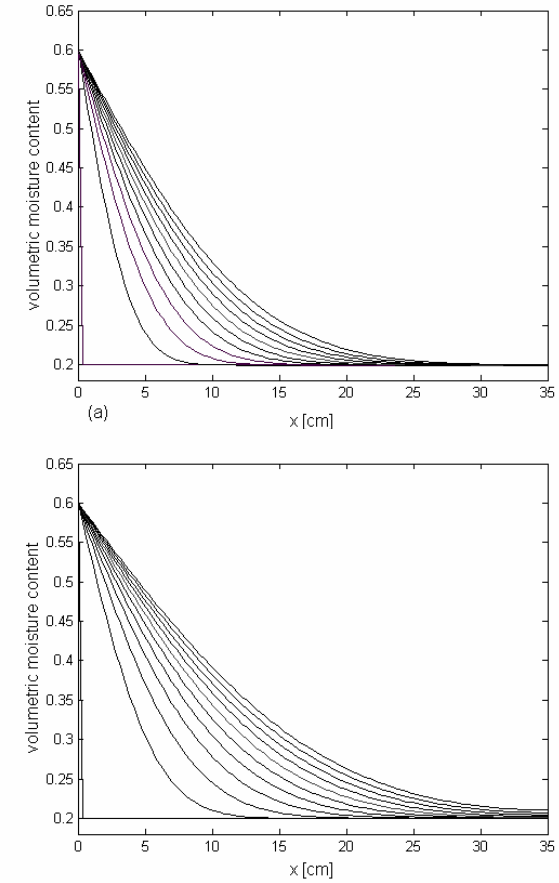


Figure 3: Isochrones of the volumetric moisture content $\theta(t,x)$ as a function of x for t=0, 100, …, 1000 min and initial conditions given by eq. (29). The uppermost curve corresponds to t=1000 min, whereas the one closest to the axis to the initial conditions. (a) Fractional Richards eq. (22) with $D_\alpha(\theta) = 0.1\,cm^2/min^\alpha$; (b) Richards equation eq.(6) with $D(\theta) = 0.1\,cm^2/min$.

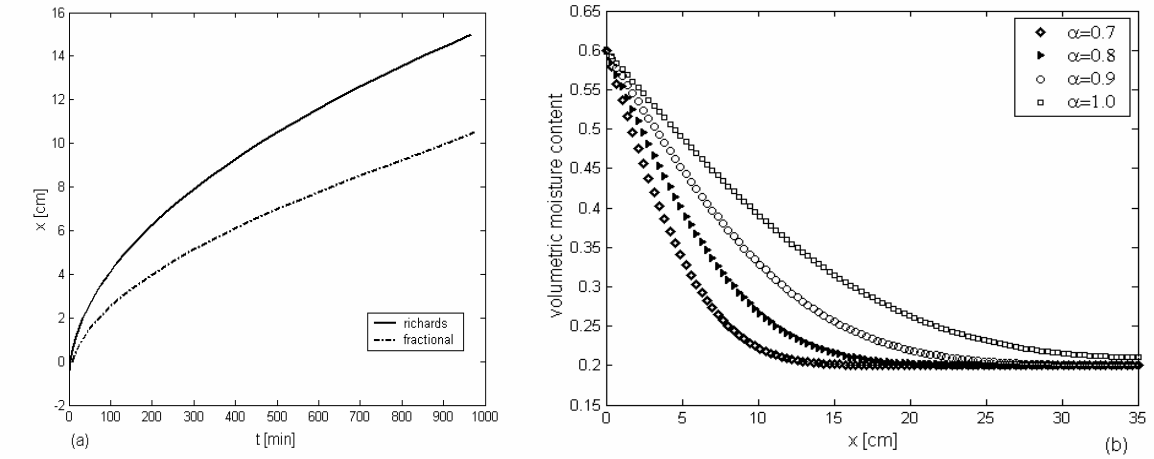


Figure 4: (a) Plot of the positions in time and space at which a volumetric moisture content $\theta(t,x)$ equal to 0.3 was observed for the Richards and the fractional diffusion equation, that is the solution of eq.(30); (b) The

volumetric moisture content $\theta(t,x)$ as a function of x is exhibited for different values of the fractional derivative α, with initial conditions given by eq. (29), $D_\alpha(\theta)$ given in eq. (28) and t equal to 1000 *min*.

8 CONCLUSIONS

The present study has shown that fractional calculus could be used to model the saturation front in partially saturated porous media in cases of subdiffusive behavior. It is important however that the right type of fractional derivative is introduced and the proper initial conditions are considered. Further on, it must be kept in mind that this approach has so far no proven connection neither to continuum mechanics nor to continuous time random walk theory and can be termed as phenomenological.

To further establish the possibility of using fractional derivatives to better model anomalous diffusion behavior of water in porous media, it is important that the relationship of the classical diffusivity to the fractional diffusivity is investigated and that work similar to the above for the case of varying diffusivity is produced, as we encountered numerical instability, when the exponential law, eq. (5), was implemented. It would also be of great interest to investigate the relationship between the movement of water in soil and the continuous time random walk with long-tailed power law distribution of waiting times.

ACKNOWLEDGEMENT

This paper is a partial result of the Project Pythagoras II/ EPEAEK II, supported by the European Union and the Greek Ministry of National Education

REFERENCES

[1] Jacques, D., Šimůnek, J., Timmerman, A., Feyen, J. (2002), "Calibration of Richards' and convection–dispersion equations to field-scale water flow and solute transport under rainfall conditions", Journal of Hydrology 259, No 1-4, pp. 15-31.

[2] Lewandowska, J., Auriault, J.-L. (2004), "Modelling of unsaturated water flow in soils with highly permeable inclusions", Comptes Rendus Mecanique 332, No 1, pp. 91-96.

[3] Van Dam, J.C., Feddes R. A. (2000), "Numerical simulation of infiltration, evaporation and shallow groundwater levels with the Richards equation", Journal of Hydrology 233, No 1-4, pp. 72-85.

[4] Gerolymatou E. (2004), *Simulation of the saturation front in partially saturated porous media,* Diploma Thesis, Faculty of Applied Mathematics and Physics, National Technical University of Athens.

[5] Pachepsky Y., Timlin D., Rawls W. (2003), "Generalized Richard's equation to simulate water transport in unsaturated soils", Journal of Hydrology 272, pp. 3-13.

[6] Ross, B. (Editor) (1975), *Fractional calculus and its applications, Lecture Notes in Mathematics #457,* Springer Verlag.

[7] Hilfer R. (2000), *Applications of fractional calculus in Physics*, World Scientific, Singapore.

[8] Zaslavsky G.M. (2002), "Chaos, fractional kinetics and anomalous transport", Physics Reports 371, pp. 461-580.

[9] Pachepsky Y., Timlin D., Rawls W. (2003), "Errata to generalized Richard's equation to simulate water transport in unsaturated soils", Journal of Hydrology 279, pp. 290.

[10] Philip J. R. (1969). "Theory of infiltration", Advances in Hydroscience 5, pp 215-296.

[11] Hilfer R. (1998), "Macroscopic Equations of Motion for Two Phase Flow in Porous Media", Physical Review E 58, pp 2090.

[12] Hilfer R. and Besserer H. (2000), "Macroscopic Two Phase Flow in Porous Media", Physica B 279, pp 125.

[13] Hillel D. (1980), *Applications of soil Physics*, Academic Press, New York.

[14] Liouville J. (1832), "Mémoire sur quelques questions de géométrie et de mécanique, et sur un nouveau genre de calcul pour résoudre ces questions", Journal de l'École Polytechnique 13, pp 1-66.

[15] Hilfer R. (2000), "Fractional diffusion based on Riemann-Liouville fractional derivatives", Jour. Phys. Chem. B 104, pp 3914-3917.

[16] Diethelm K. and Freed A. (1999), "On the solution of nonlinear fractional order differential equations used in the modeling of viscoplasticity", Scientific Computing in Chemical Engineering II—Computational fluid dynamics. In: F. Keil, W. Mackens, H. Voß and J. Werther, Editors, *Reaction Engineering, and Molecular properties*, Springer, Heidelberg, pp. 217–224.

Inverse Problems in Mechanics

AN APPLICATION OF THE VISCO-ELASTIC CONTINUOUS MACRO-ELEMENTS TO FAULT DETECTION AND LOCALIZATION IN RESPONSIBLE PARTS OF MACHINES AND STRUCTURES

Tomasz Szolc

Institute of Fundamental Technological Research
of the Polish Academy of Sciences
ul. Świętokrzyska 21, 00-049 Warsaw, Poland
e-mail: tszolc@ippt.gov.pl, web page: http://www.ippt.gov.pl

Keywords: Discrete-continuous models, Vibration and wave effects, Fault detection, Numerical simulation.

Abstract. *In the paper discrete-continuous mechanical models built by means of the one-dimensional visco-elastic continuous structural macro-elements are applied for dynamic investigations of mechanical systems and structures. The coupled vibration and wave effects are used for fault detection and localization in the most important components of the considered objects, i.e. in shafts, axles, beams, rods cantilevers and others.*

1 INTRODUCTION

The commonly observed fast development of machines operating under great material loadings requires more and more precise fault detection and localization of their most heavily affected and responsible parts. The dynamic methods for fault detection and localization based on the vibratory[1] and wave[2] approaches are very popular now. A practical application of these methods realized in the form of on-line dynamic diagnostics or test-rig-exciting of real objects must be supported by proper theoretical investigations carried out by means of several analytical or numerical methods and mechanical models. Currently, the finite element method belongs on the one hand to the most advanced and reliable, but on the other hand this method is not free of numerous drawbacks and disadvantages. In this paper there is proposed an alternative method of modeling and dynamic analysis of mechanical systems described in details in [3] and applied e.g. in [4,5]. This method is based on discrete-continuous models built by means of the finite structural elements with continuously distributed inertial-visco-elastic properties, called further the visco-elastic continuous macro-elements (V-ECM). The main purpose of the research carried out in this paper is a fault detection and localization in responsible parts of machines and structures by means of dynamic analyses performed using the discrete-continuous mechanical models.

2 GENERAL ASSUMPTIONS FOR THE DISCRETE-CONTINUOUS MODELS BUILT BY MEANS OF THE V-ECM

The most heavily affected and responsible components of a broad class of machines and mechanisms are shafts, axles, rods, cantilevers, guideways and others. For dynamic analyses they can be usually regarded as one-dimensional media. Here, each segment of such real component of constant or almost constant cross-section can be substituted in the model by the structural macro-element of continuously distributed inertial-visco-elastic properties of the length l_i, cross-sectional area A_i and polar and diametral moments of inertia J_{0i} and I_i, i=1,2,…,n, as shown in Figure 1, where n denotes the entire number of such macro-elements in the assumed model.

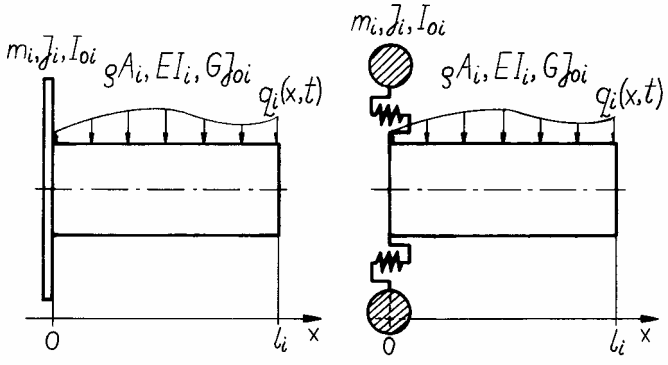


Figure 1. The visco-elastic continuous macro-elements (V-ECM)

In order to represent inertial properties of the components remaining almost undeformed during studied dynamic processes, such as rotors, impellers, coupling disks, gears, flywheels and others, to the extreme cross section of the V-ECM the rigid disk of mass m_i and diametral and polar mass moments of inertia J_i and I_{0i} can be attached directly or by means of the massless elastic membrane. The defined in this way macro-elements can be applied for investigations of bending, torsional and longitudinal vibrations, which usually are the most important in majority of machines and mechanical systems. Torsional and longitudinal motion of cross-sections of the V-ECM is described by the following partial differential equations

$$\Gamma\left(1+\tau\frac{\partial}{\partial t}\right)\frac{\partial^2\theta_i(x,t)}{\partial x^2}-\rho\frac{\partial^2\theta_i(x,t)}{\partial t^2}=q_i(x,t), \tag{1}$$

where in the i-th macro-element $\theta_i(x,t)$ denotes the angular or longitudinal displacement of the cross-section of the co-ordinate x, the symbol Γ is equal to the Kirchhoff modulus G in the case of torsional vibrations or to the Young modulus E in the case of longitudinal vibrations and ρ, τ denote the material density and the retardation time of material damping, respectively. The function $q_i(x,t)$ describes the external excitation continuously distributed along the macro-element. Flexural motion of cross-sections of the V-ECM can be described by the partial differential equations derived using various beam theories, e.g. the Bernoulli-Euler, Rayleigh or the Timoshenko one. For example, motion of the rotating with the speed Ω beam macro-element described by the Rayleigh beam theory is governed by the following equation

$$EI_i\left(1+\tau\frac{\partial}{\partial t}\right)\frac{\partial^4 v_i(x,t)}{\partial x^4}-\rho I_i\left(\frac{\partial^4 v_i(x,t)}{\partial x^2\partial t^2}-2j\Omega\frac{\partial^3 v_i(x,t)}{\partial x^2\partial t}\right)+\rho A_i\frac{\partial^2 v_i(x,t)}{\partial t^2}+$$
$$+c_{si}\frac{\partial v_i(x,t)}{\partial t}+k_{si}v_i(x,t)=q_i(x,t), \tag{2}$$

where $v_i(x,t)$ denotes the complex transverse displacement in two perpendicular planes and k_{si}, c_{si} are respectively the stiffness and damping coefficient of the visco-elastic continuous foundation. These macro-elements are mutually connected directly according to the structure of the real object or coupled with each other by means of discrete oscillators which can also describe the assumed imperfections in the considered systems in the form of cracks, bearing support misalignments, local rotor-shaft anisotropies or rigid coupling misalignments. In the discrete-continuous models such connections are expressed in the form of geometric and dynamic boundary conditions for the motion equations, e.g. (1) and (2). The geometric boundary conditions describe displacement identities of the extreme cross-sections of the directly connected adjacent macro-elements. The dynamic boundary conditions are equations of equilibrium of the inertial, elastic, damping, external, gyroscopic, support reaction forces and moments formulated for the directly connected adjacent macro-elements as well as for the macro-elements mutually connected by means of discrete oscillators.

3 MATHEMATICAL FORMULATION AND SOLUTION OF THE PROBLEM

The mathematical formulation of the problem investigated by the use of the assumed discrete-continuous models is based on the local analytical solutions of the partial differential equations of motion of the V-ECM cross-sections. In the paper the Fourier solution in the form of infinite series in orthogonal eigenfunctions is used, which leads to the following system of ordinary differential equations in the modal co-ordinates[3,4,5]

$$\mathbf{M}(t)\ddot{\mathbf{r}}(t)+\mathbf{C}(\Omega,t,\dot{r}(t),r(t))\dot{\mathbf{r}}(t)+\mathbf{K}(t,r(t))\mathbf{r}(t)=\mathbf{F}(t,\Omega^2,\dot{r}(t),r(t)). \tag{3}$$

These equations can be uncoupled in the linear case or mutually coupled by the taken into consideration nonlinear, parametric and skew-symmetrical terms contained in the variable or response-dependent mass, damping-gyroscopic and stiffness matrices **M**, **C** and **K**, respectively. In (3) $\mathbf{r}(t)$ denotes the vector of unknown modal co-ordinates $r(t)$ and **F** is the time- and response-dependent external excitation vector. Moreover, the d'Alembert solution in the form of travelling longitudinal and shear elastic waves has been applied leading to the mixed systems of algebraic and ordinary differential equations with a retarded argument[3]. Numerical integration of both mentioned above types of equations enables us to obtain dynamic responses of the considered mechanical systems or structures with assumed imperfections. It is to emphasize that the computational algorithms built on the basis of mathematical relations derived in a consequence of the applied Fourier and d'Alembert solution are characterized, in a comparison with the traditional methods, by very strong numerical efficiency and stability – so important for simulations being performed many times, in particular, when the fault detection and localization procedure reduces to an inverse problem investigation.

4 EXAMPLES OF APPLICATION

In the computational examples there is considered the multi-bearing rotor-shaft system with a transverse crack and with a local shaft anisotropy associated by a bearing misalignment. For this system the vibratory approach for fault detection is applied. Here, by means of numerical simulations of coupled lateral-torsional-axial shaft vibrations the cause-symptom relationships have been collected in order to detect and localize the imperfection during operation monitoring of the real object. The wave approach is applied in the case of cracked cantilever rod and beam, where the determined reflected waves inform us about a fault existence and position.

4.1 The rotor-shaft system with the cracked shaft

The discrete-continuous model of the rotor-shaft system with the transverse crack and with the coupling disk misalignment is presented in Fig. 2. The investigations have been performed for the rotor-shaft system with two identical bladed disks supported on three journal bearings. Each bearing is represented by means of the dynamic oscillator of two degrees of freedom, where beyond the oil-film interaction also visco-elastic properties of the bearing housing and foundation are taken into consideration. With a reasonable for practical purposes accuracy the bladed disks can be substituted in the discrete-continuous model by the system of dynamic oscillators in the form of rigid rings mutually attached to the rotor-shaft by means of the visco-elastic mass-less membranes enabling rotations of these rings as well as their translational displacements in the shaft axial direction, as shown in Fig. 2. Parameters of these oscillators have been determined by the use of the proper modal reduction method. In the considered rotor-shaft system the transverse crack of depth a/D=0.3 in the shaft segment between the second bearing #2 and the first bladed disk was assumed, which is demonstrated in Fig. 2. An additional local shaft flexibility caused by the crack is represented by the mass-less spring connecting the adjacent shaft segments, coupling shaft lateral motion with torsional and axial one as well as realizing crack "breathing" process. The 6×6 stiffness matrix of this spring has been determined similarly as in [1], i.e. using the fundamentals of fracture mechanics.

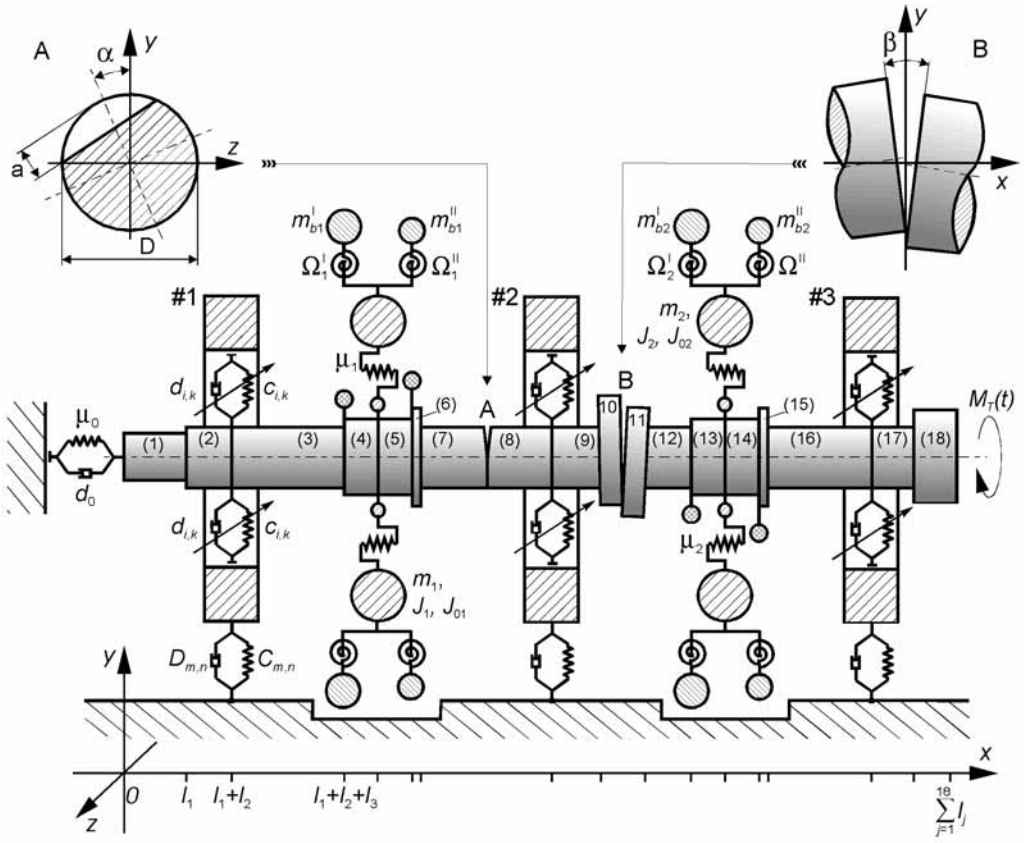
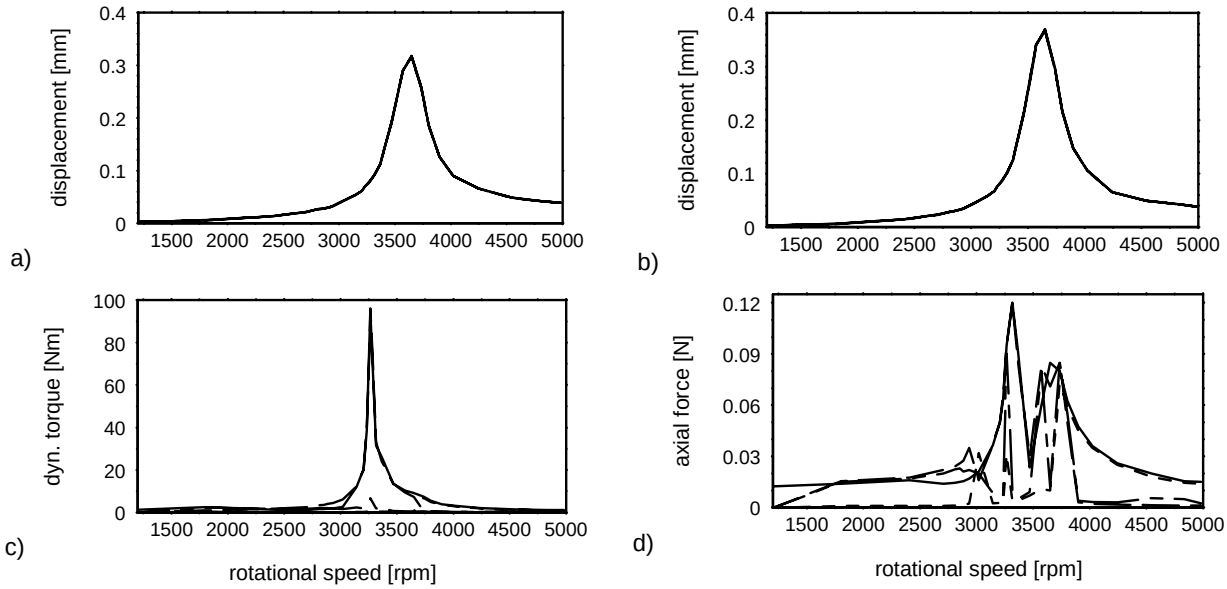


Figure 2. Discrete-continuous model of the rotor-shaft system with the transverse crack and with the coupling disk misalignment

The system non-linear dynamic response has been obtained by solving (3) in the range of system rotational speeds 1200÷5000 rpm for four circumferential crack positions on the shaft α_p=0, 90, 180 and 270 deg in order to investigate a severity of coupling effects of the rotor-shaft bending-torsional-axial vibrations indicating the presence of the crack. For each considered rotational speed and the gravitational load acting on the considered

system the proper mean values of the oil film stiffness and damping coefficients have been determined. In this way the bearing anisotropic and non-symmetric properties were introduced. The only assumed source of dynamic external excitation are the static residual unbalances of the bladed disks mutually shifted by the phase angle Δ=180 deg. Thus, the torsional-axial vibrations can be regarded here as an output effect caused by bending vibrations of the rotor-shaft. The quantities of particular interest in these investigations are bending vibration displacements of the shaft at the bladed disk locations, the dynamic torque transmitted by the shaft between the bladed disks and the thrust bearing axial force. The results of simulations are presented in Fig. 3 in the form of amplitude characteristics of the listed quantities expressed as functions of the constant rotational speed values. The amplitude of the transverse shaft displacements is defined as one half of the longer diameter of the ellipse featuring the disk center orbit. The amplitudes of the remaining quantities are regarded as maximum fluctuation values with respect of their average values of the steady state dynamic response. In Figs. 3a,b there are shown the transverse displacement amplitude characteristics for the 1st and 2nd disk, respectively. Both plots are similar to each other and they are characterized by one resonance peak at the rotational speed ca. 3648 rpm corresponding to the system 2nd bending natural frequency 60.8 Hz determined by means of the Rayleigh- and confirmed by the Timoshenko-beam theory. It is to remark that respective plots obtained for four considered crack position angles α_p almost overlay, which means that in the considered system the circumferential crack location does not influence the rotor-shaft bending vibrations. Fig. 3c presents the amplitude characteristics of the dynamic torque induced by the shaft bending vibrations. These plots are also characterized by one resonance peak at the rotational speed ca. 3264 rpm corresponding to the system 1st torsional natural frequency 54.4 Hz. It is to remark that the not presented in a graphical form time histories of this torque oscillate with the fundamental synchronous frequency $\Omega/2\pi$. This means that the coupling between shaft bending vibrations and torsional-axial vibrations of the shaft-bladed disks system indicates the greatest magnitude in the conditions of ordinary parametric resonance of the first order, in a contradistinction to such coupling studied in [4] caused by residual disk unbalance only. At this juncture, the combined parametric resonance is observed and the dynamic torque fluctuates with double-synchronous frequency Ω/π. In the considered case the system axial response in the form of thrust bearing force indicates the parametric ordinary resonance of the frequency 54.4 Hz and the ordinary "bending" resonance of the frequency 60.8 Hz, which follows from the respective peaks of the amplitude characteristic shown in Fig. 3d. The time histories of this force are characterized on the one hand by very small extreme values, but on the other hand by much richer frequency spectra including synchronous, double-synchronous and higher-frequency components of comparable magnitudes. From the respective plots in Fig. 3c it follows that the torsional-axial responses of the shaft-bladed disks system are particularly severe for the crack location angles α_p=0 and 90 deg, when the crack remained all of the time continuously open. Then, one obtains the extreme values which in the case of dynamic torque are important not only from the diagnostic viewpoint. However, the analogous plots in Fig. 3c obtained for α_p=180 and 270 deg have no importance. This result confirms rightness of the commonly applied assumption that for practical purposes in the field a crack can be treated as a permanently open crack.



(———) for $\alpha_p=0^0$, (— — —) for $\alpha_p=90^0$, (– · – · –) for $\alpha_p=180^0$, (- - - - -) for $\alpha_p=270^0$

Figure 3. Amplitude characteristics of the system coupled dynamic responses

4.2 The rotor-shaft system with the local shaft anisotropy and with the bearing misalignment

Analogous amplitude characteristics of the coupled bending-torsional-axial dynamic responses have been obtained for the similar double-span rotor shaft system characterized by the local shaft anisotropy associated with the bearing misalignment. The discrete-continuous model of this system is presented in Fig. 4. In order to collect

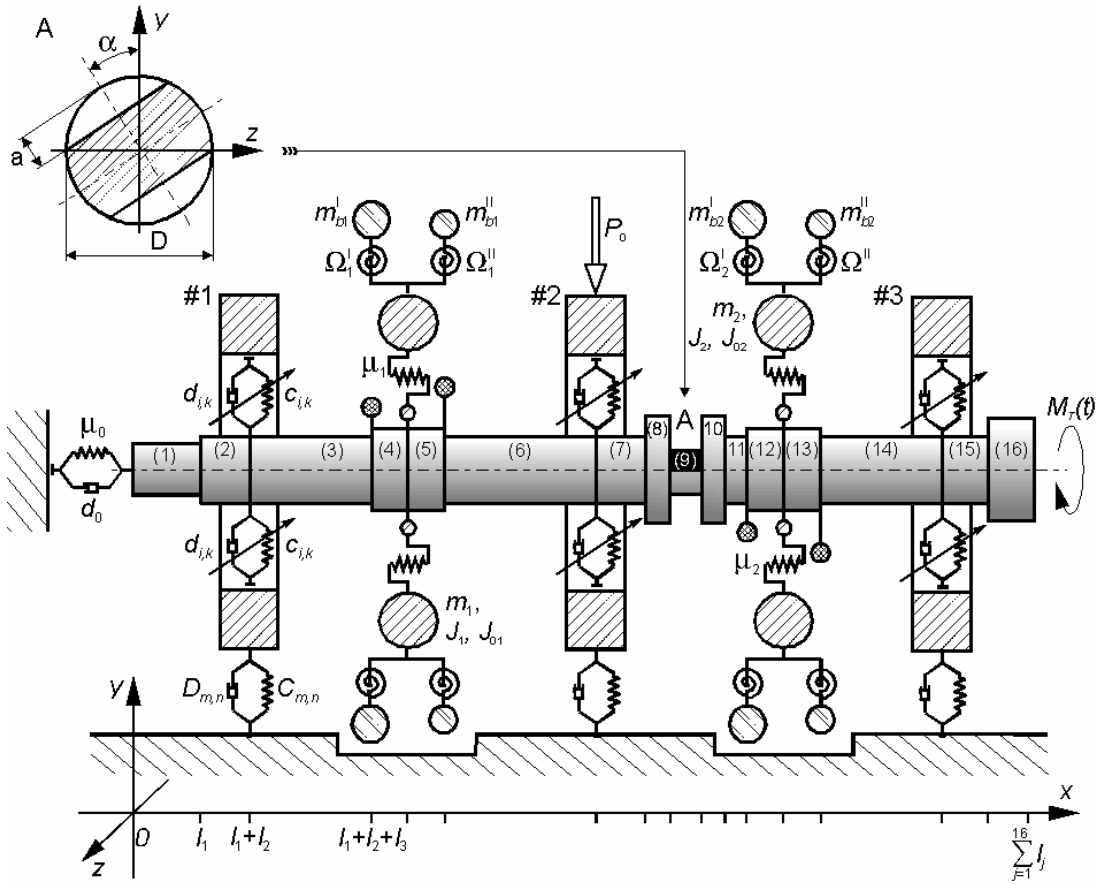


Figure 4. Discrete-continuous model of the rotor-shaft system with the local shaft anisotropy and with the bearing misalignment

for the two mentioned rotating systems with the assumed imperfections the cause-symptom relationships for fault detection and localization during real object monitoring, it is necessary to perform a qualitative analysis of these amplitude characteristics. For the transverse forces in bearings #1, #2, #3, the dynamic torques in the shaft between the bladed disks (T2) and between the second bladed disk and the right-hand shaft end (T1), the blade rim retarding transverse forces acting on the disks (D1), (D2) and for the axial force in the thrust bearing in Figs. 5 and 6 there are shown results of the FFT analysis obtained for two exemplary rotational speeds n=1740 and 2200 rpm of the rotating system with the local shaft anisotropy and the bearing misalignment. In these figures there are plotted response amplitude components in the domain of frequency related to the synchronous frequency X. Each amplitude peak excited by natural multiple of the synchronous frequency X corresponds to the induced bending (B) or torsional-axial (T-A) eigenform of the investigated mechanical system. For the four mentioned above studied quantities in Table 1 such relationships between the excitation frequency and the number of induced eigenform have been collected for the most important amplitude peaks at four selected rotational speed values n=1740, 2200, 2610 and 3012 rpm. In the all cases of rotational speeds the resonance phenomena are observed. For n=1740 rpm the parametric resonance of the coupled bending-torsional-axial vibrations occurs yielding the response components of the excitation frequency 4X. At n=2610 rpm the ordinary resonance is observed with the external excitation frequency 2X and at n=3012 rpm an increase of response amplitudes has a character of the parametric resonance with the excitation frequency components 1X, 2X and 3X. The case of n=2200 rpm can be regarded as the "mixed" one, in which the bending response is induced by the ordinary resonances of the external excitation frequency 2X and the torsional-axial response is characterized by the induced components of frequencies 1.4X, 3X and 4X.

The collected in this way excitation frequency – induced eigenform relationships for greater number of rotational speeds within the exploitation range of the rotor-machine inform us about possible coupling effects between the bending, torsional and axial vibrations caused by the assumed fault, e.g. the crack, shaft anisotropy,

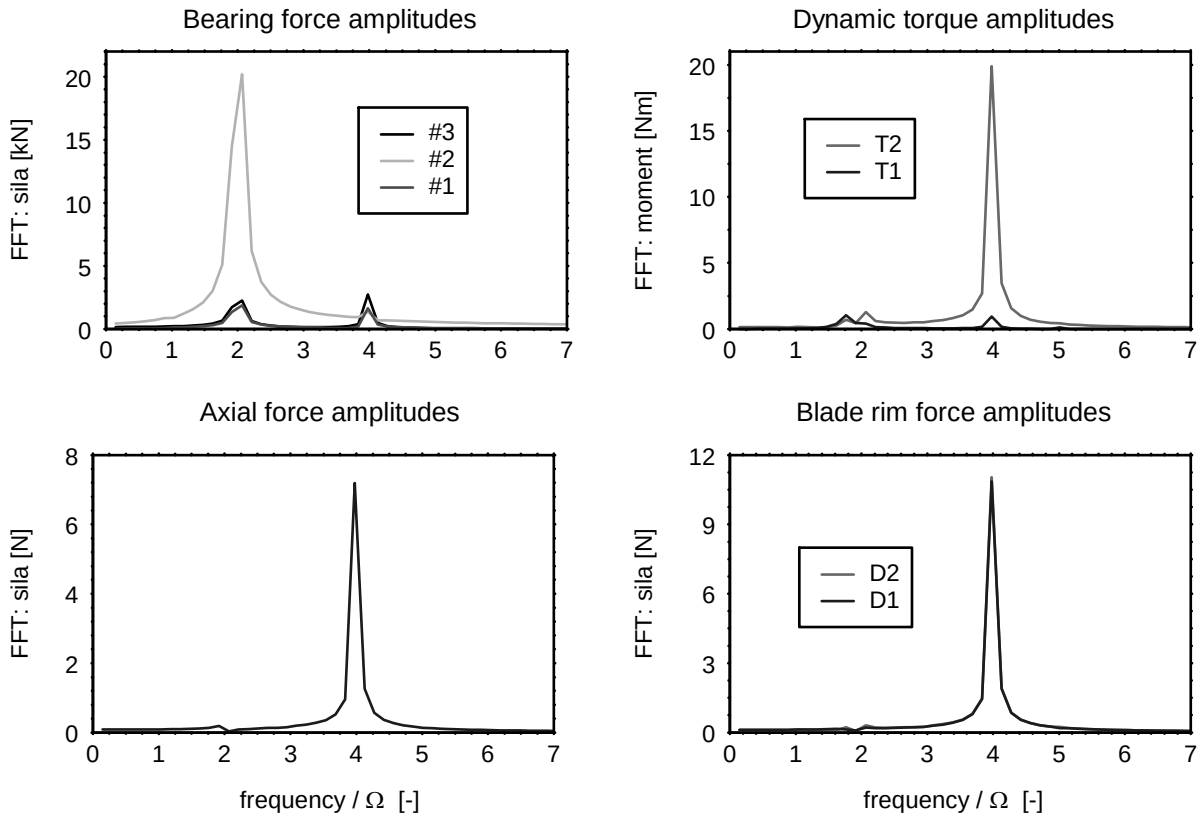


Figure 5. Results of the FFT analysis of the system dynamic response for n=1740 rpm

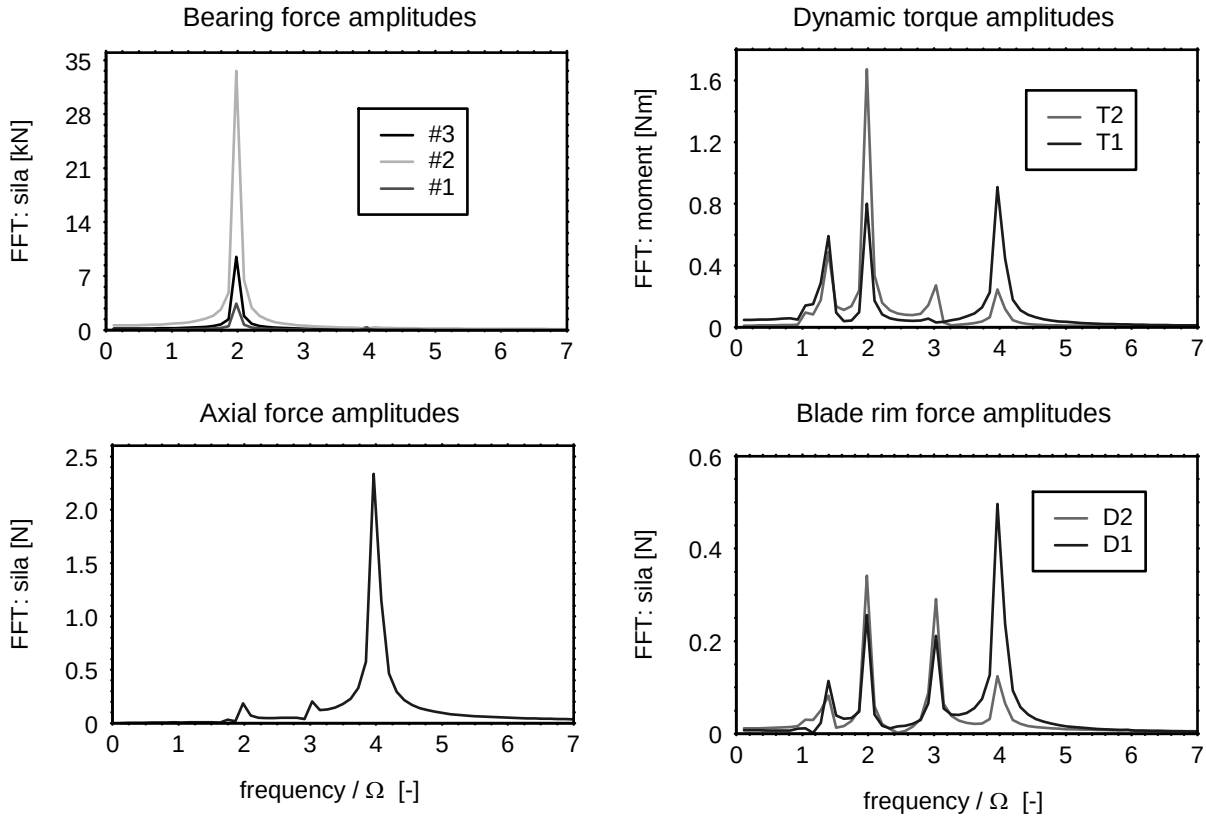


Figure 6. Results of the FFT analysis of the system dynamic response for n=2200 rpm

bearing or coupling misalignment and others. Then, proper comparisons of these relationships with the analogous monitored on-line responses of the real object enable us to detect and localize the given type of imperfection.

Rotational speed n [rpm]	**1740**	**2200**	**2610**	**3012**
Synchronous freq. X [Hz]	**29.0**	**36.667**	**43.5**	**50.2**
Bearing forces	2X→1 B 4X→5 T-A	2X→5 B	2X→7 B	2X→ excitation
Dynamic torques in the shaft	4X→5 T-A	~1.4X→1 T-A 2X→ excitation 3X→5 T-A 4X→ 7 T-A	2X→3 T-A	1X→1 T-A
Blade rim transverse retarding force	4X→5 T-A	~1.4X→1 T-A 2X→ excitation 3X→5 T-A 4X→7 T-A	2X→3 T-A	1X→1 T-A 2X→ 4 T-A
Axial force in the thrust bearing	4X→5 T-A	4X→7 T-A	2X→3 T-A	2X→4 T-A 3X→7,8 T-A

Table 1 : Cause-symptom relationships due to the imperfections in the rotor-shaft system

4.3 Crack detection in the cantilever rod and beam using the wave approach

The transverse cracks have been also detected and localized in the cantilever rod and in the cantilever beam by means of the analysis of traveling elastic waves propagating and reflecting in these objects due to the diagnostic high-frequency impulses exciting respectively longitudinal waves in the rod and shear waves in the beam. The cracked cantilever rod and beam is represented by two longitudinally or transversally deformable visco-elastic continuous macro-elements (1) and (2), as shown in Fig. 7. Motion of their cross-sections is described by the homogeneous partial differential equations (1) for $\Gamma=E$ in the case of the longitudinal wave analysis in the rod and for $\Gamma=sG$ in the case of the shear wave investigation in the Timoshenko beam of the cross-sectional shear ratio s. The equations of motion (1) are solved with boundary conditions describing the support properties, external loading as well as the presence of the crack. The introduced by this crack an additional longitudinal or transverse flexibility is represented by the mass-less spring of stiffness k_0 determined by means of the Castigliano theorem using the fundamentals of fracture mechanics, in an analogous way as for the cracked rotor-shaft discussed in Chapter 4.1. The diagnostic excitation force $P(t)$ has been applied in the form of the triangular impulse modulated by the high-frequency sinusoid, i.e. in the identical way as in [6], where the similar problem was considered by means of the spectral finite element method.

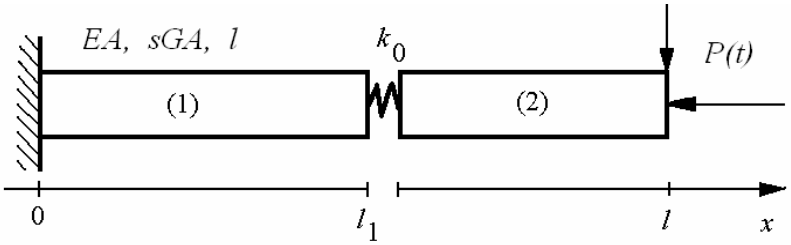


Figure 7. Discrete-continuous model of the cracked cantilever rod or beam

For the rod and beam model presented in Fig. 7 the Fourier and d'Alembert solutions of the motion equations (1) can be applied. The former, in the considered here linear case, leads to the infinite set of uncoupled modal equations (3). The latter leads to the above mentioned set of algebraic and ordinary differential equations with the retarded space-time argument z in the following form

$$g_2(z) = -f_2(z-2l_2), \quad Kg_1'(z) + g_1(z) = Kf_1'(z-2l_1) - f_1(z-2l_1) + f_2(z-l_1) + g_2(z-l_1),$$
$$f_2'(z)=f_1'(z-l_1), \quad f_1'(z)=F(z) + g_1'(z), \tag{4}$$

where the functions f_i and g_i, i=1,2, represent longitudinal or shear waves propagating along the x-axis respectively in positive and negative sense, Fig. 7, $K=\Gamma A/(k_0 l_s)$, $F(z)=P(t)l_s/(\Gamma A)$, $l_2=l-l_1$, and l_s [m] is the reference distance. An integration of the appropriately truncated number of the modal equations (3) or sequential solving of equations (4) yield in time domain system dynamic responses in the form of travelling elastic longitudinal or shear waves. For sufficiently small integration step both solutions lead to almost identical results.

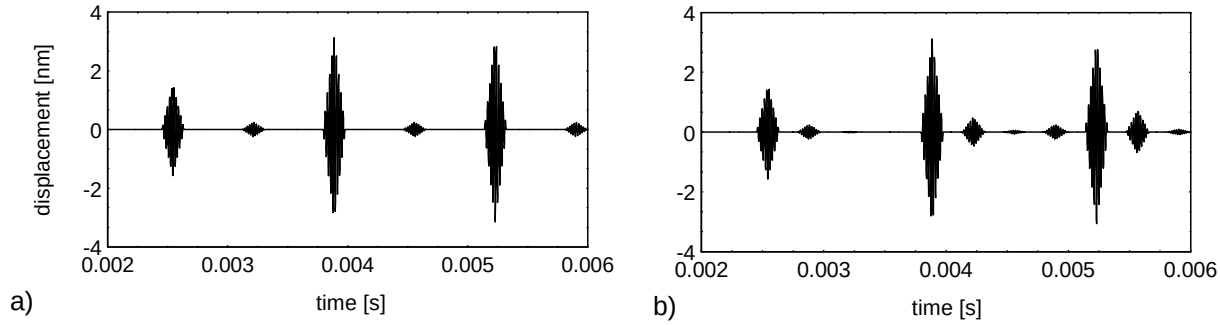


Figure 8. Transverse displacement of the free end of the cantilever beam with a crack in the middle (a) of its entire length and in the cross-section distant 75% of the entire length from the clamped end (b)

In Fig. 8 there are presented exemplary simulation results of shear wave propagation in the form of transverse displacements of the cantilever cracked beam, where the crack has been assumed in the cross-section distant respectively one half of the beam entire length l, i.e. for $x=l_1=0.5l$, Fig. 8a, and 75% of the beam length, i.e. for $x=l_1=0.75l$, from its clamped end, Fig. 8b. In these figures several "great" displacement peaks of the beam free end are observed, where the "first" peak corresponds directly to the action of the external force $P(t)$ and the "next" two "great" peaks are caused by successive reflections of the shear waves from the clamped end of the beam. "Between" the "great" peaks in Figs. 8a,b several additional "small" peaks occur. These "small" peaks appear regularly after successive wave reflections from the beam cross-section, in which the crack has been assumed. Here, the simulated effect of wave reflections as well as the time delays, after which the waves reflected from the crack are recorded, are the most essential. An appearance of these additional reflected waves informs us about an existence of the imperfection as well as the known value of wave propagation velocity multiplied by the proper time delay enables us to localize this imperfection in the considered continuous medium.

5 CLOSING REMARKS

In the paper the rod- and beam visco-elastic continuous macro-elements (V-ECM) have been used to build structural discrete-continuous physical models of mechanical systems and structures applied for fault detection and localization in the rotor-shafts, cantilever rods and beams. The continuous distribution of inertial-visco-elastic properties of the V-ECM enables us a fault detection by means of the vibratory and the wave approach. The vibratory approach for fault detection applied here to the rotor machine for a determination of cause-symptom relationships necessary for dynamic diagnostic of the real object is on the one hand rather labour-consuming and not very sensitive to small defects. But on the other hand, this method can be applied to complex mechanical systems and structures for an analysis of parametric and non-linear coupled vibrations. However, the wave approach for fault detection and localization reducing to investigation of travelling elastic wave propagation and reflections in time domain is very simple in applications and much more sensitive to small defects than the vibratory one. But on the presented level, the wave method can be applied to relatively simple models of elements of machines and structures. Nevertheless, in the case of both approaches the proposed here method of discrete-continuous modeling of mechanical systems and structures based on the V-ECM is a reliable and computationally very effective tool for dynamic analysis and fault detection in numerous cases of practical applications and theoretical studies.

REFERENCES

[1] Papadopoulos, C. A., Dimarogonas, A. D. (1987), "Coupling of bending and torsional vibration of a cracked Timoshenko shaft", Ingenieur-Archiv 57, pp. 257-266.

[2] Doyle, J. F. (1997), *Wave propagation in structures*, Springer-Verlag, New York.

[3] Szolc, T. (2003), "Dynamic analysis of complex, discrete-continuous mechanical systems", Polish Academy of Sciences – IFTR Reports 2/2003, *Habilitational dissertation (in Polish).*

[4] Szolc, T. (2000), "On the discrete-continuous modeling of rotor systems for the analysis of coupled lateral-torsional vibrations", Int. J. of Rotating Machinery 6, 2, pp. 135-149.

[5] Szolc, T., Kiciński, J., Rządkowski, R. (2003), "Analytical-numerical approach for non-linear dynamic modelling of the bearing-rotor-shaft systems to design and fault diagnostics", *Proc. of the International ISCORMA-2 Conference, Max-Media Poland by BPBC USA*, Gdańsk, Poland, 4-8 August, pp.528-538.

[6] Krawczuk, M., Palacz, M., Ostachowicz, W. (2003), "The dynamic analysis of a cracked Timoshenko beam by the spectral element method", Journal of Sound and Vibration 264, pp. 1139-1153.

RISK ASSESSMENT OF STRUCTURES USING NEURAL NETWORKS

Michalis Fragiadakis, Nikos D. Lagaros, and Manolis Papadrakakis

Institute of Structural Analysis & Seismic Research
National Technical University of Athens
Zografou Campus, Athens 15780, Greece
e-mail: {mfrag, nlagaros, mpapadra}@central.ntua.gr

Keywords: Safety Analysis, Neural Networks, Seismic Risk, Reliability.

Abstract. *Earthquake-resistant design of structures using probabilistic analysis is an emerging field in structural engineering. This paper examines the application of Neural Networks (NN) to the probabilistic analysis of real-world structural systems under seismic loading. Probabilistic analysis is carried out using the Monte Carlo Simulation (MCS) method incorporating the improved Latin Hypercube Sampling technique. Limit-state fragilities for steel moment resisting frames are determined by means of nonlinear time history analysis with a suite of ground motion records. The use of NN is motivated by the large number of time consuming nonlinear time history analyses required for MCS. The Rprop algorithm is implemented for training the NN utilizing available information extracted from each record. The trained NN is used to compute the level of damage which is expressed as maximum interstorey drift values, thus leading to a close prediction of the limit state fragility curves.*

1 INTRODUCTION

Extreme earthquake events may produce extensive damage to structural systems. It is therefore essential to establish a reliable procedure for assessing the seismic risk of real-world structural systems. Probabilistic analysis provides the framework to model the various sources of uncertainty that may influence structural performance under seismic loading conditions. Probabilistic analysis is performed in order to obtain fragility curves, which provide a measure of the safety margin of a structural system for different limit states.

The theory and the methods of structural reliability have been developed significantly during the last twenty years and are documented in a large number of publications[1]. In this work the probabilistic safety analysis of framed structures under seismic loading conditions is investigated. Randomness of ground motion excitation (that influences seismic demand) and of material properties (that affect structural capacity) are taken into consideration using Monte Carlo Simulation. The capacity assessment of steel frames is determined using nonlinear timehistory analysis. The probabilistic safety analysis using Monte-Carlo Simulation and nonlinear time history analysis results in a computationally intensive problem. In order to reduce the excessive computational cost, techniques based on Neural Networks (NN) are implemented. For the training of the NN a number of Intensity Measures (IMs) are derived from each earthquake record, for the prediction of the level of damage, which is measured by means of maximum interstorey drift values $\theta_{\max}$.

2 FRAGILITY ANALYSIS USING MONTE CARLO SIMULATION

The seismic fragility of a structure $F_R(x)$ is defined as its limit-state probability, conditioned on a specific peak ground acceleration, spectral velocity, or other control variable that is consistent with the specification of seismic hazard

$$F_R(x) = P[LS_i / PGA \geq x] \tag{1}$$

where LS_i represents the corresponding ith limit state and the peak ground PGA is the control variable. If the annual probabilities of exceedance $P[PGA \geq x]$ of specific levels of earthquake motion are known, then the mean annual frequency of exceedance of the ith limit state is calculated as follows:

$$P[LS_i] = \int F_R(x) P[PGA \geq x]\, dx \tag{2}$$

Equation (2) can be used to make decisions about, for example, the adequacy of a design or the need to retrofit. In the present study we seek the fragility $F_R(x)$. Once the fragility is calculated the extension to eq. (2) is straightforward.

Often $F_R(x)$ is modelled with a lognormal probability distribution, which leads to an analytic calculation. In the present study Monte Carlo Simulation (MCS) with improved Latin Hypercube Sampling (LHS) is adopted for the numerical calculation of $F_R(x)$. Numerical calculation of eq. (1) provides a more reliable estimate of the limit state probability, since it is not necessary to assume that seismic data follow a lognormal distribution. However, in order to calculate the limit state probability, a large number of nonlinear dynamic analyses are required for each hazard level, especially when small probabilities are sought.

The proposed methodology requires that MCS has to be performed at each hazard level. Earthquake records are scaled to a common intensity level that corresponds to the hazard level examined. Scaling is performed using the first mode spectral acceleration of the 5% damped spectrum ($Sa(T_1,5\%)$). Therefore, all records are scaled in order to represent the same ground motion intensity in terms of $Sa(T_1,5\%)$. Earthquake loading is considered as two separate sources of uncertainty, ground motion intensity and the details of ground motion. The first uncertainty refers to the general severity of shaking at a site, which may be measured in terms of any IM such as PGA, $Sa(T_1,5\%)$, Arias intensity, etc. The second source refers to the fact that, although different acceleration time histories can have their amplitudes scaled to a common intensity, there is still uncertainty in the performance, since IMs are imperfect indicators of the structural response. The first source is considered by scaling all records to the same intensity level at each limit state. The second source is treated by selecting natural records as random variables from a relatively large suite of scenario based records. The concept of considering separately seismic intensity and the details of ground is the backbone of the Incremental Dynamic Analysis method[2], while Porter et al.[3] have also introduced intensity and different records as two separate uncertain parameters in order to evaluate the sensitivity of structural response to different uncertainties.

The random parameters considered in this study are the material properties and more specifically the modulus of elasticity E and the yield stress f_y, as well as and the details of ground motion where a suite of scenario based earthquake records is used. The material properties are assumed to follow the normal distribution while the uniform distribution is assumed for the records in order to select them randomly from a relatively large bin of natural records. The first two variables are sampled by means of the iLHS technique in order to increase the efficiency of the sampling process.

In reliability analysis the MCS method is often employed when the analytical solution is not attainable and the failure domain can not be expressed or approximated by an analytical form. This is mainly the case in problems of complex nature with a large number of basic variables where all other reliability analysis methods are not applicable. Expressing the limit state function as $\mathbf{G}(\mathbf{x})<0$, where $\mathbf{x}=(x_1,x_2,..,x_M)^T$ is the vector of the random variables, the probability of exceedance can be obtained as

$$P_{LS} = \int_{G(x)\geq 0} f_x(x)dx \tag{3}$$

where $f_x(\mathbf{x})$ denotes the joint probability of failure for all random variables. Since MCS is based on the theory of large numbers (N_∞) an unbiased estimator of the probability of failure is given by

$$P_{LS} = \frac{1}{N_\infty}\sum_{j=1}^{N_\infty} I(x_j) \tag{4}$$

where $I(\mathbf{x}_j)$ is a Boolean vector indicating "successful" and "unsuccessful" simulations. In order to estimate P_{LS} an adequate number of N_{sim} independent random samples is produced using a specific probability density function for the vector $\mathbf{x}$. The value of the failure function is computed for each random sample x_j and the Monte Carlo estimation of P_{LS} is given in terms of the sample mean by

$$P_{LS} \cong \frac{N_H}{N_{sim}} \tag{5}$$

where N_H is the number of simulations where the maximum interstorey drift value exceeds a threshold drift for the limit state examined. In order to calculate eq. (5) N_{sim} nonlinear time history analyses have to be performed at each hazard level. Clearly the computational cost of performing so many nonlinear dynamic analyses is prohibitive. In order to reduce the computational cost, properly trained Neural Networks are implemented.

3 FEED-FORWARD NEURAL NETWORKS

A feed-forward neural network consists of a number of units linked together and attempts to create a desired

mapping between the inputs and the target of the training set. The training set is a set of input-target pairs $\mathbf{D} = \{\mathbf{x}^m, \mathbf{t}^m\}$, where m is the number of the pairs and $\mathbf{x}^m$, $\mathbf{t}^m$ are the input and target vectors, respectively. A neural network architecture $\mathcal{A}$ consists of a specific number of layers, a number of units in each layer and a type of activation functions. If a set of values $\mathbf{w}$ is assigned to the connections of the network a mapping $\mathbf{y}(\mathbf{x}; \mathbf{w}, \mathcal{A})$ is defined from the inputs $\mathbf{x}$ to the outputs $\mathbf{y}$. The quality of this mapping, with respect to the training set, is measured by an error function:

$$E_D(D|\mathbf{w},\mathcal{A}) = \sum_m \frac{1}{2}\left(\mathbf{y}(\mathbf{x}^m;\mathbf{w},\mathcal{A}) - \mathbf{t}^m\right)^2 \tag{6}$$

A learning algorithm tries to determine the values of $\mathbf{w}$, in order to achieve the right response for each input vector applied to the network minimizing the value of E_D. The numerical minimization algorithms used for the training generate a sequence of weight parameters $\mathbf{w}$ through an iterative procedure. To apply an algorithmic operator $\mathcal{O}$ we need the starting weight parameters $\mathbf{w}^{(0)}$, while the iteration formula can be written as follows

$$\mathbf{w}^{(t+1)} = \mathcal{O}(\mathbf{w}^{(t)}) = \mathbf{w}^{(t)} + \Delta\mathbf{w}^{(t)} \tag{7}$$

The changing part of the algorithm $\Delta\mathbf{w}^{(t)}$ is further decomposed into two parts as

$$\Delta\mathbf{w}^{(t)} = a_t \mathbf{d}^{(t)} \tag{8}$$

where $\mathbf{d}^{(t)}$ is a desired search direction vector of the move and a_t the step size in that direction.

Regularization

In the error function E_D various modifications have been applied, like the addition of the momentum term and the inclusion of noise in the learning process. In this work an extra regularizing term $E_W(\mathbf{w})$ is added to the E_D, which penalizes the large values of the weights in order to achieve a smoother or simpler mapping:

$$E_W(\mathbf{w},\mathcal{A}) = \sum_i \sum_j \frac{1}{2} w_{ij}^2 \tag{9}$$

the E_W is called weight energy term, and the error function to be minimized becomes:

$$E = \alpha E_W(\mathbf{w},\mathcal{A}) + \beta E_D(D|\mathbf{w},\mathcal{A}) \tag{10}$$

The constant α is called regularizing constant and should not be confused with the momentum term. The two constant α and β are determined using the following two rules[4]:

$$\begin{aligned} 2\alpha E_W &= \gamma \\ 2\beta E_D &= N - \gamma \end{aligned} \tag{11}$$

with

$$\gamma = \sum_i^k \frac{\lambda_i}{\lambda_i + i} \tag{12}$$

where λ_i are the eigenvalues of the quadratic form βE_D, N is the number of output units times the number of the training pairs and k is the total number of the weight parameters.

Learning algorithm

Learning algorithms can be divided to two categories. (i) Algorithms that use global knowledge of the state of the entire network, such as the direction of the overall weight update vector, which are referred as *global* techniques. In the conventional back-propagation learning algorithm the gradient descent algorithm is used belonging to the global learning algorithms. (ii) *Local* adaptation strategies based on weight specific information only such as the temporal behaviour of the partial derivative of the corresponding weight. The local approach is more closely related to the neural network concept of distributed processing in which computations can be made independent to each other. Furthermore, it appears that for many applications local strategies achieve faster and more reliable prediction than global techniques despite the fact that they use less information[5]. In this work the

Rprop[6] algorithm has been adopted since it has been proved very efficient[7].

The *Rprop* learning algorithm with locally adaptive learning rates is based on an adaptive version of the Manhattan-learning rule and has been developed by Riedmiller and Braun[6]. The abbreviation *Rprop* algorithm stands for the Resilient backpropagation algorithm. The weight updates can be written as

$$\Delta w_{ij}^{(t)} = -\eta_{ij}^{(t)} \operatorname{sgn}\left(\frac{\partial E_t}{\partial w_{ij}}\right) \tag{13}$$

with

$$\eta_{ij}^{(t)} = \begin{cases} \min(\alpha \cdot \eta_{ij}^{(t-1)}, \eta_{max}), & \text{if } \dfrac{\partial E_t}{\partial w_{ij}} \cdot \dfrac{\partial E_{t-1}}{\partial w_{ij}} > 0 \\ \max(b \cdot \eta_{ij}^{(t-1)}, \eta_{min}), & \text{if } \dfrac{\partial E_t}{\partial w_{ij}} \cdot \dfrac{\partial E_{t-1}}{\partial w_{ij}} < 0 \\ \eta_{ij}^{(t-1)}, & \text{otherwise} \end{cases} \tag{14}$$

where α=1.2, b= 0.5, η_{max}=50 and η_{min}=0.1[7]. The learning rates are bounded by upper and lower limits in order to avoid oscillations and arithmetic underflow. It is interesting to note that, in contrast to other algorithms, *Rprop* employs information about the sign and not the magnitude of the gradient components.

4 PREDICTIONS OF THE SEISMIC RESPONSE USING NEURAL NETWORKS

As already mentioned feed-forward Neural Networks are used in order to reduce the number or earthquake simulations required for the calculation of the probability of eq. (5). The principal advantage of a properly trained NN is that it requires a trivial computational effort to produce an acceptable approximate solution. Such approximations appear to be valuable in situations where actual response computations are CPU intensive and quick estimations are required. Neural Networks have been applied in the past by Papadrakakis et al.[8] in order to calculate the probability of failure for steel moment frames using inelastic static analysis. In recent studies NN have been adopted for the reliability analysis of structures by Nie and Ellingwood[9] and Hurtado[10]. However, in the present study the NN are implemented in order to predict the maximum seismic response with natural earthquake records replacing the time consuming nonlinear time history analysis. The NN are trained in order to predict the maximum interstorey drift θ_{max} for different earthquake records which are identified by NN using a set of Intensity Measures (IM).

Table 1 : Intensity measures

No	Intensity Measure
1	PGA (g)
2	PGV (m)
3	PGD (m)
4	V/A (sec)
5	Arias intensity (m/sec)
6	Significant duration (5 to 95 % of Arias) (sec)
7	RMS acceleration (g)
8	Characteristic Intensity
9	CAV
10	Spectral Intensity
11	Total Duration (sec)
12	$Sa(T_1)$ (g)
13	$Sv(T_1)$ (cm)
14	SaC, c=2 (g)
15	SaC, c=3 (g)

The term Intensity Measure is used to denote a number of common ground motion parameters which represent the amplitude, the frequency content, the duration or any other ground motion parameter. A number of different IMs has been presented the literature[11], while various attempts to relate an IM with a damage measure such as maximum interstorey drift values exist[12]. The IMs adopted can be classified as structure-independent

(e.g. PGA, Arias Intensity) or as both structure and record dependent (e.g. SaT_1). The complete list of the IMs used in this study is given in Table 1.

It can be seen that the IMs selected, vary from widely used ground motion parameters such as peak ground acceleration (*PGA*) to more sophisticated measures such as *SaC*. The definitions and further discussion on the first thirteen measures of Table 1 is given by Kramer[11]. The last two IMs refer to the measure proposed by Cordova et al.[13] which is defined as:

$$SaC = Sa(T_1)\sqrt{\frac{Sa(c \cdot T_1)}{Sa(T_1)}} \tag{15}$$

The parameter c takes the value 2 and 3 for the 14th and the 15th parameter of Table 1, respectively. These IMs were introduced in order to assist the NN to capture the effects of inelasticity by considering the elastic spectrum at an "effective" period longer than T_1, thus reflecting the reduction in stiffness.

For each hazard level separate training of the NN is performed by means of the above IMs. The training process is based on the fact that the trained NN will assign small weights to the IMs which have poor correlation with the damage measure selected. Instead of using the whole set it was examined the suitability of using only some of the IMs of Table 1. The parametric study was performed for various intensity levels since the performance of an IM depends also on the level of nonlinearity that the structure has undergone. The 10 combinations of IMs, shown in Table 2, were compared.

Table 2: Intensity measures combinations

ID	IM combinations
A	1
B	1,2
C	1,2,3
D	1,2,3,5
E	1,2,3,5,9
F	1,2,3,5,9,10
G	1,2,3,5,9,10,12
H	1,2,3,5,9,10,12,14
I	1,2,3,5,9,10,12,14,15
J	ALL

Table 3: Prediction errors (%) on the maximum interstorey drift θ_{max}

	IM Combination									
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**ALL**
					PGA = 0.05g					
MAX	49.7	39.6	19.6	32.6	15.9	14.8	4.9	27.0	23.6	9.0
MIN	4.4	0.2	0.9	0.3	0.3	0.6	1.0	1.8	0.6	0.3
MEDIAN	29.1	9.9	4.8	5.2	5.9	5.5	2.9	5.3	4.7	4.0
					PGA = 0.27g					
MAX	32.6	28.2	21.4	26.6	46.7	23.6	9.5	24.5	35.9	9.6
MIN	0.9	0.1	1.1	0.5	1.4	0.1	0.6	0.7	0.3	1.5
MEDIAN	16.9	16.1	9.8	17.2	9.0	7.2	4.9	7.2	4.4	4.4
					PGA = 0.56g					
MAX	62.0	67.2	28.8	42.2	35.4	28.0	33.2	29.3	16.4	9.2
MIN	6.7	3.2	0.2	0.6	0.1	0.2	0.7	1.3	2.2	0.8
MEDIAN	18.5	20.4	9.7	12.6	19.2	15.5	9.0	9.2	7.1	4.3
					PGA = 0.90g					
MAX	72.1	45.2	51.0	23.3	13.3	16.9	12.0	8.7	12.5	9.2
MIN	3.0	6.1	1.4	0.5	1.2	0.9	0.5	0.6	0.2	0.9
MEDIAN	36.5	15.2	15.7	3.0	2.8	7.8	1.8	3.8	2.9	3.5

The performance of each combination is shown in Table 3. The efficiency of the NN is evaluated for the ten-storey steel moment resisting frame described in the next section. For this parametric study the material random variables were considered with their mean values. The efficiency is measured by means of the error on the

prediction of θ_{max}, obtained by means of eq. (10). From Table 3 it is clear that the use of record dependent only measures, such as *PGA*, lead to increased error values, while more refined measures help to reduce the error considerably. The use of the complete set of IMs in Table 1 is preferred since it performed equally well for all four hazard levels examined in the parametric study.

5 NEURAL NETWORKS BASED FRAGILITY ASSESMENT

A suite of 95 scenario-based natural records were used in this study. All records correspond to relatively large magnitudes of 6.0–6.9 and moderate distances, all recorded on firm soil and bearing no marks of directivity. In order to obtain the fragility curves, sixteen hazard levels expressed in PGA terms ranging from 0.05g to 1.25g were used. For each hazard level, risk assessment is performed and five limit state fragilities are calculated. Each limit state is defined by means of a corresponding maximum interstorey drift θ_{max} value. In the present study the five limit states considered range from serviceability, to life safety and finally to the onset of collapse. The corresponding θ_{max} threshold values range from 0.2 to 6 percent.

The test example considered to demonstrate the efficiency of the proposed procedure is the five-bay, ten-storey moment resisting plane frame of Figure 1. The mean values of the modulus of elasticity is equal to 210GPa and the yield stress is f_y=235MPa. The coefficients of variation for E and f_y are considered as 5% and 10%, respectively, while both variables are assumed to follow the normal distribution. The constitutive law is bilinear with a strain hardening ratio of 0.01, while the frame is assumed to have rigid connections and fixed supports. The permanent load is equal to 5kN/m^2 and the live load is taken as Q = 2kN/m^2. The gravity loads are contributed from an effective area of 5m. All analyses were performed using a force-based fiber beam-column element[14] that allows the use of a single element per member, while the same material properties are used for all the members of the frame. Geometric nonlinearities were taken into consideration. Therefore, the FE model allows collapse to take place only due to second order effects, which however are not so pronounced for the frame considered in this study.

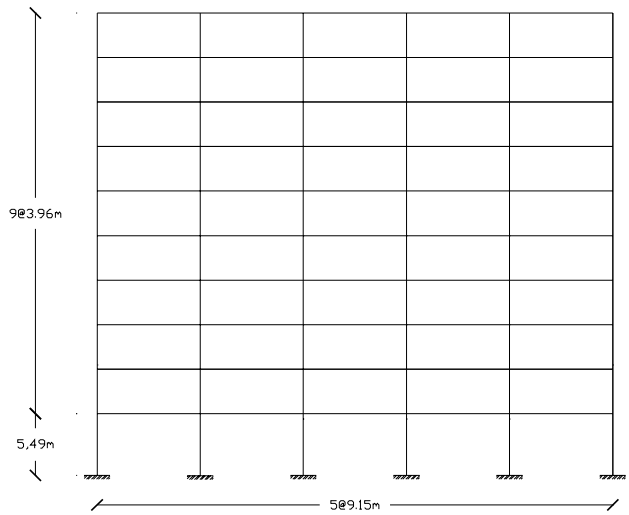


Figure 1. Ten-storey steel moment frame

For training the NN both training and testing sets have to be selected for each hazard level. The selection of the sets is based on the requirement that the full range of possible results has to be taken into account in the training step. Therefore, training/testing triads of the material properties and the records are randomly generated using the Latin Hypercube sampling. In the case of earthquake records the selection has to take into account that the scaling factor should be between 0.2 and 5. This restriction is applied because large scaling factors are likely to produce unrealistic earthquake ground motions. Furthermore, the records selected for generating the training set have to cover the whole range of structural damage for the hazard level in consideration. Thus, nonlinear time history analyses were performed, for mean E and f_y values, where the θ_{max} values of each record that satisfy the previous requirement were determined for each hazard level. In total 30 records are selected for generating the training set of each hazard level taking into account that the selection has to cover the whole range of θ_{max} values. Therefore, training sets with 90 triads of E, f_y and record number, all sampled as discussed above, are generated. Finally, a testing sample of 10 triads is also selected in a similar way in order to test the performance of the NN.

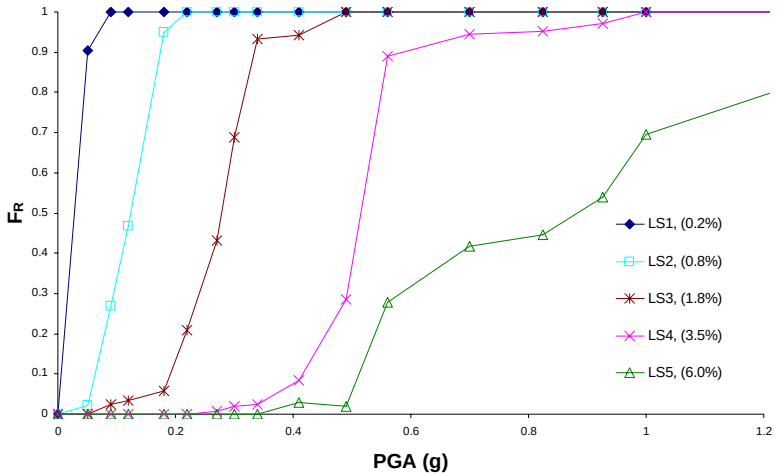


Figure 2. Fragility curves

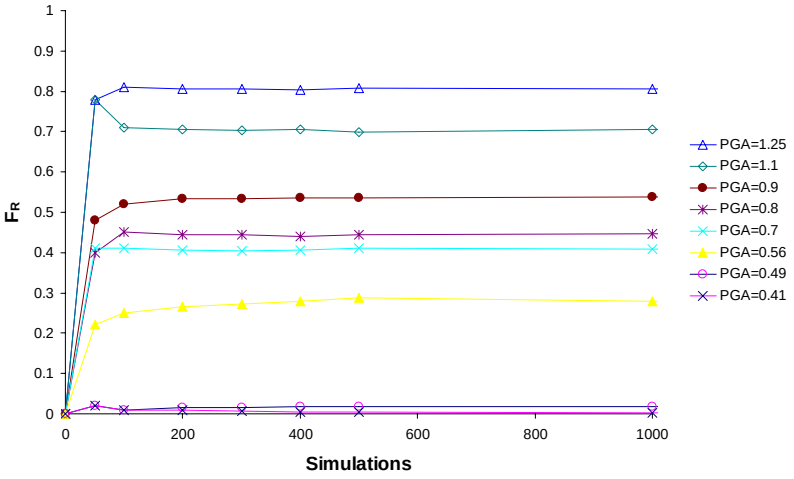


Figure 3. Number of NN-Simulations required (Near collapse limit state, $\theta_{max} \geq 6.0\%$)

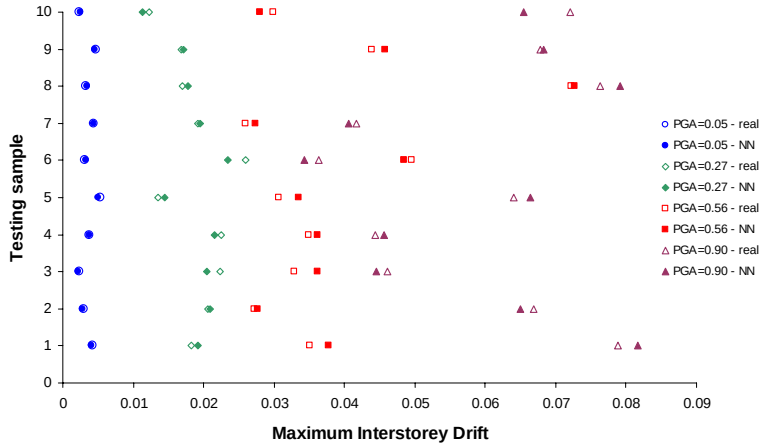


Figure 4. Prediction of θ_{max} for the testing sample

The fragility curves obtained for the five limit states considered are shown in Figure 2. Figure 3 shows the number of MCS simulations required for the fragility curve of a one limit state, in particular the Near Collapse limit state (θmax≥6.0%). It can be seen that depending on the calculated probability of exceedance the number simulations required for a single point of the fragility curve, ranges from 50 to 1000. The validity of the prediction obtained with the NN is shown in Figure 4. The maximum interstorey drift values predicted for the 10 components of the testing set compared to the values obtained with nonlinear timehistory analysis are shown in

Figure 4 for four hazard levels. Figure 4, gives the impression that better prediction is obtained for the lower intensity levels, however this occurs because the horizontal axis of Figure 4 corresponds to θ_{max} values and not on the relative error on θ_{max}, which practically remains constant for all four intensity levels.

6 CONCLUSIONS

A very efficient procedure for the fragility analysis of structures based on properly trained Neural Networks is presented. The NNs are trained by means of a set of intensity measures that can be easily extracted from the earthquake records. The proposed methodology allows the use of Monte Carlo simulation for the calculation of the limit state fragility curves, where the only simplifying assumptions made are the distributions of the uncertain parameters. The proposed formulation may be more complicated compared to other simplified approaches, however it offers a different approach to an emerging problem in earthquake engineering leading to reduction of the computational cost. The results obtained once combined with regional hazard curves can be directly applied to the performance-based design of steel frames.

REFERENCES

[1] Schueller, G.I. (ed.) (2005), "Computational Stochastic Structural Mechanics and Analysis as well as Structural Reliability", *Comput. Methods Appl. Mech. Engrg. - Special Issue* 194(12-16), pp. 1251-1795.

[2] Vamvatsikos, D., Cornell, C.A. (2002), "Incremental dynamic analysis", Earthquake Engineering and Structural Dynamics 31, pp. 491-514.

[3] Porter, K.A., Beck, J.L., and Shaikhutdinov, R.V. (2002), "Sensitivity of Building Loss Estimates to Major Uncertain Variables", Earthquake Spectra 18, pp. 719-743.

[4] MacKay, D.J.C. (1992), "A practical Bayesian framework for backprop networks", Neural Computation 4, pp. 448-472.

[5] Schiffmann, W., Joost, M., Werner, R. (1993), *Optimization of the back-propagation algorithm for training multi-layer perceptrons*, Technical report, University of Koblenz, Institute of Physics.

[6] Riedmiller, M., Braun, H. (1993), "A direct adaptive method for faster back-propagation learning: the RPROP algorithm", *Proceedings of the IEEE International Conference on Neural Networks (ICNN)*, San Francisco, pp. 586–591.

[7] Riedmiller, M. (1994), "Advanced supervised learning in multi-layer perceptrons - from backpropagation to adaptive learning algorithms", Int. Journal of Computer Standards and Interfaces - Special Issue on Neural Networks 16, pp. 265 - 278.

[8] Papadrakakis, M., Papadopoulos, V., and Lagaros, N.D. (1996), "Structural Reliability analysis of elastic-plastic structures using neural networks and Monte Carlo simulation", Computer Methods in Applied Mechanics and Engineering 136, pp. 145-163.

[9] Nie, J., Ellingwood, B.R. (2004), "A new directional simulation method for system reliability. Part II: application of neural networks", Probabilistic Engineering Mechanics 19(4), pp. 437-447.

[10] Hurtado, J.E. (2001), "Neural network in stochastic mechanics", Archives of Computational Methods in Engineering (State of the art reviews) 8(3), pp 303-342.

[11] Kramer, S.L. (1996), *Geotechnical Earthquake Engineering*, Prentice-Hall, Englewood Cliffs, NJ.

[12] Shome, N., Cornell, C.A. (1999), "Probabilistic seismic demand analysis of non-linear structures" *Report No. RMS-35, RMS Program*, Stanford University, Stanford.

[13] Cordova, P.P., Deirlein, G.G., Mehanny, S.S., and Cornell, C.A. (2000), "Development of a two-parameter seismic intensity measure and probabilistic assessment procedure", *Proceedings of the 2nd U.S.-Japan Workshop on Performance-Based Earthquake Engineering of Reinforced Concrete Building Structures*, Saporo, Japan, 11-13 September.

[14] Papaioannou, I., Fragiadakis, M., and Papadrakakis, M. (2005), "Inelastic Analysis of Framed Structures using the Fiber Approach", *Proceedings of the 5th International Congress on Computational Mechanics (GRACM 05)*, Limassol, Cyprus, 29 June – 1 July.

STRUCTURAL HEALTH MONITORING OF COMPOSITE PATCH REPAIRS USING EMBEDDED FIBER BRAGG GRATING SENSORS AND NEURAL NETWORK TECHNIQUES: DAMAGE IDENTIFICATION

G.J. Tsamasphyros, N.K. Fournarakis , and G.N. Kanderakis

National Technical University of Athens
Dept of Applied Mathematics and Physical Sciences
Section of Mechanics
Athens, Greece
E-mail: nfurnar@central.ntua.gr

ABSTRACT

Modern developments in the space of aeronautics, particularly since the incident of 11th of September 2002 in the U.S.A. which affected negatively the economic potential of air companies worldwide, render explicit the need for exploitation of air structures up to their designed structural life or even further and consequently the need for adoption of maintenance techniques supporting their structural integrity in an acceptable, according to their designed criteria, manner. One of the most promising techniques is the so called "composite patch repair method", a method capable to repair cracked metallic structures using adhesively bonded composite patches, that constitutes an effective method for re-establishment of structural integrity. However, even though this method has important advantages, problems concerning the long-term behavior of the adhesive, the appearance of debonds, etc. appear and have led to the considerable delayed widespread acceptance of this method.

In the present study the realization and implementation of elements of a structural health monitoring system for composite patch repairs is being performed, combining elements from the theory of structural repair mechanics, the theory of intelligent materials and structures and particularly the technology of optical sensors and neural networks.

The inverse problem of fault detection, qualification / quantification and life expectancies are studied using neural network techniques. Various network learning algorithms are evaluated for the corresponding failure cases and comparison results are developed. The data taken from the experimental test series (part I of the study) are used as network exemplars for training/test reasons and the corresponding deviations are calculated as well as the network performance metrics.

INTRODUCTION

Current economic world conditions are forcing to the operation of all structural types, especially for complex, expensive and multidisciplinary structures, beyond their design life. This requirement has led to a maximization effort of the performance vs. cost ratio, in order to achieve the optimum Return of Investment. In that respect, alike the philosophy for human body monitoring using periodical check ups, modern structural elements do require the same attention as it comes to their structural integrity. Current trends are dealing with on line health monitoring systems capable to trace the external stimulus and the subsequent failures this stimulus can lead to. These trends are expressed mainly in the aeronautic technological field due to the requirement of high performance – load budget operation of air structures. The modern systems expressing this structural health monitoring (SHM) philosophy are split in two main categories [1]:

- Monitoring using load spectrum: these systems are capable to estimate the cumulative damage due to fatigue loading of a structure using elements of load spectrum, operational usage and field experience. These systems are based on prior knowledge of analytical/numerical models of the structure and are not capable to identify an existing failure but are limited to damage prognosis.

- Health monitoring systems: these systems are capable to trace a number of predetermined field variants and usually are embedded on the structure under consideration. Most of the times these systems tend to be integrated, using sensing, adaptation and damage logic, or semi-integrated, having only sensing and damage logic.

Moreover, the level of integration of the system defines the model of sensing and damage logic required for a specific application. For example a system designed to monitor the health of a large scale structural element (a bridge) should use global approach techniques using distributed sensing elements, due to the fact that a point-

wise system resolution is not required. On the other hand, for the case of monitoring a crack propagating on the skin of an aircraft, where high strain resolution is required, a local sensing technique should be preferred combined with the proper damage logic.

The damage logic itself is expressed by the two following techniques [2]:

- Logic based on analytical/numerical models that apply to the specific structure or a family of identical structures, having predetermined field distributions for each discriminated fault signature.

- Logic based on decoupled from the structure itself numerical techniques, that do not take into account the geometry or the constituents of the structure but are based on existing load experience and field mapping. Such techniques are based on statistical models, on neural networks, on genetic algorithms or any combination of these.

According to the level of detail required by a health monitoring and damage identification system, four levels of logic implementation can be used [3]:

- Level 1: Damage verification and identification
- Level 2: Damage localization
- Level 3: Damage quantification
- Level 4: Life expectancy

In the present work, the use of neural networks and genetic algorithms is studied with specific application in a composite patch adhesively bonded to a crack metallic structure, in order to implement a structural integrity logic (four levels of implementation) concept in terms of health monitoring of the repair. In that respect, various learning algorithms are examined and compared using the fiber optic behavior of [4].

LEARNING ALGORITHM

In order to design the best solutions available for each level of implementation, two main network types were used: a classification network and a function approximation network. Each network was used for a specific level of implementation. Thus, for levels 1 and 2 a classification network was used while for levels 3 and 4 a function approximation multilayer perceptron network was used. For further information on these types of networks, the reader is advised to refer to [5] and [6].

In order to secure the best possible results in terms of output error, convergence to a solution, etc. the following five network learning algorithms [5] were examined and verified:

- Momentum
- Step
- Delta-Bar-Delta
- Quick Prop
- Conjugate Gradient

A portion of the data set used as a learning/testing/validation set for the network performance, is provided in Table 1.

Specimen	Sensor A	Sensor B	Crack Length
B	0.455	0.324	10
B	0.461	0.323	11
B	0.482	0.322	12
B	0.501	0.321	14
B	0.535	0.320	13
B	0.650	0.340	15

Table 1: Input data set for network learning algorithm evaluation

The data set was consisted of 400 exemplars (data points). The network used for the algorithm evaluation was consisted of 10 to 30 hidden processing elements and the test was performed with 1000 to 5000 epochs (iterations). The network architecture used is presented in figure 1.

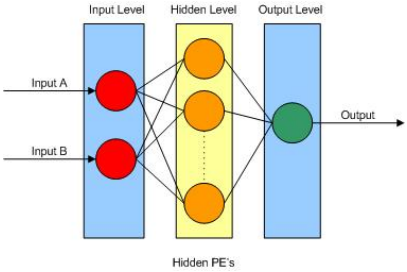


Figure 1: Network Architecture

The results of the test for each specific algorithm in terms of learning and test error are presented graphically in figure 2, where as *L* is denoted the learning error whereas *T* is the testing error, for the range of epochs and PE's.

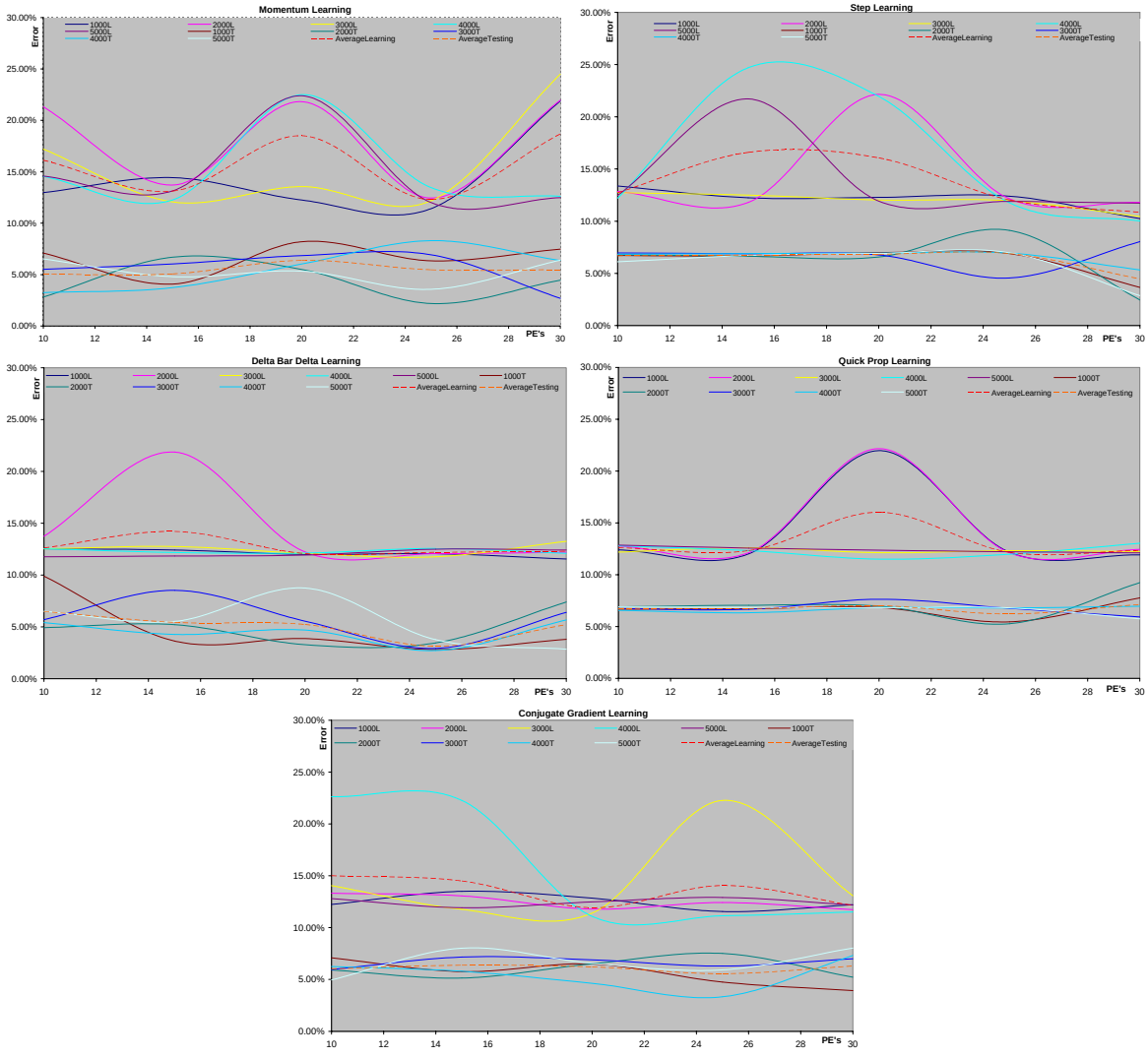


Figure 2: Learning and test error vs number of PE's

From the above results, it was obvious that the best algorithm available, in terms of minimum learning and testing error was the momentum learning algorithm. On the contrary, the most stable and fast convergent algorithm was QProp and Step. Considering that:

- the data to be used are generally a small set, the required convergence would occur soon
- the data variance is negligible

the momentum algorithm was chosen for the developed network algorithms.

SHM LEVEL1

For the implementation of the capability to identify and verify fault that occurred in a composite patch repair, due to the lack of experimental data from [4], the learning data set used was produced using the finite element method, for the model presented on figure 3.

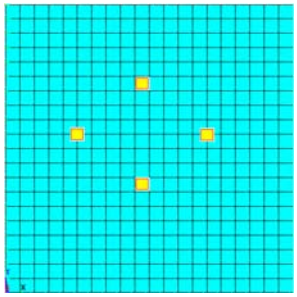


Figure 3: The structure FE model and the assumed sensor positions

Three types of failure were model as typical failures that can occur in a composite patch, which are presented in figure 4:

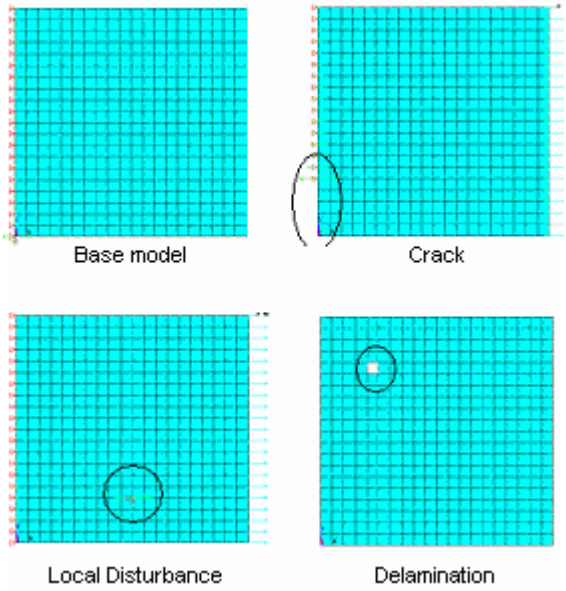


Figure 4: Modeled failure types

The data set was consisted of 500 exemplars (data points). The network used for the algorithm evaluation was consisted of 10 to 30 hidden processing elements and the test was performed with 1000 to 5000 epochs (iterations). Using the figure 3 strain measurements as inputs, for the various load cases, for a typical classification multi layer perceptron network using the momentum learning algorithm, the following results were achieved for the considered network output:

Network	Epochs	Hidden PE's	Test Error	MSE	NMSE	r
1-MOM05-1000	1000	5	14.14%	0.488	0.816	0.441
1-MOM10-1000	1000	10	14.68%	0.502	0.841	0.434
1-MOM15-1000	1000	15	13.50%	0.5	0.838	0.431
1-MOM20-1000	1000	20	13.73%	0.5	0.847	0.431
1-MOM25-1000	1000	25	14.33%	0.496	0.83	0.429
1-MOM30-1000	1000	30	13.98%	0.499	0.836	0.432
1-MOM05-2000	2000	5	14.01%	0.486	0.813	0.446
1-MOM10-2000	2000	10	14.48%	0.492	0.823	0.433
1-MOM15-2000	2000	15	14.16%	0.493	0.826	0.434
1-MOM20-2000	2000	20	14.15%	0.495	0.828	0.429
1-MOM25-2000	2000	25	15.91%	0.568	0.951	0.315

1-MOM30-2000	2000	30	16.12%	0.572	0.958	0.307
1-MOM05-3000	3000	5	14.64%	0.485	0.812	0.456
1-MOM10-3000	3000	10	13.84%	0.491	0.822	0.44
1-MOM15-3000	3000	15	14.29%	0.489	0.819	0.436
1-MOM20-3000	3000	20	15.94%	0.61	1.03	0.343
1-MOM25-3000	3000	25	13.44%	0.522	0.873	0.429
1-MOM30-3000	3000	30	16.34%	0.6	1	0.377

Table 2: Results from SHM Level 1 network

where the network codification used, had the format (SHM Level) – (Learning Algorithm) (PE's) - (Epochs), MSE stands for Mean Square Error, NMSE is Nominal Mean Square Error and r is the correlation coefficient [6]. From the results provided, the best test error was achieved for the 1-MOM25-3000 network and was 13.44%. This error could be misinterpreted as high enough, but for the case of a classification network is acceptable since the major concern is the number of misclassifications (how many misclassifications happened in the test sample) and not the variations of the probability of classification (which for the specific case the mean probability to belong to the specific class of failure was 85% that is an acceptable result).

SHM LEVEL 2

Using the same approach as for the case of SHM Level 1, a typical classification multilayer perceptron was used to locate the potential damage occurred within a composite patch. Due to the monitoring capabilities provided in [4], three classes of failures were assumed for the case of a composite patch, which are presented in figure 5.

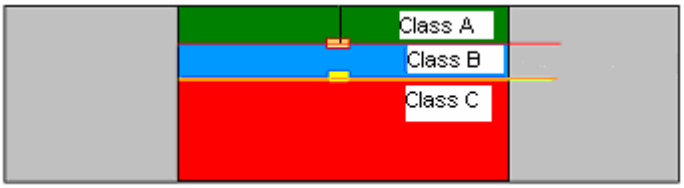


Figure 5: Classes of failure for SHM Level2

The data set was consisted of 850 exemplars (data points). The network used for the algorithm evaluation was consisted of 10 to 30 hidden processing elements and the test was performed with 1000 to 5000 epochs (iterations). Following the very same approach as in the case of SHM Level 1, the network output results are presented in table 3:

Network	Epochs	Hidden PE's	Test Error	MSE	NMSE	r
2-MOM05-1000	1000	5	8.25%	0.39	0.96	0.54
2-MOM10-1000	1000	10	14.41%	0.83	6.64	0.662
2-MOM15-1000	1000	15	3.60%	0.9	2.22	0.56
2-MOM20-1000	1000	20	3.82%	0.92	2.28	0.52
2-MOM25-1000	1000	25	15.30%	0.86	6.88	0.65
2-MOM30-1000	1000	30	3.22%	0.81	2	0.6
2-MOM05-2000	2000	5	9.27%	0.46	1.15	0.6
2-MOM10-2000	2000	10	7.72%	0.34	0.84	0.66
2-MOM15-2000	2000	15	7.75%	0.41	1.01	0.63
2-MOM20-2000	2000	20	7.65%	0.32	0.8	0.67
2-MOM25-2000	2000	25	8.38%	0.5	1.23	0.6
2-MOM30-2000	2000	30	9.39%	0.45	1.13	0.57
2-MOM05-3000	3000	5	9.30%	0.47	1.16	0.6
2-MOM10-3000	3000	10	10.18%	0.49	1.21	0.57
2-MOM15-3000	3000	15	8.24%	0.41	1.02	0.62
2-MOM20-3000	3000	20	8.37%	0.46	1.15	0.61
2-MOM25-3000	3000	25	9.20%	0.42	1.04	0.59
2-MOM30-3000	3000	30	8.11%	0.37	0.91	0.64

Table 3: Results from SHM Level 2 network

From these results, it was found out that the minimum error was achieved using the 2-MOM30-1000 network and, generally, the error varied from 3.22% to 15.3%.

SHM LEVEL 3

Using the same approach with SHM Level 1& 2 networks, but using instead a function approximation network, for the mapping

$$f(\varepsilon_z^{sensor1}, \varepsilon_z^{sensor2}) \rightarrow cracklength$$

the results achieved are presented in table 3:

Network	Epochs	Hidden PE's	Test Error	MSE	NMSE	r
3-MOM10-1000	1000	10	7.09%	0.097	0.1385	0.971
3-MOM15-1000	1000	15	4.08%	0.038	0.053	0.975
3-MOM20-1000	1000	20	8.20%	0.01	0.152	0.971
3-MOM25-1000	1000	25	6.34%	0.05	0.07	0.997
3-MOM30-1000	1000	30	7.48%	0.006	0.097	0.974
3-MOM10-2000	2000	10	2.81%	0.003	0.052	0.975
3-MOM15-2000	2000	15	6.69%	0.005	0.084	0.975
3-MOM20-2000	2000	20	5.50%	0.005	0.077	0.974
3-MOM25-2000	2000	25	2.19%	0.007	0.1	0.978
3-MOM30-2000	2000	30	4.46%	0.005	0.07	0.974
3-MOM10-3000	3000	10	5.52%	0.004	0.06	0.975
3-MOM15-3000	3000	15	6.03%	0.005	0.07	0.975
3-MOM20-3000	3000	20	6.84%	0.006	0.08	0.975
3-MOM25-3000	3000	25	6.93%	0.005	0.08	0.978
3-MOM30-3000	3000	30	2.69%	0.004	0.05	0.974
3-MOM10-4000	4000	10	3.25%	0.003	0.04	0.975
3-MOM15-4000	4000	15	3.73%	0.003	0.05	0.975
3-MOM20-4000	4000	20	6.03%	0.005	0.07	0.975
3-MOM25-4000	4000	25	8.30%	0.007	0.113	0.975
3-MOM30-4000	4000	30	6.37%	0.005	0.07	0.978
3-MOM10-5000	5000	10	6.53%	0.005	0.08	0.975
3-MOM15-5000	5000	15	4.80%	0.004	0.05	0.975
3-MOM20-5000	5000	20	5.33%	0.004	0.06	0.975
3-MOM25-5000	5000	25	3.57%	0.003	0.05	0.975
3-MOM30-5000	5000	30	6.41%	0.005	0.07	0.975

Table 4: Results from SHM Level 3 network

Whereas, it is obvious that the best network output is achieved using the 3-MOM25-2000 network architecture, while the error in all configurations varied between 2.19% and 8.3%.

SHM LEVEL 4

Finally, for the case of life expectancy of the repaired crack, having a propagating failure, suing the same techniques as those in SHM Level 3 networks for the mapping

$$f(\varepsilon_z^{sensor1}, \varepsilon_z^{sensor2}, \mathrm{N}_{consumed}, cracklength) \rightarrow \mathrm{N}_{\exp ected}$$

the following results were achieved using a data set of 2000 epochs:

Network	Epochs	Hidden PE's	Test Error	MSE	NMSE	r
4-MOM10-1000	1000	10	10.46%	0.0001	0.001	0.999
4-MOM15-1000	1000	15	13.79%	7.1E-05	0.00044	0.999
4-MOM20-1000	1000	20	16.69%	0.0001	0.001	0.999
4-MOM25-1000	1000	25	15.96%	0.00052	0.003	0.998
4-MOM30-1000	1000	30	17.32%	0.0003	0.002	0.998
4-MOM10-2000	2000	10	10.5%	0.00002	0.001	0.999
4-MOM15-2000	2000	15	13.94%	7.9E-05	0.0005	0.999
4-MOM20-2000	2000	20	7.14%	0.0001	0.0006	0.999
4-MOM25-2000	2000	25	3.82%	0.00008	0.0004	0.999
4-MOM30-2000	2000	30	14.90%	0.00013	0.00083	0.999
4-MOM10-3000	3000	10	10.97%	0.00003	0.0002	0.999
4-MOM15-3000	3000	15	3.53%	0.00038	0.0002	0.999
4-MOM20-3000	3000	20	4.34%	0.00004	0.0002	0.999

4-MOM25-3000	3000	25	8.38%	0.0001	0.0008	0.999
4-MOM30-3000	3000	30	8.49%	0.0004	0.0005	0.999

Table 5: Results from SHM Level 4 network

Whereas, it is obvious that the best network output is achieved using the 4-MOM20-3000 network architecture, while the error in all configurations varied between 3.53% and 17.32%.

RESULTS

Having completed the training process as well as the validation of the network architectures for the given data set, the optimum network architecture for each SHM Level was submitted to a "live" process, during which real strain data (not training or validation set data) taken from the sensors were fed on the networks. The network outputs for this "unknown" data set were very good, close enough to the expected causes. For economy of space, the most interested cases of network architectures for SHM Level 3 and 4 are presented in figure 6.

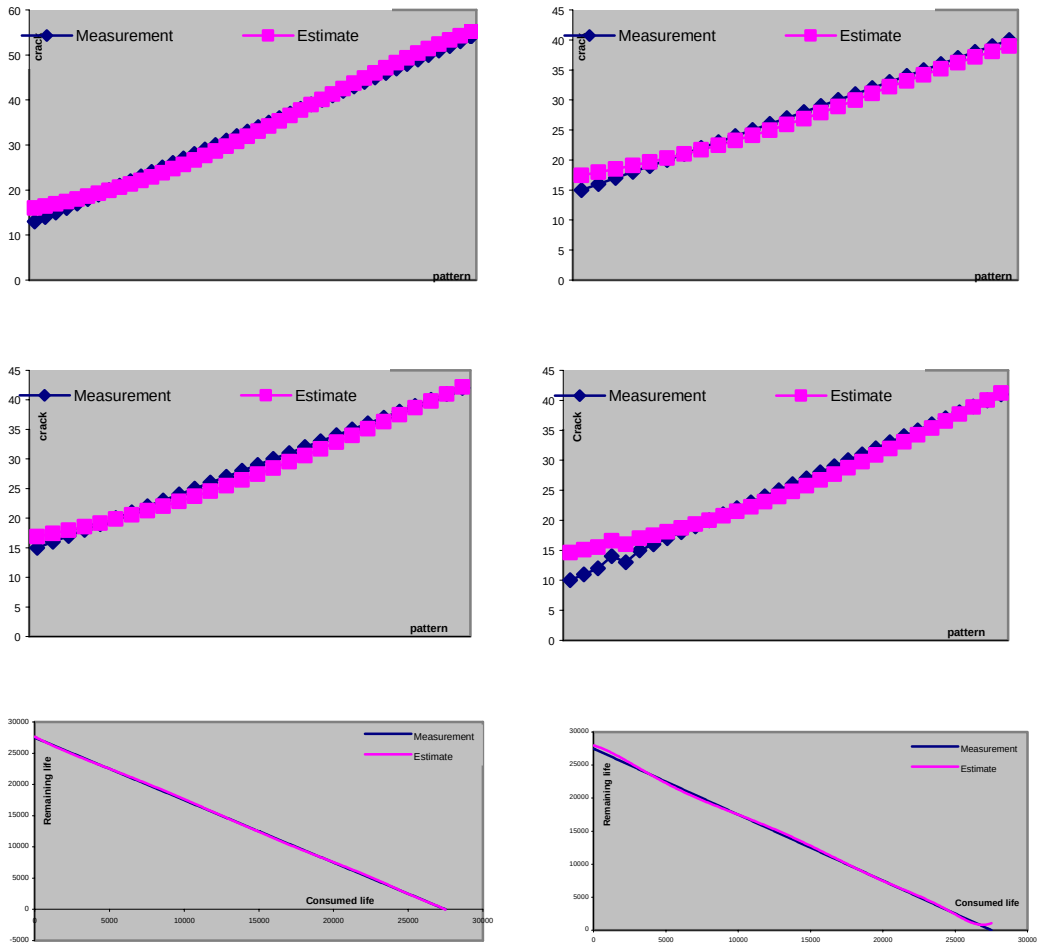


Figure 6: Network output using "unknown" data set for SHM Level 3 and 4 on various test specimens

Similar results were achieved for the rest of the network architectures used for SHM Level 1, 2. Moreover, in order to check the stability of the solutions due to small data perturbations, uniform "noise" (from 1 to 5%) was added to the data input of the networks in order to evaluate their response. Due to the fact that a uniform type "noise" filter was added to the design of the networks, these small perturbations did not result to solution instability, but on the contrary, validated the stable performance of these networks as presented on figure 7.

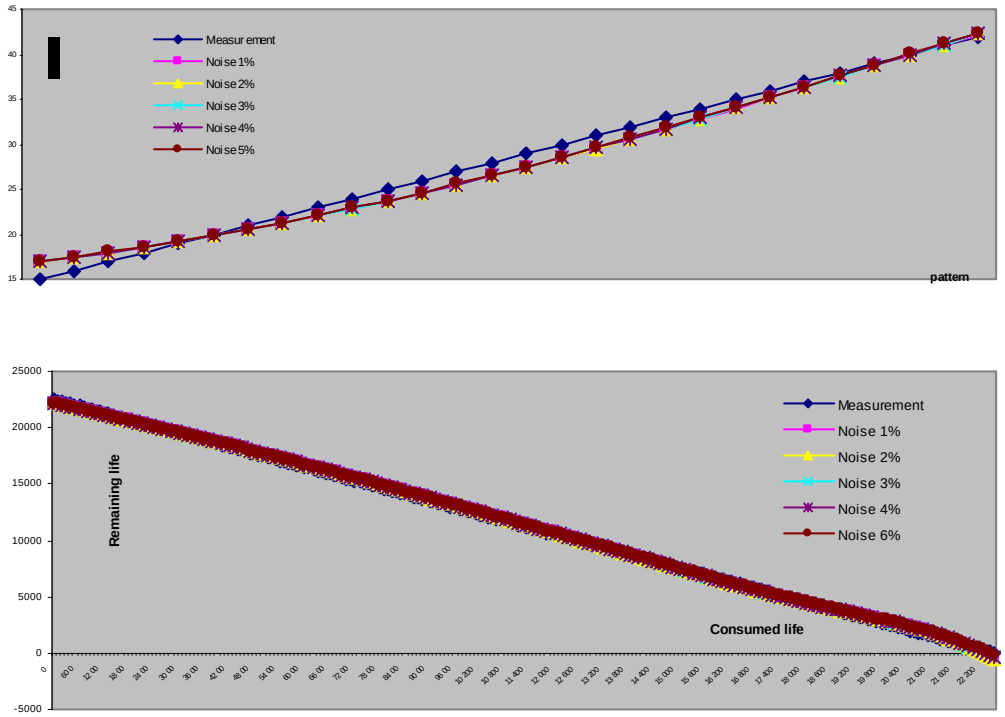


Figure 7: Network output stability testing SHM Level 3 and 4

CONCLUSIONS

Following the here presented analysis, it was found out that, the inverse problem of fault detection, qualification / quantification and life expectancy of a composite patch repair, during its operational usage, can be treated using neural network techniques and fiber optic sensors. The results achieved showed that the combination of fiber optic sensors and neural network damage prognosis and diagnosis capabilities can be considered a semi-integrated health monitoring system for the composite patch repair technique, ensuring the long term stability and durability of the repair.

ACKNOWLEDGEMENTS

The authors would like to thank the General Secretariat of Research and Development, Ministry of Education, Hellas for the financial support during the present work.

REFERENCES

[1] W. Stawenski et al, Health Monitoring of Aerospace structures, Wiley, 2004.

[2] W. Stawenski, MONITOR Research Program, EU IV Framework Program, 2001.

[3] A. Rytter, Vibration Based Inspection of Civil Engineering Structures PhD Dissertation, Department of Building Technology and Structural Engineering, Aalborg University, Denmark, 1993

[4] Z. Marioli-Riga, A. Karanika, S. Panagiotopoulos, G. Tsamasphyros, N. Fournarakis, G. Kanterakis, Structural health monitoring of composite patch repairs using embedded fiber Bragg grating sensors, GRACM 05, Limassol 2005.

[5] R. Rojas, Neural Networks: A Systematic Introduction, Springer,1996

[6] J. Proncipe, N. Euliano, Neural and Adaptive Systems: Fundamentals through Simulations, Wiley, 2000.

STRUCTURAL HEALTH MONITORING OF COMPOSITE PATCH REPAIRS USING EMBEDDED FIBER BRAGG GRATING SENSORS

Z. Marioli-Riga, A. Karanika, and S. Panagiotopoulos
Hellenic Aerospace Industry
Research and Product Development Department
Shimatari, Viotia, Greece

G.J. Tsamasphyros, N.K. Fournarakis, and G.N. Kanderakis

National Technical University of Athens
Dept of Applied Mathematics and Physical Sciences
Section of Mechanics
Athens, Greece
E-mail: nfurnar@central.ntua.gr

ABSTRACT

Modern developments in the space of aeronautics, particularly since the incident of 11th of September 2002 in the U.S.A. which affected negatively the economic potential of air companies worldwide, render explicit the need for exploitation of air structures up to their designed structural life or even further and consequently the need for adoption of maintenance techniques supporting their structural integrity in an acceptable, according to their designed criteria, manner. One of the most promising techniques is the so called "composite patch repair method", a method capable to repair cracked metallic structures using adhesively bonded composite patches, that constitutes an effective method for re-establishment of structural integrity. However, even though this method has important advantages, problems concerning the long-term behavior of the adhesive, the appearance of debonds, etc. appear and have led to the considerable delayed widespread acceptance of this method.

In the present study the realization and implementation of elements of a structural health monitoring system for composite patch repairs is being performed, combining elements from the theory of structural repair mechanics, the theory of intelligent materials and structures and particularly the technology of optical sensors and neural networks.

This investigation was carried out within the purposes of a Research project by the Research and Product Design Department of Hellenic Aerospace Ind.

INTRODUCTION

Current economic world conditions are forcing to the operation of both military and civilian aircraft well beyond their original design life, resulting in innovative repair techniques. The recent development of high strength fibres and adhesives has led to the invention of a new methodology for the repair of metallic structures by the adhesive bonding of patches manufactured by composite materials. Bonded repairs are mechanically efficient, cost effective and can be applied rapidly to produce an inspectable damage tolerant repair. The actual objective of the repair of a cracked or corroded metallic structure by an adhesively bonded composite patch is, practically, the transfer of loads from the one side of the sound material to the other via the patch, deviating the damaged area. With this technique the patch is usually manufactured using carbon / epoxy or boron / epoxy composite materials, while its bonding on the structure is achieved using high strength adhesives. The load transfer from the component to the patch and vice versa is achieved by the shear stresses applied on the adhesive layer. Even though the technique presents great advantages from a life cycle cost point of view for the aeronautic structures, the certification of the method for operational usage is not yet completed. Problems such as the long term stability of the adhesive bonding, precludes the method from a widespread acceptance as a field repair technique.

In order to advance the potential spreading of the composite patch repair technique, on line health monitoring should be adapted to secure the long term stability and the structural integrity of the repair and the structure as a whole. In that respect and in order to enable on line monitoring of the local stress field into a composite patch during a potential failure, such as crack or debond propagation, optical fiber sensors can be structurally integrated into it. Fiber optic sensors became lately a main research area in the field of "Smart Structures" because of the significant advantages they offer compared to previous efforts in the area of stress-strain monitoring (e.g. strain gages, etc.). Various types of sensors can be intrinsically embedded in a composite material, like the Bragg Grating, Fabry-Perrot, the Polarimetric, etc. On the contrary, there are not many candidate fiber sensors in the case of a

composite patch due to the miniature dimensions used, the single fiber embedding and interface requirement, the requirement to trace a single, decoupled strain component, etc. Therefore, in the present study, Fiber Bragg grating sensors were embedded in the composite patches, to trace the mechanical field variations. The field variations, for simplicity reasons, were assumed mechanical only, decoupled from any thermal effect, by keeping the environmental conditions stable during the experimental study.

SPECIMEN DESIGN

In order to proceed to the experimental studies, various specimens were designed and manufactured. First of all and before any other experiment, the behavior and the repeatability of the strain measurements taken from the Bragg sensors should be evaluated as well as the durability of the sensor during the test process. Three specimens (Type I) were manufactured, according to the geometrical characteristics presented in figure 1.

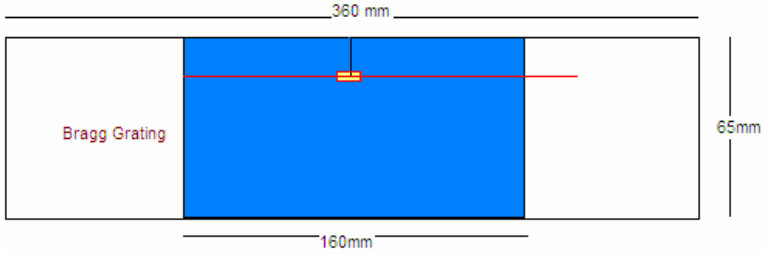


Figure 1: Measurement repeatability specimen (Type I)

The materials used, considered isotropic, are presented on Table 1.

Material	Thickness(mm)	E (MPa)	G (MPa)	ν
Aluminium 2024-T3	6	72000	26900	0.3
Textron 5521 Prepreg	0.125 per lamina	207000	4800	0.21
FM73 Film Adhesive	0.2	---	750	---
Optical Fiber	Diameter 0,1mm	70000		0.29

Table 1: Material properties

The vertical projection of the crack tip of the metallic structure was chosen as the sensor location, based on the numerical simulation results presented in [1]-[5]. The composite patch used was manufactured using six laminates of carbon epoxy prepreg. The sensor was embedded between the third and fourth lamina, based on the results of [4]. The physical characteristics of the sensors were the following:

- Center wavelength λ_B= 1535±10 nm
- $\Delta\lambda$=~0.7nm
- Reflectivity R= 85~97 %
- Sensor length L= 2mm

The specimens were submitted to a tensile testing process using an Instron test machine. The specimens were gradually loaded to a range of 1 to 10 KN tensile load and measurements for each load condition were recorded from the optical fiber sensors. A COTS Micron Optics Bragg Interrogator had been used for the acquisition of measurements, having the capability to store digitally the wavelength shifts of the sensors during the loading process.

Moreover, specimens were manufactured in order to study potential fault propagation. Taking into account that the faults studied were the crack or debond propagation, the specimens of figure 2 were manufactured (Type II). Each of these specimens was representing a potential debond (yellow area) developed in the area of the crack tip, between the composite patch and the repaired metallic area. The aim of the test series was to examine the possible propagation patterns of the cracked and/or debond area, using NDI techniques. Fiber optic sensors were not embedded in these specimens, due to the fact that measurements should not be taken. The specimens, manufactured using the materials of table 2 with six ply composite patch, were submitted to fatigue testing, using the following load data:

- Mean load : 2,1 tn
- Amplitude : 1,9 tn
- Load frequency : 10 Hz

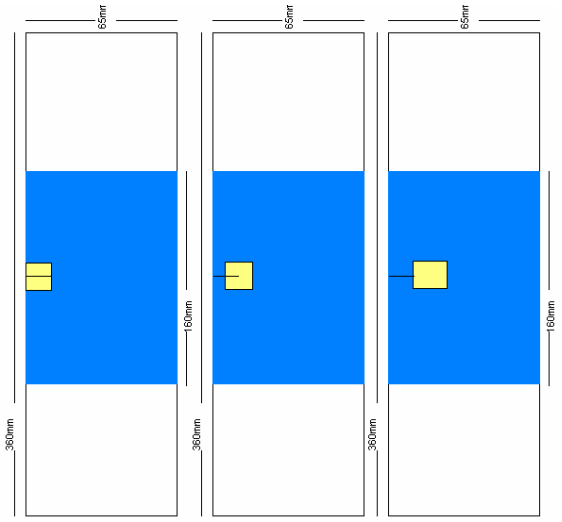


Figure 2: Fault propagation study specimens (type II-1, 2 and 3 respectively)

Finally, in order to monitor the propagation of a failure in a composite patch repair, more specimens were manufactured, having embedded optical fiber sensors, as presented in figure 3. Each of these specimens had two embedded optical sensors, in positions determined after the experimental study of specimens type II, using the same materials and sensors of specimens type I and II. The crack tip sensor was called as sensor “a” while the second sensor was called “b”. The loading conditions of the specimens were identical with the conditions of specimens type II.

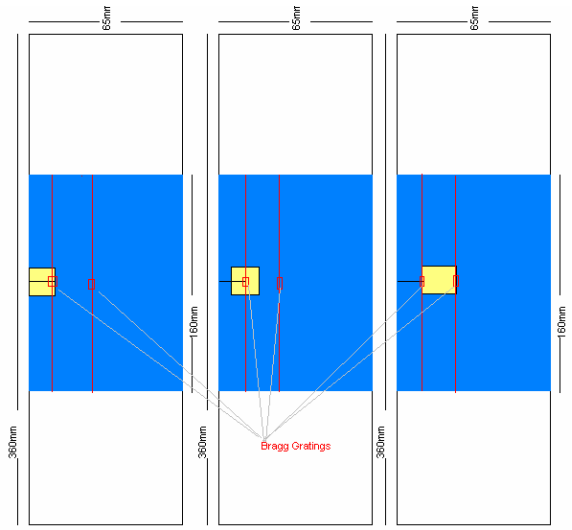


Figure 3: Monitored specimens (type III-1, 2 and 3 respectively)

For each of the above, three specimens were manufactured in order to secure the experimental results. The exact specimen details are presented in Table 2.

Specimen Code	III-1			III-2			III-3		
	A	D	G	B	E	H	C	F	I
FBG a – Center Wavelength (nm)	1551,24	1551,28	1563,47	1551,12	1551,59	1562,36	1551,8	1551,34	1551,51
FBG b - Center Wavelength (nm)	1551,32	1551,26	1530,14	1551,65	1551,52	1551,82	1551,4	1551,89	1551,39
FBG a – Reflectivity (%)	99	98	97,8	98,6	98,8	97,1	99,3	98,9	97,6
FBG b - Reflectivity (%)	97,5	98,75	32,9	98,4	98,9	97,9	93	98,75	98,8

Table 2: Specimen type III details

The data acquisition during the fatigue loading of specimens type III, was based on the following technique:

- Measurement of crack length at 10K cycles and every 2.5K cycles with simultaneous sensor wavelength shift recording
- Measurement of debond area using C-Scan NDI every 10K cycles
- Ramp type tensile loading (figure 4) every 10K cycles with simultaneous sensor wavelength shift recording

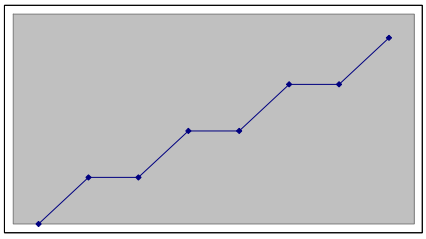


Figure 4: Ramp type loading of specimens type III

EXPERIMENTAL RESULTS – SPECIMENS TYPE I & II

The results of the tensile loading of specimens type I are presented in figure 4. It is shown that the repeatability of measurements is very satisfying and the strain measuring capability of the sensor is accurate, therefore the sensors were appropriate for the experiment.

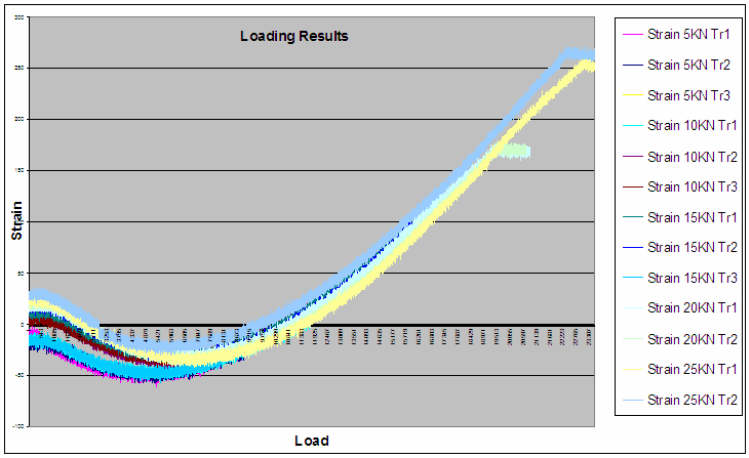


Figure 5: Load vs Strain results during the repeatability test series

It was also found that during the initial tensile loading, compressive loads are developed near the crack tip, due to the fact that the specimen has a resulted curvature from the curing process because of the thermal coefficient mismatch of the patch and the aluminum material.

The results of the specimen type II loading, with respect to crack extension, are presented in figure 5.

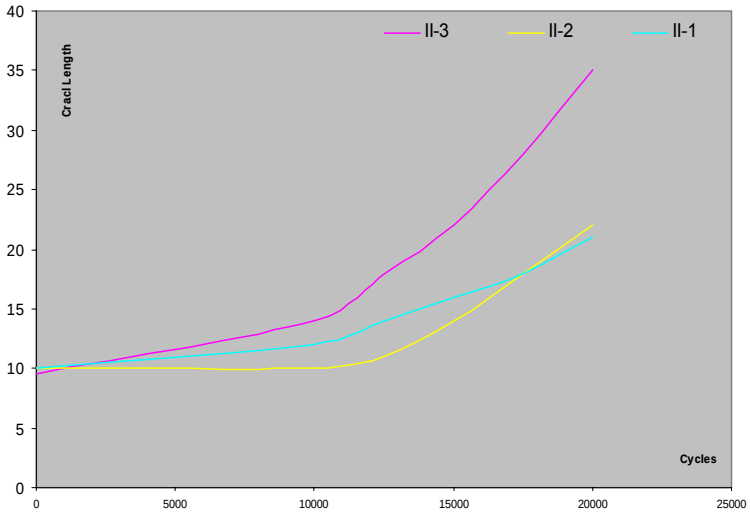


Figure 6: Crack extension of specimens type II

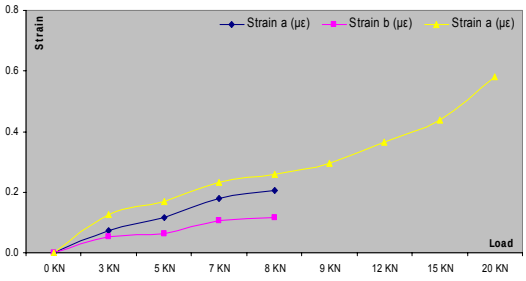
Moreover, C-Scan NDI was performed on these specimens in order to check the debond propagation due to the fatigue loading. From the results of the studies it was found out that the crack propagated faster on the specimen type II-3 compared to the specimen type II-1 or 2. From these results it was obvious that specimens type II-3 were more prone to damage compared to the other two specimens. These results are self explanatory since the position of the debond of specimen type II-3 results in a more stress intensive crack tip area as compared to the other specimens, due to the fact that all loads are undertaken by the structure and not the patch at the specific area.

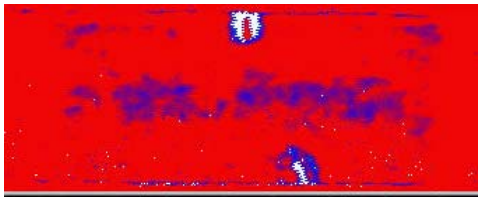
EXPERIMENTAL RESULTS – SPECIMENS TYPE III

Following the above described test procedure, various results were taken form the fiber optic sensors during the testing of specimens type III. The results are split in two major categories: results related to the debond extension and results related to the crack propagation.

Debond Extension

The measured results for each specimen of Table 2, in accordance with the bond extension monitoring, are presented in figure 7.





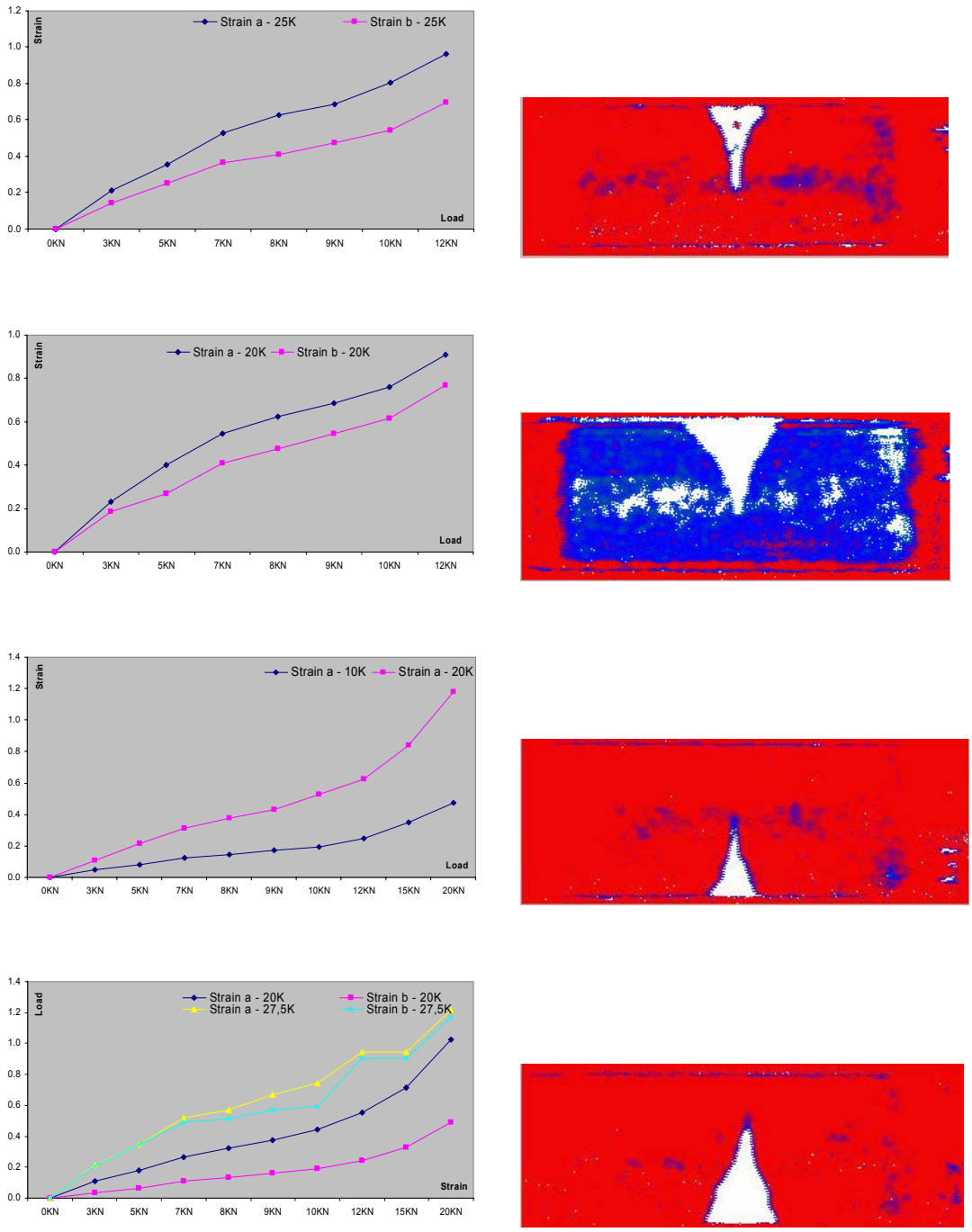


Figure 7: Debond extension monitoring and final debond area for specimens type III A, B, C, E and G

From the above results it is obvious that, during the ramp loading, there is a shift in strain measurement due to the fact that the debond has propagated and resulted in a field alternation near the fiber optic sensors. The amount of debond extension determines the strain shift for the sensors.

Crack Propagation

The measured results for each specimen of Table 2, in accordance with the crack propagation monitoring, are presented in figure 8.

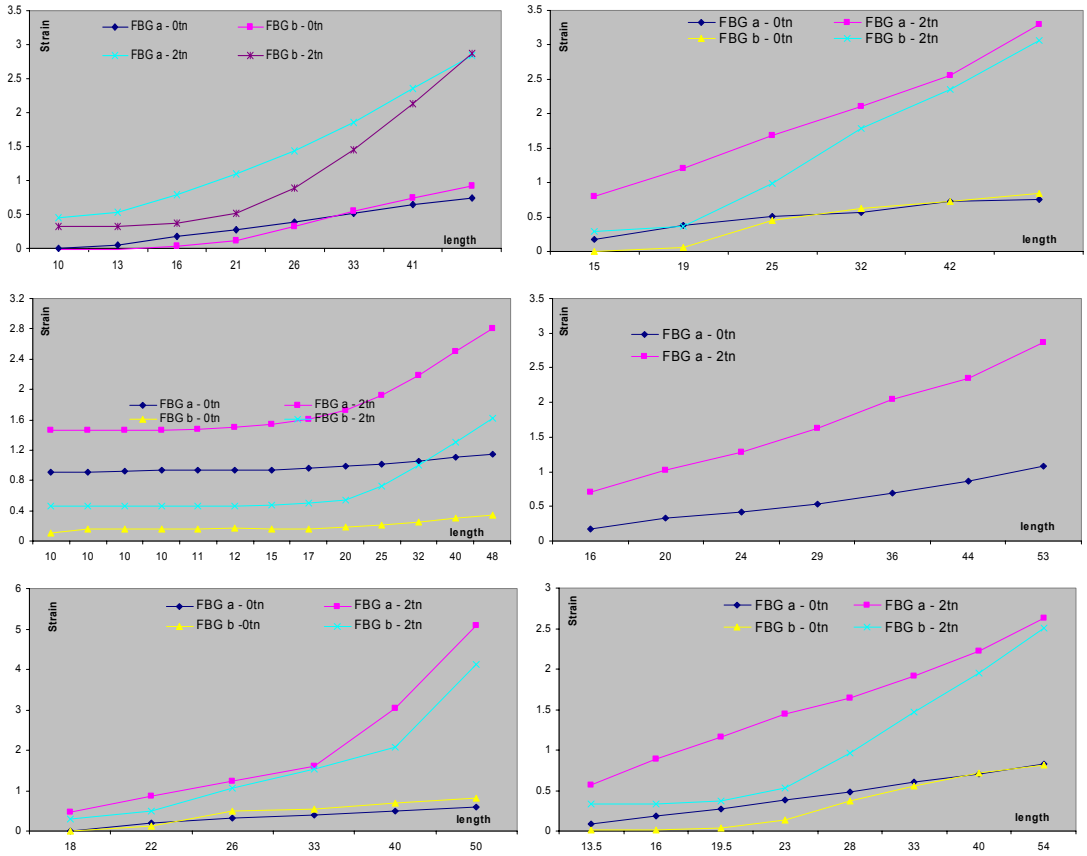


Figure 8: Crack Propagation monitoring for specimens type III B, C, D, E, F and G

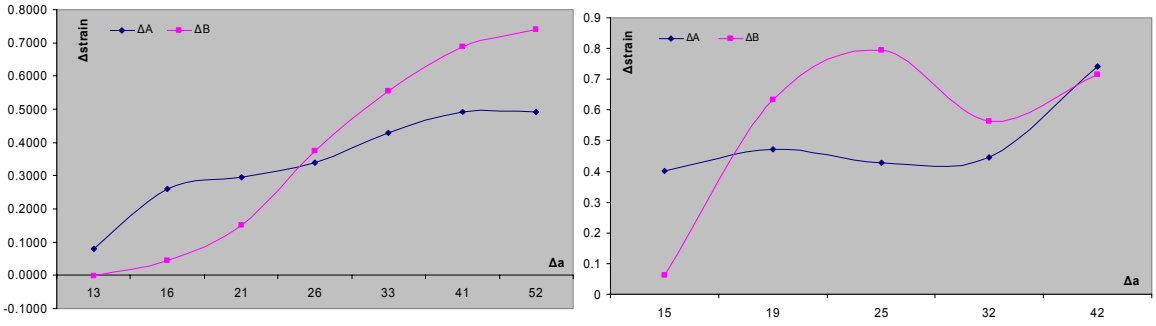
From the above results, a strain increment is obvious during the crack propagation. Moreover, for the sensor "b" of each specimen, a sudden strain increment was noticed when the crack was passing through the vertical level of the sensor.

RESULT ANALYSIS

Taking into account the above results, a study was made to relate the measured results with the propagating failure. Considering the crack propagation as the major failure that could lead to the degradation of the repair strength and the final structure failure, the above data were further analyzed, assuming that:

- A: strain of sensor a
- B: strain of sensor b
- ΔA: strain increase of sensor a
- ΔB: strain increase of sensor b
- Δa: crack length increase

Using this notation, the test results for the crack propagation failure case, are presented in figure 9.



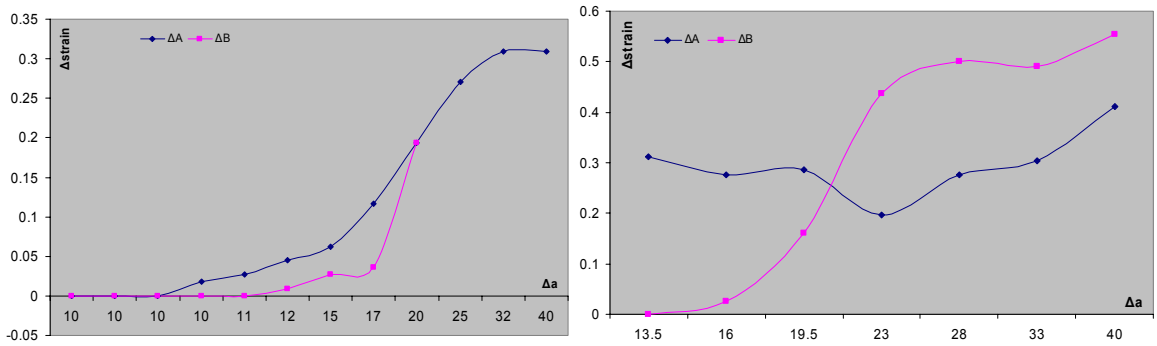


Figure 9: Strain increment due to crack propagation for specimens type III B, C, D and G

From the above figure, it is obvious that during the crack propagation and when the crack "passes" from the sensor "b", the increment curves of the two sensors cross each other, giving a notion of the crack length on that time. Relevant curves can be developed for the case of the debond propagation, presenting the capability of the sensors to trace a failure propagation in the repaired structure.

CONCLUSIONS

Following the here presented experimental analysis, it was found out that optical fiber sensors can be used efficiently to monitor the health of a composite patch repaired structure. The sensors presented very good measurement stability, great sensitiveness and the capability to trace effectively propagating failures in the repaired area.

ACKNOWLEDGEMENTS

The authors would like to thank the General Secretariat of Research and Development, Ministry of Education, Greece as well as the Hellenic Aerospace Industry for the financial support and the facilities used during the present work.

REFERENCES

[1] G. Tsamasphyros, G. N. Kanderakis, N.K. Furnarakis, Z. P. Marioli-Riga,Optimization of Embedded Optical Sensor Locations in Composite Repairs Applied Composite Materials Volume 10 No3: 129-140, May 2003, Kluwer Academic Publishers, Netherlands.

[2] G.J. Tsamasphyros, N.K. Furnarakis, G.N. Kanderakis, Three Dimensional Finite Element Analysis of Composite Patches with Embedded Optical Fibres – Through Thickness Optimization, Z. P. Marioli-Riga "International Conference on Computational Engineering & Sciences, ICES'01, Puerto Vallarta, Mexico (2001).

[3] G. J. Tsamasphyros, G. N. Kanderakis, N. K. Furnarakis, Z. P. Marioli-Riga, R. Chemama, R. Bartolo, Three – Dimensional Finite Element Analysis of composite patches with embedded optical fibers – Selection of Optical Fibers Paths and Sensors Locations, Structural Health Monitoring 2002, Daniel L. Balageas, 1203-1210, Destech Publications, ENS Cachan France, July 10-12, 2002.

[4] G. J. Tsamasphyros, N. K. Furnarakis, G. N. Kanderakis, Z. P. Marioli-Riga,Three – Dimensional Finite Element Analysis of composite patches with embedded optical fibers – Optimizing Optical Fiber Embedding Location: Structural Health Monitoring 2002, Daniel L. Balageas, 1219-1226, Destech Publications, ENS Cachan France, July 10-12, 2002.

[5] G. J. Tsamasphyros, G. N. Kanderakis, N. K. Fournarakis, Z. P. Marioli-Riga, Detection of patch debonding in composite repaired cracked metallic specimens, using optical fibers and sensors SPIE Optical Metrology Conference, 23-26 June 2003, Munich, Germany.

[6] W. Stawenski et al, Health Monitoring of Aerospace structures, Wiley, 2004.

[7] A.A. Baker, R. Jones, *Bonded Repair of Aircraft Structures*, Martinus Nijhoff Publishers, Dordrecht (1988).

Numerical Methods and Algorithms

ULTIMATE STRENGTH ANALYSIS OF ARBITRARY CROSS SECTIONS UNDER BIAXIAL BENDING AND AXIAL LOAD BY FIBER MODEL AND CURVILINEAR POLYGONS

Aristotelis E. Charalampakis and Vlasis K. Koumousis

Institute of Structural Analysis and Aseismic Research
National Technical University of Athens
NTUA, Zografou Campus GR-15773, Athens, Greece
e-mail: vkoum@central.ntua.gr, web page: http://users.ntua.gr/vkoum

Keywords: biaxial bending, fiber model, failure surface, composite structure.

Abstract. *A fiber model algorithm for the analysis of arbitrary cross sections under biaxial bending and axial load is presented. The method can be applied to complex cross sections of irregular shape and curved edges, with or without openings and consisting of various materials. The only assumption is that plane sections remain plane. The cross section is described by curvilinear polygons. The material properties are user – defined; the stress – strain diagrams of all materials consist of any number of consecutive polynomial segments (up to cubic). Various effects such as concrete confinement, concrete tensile strength, strain hardening of the reinforcement etc. may be taken into account. Apart from ultimate strength analyses of cross sections, the algorithm can be applied to other problems in which the Bernoulli – Euler assumption is valid. Based on the proposed algorithm, a computer program with full graphical interface was developed.*

1 INTRODUCTION

The analysis of an arbitrary cross-section under biaxial bending and axial load has received extensive attention in the literature lately [6], [4], [9]. With the advent of inexpensive computer systems, the generation of the failure surface has been made possible using the "fiber" approach. This approach produces consistent results that agree closely with experimental results [2].

The failure of the cross section corresponds to the top of the moment – curvature diagram. However, the conventional failure, defined by design codes, occurs when any of the materials reaches its predefined maximum allowable compressive or tensile strain.

The results of such an analysis are important as they can be used in non-linear analyses of structures where the plastic deformations of a structural element are functions of the load history and the distance of the load vector from the surface. Moreover, it provides grounds for the damage analysis of the cross section.

2 GENERATION OF FAILURE SURFACE

There exist three different techniques to generate the failure surface of an arbitrary cross section: (1) interaction curves for a given bending moments ratio, (2) load contours for a given axial load and (3) isogonic or 3D curves.

The first two techniques require the calculation of the exact position of the neutral axis. The set of equilibrium equations are non-linear and coupled and an iterative approach such as the quasi-Newton method is needed to determine the position of the neutral axis, as proposed by Yen [10]. These procedures are not straightforward to implement and, in many cases, are sensitive to the selection of the origin of the reference system. Moreover, these algorithms usually become unstable near the state of pure compression.

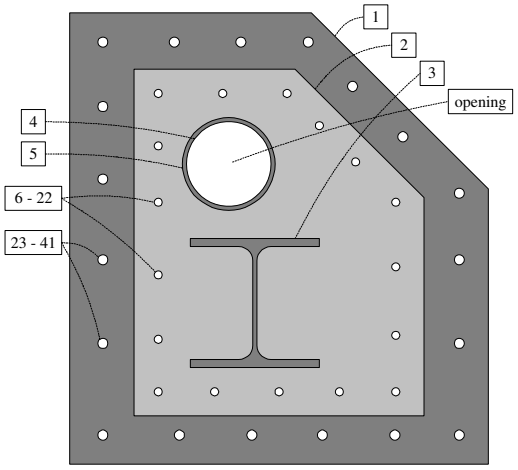
On the other hand, the third technique, which is used in the method presented, is more direct because the direction of the neutral axis is assumed from the very beginning. The produced points describe a more complex 3D plot, because the meridians of the failure surface, in general, are not plane. This is due to the asymmetry of the cross section, as described later.

3 CROSS SECTION

The curvilinear polygon is the only type of graphical entity that is used for the description of all cross sections. A curvilinear polygon has edges that may be straight lines and/or circular arcs. Since these polygons can be nested in any depth, it is obvious that almost any cross section can be described accurately. Circles are taken

into account as two-sided curvilinear polygons with curved edges. Notice that even small objects, such as the reinforcement bars, are treated as actual graphical entities and not dimensionless individual fibers.

In order to significantly reduce the expensive calculations required to identify the various regions in a complex cross section with several materials, each curvilinear polygon is treated separately. Two material properties are defined: the "foreground" material and the "background" material. The foreground material is taken into account with a positive sign, whereas the background material is taken into account with a negative sign. Therefore, almost any cross section can be described, as shown in the example of Figure 1:

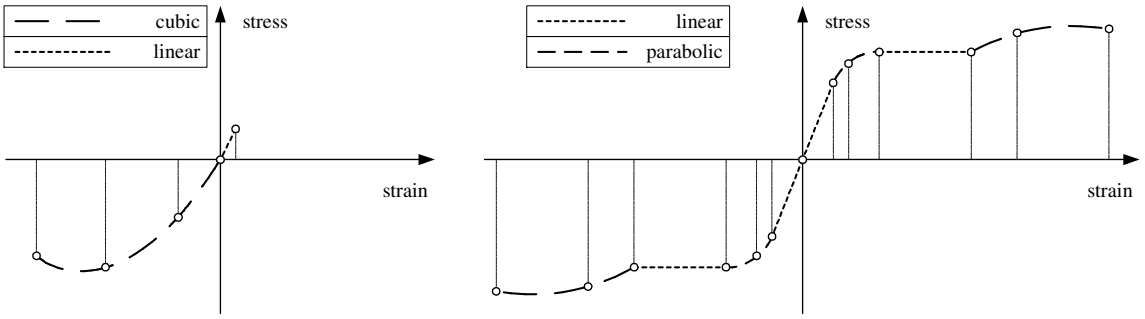


Entity	Number of Nodes	Foreground material	Background material
1	5	Unconfined (outer) concrete	None
2	5	Confined (inner) concrete	Unconfined (outer) concrete
3	16	Structural steel	Confined (inner) concrete
4	2	None	Structural steel
5	2	Structural steel	Confined (inner) concrete
6 – 22	2	Reinforcement	Confined (inner) concrete
23 – 41	2	Reinforcement	Unconfined (outer) concrete

Figure 1. Example of complex cross section consisting of various materials

4 MATERIAL PROPERTIES

The stress – strain diagrams of all materials are composed of any number of consecutive segments. Each segment is a polynomial expression (up to cubic), which is automatically defined by an appropriate number of points; for example, a cubic segment is defined by four consecutive points. Therefore, the stress strain diagrams of a certain kind of concrete and steel may be defined as shown in Figure 2:



(a) Example of stress - strain diagram (concrete)

(b) Example of stress - strain diagram (steel)

Figure 2. Example of stress – strain diagrams (tension positive)

5 CALCULATIONS

5.1 Rotation

We assume that the X axis is parallel to the central axis of the element. Any convenient point may be used as the origin for the calculations. Since the direction of the neutral axis is assumed from the beginning, it is convenient to express all coordinates in another YZ Cartesian system with Y axis parallel to the chosen direction of the neutral axis. Therefore, the Cartesian system is rotated counter-clockwise around the origin by an angle θ, as shown in Figure 3. In this way, the strains and therefore the stresses vary only in Z axis.

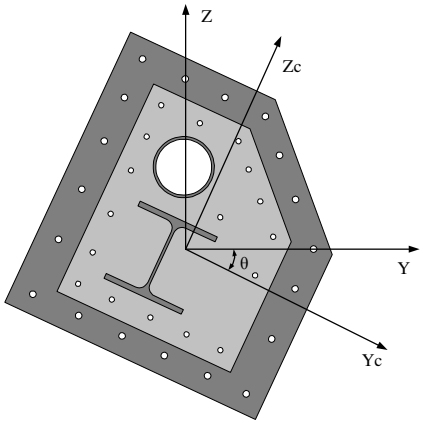


Figure 3. Rotation of cross section

5.2 Decomposition of curvilinear polygons

The next step is the decomposition of all curvilinear polygons into curvilinear trapezoids. The top and bottom edges of the curvilinear trapezoids are straight lines parallel to the neutral axis whereas the left and right edges may be straight lines or arcs. This procedure is required only once for every direction of the neutral axis; this basic set of trapezoids may be stored in memory and retrieved when needed.

Figure 4 shows an example of decomposition of a steel section and some of the produced curvilinear trapezoids. Note that the section is described exactly by a 16-node curvilinear polygon:

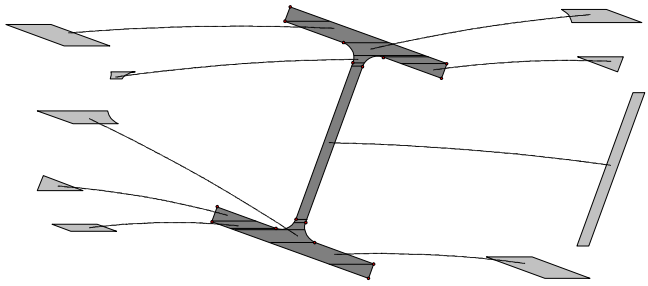


Figure 4. Decomposition of a steel section into curvilinear trapezoids

For reasons of simplicity we will drop the term “curvilinear” for both the curvilinear polygons and the curvilinear trapezoids.

5.3 Calculation of integrals

The next step is the calculation of the basic integrals of the trapezoids. These integrals are of the form $y^m \cdot z^n$, where m, n, are specific integers (equation (1)). All expressions for the integrals are analytical. Again, the integrals need to be evaluated only once for every direction of the neutral axis and the results can be stored in memory for later use. Therefore, the overhead for using analytical expressions is minimal.

$$I^{j}_{(m,n)} = \int_{trapezoid\ j} \left(y^m \cdot z^n \right)$$
$$(m,n) = (0,0..4),\ (1,1..4) \tag{1}$$

This method is also used for the exact calculation of cross sectional properties, such as area, first moments of area, centroids, moments of inertia, products of inertia, principal axes etc.

5.4 Strain distribution

As mentioned before, the main assumption is that plane sections remain plane. Therefore, three parameters are needed to define the deformed plane, namely the direction of the neutral axis (angle θ), the curvature k and the strain ε_0 at the origin, as shown in Figure 5:

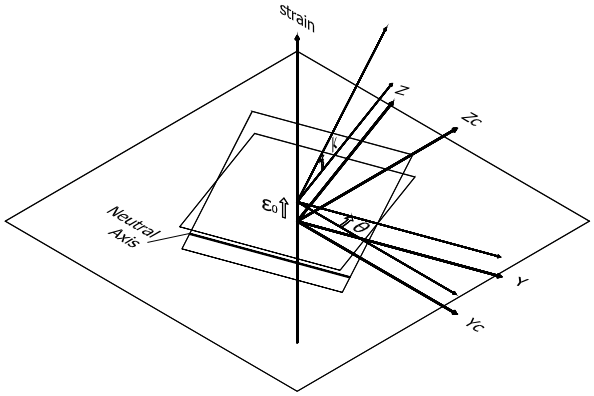


Figure 5. Deformed plane (expressed in terms of strains)

Since the Cartesian system has been rotated by an angle θ, the strain is a function of z only (equation (2)):

$$\varepsilon(z) = \varepsilon_0 + k \cdot z \tag{2}$$

Given k, ε_0, the neutral axis is defined as the line parallel to the Y axis at a distance given by equation (3):

$$z_{na} = -\frac{\varepsilon_0}{k} \tag{3}$$

5.5 Calculation of stress resultants

The calculation of stress resultants will take place for an imposed deformed configuration defined by a set of given values of the parameters θ, k, ε_0. In general, we assume that the segment of the stress – strain diagram covering the specific trapezoid is a cubic polynomial expression of the following form:

$$\sigma(\varepsilon) = \sum_{i=0}^{3}\left(a_i \cdot \varepsilon^i\right) \tag{4}$$

The coefficients α_i are known from the properties of the material. Substituting relation (2) into (4) we obtain:

$$\begin{aligned}
&\sigma(k, \varepsilon_0, z) = \sum_{i=0}^{3}\left(b_i \cdot z^i\right) \\
&b_0 = \alpha_3 \cdot {\varepsilon_0}^3 + \alpha_2 \cdot {\varepsilon_0}^2 + \alpha_1 \cdot \varepsilon_0 + \alpha_0 \\
&b_1 = 3 \cdot \alpha_3 \cdot {\varepsilon_0}^2 \cdot k + 2 \cdot \alpha_2 \cdot \varepsilon_0 \cdot k + \alpha_1 \cdot k \\
&b_2 = 3 \cdot \alpha_3 \cdot \varepsilon_0 \cdot k^2 + \alpha_2 \cdot k^2 \\
&b_3 = \alpha_3 \cdot k^3
\end{aligned} \tag{5}$$

The stress resultants of trapezoid j are calculated by integration of equation (5) as:

$$N_x^j = \int_{trapezoid\ j}\left(\sum_{i=0}^{3}\left(b_i \cdot z^i\right)\right) = \sum_{i=0}^{3}\left(b_i \cdot \int_{trapezoid\ j}\left(z^i\right)\right) = \sum_{i=0}^{3}\left(b_i \cdot I_{(0,i)}^j\right) \tag{6}$$

$$M_y^j = \int_{trapezoid\ j}\left(\sum_{i=0}^{3}\left(b_i \cdot z^i\right) \cdot z\right) = \sum_{i=0}^{3}\left(b_i \cdot \int_{trapezoid\ j}\left(z^{i+1}\right)\right) = \sum_{i=0}^{3}\left(b_i \cdot I_{(0,i+1)}^j\right) \tag{7}$$

$$M_z^j = \int_{trapezoid\ j}\left(\sum_{i=0}^{3}\left(b_i \cdot z^i\right) \cdot y\right) = \sum_{i=0}^{3}\left(b_i \cdot \int_{trapezoid\ j}\left(z^i \cdot y\right)\right) = \sum_{i=0}^{3}\left(b_i \cdot I_{(1,i)}^j\right) \tag{8}$$

Note that the integrals $I_{(m,n)}^j$ have already been calculated and are independent of k, ε_0. By a simple summation of the stress resultants of all trapezoids, the overall forces and bending moments required to impose

the specific deformed configuration are obtained. Therefore, the steps to calculate the stress resultants are as follows:

- Pick k, $\varepsilon_{0.}$
- For trapezoid j:
  Calculate b_i as functions of α_i (equations (5))
  Calculate the stress resultants using the known values of $I_{(m,n)}^{j}$ (equations (6) to (8)).
- Sum the results from all trapezoids to obtain overall stress resultants N_X, M_Y, M_Z.

6 CONSTRUCTION OF MOMENT-CURVATURE DIAGRAM

For a specified axial load and direction of neutral axis (angle θ), a full moment – curvature diagram can be constructed. After the initialization (rotation of cross section, decomposition of polygons into trapezoids, calculation of the basic integrals of the trapezoids), small increments of Δk are applied as imposed curvature. Since θ, k are given, the deformed plane is only a function of strain ε_0 at the origin. The algorithm calculates an upper and lower bound of ε_0 and uses a fast Van Wijngaarden – Dekker – Brent method to calculate ε_0 in order to achieve axial equilibrium to a specified accuracy.

As the curvature increases, the neutral axis moves perpendicular to its direction. This incremental procedure continues until the moment reaches a maximum (failure), or until one of the materials reaches the maximum compressive or tensile strain specified by the user (conventional failure). Thus, the complete moment – curvature diagram can be obtained, both for the primary moment M_Y and for secondary moment M_Z. In general, the secondary moment M_Z, expressed by equation (8), is small as compared to the primary moment M_Y. Finally, the moments can be expressed in the global reference system with an inverse rotational transformation.

The algorithm uses a variable curvature step which is adjusted automatically; therefore, the final result is independent of the initial curvature step (specified by the user). A small initial curvature step produces a smooth moment – curvature diagram.

7 CONSTRUCTION OF INTERACTION CURVES AND FAILURE SURFACES

By repeating the procedure described previously for different directions θ of the neutral axis in the range of 0^0 – 360^0, we are able to construct the interaction curve for a given axial load. In addition, by constructing interaction curves for various axial loads, we are able to construct the full failure surface of the cross section equator – by – equator.

8 DEFORMED CONFIGURATION UNDER GIVEN LOADS

The algorithm can also be used for calculating the deformed configuration of a cross section under given loads. The calculation is a trial and error procedure (Figure 6). The task is to calculate the parameters θ, k, ε_0 of the deformed plane for which the cross section is in equilibrium with the external loads N_{Xc}^{T}, M_{Yc}^{T}, M_{Zc}^{T}.

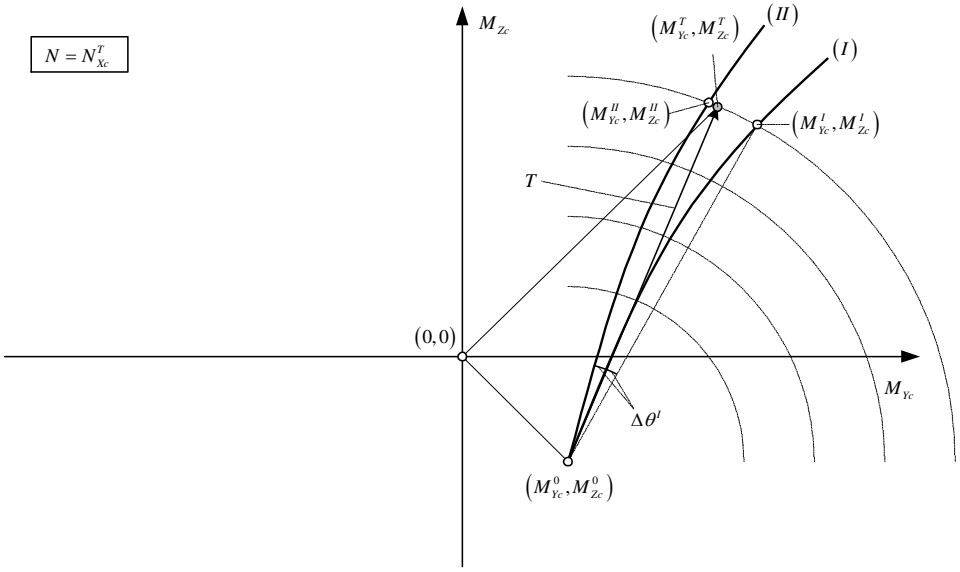


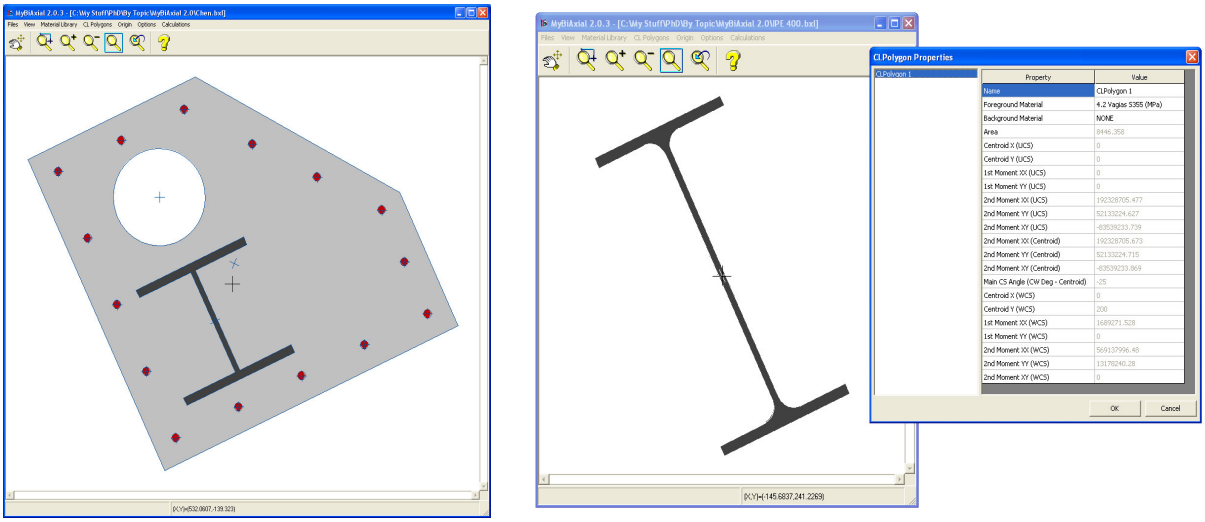
Figure 6. Calculation of deformed configuration under given loads

All calculations are performed with axial load equal to N_{Xc}^{T}. The origin can be any point; therefore, we first have to calculate the bending moments M_{Yc}^{0}, M_{Zc}^{0} required for a deformed plane with no curvature (k=0). Since curvature is always increased from zero until failure, this bending moment vector is the first result for any direction θ of the neutral axis. Therefore, the paths of all analyses stem from (M_{Yc}^{0}, M_{Zc}^{0}). The target vector T connects (M_{Yc}^{0}, M_{Zc}^{0}) with (M_{Yc}^{T}, M_{Zc}^{T}).

As first attempt (I), we set the direction θ^{I} of the neutral axis equal to the direction of the target vector T. As curvature is increased, the path of the analysis deviates because of the secondary moment M_Z; when the norm reaches the norm of the target vector, the analysis stops and the result (M_{Yc}^{I}, M_{Zc}^{I}) may differ from the target values (M_{Yc}^{T}, M_{Zc}^{T}). The direction of the neutral axis is then corrected by the difference $\Delta\theta^{I}$ found in the first iteration. In the second attempt, the results (M_{Yc}^{II}, M_{Zc}^{II}) are much closer to the target values. The procedure stops when a specified accuracy is achieved.

8 COMPUTER IMPLEMENTATION

A computer program, called myBiAxial, which implements the method proposed, has been developed. The program features a full graphical interface. Some screen shots are shown in Figure 7:



(a) (b)

Figure 7. MyBiAxial computer program

9 VALIDATION - EXAMPLES

9.1 Example 1

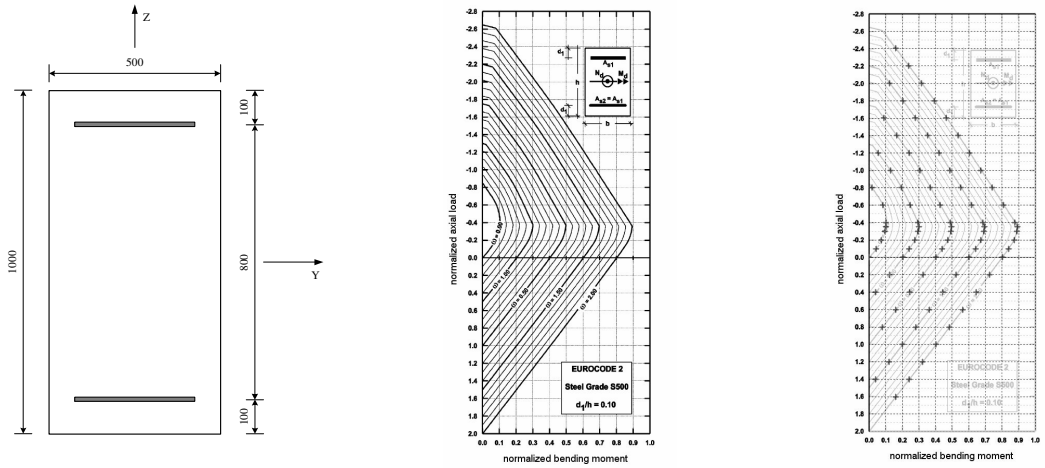
Eurocode 2 provides design charts for common reinforced concrete cross sections. These charts provide combinations of axial loads and their respective ultimate bending moment capacities for a range of longitudinal reinforcement expressed by the mechanical reinforcement percentage ω.

The axial load and bending moment are in a non-dimensional form with respect to the concrete properties and the cross sectional dimensions; therefore, a single chart covers all cases for a certain steel grade.

Eurocode 2 specifies the value of 0.020 as the ultimate strain limit for longitudinal steel reinforcement. Also, for large compressive axial loads, it reduces the ultimate curvature capacity by imposing the rotation of the strain profile around point C which is located at a distance 3/7·h from the most compressed fiber and has a strain of ε_0=-0.002. This restriction is included easily in the algorithm; however, it is of little practical interest since large compressive axial loads in concrete cross sections must be avoided for other reasons i.e. creep.

The developed computer program was used to calculate pairs of axial loads and bending moments for the rectangular cross section of Figure 8a. The characteristic strengths and partial safety factors for concrete and reinforcement bars were taken as follows: f_{ck}=20MPa, γ_c=1.5, f_y=500MPa, γ_r=1.15.

Five different cases of longitudinal reinforcement were considered, i.e. ω=0.00, 0.50, 1.00, 1.50, 2.00. The computed results follow the corresponding curve exactly, as shown in Figure 8c.



(a) (b) (c)

Figure 8. (a) Rectangular reinforced concrete cross section (distances in mm) (b) Corresponding EC2 design chart (steel grade S500) (c) Results from proposed algorithm superimposed over the design chart

9.2 Example 2

This is an example presented by Chen et al. [2], which invokes the polygonal composite column cross section of Figure 9. The cross section consists of a concrete core, an asymmetrically placed H – shaped steel section, 15 reinforcement bars of diameter 18mm and a circular opening.

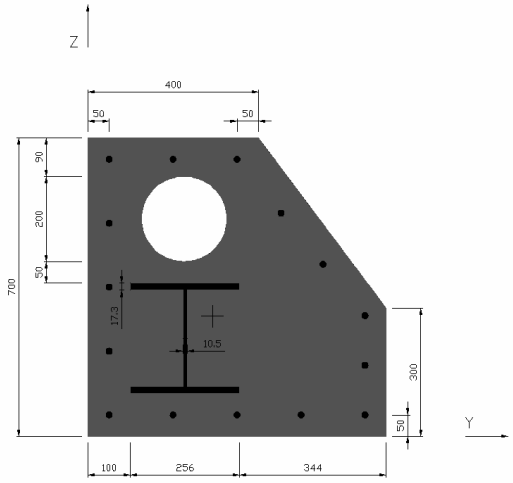


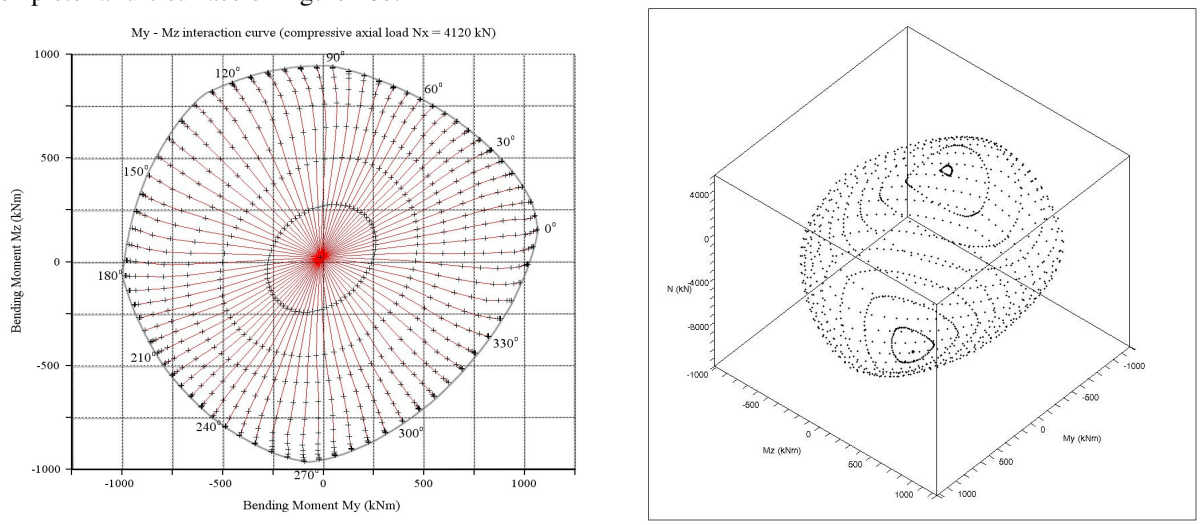
Figure 9. Composite column cross section

Chen et al. use a quasi – Newton method [10] to analyze the cross section. However, the convergence of the iterative process invoked by this algorithm cannot be guaranteed when dealing with large axial loads i.e. loads that approach the axial load capacity under pure compression. In order to ensure the stability of Chen's algorithm, the plastic centroid must be used as the origin of the Cartesian system. In this case, the coordinates of the plastic centroid with respect to the bottom left corner are [2] Y_{pc}=292.2mm, Z_{pc}=281.5mm.

The stress – strain curve for concrete (CEC 1994) which consists of a parabolic and a linear (horizontal) part was used in the calculation, with $f_{cc}=0.85\cdot f_{ck}/\gamma_c$, ε_0=0.002 and ε_{cu}=0.0035. The Young modulus for all steel sections was 200GPa while the maximum strain was $\varepsilon_u=\pm 0.010$. The characteristic strengths and partial safety factors for concrete, structural steel and reinforcement bars were taken as follows: f_{ck}=30MPa, γ_c=1.5, f_s=355MPa, γ_s=1.1, f_y=460MPa, γ_r=1.15. The analysis was carried out with an angle step of 5 degrees and an initial curvature step of 1e-05.

Figure 10a shows the interaction curve produced by the proposed algorithm for compressive axial load 4120kN. The image is superimposed over the results taken from [2]; it is obvious that the curves almost coincide.

The same figure also shows the paths of the analyses and the directions of the neutral axes that correspond to each spike. Note that the data for each spike becomes denser near failure; this is because the curvature step is decreased in order to achieve accuracy. By repeating this procedure for various axial loads we obtain the complete failure surface of Figure 10b.



(a) (b)

Figure 10. (a) Interaction curve for compressive axial load 4120 kN (b) Complete failure surface

10 CONCLUSIONS

A generic algorithm for the analysis of arbitrary cross sections under biaxial bending and axial load is presented. The algorithm has some unique features as compared to the literature. The cross section is described by curvilinear polygons, i.e. closed polygons with straight or curved edges; the material stress – strain diagrams are fully user – defined as piecewise functions of polynomial segments; the integration of the stress field is analytical. Apart from producing interaction curves and failure surfaces, the algorithm can be used for the calculation of the deformed state of the cross section under given loads.

The algorithm has proved to be very stable and fast, while providing accurate results for non-linear analysis.

REFERENCES

[1] Bonet, J.L., Romero, M.L., Miguel, P.F., Fernandez, M.A. (2004), "A fast stress integration algorithm for reinforced concrete sections with axial loads and biaxial bending", Computers and Structures 82 213-225

[2] Chen, S. F., Teng, J. G., Chan, S. L. (2001), "Design of biaxially loaded short composite columns of arbitrary cross section", J. Struct Engng, ASCE; 127(6):678-685.

[3] De Vivo, L, Rosati, L. (1998), "Ultimate strength analysis of reinforced concrete sections subject to axial force and biaxial bending", Comput. Methods Appl. Mech. Engrg. 166:261-287.

[4] Fafitis, A (2001), "Interaction surfaces of reinforced-concrete sections in biaxial bending", J Struct Engng, ASCE; 127(7):840-6

[5] Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. (2002), *Numerical recipes in C++: the art of scientific computing*, Cambridge University Press.

[6] Rodriguez, J. A., Aristizabal-Ochoa, J. Dario (1999), "Biaxial interaction diagrams for short RC columns of any cross section", J Struct Engng, ASCE; 125(6):672-683.

[7] Rodriguez, J. A., Aristizabal-Ochoa, J. Dario (2001), "M-P-φ diagrams for reinforced, partially and fully prestressed concrete sections under biaxial bending and axial load", J Struct Engng, ASCE; 127(7):763-773.

[8] Rodriguez, J. A., Aristizabal-Ochoa, J. Dario (2001), "Reinforced, partially and fully prestressed concrete columns under biaxial bending and axial load", J Struct Engng, ASCE; 127(7):774-783.

[9] Sfakianakis, M. G. (2002), "Biaxial bending with axial force of reinforced, composite and repaired concrete sections of arbitrary shape by fiber model and computer graphics", Advances in Engineering Software 33 227 - 242.

[10] Yen, J. Y. R. (1991), "Quasi-Newton method for reinforced concrete column analysis and design", J Struct Engng, ASCE; 117(3):657-666.

NUMERICAL ADVANCES IN THE LIMIT ANALYSIS OF STRUCTURES

Konstantinos V. Spiliopoulos, and Agis G. Politis

Department of Civil Engineering
National Technical University of Athens, 157-73 Athens, Greece
e-mail: kvspilio@central.ntua.gr

Keywords: Direct Methods, Limit Analysis, Spatially Varying Young's Modulus, Perforated Plate

Abstract. *Based on the two theorems of plasticity, direct methods of limit analysis provide a better alternative than time stepping procedures for the estimation of the strength of a structure. Two methods, based on the upper bound theorem, are considered in the present work and are applied to plane structures. The first method considers lumped plastic deformation along the edges of rigid finite elements. With a linearised von Mises yield criterion the problem is converted to a linear programming problem. On the other hand, a nonlinear optimization problem must be solved with the yield criterion in its proper nonlinear form. Both programs are solved using a nonlinear programming algorithm, which given a good initial guess, converges rapidly. The second method determines the limit load by a sequence of linear elastic solutions. Within an elastic solution, the stresses in regions of the body exceed the yield condition. The stresses in such locations are then reduced to a yield value by changing the elastic properties, assuming the strain field remains unchanged. A new resolution then may take place since the new stress field does not satisfy equilibrium. The procedure results to a sequence of load factors which converge to the true one. Examples of application are included in the paper.*

1 INTRODUCTION

Plastic analysis of structures is nowadays an important tool in civil engineering used to accomplish precise predictions of the limit load carrying capacity which contributes to the economy of a structure. For planar structures in particular, these types of problems occur in the plastic analysis of discs, in soil mechanics for the load carrying capacity of footings, etc.

Usually plastic analysis is carried out through the cumbersome time-stepping finite element method. There are however methods depending on the kinematic theorem of plasticity for determining the upper bound of the limit load of a structure[1-3] (called direct methods), which may give a very good approximation to the true limit load. These methods need much less computing time and also give a better insight to the inelastic behaviour of structures.

In this work two classes of direct methods are investigated. In the first part (sections 2-3), the direct methods which are based on the minimization of the load factor through a sequence of compatible modes of collapse are investigated. The methods are applied to a 2D structure. In section 4 an alternative method based on an iterative procedure which updates the modulus of elasticity, at each iteration, so that the stresses lie within the yield surface is presented. An application of the method to a simple 1D example is also included.

2 DIRECT METHODS

2.1 Kinematical equations

The plane structure is divided into an adequate number of elements, (triangular or quadrateral), whose common edges may serve as possible yield lines of the structure. It is assumed that these elements have the capability of motion and not rotation i.e. they have the capability to move like rigid bodies in the plane without rotating round their mass axis. Every possible collapse mechanism of the structure may be described by the velocity rate of the horizontal and perpendicular component of displacement on the yield line. It is also assumed that the total plastic deformation of the structure is concentrated between adjacent elements and there are no gaps or overlaps between these elements[4-5].

The collapse mechanism is kinematically acceptable if the adjacent to the yield line elements may be separated either tangentially or perpendicularly or finally by a combination of both. The perpendicular and tangential relative velocity between elements i and j which are separated by the line *m* are described through the displacement rate of the two elements $\dot{u}_i$, $\dot{v}_i$, $\dot{u}_j$, and $\dot{v}_j$ and are given by the following relation:

$$\dot{e}_n = -(\dot{u}_j - \dot{u}_i)\cdot\sin\alpha + (\dot{v}_j - \dot{v}_i)\cdot\cos\alpha \tag{1}$$

$$\dot{e}_t = (\dot{u}_j - \dot{u}_i)\cdot\cos\alpha + (\dot{v}_j - \dot{v}_i)\cdot\sin\alpha \tag{2}$$

where α is the angle between the horizontal principle axis and the common edge of the two elements, (Fig. 1).

Grouping all the above kinematical equations for all the elements which constitute the structure we may obtain the following matrix equations:

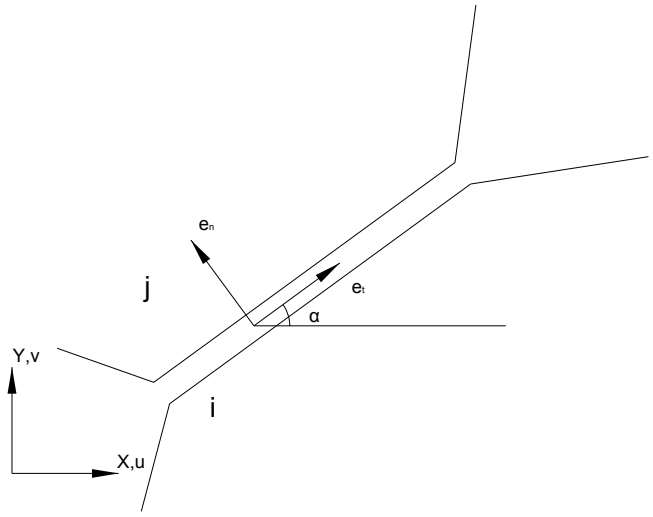


Figure 1. Plasticity lumped along common edges.

$$\dot{\mathbf{e}}_n = \mathbf{B}_n \cdot \dot{\mathbf{d}} \tag{3}$$

$$\dot{\mathbf{e}}_t = \mathbf{B}_t \cdot \dot{\mathbf{d}} \tag{4}$$

The vectors $\dot{\mathbf{e}}_n$ and $\dot{\mathbf{e}}_t$ contain the vertical and tangential rates of the relative displacements. The vector $\dot{\mathbf{d}}$ contains the rate of the displacements.

2.2 Yield criterion

The direct method uses the von Mises yield criterion which is expressed for plane structures as:

$$F = \sigma_m^2 + 3 \cdot \tau_m^2 = {\sigma_y}^2 \tag{5}$$

where σ_m is the normal stress, τ_m is the shearing stress along a yield line and σ_y is the yield stress of the material.

2.2.1 Linearization of the yield criterion

If we replace $\sigma_m = X$, $\sqrt{3} \cdot \tau_m = Y$ and $\sigma_y = R$ the above equation is transformed to an equation of a circle whose center lies at the start of the principal axes σ_m and τ_m. Instead of using the full circle one may use the circumscribed polygon. The coordinates of the vertices of this polygon are given by the following equations:

$$X_k = \frac{R \cdot \cos(a_k - \beta)}{\cos\beta} \tag{6}$$

$$X_{k+1} = \frac{R \cdot \cos(a_k + \beta)}{\cos\beta} \tag{7}$$

$$Y_k = \frac{R \cdot \sin(a_k - \beta)}{\cos\beta} \tag{8}$$

$$Y_{k+1} = \frac{R \cdot \sin(a_k + \beta)}{\cos\beta} \tag{9}$$

with $a_k = 2 \cdot k \cdot \pi$, $k = 1, 2, ..., K$, $\beta = \frac{\pi}{K}$

K is the total number of the sides of the circumscribed polygon.

With the help of the above equations we obtain the following relation which is actually the equation which describes each side of the polygon.

$$\cos a_k \cdot X + \sin a_k \cdot Y = R \tag{10}$$

Therefore the linearized yield criterion is expressed by the following equation:

$$F_k = \cos a_k \cdot \sigma_m + \sqrt{3} \cdot \sin a_k \cdot \tau_m = \sigma_y \tag{11}$$

The components of the plastic flow rate on the sides of the linearized equation are obtained by differentiating the above equation with respect to the normal and shearing stresses respectively which results to the following relations:

$$\dot{e}_{n,m} = \dot{\lambda}_m^k \cdot \frac{\partial F_k}{\partial \sigma_m} = \dot{\lambda}_m^k \cdot \cos a_k \tag{12}$$

$$\dot{e}_{t,m} = \dot{\lambda}_m^k \cdot \frac{\partial F_k}{\partial \tau_m} = \sqrt{3} \cdot \dot{\lambda}_m^k \cdot \sin a_k \tag{13}$$

2.2.2 Nonlinear yield criterion

In this case, the components of the plastic flow rates are derived straight away by differentiating (5) with respect to the normal and shearing stresses and so the following equations are obtained:

$$\dot{e}_{n,m} = \dot{\lambda}_m \cdot \frac{\partial F}{\partial \sigma_m} = 2 \cdot \dot{\lambda}_m \cdot \sigma_m \tag{14}$$

$$\dot{e}_{t,m} = \dot{\lambda}_m \cdot \frac{\partial F}{\partial \tau_m} = 6 \cdot \dot{\lambda}_m \cdot \tau_m \tag{15}$$

3 GOVERNING EQUATIONS

Denoting herein with bold letters vectors and matrices, the rate of the work produced by the external forces which act on the structure is given by the following equation:

$$\dot{W} = \mu \cdot \bar{\mathbf{f}}^T \cdot \dot{\mathbf{d}} \tag{16}$$

where μ is the load factor and $\bar{\mathbf{f}}$ is the vector of the external loads. The rate of dissipation of the plastic work along the common lines of adjacent elements is given by the relation:

$$\dot{D} = \sum_{m=1}^{M} \int_0^{l_m} (\sigma_m \cdot \dot{\varepsilon}_m + \tau_m \cdot \dot{\gamma}_m) dl_m = \sum_{m=1}^{M} \int_0^{l_m} (\sigma_m \cdot \dot{e}_{n,m} + \tau_m \cdot \dot{e}_{t,m}) dl_m \tag{17}$$

where l_m denotes the length of a common edge, M is the total number of the common edges. Using the linearized von Mises yield criterion, the above equation is transformed to the following relation:

$$\begin{aligned} &\dot{D} = \sum_{m=1}^{M} \int_0^{l_m} (\sigma_m \cdot \dot{e}_{n,m} + \tau_m \cdot \dot{e}_{t,m}) dl_m = \sum_{k=1}^{K} \sum_{m=1}^{M} \int_0^{l_m} (\sigma_m \cdot \dot{\lambda}_m^k \cdot \cos a_k + \tau_m \cdot \sqrt{3} \cdot \dot{\lambda}_m^k \cdot \sin a_k) dl_m \Leftrightarrow \\ &\Leftrightarrow \dot{D} = \sum_{k=1}^{K} \sum_{m=1}^{M} \dot{\lambda}_m^k \cdot \int_0^{l_m} (\cos a_k \cdot \sigma_m + \sqrt{3} \cdot \sin a_k \cdot \tau_m) dl_m = \sum_{k=1}^{K} \sum_{m=1}^{M} \dot{\lambda}_m^k \cdot \int_0^{l_m} \sigma_y dl_m \Leftrightarrow \\ &\Leftrightarrow \dot{D} = \sigma_y \cdot \sum_{k=1}^{K} \sum_{m=1}^{M} \dot{\lambda}_m^k \cdot \int_0^{l_m} dl_m = \sigma_y \cdot \sum_{k=1}^{K} \sum_{m=1}^{M} \dot{\lambda}_m^k \cdot l_m \end{aligned} \tag{18}$$

Equating (16) to (18) and requiring the extra constraint $\bar{\mathbf{f}}^T \cdot \dot{\mathbf{d}} = 1$, so that plastic mechanisms may exist, the problem is converted to a linear programming problem:

$$\text{Min } \mu = \sigma_y \cdot \sum_{k=1}^{K} \sum_{m=1}^{M} \dot{\lambda}_m^k \cdot l_m$$

$$\text{Subject to } \dot{\mathbf{e}}_n - \mathbf{B}_n \cdot \dot{\mathbf{d}} = \mathbf{0}$$

$$\dot{\mathbf{e}}_t - \mathbf{B}_t \cdot \dot{\mathbf{d}} = \mathbf{0} \tag{19}$$

$$\bar{\mathbf{f}}^T \cdot \dot{\mathbf{d}} = 1$$

$$\dot{\lambda}_m^k \geq 0$$

The elements of $\dot{\mathbf{e}}_n$, $\dot{\mathbf{e}}_t$ are given by equations (12) and (13) respectively. Thus the variables of the program are the rate of displacements $\dot{u}_i$ and $\dot{v}_i$ and the plastic multipliers λ_m^k .

When using the von Mises yield criterion in its proper form, the rate of the plastic work dissipation becomes:

$$\begin{aligned} &\dot{D} = \sum_{m=1}^{M} \int_0^{l_m} (\sigma_m \cdot \dot{e}_{n,m} + \tau_m \cdot \dot{e}_{t,m}) dl_m = \sum_{m=1}^{M} \int_0^{l_m} (\sigma_m \cdot 2 \cdot \dot{\lambda}_m \cdot \sigma_m + \tau_m \cdot 6 \cdot \dot{\lambda}_m \cdot \tau_m) dl_m \Leftrightarrow \\ &\Leftrightarrow \dot{D} = \sum_{m=1}^{M} 2 \cdot \dot{\lambda}_m \cdot \int_0^{l_m} (\sigma_m{}^2 + 3 \cdot \tau_m{}^2) dl_m = 2 \cdot \sum_{m=1}^{M} \dot{\lambda}_m \cdot \int_0^{l_m} \sigma_y{}^2 dl_m \Leftrightarrow \\ &\Leftrightarrow \dot{D} = 2 \cdot \sum_{m=1}^{M} \dot{\lambda}_m \cdot \sigma_y{}^2 \cdot \int_0^{l_m} dl_m = 2 \cdot \sigma_y{}^2 \cdot \sum_{m=1}^{M} \dot{\lambda}_m \cdot l_m \end{aligned} \tag{20}$$

In this case the problem is converted to a nonlinear program with a linear objective function and with constraints both linear and nonlinear:

$$\text{Min } \mu = 2 \cdot \sigma_y{}^2 \cdot \sum_{m=1}^{M} \dot{\lambda}_m \cdot l_m$$

$$\text{Subject to } \dot{\mathbf{e}}_n - \mathbf{B}_n \cdot \dot{\mathbf{d}} = \mathbf{0}$$

$$\dot{\mathbf{e}}_t - \mathbf{B}_t \cdot \dot{\mathbf{d}} = \mathbf{0} \tag{21}$$

$$\sigma_m{}^2 + 3 \cdot \tau_m{}^2 = \sigma_y{}^2$$

$$\bar{\mathbf{f}}^T \cdot \dot{\mathbf{d}} = 1$$

$$\dot{\lambda}_m \geq 0$$

The elements of $\dot{\mathbf{e}}_n$, $\dot{\mathbf{e}}_t$ are this time given by equations (14) and (15) respectively. The variables of the program are the quantities $\dot{\lambda}_m, \sigma_m, \tau_m$ along the yield lines and the rates of the displacements $\dot{u}_i, \dot{v}_i$.

3.1 SOLUTION OF THE MATHEMATICAL PROGRAMS

A critical issue is the amount of computing time for the solution of any of the two programs described above. Non-linear optimization techniques are used in this work, even for the linear program, since when using a good starting solution we get a quick convergence. An alternative method for the linear programming problem (19), the simplex method, despite its convergence in a finite number of steps, a lot of extra (artificial) variables must be introduced, for its solution, at the expense of extra computing time.

A software package[6] is used to solve the generally constrained minimization problem:

$$\text{Min } f(\mathbf{x})$$

Subject to the general (possibly nonlinear) constraints

$$c_r(\mathbf{x}) = 0, \quad 1 \leq r \leq m \tag{22}$$

And the simple bounds

$$\ell_i \leq x_i \leq u_i$$

Here f and c_r are assumed to be twice-continuously differentiable and any of the above bounds may be infinite. In our problem the bounds of our variables are either 0, or they are unbounded.

To solve the above problem the objective function and the general constraints are combined into a composite function which is called the *augmented Lagrangian function*,

$$\Phi(\mathbf{x},\nu,S,q) = f(\mathbf{x}) + \sum_{i=1}^{m} \nu_i c_i(\mathbf{x}) + \frac{1}{2q}\sum_{i=1}^{m} s_{ii}[c_i(\mathbf{x})]^2 \tag{23}$$

where the components ν_i are the Lagrange multipliers estimates, the entries s_{ii} are positive scaling factors, and q is known as the penalty parameter.

An iterative scheme is employed to accomplish the minimization of this function. This iterative scheme is made up of three steps. At the start of the j+1 iteration, Lagrange multipliers estimates $\nu_i^{(j)}$, constraint scaling factors $s_{ii}^{(j)}$ and a penalty parameter $q^{(j)}$ are given. The steps performed may be summarized, in order, as follows:

1. Test for convergence: The derivative of the Lagrangian function and the constraints are checked against some tolerances:

$$\left\|\nabla_x L(\mathbf{x}^{(j)})\right\| \le q_L$$

And (24)

$$\left\|\mathbf{c}(\mathbf{x}^{(j)})\right\| \le q_c$$

where $L(\mathbf{x},\nu) = f(\mathbf{x}) + \sum_{i=1}^{m} \nu_i c_i(\mathbf{x})$ is the Lagrangian function.

2. A quadratic model of the, generally, nonlinear objective function $\Phi(\mathbf{x})$ is built. This model contains the function, its derivative and the Hessian matrix. A trust region is also used within which we trust that the values of the quadratic model and the function will generally agree.
   The convergence of the augmented Lagrangian is guaranteed if the penalty parameter is gradually reduced to zero, almost regardless of the values of the Lagrange multiplier estimates. We may arrive at a good estimate of $\mathbf{x}^{(j+1)}$ if:

$$\left\|\nabla_x \Phi(\mathbf{x}^{(j+1)})\right\| \le \omega^{(j)} \tag{25}$$

   is satisfied for some tolerance $\omega^{(j)}$. A further test then occurs whether

$$\left\|\mathbf{c}(\mathbf{x}^{(j+1)})\right\| \le \eta^{(j)} \tag{26}$$

   holds for some other tolerance $\eta^{(j)}$. If (26) is satisfied, the penalty parameter is left unchanged but the Lagrange multiplier estimates are updated. Otherwise the penalty parameter is reduced while leaving the Lagrange multipliers estimates as they are.

3. Update the Lagrange multipliers.

3.2 NUMERICAL EXAMPLE

The procedure is applied to the limit analysis of a square plate with a circular hole of diameter of 1/10 of the side of the square.

The dimensions of the plate are 20m x 20 m x 1m. Two uniformly distributed loads parallel to the horizontal and to the vertical axes having a maximum value equal to $Pl=P2=\sigma_y$ are applied at the far ends of the plate. Due to the symmetry of the problem, only a quarter of the plate was discretized with 98 quadrilateral elements (Fig.2). The relevant boundary conditions along the two sides can be seen in the same figure.

The sequence of loading is the following: First P1 is applied as a whole with its maximum value. After that, loading P2 is augmented every time by 10% of the maximum value until it also reaches its maximum value. P1 is then decreased by portions of 10% of its maximum value until it is zeroed.

It can be seen from Fig.3 that there is little difference in the results of the two programs (19) and (21). It must be noted, nevertheless that the non-linear program (21) uses fewer amounts of variables and constraints and therefore needs less computing time. The program (19), however, seems to converge from any starting point (e.g. null variables). This is not the case for the program (21) for which some of the variables may be chosen so as to

satisfy the yield constraint in an arbitrary way, i.e. all the normal stresses are put equal to σ_y and all the other variables are zeroed.

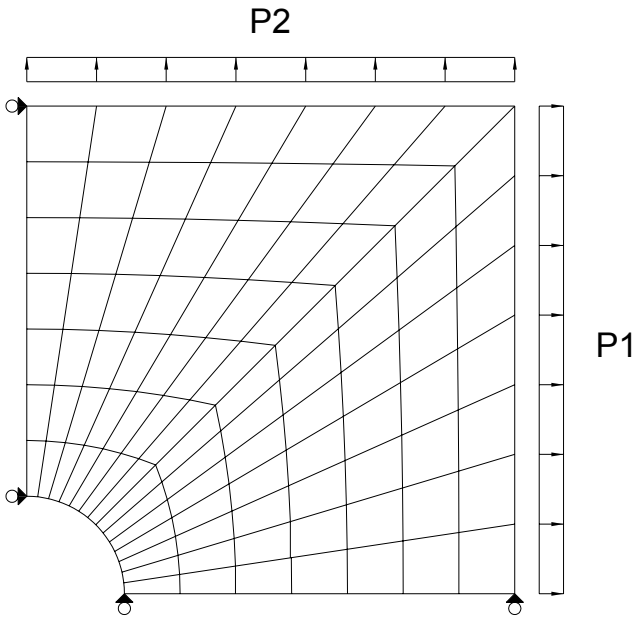


Figure 2. Finite element dicretization of the plate.

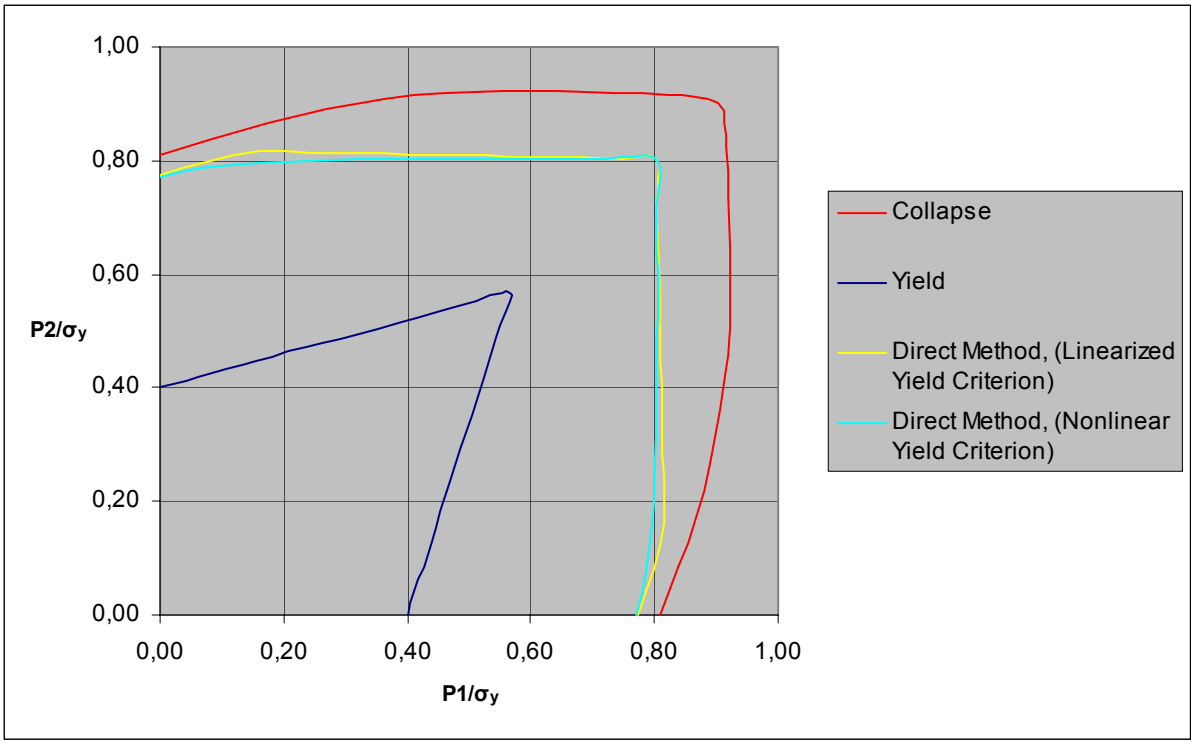


Figure 3. Results for the plate problem.

Results were compared (Fig. 3) against a time-stepping program (ABAQUS[7]) which uses an arc-length method. The same discretization was employed for both methods but the running time was approximately 80 times less for the direct methods. The difference in the value of the limit load as compared with the one of the time-stepping procedure varied between 3.7% and 12.0%. This discrepancy is expected due to the fact the direct method pre-assumes a collapse mode along the edges of the elements whereas the time-stepping method takes

into account plastification inside the elements. This discrepancy is expected to decrease with more refined discretization.

4 AN ALTERNATIVE DIRECT METHOD

There has been some interest in the recent structural engineering literature to another class of direct methods of computing limit load solutions to plasticity. These methods are also based on the kinematical theorem of plasticity and produce a sequence of lower bounds of increasing accuracy. The main essence is to adjust the elastic modules within a finite element scheme so that the stresses are brought within the yield condition at a fixed strain distribution. The elastic problem is then resolved using the new spatial distribution of elastic moduli. At each stage, a lower bound on the limit load can be found by scaling the solutions so that the stresses lie within yield for the current elastic solution. Experience has shown that a monotonically increasing sequence of lower bounds is usually obtained[8].

4.1 Illustrative example

Consider the simple problem shown in Fig. 4 where a uniaxial rod of 2l consists of two sections of length l with cross sectional areas of 2A and A and of common Young's module. The problem is treated as a displacement boundary condition and it is assumed that the total deflection rate δ is fixed. The problem has a single degree of freedom u.

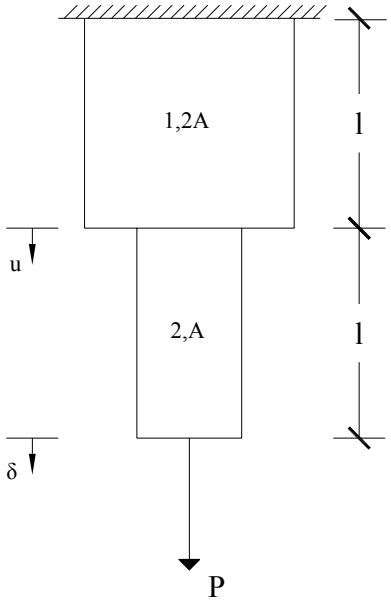


Figure 4. Geometry of the rod.

The strains for the two sections are given by:

$$\varepsilon_1 = u/l\,,\ \varepsilon_2 = (\delta - u)/l \tag{22}$$

and for an arbitrary u, through the principle of virtual work, the upper bound of the critical load for δ>0 is given by:

$$P_{UB}\cdot\delta = \begin{bmatrix} 2\cdot A\cdot l\cdot\sigma_y\cdot(u/l)+A\cdot l\cdot\sigma_y\cdot((\delta-u)/l) & \delta>u \\ 2\cdot A\cdot l\cdot\sigma_y\cdot(u/l)+A\cdot l\cdot\sigma_y\cdot((u-\delta)/l) & \delta<u \end{bmatrix} \tag{23}$$

Employing the procedure described above, the modules of elasticity have to change in every iteration. If we denote by E_1^j and E_2^j the Young's modules in the kth iteration in section 1 and 2 respectively, we obtain the following relations:

$$\sigma_2^j = 2\cdot\sigma_1^j\,,\ E_1^{j+1} = E_1^j\cdot\frac{\sigma_y}{\sigma_1^j}\,,\ E_2^{j+1} = E_2^j\cdot\frac{\sigma_y}{\sigma_2^j} \tag{24}$$

With the help of the above equations we obtain:

$$\frac{E_1^{j+1}}{E_2^{j+1}}=2\cdot\frac{E_1^{j}}{E_2^{j}} \tag{25}$$

Since $E_1^0=E_2^0=E$ then $\frac{E_1^j}{E_2^j}=2^j$

The full elastic solution at the j^{th} iteration is given by:

$$\varepsilon_1^j+\varepsilon_2^j=\frac{\delta}{l},\ 2\cdot\sigma_1^j=\sigma_2^j=\frac{P}{A},\ \varepsilon_1^j=\frac{\sigma_1^j}{E_1^j}\ \text{and}\ \varepsilon_2^j=\frac{\sigma_2^j}{E_2^j} \tag{26}$$

This may be solved to give:

$$\varepsilon_1^j=\left(\frac{1}{1+2^{j+1}}\right)\cdot\frac{\delta}{l}\ \text{and}\ \varepsilon_2^j=\left(\frac{2^{j+1}}{1+2^{j+1}}\right)\cdot\frac{\delta}{l} \tag{27}$$

So the j^{th} limit load upper bound is given by:

$$P_{UB}^j=\left(A\cdot\sigma_y\right)\cdot\left(\frac{2+2^{j+1}}{1+2^{j+1}}\right) \tag{29}$$

If $j\to\infty$ then $P_{UB}^j\to A\cdot\sigma_y$

The convergence is geometric and the error is reduced in each iteration by 50%.

The above described method seems to be a good alternative to the orthodox direct methods which are based on mathematical programming techniques as it concerns only with the solution of a sequence of elastic problems. The applicability of the method to general structures must be further investigated.

5 CONCLUSIONS

The numerical implementation of the direct methods to the limit analysis of plane structures is investigated. These methods offer a good alternative to the time consuming step-by-step methods, whenever one only requires an estimate of the strength of a structure.

The first class of these methods use a rigid finite element formulation with the plasticity lumped along the edges of the elements. Employing a nonlinear von Mises yield criterion, two different formulations are presented: one using the yield criterion in its proper form and one using it in its linearised form. Both formulations are being solved with the aid of a non-linear programming algorithm. The first formulation needs for convergence a different from the null starting solution, whereas the second one may converge from any starting solution.

The second class of direct methods is based on an iterative procedure which updates the elastic modulus at any part of the structure by the amount that the von Mises stress exceeds the yield stress under a fixed strain distribution. This approach has been applied to a simple one-dimensional structure and has shown a geometric convergence. Further investigation will be needed to clarify its applicability to general structures.

6 REFERENCES

[1] Belytschko, T. and Hodge, P. G. Jr. (1970), "Plane stress limit analysis by finite elements", ASCE Journal of the Engineering Mechanics Division Vol. 96, No. EM6, pp. 931-944.

[2] Sloan, S. W. (1989), "Upper bound limit analysis using finite elements and linear programming", Int. Jnl. for Numerical and Analytical Methods in Geomechanics 13, pp. 263-282.

[3] Sloan, S. W. and Kleeman P. W. (1995), "Upper bound limit analysis using discontinuous velocity fields", Comput. Methods Appl. Mech. Engrg. 127, pp. 293-314.

[4] Jirásek, M. and Bažant Z. (2002), *Inelastic Analysis of Structures*, John Wiley & Sons, Ltd., West Sussex, England.

[5] Spiliopoulos, K. V., Politis, A. G. (2004), "On the limit analysis of plane structures", *Proceedings of the 7th National Congress on Mechanics*, Chania, Crete, June 24 – 26, Vol. I, pp. 238 – 243.

[6] Conn, A. R., Gould, N. I. M. and Toint Ph. L. (1991), *LANCELOT: A Fortran Package for Large-Scale Nonlinear Optimization (Release A)*, Springer-Verlag, Berlin.

[7] Hibbitt, Karlsson & Sorensen Inc. (1993), *Abaqus Manual*.

[8] Ponter, A.R.S., Carter, K.F. (1997), "Limit state solutions, based upon linear elastic solutions with a spatially varying elastic modulus", Comput. Methods Appl. Mech. Engrg. 140, pp. 237-258.

NUMERICAL SIMULATION OF MULTI-BUBBLE GROWTH IN FILAMENTS UNDERGOING STRETCHING

Katerina Foteinopoulou[1,2], Vlasis G. Mavrantzas[1,2], and John Tsamopoulos[1]

[1]Department of Chemical Engineering, University of Patras, GR 26504, Patras, Greece

[2]Institute of Chemical Engineering and High Temperature Chemical Processes (FORTH-ICE/HT), GR 26504, Patras, Greece

e-mails: kfotein@iceht.forth.gr, vlasis@chemeng.upatras.gr, tsamo@chemeng.upatras.gr

Keywords: Filament stretching, cavitation, bubble growth, finite elements, bubble coalescence.

Abstract: *We present results concerning bubble deformation in Newtonian filaments confined between two disks and undergoing stretching by pulling the upper disk along its axis with a constant velocity. The governing equations consist of the momentum, continuity and constitutive equations and the free surface boundary conditions at the bubble-liquid and liquid-air interfaces. These are solved by a finite element/Galerkin method coupled with an implicit Euler for the time integrations. In addition, a robust mesh generation scheme that solves a set of partial, elliptic differential equations for the nodal point is used, which has allowed us to study multiple bubbles simultaneously deforming in the stretched filament. At each time step the flow equations are solved with the mesh equations using Picard iterations. Numerical results are presented concerning the dependence of bubble growth and deformation on the dimensionless numbers of the problem (i.e. Ca), the geometry aspect ratio, relative bubble size, and bubble-bubble separation. We also report results for the case where deforming bubbles are traveling close enough to each other to coalesce into a single bubble, and the progressive evolution of the emerging bubble.*

1. INTRODUCTION

In the bulk of materials made up by block copolymers (such as those based on styrene-isoprene triblocks or acrylates) that are extensively used nowadays as pressure sensitive adhesives (PSAs), or at their interface with a substrate, small cavities or air bubbles often develop. In real-life applications where PSA's undergo large deformations (corresponding to large values of adhesive energies), such cavities propagate along the material and, at high deformation levels, can cause fracture. Understanding therefore how these cavities affect the rheological response of the material to an imposed flow field is critical in our ability to design new polymeric materials with optimal tacky properties.[1]-[6] We recently addressed[7] the issue of single bubble growth in a Newtonian or viscoelastic medium undergoing stretching through a numerical algorithm based on the finite-element/Galerkin method and an algebraic transformation of the physical domain to the computational space. In the present study, the analysis is extended to filaments containing more (than one) bubbles closely spaced along the filament axis to allow for bubble-bubble interactions to develop during the flow. Our aim is to investigate how this interaction affects the rheological response to the imposed stretching flow, especially in filaments characterized by small aspect ratios as encountered in practice.

To account for the multiple moving boundaries accompanying the presence of a large number of bubbles in the filament, a robust coordinate transformation has been used, based on the solution of a set of elliptic partial differential equations.[8],[9],[10] The method is capable of generating a discretized mesh that optimally conforms to an entire domain that undergoes large deformations in primarily one direction (anisotropic deformations), thus producing exceptionally smooth mappings and practically quasi-orthogonal grids; further, it gives us the freedom to concentrate the coordinate curves in regions of interest where special resolution is needed. By combining the new technique with the mixed finite element method, a robust algorithm is formulated for the solution of moving boundary flow problems undergoing large deformations, such as those associated with the deformation of interacting bubbles in media subject to stretching addressed here. At the present stage of the study, the analysis is restricted to bubbles growing in Newtonian liquids.

2. PROBLEM FORMULATION- GOVERNING EQUATIONS

The mathematical formulation of the problem is described in the next figure:

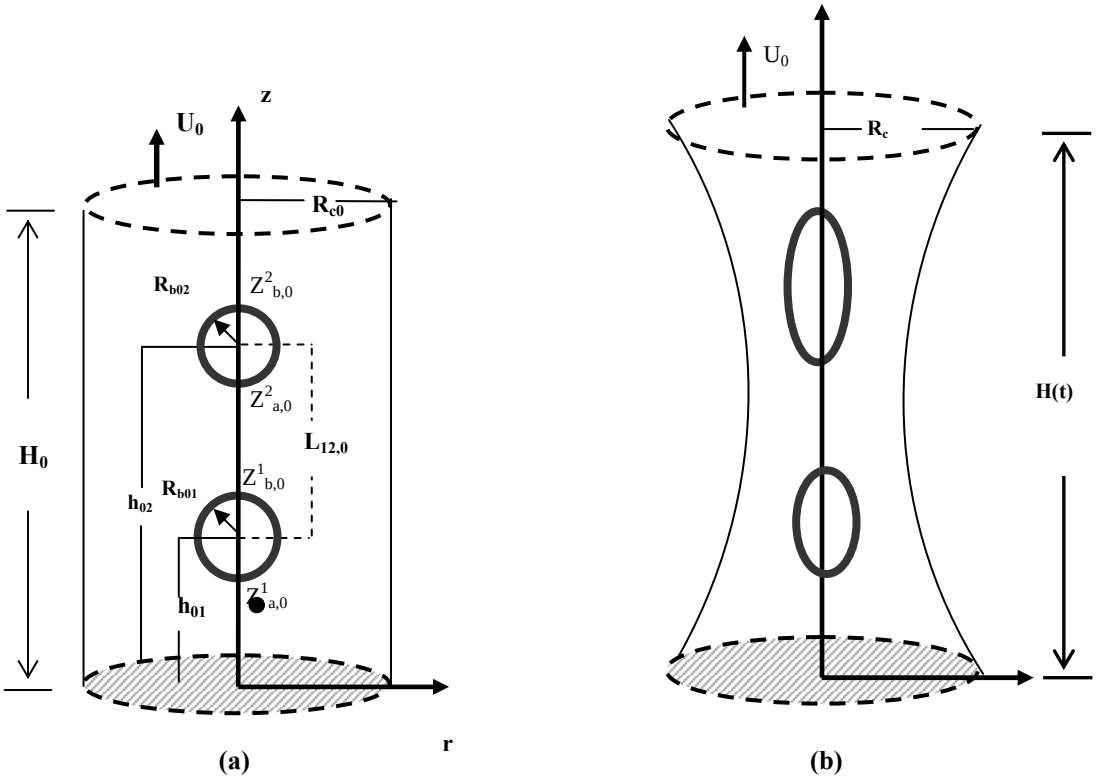


Figure 1: Geometric illustration of the system considered in this work containing two bubbles in the liquid filament: a) before stretching (t=0), b) after stretching (t>0).

The liquid filament is assumed to have initially a cylindrical outer surface with a uniform radius R_{c0} and be confined between two solid and coaxial disks, of radius also R_{c0}, separated initially by H_0. The filament undergoes deformation by pulling the upper disk with a constant velocity U_0, while the lower disk, onto which the filament is permanently bonded, remains always stationary. Inside the filament, we assume the existence of N bubbles having initially the shape of a sphere of radius R_{b10} , $R_{b20\ ,\ ...,}$ R_{bN0}, and located at distances h_{01}, h_{02} , …, h_{0N}, above the solid substrate, respectively. The corresponding successive, initial separations of the bubbles are denoted as $L_{ij,0}$, i,j=1,2,.., N, with $i \neq j$. Based on the above definitions, the following geometric dimensionless numbers arise: $e_i=R_{bi0}/R_{c0}$, $e_{ij}=R_{bi0}/R_{bj0}$, and $h_i=h_{0i}/H_0$, $a_i= R_{bi0}/H_0$, i=1,2,…N. A filament aspect ratio $\Lambda=H_0/R_{c0}$ is also defined. Due to the assumption of axial symmetry (see below), all N bubbles are assumed to lie along the axis of symmetry of the filament. A schematic of the system considered is provided in Fig. 1.

As the upper plate is being pulled, the height $H(t)$ of the filament increases, the bubbles (following the deformation of the surrounding medium) deform or translate along the filament axis, and all (bubble-liquid and liquid-air) interfaces are distorted. The largest deformations are anticipated to appear in the regions closest to the two disks (in particular, near the corners) and around the surfaces of the bubbles. The problem is formulated in cylindrical coordinates (r,θ,z) with the centre of the coordinate system placed at the centre O of the lower disk of the cylinder.

The governing equations consist of the conservation equations for the filament momentum and mass, and the evolution or constitutive equation for the stress tensor. In dimensionless form and under the assumptions of negligible gravitational forces and fluid incompressibility, they are written as:

$$\nabla \cdot \boldsymbol{u} = 0 \tag{1}$$

$$Re\left(\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u} \cdot \nabla \boldsymbol{u}\right) = -\nabla P + \nabla \cdot \boldsymbol{\tau} \tag{2}$$

$$\boldsymbol{\tau} = \nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^{\mathrm{T}} \tag{3}$$

where $\boldsymbol{\tau}$ denotes the deviatoric part of the total stress tensor $\boldsymbol{\sigma}$ ($\boldsymbol{\sigma}$ = - $P\mathbf{I}$+ $\boldsymbol{\tau}$), and Re is the Reynolds number.

The above set of equations is closed with the specification of the appropriate set of initial and boundary conditions. The shapes of the moving boundaries are determined by invoking the kinematic condition:

$$\frac{\partial \boldsymbol{R}(\mathrm{z},t)}{\partial t} + \boldsymbol{u} \cdot \nabla \boldsymbol{R}(\mathrm{z},t) = \boldsymbol{u} \tag{4}$$

where $\boldsymbol{R}(z,t)$ describes either the bubble or the liquid-air free surfaces. At the bubble-liquid and liquid-air interfaces, a normal force balance results into:

$$\boldsymbol{n} \cdot (\boldsymbol{\tau}_G - \boldsymbol{\tau}_L) = -\left(2\Re/Ca\right)\, \boldsymbol{n} + (P_G - P_L)\, \boldsymbol{n} \tag{5}$$

where $\Re$ denotes the curvature of the free surface, $\boldsymbol{n}$ is the unit vector normal to the free surface, while $Ca=(\eta U_0)/\sigma$ is the capillary number (σ: surface tension, η: medium viscosity, U_0: pulling velocity). Initially, we assume that the liquid and the gas in the bubbles are under conditions of hydrostatic and thermodynamic equilibrium.

An equation of state describing a reversible adiabatic axpansion is also assumed for the state of the gas inside the bubble:

$$P_b\, V^{\gamma} = \text{const.} \tag{6}$$

3. NUMERICAL METHOD

In the framework of the elliptic grid generation scheme, the two dimensional time-varying physical domain (r,z,t) is mapped onto a constant computational domain(ξ,η,τ) which is chosen to be the volume occupied by the fluid in a filament of radius R_{c0} and height H_0 in the absence of any bubbles ($0\le\xi\le R_{c0}$, $0\le\eta\le H_0$). Following Dimakopoulos and Tsamopoulos[10] this mapping is achieved by solving at each time step the following PDEs:

$$\nabla \cdot \nabla \xi = Q \qquad \text{(equation generating the } \eta\text{-curves)} \tag{7}$$

$$\nabla\left(\varepsilon_1 \sqrt{\frac{r_\eta^{\,2} + z_\eta^{\,2}}{r_\xi^{\,2} + z_\xi^{\,2}}} + (1-\varepsilon_1) \right) \nabla \eta = 0 \quad \text{(equation generating the } \xi\text{-curves)} \tag{8}$$

The Q term in equation (7) is used to densify the mesh in the regions near the bubbles. The mesh equations (7) and (8) are solved with appropriate boundary conditions that can also affect the quality of the generated mesh. The above transformation equations generates a duscretization mesh that optimally conforms to an entire physical domain that undergoes large anisotropic deformations.

To decouple flow (1)-(3) and mesh equations (7)-(8), a Picard type Gauss-Seidel iterative method was used with each sub-problem being solved with Newton-Raphson iterations and a 10^{-9} tolerance for the absolute norm of the residual vector. The tolerance in the Gauss-Seidel technique was set to 10^{-8}. Implicit Euler was employed for the time integration, whose time step was adapted in the course of the calculations based on the estimate of the local truncation error (defined by the difference between accurately approximated and explicitly predicted values). The flow and mesh equations are discretized by a mixed Galerkin finite element method, in the sense that the velocity field is approximated by using biquadratic, the pressure field bilinear and the position vector (r,z) at each nodal point linear Langrangian basis functions.[11]

4. RESULTS

With the numerical method described above, numerical results have been obtained[12] concerning the time evolution of the bubbles and the filament as a function of the parameters defining the kinematics of the imposed flow, such as the aspect ratio of the geometry, the value of the capillary number, etc. It was observed that for large enough aspect ratios, the bubbles experience large deformations mainly along the axial direction, deforming into fibrils. At moderate aspect ratios ($\Lambda\sim0.5$), in addition to axial deformation, a weak radial expansion is also observed. In fact, as the aspect ratio is decreased to rather small values ($\Lambda<0.2$) large bubble deformations are recorded both radially and axially. This result, which is also confirmed experimentally,[2] is explained in terms of the high values of the lateral stress σ_{rr} developing in filaments with small aspect ratios. Typical instantaneous snapshots obtained from the present simulations with a system containing 5 bubbles in a filament with aspect ratio $\Lambda=0.5$ are shown in Fig. 2. The corresponding force-vs.-time profile required to maintain the imposed flow is shown in Fig. 3.

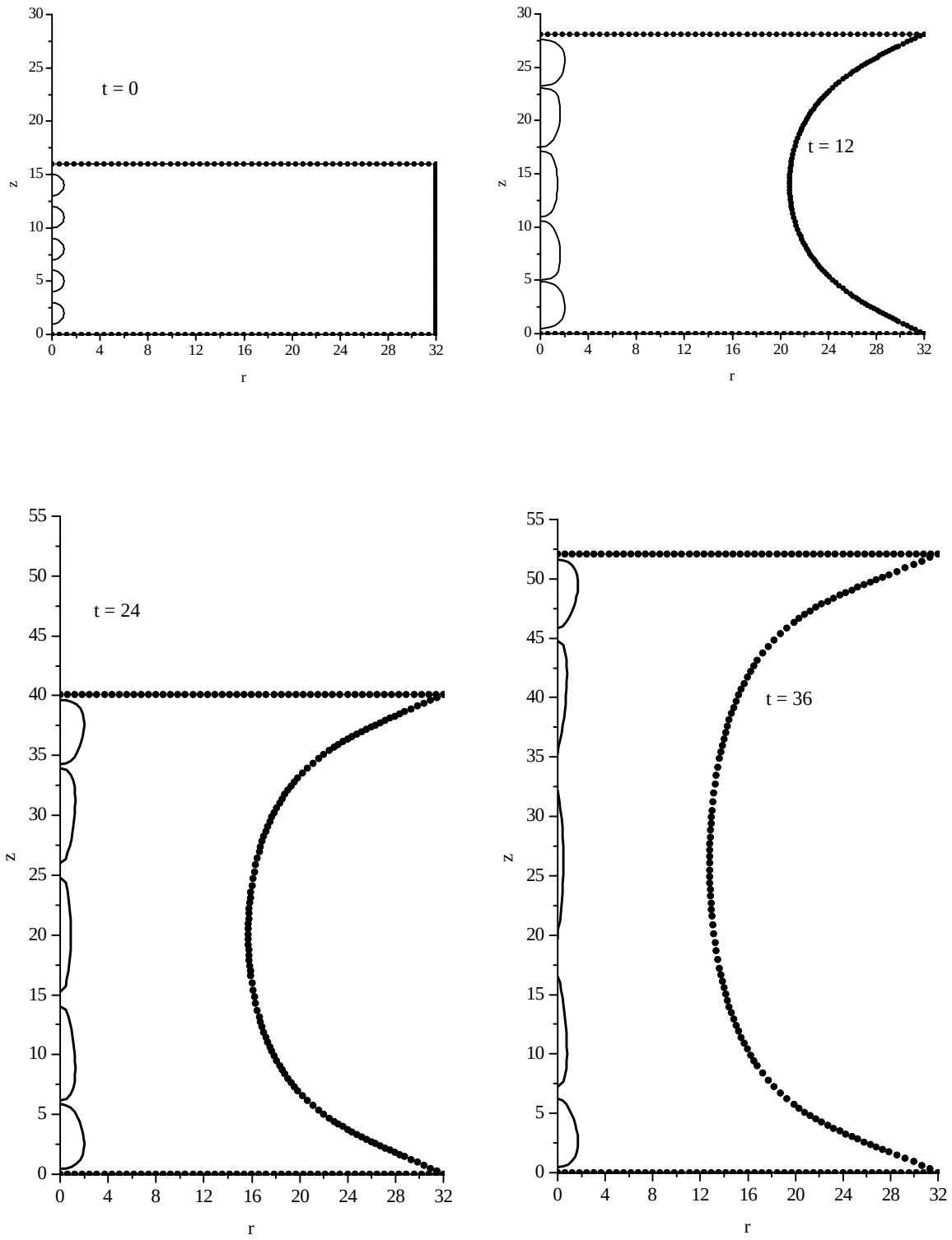


Figure 2. Typical plots of the instantaneous bubble-liquid and filament-air interfaces as obtained from the present calculations with a system containing 5 bubbles. In all cases: Λ=0.5, *Ca* = 10, a_1=a_2= …=a_5=0.0625, *Re*=0.

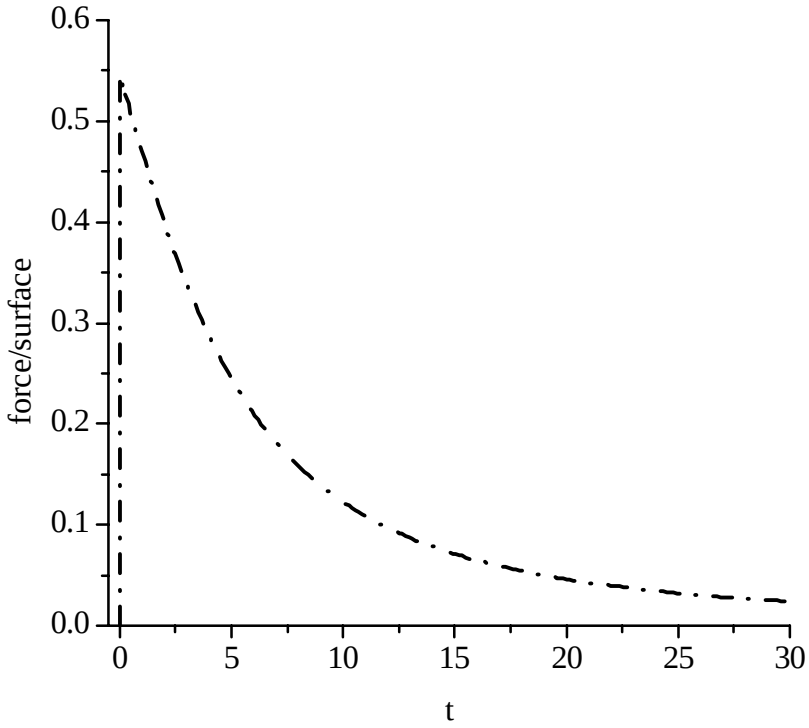


Figure 3. Time profile of the axial force on the upper disc required to keep the flow field discussed in Figure 2.

At rather high values of the applied pulling velocity, neighboring bubbles are observed to come so close that they usually coalesce to form a single, larger bubble.[1],[3],[13] In our computational experiments, bubble coalescence was allowed only for bubbles whose separation L_{ij} in the course of the "experiment" attained values smaller than about $0.2L_{ij,0}$, where $L_{ij,0}$ is their initial separation. Once two bubbles are united, the corresponding nodes in the physical domain are reconstructed by solving the mesh equations (7)-(8), together with appropriate boundary conditions. The new value of the pressure in the daughter bubble was calculated by invoking the assumption of isothermal expansion for ideal gases.

$$P_b = \frac{P_{b1}V_1 + P_{b2}V_2}{\left(V_1 + V_2\right)} \tag{9}$$

A typical example of bubble coalescence is shown in Fig. 4. How the emerging bubble evolves evolves is also shown in the figure.

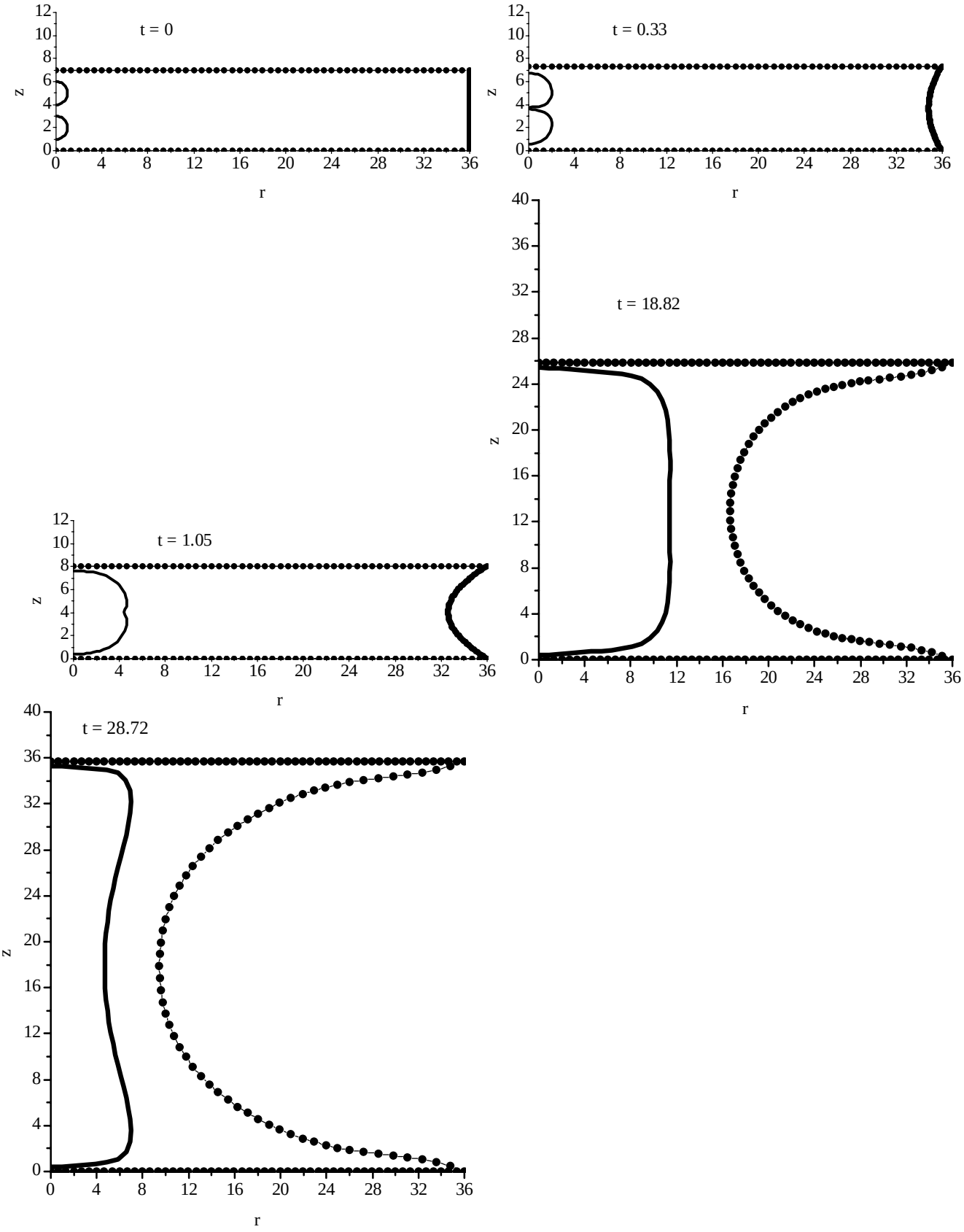


Figure 4: Time evolution of a system containing two (N=2) bubbles before and after bubble coalescence [Ca=1, Λ=0.195, h_1=0.286 , h_2=0.714, a_1=a_2= 0.143].

5. CONCLUSIONS

We have developed an efficient and robust numerical algorithm, based on the solution of a set of elliptic partial differential equations for the mesh discretization and a mixed finite element method for the solution of the discrete equations, that has allowed us to address multi-bubble growth and dynamics in Newtonian liquids undergoing stretching, under the assumption of axial symmetry (two-dimensional flow). The method has allowed us to address important issues governing or connected with bubble dynamics, such as the effect of the geometry aspect ratio. For a given value of the capillary number, a characteristic value of the aspect ratio is calculated below which bubbles exhibit significant deformations not only axially (along the axis of pulling) but also radially. As the filament aspect ratio decreases, the adhesion strength, quantified, e.g., by the maximum in the force measured on the upper plate, exhibits a substantial increase.

We have also investigated issues related to bubble interaction and bubble coalescence during stretching and how this affects their growth and the measured force. Interestingly enough, the presence of more than one bubble in the filament was not seen to affect the force needed to maintain the flow. The force-versus-time profiles were also seen to be insensitive to bubble coalescence. Whether or not this persists in viscoleastic liquids too and/or is an artifact of the axial symmetry assumption (that renders the flow two-dimensional) will be addressed in a future work. Relaxing the assumption associated with the symmetry with respect to the azimuthal direction, in particular, requires the development of parallel computational methodologies,[14]-[16] and will also be left as a future plan.

ACKNOWLEDGEMENT
The project is co-funded by the European Social Fund and National Resources.

REFERENCES

[1] Zosel A. (1989), “Adhesive Failure and Deformation Behaviour of Poymers”, J. Adh. 30, pp. 135-149.
[2] Crosby A.J., Shull K.R., Lakrout H., Creton C. (2000), “Deformation and failure Modes of adhesively Bonded Elastic Layers”, J. Appl. Phys. 88, pp. 2956-2966.
[3] Lakrout H., Sergot P., Creton C. (1999), “Direct Observation of Cavitation and Fibrillation in Probe Tack Experiment on Model Pressure-Sensitive-Adhesives”, J. of Adh. 69, pp. 307-359.
[4] Gay C., Leibler L. (1999), “Theory of Tackiness”, Phys. Rev. Lett. 82, pp. 936-939.
[5] McKinley G.H., Sridhar T. (2002), “Filament-Stretching Rheometry of Complex Fluids”, Annu. Rev. Fluid Mech. 34, pp. 375-415.
[6] Kezios P.S., Schowalter W.R. (1986), Rapid Growth and collapse of single bubbles in polymer solutions undergoimg shear, Phys. Fluids 29, pp. 3172-3181.
[7] Foteinopoulou K., Mavrantzas V.G., Tsamopoulos J. (2004), “Numerical simulation of bubble growth in Newtonian and viscoelastic filaments undergoing stretching”, J. Non-Newt. Fluid. Mech. 122, pp. 177-200.
[8] Christodoulou K.N., Scriven L.E. (1992), “Discretization of Free Surface Flows and Other Moving Boundary Problems”, J. Comp. Phys. 99, pp. 39-55.
[9] Tsiveriotis K., Brown R.A. (1992), “Boundary-conforming mapping applied to computations of highly deformed solidification interfaces”, Int. J. Num. Meth. Fluids 14, pp. 981-1003.
[10] Dimakopoulos Y., Tsamopoulos J. (2003), “A quasi-elliptic transformation for moving boundary problems with large anisotropic deformations”, J. Comp. Phys. 192, pp. 494-522.
[11] Zienkewicz O.C., Taylor R.L.(1994), *The Finite Element Method, Solid and Fluid Mechanics, Dynamics and Non-Linearity*, Vol. 2, McGraw-Hill, London.
[12] Foteinopoulou K., Mavrantzas V.G., Dimakopoulos Y., Tsamopoulos J., to be submitted.
[13] Roos A., Creton C., Novikov M.B., Feldstein M.M. (2002), “Viscoelasticity and Tack on Poly(Vinyl Pyrrolidone)-Poly(Ethylene Glycol) Blends”, J. Polym. Sci.; Part B: Polym. Phys. 40, pp. 2395-2409.
[14] Liou T.M., Shih K.C., Chau S.W., Chen S.C. (2002), “Three-dimensional simulations of the droplet formation during the inkjet printing process”, Int. Comm. Heat Mass Transfer 29, pp. 1109-1118.
[15] Bogaerds A.C.B., Verbeeten W.M.H., Peters G.W.M., Baaijens F.P.T (1999), “3D viscoelastic analysis of a polymer solution in a complex flow”, Comp. Meth. Appl. Mech. Eng. 180, pp. 413-430.
[16] Dolean V., Lanteri S. (2004), “Parallel multigrid methods for the calculation of unsteady flows on unstructured grids: algorithm aspects and parallel performances on clusters of PCs”, Parallel Computing 30, pp. 503-525.

FE-BUI – FINITE ELEMENT BEOWULF USER INTERFACE: A HOMEMADE PACKAGE FOR AUTOMATED PARALLEL FINITE ELEMENT COMPUTATIONS

Antony N. Spyropoulos, Athanasios G. Papathanasiou, John A. Palyvos, and Andreas G. Boudouvis*

School of Chemical Engineering
National Technical University of Athens
Zografou Campus, 15780 Athens, Greece
*e-mail: boudouvi@chemeng.ntua.gr, web page: http://www.chemeng.ntua.gr/dep/boudouvis

Keywords: Beowulf Clusters, Parallel Computations, Finite Element, Krylov, Preconditioning, Message Passing

Abstract. *A computational package, the FE-BUI, is introduced for automated parallelization of finite element codes. The package has been developed at the Computer Center of the School of Chemical Engineering of NTUA. Its scope is to provide the ability for parallel execution of serial codes on Beowulf clusters, in an easy and efficient way.*

1 INTRODUCTION

The main computational cost of the finite element codes comes from the solution of large linear algebraic equation systems. Direct (e.g. frontal-type) solvers require large memory and computational cost, and most importantly, their serial parts lack the advantage of exhibiting appreciable parallel efficiency. Recent products of the development of parallelization tools for finite element codes are freely available packages such as the partitioner parMETIS[1] and the solvers Aztec[2] and PETSc[3]. Even with these tools, the effort and cost required for parallelization of a serial code might be prohibitively high. The other alternative, the automated parallelization of serial codes with parallel compilers, yields no more than 10% reduction of the computational cost.

In this paper is presented a new, homemade, parallel package for the solution of finite element problems. This package offers convenience and effectiveness in doing large scale computations since there is no need for the user to learn and implement suitable solvers and communication protocols in parallel computer architectures. The user simply calls, from the serial code, the parallel solver, which takes care of the mesh partitioning, of the load assignment to the available processors and of the parallel solution of the resulting linear systems. The package uses parallel iterative solvers that are based on Krylov projection methods[4] and exploits the architecture of Beowulf clusters using the MPI (Message Passing Interface)[5] for the processors communication. Typical runs with 3D finite element problems on a small, 4-processor cluster yield a reduction of the computational cost by a factor of 3.

2 BEOWULF CLUSTERS

The outcome of the evolution, during the last decade, of the hardware of the personal computers (PCs), mainly in processors, driven by the major companies in this field, Intel and AMD, but also in motherboards and memories, is low-cost and high-performance personal computing. Moreover, commodity computer networks offer high bandwidth and low latency, like Fast Ethernet, Gigabit Ethernet and the more advanced networks Myrinet (http://www.myri.com) and SCI (http://www.dolphinics.com). This progress in conjunction with the development of Linux (http://www.linux.org), a freely available, stable and reliable operating system, enables large scale computations on Beowulf clusters (http://www.beowulf.org). These clusters are computational systems that consist of PCs which are interconnected with a private network.

Beowulf clusters are distributed memory parallel computers, where each processor has a private memory and does not have direct access to the memory of the other processors. Thus, a two-processor communication is required when a processor needs data residing in the memory of another processor. This communication can be done with the MPI, which is a library of subroutines that a programmer calls from a C or a Fortran code. In this case, the parallel execution of a serial code on Beowulf clusters needs the explicit programming of the communication between the processors. This is the main obstacle for the user in converting a serial code to a suitable, for Beowulf clusters, parallel code.

3 THE FE-BUI PACKAGE

The FE-BU is programmed in Fortran 77. Its installation requires the freely available libraries: BLAS[6], LAPACK[7] and MPICH[8] or LAM/MPI[9]. The main components are shown in Figure 1.

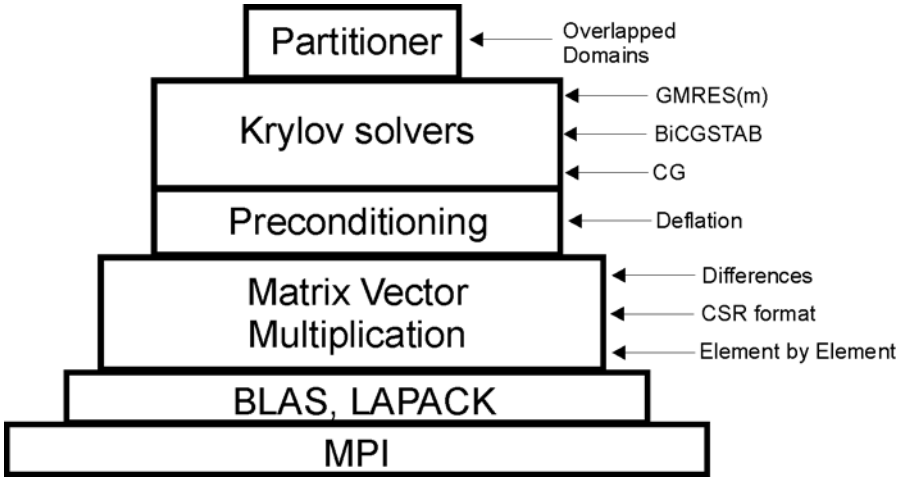


Figure 1. The FE-BUI components

3.1 KRYLOV ITERATIVE SOLVERS

The computational kernel of the FE-BUI is based on the preconditioned Krylov iterative solvers for the solution of large and sparse linear systems, such as the GMERS(m), BiCGSTAB and CG; currently, only the GMRES(m) solver is employed.

The GMRES is preferred for the iterative solution of large algebraic equation sets with non-symmetric matrices, on the basis of its parallel efficiency[10]. Starting from an initial guess, x_o, of the solution of the linear system:

$$Ax = b \tag{1}$$

where $A \in \mathbf{R}^{NxN}$ and $x, b \in \mathbf{R}^{NxN}$, GMRES uses Arnoldi's method[11], combined with an orthogonalization technique – the Modified Gram-Schmidt method is used here – to construct an orthonormal basis $V_m \in \mathbf{R}^{Nxm}$ of the m-dimensional Krylov subspace

$$K_m(A, v) = \text{span}\left\{v, Av, A^2v, ..., A^{m-1}v\right\} \tag{2}$$

where $v \equiv r_o / \|r_o\|_2$, $r_o \equiv b - Ax_o$. The new approximation of the solution is

$$x_m = x_o + V_m y_m \tag{3}$$

where y_m is a vector of size m and it is computed from the solution of the least squares problem

$$y_m = \underset{y}{\text{argmin}} \left\| \beta e_1 - \overline{H}_m y \right\|_2, \; y \in \mathbf{R}^m \tag{4}$$

In eq. (4), $\beta \equiv \|r_o\|_2, e_1 = [1,0,...,0]^T$ and $\overline{H}_m \in \mathbf{R}^{(m+1)xm}$ is an upper Hessenberg matrix, such as

$$AV_m = V_{m+1}\overline{H}_m \Rightarrow V_m^T A V_m = H_m \tag{5}$$

$H_m \in \mathbf{R}^{mxm}$ is an upper Hessenberg matrix obtained from the $\overline{H}_m$ by deleting its last row. The least squares problem (4) is solved by transforming $\overline{H}_m$ into an upper triangular matrix $R_m \in \mathbf{R}^{mxm}$ using plane rotations[4].

The storage requirements and the computational cost of Arnoldi's method increase rapidly with m and, thus, a restarting variant of the GMRES – the GMRES(m) – is used in practice. When m reaches a certain preset value, the algorithm restarts, using the last approximation x_m from eq. (3) as a new initial guess. Thus, two iterations are performed: the "inner" m iterations and the "outer" iterations that correspond to the restarts of the

GMRES(m).

3.2 PRECONDITIONING

A preconditioner is essential in enhancing the convergence rate of a Krylov iterative solver. Thus, the original linear system (1) must be transformed to an equivalent one that has better convergence properties. In the FE-BUI the linear system (1) is preconditioned from the right

$$AM^{-1}z = b,\ x = M^{-1}z \tag{6}$$

In eq. (6), z is a vector of size N and $M^{-1} \in \mathbf{R}^{NxN}$ is the preconditioner matrix which is constructed from a deflation technique[12] and it is given by

$$M^{-1} = I_N + U(|\mu|T^{-1} - I_r)U^T \tag{7}$$

where $\mu \in \mathbf{R}$ is the largest eigenvalue of the matrix A, $I_N \in \mathbf{R}^{NxN}, I_r \in \mathbf{R}^{rxr}$ are identity matrices, $U \in \mathbf{R}^{Nxr}$ is an orthonormal basis of the r-dimensional invariant subspace, P_r, corresponding to the r smallest eigenvalues (in terms of the absolute value of their real parts) of the matrix A and $T \in \mathbf{R}^{rxr}$ such as

$$T = U^T AU \tag{8}$$

The largest eigenvalue and the Schur vectors of the matrix A, needed in eq. (7), are approximated by those of the Hessenberg matrix H_m. Thus, at each restart of the GMRES(m), a Schur decomposition of the Hessenberg matrix is performed to approximate the largest eigenvalue and the Schur vectors corresponding to the smallest eigenvalues of the matrix A. These vectors are added to P_r, increasing its dimension. In order to save on computational cost and memory requirements arising from the preconditioning operations, an upper limit, r_{max}, on r is set; when it is reached, the update of the preconditioner stops and the GMRES(m) continues with the same preconditioner. The key idea of this preconditioning technique is to remove by deflation the smallest eigenvalues of the matrix A that cause slow or no convergence of the GMRES(m)[13].

3.3 PARTITIONING

The partitioner of the FE-BUI package is based on an overlapped domains[14] partitioning technique. The original domain, tessellated by the finite element method, is divided into subdomains. A subdomain is defined as a separate group of elements and it is assigned to a processor.

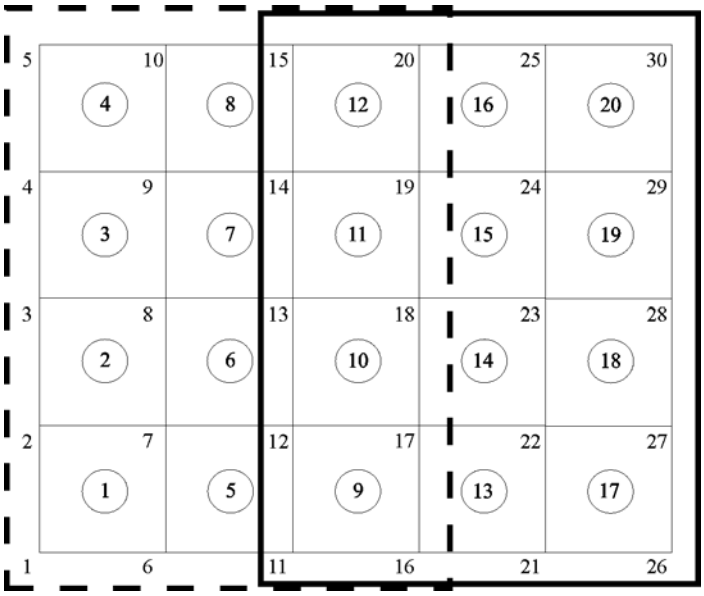


Figure 2. A sample 2D finite element mesh assigned to 2 processors. Dashed and solid lines depict the 2 overlapped subdomains.

In Figure 2 is shown a 2D finite element mesh that is assigned to 2 processors. According to this partitioning

technique, each processor takes 15 nodes, i.e. 15 rows of the matrix A, 1 to 15 for the first processor and 16 to 30 for the second processor. Each processor in order to fully assemble every local node contribution to the matrix A, makes some extra computations to the common elements 9 to 12. The nodes of these elements are called communication nodes.

Thus, in FE-BUI the decomposition of the finite element mesh corresponds to the distribution of the matrix A rows to the processors. It is known that this technique leads to smaller parallel efficiency than that of other domain decomposition techniques[14],[10] (see also section 5), but offers flexibility and usage convenience to the FE-BUI package.

3.4 PARALLEL OPERATIONS IN THE FE-BUI

The basic operations of the GMRES method are: (i) Vector updates, (ii) inner products, (iii) matrix-vector products. Moreover, preconditioning operations are needed for the preconditioned GMRES(m). The performance of these operations depend on the choice of the preconditioner. The deflation preconditioning technique can be analyzed in the same basic operations as the GMRES method. All these operations can be decomposed in tasks and each task can be independently executed on each processor. More details about the parallel implementation of these operations on Beowulf clusters are available in [10]. Briefly, vector updates can be done in parallel without communication between the processors. In order to compute an inner product each processor computes a local inner product. The latter operation is completed through a global communication between the processors to sum up the calculated local inner products. During the global communication, the processors exchange data of length of 8 bytes (the scalar local inner product). For the matrix-vector product, a communication between its processor and its neighbors is required. During the neighboring communication the processors exchange arrays of length equal to the communication nodes. For example, with reference to figure 2, the processors exchange data of length of $5 \cdot 8 = 40$ bytes.

In FE-BUI the matrix-vector product can be done using three methods: a) Compressed Sparse Row Format (CSR) were only the nonzero elements of the matrix A are stored. b) The element-by-element matrix-vector product and c) With a matrix-free approach[15], that is by approximating the elements of A by differencing; this is often the case in nonlinear problems, where A is the Jacobian matrix corresponding to the discretized equations.

4 USAGE OF THE FE-BUI

As a first step with the FE-BUI package, the new user can simply call from his/her serial code the driver subroutine FEBUIdrv replacing the serial solver call. This driver takes care of the solution of the linear system by calling the default partitioning and solver subroutines. At a more advanced level, the user can call selectively the appropriate subroutines.

The required input data of the FEBUIdrv are already computed in most finite element codes. Typical input data are:

a) the total number of the nodes (NN) and the elements (NE) of the mesh,
b) the array NpE of dimension NE – NpE(NE) – that contains the number of nodes of each element,
c) the array NOP(NE,max(NpE)) that associates the local (element level) and global (mesh level) numbering of nodes,
d) the arrays NCOD(NN) and BC(NN), for distinguishing nodes bearing Dirichlet boundary conditions,
e) an approximation to the solution u(NN), in case of nonlinear equation systems.

Also the user must supply the subroutine that computes the element contributions and the right hand side of the linear system.

5 RESULTS

The parallel efficiency of an algorithm is measured by the parallel speedup[14],[16], S, which indicates how faster the algorithm runs using p processors compared to the performance on one processor:

$$S = \frac{\text{Execution time on 1 processor}}{\text{Execution time on p processors}} \tag{9}$$

Ideally, a parallel algorithm must run p times faster when executed on p processors. However, the speedup is limited by the ever-present serial tasks in a parallel algorithm, by the load balancing and by the communication between the processors that is the main factor of a reduced parallel speedup.

The FE-BUI has been tested on the solution of a three-dimensional, nonlinear and free boundary problem of intefacial magnetohydrostatics[17]. The achieved speedup of the preconditioned GMRES(m) is shown in table 1 for two cases:

(i) parallel computations with FE-BUI,
(ii) parallel computations with a parallel code that was developed in [18].

The computations were done on a linear system of 254,857 nodes at a small Beowulf cluster of 4 nodes (http://www.chemeng.ntua.gr/yk/cluster)

CPUs	Case (i)	Case (ii)
1	1	1
2	1.8	1.9
3	2.7	-
4	3.4	3.9

Table 1: Parallel speedup

The achieved speedup is smaller than the ideal, in both cases, due to the communication between the 4 processors. The speedup in (i) is smaller than in (ii) since FE-BUI "ignores" particular aspects[18] of the mesh, resulting in a slightly unbalanced distribution of the mesh to the available processors; in such a case, the under-loaded processors have to remain idle until the over-loaded processors finish their tasks.

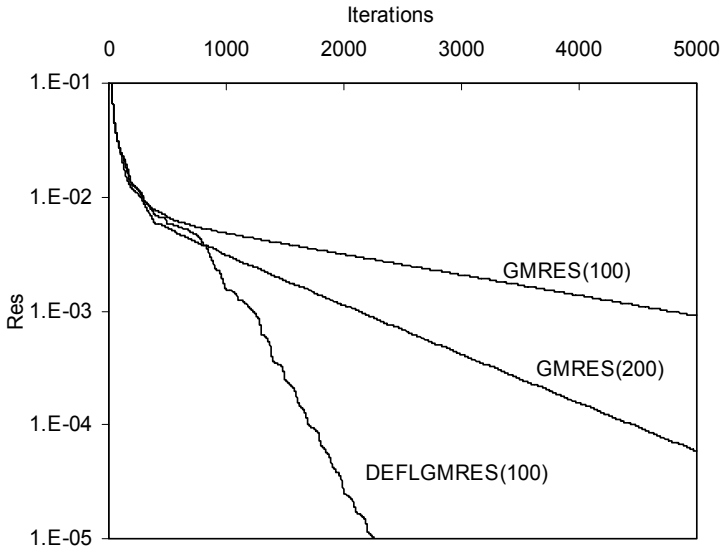


Figure 3. The convergence of the GMRES(m), m = 100, 200 and the DEFLGMRES(100), $r_{max} = 20$.

In Figure 3 is shown the convergence of the GMRES(m) and the preconditioned GMRES(m) by deflation – DEFLGMRES(m). The horizontal axis is the product of m by the restarts of the GMRES(m) or DEFLGMRES(m). The vertical axis is the residual $\text{Res} = \frac{\left\|b - Ax_m\right\|_2}{\left\|b\right\|_2}$.

Another important feature of the preconditioner of the FE-BUI is that the additional communication load coming from the extra preconditioning operations, has no appreciable effect on the speedup of the GMRES(m) – a significant advantage of the chosen preconditioner compared to a commonly used ILU-type preconditioner[10].

The achieved speedup versus the number of processors for two problem sizes, N=68,377 and N=254,857 is shown in figure 4. The speedup increases with the number of processors significantly faster in big problems than in smaller ones, because in the former case the computational time increases faster than the communication time, as it is noticed also from figure 5. The latter figure shows the relative communication (global and neighbouring) and computational time as percentages of the total execution time of the parallel preconditioned GMRES(m) versus the problem size, when the solver runs on 4 processors.

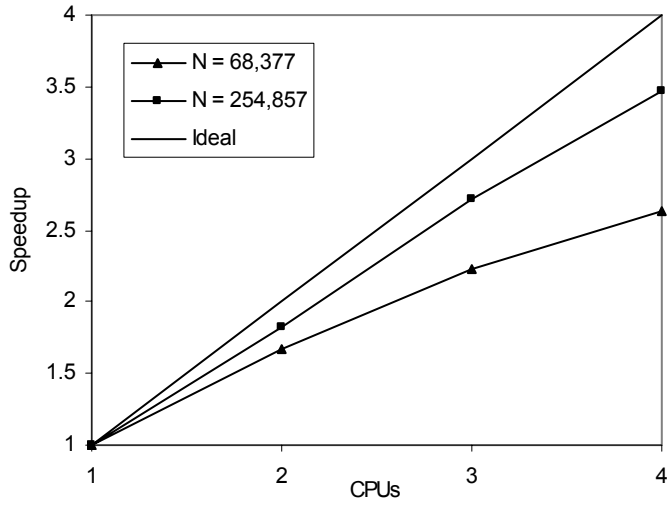


Figure 4. Parallel speedup versus the number of processors for two problem sizes.

The main network overhead comes from the global communication time, although the message length is too small. This is due to the high latency of the Ethernet network of our cluster. The time (t_{comm}) that is spend for a message of n bytes in length is given from:

$$t_{comm} = \mathbf{a} + \mathbf{b} \cdot n \tag{10}$$

where **a** is the latency of the network and **b** is the time for sending 1 byte. Thus, two network related factors limit the communication time: the latency and the bandwidth. Latency limits the exchange of small messages, mainly required in global communication whereas bandwidth limits the exchange of large messages, as happens in the neighboring communication. Thus, for finite element parallel computations with iterative Krylov solvers, a network with small latency is strongly preferred.

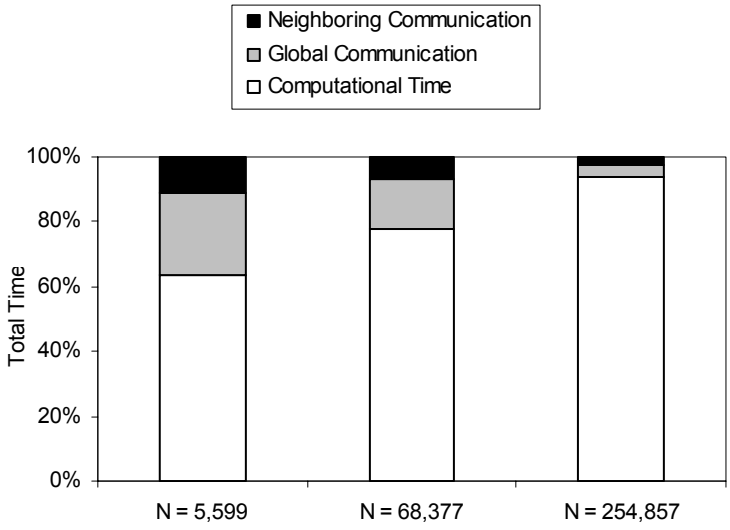


Figure 5. Computational vs Communication time.

More information about the usage and the availability of the FE-BUI, is available at http://www.chemeng.ntua.gr/yk/cluster.

ACKNOWLEDGMENTS

Financial support for this work was provided by the Ministry of Education through the Research Program "Pythagoras" and by the General Secretariat for Research and Technology through the "ENTEP" Program.

REFERENCES

[1] http://www-users.cs.umn.edu/~karypis/metis/index.html
[2] http://www.cs.sandia.gov/CRF/aztec1.html
[3] http://www-unix.mcs.anl.gov/petsc/petsc-2
[4] Saad, Y. (1996), *Iterative methods for sparse linear systems*, PWS Publishing Company. Available to http:// www-users.cs.umn.edu/~saad/books.html
[5] http://www-unix.mcs.anl.gov/mpi/
[6] http://www.netlib.org/blas
[7] http://www.netlib.org/lapack
[8] http://www-unix.mcs.anl.gov/mpi/mpich
[9] http://www.lam-mpi.org
[10] Spyropoulos, A. N., Palyvos, J. A. and Boudouvis, A. G. (2000) "Finite element computations on cluster of PC's and workstations." In *Proceedings of the 8th Euromicro Workshop on Parallel and Distributed Processing – EURO-PDP'2000*, pp. 56-61, Rhodos, Greece, January 2000 (IEEE Computer Society, Los Alamitos, CA, USA).
[11] Arnoldi, W.E. (1951), "The principle of minimized iterations in the solution of the matrix eigenvalue problem", Q. Appl. Math., Vol. 9, pp. 17-29.
[12] Erhel, J., Burrage, K. and Pohl, B. (1996), "Restarted GMRES preconditioned by deflation", J. Comput. Appl. Math., Vol. 69, pp. 303-318.
[13] Van der Vorst, H.A. and Vuik, C. (1993), "The superlinear convergence behaviour of GMRES", J. Comput. Appl. Math., Vol. 48, pp. 327-341.
[14] Smith, B., Bjorstad, P. and Gropp, W. (1996), *Domain Decomposition. Parallel multilevel methods for elliptic partial differential equations*, Cambridge University Press.
[15] Dennis, J. E. and Schnabel, R. B. (1996), Numerical *methods for unconstrained optimization and nonlinear equations*, SIAM, Classics in Applied Mathematics, 16, Philadelphia.
[16] Buyya, R. (ed.) (1999), *High Performance Cluster Computing: Programming and Applications*, Vol. 2, Prentice Hall, NJ, USA.
[17] Spyropoulos, A. N., Palyvos, J. A. and Boudouvis, A. G. (2004), "Bifurcation detection with the (un)preconditioned GMRES(m)", Comput. Methods Appl. Mech. Engrg., Vol. 193, pp. 4707-4716.
[18] Spyropoulos, A. N. (2003), *Large scale computations with parallel processing methods in nonlinear problems of interfacial magnetohydrostatics*, Doctoral Thesis (in Greek), NTUA.

ON THE COMPUTATION OF SELF-SIMILAR AND "COARSE" SELF-SIMILAR SOLUTIONS

Michail E. Kavousanakis[*], Constantinos I. Siettos[†], Andreas G. Boudouvis[*] and Ioannis G. Kevrekidis[‡]

[*] School of Chemical Engineering
National Technical University of Athens
9 Heroon Polytechniou St., Zografos, Athens, Greece, GR-15780

[†]School of Applied Mathematics and Physics
National Technical University of Athens
9 Heroon Polytechniou St., Zografos, Athens, Greece, GR-15780

[‡]Department of Chemical Engineering & Program in Applied and Computational Mathematics
Princeton University
Princeton, NJ 08544, USA

Keywords: dynamic renormalization, coarse self-similar, template function method, pinning condition

Abstract. *We present and discuss a computational approach to the study of partial differential equations with self similar solutions. The solutions of such equations can explode or decay preserving a constant (rescaled) shape. The key idea is to introduce a "dynamic pinning condition" through the use of which the solutions are studied in a co-exploding, or a co-collapsing frame and self-similarity can be factored out[1,2]. We show how to construct dynamic renormalization algorithms for the location of self similar solutions and the corresponding similarity exponents[3]. The dynamic renormalization algorithm can be implemented in system models at a microscopic level (kinetic Monte Carlo, Molecular Dynamics, Brownian Dynamics, Cellular Automata), where the statistics of the problem satisfy a macroscopic equation with self-similar solutions[4,5]. The template based approach for the computation of "coarse" self-similar solutions is implemented in the illustrative example of random walker diffusion.*

1 INTRODUCTION

In problems with translational invariance one encounters travelling wave solutions, which preserve their shape moving in space at a constant speed. It is convenient to study these problems in a co-travelling frame where the solution appears stationary and it is much easier to study its stability unencumbered by its constant motion. There are good techniques for the computational locating of travelling solutions along with the speed they move in space, as a nonlinear eigenvalue problem[6]. However, during transient simulations the solutions both travel and approach their ultimate translationally invariant shape; the "right" speed with which the travelling solution moves may change from moment to moment and the best way to choose it is not transparent. Rowley and Marsden proposed a template based approach[1] which allows to recover such an appropriate instantaneous speed. As the transient solution approaches its ultimate translationally invariant shape the speed also converges to the correct travelling speed.

The template based method can be straightforwardly adapted[2,3,7] for the computation of scale-invariant solutions, i.e. solutions of dynamic equations which evolve across scales. Self-similar solutions[6] are an important class of scale-invariant solutions. In the same sense that a travelling solution is more convenient to study in a co-travelling frame, a self-similar solution is more convenient to study in a co-collapsing (decaying solutions) or co-exploding frame (blow up solutions). Consider the general form of a partial differential equation:

$$u_t = D_x(u) \tag{1}$$

We restrict our interest in operators which satisfy the scaling relation:

$$D_x(Cu(\frac{x}{A})) = C^{\gamma} A^{\delta} D_y(u(y)), \forall A, C > 0, \text{ where } y = \frac{x}{A} \tag{2}$$

We assume that there exists a family of self-similar solutions for (1):

$$u(x,t) = \left|t - t^*\right|^a U(\frac{x}{\left|t - t^*\right|^b}) \tag{3}$$

where a and b are the similarity exponents and U is a member of the family of self similar solutions for (1). The critical time t^* is when the solution "explodes" either forward or backward in time. If we use the expression (3)

in (1) then U satisfies the following ordinary differential equation:

$$\sigma(aU - by\frac{\partial U}{\partial y}) = D_y(U) \tag{4}$$

where: $y = \frac{x}{\left|t - t^*\right|^b}$ and $\sigma = sign(t - t^*)$. The similarity exponents satisfy:

$$-1 = (\gamma - 1)a + \delta b \tag{5}$$

The cases where similarity exponents can be obtained using scaling arguments and conservation laws are referred to as self-similar of the first kind. When the similarity exponents cannot be obtained a priori and are usually obtained by solving a nonlinear eigenvalue problem for U, the self-similar solution is characterized as of the second kind [6]. An alternative approach for the computation of self-similar solutions of both kinds in a dynamical frame is the proposed template function method.

2 TEMPLATE FUNCTION METHOD

In this section we describe the template based approach for the computation of self-similar solutions of dynamic partial differential equations. We start with a general scaling:

$$u(x,t) = B(\tau)W(\frac{x}{A(\tau)}, \tau), \ \tau = \tau(t) \tag{6}$$

where A, B and τ are unknown functions; A and B correspond to "width" and "amplitude" of the solution. Setting $\tau = A^{\delta}B^{\gamma-1}$ equation (1) becomes:

$$\frac{\partial W}{\partial \tau} = D_y(W) + \frac{1}{A}\frac{\partial A}{\partial \tau}y\frac{\partial W}{\partial y} - \frac{1}{B}\frac{\partial B}{\partial \tau}W, \text{ where } y = \frac{x}{A} \tag{7}$$

This is the co-exploding / co-collapsing equation which for self-similar solutions is analogous to the co-traveling equation for translationally invariant ones. To obtain a unique self-similar solution for (7) we need to specify two additional constraints, "pinning conditions" which set the appropriate scales for $W(y,\tau)$. In the spirit of Rowley and Marsden these pinning conditions can be constructed imposing relationships between the solution and essentially arbitrary template functions[3]. These pinning conditions have the general form:

$$\begin{aligned} g_1(W, T_1(y)) &= 0 \\ g_2(W, T_2(y)) &= 0 \end{aligned} \tag{8}$$

where g_1, g_2 are typically algebraic expressions. Once the pinning conditions have been imposed, equation (7) describes the evolution of both the reconstructed solution W and the scaling factors A and B. It is actually possible to eliminate the $\frac{1}{A}\frac{\partial A}{\partial \tau}$ and $\frac{1}{B}\frac{\partial B}{\partial \tau}$ terms in equation (7) using the pinning conditions (8), to end up with a "co-exploding" PDE which is called MN-dynamics[3]. If the steady state of (7) exists then one can find the values of $\frac{1}{A}\frac{\partial A}{\partial \tau}$ and $\frac{1}{B}\frac{\partial B}{\partial \tau}$ terms and obtain one additional algebraic equation for the similarity exponents. Comparing the coefficients in (4) and the steady state of (7):

$$\frac{a}{b} = (\frac{1}{B}\frac{\partial B}{\partial \tau})/(\frac{1}{A}\frac{\partial A}{\partial \tau})\Big|_{\tau\to\infty} \tag{9}$$

From the simple system of equations (5) and (9) one can get the values of similarity exponents a and b. This template based approach can be implemented to locate both types of self-similar solutions. Next we present the application of the template based approach for the computation of self-similar solutions for the porous medium equation.

3 SELF-SIMILAR SOLUTIONS FOR THE POROUS MEDIUM EQUATION.

We present the template based approach for the computation of self-similar solutions for the following porous medium equation:

$$\frac{\partial u}{\partial t} = \nabla^2 u^m, \text{ with } m = 2 \tag{10}$$

The 1D Barenblatt solution is the self-similar solution of the first kind to the porous medium equation whose initial datum is a Dirac mass at the origin. This solution is known[8] and the similarity exponents are *b=-a=1/3.* The operator in this case is $D_x(u) = \frac{\partial^2 u^2}{\partial x^2}$ and the scaling condition for the similarity exponents is (from eq.(5)):

$$-1 = a - 2b \tag{11}$$

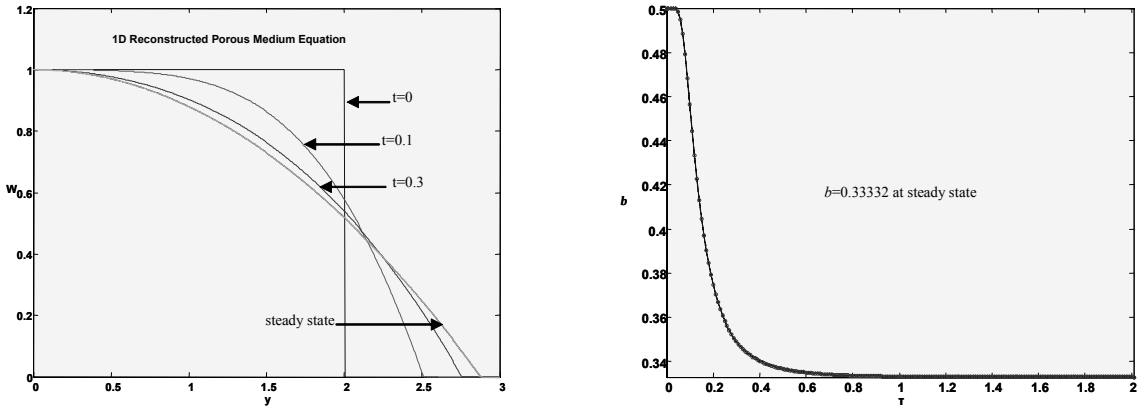
The equation is discretized with central finite differences and time integration is carried out with an explicit Euler scheme on a domain [0,3] with 1000 nodes. At the boundaries zero flux conditions are imposed. For the initial condition we take:

$$w(y,0) = \begin{cases} 1, y < 2 \\ 0, y \geq 2 \end{cases} \tag{12}$$

The two pinning conditions we select are:

$$\begin{aligned} &< w, T_1 >= 0 \\ &< w, T_2 >= \text{constant} = w(0.003,0) \end{aligned} \tag{13}$$

where:< > denotes inner product, $T_1 = \begin{cases} 1, y < 1 \\ -1, y \geq 1 \end{cases}$ and $T_2 = \delta(y - 0.003)$. The first condition is an orthogonality condition between the reconstructed solution and the template T_1, and the second forces W to have a constant value, equal to that of the initial condition value, at $y_o=0.003$. The evolution of W and similarity exponent b, up to steady state are depicted in Fig.1.



(i) (ii)

Figure 1. Time evolution of (i):reconstructed *W* towards the Barenblatt steady state self-similar solution and (ii): similarity exponent *b* for 1D porous medium equation with m=2.

The numerical value of similarity exponent *b* at steady state is *b=0.33332* and it is very close to the theoretical value of 1/3.

The Graveleau solutions to the porous medium equation form a 1-parameter family of axi-symmetric focusing solutions which are self-similar of the second kind[9].The partial differential equation examined is:

$$u_t = D_r(u) = \frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial u^2}{\partial r}) \tag{14}$$

It is trivial to verify that the differential operator *D* satisfies the exponential relation (2) with *γ=2* and *δ=-2.* Then the first algebraic equation for the similarity exponents *a* and *b*, from (5), is:

$$1 = -a + 2b \tag{15}$$

Equation (14) is solved in $r \in [0,10]$ with boundary conditions $\left.\frac{\partial u}{\partial r}\right|_{r=10}, u|_{r=0} = 0$ .The initial condition we use here is:

$$u(r,0) = \begin{cases} 1, x > 4 \\ 0, x \le 4 \end{cases} \tag{16}$$

The initial condition for *u* coincides with the initial condition for *W,* as does the physical space *r* with the rescaled space *y*.

The discretization scheme used is central finite different and the number of discretization nodes 2000. For the time stepper we used a fully implicit time integration scheme. The time evolution of (14) is depicted in Fig.2.

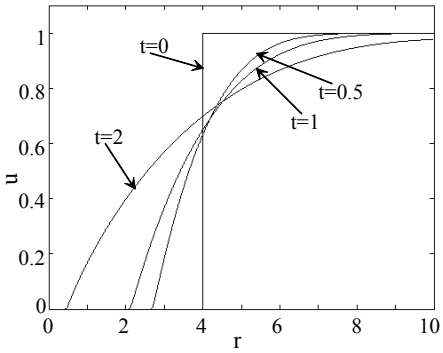


Figure 2. Time evolution of the radially symmetric solution of the porous medium equation.

In order to apply the dynamic renormalization algorithm we need two additional template conditions (pinning conditions). For this problem one choice of pinning conditions is the following:

$$\begin{aligned} &< W, T_1 >= 0 \\ &< W, T_2 >= \text{constant} \end{aligned} \tag{17}$$

where:

$$T_1 = \begin{cases} -1, y < 7 \\ 1, y \ge 7 \end{cases} \text{ and } T_2 = \delta(y - 9.98) \tag{18}$$

It can be easily verified that the first pinning condition is valid for *W=u(r,0)* while the second imposes a constant value for the reconstructed solution at *y*=9.98. This constant can be taken equal to *u(r=9.98,0).*

We denote with *W* the dynamically renormalized solution for the porous medium equation and $y = \frac{r}{A}$ the rescaled space. The co-collapsing equation in this case is according to (7):

$$\frac{\partial W}{\partial \tau} = \frac{1}{y}\frac{\partial}{\partial y}(y\frac{\partial W^2}{\partial y}) + Gy\frac{\partial W}{\partial y} + CW \tag{19}$$

where $G = \frac{1}{A}\frac{\partial A}{\partial \tau}$, $C = -\frac{1}{B}\frac{\partial B}{\partial \tau}$ and τ which is the properly rescaled time $\frac{\partial \tau}{\partial t} = BA^{-2}$.

A unique solution of (19) will be obtained using two additional constraints, i.e. the pinning conditions (17) and (18). Then the dynamically renormalized system is the system of equations (17), (18) and (19). At steady state solution *W* corresponds to a member of the family of self-similar solutions for the 2D porous medium equation. The similarity exponents can also be evaluated using (15) and:

$$\frac{a}{b} = (\frac{1}{B}\frac{\partial B}{\partial \tau})/(\frac{1}{A}\frac{\partial A}{\partial \tau})\Big|_{\tau\to\infty} = C^*/G^* \tag{20}$$

where C^* and G^* are the steady state values of *C* and *G*, respectively, for the system (17), (18) and (19). The numerical value of exponent *b* is: *b≈0.858,* and the theoretical value is approximately 0.856, which is in good

agreement with the numerical one.

The evolution of reconstructed *W* towards steady state is depicted in Fig.3.

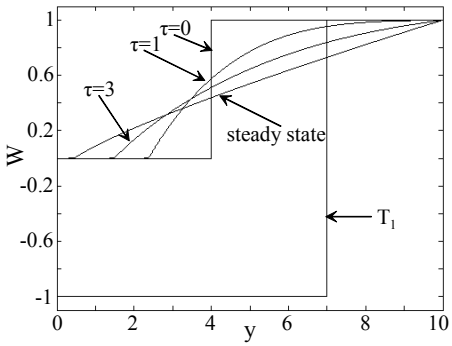


Figure 3. Dynamic renormalization for the computation of Graveleau self-similar solution.

4 COARSE DYNAMIC RENORMALIZATION

As a first step, it is interesting to consider the case where the only available tool is a "legacy dynamic code", i.e. a code which simulates the evolution of the original equation but cannot be modified. The so-called numerical analysis of legacy codes[10] allows us to utilize the legacy simulator and perform a different set of tasks for which it was not designed, by wrapping around the code a computational superstructure. One of these tasks is the computation of self-similar solutions for systems where the only available information comes from runs of a "black box" simulator. One can evolve in physical variables and then rescale the results as opposing to first recalling the equation and then evolving the rescaled equation.

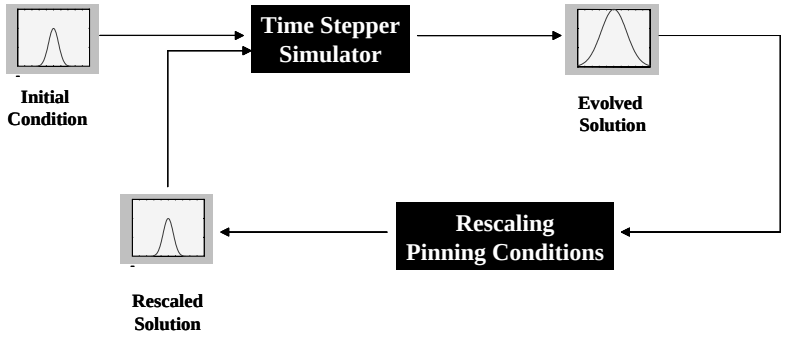


Figure 4. Schematic procedure of dynamic rescaling using a legacy simulator.

Starting with an initial condition one evolves forward for a finite time. Then the template conditions are applied in order to rescale the space variable for the final profile. The rescaled profile is given as an initial condition for the time stepper, it is evolved, and the new final profile is again rescaled by applying the template conditions. The iterated procedure may converge to a member of a family of self-similar solutions. The key idea is that rescaling and then evolving the rescaled equation for a finite time commutes with evolving in physical space and then rescaling the result[4].

Only stable self-similar shapes can be obtained by such a dynamic rescaling and forward integration. Unstable self-similar solutions can be found through fixed point algorithms (like Newton-Raphson). If we call $R_T(w(y))$ the result of integrating the rescaled solution with initial condition $w(y)$ for a finite time T, the self-similar solution satisfies:

$$w(y) - R_T(w(y)) = 0 \tag{21}$$

Matrix-free iterative linear algebra techniques[11] can be used to converge to steady state solutions of equation (21). The only available tool is a subroutine that numerically computes $R_T(w(y))$, yet it is the only necessary tool for matrix-free iterative techniques, like Newton-GMRES.

4.1 "Coarse" self-similar solutions

In contemporary problems of scientific and engineering interest the best system models are often available at

a fine (microscopic) scale. We need to predict the macroscopic behavior of such a system. In traditional continuum modeling, successful closures allow us to write models at a coarse grained - macroscopic level. However, such closures are not always available and the range of time and space scales precludes the effective macroscopic prediction through direct atomistic simulation. The recently developed "equation-free" approach[10] to coarse-grained, computer assisted analysis, attempts to bridge the enormous scale gap (microscopic to macroscopic) when macroscopic equations conceptually exist but are not available in a closed form. The main idea is to consider the microscopic simulation as a computational experiment which can be initialized at will. One can then estimate numerical quantities, such us functional and parametric derivatives, using the results of appropriately initialized short bursts of computational experiments, which in turn are used when one applies a traditional numerical method.

In many cases the macroscopic dynamics do not involve stationary solutions but rather traveling or self-similar ones. We will show how to construct dynamic renormalization algorithms for the location of self-similar solutions when the only available tool is a simulator (time-stepper) at a fine scale (e.g. kinetic Monte Carlo, Brownian Dynamics, Molecular Dynamics, e.t.c.). The procedure consists of the following steps[4]:

1. We choose the statistics of interest for describing the coarse-grained behavior of the system and an appropriate representation of them. If the system we study is molecular diffusion then the appropriate macroscopic observable is the concentration. It is more convenient to use the Cumulative Distribution Function (CDF) of particles at each moment. If the CDF is smooth enough then one can use a low-dimensional description based on a sequence e.g. of low order orthogonal polynomials. We denote this macroscopic description with u. These choices determine a restriction operator $\mathtt{M}$ from the microscopic-level description (U: coordinates of a large number of particles) to the macroscopic description, .i.e. $u=\mathtt{M}U$.
2. Choose an appropriate lifting operator μ from the macroscopic to microscopic description, e.g. we can assign the position coordinates of each particle using the functional form of the particle CDF. The lifting operator should have the property: $\mu\mathtt{M}=\mathrm{I}$. In other words lifting from the microscopic to macroscopic description and then restricting should have no effect.
3. From an initial value at microscopic level ($U(t_o)$) run the microscopic simulator for a relatively short time horizon T and obtain the new microscopic description $U(T)$. We may need several microscopic initial conditions (copies), all consistent with the same macroscopic one, for variance reduction purposes.
4. Obtain the (average) restriction of the new microscopic description: $u(T)=\mathtt{M}U(T)$.
5. Rescale the obtained macroscopic description using the template conditions. If $\mathtt{g}$ is the rescaling operator applying the template conditions, then the following relation holds between the macroscopic and rescaled macroscopic description: $u_R(T)=\mathtt{g}(u(T))$.
6. Lift the rescaled restricted description ($U=\mu u_R$) and repeat steps 3-6.

The procedure above is the backbone of the so called "coarse dynamic renormalization" which we expect to converge to stable "coarse" self-similar solutions, i.e. self-similar solutions of the unavailable macroscopic equation.

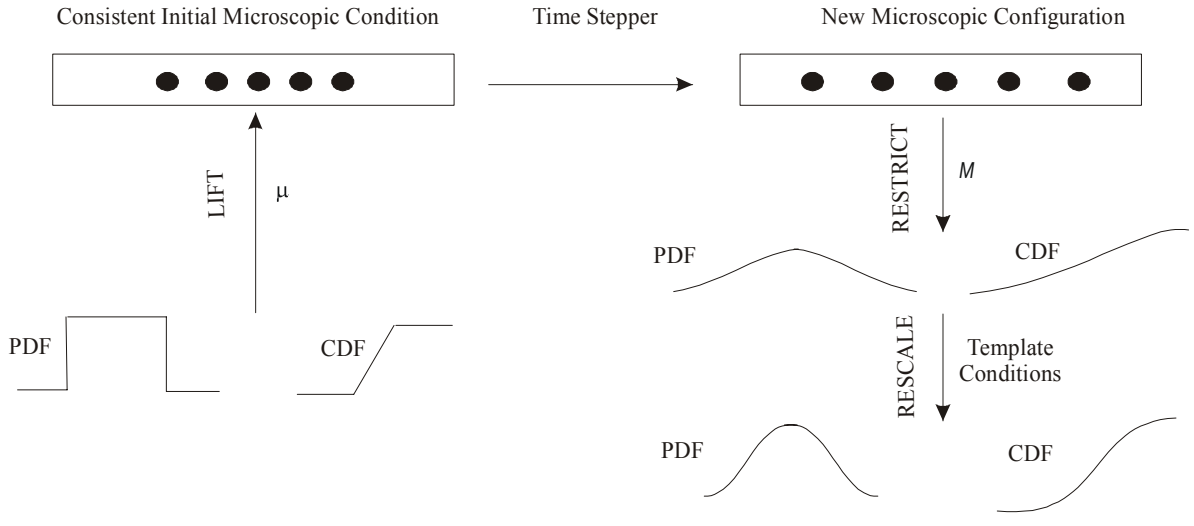


Figure 5. Schematic depiction of the coarse dynamic renormalization procedure.

5 EQUATION-FREE COMPUTATION OF COARSE SELF-SIMILAR SOLUTIONS FOR THE SIMPLE 1D PARTICLE DIFFUSION.

The simple diffusion process can be simulated by a large number of random walkers. In our micro-simulation random walkers move on a 1D domain at discrete time steps: $t_n = nh$. Every walker-particle is moved by $dx \in N(0, 2Dh)$ -the normal distribution with 0 mean value and $2Dh$ variance-, where D is the diffusion coefficient[12].

We are going to work with a single-variable cumulative distribution function (CDF). We use the CDF rather than the particle density because it is easy to obtain from the particles, while the computation of the empirical density function is numerically sensitive. After each micro-simulation, the new positions of the random walkers are sorted and plotted versus the CDF value. In fact it is easier to work with the inverse CDF, the ICDF, which gives the x coordinate of a given particle, i. The ICDF of the obtained CDF, whose support lies between [0,1] can be easily approximated by orthogonal polynomials[13].

For this computation we considered a 5^{th} order polynomial approximation of the ICDF, where the macroscopic observables are the appropriately computed coefficients of the orthogonal polynomials. The constancy of the number of particles provides the first template condition. The second template condition demanded that the coefficient of the first order term in the polynomial expansion be unity. The dynamic renormalization procedure converges to the expected Gaussian distribution, which corresponds to the self-similar shape solution for the simple diffusion problem.

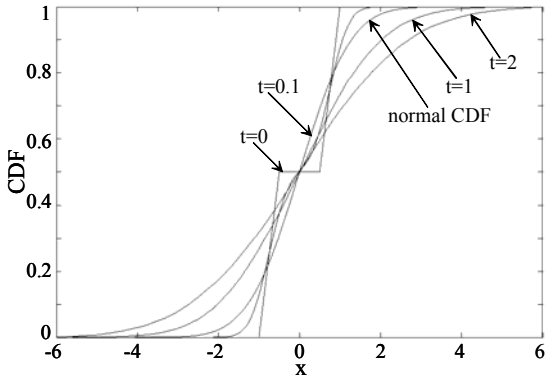


Figure 6. Time evolution of CDF for the 1D simple diffusion using N_{tot} random walkers.

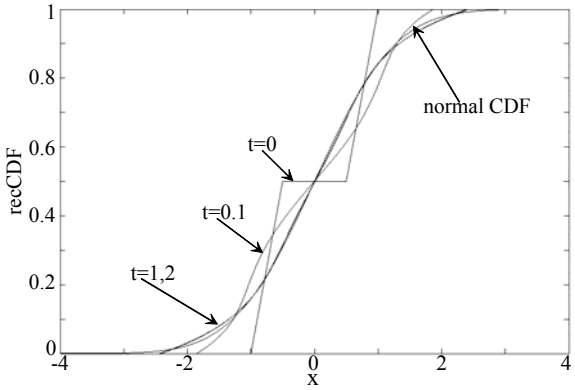


Figure 7. Coarse evolution of the CDF using coarse renormalized timestepping.

6 CONCLUSIONS

We have described a systematic approach for the computation of self-similar and coarse self-similar solutions. Locating self-similar solutions becomes a steady state problem for the co-exploding PDE. Therefore, tools from numerical bifurcation analysis can be brought to bear on this study. It is possible to implement bifurcation / continuation /stability calculations and attempt to locate branches of self-similar solutions (stable, unstable, bifurcating). In[7] the template based approach applied on the focusing nonlinear Schrödinger equation yields a bifurcation view of the onset of focusing.

In case of "coarse" self-similar solutions the only available information comes from microscopic simulations. The procedure of coarse dynamic renormalization uses short bursts of appropriately initialized simulations. Matrix-free techniques can be implemented for the computation of both stable and unstable self-similar solutions, compute finite times to blow up and estimate similarity exponents. We can also check for the stability

of solutions through matrix-free eigenvalue analysis.

It is also possible to combine the equation-free approach with the so called "gaptooth scheme" and "patch dynamics"[10,14,15]. The microscopic simulations are not performed in the whole physical space but in small domains (teeth) which are separated by gaps and connected with appropriate boundary conditions. This approach exploits smoothness of macroscopic observables (e.g. particle density) in both space and time and, for certain problems, has the potential to reduce the necessary microscopic simulation time for the computation of coarse self-similar solutions.

AKNOWLEDGEMENTS

This work was supported by the Basic Research Program "PROTAGORAS" of the National Technical University of Athens.

REFERENCES

[1] Rowley C.W. and Marsden J.E. (2000), "Reconstruction equations and the Karhunen-Loève expansion for systems with symmetry", Physica D 142, pp.1-19.
[2] Rowley C.W., Kevrekidis I.G., Marsden J.E. and Lust K. (2003), "Reduction and reconstruction for self-similar dynamical systems", Nonlinearity 16, pp.1257-1275.
[3] Aronson D.G., Betelu S. and Kevrekidis I.G. (2001), "Going with the flow: a Lagrangian approach to self-similar dynamics and its consequences", oai:arXiv.org:nlin/0111055.
[4] Chen L., Debenedetti P.G., Gear C.W. and Kevrekidis I.G. (2004), "From molecular dynamics to coarse self-similare solutions: a simple example using equation-free computation",Journal of non Newtonian Fluid Mechanics 120, pp:215-223.
[5] Chen L., Kevrekidis I.G. and Kevrekidis P.G. (2004), "Equation-free dynamic renormalization in a glassy compaction model",oai:arXiv.org:cond-mat/0412773.
[6] Barenblatt I. (1996), Scaling, Self-Similarity and Intermediate Asymptotics, Cambridge, University Press, Cambridge.
[7] Siettos C.I., Kevrekidis I.G. and Kevrekidis P.G. (2003), "Focusing revisited: a renormalization / bifurcation approach", Nonlinearity 16, pp:497-506.
[8] Aronson D.G. (1986), "The porous medium equation". In "Some problems is Nonlinear Diffusion" (eds. Fasano A. and Primicerio M.), Lecture Notes in Math. No.1224, Springer-Verlag.
[9] Aronson D.G. and Graveleau J. (1993), "A self-similar solution to the focusing problem for the porous medium equation", European Journal of Applied Mathematics 4, pp: 65-81.
[10] Kevrekidis I.G., Gear C.W., Hyman J.M., Kevrekidis P.G., Runborg O and Theodoropoulos K. (2003), "Equation-free multiscale computation: enabling microscopic simulators to perform system level-tasks", Comm. Math. Sci. 14, pp: 715-762.
[11] Kelley C.T., Iterative Methods for Linear and Nonlinear Equations, SIAM Publications, Philadelphia, 1995.
[12] Hughes B.D. (1995), "Random Walks and Random Environments Vol.1: Random Walks", New York: Oxford University Press.
[13] Gear C.W. (2001), "Projective integration methods for distributions", NEC TR 130.
[14] Gear C.W., Li J. and Kevrekidis I.G. (2003), "The gap-tooth method in particle simulations", Phys. Lett. A 316, pp: 190-195.
[15] Samaey G., Roose D. and Kevrekidis I.G. (2003), "Patch dynamics for Homogenization problems", SIAM MMS, in press, also physics/0312004 at arXiv.org.

A FRESH VIEW: QUADRATIC ASSIGNMENT PROBLEM

Arif Akkeleş, Ahmet Ünveren, Adnan Acan

Department of Applied Mathematics and Computer Science
Eastern Mediterranean University
Famagusta, Cyprus
e-mail: arif.akkeles@emu.edu.tr, web: http://www.emu.edu.tr/~akkeles

Keywords: QAP, linear programming, Dual Problem, Genetic Algorithm, Nugent.

Abstract. *The Quadratic Assignment Problem (QAP) is a well-known combinatorial optimization problem (NP-Hard) with a wide variety of applications. Many applications and some semi-numerative procedures have been proposed for QAP but there is no algorithm on this problem, which is dominant. In this work we survey some of most important applications, methods and techniques about the QAP. Some recent developments with our new applications, such as Genetic algorithm approach and some algorithms to solve the problem (QAP) optimality will be explained. Also some computational results which show that our approach finds solutions competitive with those of which are well known heuristics are given.*

1 INTRODUCTION

Consider the problem of assigning objects to positions in such a way that each object is designated to exactly one position and reciprocally. The demand flows of among the objects, the distance between the positions and the assignment cost (the object versus positions) are known. In general, we are given a set of n locations and n facilities, and told to assign each facility to a location. To measure the cost of each possible assignment we multiply the prescribed flow between each pair of facilities by the distance between their assigned locations, and sum over all the pairs. The objective is to find an assignment to all facilities to all locations, such that the total cost of the assignment is minimized. This problem is precisely a quadratic assignment problem. If we approach mathematically, we can formulate the problem by defining two n by n matrices $(n \times n)$: a flow matrix $F = f_{ij}$ element represents the flow between facilities i and j, and a distance matrix $D = d_{ij}$ element represents the distance between locations i and j. We represent an assignment by the vector p, which is a permutation of the numbers $N=\{1, 2, \ldots, n\}$, $p(j)$ is the location to which facility j is assigned.

With these definitions, the QAP can be written as

$$\min_{p \in \Pi} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} d_{p(i)p(j)}$$

The main idea of this paper is to make some survey of QAP, applications and solution techniques, algorithms which is used to sole QAP, some researcher's solution theory, genetic algorithm application of the problem and comparison of the different algorithms to show that our solution has better approach from many known algorithms in literature.

The QAP was originally introduced in 1957 by T.C. Koopmans and Martin Beckman [45] as a mathematical model related to economic activities. Since then, it has among the most studied problem in combinatorial optimization.

The most effective algorithms for optimally solving quadratic assignment problems are based on branch and bound algorithm. To solve the QAP with on branch and bound algorithm, there are 3 types of algorithms. These are single assignment algorithm, pair assignment algorithms and relative positioning algorithm [4]. Some of earliest methods on branch and bound algorithms for solving QAPs can be found in detailed in [4] [12] [31] [48] [49] [51] [56].

This paper will present a general overview of the QAP. An introduction providing a definition will be given first together with a brief historical overview. Formal problem description and mathematical formulations will be provided.

In section 2 some formulations and specifications will be introduced. Lower bounds and exact algorithms will be the focus of section 3. Concluding remarks are given in section 4.

2 GENERAL FOMULATIONS OF QAP

The term "quadratic" comes from the reformulation of the problem as an optimization problem with a quadratic objective function.

2.1 Some Formulations of QAP

There is an equivalent formulation of QAP as Integer Linear Programming formulation or a quadratic 0-1 programming formulation. Let $X = (x_{ij})$ represents a permutation matrices where the size of this matrix is $n \times n$ and Π_N represents set of all permutations of N, then there is a one-to-one correspondence between Π_N and X . Then the entries of the each such matrices satisfies the following constraints;

$$\sum_{j=1}^{n} x_{ij} = 1, \quad \mathrm{i} = 1,\ldots,\mathrm{n}, \tag{1}$$

$$\sum_{j=1}^{n} x_{ij} = 1, \quad \mathrm{j} = 1,\ldots,\mathrm{n}, \tag{2}$$

$$x_{ij} \in \{0,1\}, \quad i = 1,\ldots,n, \quad j = 1,\ldots,n,$$

where
$$x_{ij} = \begin{cases} 1 & \text{if facility i assigned to location j} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

If the above conditions are met , the QAP can be formulates as follows; where we consider f_{ij} the flow between objects *i* and *j,* and d_{kl} the distance between positions *k* and *l*. Then our aim is to calculate the following

$$\min \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} f_{ij} d_{kl} x_{ik} x_{jl} \tag{4}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_{ij} = 1, \quad \mathrm{j} = 1,\ldots,\mathrm{n},$$

$$\sum_{i=1}^{n} x_{ij} = 1, \quad \mathrm{j} = 1,\ldots,\mathrm{n},$$

$$x_{ij} \in \{0,1\}, \quad i = 1,\ldots,n, \quad j = 1,\ldots,n,$$

With the conditions (1)-(3) we may have an equivalent formulation for the QAP, by working on the space of permutation matrices. A general form of QAP instance of order *n* is given by flow matrix, distance matrix and cost of assignment matrix (cost of the objects to a positions) can be defined as ;

$$\min \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} f_{ij} d_{kl} x_{ik} x_{jl} + \sum_{i=1}^{n}\sum_{k=1}^{n} c_{ik} x_{ik} \tag{5}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_{ij} = 1, \quad \mathrm{j} = 1,\ldots,\mathrm{n},$$

$$\sum_{i=1}^{n} x_{ij} = 1, \quad \mathrm{j} = 1,\ldots,\mathrm{n},$$

$$x_{ij} \in \{0,1\}, \quad i = 1,\ldots,n, \quad j = 1,\ldots,n,$$

Then problem become easier to solve with the linear system (5). A more general form of QAP involves costs that do not necessarily correspond to products of flows by distances. This version of QAP was introduced by Lawler (1963) [46] and his formulation is as follows:

$$\min \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} c_{ijlk} x_{ik} x_{jl}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_{ij} = 1, \quad \mathrm{j} = 1,\ldots,\mathrm{n},$$

$$\sum_{i=1}^{n} x_{ij} = 1,\ \mathrm{j} = 1,\ldots,\mathrm{n},$$

$$x_{ij} \in \{0,1\},\ i = 1,\ldots,n,\quad j = 1,\ldots,n,$$

This model was used by many researchers [11] [21] [27] [36] [41] [63].

By replacing the quadratic terms by linear terms we have an integer programming formulation of QAP that is called mixed integer linear programming formulation (MILP). This can be found in different formats in some paper (depending number of variables that is used).

As in [46] and in [50],

$$c_{ij} = f_{ij} d_{kl} \text{ and } y_{ijkl} = x_{ik} x_{jl} \text{ where } 1 \le i, \le j, k, l \le n$$

and he has used n^4 variables in his format. For the original problem, some other formulation use relaxations as in [11] [17] [48] [50] [55] [57]. There can be huge number of variables and constraints for linearization of QAP on MILP. Thus this makes problem more complicated and this is the fact that makes this approach avoided in many cases. However, together with some properties the arise from the linearization of the objective function and together with some constraint relaxations, the lower bounds for the optimal solution is achieved. Research about linearization can be found in [1] [20] [32] [55] (for achieving the best lower bound via relaxation via Lagrangian relaxation) [26], which has the advantage of smaller number of restrictions [55] (owing to its polytope description).

The authors show that the following equations are equivalent to the formulations (1)-(5) [20].

$$\min \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} f_{ij} d_{kl} y_{ijkl} \tag{6}$$

$$\sum_{i=1}^{n} x_{ik} = 1 \qquad k = 1,\ldots,n, \tag{7}$$

$$\sum_{k=1}^{n} x_{ik} = 1 \qquad i = 1,\ldots,n, \tag{8}$$

$$\sum_{i=1}^{n} y_{ijkl} = x_{jl} \qquad i, j, l = 1,\ldots,n, \tag{9}$$

$$\sum_{j=1}^{n} y_{ijkl} = x_{ik} \qquad j, k, l = 1,\ldots,n, \tag{10}$$

$$\sum_{k=1}^{n} y_{ijkl} = x_{jl} \qquad i, j, l = 1,\ldots,n, \tag{11}$$

$$\sum_{l=1}^{n} y_{ijkl} = x_{jk} \qquad i, j, k = 1,\ldots,n, \tag{12}$$

$$y_{ijkk} = x_{iikk} \qquad i, k = 1,\ldots,n \tag{13}$$

$$x_{ik} \in \{0,1\} \qquad i, k = 1,\ldots,n \tag{14}$$

$$0 \le y_{ijkl} \le 1 \qquad i, j, k, l = 1,\ldots,n. \tag{15}$$

By taking a simple approach to the pairwise allocation of object costs to adjacent positions is proportional to flows and to distances between them. The QAP formulation uses the permutation concept can be found in [48] [55] [59].. There are also some formulations about non-symmetrical QAP through a directed graph.

2.2 Linear Programming and Dual LP Bounds

In section 2.1 we have introduced some equations with some constraints. It is known that large classes of bounds for the QAP are related to linear programming relaxations [6]. Defining variables ($y_{ijkl} = x_{ij} x_{kl}$) and by dropping the integrality conditions, resulting in linear programming relaxation [6] [32], we have the following formulation with some constraints.

$$\min \sum_{i=1}^{n} y_{ijkl} d_{ijkl}$$

$$\text{s.t} \quad \sum_{j=1}^{n} x_{ij} = 1, \qquad i = 1,\ldots,n,$$

$$\sum_{l=1}^{n} x_{ij} = 1, \qquad j = 1,\ldots,n,$$

$$\sum_{l=1}^{n} y_{ijkl} = x_{ij} \qquad i, j, k = 1,\ldots,n,$$

$$\sum_{k=1}^{n} y_{ijkl} = x_{ij} \qquad i, j, l = 1,\ldots,n,$$

$$y_{ijkl} = y_{klij} \qquad i, j, k, l = 1,...,n,$$

$$x_{ij} \geq 0, \quad y_{ijkl} \geq 0 \quad i, j, k, l = 1,...,n.$$

The constraints in $y_{ijkl} = y_{klij},\ i, j, k, l = 1,...,n,$ are called symmetry conditions. And they show that linear programming for QAP can be formulated by using $y_{ijkl} = y_{klij}$, $i \leq k$ .. Together with the variables that can be eliminated using the fact that *x* is feasible in QAP, the linear programming QAP can be written as a linear programming problem. To reduce number of equality constraints more analysis are needed. They can be found in [55]. The advantage of those problems which have symmetric data is that the number of variables and equality constrains can be halved [55]. It is proved that if the symmetry conditions are dropped the value in LPQAP is exactly GLP Adams and Johnson [1]. Many bounding schemes for QAP can be viewed as Lagrangian procedures that try to solve dual of Linear Formulation of QAP approximately. A bounding procedure for the QAP extends the Hungarian method for linear assignment problem (LAP) to QAPs, and these are computationally most successful schemes [40] [41] [42] [43].

3 LOWER BOUNDS AND EXACT ALGORITHMS

Calculation of lower bounds of QAP is one of the most studied topics for the solution of QAP Çela, E. [20].. In practice, the lack of efficiently computable, tight lower bounds for QAP has been key factor in the problem's difficulty. They are essential components of branch and bound procedures and they are also used to evaluate the goodness of solutions produced by heuristics. There are many approaches to obtaining lower bounds. Gilmore-Lawler bound (GLB) is one of the oldest methods, which is widely used. We have three main classes of lower bounds. Gilmore-Lawler bound (GLB), eigenvalue related bounds, and bounds based formulations. More explanations about these methods can be found in [1] [3] [5] [6] [8] [9] [11] [12] [27] [30] [38] [39] [41] [43] [48] [61].

3.1 Exact Algorithms

The different methods used to find the global optimum solution for a given QAP. Branch-and-bound, cutting planes and dynamic programming methods are some of these techniques. Branch-and-bound is the most known technique and it is redefined from allocation and cutting rules [10] [11] [14].. These rules define lower bounds for the problem. Branch-and-bound procedures are generally the most helpful for solving QAP. The first enumerative schemes, which use lower bounds to eliminate, undesired solutions can find in [33] [46] [47] [49], in Pitsoulis paper [58] an excellent description can be found and there are many references concerning QAP branch-and-bound in [1] [33] [58] [59].. Although many procedures that combine branch and bound techniques together with parallel implementations are being used, for bigger sizes successes depend on the hardware of the computer [19] [54].

When the flow matrix is the adjacency matrix that formed from tree, the dynamic programming technique used. Urban [64] and Christopher and Benavent [22] used this technique for MILP approach for relaxation. They used the advantage of the polynomial complexity of instances.

Cutting plane methods is used formulation of some heuristics that use MILP [14].. When method introduced by Bazaraa and Sherali [12], results were not satisfactory. It is proper for small instances [44] [19]. Recently, Bender's algorithm is used to deal with motherboard-designed problem, including linear cost in the formulation see [53].

4 HEUIRISTICS

Heuristic algorithms do not guarantee the best solution for the optimality. In this category approximated methods can be included where the worst-case guarantees of properties are well known. Here we are introducing some known methods. These are the most recent methods that can be adapted wide range of method. These procedures do not provide a global optimal solution but can produce good answers within reasonable constraints. Construction methods, Limited enumeration method, Improvement Methods, Simulated annealing techniques, and Genetic algorithms are five of these basic categories of heuristics of QAP. In this section we will concentrate on Genetic algorithms. Some useful results are included for genetic algorithm to show that genetic algorithm is competitive to other algorithms that are in the same category.

4.1 Genetic Algorithms

J. Holland has developed genetic algorithms in 1975. Genetic algorithms store a set of solutions and then it work to replace these solutions with better ones based on some fitness criterion, usually the objective function value [36] [57].. Genetic algorithms are based on the natural selection and natural adaptation [36].. These algorithms are parallel and helpful when applied in suitable environment [57]. In [19] is used to solve the different instances of QAP. In this paper we solved Nugent problem for different instances of it. Genetic algorithms solved the problem in very short time considering other algorithms solution with better optimal solutions for the same costs and distances.

Nugent, Vollman, and Ruml posed a set of problem instances of size 5, 6, 7, 8, 12, 15, 20, and 30, noted for their difficulty [1] [19]. These QAP samples have multiple global optima. Even worse, these globally optimal solutions are at the maximally possible distance from other globally optimal solutions.

By enumerating all possible solutions Nugent, Vollman, and Ruml solved the first four of these samples [16] [23]. Then, Nug12 and Nug15 were solved using branch and bound heuristics on. In past 14 years, techniques such as simulated annealing, genetic algorithms, randomized adaptive search, etc were used to solve instances up to Nug24.

In 2000, Nug30 is solved exactly [6] [7].. The solution was found by applying a branch and bound algorithm, [5]. The involved bound was based on convex quadratic programming [5].

T. Mautor [19] focuses on parallel implementations and exploits the metric structure of the Nugent instances to reduce the branching tree considerably. In table 1 and table 2 we used those instances that are probably mostly used. The distance matrix contains Manhattan distances of rectangular grids and they are taken form QAPLIB.

The instances of size n = {14, 16, 17, 18, 21, 22, 24, 25} were constructed out of the larger instances by deleting certain rows and columns, [25]. For the optimal solutions [64] is concerned. Optimality of Nug21 and Nug22 was proved by [1] [15], for Nug24 by [19] [26].. The instances of size n =27 and n =28 were constructed out of the instance of size n =30 by deleting the three or two last facilities, respectively, and were solved by Anstreicher, Brixius, Linderoth and Groux [7] in 2002.

In the following two tables, for the different instances of Nugent problem T. Mautor's solution [15] by using branch and bound algorithm and Genetic Algorithm is used. According to [19] genetic algorithm solutions are much shorter and less iteration is used to get the feasible solution of the problem.

Name	Fitness	Permutation
Nug12	578	(12,7,9,3,4,8,11,1,5,6,10,2)
Nug14	1014	(9,8,13,2,1,11,7,14,3,4,12,5,6,10)
Nug15	1152	(1,2,13,8,9,4,3,14,7,11,10,15,6,5,12)
Nug16a	1610	(9,14,2,15,16,3,10,12,8,11,6,5,7,1,4,13)
Nug16b	1240	(16,12,13,8,4,2,9,11,15,10,7,3,14,6,1,5)
Nug17	1732	(16,15,2,14,9,11,8,12,10,3,4,1,7,6,13,17,5)
Nug18	1930	(10,3,14,2,18,6,7,12,15,4,5,1,11,8,17,13,9,16)
Nug20	2570	(18,14,10,3,9,4,2,12,11,16,19,15,20,8,13,17,5,7,1,6)
Nug21	2438	(4,21,3,9,13,2,5,14,18,11,16,10,6,15,20,19,8,7,1,12,17)
Nug22	3596	(2,21,9,10,7,3,1,19,8,20,17,5,13,6,12,16,11,22,18,14,15)
Nug24	3488	(17,8,11,23,4,20,15,19,22,18,3,14,1,10,7,9,16,21,24,12,6,13,5,2)
Nug25	3744	(5,11,20,15,22,2,25,8,9,1,18,16,3,6,19,24,21,14,7,10,17,12,4,23,13)
Nug27	5234	(23,18,3,1,27,17,5,12,7,15,4,26,8,19,20,2,24,21,14,10,9,13,22,25,6,16,11)
Nug28	5166	(18,21,9,1,28,20,11,3,13,12,10,19,14,22,15,2,25,16,4,23,7,17,24,26,5,27,8,6)
Nug 30	6124	(14,5,28,24,1,3,16,15,10,9,21,2,4,29,25,22,13,26,17,30,6,20,19,8,18,7,27,12,11,23)

Table 1: Results of Branch and Bound algorithm for Nugent problem [15]..

Name	Fitness	Permutation
Nug12	578	(3,9,7,12,1,11,8,4,2,10,6,5)
Nug14	1014	(9,8,13,2,1,11,7,14,3,4,12,5,6,10)
Nug15	1152	(11,8,7,12,10,9,13,3,5,15,1,2,4,14,6)
Nug16a	1610	(9,14,2,15,16,3,10,12,8,11,6,5,7,1,4,13
Nug16b	1240	(14,15,4,16,6,10,2,12,1,7,9,13,5,3,11,8)
Nug17	1732	(16,15,2,14, 9,11,8,12,10,3,4,1,7,6,13,17,5)
Nug18	1930	(10,3,14,2,18,6,7,12,15,4,5,1,11,8,17,13,9,16)
Nug20	2570	(18,14,10,3,9,4, 2,12,11,16,19,15,20, 8,13,17,5,7,1,6)
Nug21	2438	(4, 21,3,9,13,2,5,14,18,11,16,10,6,15,20,19,8,7,1,12,17)
Nug22	3596	(2, 21,9,10,7,3,1,19,8,20,17,5,13,6,12,16,11,22,18,4,14,15)
Nug24	3488	(17, 8,11,23, 4,20,15,19,22,18, 3,14,1,10, 7, 9,16,21,24,12, 6,13, 5, 2)
Nug25	3744	(5,11,20,15, 22,2, 25,8, 9,1,18,16, 3,6,19,24,21,14,7,10,17,12,4,23,13)
Nug27	5234	(23,18,3,1,27,17,5,12,7,15,4,26,8,19,20,2,24,21,14, 10, 9,13,22, 25, 6, 16, 11)
Nug28	5166	(18,21,9,1,28,20,11,3 ,13,12,10,19,14,22,15,2,25,16,4,23,7,17,24, 26, 5, 27, 8, 6)
Nug30	6124	(14,5,28,24,1,3,16,15,10,9,21,2,4,29,25,22,13,26,17,30,6,20,19,8,18,7,27,12,11,23)

Table 2: Results of Genetic Algorithm for QAP.

For the same cost matrices and distance matrices we have different optimal solutions. The advantage of genetic algorithm is to have less iteration and with less time have optimal solution which is better than the used method in table 1.

5 CONCLUSION

In this paper we gave a short survey and some methods to solve QAP. We concentrate on linearization of the problem and application of genetic algorithm of the problem. We have some other uncompleted solution of the QAP by using Convex Quadratic Optimization application. New matrix formats and redefined algorithm is defined for the problem [15].. It is still in test mode.

The results obtained from Nugent problem form different instances show that genetic approach is competitive with other recent heuristic methods.

REFERENCES

[1] Adams, W.P. and Johnson, T.A. (1994), "Improved linear programming-based lower bounds for the quadratic assignment problem," DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 16, pp. 43-75.

[2] Anstreicher, K.M. and Brixius, N.W. (2001), "Solving Quadratic Assignment Problems Using Convex Quadratic Programming Relaxations," Optimization Methods And Software, 16, pp. 49-68.

[3] Anstreicher, K.M. and Brixius, N.W. (2001), "A New Bound for the Quadratic Assignment Problem Based On Convex Quadratic Programming," Mathematical Programming, 89(3), pp. 341-357.

[4] Anstreicher, K.M., Brixius, N.W., Goux J.P., and Linderoth, J. (2002), "Solving Large Quadratic Assignment Problems on Computational Grids," Mathematical Programming 91(3), pp. 563-588.

[5] Anstreicher, M. K. and Brixius, N.W. (1999), "A New Bound for the Quadratic Assignment Problem Based on Convex Quadratic Programming," http:/www.biz.uiowa.edu/faculty/anstricher/qapqp.ps2

[6] Anstreicher,K.M.(2003), "Recent Advances in The Solution of Quadratic Assignment Problems," Mathematical Programming, Series B97, pp. 27-42.

[7] Anstreicher,K.M., Brixius, N.W., J.Linderoth, and J.-P.Goux(2002), "Solving Large Quadratic Assignment Problemson Computational Grids," Mathematical Programming, Series B91, pp. 563-588.

[8] Armour, G.C. and Buffa, E.S. (1963, "Heuristic Algorithm and Simulation Approach to Relative Location of Facilities," Management Science, 9, pp. 294-309.

[9] Assad, A.A. and W. Xu (1985), "On Lower Bounds for a Class of Quadratic {0,1} Programs," Operations Research Letters 4, pp. 175-180.

[10] Bazaraa, M. S. and Kirca, O. (1983), "A branch-and-bound based heuristic for solving the quadratic assignment problem," Naval Research Logistics Quarterly, 30, pp. 287-304

[11] Bazaraa, M.S. and Elshafei, A.N. (1979), "An exact branch-and-bound procedure for the quadratic assignment problem," Naval Research Logistics Quarterly, 26, pp. 109-121.

[12] Bazaraa, M.S. and Sherali, H.D. (1979). "New approaches for solving the quadratic assignment problem," Operations Research Verfahren, 32, pp. 29-46.
[13] Bazaraa, M.S. and Sherali, H.D. (1980), "Benders' partitioning scheme applied to a new formulation of the quadratic assignment problem," Naval Research Logistics Quarterly, 27, pp. 29-41.
[14] Bazaraa, M.S. and Sherali, H.D. (1982), "On the use of exact and heuristic cutting plane methods for the quadratic assignment problem," Journal of the Operational Research Society, 33, pp. 991-1003.
[15] Brixius, N.W. and Anstreicher,, K.P. (2000). "Solving Quadratic Assignment Problems Using Convex Quadratic Programming Relaxations," http:/www.biz.uiowa.edu/faculty/anstricher/qapqp2.ps.
[16] Bruengger, A., Clausen J., Marzetta A. and Perregaard, M. (1996), "Joining Forces In Solving Large-Scale Quadratic Assignment Problems In Parallel," Diku Technical Report, University Of Copenhagen.
[17] Burkard, R.E and Bonniger, T. (1983), "A heuristic for quadratic Boolean programs with applications to quadratic assignment problems," European Journal of Operation Research, 13, pp. 374-386.
[18] Burkard, R.E. and Derigs, U. (1980). "Assignment And Matching Problems: Solution Methods With Fortran Programs," Lecture Notes In Economics And Mathematical Systems, Springer, 184.
[19] Burkard, R.E., S.E. Karisch, and F. Rendl (1997), "QAPLIB – A Quadratic Assignment Problem Library," Journal of Global Optimization, 10, pp. 391-403.
[20] Çela, E. (1998), "*The Quadratic Assignment Problem: Theory and Algorithms*," Kluwer Academic Publishers, Dordrecht, the Netherlands.
[21] Chiang, W.C. and Chiang, C. (1998), "Intelligent local search strategies for solving facility layout problems with the quadratic assignment problem formulation," European Journal of Operational Research, 106(2-3), pp. 457-488.
[22] Christofides, N. and Benavent, E. (1989), "An Exact Algorithm for the QAP," Operations Research, 37(5), pp. 60-768.
[23] Christofides, N., Mingozzi, A. and Toth, P. (1980), "Contributions to the quadratic assignment problem," European Journal of Operations Research, 4, pp. 243-247.
[24] Clausen, J. and Perregaard, M. (1997), "Solving Large Quadratic Assignment Problems In Parallel," Computational Optimization And Applications, 8(2), pp. 11-127.
[25] Clausen, J. and Perregaard, M. (1997), "Solving Large Quadratic Assignment Problems in Parallel," Computational Optimization and Applications, 8(2), pp. 11-127.
[26] Clausen, J., Espersen, T., Karisch, S.E., Perregaard, M., Sensen, N. and Tschöke, S. (1996), "Benchmark Testing For Quadratic Assignment Problems on a Portable Parallel Branch-And-Bound Library".. In progress.
[27] Drezner, Z. (1995), "Lower bounds based on linear programming for the quadratic assignment problem," Computational Optimization and Applications, 4(2), pp. 159-165.
[28] Drezner, Z. (2003), "A new genetic algorithm for the quadratic assignment problem," Informs Journal on Computing, 15(3), pp. 320-330.
[29] Drezner, Z. and Marcoulides, G.A. (2003), "A distance-based selection of parents in genetic algorithms. In: Metaheuristics: Computer Decision-Making," [edited by M.G.C. Resende and J.P. de Sousa], Combinatorial Optimization Book Series, Kluwer Academic Publishers, pp. 257-278.
[30] Elshafei, A.N. (1977), "Hospital Layout as a Quadratic Assignment Problem," Operations Research Quarterly 28, pp. 167-179.
[31] Finke, G., R.E. Burkard, and F. Rendl, (1987), "Quadratic Assignment Problems," Annals of Discrete Mathematics 31, pp. 61-82.
[32] Frieze, A.M. and Yadegar, J. (1981), "On the QAP," Discrete Applied Mathematics, 10, pp. 305-313.
[33] Gilmore, P.C. (1962), "Optimal and Suboptimal Algorithms for the Quadratic Assignment Problem," SIAM Journal on Applied Mathematics 10, pp. 305-313.
[34] Glover, F. (1989), "Tabu Search - Part 2," ORSA Journal on Computing, 2, pp. 4-32.
[35] Glover, F. (1989), "Tabu Search - Part 1," ORSA Journal on Computing, 1, pp. 190-206.
[36] Gong, D., Yamazaki, G., Gen, M. and Xu, W. (1999), "A genetic algorithm method for one dimensional machine location problems," International Journal of Production Economics, 60-1, pp. 337-342.
[37] Goux, J.P., Kulkani, S., Linderoth, J. and Yoder, M. (2000), "An Enabling Framework for Master-Worker Computing Applications on the Computational Grid," an online version is available via the World Wide Web at the following URL:http:/www.cs.berkely.edu/~sonesh/cs267/papers/mw2.ps
[38] Hadley, S.W., Rendl, F. and Wolkowicz, H. (1990), "Bounds for the Quadratic Assignment Problem Using Continuous Optimization Techniques," *Proceedings of 1st International Integer Programming and Combinatorial Optimization Conference (IPCO)*, Waterloo, Canada, May 28-30, pp. 237-248. 1
[39] Hadley, S.W., Rendl, F. and Wolkowicz, H. (1993). "A New Lower Bound via Projection for the Quadratic Assignment Problem," Mathematics of Operations Research 17, pp. 727-739.
[40] Hahn, P. and Kurtzberg, J.M. (1972). "A Review of the Placement and Quadratic assignment Problems," SIAM Review 14, pp. 324-342.

[41] Hahn, P. and Grant, T. (1998). "Lower Bounds for the Quadratic Assignment Problem Based upon a Dual Formulatio." Operations Research, 46, pp. 912-922.
[42] Hahn, P.M. and Krarup, J. (2001), "A Hospital facility layout problem finally solved," Journal of Intelligent Manufactoring, 12, pp. 487-496.
[43] Hahn, P., Grant, T. and Hall, N. (1998), "A Branch-and Bound Algorithm for The Quadratic Assignment Problem Based on The Hungarian Method," to appear in European Journal of Operational Research..
[44] Kaufman, L. and Broeckx, F. (1978), "An algorithm for the quadratic assignment problem using Bender's decomposition," European Journal of Operation Research, 2, pp. 204-211.
[45] Koopmans, T.C., and Beckman, M. (1957), "Assignment problems and the location of economic activities," Econometric, 25, pp. 53-76.
[46] Lawler, E.L. (1963), "The Quadratic Assignment Problem," Management Science, 9, pp. 586-599.
[47] Li, Y., Pardalos, P.M. and Resende, M. (1994), "A Greedy Randomized Adaptive Search Procedure for the Quadratic Assignment Problem," in Quadratic Assignment and Related Problems, P.M. Pardalos and H. Wolkowicz, eds., DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 16, pp 237-261.
[48] Li, Y., Pardalos, P.M., Ramakrishnan, K.G. and Resende, M. (1994), "Lower Bounds for the Quadratic Assignment Problem," Annals of Operations Research, 50, pp. 387-410.
[49] Li.Y and Pardalos, P.M. (1992), "Generating Quadratic Assignment Test Problems with Known Optimal Permutations," Computational Optimization and Applications, 1, pp. 163-184.
[50] Loiola, M. E., Aberu, M.M.N., Netto, B.O.P., Hann P. and Querido T. (2004), "An Analytical Survey for the QAP," technical report.
[51] Love, R.F. and Wong, J.Y. (1976a), "Solving quadratic assignment problems with rectangular distances and integer programming," Naval Research Logistics Quarterly, 23, pp. 623-627.
[52] Love, R.F. and Wong, J.Y. (1976b), "On solving a one-dimensional space allocation problem with integer programming," INFOR, 14, pp. 139-143.
[53] Miranda, G., Luna, H.P.L., Mateus, G.R. and Ferreira, R.P.M. (2004), "A performance guarantee heuristic for electronic components placement problems including thermal effects," Computers & Operational Research, Article in Press.
[54] Nugent, C.E., Vollman, T.E., and Ruml, J. (1968), "An Experimental comparison of techniques for the assignment of facillities to locations," Operations Research, 16, pp. 150-173.
[55] Padberg, M.W. and Rijal, M.P. (1996), *Location Scheduling, Design and Integer Programming*, Kluwer's Academic Publishers, MA.
[56] Pardalos, P.M. and Wolkowicz H., (1994), "Quadratic Assignment and Related Problems," Dimacs Series In Discrete Mathematics And Theoretical Computer Science, 16.
[57] Pardalos, P.M., Rendl, F. and Wolkowicz, H. (1994). "The Quadratic Assignment Problem: A Survey and Recent Developments, in Quadratic Assignment and Related Problems," P.M. Pardalos and H. Wolkowicz, eds., DIMACS Series in Discrete Mathematics and Theoretical Computer Science 16, pp. 1-42.
[58] Pitsoulis, L. (1994), "A Sparse GRASP for solving the QAP," thesis, University of Florida.
[59] Ramakrishnan, K.G., Resende, M.G.C., Ramachandran, B. and Pekny, J.F. (2002), "Tight QAP bounds via linear programming," Combinatorial and Global Optimization [edited by P.M. Pardalos, A. Migdalas, and R.E. Burkard], World Scientific Publishing, pp. 297-303.
[60] Ramakrishnan, K.G., Resende, M.G.C., Ramachandran, B., and Pekny, J.F. (2002). "Tight QAP bounds via linear programming, Combinatorial and Global Optimization", P.M. Pardalos, A. Migdalas, and R.E. Burkard, eds., World Scientific Publishing Co., pp. 297-303.
[61] Rendl, F. and Wolkowicz, H. (1992), "Applications of Parametric Programming and Eigenvalue Maximization to the Quadratic Assignment Problem," Mathematical Programming, 53, pp. 63-78.
[62] Resende, M.G.C., Ramakrishnan, K.G. and Drezner, Z. (1995), "Computing Lower Bounds for the Quadratic Assignment Problem with an Interior Point Algorithm for Linear Programming," Operations Research, 43, pp. 781-791.
[63] Rossin, D.F., Springer, M.C. and Klein, B.D. (1999), "New complexity measures for the facility layout problem: an empirical study using traditional and neural network analysis," Computers and Industrial Engineering, 36(3), pp. 585-602.
[64] Urban, T.L. (1998), "Solution procedures for the dynamic facility layout problem," Annals of Operations Research, 76, pp. 323-342.
[65] Urban, T.L., Chiang, W.C. and Russell, R.A. (2000). "The integrated machine allocation and layout problem," International Journal of Production Research, 38(13), pp. 2911-2930.

A HIGH ACCURACY STAGGERED GRID METHOD FOR INCOMPRESSIBLE FLOW

N. A. Kampanis[1], and J. A. Ekaterinaris[2]

Institute of Applied and Computational Mathematics
Foundation for Research and Technology
P.O. BOX 1527, 71110 Heraklion, Crete, Greece
Web page: http://www.iacm.forth.gr

[1] Senior Researcher, e-mail: kampanis@iacm.forth.gr,
[2] Researcher Director, e-mail: ekaterin@iacm.forth.gr

Keywords: Incompressible Navier-Stokes equation, High-order methods, Compact schemes, Runge-Kutta time stepping, Pressure correction.

Abstract. *The incompressible Navier-Stokes equations are discretized using a high-order accurate in space and time numerical method. The momentum equations are discretized on a staggered mesh using fourth-order accurate explicit or compact finite-difference schemes. High-order accuracy in time is also obtained by marching the solution with the explicit, four-stage Runge-Kutta method. At each stage of the Runge-Kutta method, incompressibility is enforced iteratively using two variants of a pressure-velocity correction technique. The first variant is applied locally and performs corrections cell by cell satisfying a fourth-order accurate discrete continuity equation. The second variant performs corrections globally by computing pressure updates from a fourth-order accurate numerical solution of a Poisson-type equation. The accuracy and efficiency of the proposed high-order method is demonstrated for various incompressible flow problems.*

1 INTRODUCTION

In many industrial flows, such as flows over aircraft wings during takeoff and landing, wind-turbine blades, and hydrofoils, accurate calculation of separated flow occurring at low speeds $M_\infty<0.1$, which may practically be considered as incompressible, is important. Incompressible flow is also encountered in biomedical applications, such as blood flow. Furthermore, the development of high-speed trains and naval transportation require detailed knowledge of the flowfield because the harsh environment of such endeavors imposes high structural and propulsive loads. In most cases, full-scale testing is difficult or impossible. Therefore, severe constraints for the evolution of new concepts are encountered. All these constitute a strong motive for the development of highly-accurate and efficient numerical methods for the simulation of incompressible flows.

An important problem for the numerical simulation of incompressible flows is coupling of the velocity and the pressure fields obtained through the momentum equations while at the same time the continuity equation, which has non-evolutionary character, is satisfied by the velocity field. Alternative formulations, such as the stream-function-velocity and the vorticity-velocity formulations[21] which remove the continuity equation and overcome this problem, are cumbersome to extent to the three dimensions. Traditional methods used for the numerical solutions of the incompressible Navier-Stokes equations are either the pressure correction based methods[5], [6] or the fractional time-step methods[2], [11], [20], [9], [16], [7], [19]. The artificial compressibility or pseudocompressibility method[1], [14], [18] is an attractive alternative to the previous methods and was used for the numerical solutions of time-dependent flows in complex domains [19], [3].

Improved aerodynamic design and newly developed conventional and unconventional flow control techniques require detailed information of the near wall flowfield, nonlinear instabilities and fine turbulent structures. Traditional, second-order accurate in space methods employed for the numerical solution of the incompressible Navier-Stokes equations[5], [6], [18] require a large number of grid points to resolve these flows. High-order upwind methods on the other hand often produce unacceptable solutions by introducing excessive amount of numerical dissipation. Spectral methods which do not suffer of such deficiencies do not apply easily for complex domains or for flows with unsteadiness caused by wall motion.

In recent years, highly accurate and robust methods have been developed. High-order accurate finite element methods based on high-order expansions of the discrete solution have been proposed[10].

Efficient, highly-accurate finite-difference and finite-volume methods on staggered grids have been also pursued for the numerical solution of incompressible flow equations[15], [13]. The important issue of conservation for various high-order finite-difference schemes on regular or staggered grids for space discretization was studied[15]. In this spirit, a conservative, finite volume approach on staggered grid was proposed[13].

In the present paper, a high-order accurate in space and time method is proposed for the numerical solution of the incompressible Navier-Stokes equations. High-order explicit finite difference formulas or compact schemes on a staggered-grid are used to discretize the convective and viscous fluxes in the momentum equations. High-order accurate compact schemes shown in Ref. [4] and analyzed in Ref. [12] are also employed as alternative to explicit finite-differences. The momentum equations are advanced in time using the explicit, fourth-order Runge-Kutta method. In the proposed algorithm, incompressibility is enforced using two different approaches; a local pressure correction technique analogous to that used in Ref. [6], or a Poisson-equation based global pressure correction method. In both cases, fourth-order accurate discretizations are used for the space derivatives involved in the continuity equation or the Poisson-type equation. The efficiency and accuracy of the proposed fourth-order accurate method is evaluated and its conservation properties are discussed and demonstrated with the numerical solution of test problems.

The rest of this paper is organized as follows: In the second section the governing equations in Cartesian coordinates are shown. In the third section the space-time discretization and the procedure to enforce incompressibility to fourth-order accuracy are presented. In the fourth section the numerical results that demonstrate the efficiency and accuracy of the proposed method are shown. The paper is closed by a concluding remarks section.

2 GOVERNING EQUATIONS

The incompressible Navier-Stokes equations in Cartesian coordinates (x,y) are

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0 \tag{1}$$

$$\frac{\partial \mathrm{u}}{\partial t}+\frac{\partial \mathrm{F}}{\partial x}+\frac{\partial \mathrm{G}}{\partial y}=-\nabla p+\frac{1}{\mathrm{Re}}\left(\frac{\partial \mathrm{F_v}}{\partial x}+\frac{\partial \mathrm{G_v}}{\partial \mathrm{y}}\right) \tag{2}$$

where $\mathrm{u}=[u,v]^{\mathrm{T}}$ is the velocity vector and p the pressure, **F** and **G** are the inviscid flux vectors given by $\mathrm{F}=[u^2,uv]^{\mathrm{T}}$, $\mathrm{G}=[uv,v^2]^{\mathrm{T}}$, Re is the Reynolds number and $\mathbf{F}_{\mathrm{v}}$ and $\mathbf{G}_{\mathrm{v}}$ are the viscous fluxes given by $\mathrm{F_v}=\left[\frac{\partial u}{\partial x},\frac{\partial v}{\partial x}\right]^{\mathrm{T}}$, $\mathrm{G_v}=\left[\frac{\partial u}{\partial y},\frac{\partial v}{\partial y}\right]^{\mathrm{T}}$.

3 SPACE-TIME DISCRETIZATION AND INCOMPRESSIBILITY ENFORCEMENT

In this paper, a high-order accurate algorithm is proposed based on the staggered grid and pressure correction procedure of Ref. [6]. Fourth-order accuracy in space is obtained using explicit or compact centered finite difference schemes. Incompressibility is enforced using two different approaches, a local pressure correction technique analogous to that used in Ref. [6], or a global pressure correction through the solution of a Poisson-equation. Both methods are applied iteratively by computing pressure corrections, which are then introduced to the discrete momentum equations to update the velocities. The pressure corrections terminate once convergence to a divergence free velocity field is reached within some tolerance. The first approach computes pressure corrections locally on each cell by requiring that the perturbed velocities obtained from a fourth-order accurate discretization of the momentum equations (2), satisfy a fourth-order accurate discrete analog of the continuity equation (1). The second approach computes pressure corrections globally, i.e. simultaneously on all cells, using a Poisson-type equation which results by enforcing zero divergence to the updated velocities. The Poisson-type equation is discretized to fourth-order accuracy. Time-marching the solution of (1)-(2) is performed using the explicit fourth-order accurate Runge-Kutta method.

3.1 Space discretization

The computational domain is divided into rectangular cells. An exterior fictitious layer of three cells, adjacent on each side of the physical domain is added to allow imposition of fourth-order accurate discrete boundary conditions. A staggered grid configuration is used to discretize the dependent variables. On each cell, the pressure p is computed at the cell center, the velocity component

u is computed at the mid of the vertical edges, and the velocity component *v* is computed at the mid of the horizontal edges. In the associated formulation the *u*-momentum equation (the first of (2)) is discretized at the mid of the right vertical edge, the *v*-momentum equation (the second of (2)) is discretized at the mid of the top horizontal edge, and the continuity equation (1) is discretized at the cell center.

The Navier-Stokes equations (1)-(2) are discretized using fourth-order accurate, explicit or compact finite-difference schemes. Explicit fourth-order accurate finite differences[15] approximate the first- and second-order derivatives at midpoints as follows:

$$\left(\frac{df}{dx}\right)_{i-\frac{1}{2}} \approx \frac{-f_{i+3/2}+8f_{i+1/2}-8f_{i-3/2}+f_{i-5/2}}{12\Delta x},$$

$$\left(\frac{d^2f}{dx^2}\right)_{i-\frac{1}{2}} \approx \frac{-f_{i+3/2}+16f_{i+1/2}-30f_{i-1/2}+16f_{i-3/2}-f_{i-5/2}}{12(\Delta x)^2}.$$

First derivatives are evaluated to fourth-order accuracy using functional values at integer points as

$$\left(\frac{df}{dx}\right)_{i-\frac{1}{2}} \approx \frac{-f_{i+1}+27f_i-27f_{i-1}+f_{i-2}}{24\Delta x}.$$

Compact, fourth-order accurate schemes[12], [4] compute first-order derivatives at integer points in a coupled fashion using functional values at integer points, by the solution of the following tridiagonal linear system

$$h\left(\phi'_{i-1}+4\phi'_i+\phi'_{i+1}\right)=3\left(\phi_{i+1}-\phi_{i-1}\right).$$

First-order derivatives at integer points are computed using functional values at midpoints by solving the following tridiagonal linear system

$$h\left(\phi'_{i-1}+22\phi'_i+\phi'_{i+1}\right)=24\left(\phi_{i+1/2}-\phi_{i-1/2}\right).$$

3.2 Time discretization

The discrete in space momentum equations resulting from (2) are written as:

$$\frac{d\mathrm{U(t)}}{dt}=\mathrm{R}\left(\mathrm{U,P}\right), \tag{3}$$

where $\mathrm{R}\left(\mathrm{U,P}\right)=-\nabla\mathrm{P}+\mathrm{A}\left(\mathrm{U}\right)$, and **A**(**U**) denotes the space discretization of the viscous and convective fluxes in (1.2). The fourth-order, explicit Runge-Kutta method when applied to (3) is written as:

$$\begin{aligned}&\mathrm{U}^{n,1}=\mathrm{U}^n,\ P^{n,1}=P^n,\\ &\mathrm{U}^{n,2}=\mathrm{U}^n+\frac{\Delta t}{2}\mathrm{R}^{n,1},\ \mathrm{U}^{n,3}=\mathrm{U}^n+\frac{\Delta t}{2}\mathrm{R}^{n,2},\ \mathrm{U}^{n,4}=\mathrm{U}^n+\Delta t\mathrm{R}^{n,3},\end{aligned} \tag{4}$$

$$\mathrm{U}^{n+1}=\mathrm{U}^n+\frac{\Delta t}{6}\left(\mathrm{R}^{n,1}+2\mathrm{R}^{n,2}+2\mathrm{R}^{n,3}+\mathrm{R}^{n,4}\right), \tag{5}$$

with $t^n=n\Delta t$, $t^{n,1}=t^n$, $t^{n,2}=t^{n,3}=t^n+\Delta t/2$, $t^{n,4}=t^n+\Delta t$, and $\mathbf{R}^{n,l}=\mathbf{R}(\mathbf{U}^{n,l}, P^{n,l})$, for l=2,3,4. The quantities $P^{n,l}$, for l=2,3,4, required for each stage of the Runge-Kutta method are determined by enforcing the incompressibility condition (1) on the velocity vectors $\mathbf{U}^{n,l}$, l=2,3,4, and $\mathbf{U}^{n+1}$ by the pressure correction procedures described in the sequel.

3.3 Incompressibility enforcement

The first approach[8], which is a fourth-order analog of the pressure correction method of Ref. [6], computes pressure corrections locally on each cell by sweeping them in a predefined order and then computing velocity updates on each cell. For each stage of the Runge-Kutta procedure, a corrected pressure $\left(p_{new}\right)_{ij}=\left(p_{old}\right)_{ij}+\Delta p$ is defined on the ij-cell. Substituting this corrected pressure in the momentum equations (2) discretized to fourth-order accuracy, a locally perturbed velocity field is obtained, and the updated velocity components are given by

$$\left(u_{new}\right)_{i\pm1/2,j}=\left(u_{old}\right)_{i\pm1/2,j}\pm a_l\Delta t\frac{27\Delta p}{24\Delta x},\ \left(u_{new}\right)_{i\pm3/2,j}=\left(u_{old}\right)_{i\pm3/2,j}\mp a_l\Delta t\frac{\Delta p}{24\Delta x},$$

$$\left(v_{new}\right)_{i,j\pm1/2}=\left(v_{old}\right)_{i,j\pm1/2}\pm a_l\Delta t\frac{27\Delta p}{24\Delta y},\ \left(v_{new}\right)_{i,j\pm3/2}=\left(v_{old}\right)_{i,j\pm3/2}\mp a_l\Delta t\frac{\Delta p}{24\Delta y}.$$

Here $a_2=a_3=1/2$ and $a_4=1$, and correspond to the coefficients of the Runge-Kutta stages. The pressure correction Δp is computed by requiring that the updated velocities satisfy locally on the ij-cell a fourth-order accurate discrete analog of (1). All cells are swept in a predefined order. The entire procedure is repeated iteratively until global incompressibility is achieved.

The second approach[8] introduces a globally corrected pressure for all cells

$$\left(p_{new}\right)_{ij}=\left(p_{old}\right)_{ij}+\Delta p_{ij}.$$

Using this correction, perturbed velocities are defined globally on all cells as follows:

$$\left(u_{new}\right)_{i+1/2,j}=\left(u_{old}\right)_{i+1/2,j}-a_l\Delta t\left(\partial_x\Delta p\right)_{i+1/2,j},\ \left(v_{new}\right)_{i,j+1/2}=\left(v_{old}\right)_{i,j+1/2}-a_l\Delta t\left(\partial_y\Delta p\right)_{i,j+1/2}.$$

When the perturbed velocities are subjected to a fourth-order accurate approximation of the incompressibility condition (1) the updates Δp_{ij} can be determined simultaneously by solving the following discrete Poisson-type equation

$$\left(\partial_x^2+\partial_y^2\right)\Delta p_{ij}=\frac{\left(D_{old}\right)_{ij}}{a_l\Delta t},\tag{6}$$

where $\partial_x^2\Delta p_{ij}$, $\partial_y^2\Delta p_{ij}$ are evaluated by fourth-order accurate finite difference schemes, and $\left(D_{old}\right)_{ij}$ is the discrete divergence (not necessarily zero) of the original velocities u_{old}, v_{old}. The whole procedure is also repeated iteratively until global incompressibility is achieved.

4 NUMERICAL RESULTS

In this section the efficiency and accuracy of the proposed space-time discretization method are evaluated for the steady-state solution of low Reynolds number driven cavity flow and for the time-dependent solution of the decaying Oseen's vortex. The solutions computed with the proposed high-order method are compared with the exact solutions and solutions obtained with an established second-order method[6], [18]. Fourth-order of accuracy is demonstrated for the proposed method. Fourth-order in space is achieved by both the explicit and the compact fourth-order accurate differencing schemes. The computations were performed with CFL numbers below the stability limit of the Runge-Kutta method. The discrete divergence free condition is enforced to machine zero accuracy in all cases.

4.1 Driven cavity flow

For the numerical solution of the square driven cavity problem a zero normal pressure gradient is assumed for all walls and a non slip condition for all walls is imposed except for the top wall where u=1 and v=0. Fully converged solutions were obtained by the proposed fourth-order accurate method at a Reynolds number Re=100 on the series of 111x111, 121x121, 131x131, 141x141, 151x151, 161x161, 181x181, 191x191, 201x201, and 221x221 point uniform grids. For comparison, numerical solutions were also obtained by a second-order accurate method[6] on the same grids. In addition, a solution was computed by the second-order accurate method on a very fine 301x301 point grid.

In Fig. 1, a comparison of the u-velocities computed by the fourth-order accurate method on the 201x201 and 221x221 point grids and by the second-order accurate method on the 221x221 and 301x301 point grids and the results of Ref. [18] is shown. In Fig. 2, an analogous comparison for the v-velocities is shown. Second-order accurate solutions on finer grids show larger deviations to the solution of Ref. [18] than those of the fourth-order accurate solution on coarser grids. The solutions shown in Figs. 1 and 2, were obtained after integrating up to 10 time units. Since no difference was fond by integrating up to 15 or 20 time units, the steady-state solution was safely reached for 10 time units.

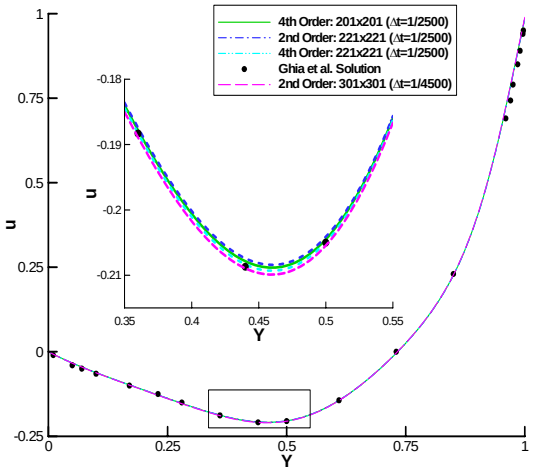


Figure 1: Driven cavity flow (Re=100). Comparison of u-velocity component computed by the fourth- and second-order accurate methods with a reference solution of Ref. [18].

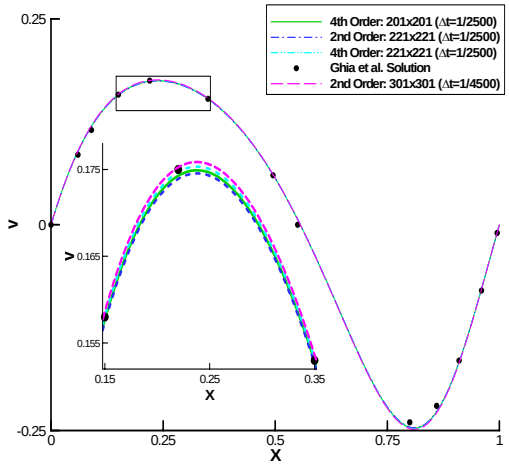


Figure 2: Driven cavity flow (Re=100). Analogous comparison for the v-velocity component.

An a posteriori estimation of the spatial discretization error of the proposed high-order method is shown in Fig. 3. A very fine grid solution computed on the 301x301 point grid by an established second order accurate method[6] is used as reference solution u_R. The L^2 norms of the error u_N- u_R , where u_N denotes the fourth-order accurate solutions computed on the NxN point grids, with N=111, 121, 131, 141, 151, 161, 181, 191, 201, and 221, are shown in Fig. 3. It is assumed that the reference solution u_R has negligible error compared to the 'exact' solution u_E in the L^2 norm. Therefore, the L^2 norms of the error u_N- u_R are good approximations of the L^2 norms of the error u_N- u_E (the error between the computed solutions and the 'exact' solution). It is evident that the observed order of accuracy of the proposed fourth-order method is fair.

The numerical solutions obtained on the series of grids considered by using compact schemes for the space discretization in the proposed fourth-order accurate method coincide. Regarding the two different approaches for enforcing incompressibility the following behavior was observed. For the low Reynolds number Re=100 numerical solution, the local pressure correction method required less computing time than the Poison-equation, global pressure correction method to enforce incompressibility to machine accuracy (10^{-16}). However, it was observed that the number of corrections per stage/ time-step of the Runge-Kutta method, required by the Poisson-equation pressure correction method, is smaller compared to that for the local pressure correction method on the same grid. The local correction method in total required smaller computing time because for each correction sweep the computing time for solving numerically the Poisson-type equation is larger than that for the simple local correction procedure.

The observation that the number of iterations for the Poisson-equation, global pressure correction method was in all cases less that the number of iterations required by the local pressure correction method prompted further comparisons the higher Reynolds numbers Re=500, 1000, 2000 and 10000. An important outcome of this study was that the increase of the Reynolds number comparatively reduces the number of iterations required per time-step by the Poisson-equation pressure correction to

the number of iterations required per time-step by the local pressure correction. Therefore, it appears that with the progressive increase of the Reynolds number, the Poisson-equation pressure correction method becomes more attractive alternative in terms of efficiency than the local pressure correction method.

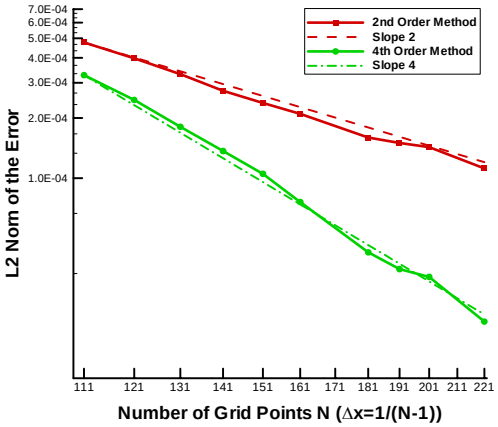


Figure 3: Driven cavity flow (*Re*=100). Grid convergence shown by the L^2 norm of the error obtained by the fourth- and second-order accurate methods.

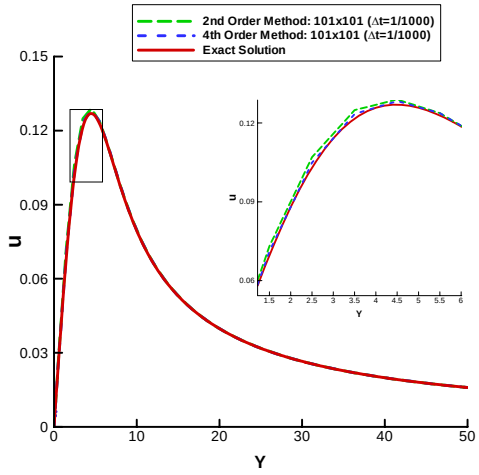


Figure 4: Oseen vortex decay. Comparison of the *u*-velocity computed by the fourth- and second-order methods on the 101x101 point grid with the exact solution at *T*=4.

4.2 Oseen vortex decay

The decay of an ideal vortex is an unsteady flow problem of interest to numerical simulations of trailing vortices, (LES), and (DES) simulations. The initial velocity profile of the Oseen vortex is $v_\theta(r,t=0)=\Gamma/2\pi r$, where Γ is the strength of the vortex and *r* is the distance from the origin. The initial pressure distribution is constant. Under the effects of viscous forces this vortex decays and the velocity at time *t* is given by

$$v_\theta(r,t)=\frac{\Gamma}{2\pi r}\left[1-e^{\left(-\frac{r^2 Re}{4t}\right)}\right]. \tag{7}$$

The time-dependent flow for the numerical experiments shown below was computed with Γ=5. In Fig. 4 the numerical solutions obtained on a uniform 101x101 point Cartesian grid, using the fourth- and the second-order accurate methods were compared with the exact solution (7) for time *T*=4. The superiority of the fourth-order accurate method when compared to the second-order accurate method is evident. In Fig. 5 numerical solutions obtained on a 1001x1001 point grid by the fourth- and the second-order accurate methods for a very large Reynolds number *Re*= 10^{10}, i.e. for practically inviscid flow, are compared with the exact solution. In both cases, a small time step of 1/1000, well below the CFL stability limit of the Runge-Kutta method, was used to ensure time accuracy.

The kinetic energy conservation of the proposed fourth-order accurate method is shown in Fig. 6. The kinetic energy is defined as $\mathrm{E}(t)=\iint u^2(r,\theta,t)rdrd\theta$, is plotted for *Re*= 1, for the exact solution and the fourth-order accurate numerical solution computed on the 101x101 point grid. For the space discretization the explicit fourth-order accurate finite difference schemes of subsection 3.1 are used. The kinetic energy is plotted in the same figure for the inviscid limit *Re*= 10^{10}, for the exact solution and the computed fourth-order accurate numerical on the 1001x1001 point grid. In both cases, the kinetic energy conservation of the proposed fourth-order accurate method is evident. The same behavior was exhibited by the fourth-order accurate compact schemes of subsection 3.2, when used for the space discretization in the proposed time-space discretization method.

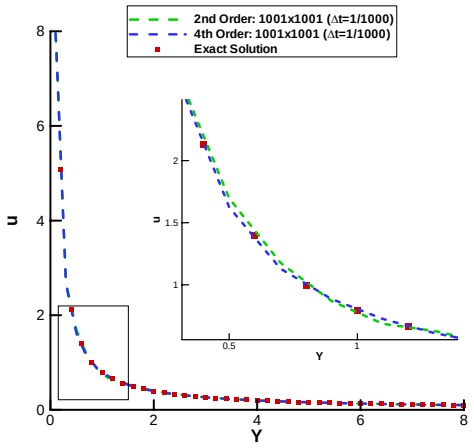


Figure 5: Oseen vortex decay. Comparison of the *u*-velocity computed by the fourth- and second-order methods on the 1001x1001 point grid with the exact solution for $Re=10^{10}$, at $T=4$.

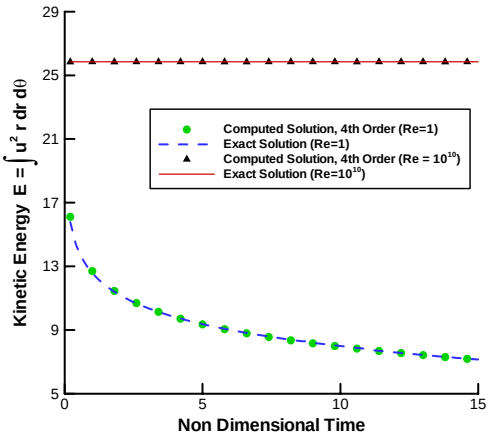


Figure 6: Oseen vortex decay. Comparison of the kinetic energy **E** as a function of time obtained from the *u*-velocity computed by the fourth-order accurate method for $Re=1$ (on the 101x101 point grid) and $Re=10^{10}$ (on the 1001x1001 point grid) and the corresponding exact solution.

5 CONCLUSIONS

A fourth-order accurate in space and time method, for the numerical solution of the incompressible Navier-Stokes equations was shown. The method uses fourth-order accurate explicit or compact schemes for the space discretization. The fourth-order explicit Runge-Kutta method is used for time marching the numerical solution. The computed velocities at each stage/ time-step of Runge-Kutta procedure are subjected to the incompressibility condition using two different approaches. The first approach iteratively computes pressure updates locally on each cell which then proceeds in to local velocity corrections until incompressibility is enforced. The second method solves a global Poisson-type equation to compute pressure updates simultaneously for all cells; velocity is recomputed globally. The space derivates in the continuity equation (1) or the Poisson-type equation (4) used by the two approaches for incompressibility enforcement, are discretized to fourth-order accuracy.

The accuracy of the proposed fourth-order method is accessed. Computed solutions by the proposed method are compared with exact solutions for incompressible flow problems. Numerical results for the steady-state flow in a driven cavity and the unsteady flow for the decay of the Oseen vortex are presented. The proposed fourth-order accurate method is more efficient than the second-order accurate method. Fourth-order accuracy in space was established for the proposed method both when fourth-order accurate explicit or compact finite difference schemes were used for the space discretization. Fourth-order accuracy in time was also achieved. The local pressure correction method was observed to perform better for low Reynolds numbers than the Poisson-equation pressure correction method. The later appears to become a more attractive alternative as the Reynolds number increases.

REFERENCES

[1] Chorin, A. J. (1967), "A numerical method for solving incompressible, viscous flow problems," J. Comput. Physics 2, pp. 12-26.

[2] Chorin, A. J. (1968), "Numerical solution of the Navier-Stokes equations," Math. Comp. 22, pp. 745-762.

[3] Ekaterinaris, J. A. (1998), "Numerical simulation of incompressible two-blade rotor flowfields," AIAA Journal of Propulsion and Power 14, pp. 367-374.

[4] Gaitonde, D. V. and Visbal, M. R. (1998), *High--order schemes for Navier-Stokes equations: Algorithms and implementation into FDL3DI*, Air Vehicles Directorate, Air Force Research Laboratory, Write-Patterson Air Force Base, Ohio 45433-7913, AFRL-VA-WP-TR-1998-3060.

[5] Harlow, F. H. and Welch, J. E. (1965), "Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface," Physics of Fluids 8, pp. 2182-2189.

[6] Hirt, C. W., Nichols, B. D., and Romero, N. C. (1975), *SOLA-A numerical solution algorithm for transient fluid flows*, Los Alamos Report, LA-5852.

[7] Issa, R. I. (1985), "Solution of the implicitly discretised fluid flow equations by operator-splitting," J. Comput. Physics 62, pp. 40-65.

[8] Kampanis, N. A. and Ekaterinaris, J. A. (2004), "A staggered grid high-order accurate method for incompressible viscous flow," AIAA-2004-0432.

[9] Karniadakis, G. E., Israeli, M., and Orszag, S. (1991), "High-order splitting methods for the incompressible Navier-Stokes equations," J. Comput. Physics 97, pp. 414-443.

[10] Karniadakis, G. E. and Sherwin, S. J (1999), *Spectral/hp Element Method for CFD*, Oxford University Press.

[11] Kim, J. and Moin, P. (1985), "Application of a fractional-step method to incompressible Navier-Stokes equations," J. Comput. Physics 59, pp. 308-323.

[12] Lele, S. K. (1992), "Compact finite difference schemes with spectral-like resolution," J. Comput. Physics 103, pp. 17-42.

[13] Mahesh, K., Constantinescu, G., and Moin, P. (2004), "A numerical method for large-eddy simulation in complex geometries," J. Comput. Physics 197, pp. 215-240.

[14] Merkle, C. L. and Athavale, M. (1987), "Time-accurate unsteady incompressible flow algorithms based on artificial compressibility," AIAA Paper 87-1137.

[15] Morinishi, Y., Lund, T. S., Vasilyev, O. V., and Moin, P. (1998), "Fully conservative higher order finite difference schemes for incompressible flow," J. Comput. Physics 143, pp. 90-124.

[16] Patankar, S. V. (1980), *Numerical Heat Transfer and Fluid Flow*, Hemisphere Publishing Co.

[17] Rhie, C. M. and Chow, W. L. (1983), "Numerical study of the turbulent flow past an airfoil with trailing edge separation," AIAA Journal 21, pp. 1525-1532.

[18] Rogers, S. E. (1990), *Numerical solution of the incompressible Navier-Stokes equations*, NASA TM 102199, Ames Research Center, Moffett Field, CA.

[19] Rogers, S. E., Kwak, D., and Kiris, C. (1991), "Steady and unsteady solutions of the incompressible Navier-Stokes equations," AIAA Journal 29, pp. 603-610.

[20] Rosenfeld, M., Kwak, D., and Vinokur, M. (1991), "A fractional step solution method for the unsteady incompressible Navier-Stokes equations in generalized curvilinear coordinate systems," J. Comput. Physics 94, pp. 102-137.

[21] Wu, J. C. (1981), "Theory of aerodynamic force and moment in viscous flow," AIAA Journal 19, pp. 432-441.

THE MOSAIC OF HIGH PERFORMANCE DOMAIN DECOMPOSITION METHODS FOR SOLID AND STRUCTURAL MECHANICS – THEORY AND PRAXIS

Yannis Fragakis[a,b], Manolis Papadrakakis[b]

[a]International Center for Numerical Methods in Engineering (CIMNE)
Technical University of Catalonia
Edificio C1, Campus Norte, Gran Capitan s/n, Barcelona, 08034, Spain
e-mail: fragayan@cimne.upc.edu, web page: http://users.ntua.gr/fragayan

[b]Institute of Structural Analysis & Seismic Research
National Technical University Athens
9, Iroon Polytechniou, Zografou Campus, GR-15780 Athens, Greece
e-mail: mpapadra@central.ntua.gr, web page: http://users.civil.ntua.gr/papadrakakis

Keywords: High-Performance Computing, Domain Decomposition Methods, P-FETI Methods.

Abstract. *In the past years, Domain Decomposition Methods (DDM) emerged as advanced solvers in several areas of Computational Mechanics. In particular, during the last decade, in the area of Solid and Structural Mechanics, they reached a considerable level of advancement and were shown to be more efficient than popular solvers, like advanced sparse direct solvers. The present contribution follows the lines of a series of recent publications by the authors on DDM. In these papers, the authors developed a unified theory of primal and dual methods and presented a family of DDM that were shown to be more efficient than previous methods. The present paper extends this work, presenting a new family of related DDM, thus enriching the theory of the relations between primal and dual methods.*

1 INTRODUCTION

In the last decade Domain Decomposition Methods (DDM) have undergone a significant progress leading to a large number of methods and techniques, capable of giving solution to various problems of Computational Mechanics. In the field of Solid And Structural Mechanics, in particular, this fruitfull period led to the extensive parallel development of two large families of methods: (a) the Finite Element Tearing and Interconnecting (FETI) methods and (b) the Balancing Domain Decomposition (BDD) methods. Both introduced at the beginning of the 90s[1,2], these two categories of methods today include a large number of variants. However, their distinct theories led to the lack of extensive studies to interconnect them in the past. Thus, in the present decade two studies[3,4] attempted to determine the relations between the two methods.

In particular, the studies [4,5] set the basis of a unified theory of primal and dual DDM. This effort also led to the introduction of a new family of methods, under the name "Primal class of FETI methods", or in abbreviation "P-FETI methods". These methods are derived from the Dirichlet preconditioned FETI methods. They, thus, inherit the high computational efficiency properties of these methods, while their primal flavour gives them increased efficiency and robustness in ill-conditioned problems. However, so far there has not been presented a primal alternative for the lumped preconditioned FETI methods. Filling this hole is the object of the present study and even though the new formulations do not appear to share the same advantages as the P-FETI formulations, they serve the purpose of diversifying our knowledge of the relations of primal and dual methods.

This paper, thus, presents the primal alternatives of the lumped preconditioned FETI methods and is organised as follows: Section 2 presents the base formulation of the introduced methods. Section 3 sets up some algebraic relations that lead to the transformation of the algorithms in a more economical form, which is derived in the section that follows. Finally, section 5 presents numerical results for comparing the new formulation with previous ones and section 6 gives some concluding statements.

2 BASIC FORMULATION OF THE PRIMAL ALTERNATIVES OF THE FETI METHODS EQUIPPED WITH THE LUMPED PRECONDITIONER

The P-FETI methods were built on the concept of preconditioning the Schur complement method with the first estimate of displacements obtained during the FETI methods. Accordingly, the primal counterparts of the lumped preconditioned methods will be obtained by similarly preconditioning the intact global problem. Thus, the following equation

$$Ku = f \Leftrightarrow L^{\mathrm{T}} K^{s} L u = L^{\mathrm{T}} f^{s} \tag{1}$$

will be preconditioned with the first displacement estimate of a FETI method. In eq. (1), K, u, and f represent the global stiffness matrix, displacement and force vectors, respectively, while

$$K^{s} = \begin{bmatrix} K^{(1)} & & \\ & \ddots & \\ & & K^{(n_s)} \end{bmatrix} \quad , \quad u^{s} = \begin{bmatrix} u^{(1)} \\ \vdots \\ u^{(n_s)} \end{bmatrix} \quad \text{and} \quad f^{s} = \begin{bmatrix} f^{(1)} \\ \vdots \\ f^{(n_s)} \end{bmatrix} \tag{2}$$

are the matrix block-diagonal assemblage of the correponding quantities of subdomains $s = 1,..., n_s$ and L is a Boolean restriction matrix, such that $u^{s} = Lu$. Using the original FETI formulation, usually refered to as "one-level FETI" or "FETI-1", the following preconditioner for (1) is derived (this equation is obtained following an analysis almost identical to [4, section 6]):

$$\tilde{A}^{-1} = L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} L_p \tag{3}$$

where: $$\tilde{A}^{s^{-1}} = H^{\mathrm{T}} K^{s^{+}} H \quad , \quad H = I - B^{\mathrm{T}} Q G (G^{\mathrm{T}} Q G)^{-1} R^{s^{\mathrm{T}}} \quad \text{and} \quad G = B R^{s} \tag{4}$$

Here, R^{s} and $K^{s^{+}}$ are the block-diagonal assemblage of subdomain zero energy modes and generalized inverses of subdomain stiffness matrices, respectively, B is a mapping matrix such that $\mathrm{null}(B) = \mathrm{range}(L)$, Q is a symmetric positive definite matrix used in the FETI-1 coarse projector (see for instance [6]), while L_p and B_p are scaled variants of L and B (see the expressions gathered from various DDM papers in [4]). Similar ideas lead to the corresponding preconditioners that are derived from other FETI variants. Comparing the lumped preconditioned FETI-1 method with the method of this section, it is noted that the present method has a significantly higher computational cost, because it operates on the full displacement vector u of the structure and also needs multiplications with the full stiffness matrices of the subdomains. In order to diminish its cost, this algorithm will be transformed into a more economical version, by respresenting its primal variables with dual variables.

3 AUXILIARY EQUATIONS LEADING TO THE CHANGE OF VARIABLES OF THE ALGORITHM

In order to perform the transformation of variables, we need to prove some relations of the introduced matrices. First, the L and B matrices satisfy the (some of the most complete studies on these equations can be found in [3, 7]:

$$\mathrm{range}(L) = \mathrm{null}(B) \quad \text{and} \quad \mathrm{range}(L_p) = \mathrm{null}(B_p) \tag{5}$$

$$(L_p L^{\mathrm{T}})^2 = L_p L^{\mathrm{T}} \quad \text{and} \quad (B^{\mathrm{T}} B_p)^2 = B^{\mathrm{T}} B_p \tag{6}$$

$$L^{\mathrm{T}} L_p = I \quad \text{and} \quad L_p L^{\mathrm{T}} + B^{\mathrm{T}} B_p = I \tag{7}$$

H is a projector satisfying

$$R^{s^{\mathrm{T}}} H = 0 \quad , \quad H B^{\mathrm{T}} Q G = 0 \quad , \quad H^2 = H \quad \text{and} \quad H^{\mathrm{T}} L = L \tag{8}$$

The stiffness matrices satisfy the relations:

$$K^{s} R^{s} = 0 \quad \text{and} \quad H K^{s} = K^{s} \tag{9}$$

and the property: There is a matrix Y such that:

$$K^{s^{+}} K^{s} = I + R^{s} Y \tag{10}$$

Using eqs. (8) - (10), we obtain:

$$\tilde{A}^{s^{-1}} K^{s} = H^{\mathrm{T}} K^{s^{+}} H K^{s} = H^{\mathrm{T}} K^{s^{+}} K^{s} = H^{\mathrm{T}} \left(I + R^{s} Y \right) = H^{\mathrm{T}} + H^{\mathrm{T}} R^{s} Y = H^{\mathrm{T}} \tag{11}$$

Using eqs. (5), (7), (8) and (11), we obtain the following equations:

$$L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} K^{s} L = L_p^{\mathrm{T}} H^{\mathrm{T}} L = L_p^{\mathrm{T}} L = I \quad , \quad B \tilde{A}^{s^{-1}} K^{s} L = B H^{\mathrm{T}} L = B L = 0 \tag{12}$$

$$K\tilde{A}^{-1} = L^{\mathrm{T}} K^{s} L\, L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} L_p = L^{\mathrm{T}} K^{s} (I - B_p^{\mathrm{T}} B) \tilde{A}^{s^{-1}} L_p =$$
$$L^{\mathrm{T}} K^{s} \tilde{A}^{s^{-1}} L_p - L^{\mathrm{T}} K^{s} B_p^{\mathrm{T}} B \tilde{A}^{s^{-1}} L_p = I - L^{\mathrm{T}} K^{s} B_p^{\mathrm{T}} B \tilde{A}^{s^{-1}} L_p \tag{13}$$

$$L^{\mathrm{T}} K^{s} L L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} = L^{\mathrm{T}} K^{s} \left(I - B_p^{\mathrm{T}} B \right) \tilde{A}^{s^{-1}} B^{\mathrm{T}}$$
$$= L^{\mathrm{T}} K^{s} \tilde{A}^{s^{-1}} B^{\mathrm{T}} - L^{\mathrm{T}} K^{s} B_p^{\mathrm{T}} B \tilde{A}^{s^{-1}} B^{\mathrm{T}} = -L^{\mathrm{T}} K^{s} B_p^{\mathrm{T}} B \tilde{A}^{s^{-1}} B^{\mathrm{T}} \tag{14}$$

In addition, using eqs. (7) and (8):

$$HB^{\mathrm{T}} B_p H = H(I - L_p L^{T}) H = HH - HL_p L^{T} H$$
$$= H - HL_p L^{T} = H(I - L_p L^{T}) = HB^{\mathrm{T}} B_p \tag{15}$$

$$\tilde{A}^{s^{-1}} K^{s} B_p^{\mathrm{T}} B H^{\mathrm{T}} = H^{\mathrm{T}} B_p^{\mathrm{T}} B H^{\mathrm{T}} = B_p^{\mathrm{T}} B H^{\mathrm{T}} \quad \text{and} \quad L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} K^{s} B_p^{\mathrm{T}} B H^{\mathrm{T}} = L_p^{\mathrm{T}} B_p^{\mathrm{T}} B H^{\mathrm{T}} = 0 \tag{16}$$

Also, using eqs. (5) and (7), it follows from eq. (16):

$$B\tilde{A}^{s^{-1}} K^{s} B_p^{\mathrm{T}} B H^{\mathrm{T}} = B B_p^{\mathrm{T}} B H^{\mathrm{T}} = B(I - L L_p^{\mathrm{T}}) H^{\mathrm{T}} = B H^{\mathrm{T}} - B L L_p^{\mathrm{T}} H^{\mathrm{T}} = B H^{\mathrm{T}} \tag{17}$$

and using eqs. (7), (16):

$$L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} L_p L^{\mathrm{T}} K^{s} B_p^{\mathrm{T}} B H^{\mathrm{T}} = L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} \left(I - B^{\mathrm{T}} B_p \right) K^{s} B_p^{\mathrm{T}} B H^{\mathrm{T}}$$
$$= L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} K^{s} B_p^{\mathrm{T}} B H^{\mathrm{T}} - L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} B_p K^{s} B_p^{\mathrm{T}} B H^{\mathrm{T}} = -L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} B_p K^{s} B_p^{\mathrm{T}} B H^{\mathrm{T}} \tag{18}$$

4 FINAL FORM OF THE ALGORITHM

In this section, it will be shown that if the initial solution vector of the PCG algorithm applied for the solution of eq. (1) , with the preconditioner of eq. (3), is set equal to (In the following of this section we use the notation and steps of Algorithm 1):

$$u^{0} = \tilde{A}^{-1} f \tag{19}$$

then there exist suitable vectors (denoted below with the subscript "1"), such that the following variables of the PCG can be written in the forms ($k = 0,1,\ldots$):

$$z^{k} = -L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} z_1^{k} \quad , \quad p^{k} = -L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} p_1^{k} \quad , \quad r^{k} = L^{\mathrm{T}} K^{s} B_p^{\mathrm{T}} r_1^{k} \quad \text{and} \quad q^{k} = L^{\mathrm{T}} K^{s} B_p^{\mathrm{T}} q_1^{k} \tag{20}$$

Eqs. (20) allow expressing the PCG vectors, which have the size of the total number of degrees of freedom (d.o.f.), with respect to vectors whose size is equal to the row size of matrix B (which in turn is equal to the number of Lagrange multipliers used in dual DDM). They thus allow reducing the cost of the algorithm. Before proceeding with the proof, we will study how the linear combinations and dot products performed by the PCG algorithm are transformed due to eqs. (20).

The linear combinations of the PCG vectors are simply transformed to linear combinations of Lagrange multiplier vectors. For instance: using eq. (20) the PCG residual becomes:

$$r^{k} = r^{k-1} - \eta^{k-1} q^{k-1} = L^{\mathrm{T}} K^{s} B_p^{\mathrm{T}} r_1^{k-1} - \eta^{k-1} L^{\mathrm{T}} K^{s} B_p^{\mathrm{T}} q_1^{k-1}$$
$$= L^{\mathrm{T}} K^{s} B_p^{\mathrm{T}} (r_1^{k-1} - \eta^{k-1} q_1^{k-1}) \Rightarrow r_1^{k-1} = r_1^{k-1} - \eta^{k-1} q_1^{k-1} \tag{21}$$

In order to discuss the transformation of the PCG dot products, we define vectors:

$$z_2^{k} = B\tilde{A}^{s^{-1}} B^{\mathrm{T}} z_1^{k} \quad \text{and} \quad z_3^{k} = B_p K^{s} B_p^{\mathrm{T}} z_2^{k} \quad , \quad p_2^{k} = B\tilde{A}^{s^{-1}} B^{\mathrm{T}} p_1^{k} \quad \text{and} \quad p_3^{k} = B_p K^{s} B_p^{\mathrm{T}} p_2^{k} \tag{22}$$

$$r_2^{k} = B_p K^{s} B_p^{\mathrm{T}} r_1^{k} \quad \text{and} \quad r_3^{k} = B\tilde{A}^{s^{-1}} B^{\mathrm{T}} r_2^{k} \quad , \quad q_2^{k} = B_p K^{s} B_p^{\mathrm{T}} q_1^{k} \quad \text{and} \quad q_3^{k} = B\tilde{A}^{s^{-1}} B^{\mathrm{T}} q_2^{k} \tag{23}$$

Then for instance (using eqs. (7) and (20)), the following dot product becomes:

$$p^{k^{\mathrm{T}}} q^{k} = -p_1^{k^{\mathrm{T}}} B\tilde{A}^{s^{-1}} L_p L^{\mathrm{T}} K^{s} B_p^{\mathrm{T}} q_1^{k} = -p_1^{k^{\mathrm{T}}} B\tilde{A}^{s^{-1}} (I - B^{\mathrm{T}} B_p) K^{s} B_p^{\mathrm{T}} q_1^{k}$$
$$= -p_1^{k^{\mathrm{T}}} B\tilde{A}^{s^{-1}} K^{s} B_p^{\mathrm{T}} q_1^{k} + p_1^{k^{\mathrm{T}}} B\tilde{A}^{s^{-1}} B^{\mathrm{T}} B_p K^{s} B_p^{\mathrm{T}} q_1^{k} \tag{24}$$

- Initialize

$$r^0 = b - Ku^0 \quad , \quad z^0 = \tilde{A}^{-1} r^0 \quad , \quad p^0 = z^0 \quad , \quad q^0 = Kp^0 \quad , \quad \eta^0 = \frac{p^{0^{\mathrm{T}}} r^0}{p^{0^{\mathrm{T}}} q^0}$$

- Iterate $k = 1, 2, \ldots$ until convergence

$$u^k = u^{k-1} + \eta^{k-1} p^{k-1} \quad , \quad r^k = r^{k-1} - \eta^{k-1} q^{k-1} \quad , \quad z^k = \tilde{A}^{-1} r^k$$

$$p^k = z^k - \sum_{i=0}^{k-1} \frac{z^{k^{\mathrm{T}}} q^i}{p^{i^{\mathrm{T}}} q^i} p^i \quad , \quad q^k = Kp^k \quad , \quad \eta^k = \frac{p^{k^{\mathrm{T}}} r^k}{p^{k^{\mathrm{T}}} q^k}$$

Algorithm 1. The PCG algorithm for solving system $Ku = f$ preconditioned with $\tilde{A}^{-1}$ (full reorthogonalization)

Here, it is needed to make the following assumption, which will also be verified in the proof that follows: $q_1^k \in$ range(BH^{T}). Thus, there is a vector y such that $q_1^k = BH^{\mathrm{T}} y$ and the first term of eq. (24) becomes (using eq. (17)):

$$-p_1^{k^{\mathrm{T}}} B\tilde{A}^{s^{-1}} K^s B_p^{\mathrm{T}} q_1^k = -p_1^{k^{\mathrm{T}}} B\tilde{A}^{s^{-1}} K^s B_p^{\mathrm{T}} BH^{\mathrm{T}} y = -p_1^{k^{\mathrm{T}}} BH^{\mathrm{T}} y = -p_1^{k^{\mathrm{T}}} q_1^k \tag{25}$$

For the second term, there are the following three choices (taking into account the assumptions (20) and the definitions (22) - (23)):

$$p_1^{k^{\mathrm{T}}} B\tilde{A}^{s^{-1}} B^{\mathrm{T}} B_p K^s B_p^{\mathrm{T}} q_1^k = p_2^{k^{\mathrm{T}}} q_2^k \tag{26}$$

or:
$$p_1^{k^{\mathrm{T}}} B\tilde{A}^{s^{-1}} B^{\mathrm{T}} B_p K^s B_p^{\mathrm{T}} q_1^k = p_1^{k^{\mathrm{T}}} q_3^k \tag{27}$$

or:
$$p_1^{k^{\mathrm{T}}} B\tilde{A}^{s^{-1}} B^{\mathrm{T}} B_p K^s B_p^{\mathrm{T}} q_1^k = p_3^{k^{\mathrm{T}}} q_1^k \tag{28}$$

Thus, using each of these choices and expression (25), the dot product (24) takes the following three expressions:

$$p^{k^{\mathrm{T}}} q^k = p_2^{k^{\mathrm{T}}} q_2^k - p_1^{k^{\mathrm{T}}} q_1^k \tag{29}$$

or:
$$p^{k^{\mathrm{T}}} q^k = -p_1^{k^{\mathrm{T}}} q_1^k + p_1^{k^{\mathrm{T}}} q_3^k = p_1^{k^{\mathrm{T}}} (q_3^k - q_1^k) \tag{30}$$

or:
$$p^{k^{\mathrm{T}}} q^k = -p_1^{k^{\mathrm{T}}} q_1^k + p_3^{k^{\mathrm{T}}} q_1^k = (p_3^{k^{\mathrm{T}}} - p_1^{k^{\mathrm{T}}}) q_1^k \tag{31}$$

where it is noted that options (30) and (31) are more cost effective than option (29). Like dot product $p^{k^{\mathrm{T}}} q^k$ which was used here as an example, all dot products of the PCG can be expressed in the form of dot products of Lagrange multiplier vectors, like in eqs. (29) - (31).

We are now ready to proceed to the proof of eqs. (20), proving also that there are vectors y such that $q_1^k = BH^{\mathrm{T}} y$ and $r_1^k = BH^{\mathrm{T}} y$ and obtaining the transformed algorithm with respect to Lagrange multiplier vectors. We simply follow the steps of Algorithm 1. Thus, from eq. (19) it follows (using eq. (13)):

$$r^0 = f - Au^0 = f - A\tilde{A}^{-1} f = f - (I - L^{\mathrm{T}} K^s B_p^{\mathrm{T}} B\tilde{A}^{s^{-1}} L_p) f = L^{\mathrm{T}} K^s B_p^{\mathrm{T}} B\tilde{A}^{s^{-1}} L_p f \tag{32}$$

and:
$$r_1^0 = B\tilde{A}^{s^{-1}} L_p f = BH^{\mathrm{T}} (K^{s^+} HL_p f) \tag{33}$$

Computing the residual r^0 from the above equations, we get:

$$r^0 = L^{\mathrm{T}} K^s B_p^{\mathrm{T}} r_1^0 \Rightarrow \begin{bmatrix} r_b^0 \\ r_{i(b)}^0 \end{bmatrix} = \begin{bmatrix} L_b^{\mathrm{T}} \\ I \end{bmatrix} \begin{bmatrix} K_{bb}^s & K_{ib}^s \\ K_{ib}^s & K_{ii}^s \end{bmatrix} \begin{bmatrix} B_{p_b}^{\mathrm{T}} \\ 0 \end{bmatrix} r_1^0 = \begin{bmatrix} L_b^{\mathrm{T}} K_{bb}^s \\ K_{ib}^s \end{bmatrix} B_{p_b}^{\mathrm{T}} r_1^0 \tag{34}$$

where subscripts b and i restrict the matrices to interface or internat d.o.f. of the subdomains, respectively. In eq. (34) it is worth noting that the residual vanishes in internal d.o.f. of the subdomains, when these d.o.f. are not adjucent to interface d.o.f., which is also observed in the lumped preconditioned FETI methods. Futhermore:

$$r_2^0 = B_p K^s B_p^{\mathrm{T}} r_1^0 = \begin{bmatrix} B_{p_b}^{\mathrm{T}} \\ 0 \end{bmatrix} \begin{bmatrix} K_{bb}^s & K_{ib}^s \\ K_{ib}^s & K_{ii}^s \end{bmatrix} \begin{bmatrix} B_{p_b}^{\mathrm{T}} \\ 0 \end{bmatrix} r_1^0 = B_{p_b}^{\mathrm{T}} K_{bb}^s B_{p_b}^{\mathrm{T}} r_1^0 \tag{35}$$

where it is worth noting that matrix $B_{p_b}^{\mathrm{T}} K_{bb}^s B_{p_b}^{\mathrm{T}}$ is identical to the lumped preconditioner of FETI-1. Then, using eq. (18) it follows that there is a y such that (here $y = K^{s^+} H L_p f$ from eq. (33))

$$\begin{aligned} z^0 = \tilde{A}^{-1} r^0 &= L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} L_p L^{\mathrm{T}} K^s B_p^{\mathrm{T}} r_1^0 = L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} L_p L^{\mathrm{T}} K^s B_p^{\mathrm{T}} B H^{\mathrm{T}} y \\ &= -L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} B_p K^s B_p^{\mathrm{T}} B H^{\mathrm{T}} y = -L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} B_p K^s B_p^{\mathrm{T}} r_1^0 = -L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} r_2^0 \end{aligned} \tag{36}$$

and: $$z_1^0 = r_2^0 \quad \text{and} \quad z_2^0 (= r_3^0) = B \tilde{A}^{s^{-1}} B^{\mathrm{T}} z_1^0 \tag{37}$$

Then:

$$p^0 = z^0 \Rightarrow \begin{cases} p_1^0 = z_1^0 \\ p_2^0 = z_2^0 \end{cases} \tag{38}$$

and using eq. (14):

$$q^0 = A p^0 = -L^{\mathrm{T}} K^s L L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} p_1^0 = L^{\mathrm{T}} K^s B_p^{\mathrm{T}} B \tilde{A}^{s^{-1}} B^{\mathrm{T}} p_1^0 = L^{\mathrm{T}} K^s B_p^{\mathrm{T}} p_2^0 \tag{39}$$

Therefore:

$$q_1^0 (= z_3^0) = p_2^0 = B \tilde{A}^{s^{-1}} B^{\mathrm{T}} p_1^0 \Rightarrow q_1^0 \in \text{range}(B H^{\mathrm{T}}) \quad \text{and} \quad q_2^0 (= p_3^0) = B_p K^s B_p^{\mathrm{T}} q_1^0 \tag{40}$$

Then (similar to eq. (31) and using the fact that both r_1^0, $q_1^0 \in \text{range}(B H^{\mathrm{T}})$:

$$\eta^0 = \frac{p^{0^{\mathrm{T}}} r^0}{p^{0^{\mathrm{T}}} q^0} = \frac{(p_3^{0^{\mathrm{T}}} - p_1^{0^{\mathrm{T}}}) r_1^0}{(p_3^{0^{\mathrm{T}}} - p_1^{0^{\mathrm{T}}}) q_1^0} \tag{41}$$

Then:

$$r^k = r^{k-1} - \eta^{k-1} q^{k-1} \quad , \quad r_1^k = r_1^{k-1} - \eta^{k-1} q_1^{k-1} \quad \text{and} \quad r_2^k = r_2^{k-1} - \eta^{k-1} q_2^{k-1} \tag{42}$$

which also implies that if r_1^{k-1}, $q_1^{k-1} \in \text{range}(B H^{\mathrm{T}})$ then $r_1^k \in \text{range}(B H^{\mathrm{T}})$.

Continuing, like in eq. (36) we obtain (provided that $r_1^k \in \text{range}(B H^{\mathrm{T}})$):

$$\begin{aligned} z^k = \tilde{A}^{-1} r^k &= L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} L_p L^{\mathrm{T}} K^s B_p^{\mathrm{T}} r_1^k = L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} L_p L^{\mathrm{T}} K^s B_p^{\mathrm{T}} B H^{\mathrm{T}} y \\ &= -L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} B_p K^s B_p^{\mathrm{T}} B H^{\mathrm{T}} y = -L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} B_p K^s B_p^{\mathrm{T}} r_1^k = -L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} r_2^k \end{aligned} \tag{43}$$

and: $$z_1^k = r_2^k \quad \text{and} \quad z_2^k (= r_3^k) = B \tilde{A}^{s^{-1}} B^{\mathrm{T}} z_1^k \tag{44}$$

Then (in the case of full reorthogonalization):

$$p^k = z^k - \sum_{i=0}^{k-1} \frac{z^{k^{\mathrm{T}}} q^i}{p^{i^{\mathrm{T}}} q^i} p^i \tag{45}$$

The dot product terms are written (using eq. (30) and assuming $q_1^k \in \text{range}(B H^{\mathrm{T}})$):

$$\frac{z^{k^{\mathrm{T}}} q^i}{p^{i^{\mathrm{T}}} q^i} = \frac{z_1^{k^{\mathrm{T}}} (q_3^i - q_1^i)}{p_1^{i^{\mathrm{T}}} (q_3^i - q_1^i)} \tag{46}$$

and we have:

$$p_1^k = z_1^k - \sum_{i=0}^{k-1} \frac{z^{k^{\mathrm{T}}} q^i}{p^{i^{\mathrm{T}}} q^i} p_1^i \quad \text{and} \quad p_2^k = z_2^k - \sum_{i=0}^{k-1} \frac{z^{k^{\mathrm{T}}} q^i}{p^{i^{\mathrm{T}}} q^i} p_2^i \tag{47}$$

where the dot products are given by eq. (46). Then, like in eq. (39) (using eq. (22)):

$$q^k = A p^k = L^{\mathrm{T}} K^s B_p^{\mathrm{T}} p_2^k \quad , \quad q_1^k = p_2^k = B \tilde{A}^{s^{-1}} B^{\mathrm{T}} p_1^k \quad \text{and} \quad q_2^k (= p_3^k) = B_p K^s B_p^{\mathrm{T}} p_2^k \tag{48}$$

- Initialize

$$u^0 = L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} L_p f \quad , \quad \tilde{u}^0 = 0 \quad , \quad r_1^0 = B\tilde{A}^{s^{-1}} L_p f \quad , \quad r^0 = \begin{bmatrix} L_b^{\mathrm{T}} K_{bb}^s \\ K_{ib}^s \end{bmatrix} B_{p_b}^{\mathrm{T}} r_1^0$$

$$p_1^0 = z_1^0 = B_{p_b}^{\mathrm{T}} K_{bb}^s B_{p_b}^{\mathrm{T}} r_1^0 \quad , \quad q_1^0 = p_2^0 = r_3^0 = z_2^0 = B\tilde{A}^{s^{-1}} B^{\mathrm{T}} z_1^0$$

$$q^0 = \begin{bmatrix} L_b^{\mathrm{T}} K_{bb}^s \\ K_{ib}^s \end{bmatrix} B_{p_b}^{\mathrm{T}} q_1^0 \quad , \quad p_3^0 = q_2^0 = B_{p_b}^{\mathrm{T}} K_{bb}^s B_{p_b}^{\mathrm{T}} q_1^0 \quad , \quad \eta^0 = \frac{(p_3^{0^{\mathrm{T}}} - p_1^{0^{\mathrm{T}}}) r_1^0}{(p_3^{0^{\mathrm{T}}} - p_1^{0^{\mathrm{T}}}) q_1^0}$$

- Iterate $k = 1, 2, \ldots$ until convergence ($\left\| r^k \right\| < \varepsilon$)

$$\tilde{u}_1^k = \tilde{u}_1^k + \eta^{k-1} p_1^{k-1} \quad , \quad r^k = r^{k-1} - \eta^{k-1} q^{k-1} \quad , \quad r_1^k = r_1^{k-1} - \eta^{k-1} q_1^{k-1}$$

$$z_1^k = r_2^k = r_2^{k-1} - \eta^{k-1} q_2^{k-1} \quad , \quad r_3^k = z_2^k = B\tilde{A}^{s^{-1}} B^{\mathrm{T}} z_1^k \quad , \quad q_3^{k-1} = \left(1/\eta^{k-1}\right)\left(r_3^{k-1} - r_3^k\right)$$

$$p_1^k = z_1^k - \sum_{i=0}^{k-1} \frac{z_1^{k^{\mathrm{T}}} (q_3^i - q_1^i)}{p_1^{i^{\mathrm{T}}} (q_3^i - q_1^i)} p_1^i \quad , \quad q_1^k = p_2^k = z_2^k - \sum_{i=0}^{k-1} \frac{z_1^{k^{\mathrm{T}}} (q_3^i - q_1^i)}{p_1^{i^{\mathrm{T}}} (q_3^i - q_1^i)} p_2^i$$

$$q^k = \begin{bmatrix} L_b^{\mathrm{T}} K_{bb}^s \\ K_{ib}^s \end{bmatrix} B_{p_b}^{\mathrm{T}} p_2^k \quad , \quad p_3^k = q_2^k = B_{p_b}^{\mathrm{T}} K_{bb}^s B_{p_b}^{\mathrm{T}} p_2^k \quad , \quad \eta^k = \frac{(p_3^{k^{\mathrm{T}}} - p_1^{k^{\mathrm{T}}}) r_1^k}{(p_3^{k^{\mathrm{T}}} - p_1^{k^{\mathrm{T}}}) q_1^k}$$

- After convergence

$$u^k = u^0 - L_p^{\mathrm{T}} \tilde{A}^{s^{-1}} B^{\mathrm{T}} \tilde{u}_1^k$$

Algorithm 2: The primal alternative of the FETI-1 method with the lumped preconditioner (full reorthogonalization)

which also shows that $q_1^k \in \text{range}(BH^{\mathrm{T}})$. Note that this concludes recursively the proof that r_1^k, $q_1^k \in \text{range}(BH^{\mathrm{T}})$, $k = 0, 1, \ldots$

Finally, we have:

$$\eta^k = \frac{p^{k^{\mathrm{T}}} r^k}{p^{k^{\mathrm{T}}} q^k} = \frac{(p_3^{k^{\mathrm{T}}} - p_1^{k^{\mathrm{T}}}) r_1^k}{(p_3^{k^{\mathrm{T}}} - p_1^{k^{\mathrm{T}}}) q_1^k} \tag{49}$$

In eq. (46), we use vector q_3^i that has not been computed yet. This vector will be computed using eq. (42) that implies that:

$$r_3^k = r_3^{k-1} - \eta^{k-1} q_3^{k-1} \Rightarrow q_3^{k-1} = \left(1/\eta^{k-1}\right)\left(r_3^{k-1} - r_3^k\right) \tag{50}$$

Hence, using the previous equations, the final form of the algorithm is obtained as is shown in Algorithm 2 (in the case of full reorthogonalization). It is worth noting that even though the formulation is primal, the final algrorithm is very similar to the algorithm of the FETI-1 method with the lumped preconditioner. In particular:

- The matrices $B\tilde{A}^{s^{-1}} B^{\mathrm{T}}$ and $B_{p_b}^{\mathrm{T}} K_{bb}^s B_{p_b}^{\mathrm{T}}$ that are used during the iterations are equal to the FETI-1 matrix operator and lumped preconditioner, respectively.
- The algorithm iterates on vectors of the size of the Lagrange multipliers.
- The residual vanishes in internal d.o.f. of the subdomains, when these d.o.f. are not adjucent to the interface, again as in FETI-1 with the lumped preconditioner.

On the other hand, each iteration of the present algorithm requires more linear combinations of vectors than a dual algorithm. These operations become important in the case of reorthogonalization. In this case, the required dot products $z_1^{k^{\mathrm{T}}} (q_3^i - q_1^i)$, $i = 0, \ldots, k-1$ imply the same computational cost as in FETI-1, because at each iteration $q_3^k - q_1^k$ is computed and stored. However, compared to FETI-1, this algorithm requires twice as many lin-

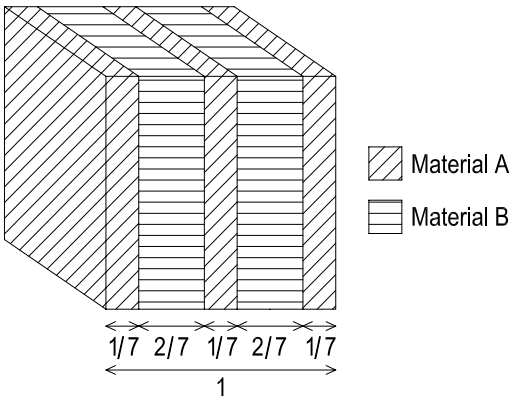


Figure 1. A cubic structure composed of two materials

ear combinations for computing the vectors p_1^k and p_2^k, that represent the direction vectors p^k. In total, in this algorithm reorthogonalization requires 50% more floating point operations than in FETI-1. In addition, while FETI-1 reorthogonalization requires storing two vectors per iteration, here it is required to store the three vectors p_1^k, p_2^k and $q_3^k - q_1^k$, which implies 50% higher memory requirements for reorthogonalization in Algorithm 2.

5 NUMERICAL RESULTS

We have implemented the FETI-1 and FETI-DP methods with the lumped preconditioner and their primal alternatives in our Matlab code and we consider the 3-D elasticity problem of Fig. 1. This cubic structure is composed of five layers of two different materials and is discretized with $28\times28\times28$ 8-node brick elements. Additionally, it is pinned at the four corners of its left surface. Various ratios $E_{\mathrm{A}}/E_{\mathrm{B}}$ of the Young modulus and ρ_A/ρ_B of the density of the two materials are considered in the paper, while their Poisson ratio is set equal to $\nu_{\mathrm{A}} = \nu_{\mathrm{B}} = 0.30$. Two decompositions P1 and P2 of this heterogeneous model of 73,155 d.o.f. in 100 subdomains, are considered (see [4] for details).

Table 1 presents the iterations required by primal and dual formulations of the lumped preconditioned FETI-1 method. The results show that like in the case of comparing dual and primal formulations of the Dirichlet preconditioned FETI methods, the iterations of the two formulations of the lumped preconditioned FETI-1 methods are comparable. More precisely, it is noted that in the more ill-conditioned cases the primal method performs slightly less iterations (up to 11%) than the dual one. In fact, judging also from many other tests that we have performed comparing the two formulations of FETI-1 and FETI-DP with the lumped preconditioner, it appears that the difference between the number of iterations of primal and dual formulations in ill-conditioned problems is more pronounced in the case of the lumped preconditioner than in the case of the Dirichlet preconditioner. A probable explanation is that the lumped preconditioned methods lead by themselves to more ill-conditioned systems than the Dirichlet ones.

On the other hand, bearing in mind that the primal formulation implies a 50% higher reorthogonalization cost, we conclude that statistically the primal formulation will be probably slower than the dual one in well-conditioned problems and probably faster in ill-conditioned problems with relatively low reorthogonalization cost. In addition, in the case of the lumped preconditioner, our results do not show the increased robustness (measured in terms of the maximum achievable solution accuracy in ill-conditioned problems) of the primal formulation that has been seen in the case of the P-FETI formulations. A probable explanation of this observation is given by the increased operations required in each iteration of the primal algorithm as oposed to the dual one and also by the fact that due to setting the initial solution vector equal to eq. (19), the initial residual of the primal methods is equal to the initial residual of the dual methods (see the residual of eq. (32), which is equal to the initial residual of the FETI-1 method). Thus, contrary to the P-FETI formulations, the residuals of the primal formulations of the lumped preconditioned FETI methods begin from relatively high values, as in the dual formulations.

6 CONCLUSIONS

The roots of the presented in this paper work can be traced back to the paper [4]. This paper introduced the P-FETI methods, as the primal alternatives of the Dirichlet preconditioned FETI methods. Compared to the original FETI formulations, the P-FETI methods present the advantage of being more robust and faster in the solution of ill-conditioned problems. [4] also introduced an open question of the existence or not of a primal alternative for the lumped preconditioned FETI methods. In the last years it has become clear that the the lumped preconditioner leads to faster solutions, in the cases where a problem needs to be decomposed in a relatively small number of subdomains. This case and also the case where the lumped precondtioner leads to less memory consumption (in large problems where memory consumption can be the main issue), appear to be the uses of the lumped preconditioner in modern DDM practice.

Ratio of Young moduli	Type of decomposition	Dual formulation	Primal formulation
10^0	P1	25	24
10^3	P1	44	41
10^3	P2	25	24
10^6	P1	30	26
10^6	P2	53	47

Table 1. Number of iterations (Tolerance: 10^{-3}) of the lumped Preconditioned FETI-1 method and its primal alternative for the solution of the example of Fig. 1

The present work introduces the primal alternatives of the lumped preconditioned FETI methods. These new formulations do not appear to present the advantages of the P-FETI formulations, since they are slightly slower or faster than their dual counterparts depending on the problem and do not exhibit higher robustness properties than the dual methods. Their principal value lies in the fact that they add a new level of completion to the theory of the relations of primal and dual methods. The fact that a primal algorithm can be turned to an algorithm which uses dual operators and vectors appears to be new in DDM literature. It is also worth noting that the same transformations used in this paper can be used in the P-FETI and the BDD methods in order to transform them into algorithms that operate on dual quantities. This and many other recent studies[3,7] show more and more that primal and dual formulations are closely connected.

REFERENCES

[1] Farhat, C. and Roux, F.-X. (1991), "A method of finite element tearing and interconnecting and its parallel solution algorithm", Int. J. Numer. Meth. Engng. 32, pp. 1205-1227.

[2] Mandel, J. (1993), "Balancing domain decomposition", Commun. Appl. Numer. Meth. 9, pp. 233-241.

[3] Klawonn, A. and Widlund, O.B. (2001), "FETI and Neumann-Neumann iterative substructuring methods: connection and new results", Communications on pure and applied Mathematics 54, pp. 57-90.

[4] Fragakis, Y. and Papadrakakis, M. (2003), "The mosaic of high performance Domain Decomposition Methods for Structural Mechanics: Formulation, interrelation and numerical efficiency of primal and dual methods", Comput. Methods Appl. Mech. Engrg. 192, pp. 3799-3830.

[5] Fragakis, Y. and Papadrakakis, M. (2004), "The mosaic of high performance Domain Decomposition Methods for Structural Mechanics – Part II: Formulation enhancements, multiple right-hand sides and implicit dynamics", Comput. Methods Appl. Mech. Engrg. 193, pp. 4611-4662.

[6] Bhardwaj, M., Day, D., Farhat, C., Lesoinne, M., Pierson, K. and Rixen. D. (2000), "Application of the FETI Method to ASCI Problems: Scalability results on one-thousand processors and discussion of highly heterogeneous problems", Int. J. Numer. Meth. Engng. 47, pp. 513-536.

[7] Mandel, J., Dohrmann, C.R., and Tezaur, R. (2003), "An algebraic theory for primal and dual substructuring methods by constraints", Applied Numerical Mathematics, to appear, *Proceedings of the 6th IMACS International Symposium on Iterative Methods*, Denver, CO, USA, 27 – 30 March.

Stability and Chaos

GLOBAL BIFURCATIONS IN LIQUID CRYSTALS UNDER SHEAR FLOW

Lucia Russo[1], Constantinos I. Siettos[2] and Ioannis G. Kevrekidis[3]

[1] University of Salerno, Department of Food and Chemical Engineering,
via Ponte don Melillo I-84084, Fisciano (SA), Italy

[2]National Technical University of Athens, School of Applied Mathematics & Physics,
9, Heroon Polytechniou str., GR 157 80, Polytechnioupoli Zografou, Athens, Greece

[3]Princeton University, Department of Chemical Engineering, Program of Applied and Computational Mathematics and Department of Mathematics, Princeton, NJ, 08544 USA

Keywords: liquid crystals, global bifurcations, nonlinear dynamics, invariant manifolds.

Abstract. *We analyze the global dynamics of a liquid crystalline sample when subjected to shear flow. The fluid is described by a simplified model derived through a closure from the rigid rod model. The analysis is based on the numerical computation of stable and unstable manifolds of both periodic and stationary solutions, and their interactions in parameter space. Several global bifurcations have been detected and the impact of the closure approximation on the global dynamics is discussed.*

1 INTRODUCTION

Liquid crystal polymer solutions are known to exhibit complex and nonlinear rheological behavior when subjected to shear flows. Theoretical and experimental work has shown that, depending on shear rate, several stationary and oscillatory regimes (periodic and even chaotic) can been observed. When the rodlike molecules are treated as rigid rods, the sample is described by an orientational distribution function which obeys a Fokker-Planck-type equation, also called the Smoluchowski equation[1]. As is now well known, the rigid rod model is capable of describing the behavior of liquid crystalline polymers in the nematic phase under moderate and large shear rates, i.e., when the effects of texture become negligible[2-4] . The predictions are, in fact, obtained under the assumption of homogeneous conditions, and thus the flow has to be strong enough to destroy the inhomogeneous structure typical of textured samples. Within this framework, the molecules are treated as rigid rods with infinite aspect ratios. The model predicts different regimes (stationary, periodic and chaotic) depending on both the shear rate and the intensity of the nematic field[5]. When the system is a strong nematic, that is, when the intensity of the nematic field is larger than a critical value, stationary flow aligning solutions are predicted at high shear rates, whereas at intermediate and low shear rates several different regimes are predicted, both stationary (log rolling and out of plane aligning) and periodic (kayaking, tumbling, wagging, out of plane oscillating). Recently, also chaotic dynamics have been discovered[5]. The rigid rod model was used as guidance in the experimental quest for oscillating solutions[6]. In that case, a nematic solution showed sustained oscillations under relatively large shear rates, that is, close to the flow-aligning regime. The evidence of sustained oscillations under "strong" flows justifies, at least in first approximation, the use of a monodomain description of the nematic phase, and as a consequence, the analysis of a homogeneous "simple" model can produce meaningful results. Several simplified models have been derived from the Smoluchowski equation writing a continuum equation for the second moment of the distribution and adopting a closure approximation for higher order moments.

We use as a model for a nematic polymer a closed equation capable of reproducing, at least qualitatively, the features of the rigid rod model. The constitutive equation considered here has been studied in great detail by Forest and Wang (2002). It is obtained from the rigid rod model for molecules with finite aspect ratio by applying a quadratic closure approximation. It is well known that the quadratic closure erases most of the rigid rod model predictions under constant shear flows when the molecules are considered of infinite aspect ratio, but, when this closure is applied to the model for rods with finite aspect ratio, the predictions one obtains are in good qualitative agreement with those obtained with the rigid rod model proposed in Doi, 1981 and formulated in terms of orientational distribution function for rods of infinite aspect ratio[1, 7-9]. Even complex and sensitive phenomena as the period doubling cascade leading to chaos are qualitatively reproduced. Critical values of shear rate at which transitions occur have been detected with nonlinear analysis of both closure and distributed models and a comparison between the predictions of different models has been performed with the detection of local bifurcations[7] .

However, local analysis only provides partial information; we still need to understand global bifurcations which can occur, for example, when stable and unstable manifolds of stationary (or periodic) solutions interact. In this work we analyze some features of the global dynamics of a liquid crystalline sample when subjected to shear flow. The analysis is based on the computation of stable and unstable manifolds of both periodic and stationary solutions, and their interactions in parameter space, with the techniques developed in Johnson et al. (1997)[10]. It should be stressed that even though the true dynamics can be much more complex if one accounts for the full coupling between the order parameter and the flow field, we believe that the use of a simple model is sufficient to illustrate that global bifurcations might occur.

2 THE MODEL

2.1 The constitutive equation

Rheological predictions of complex fluids are often based on the use of microscopic modeling techniques[1,11,12]. The molecules are usually treated as long rigid rods exhibiting phase transitions between highly oriented and isotropic states with changes of temperature and/or concentration. A simple model describing the dynamics of a single rigid-rod molecule orientational probability distribution function $\psi(\boldsymbol{u})$ under shear is given by the celebrated Smoluchowski equation[1, 3]:

$$\frac{\partial \psi(\boldsymbol{u})}{\partial t} = D\frac{\partial}{\partial \boldsymbol{u}}\cdot\left(\frac{\partial \psi(\boldsymbol{u})}{\partial \boldsymbol{u}} + \psi(\mathbf{u})\frac{\partial}{\partial \boldsymbol{u}}\left(\frac{V[\psi,\mathbf{u}]}{kT}\right)\right) - \frac{\partial}{\partial \boldsymbol{u}}\cdot(\dot{\boldsymbol{u}}\psi) \tag{1}$$

where $\boldsymbol{u}$ is a unit vector describing the orientation of a rod, k is the Boltzmann's constant, T represents the absolute temperature, D is the average rotational diffusivity while $V[\psi,\mathbf{u}]$ is the mean field nematic potential which in general is a functional of the distribution function relating the free energy associated with a molecule with orientation $\boldsymbol{u}$ interacting with its neighbors. The rate of change of the orientation, $\dot{\boldsymbol{u}}$ due to the macroscopic flow reads[7, 13]: $\dot{\boldsymbol{u}} = \boldsymbol{\Omega}\cdot\boldsymbol{u} + a\left[\mathbf{D}\cdot\boldsymbol{u} - \mathbf{D}:\boldsymbol{uuu}\right]$.

The above equation generalizes the applicability of the Smoluchowski equation to molecules with different shapes, from rod to disc-like. This is controlled by the parameter a in the above expression given by $a = \frac{r^2-1}{r^2+1}$; $a = 1$ corresponds to the thin-rod limit, $a = 0$ corresponds to spherical molecules, while $a \approx -1$ correspond to the thin-disc limit. $\mathbf{D}$ and $\boldsymbol{\Omega}$ are the tensors of the rate of strain and vorticity, given by

$$\mathbf{D} = \frac{1}{2}\left(\nabla\mathbf{v} + \nabla\mathbf{v}^T\right), \quad \boldsymbol{\Omega} = \frac{1}{2}\left(\nabla\mathbf{v} - \nabla\mathbf{v}^T\right) \tag{2}$$

where $\mathbf{v}$ is the macroscopic velocity of the flow. Here we use the simple Maier-Saupe potential given by[1]: $V[\boldsymbol{u}] = -\frac{3}{2}UkT\,\boldsymbol{uu}:\mathbf{S}$, where $\mathbf{S} = \langle\mathbf{uu}\rangle - \frac{1}{3}\mathbf{I}$ is the *tensor* order parameter, $\langle(\cdot)\rangle = \int(\cdot)\psi(\mathbf{u})d\mathbf{u}$ and U is the intensity of the nematic potential (which can be thought to be proportional to the concentration of the rods).

The velocity of the macroscopic shear flow is given by $\mathbf{v} = (Gz, 0, 0)$, where the x-coordinate shows the flow direction, the y-coordinate the vorticity direction and the z-coordinate the velocity gradient direction; G is the nondimensional shear rate.

Approximate models for macroscopic quantities such as the tensor order parameters $\mathbf{S}$ are derived using various closure approximations[1,4]. The most common is the one obtained by writing the fourth moment of the distribution in terms of the lower second order moments as $\langle\boldsymbol{uuuu}\rangle \approx \langle\boldsymbol{uu}\rangle\langle\boldsymbol{uu}\rangle$. Multiplying Eq.(1) by $\boldsymbol{uu}$ and then averaging over the orientational distribution ψ we get the following evolution equation for $\mathbf{S}$:

$$\frac{d\mathbf{S}}{dt} = -6\left[\mathbf{S} - U\left(\mathbf{S}+\frac{1}{3}\mathbf{I}\right)\cdot\mathbf{S} + U\left(\mathbf{S}:\mathbf{SS} + \frac{1}{3}\mathbf{S}:\mathbf{SI}\right)\right] - \mathbf{S}\cdot\boldsymbol{\Omega} + \boldsymbol{\Omega}\cdot\mathbf{S} + \frac{2a}{3}\mathbf{D} + a\left(\mathbf{D}\cdot\mathbf{S} + \mathbf{S}\cdot\mathbf{D}\right) - 2a\mathbf{D}:\mathbf{S}\left(\mathbf{S}+\frac{1}{3}\mathbf{I}\right) \tag{3}$$

where $\mathbf{D}$ and $\boldsymbol{\Omega}$ now read: $\mathbf{D} = \frac{1}{2}\mathrm{G}\begin{pmatrix}0&1&0\\1&0&0\\0&0&0\end{pmatrix}$; $\boldsymbol{\Omega} = \frac{1}{2}\mathrm{G}\begin{pmatrix}0&1&0\\-1&0&0\\0&0&0\end{pmatrix}$

The above is a five-dimensional system of ordinary differential equations, since the tensor S has five independent components ($\boldsymbol{S}$ is a symmetric and traceless).

2.2 Symmetry properties of the model

The 5-D system of equations exhibits some interesting equivariance properties. It is well known that dynamical systems with different symmetries can exhibit different bifurcating solutions[14]. To start with, we write Eq. (3) in the following form:

$$\frac{d\mathbf{S}}{dt} = \mathbf{F}(\mathbf{S}) \tag{4}$$

Using $\kappa = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$ as the reflection matrix, it is simple to verify that the **F** commutes with the compact Lie group of symmetry $\Gamma = \{\mathbf{I}, \kappa\} \cong Z_2$, that is: $\kappa\mathbf{F}(\mathbf{S}) = \mathbf{F}(\kappa\mathbf{S})$. This means that if **S**(t) is a solution of Eq.(4), then γ**S**(t) is also a solution for all t and $\gamma \in \Gamma$. Let's characterize the set of points that have a given symmetry by the fixed-point subspace: $Fix(\Gamma) = \{\mathbf{S} \in V : \gamma\mathbf{S} = \mathbf{S}, \forall\ \gamma \in \Gamma\}$ where V is the 5-dimensional state-space of traceless symmetric matrices. This linear subspace is flow-invariant, i.e., if an initial condition is taken on it, the orbit will remain on it for all times.

An attractor, or in general an invariant set A of the system (4) is *point-wise symmetric* if $\gamma\mathbf{S}(t) = \mathbf{S}(t)$, $\forall\ t \in R$, $\mathbf{S}(t) \in A$, $\gamma \in \Gamma$ (i.e., A belongs to $Fix(\Gamma)$); A is *set-wise symmetric* with respect to Γ if $\gamma A = A$, but $\gamma\mathbf{S}(t) \neq \mathbf{S}(t)$. When A doesn't possess any symmetry, then different invariant sets A', A'' can be found satisfying $A' = \gamma A''$.

To this end we decompose **S** as follows:

$$\mathbf{S} = \begin{pmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{xy} & S_{yy} & S_{yz} \\ S_{xz} & S_{yz} & * \end{pmatrix} = \begin{pmatrix} S_{xx} & S_{xy} & 0 \\ S_{xy} & S_{yy} & 0 \\ 0 & 0 & * \end{pmatrix} + \begin{pmatrix} 0 & 0 & S_{xz} \\ 0 & 0 & S_{yz} \\ S_{xz} & S_{yz} & 0 \end{pmatrix} = \mathbf{S}_1 + \mathbf{S}_2$$

defined in the subspace $V = V_1 \oplus V_2$, where V_1 and V_2 are mutually orthogonal subspaces of dimension 3 and 2 respectively. The projection of Eq. (4) on these two subspaces leads to the following system of equations:

$$\begin{aligned} \frac{d\mathbf{S}_1}{dt} &= \mathbf{F}_1(\mathbf{S}_1) + \mathbf{G}_1(\mathbf{S}_1, \mathbf{S}_2) \\ \frac{d\mathbf{S}_2}{dt} &= \mathbf{F}_2(\mathbf{S}_2) + \mathbf{G}_2(\mathbf{S}_1, \mathbf{S}_2) \end{aligned} \tag{5}$$

The system (4) with $\mathbf{S}_2$=**0** corresponds to the restriction of the system (3) on the invariant fixed-point subspace $V_1 = Fix(\Gamma)$.

3 RESULTS AND DISCUSSION

3.1 The Bifurcation Diagram

We start with a brief discussion of the different invariant sets of the system (4) and the local bifurcations that these sets encounter when the shear rate is varied. A 3-D construction of the bifurcation diagram reported also by Forest and Wang (2003) is shown in fig. 1. This was obtained for an aspect-ratio a=0.8 and U=6. The authors found that the 5-D macroscopic model with the quadratic Doi closure[1], for a finite aspect-ratio approximates the dynamics of the Smoluchowski equation for infinite aspect-ratio[5]. Here we chose to study the global dynamics of this particular closure model for the same parameters values aiming to give an insight of the global dynamics of the Smoluchowski equation. All the qualitative Smoluchowski equation solutions are predicted: flow-aligning, wagging, tumbling, kayaking1, kayaking2 and the period-doubling route to chaos.

Starting from high shear rate (Pe=G/6>3.979) only one stationary solution exists: the flow-aligning. This is an "in-plane" solution as the major director, i.e., the eigenvector corresponding to the largest eigenvalue of the **S** tensor, lies in the (x, y)-plane of shear and so it is a symmetric solution belonging to the fixed-point invariant subspace. All five eigenvalues are negative so the flow-aligning solution branch is a stable one in the full space (see fig. 1). The stability of these stationary states as the Pe number is decreased from 3.979 to 0 is given in Table 1.

Flow-aligning			Logrolling-1			Logrolling-2		
3.719<Pe<3.979								
	IN	OUT		IN	OUT		IN	OUT
ST	3	2	ST	2	0	ST	3	0
UNST	0	0	UNST	1	2	UNST	0	2
2.8199<Pe<3.719								
	IN	OUT		IN	OUT		IN	OUT
ST	3	1	ST	2	1	ST	3	0
UNST	0	1	UNST	1	1	UNST	0	2
0<Pe<2.8199								
	IN	OUT		IN	OUT		IN	OUT
ST	1	1	ST	2	1	ST	3	0
UNST	2	1	UNST	1	1	UNST	0	2

Table 1: Number of the eigenvectors IN and OUT (orthogonal complement) of the invariant subspace of stationary states (ST is for eigenvectors corresponding to eigenvalues with a negative real part, while UNST is for eigenvectors corresponding to eigenvalues with positive real part).

At *Pe*=3.979 two new branches of stationary states appear through a turning point. These stationary states are called "logrolling" because the major director is aligned with the vorticity axis. They are both unstable and symmetric with respect to the reflection matrix *k* and they belong to the invariant fixed-point subspace. We will refer to "logrolling-1" for the upper branch and "logrolling-2" for the lower one (see fig. 1).
The logrolling-2 branch has 5 real eigenvalues with 3 stable eigenvectors on the invariant subspace and 2 eigenvectors unstable and orthogonal to the invariant subspace. As Pe is decreased, two positive eigenvalues collide on the real axis and become complex. Logrolling-1 is unstable both in the invariant fixed-point subspace and the full space. In the invariant subspace two eigenvectors are stable and one is unstable while both the "out of plane" eigenvectors are unstable. At *Pe*=3.719 the flow-aligning solution branch loses stability through a pitchfork bifurcation. An infinite number of asymmetric stationary states, belonging to an "out of plane" circle, arise with this bifurcation. The set is represented in the right-up inset of fig.1 by a vertical branch, which ends up on the logrolling-1 branch. At this point one positive eigenvalue of logrolling-1 becomes negative and one unstable direction out of the subspace becomes stable. The stability of the stationary states on this "out-of plane" branch changes along the circle and is not related to the symmetry properties of the model.
Instead, for a different closure model, this degenerate situation is related with some additional properties of the model given by (3)[15]. We note that this situation is not present in the full PDE model[5] and it is not structurally stable with respect to small perturbations of the parameter *Pe*. We will comment on this in more detail in future work. On the circle of stationary solutions there is a Hopf bifurcation that gives birth to asymmetric periodic solutions called "kayaking-2".
Due to the model symmetry these periodic solutions come in pairs and they are symmetrically related by the reflection matrix κ. If $\mathbf{S}(t)$ is the periodic orbit of a kayaking-2 regime, the other asymmetric regime is given by $\kappa\mathbf{S}(t)$. These regimes are characterized by a major director that rotates around a fixed axis lying between the shear-plane and the vorticity axis. When the *Pe* number is decreased down from *Pe*= 3.719, these asymmetric cycles undergo a cascade of period-doubling bifurcations (the first period doubling appears at *Pe*= 3.29) leading to a couple of asymmetric chaotic attractors. An orbit projection of these attractors is reported in fig. 2a for *Pe*=3.24.

3.2 Global Bifurcations

Further decrement of the parameter leads to a global bifurcation where suddenly a set-wise symmetric chaotic attractor appears (fig. 2b). This attractor was also observed in [7]. As discussed in the previous section, the set-wise symmetry means that if *A* is the chaotic attractor, the invariance property refers to the set as a whole and guarantees that this set will never intersect with the fixed point invariant subspace. From a more physical point of view, this means that the major director rotates randomly around the vorticity axis, does not intersect the shear-plane and is never aligned with the vorticity axis. The bifurcation of a chaotic attractor, where a sudden discontinuous change is observed, is called crisis[16]. Three different type of crisis are described in [16]: in the first type, a chaotic attractor is suddenly destroyed as the parameter goes though its critical value; in the second one, the size of the chaotic attractor in phase space suddenly changes, while in the third type two or more chaotic attractors merge to one chaotic attractor slightly bigger in size compared to the union of the two asymmetric chaotic regimes. A discrete jump of the attractor size is apparent in fig. 2a, b. The time series before the crisis are shown in fig. 3a, while the ones after the crisis, for different values of the bifurcation parameter, are given in fig. 3b-f.

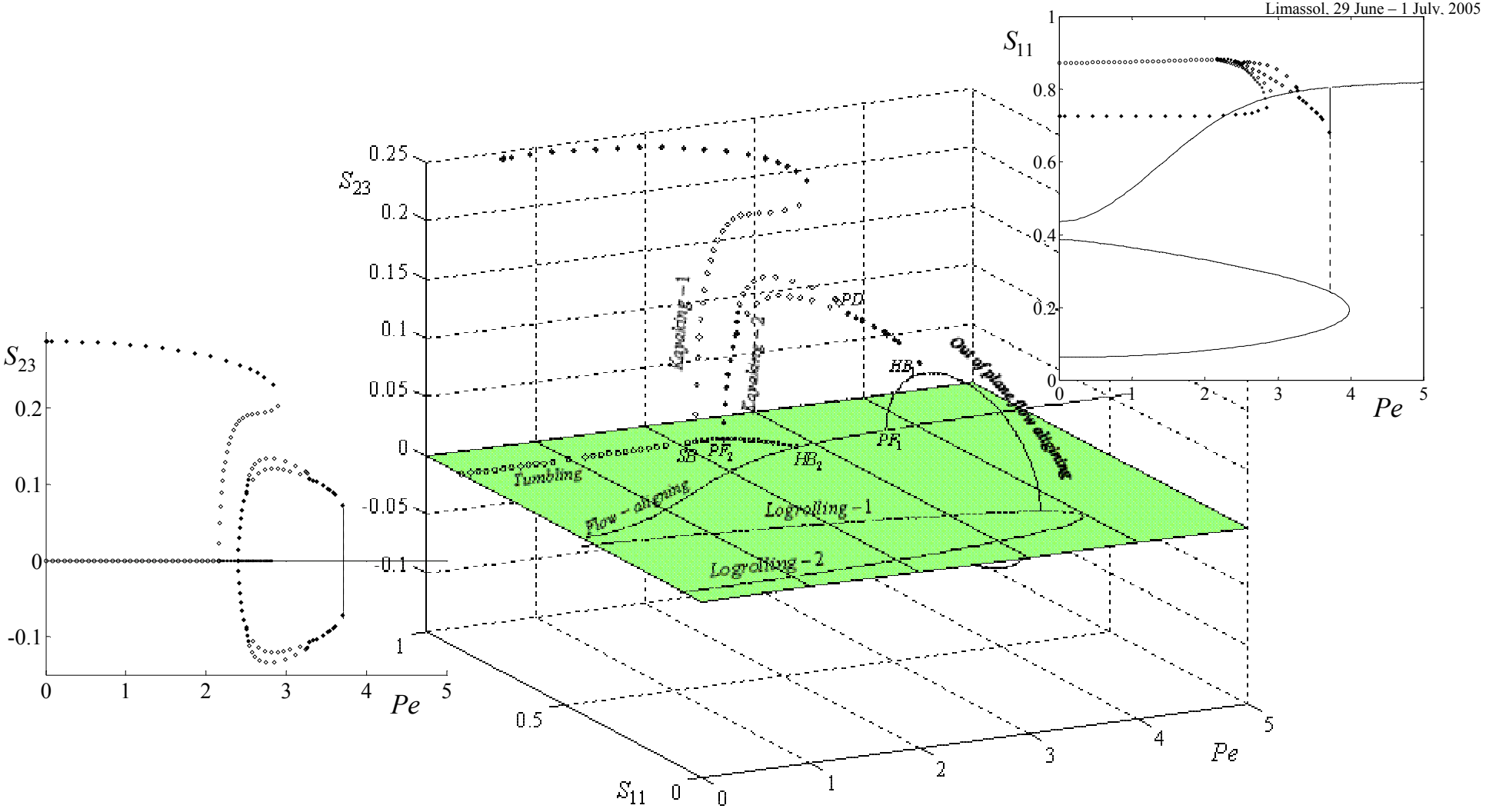
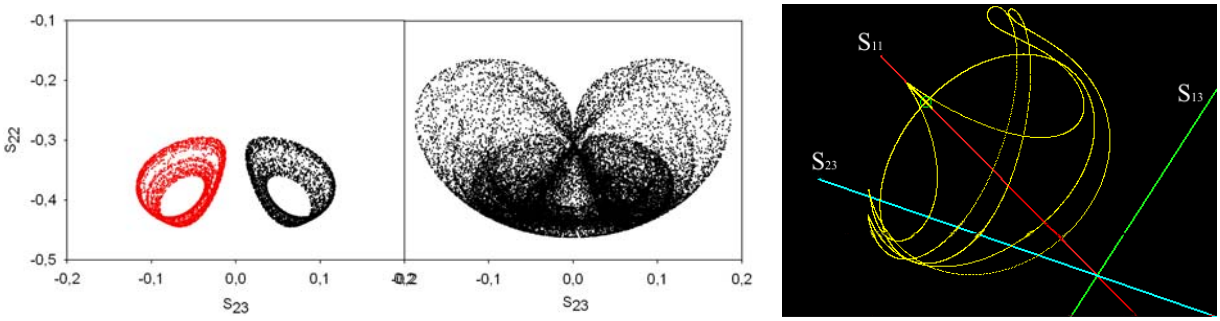


Figure 1. The 3-D Bifurcation Diagram (*Pe*=G/6). Solid lines correspond to stable stationary solutions, dotted lines to unstable ones. Limit cycles are represented as circles, squares and rhombs; empty ones correspond to unstable solutions while filled ones to stable solutions. A 2-D projection of this diagram is reported in [7].



(a) (b) (c)

Figure 2. (a) A couple of asymmetric chaotic attractors at *Pe*=3.26, (b) at *Pe*=3.23 a global bifurcation appears leading to a set-wise symmetric chaotic attractor, (c) A homoclinic loop appearing at $Pe = Pe^* \approx 3.2386715$.

It is apparent that time series after the crisis are characterized not only by the transition from "one" asymmetric chaos to " the other" but also by an intermittent bursting that leads the orbit away from the phase space neighborhood of the pre-existing chaotic attractors. Close (but for *Pe*>*Pe**) to the bifurcation point (fig. 3b) the orbit spends a long time in the neighborhood of one of the two pre-existing asymmetric attractors and switches after the intermittent bursting to the "other" chaotic attractor region. As the value of the bifurcation parameter is further decreased the intermittent bursting is more and more frequent until the ghost of the previous asymmetric chaotic attractors cannot be visually distinguished anymore (fig. 3f).

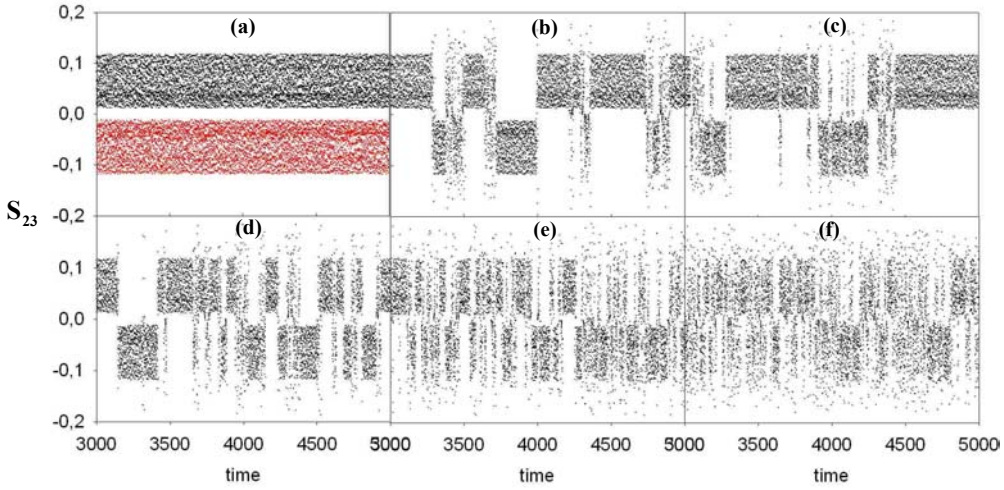
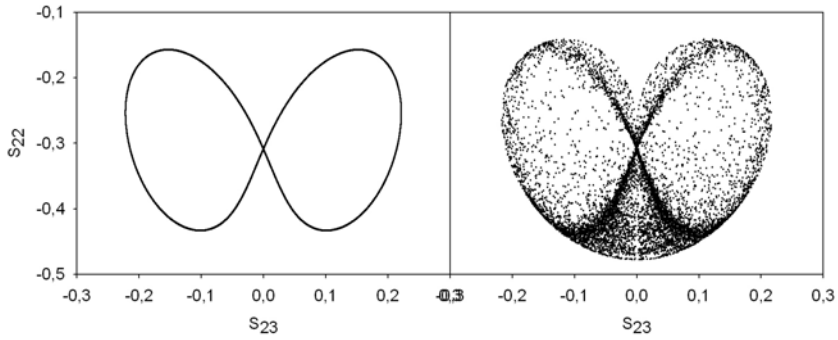


Figure 3. Time series close to the crisis appearing at $Pe^* \approx 3.2386715$. (a) Before the bifurcation point (at *Pe*=3.24) and (b), (c), (d), (e), (f) for smaller values of the *Pe* number.

Another global bifurcation is encountered following the symmetric chaotic attractor for decreasing values of *Pe* numbers. Until *Pe*=2.8199, the flow-aligning solution branch does not encounter any stability change (see Table-1). However at Pe_c=2.9233 two new limit cycles are born in a turning point (see bifurcation diagram in fig. 1). This branch of limit cycles is called kayaking1 ant its stability changes through the turning point. These solutions do not belong to the invariant fixed-point subspace but they are spatio-temporally symmetric. Indeed, they respect the following invariance property: $\kappa \mathbf{S}(t) = \mathbf{S}\left(t + T/2\right)$. These kind of limit cycles are also called discrete travelling waves or "POMs" (ponies on a merry-go-round). The major director rotates symmetrically around the vorticity axis and it is an "out of plane" solution. An orbit projection of the stable kayaking-1 orbit is shown in fig.4a. The transition from kayanking1 to the chaotic attractor is marked by the turning point on the kayaking-1 branch. The time series showing the transition are illustrated in fig. 5. This resembles an intermittency mechanism[17]. For $Pe > Pe_c$ the system response consists of long in time almost regular oscillations (laminar phases) which are intermittently interrupted by shorter in time chaotic outbreaks (turbulent burst) , see fig. 4b. During the bursts, the trajectory moves far away from the vicinity of the pre-existing periodic orbit (fig. 5). It is also clear that as the bifurcation parameter increases, so does the frequency of the bursts. Three types of generic bifurcations, namely, cyclic-fold, subcritical Hopf bifurcation and subcritical period-doubling bifurcations underlie respectively type I, type II, and type III intermittency[17]. However, the loss of stability of a periodic orbit via one of these three local bifurcations is not sufficient for intermittency to occur.



(a) (b)

Figure 4. (a) A kayaking-1 limit cycle at Pe=2.92, (b) at the turning point of the kayaking1 branch a global bifurcation appears giving rise to a set-wise symmetric chaotic attractor (simulation at Pe=2.93).

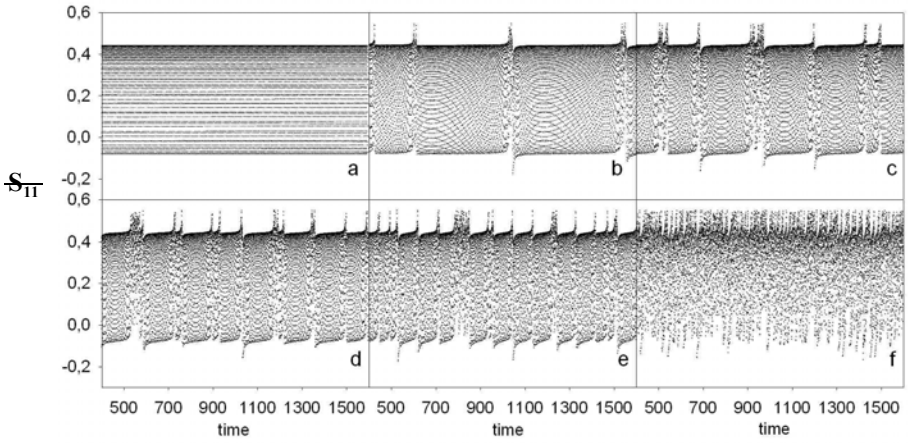
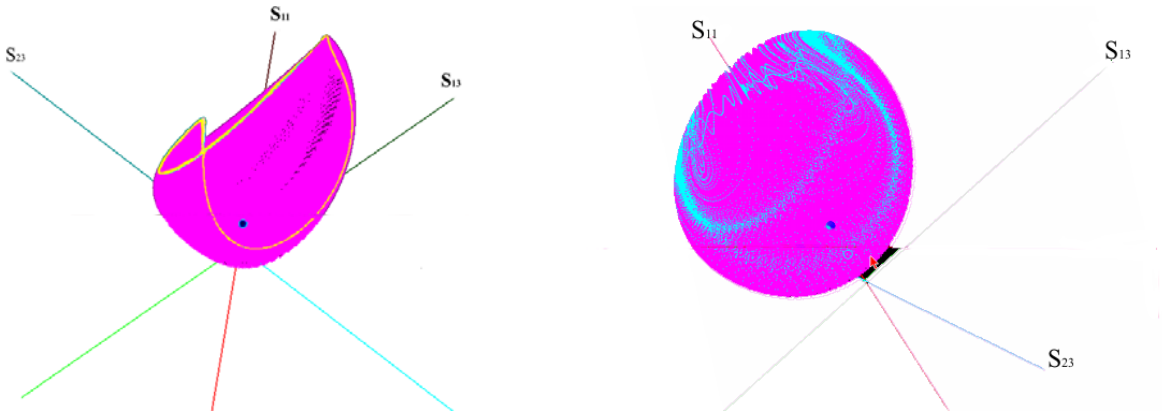


Figure 5. Time series showing the transition from kayanking1 to the symmetric chaotic attractor. a) time series before the transition at Pe=2.92<Pe_c, (b)- (f) time series for increasing values of the Pe number past the critical value Pe_c.

Attempting to explore the global phase space structure we computed the invariant manifolds of the stationary states close to this bifurcation (fig 6). The 2-dimensional unstable manifold of the logrolling2 stationary state is asymptotically attracted by the limit cycle for Pe<Pe_c (fig 6a) and by the symmetric chaotic attractor for Pe>Pe_c) (fig. 6b), while at $Pe \approx Pe_c$ it becomes homoclinic.



(a) (b)

Figure 6. Invaraint manifolds of the stationary states invlolved in the intermittency global bifurcation. The 2-dimendional unstable manifold of the logrolling2 stationary state is asymptotically attracted by (a) the limit cycle for Pe<Pe_c (b) the symmetric chaotic attractor for Pe>Pe_c, while for $Pe \approx Pe_1$ it forms a 2-dimensional homoclinic loop. The blue dot corresponds to logrolling stationary state.

REFERENCES

[1] Doi M. (1981), "Molecular Dynamics and Rheological Properties of Concentrated Solutions of Rodlike Polymers in Isotropic and Liquid Crystalline Phases", J. Polym. Sci. Polym. Phys., 19, pp. 229–243.

[2] Marrucci G. and Maffettone P.L. (1989), "Description of the liquid crystalline phase of rodlike polymers at high shear rates", Macromolecules 22, pp.4076–4082.

[3] Larson R.G. (1990), "Arrested Tumbling in Shearing Flows of Liquid Crystal Polymers", Macromolecules, 23, pp. 3983–3992.

[4] Rey A.D., Denn M.M., (2002), "Dynamical phenomena in liquid-crystalline materials", Annu. Rev Fluid Mech, 34, pp. 233-266.

[5] Grosso M, Keunings R, Crescitelli S, and Maffettone P.L.(2001),"Prediction of chaotic dynamics in sheared liquid crystalline polymers",Phys. Rev. Lett., 86,pp. 3184-3187.

[6] Grosso M, Crescitelli S, Somma E, Vermant J, Moldenaers P, Maffettone PL (2003), "Prediction and observation of sustained oscillations in a sheared liquid crystalline polymer", Phys. Rev. Lett., 90, art. no. 098304.

[7] Forest MG, Wang Q. (2003), "Monodomain response of finite-aspect-ratio macromolecules in shear and related linear flows", Rheol. Acta, 42, pp.20-46.

[8] Kupferman, R., Kawaguchi M.N., Denn M.M. (2000), "Emergence of structure in a model of liquid crystalline polymers with elastic coupling", J. Non-Newtonian Fluid Mech., 91, pp. 255-271.

[9] Faraoni V., Grosso M., Crescitelli S., and Maffettone P.L. (1999), "The rigid rod model for nematic polymers: an analysis of the shear flow problem", J. Rheol., 43, pp. 829-843.

[10] Johnson M.E., Jolly M.S. and I.G. Kevrekidis (1997), "Two-dimensional invariant manifolds and global bifurcations: Some approximation and visualization studies", Numer. Algorithms, 14, pp.125-140.

[11] Hess S. Z. (1976), "Fokker–Planck–Equation Approach to Flow Alignment in Liquid Crystals", Naturforsch., 31A, pp.1034–1037.

[12] Siettos CI, Graham MD, Kevrekidis IG. (2003), "Coarse Brownian dynamics for nematic liquid crystals: Bifurcation, projective integration, and control via stochastic simulation", J. Chem. Phys., 118, pp.10149-10156.

[13] Wang Q. (2002), "A hydrodynamic theory for solutions of nonhomogeneous nematic liquid crystalline polymers of different configurations"**,** J. Chem. Phys. ,116, pp.9120-9136.

[14] Golubitsky, M., I., Stewart & Schaeffer, D. G. (1988). Singularities and groups in bifurcation theory, Vol. 2. New York: Springer.

[15] Chillingworth DRJ, Alonso EV, Wheeler AA (2001), "Geometry and dynamics of a nematic liquid crystal in a uniform shear flow", J. Phys. A-Math. Gen., 34, pp. 1393-1404.

[16] Grebogi C., Ott E., Romeiras F. and Yorke J.A. (1987), "Critical exponents for crisis-induced intemittency", Phys. Rev. A 36, pp. 5365-5380.

[17] Pomeau, Y., & Manneville, P. (1980), "Intermittent transition to turbulence in dissipative dynamical systems," Commun. Math.Phys., 74, pp.189.

DETERMINATION OF CHAOS IN THE DIRECT SIMULATION OF TWO DIMENSIONAL TURBULENT FLOW OVER A SURFACE MOUNTED OBSTACLE

V. P. Fragos[1], S. P. Psychoudaki[1], and N. A. Malamataris[2]

[1] Department of Hydraulics, Solid Sciences and Agricultural Engineering
Aristotle University of Thessaloniki
GR-54124 Thessaloniki
e-mail: fragos@agro.auth.gr , web page: http://www.auth.gr/agro/eb/hydraul/intex.htm

[2] Department of Mechanical Engineering
TEI W. Macedonia
GR-50100 Kila, Kozani
e-mail: nikolaos@eng.auth.gr

Keywords: Quasi two dimensional turbulence, energy and enstrophy spectra, calculation of auto-correlation and cross-correlation functions, Galerkin Finite Elements, Chaos.

Abstract. *The quasi two dimensional turbulent flow over a surface mounted obstacle is studied as a numerical experiment by directly solving the transient Navier Stokes equations with Galerkin finite elements. The Reynolds number defined with respect to the obstacle height is 1304. Energy and enstrophy spectra yield the dual cascade of two dimensional turbulence and the -1 power law decay of enstrophy. Other statistical characteristics of turbulence such as Eulerian autocorrelation coefficients, longitudinal and lateral coefficients are also computed. Finally, oscillation diagrams of computed velocity fluctuations yield the chaotic behavior of turbulence*

1 INTRODUCTION

The turbulent flow over a surface mounted obstacle is a fundamental problem in fluid mechanics having a wide range of applications in all domains of engineering science, as recently reviewed by Fragos et al[3],[4] Although the flow has received a lot of attention in the engineering community, it is still an open ended problem partly due to its complicated geometry and partly due to the unresolved issues of the nature of turbulence. The term flow over a surface mounted obstacle is used ambiguously in the literature. In this work, it is examined the flow over a cubic or a prismatically shaped obstacle having a width that extends up to the walls of a wind tunnel, where the obstacle is placed. This case is a quasi two dimensional flow, that takes place in the two dimensional space, where any three dimensional effects are generated from the existence of walls or from turbulence. For this particular flow, there is some recent experimental work in the turbulent regime conducted by Acharya et al[1] and Larichkin and Yakovenko[9] for obstacles with rectangular cross section of aspect ratio 1:1. There are also attempts to study this flow computationally by Acharya et al[1], Hwang et al[7] who used k-ε models with a finite difference method. In this work, the quasi two dimensional flow over a surface mounted obstacle is studied computationally solving the unsteady Navier Stokes equations in primitive variable formulation with standard Galerkin finite elements. This approach is used for the first time for this flow. The experimental set up and the process parameters of the work of Acharya et al[1] are taken for comparison with the numerical results of this work.

In this work, the issues of both quasi two dimensional turbulence and direct numerical simulation of turbulent flows are addressed in the study of two dimensional turbulent flow over a surface mounted obstacle with square cross section. In the following, the governing equations are presented along with the computational domain and the parameters of the flow. The issues of initial condition and inflow as well as outflow boundary condition are examined next, followed by the finite element formulation and the computational details. The results of this work are subsequently discussed and finally conclusions are drawn.

2 GOVERNING EQUATIONS AND PROCESS PARAMETERS

The computational domain for the turbulent flow over a surface mounted obstacle is shown in Figure 1. A Newtonian fluid of constant viscosity and density approaches with uniform u-velocity a wind tunnel of rectangular cross-section. At the entrance of the tunnel, the fluid is decelerated along the wall due to the no-slip boundary condition. The fluid continues its motion through the tunnel and hits the surface mounted obstacle,

which has a square cross section. The flow separates forming recirculation zones both upstream and downstream the obstacle. The fluid leaves the computational domain after a certain distance downstream. It is expected, that the flow phenomena are less intense, as the fluid is distancing from the obstacle.

For a two dimensional isothermal turbulent flow of an incompressible Newtonian fluid, the dimensionless Navier Stokes equations are:

$$\nabla u = 0 \tag{1}$$

$$\frac{\partial u}{\partial t} + u\nabla u = -\nabla p + \frac{1}{\mathrm{Re}}\nabla^2 u \tag{2}$$

Equations (1) and (2) represent conservation of mass and momentum respectively.

The governing equations have been rendered dimensionless, by choosing the uniform approaching velocity of the fluid U_0 as the characteristic velocity and the height of the obstacle *h* as the characteristic length. The characteristic or reference time t_r is then the ratio h/U_0. In the governing equations, **u** = (u, v) is the velocity vector of the fluid with u and v its components in the x- and y- direction respectively, t is the time, p is the pressure and $Re= U_0h/\nu$ is the Reynolds number with respect to the height of the obstacle, with ν the kinematic viscosity of the fluid. The pressure p has been non-dimensionalized with term ρU_0^2, with ρ the density of the fluid. The time t has been non-dimensionalized with the characteristic time t_r. The dimensions of the computational domain and the Reynolds number were chosen to match the parameters of a laboratory experiment conducted by Acharya et al [1]. According to their conditions, the height of the obstacle is h = 6.35 mm and the approaching velocity of the fluid U_0 = 3.225 m/s. Air was chosen as a working fluid with a kinematic viscosity $\nu = 1.57\ 10^{-5}$ m^2/s and a density $\rho = 1.2$ kg/m^3. The Reynolds number is then 1304 with these values of process parameters. All other dimensions of the computational domain are depicted in Figure 1 as functions of the height of the obstacle h.

In order to solve the problem, appropriate boundary and initial conditions must be chosen, which are discussed in the next section.

3 BOUNDARY AND INITIAL CONDITION

The boundary conditions for this flow are depicted in Figure 1 and given in equations below:

At the entrance : $u=U_0$, v=0
Top and bottom boundaries, upstream the wind tunnel : $u=U_0$, v=0
Top and bottom walls of wind tunnel : u=0, v=0
Along walls of obstacle : u=0, v=0
At the outflow : free boundary condition.

At the entrance a uniform undisturbed velocity profile is imposed, that defines the flow rate entering the wind tunnel. Upstream the wind tunnel, tow tank boundary conditions are imposed. Along the walls of obstacle, no slip boundary conditions are imposed as well as along the solid walls f the computational domain. The concept of the free boundary condition is used for this flow problem at the outflow.

For a thorough discussion of this concept the reader is recommended to check with Fragos et al[3,4]. In this work, the steady state solution of laminar flow at Re = 1 was chosen as the initial condition. This numerical solution has been validated with laboratory experimental data, as discussed by Fragos et al[3]. The streamlines for this flow are shown in Figure 2. By executing the computer program, the flow develops from its initial laminar state to fully developed turbulence. Hence, this approach has the advantage that the transition from laminar to turbulent flow is also computed in addition to the study of turbulence. For a detailed discussion of initial and boundary conditions in turbulent flows, the reader should check the review article of Friedrich et al[5].

4 FINITE ELEMENT FORMULATION AND COMPUTATIONAL DETAILS

The computational mesh used in this work consists of rectangular finite elements of different size with nine nodes each of them. Standard Galerkin finite elements are used to solve the governing equations along with the appropriate initial and boundary conditions. Velocities and pressure are approximated with quadratic and linear basis functions. The weighted residuals, after the application of the divergence theorem, become:

$$R_C^i = \int_V \nabla u \psi^i dV \tag{3}$$

$$R_M^i = \int_V \left[\left(\frac{\partial u}{\partial t} + u\nabla u \right) \varphi^i - \left(-pI + \frac{1}{\mathrm{Re}} \nabla u \right) \nabla \varphi^i \right] dV - \int_s n \left[-pI + \frac{1}{\mathrm{Re}} \nabla u \right] \varphi^i dS \tag{4}$$

Equations (3) and (4) represent an algebraic system of nonlinear equations, which is solved with a Newton-Raphson iterative scheme. Gauss elimination is used for the inversion of the Jacobian matrix. All results that are presented in the next section are independent of time step and mesh resolution.

5 RESULTS AND DISCUSSION

5.1 Velocity fluctuations

Oscillograms of u- and v- velocity fluctuations are shown in Figure 3 at position (x=50,y=5) of the flow domain. The result is a randomness in the oscillograms of the velocity fluctuations superimposed to the periodicity due to the steady generation of vortices. This randomness is exhibited in the acanonical variation of the minima and the maxima of the amplitude of the oscillations. The period of the oscillation is undisturbed though, as it should be, due to the nature of the flow at this Reynolds number. Additionally, randomness is also observed within each period of oscillation, where the pattern of the fluctuations is in each period different and acanonical as well. This difference in the pattern may not be obvious at a first glance, due to the dominant periodical nature of the signal. However, the randomness in the pattern exists and gives the message, that turbulence develops as the flow is distancing from the obstacle. Figure 3 is reminiscent of the signals taken by many experimentalists over the last 80 years and shown in standard books of turbulence (Frish[6]), in order to invoke the intuition of the reader about the complicated nature of turbulent flow and justify the attitude of many workers in the field towards a statistical analysis of turbulence.

In this work, it is shown, that deterministic spatiotemporal chaos is predicted by directly solving the Navier Stokes equations. To the best of our knowledge, this kind of oscillograms are shown for the first time from a computational analysis. So far, even in books dealing with chaos, chaotic behavior has been limited to the subject of solving equations, where the dependent variables are only a function of time. It should be also noted, that this behavior of velocity fluctuations is representative of any other point in the flow domain.

5.2 Energy and enstrophy spectra

It is customary in turbulent research to study the energy and enstrophy spectra of the instantaneous values of velocities and vorticity. Especially the energy spectrum has been the subject of extensive experimental investigation and led to an empirical law, that has been verified by the deeply influential work of Kolmogoroff (Frish[6] Batchelor[2]and Kraichnan[8] extended these ideas to quasi two-dimensional turbulence. They found that the energy spectrum follows both the -5/3 power law of three dimensional turbulence (called inverse energy cascade in the terminology of quasi two dimensional turbulence) and a -3 power law, which has been attributed to a forward enstrophy cascade. Additionally, they studied the enstrophy cascade and they predicted a -1 power law for the dependency of enstrophy with respect to the frequency of the spectrum.

The energy spectra are shown in Figure 4. They have been calculated at the same points as the oscillograms of the velocity fluctuations in Figure 3. A specially developed computer code was used for this calculation based on

LabView, the commercial data processing computer software provided by the National Instruments Corporation. A common characteristic to the graphs in this Figure is the spike at 30 Hz, which is attributed to the periodical nature of the generation of vortices downstream the obstacle. This frequency corresponds to the period of oscillations in Figure 3, which is 16.7 dimensionless time units. The reference time h/U_0 is 1.96 msec, so that in dimensions of time, the period of oscillations obtains the value 32.88 msec. Finally the inverse of the period is 30.4 Hz, which is the frequency of the spike.

The entsrophy spectra for this flow are shown in Figure 5 at the same point as the energy spectra. The estrophy was calculated as one half of the square of vorticity. The -1 power law in the decay of enstrophy covers one order of magnitude in the frequency range. The agreement of the results of the spectral analysis of this work with theoretical results of quasi two dimensional turbulence confirm the accuracy of the numerical

predictions.

5.3 Statistical properties of the flow.

The autocorrelation function $R(\tau)$ at point (x=50,y=5) is shown in Figure 6. The dimensionless time difference t goes up to 150 time units. $R(\tau)$ has been evaluated at each time step $\Delta t = 0.01$ yielding 15000 points for the calculation of the curve in Figure 6. The autocorrelation function starts at 1, diminishes gradually, obtains negative values and finally tends to zero. This is a typically measured curve indicative of the randomness of the signal, which has already been observed in Figure 3.

The longitudinal and lateral correlation functions at point ($x=50,y_0=4$)are shown in Figure 7. The variation in the y- coordinate goes up to the upper wall of the wind tunnel of the computational domain (s. Figure 1). Both correlation functions start from 1 and gradually diminish. The function $R_u(y)$ obtains appreciable negative values like $R(\tau)$ and finally both cross correlations become zero at the wall due to the no-slip boundary condition there. Although, all three correlation functions are unable to contribute to the understanding of the physics of turbulence, they provide an accurate test for the quality of the computational results. These functions simply examine whether events that take place spatially or temporally away from a fixed point in space or time correlate in an increasingly fading fashion, as it should actually be for a random process, according to common sense. The computational results of this work satisfy this requirement of statistical analysis and verify the chaotic behavior of the oscillograms of the velocity fluctuations shown in Figure 3. It should be noted, that the compatibility of the results in figures 4-7 with the oscillograms shown in Figure 3, verify the accuracy of the numerical results and exclude any possibility of numerical noise.

6 CONCLUSIONS

In this work, the quasi two dimensional turbulent flow over a surface mounted obstacle has been studied numerically with the aim to compute fundamental turbulent characteristics.

Oscillograms of the velocity fluctuations confirm the periodic nature of the flow and exhibit the chaotic behavior of turbulence. The analysis of energy spectra yield the double cascade, which is a characteristic of quasi two dimensional turbulent flow. The decay of enstrophy with respect to the frequency of the spectrum follows the -1 power law. The computation of the autocorrelation function along with the longitudinal and lateral correlation functions yield the expected result that events far from a fixed point in space or time correlate in a fading fashion. This statistical analysis of the velocity fluctuations verify the chaotic nature of the flow and confirm the randomness in the patterns of the computed oscillograms.

The results of this work may be useful in two ways: on the one hand to colleagues who are engaged in the study of chaos that emerges from quasi two dimensional turbulence and on the other hand to colleagues who are engaged in the direct numerical simulation of turbulent flow and are interested in applying the alternative approach to initial and boundary conditions proposed here. Another challenge would be the use of the primitive variable formulation of the Navier Stokes equations in the study of flows with meteorological interest.

Acknowledgements

Finacial support for this work has been granted by the Office of the Hellenic Foundation for Research (Program: ARCHIMEDES).

REFERENCES

[1] Acharya S., Dutta S., Myrum T.A. and Baker R.S. (1994)``*Turbulent Flow Past a Surface-Mounted Two-Dimensional Rib*", ASME J. Fluids Eng., 116, pp. 238-246.

[2] Batchelor G.K. (1969) ``*Computation of the energy spectrum in homogeneous two-dimensional turbulence*", Physics of Fluids, Suppl. II, pp. 233-239.

[3] Fragos V.P., Psychoudaki S.P. and Malamataris N.A. (1997)``*Computer-Aided Analysis of Flow Past A Surface-Mounte Obstacle*", J. Numer. Meth. Fluids}, 25, pp. 495-512.

[4] Fragos V.P., Psychoudaki S.P. and Malamataris N.A.`*Direct simulation of quasi two dimensional turbulent flow over a surface mounted obstacle*", Computers & Fluids, acc.

[5] Friedrich R., Huttl T.J., Manhart M. and Wagner C. (2001) ``*Direct numerical simulation of incompressible flows", Computers \& Fluids*, 30, pp. 555-579.

[6] Frisch U. (1995) *Turbulence: The Legacy of A.N.Kolmoforov*, Cambridge University Press, Cambridge UK.

[7] Hwang R.R., Chow Y.C. and Peng Y.F. (1999) ``Numerical study of turbulent flow over two-dimensional surface-mounted ribs in a channel", Int. J. Numer. Meth. Fluids, 31, pp. 767-785.

[8] Kraichnan R.H. (1969) ``*Inertial ranges in two dimensional turbulence*", Physics of Fluids, Suppl. II, pp. 1417-1423.

[9] Larichkin V.V. and Yakovenko S.N. (2003) `` *Effect of boundary-layer thickness on the structure of a near-wall flow with a two-dimensional obstacle*",J. Appl. Mech. Techn Physics, 44, pp. 365-372.

[10]Lohasz M., Rambaud P. and Benocci C. (2003) *"LES simulation of ribbed square duct flow with FLUENT and comparison with PIV data*", CMFF'03, The 12th Int. Conf. on Fluid Flow Technologies, Budapest, Hungary.

[11] Moon F.C. (1992) *Chaotic and fractal dynamics*, J.Wiley & Sons, Inc.

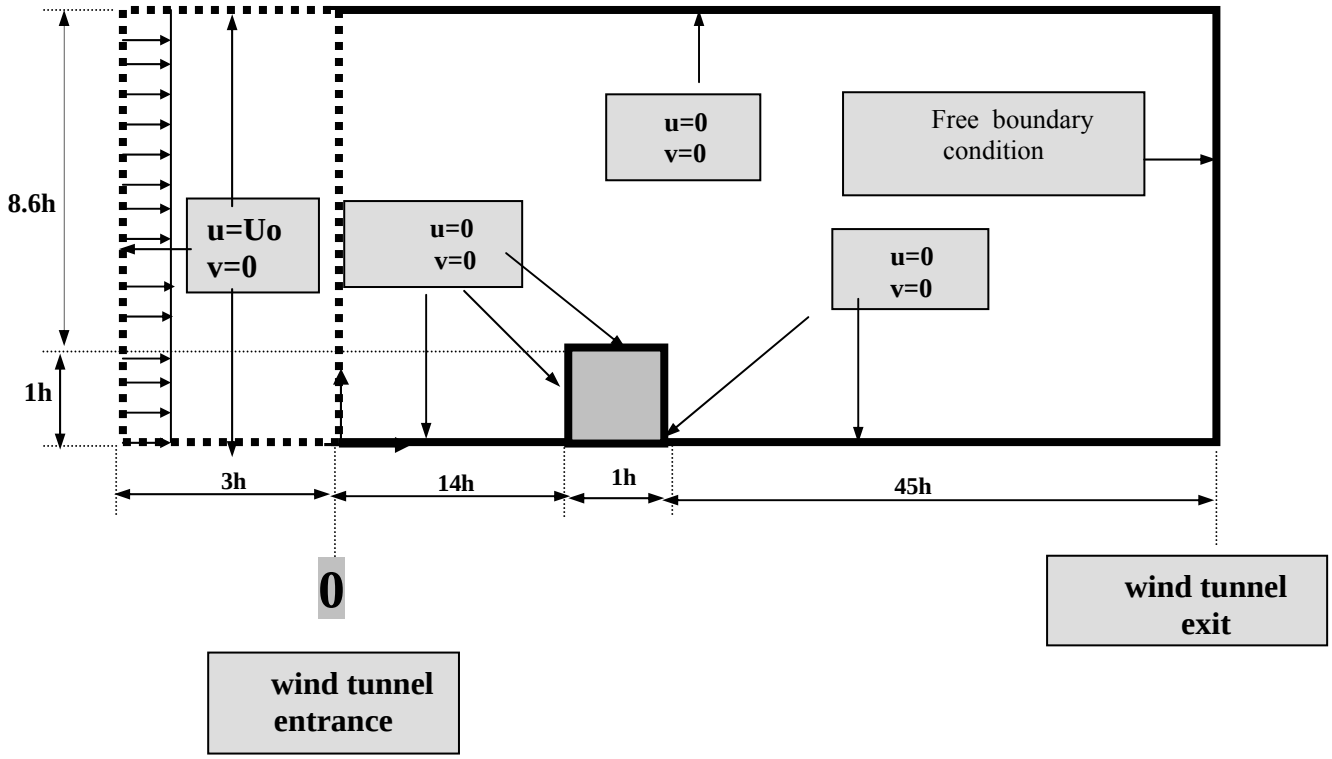


Figure 1. Computational domain of the quasi two dimensional turbulent flow over a surface mounted obstacle with a square cross section, placed in a wind tunnel.

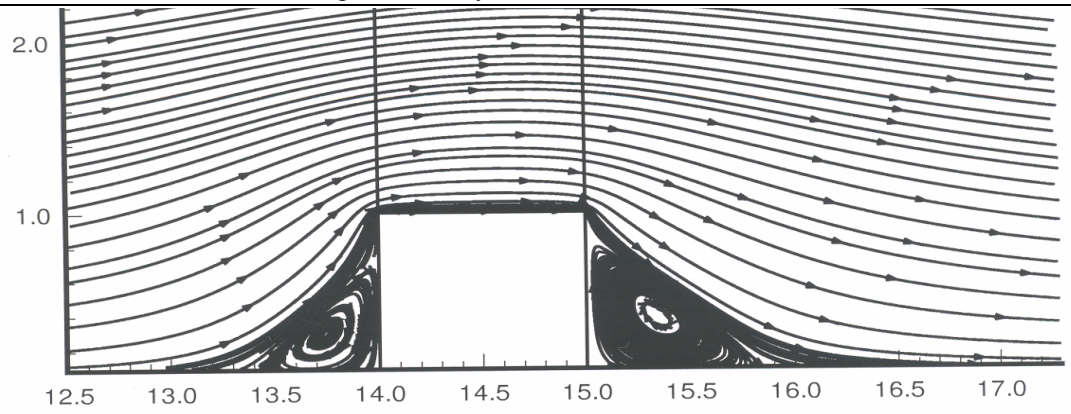
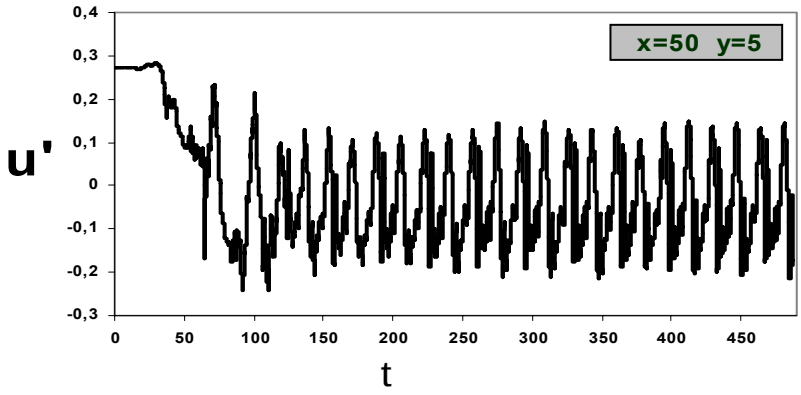
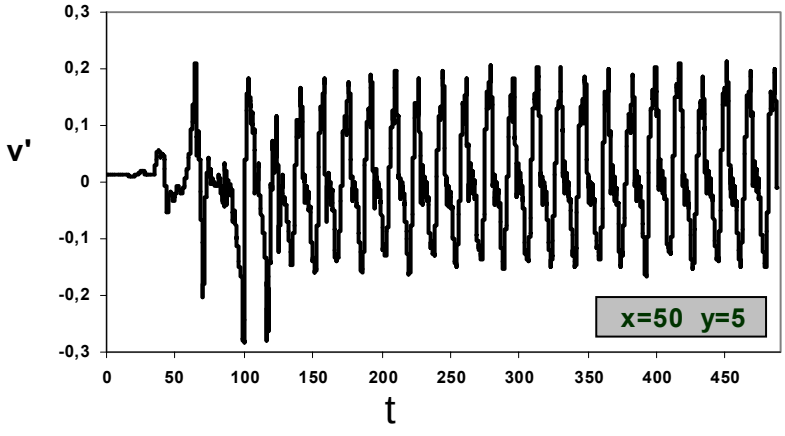


Figure 2. Streamlines of the laminar flow over a surface mounted obstacle at Reynolds number 1 with the respect to the obstacle height, chosen as the initial condition for this work.



(a)



(b)

Figure 3. Oscillations of the fluctuations of u (a) and v (b) velocity components at point (x=50,y=5) of the computational domain.

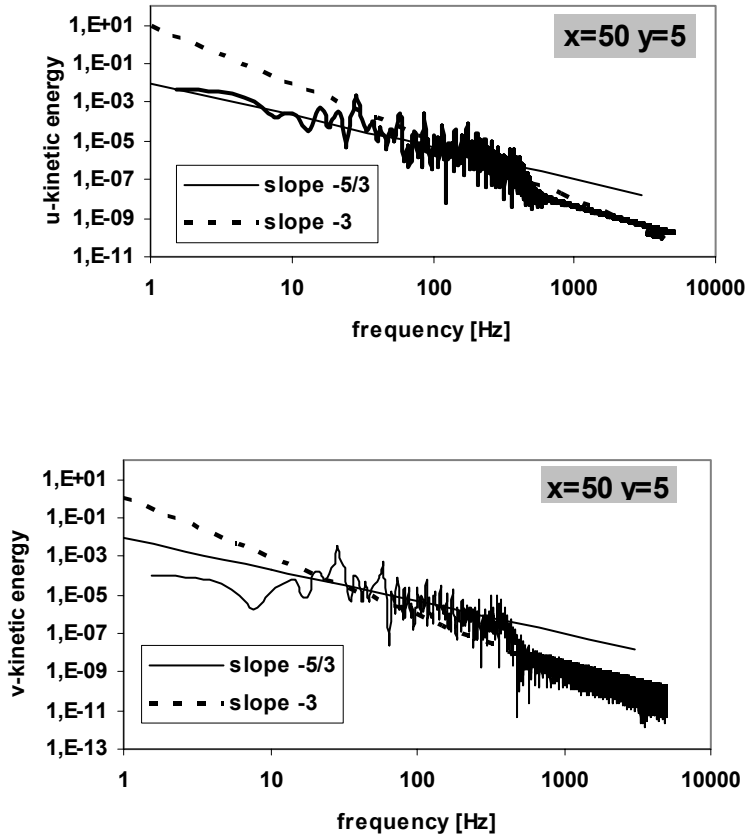


Figure 4. Energy spectra of the kinetic energy of both velocity components at point (x=50,y=5) of the computational domain.

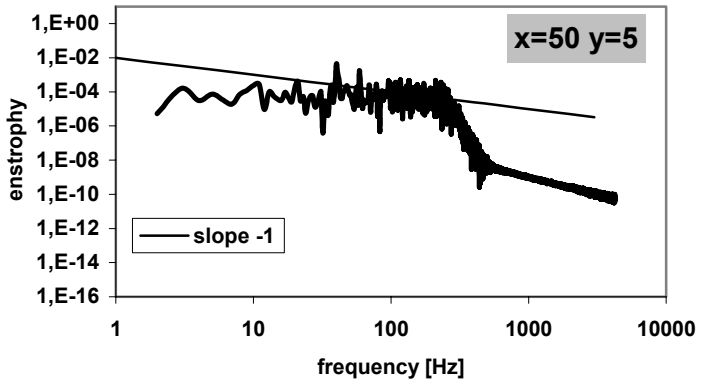


Figure 5. Enstrophy spectra at point (x=50,y=5) of the computational domain.

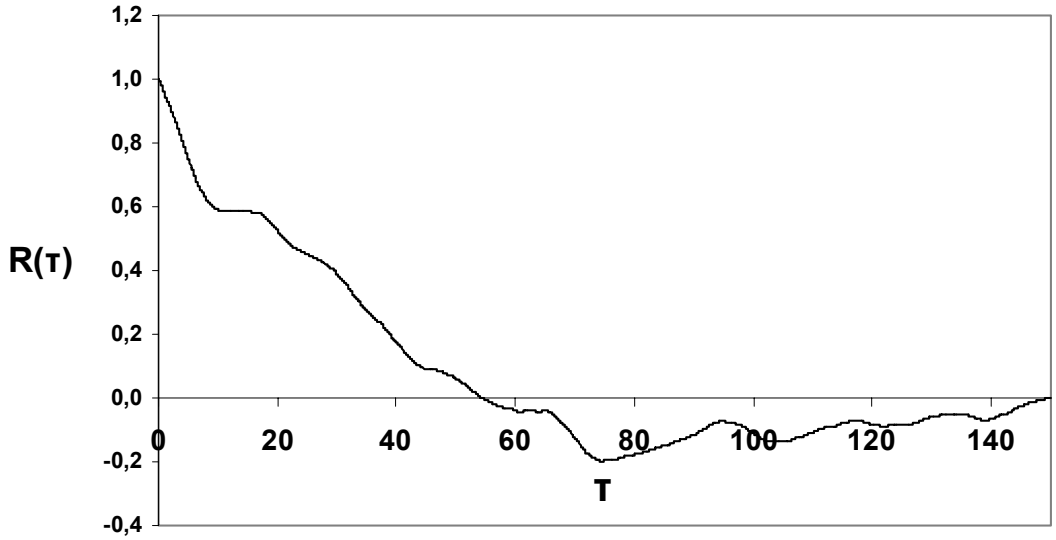


Figure 6. The Eulerian autocorrelation function at point (x=50, y=5) calculated from the computed fluctuation of the u-velocity component.

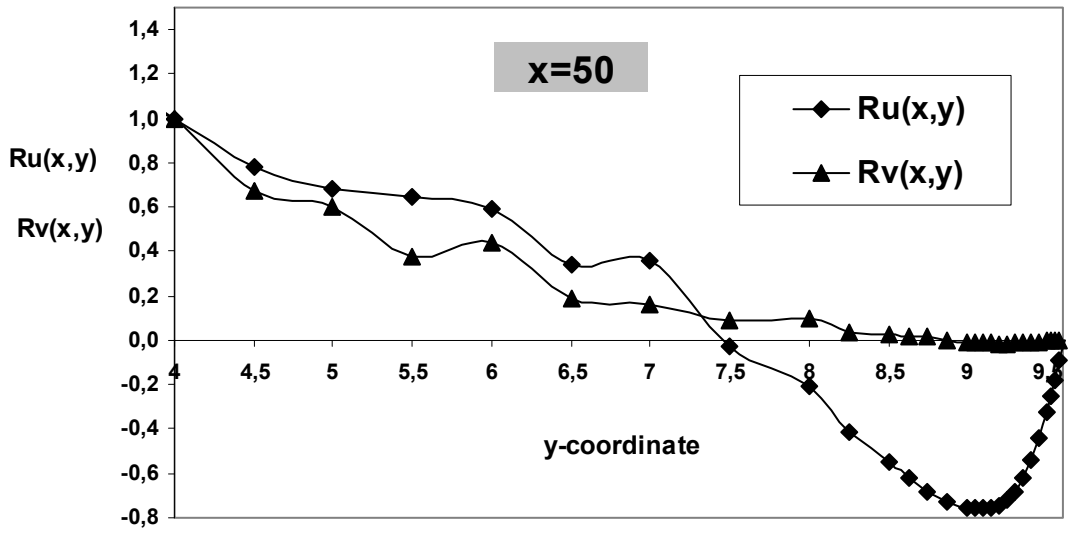


Figure 7. The longitudinal $R_u(y)$ and lateral $R_v(y)$ cross correlation function calculated from the computed fluctuations of the u- and v- velocity components respectively, with fixed point (x=50, y_0=5) and variation along y- coordinate of the computational domain up to the upper wall of the wind tunnel.

5th GRACM International Congress on Computational Mechanics
GRACM 05
Limassol, 29 June – 1 July, 2005

BUCKLING OF PILES

Ioannis N. Petrakis
Civil Engineer, MSc.
Therissou 112, 71305 Heraklion, Greece
e-mail: yanpetrakis@yahoo.gr

Keywords: Buckling, Piles, Deep Foundations.

Abstract. *An efficient method of assessing buckling response of piles is developed. The exact stiffness matrix method is employed for the stability analysis of piles embedded in elastic media. Shear deformations are taken into account. A bi-parametric linear elastic soil model is adopted. The method applies to the most common types of piles (end-bearing, friction, tapered) in either uniform, or variable (not necessarily obeying a specific variation function) subsoil conditions, and all kinds of boundary conditions. The differential equation of the elastic line for an elemental length is derived in terms of the relevant displacements and rotations. Each element stiffness matrix is composed relating forces and displacements at the ends of the element. Once the global stiffness matrix is assembled, stability analysis is performed on the basis of the solution of a transcendental eigenvalue problem. The method of converging with certainty upon the critical buckling load is based on the Wittrick-Williams algorithm. The corresponding eigenvector is obtained using Gauss elimination and then performing a normal back substitution.*

1 INTRODUCTION

Modern design practices in deep foundation engineering in conjunction with the extensive use of steel, high strength alloys and composite materials gradually point more attention at the instability problem of piled structures. Until recently, the vast amount of the published work on theoretical concepts of assessing pile capacity and behavior tended to ignore or treat as a minor sub-context the buckling problem. On the other hand, the work directly or indirectly related to the topic was focused on very specific and limited cases. The present work aims to present a consistent theoretical approach on the buckling problem of piles and develop a calculation method covering most of the cases that can be encountered in engineering practice.

Although in buckling problems only the first (critical) buckling load is of practical importance and shear deformations become progressively more important in higher numbered buckling modes, modern construction materials are much more shear sensitive and shear deformations can be quite significant, even on the lowest critical buckling load. It has been shown by Banerjee and Williams[1] that the maximum error encountered by ignoring the effect of shear deflections can be as much as 10%-12%.

The present paper gives the exact stiffness coefficients of a pile member which is embedded in a linearly elastic medium. It is assumed that the pile material is homogeneous, isotropic and elastic, the stresses remain within the elastic limits, and the displacements are infinitesimal, there are no initial imperfections or loading eccentricity and that the mass is uniformly distributed along the member. The force-displacement relationship of the surrounding medium is assumed to be linearly elastic and the Winkler soil model is adopted. From these member stiffnesses the overall stiffness matrix **K** can be assembled. Stability analysis can be performed on the basis of the solution of a transcendental eigenvalue problem. The buckling loads correspond to the values of P which satisfy the matrix equation $\mathbf{K}(P)\mathbf{D}=\mathbf{0}$, where **D** is the vector of the amplitudes of the nodal displacements and P is the axial compressive load by which the datum member forces are scaled. In structural systems with a uniform distribution of mass, an infinite number of values for the axial force, P, satisfy the stability criterion set previously. In buckling problems only the first (critical) buckling load is of practical importance, but its determination may present a number of problems, as any trial and error method, which involves computing $|\mathbf{K}|$ and observing when it changes sign, may be proved not credible. In order to overcome all these problems, the Wittrick-Williams algorithm can be employed to converge with certainty upon the critical buckling load. The eigenvector that corresponds to the critical buckling condition can be obtained by determining the response of the pile to a small disturbance when $\mathbf{K}(P)$ is evaluated at $P=P_{cr}$.

The buckling problem of piles can be performed on the same basis as the problem of buckling of axially loaded beams resting on elastic media. The classical treatise on the analysis of a finite beam resting on a Winkler medium is that of Hetenyi[2]. Since its publication, several alternative procedures have been advanced. Matrix methods have been employed by Iyengar[3], Mozingo[4], Frazer[5] and Bowles[6][7]. Timoshenko and Gere[8] treat the buckling problem taking into consideration the effects of shearing deformations through analytical and energy methods. Selvadurai[9] presents a series of solutions using the method of initial parameters, the method of superposition and the energy method. Poulos and Davies[10] have concluded their work on the buckling of piles by presenting solutions for different types of piles embedded in parameterized elastic media. Since then, numerous authors have treated very specific cases of axially loaded piles using various methods. An exhaustive reference to their works is beyond the scope of the present paper.

2 THEORY

The differential equation of the elastic line for an elemental length, *dx*, is derived by applying the condition of equilibrium of the acting forces and moments, when transmission from a stable to an unstable configuration is taking place, in terms of the relevant displacements and rotations. Figure 1 shows a typical element of the pile undergoing in-plane buckling, subjected to a static axial load *P* (positive for compression) and embedded in an elastic medium with stiffnesses per unit length k_y and k_θ restraining, respectively, the lateral displacement *v* and the rotation ψ. The vertical and moment equilibria give:

$$Q-(Q+dQ)+k_y\cdot v\cdot dx=0\Rightarrow\frac{dQ}{dx}=k_y\cdot v \tag{1}$$

and

$$Q\cdot dx+P\cdot\frac{dv}{dx}dx+M+k_y\cdot v\cdot dx\cdot\frac{dx}{2}-(M+dM)-k_\theta\cdot dx\cdot\psi=0 \tag{2}$$

where *Q* is the transverse force due to bending (see Figure 1), M is the bending moment, *v* is the lateral displacement and ψ is the angle of rotation due to bending.

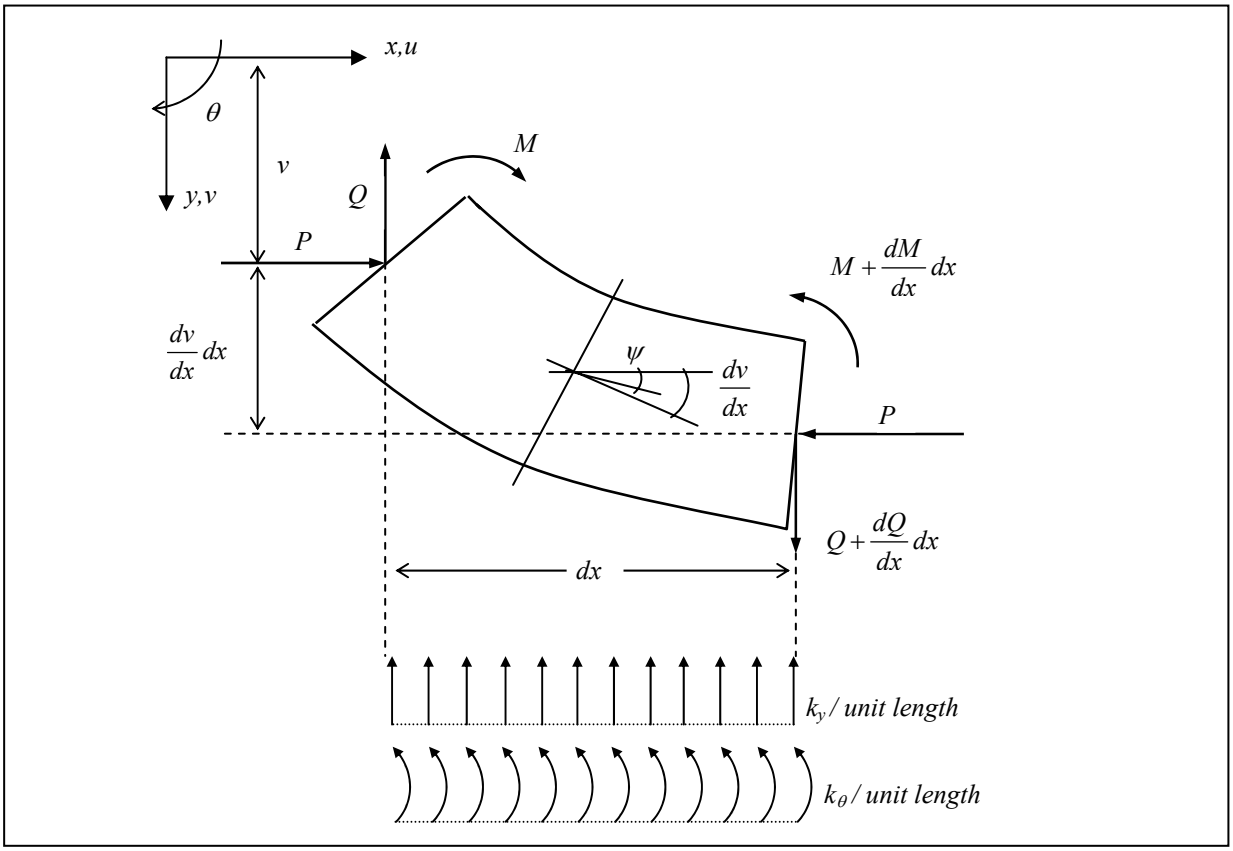


Figure 1. Forces acting on a typical elemental length of the pile

Simple bending theory gives:

$$M=-EI\cdot\frac{d\psi}{dx} \tag{3}$$

Finally, considering of the shearing of the element gives:

$$\psi=\frac{dv}{dx}\cdot(1-\frac{P}{\phi'})-\frac{Q}{\phi'} \tag{4}$$

where $\phi' = f \cdot A \cdot G$ and f is the section shape factor, A is the cross section area, E is Young's modulus, G is the modulus of rigidity (shear modulus) and I is the second moment of area. Eliminating Q, M and ψ from equations (1) – (4) yields the following fourth order differential equation:

$$EI \cdot (1-\frac{P}{\phi'}) \cdot \frac{d^4 v}{dx^4} + \left[P - \frac{k_y \cdot EI}{\phi'} - k_\theta \cdot (1-\frac{P}{\phi'}) \right] \cdot \frac{d^2 v}{dx^2} + k_y \cdot (1+\frac{k_\theta}{\phi'}) \cdot v = 0 \tag{5}$$

Changing equation (5) to a non-dimensional form using the dimensionless variable $\xi = \frac{x}{L}$, where L is the length of the element and using the parameters,

$$s^2 = \frac{EI}{\phi' L^2} \qquad p^2 = \frac{PL^2}{EI} \qquad k_y^* = -k_y \frac{L^4}{EI} \qquad k_\theta^* = -k_\theta \frac{L^2}{EI} \tag{6}$$

yields:

$$\frac{d^4 v}{d\xi^4} + \frac{p^2 + s^2 \cdot k_y^* + k_\theta^* \cdot (1 - s^2 \cdot p^2)}{1 - s^2 \cdot p^2} \cdot \frac{d^2 v}{d\xi^2} - \frac{k_y^* \cdot (1 - s^2 \cdot k_\theta^*)}{1 - s^2 \cdot p^2} \cdot v = 0 \tag{7}$$

This non-dimensional formulation is particularly convenient, since the effects of shear deformations are included if s^2 takes its natural value, and they are omitted when s^2 is set to zero.

Equation (7) is a linear differential equation with constant coefficients and hence it can be solved exactly. Its solution may be expressed in the general form as:

$$v = B_1 \sinh \theta\xi + B_2 \cosh \theta\xi + B_3 \sinh \varphi\xi + B_4 \cosh \varphi\xi \tag{8}$$

where:

$$\theta^2 = b + (b^2 - c)^{1/2}, \ \varphi^2 = b - (b^2 - c)^{1/2} \tag{9}$$

and

$$b = -\frac{p^2 + s^2 \cdot k_y^* + k_\theta^* \cdot (1 - s^2 \cdot p^2)}{2 \cdot (1 - s^2 \cdot p^2)}, \ c = -\frac{k_y^* \cdot (1 - s^2 \cdot k_\theta^*)}{1 - s^2 \cdot p^2} \tag{10}$$

Depending upon the values of b and c, the parameters θ and φ may be real, imaginary or complex. Therefore, any computing that is done on the basis of equation (8) must use a complex arithmetic routine.

The Force-Displacement relationship, for a typical element, using the familiar matrix notation could be expressed as:

$$\{\mathbf{F}\}=[\mathbf{K}]\{\mathbf{D}\} \tag{11}$$

where: $\{\mathbf{F}\}=[P_1, Q_1, M_1, P_2, Q_2, M_2]^T$ and $\{\mathbf{D}\}=[u_1, v_1, \psi_1, u_2, v_2, \psi_2]^T$. Identifiers 1, 2 refer to the ends of the element. Axial deformations, u, produced by the axial forces, P, can be neglected, as they are relatively very small. This assumption does not affect the resulting buckling behaviour of a compressed member and it reduces considerably the requisite work without dissimulating any useful information regarding the buckling response. Hence, the Force-Displacement relationship, for a typical element, could be restated as follows:

$$\begin{bmatrix} Q_1 \\ M_1 \\ Q_2 \\ M_2 \end{bmatrix} = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ & k_{22} & k_{23} & k_{24} \\ & & k_{33} & k_{34} \\ sym. & & & k_{44} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ \psi_1 \\ v_2 \\ \psi_2 \end{bmatrix} \tag{12}$$

Defining the sign convention in order to satisfy the compatibility requirements, the stiffness coefficients, k_{ij}, can be determined by the following expressions for v, ψ, Q, M, which are derived from equations (1) – (4) using the non-dimensional parameters p^2, s^2, k_y^*, k_θ^*, and ξ introduced previously:

$$\psi = \frac{1}{L}\left[\frac{s^2\left(1-s^2p^2\right)}{1-s^2k_\theta^*}\cdot\frac{d^3v}{d\xi^3} - s^2\left[\frac{-s^2k_y^*}{1-s^2k_\theta^*} - p^2 - k_\theta^*\left(1-s^2p^2\right) - \frac{1-s^2p^2}{s^2}\right]\cdot\frac{dv}{d\xi}\right] \tag{13}$$

$$Q = -\frac{EI}{L^3}\left[\frac{1-s^2p^2}{1-s^2k_\theta^*}\cdot\frac{d^3v}{d\xi^3} + \left[\frac{s^2k_\theta^*}{1-s^2k_\theta^*} + p^2 + k_\theta^*\left(1-s^2p^2\right)\right]\cdot\frac{dv}{d\xi}\right] \tag{14}$$

$$M = -\frac{EI}{L^2}\left[\left(1-s^2p^2\right)\cdot\frac{d^2v}{d\xi^2} + s^2k_\theta^*\cdot v\right] \tag{15}$$

Using the derivatives of v (8) in equations (13), (14), (15), ψ, Q and M could be expressed in terms of the B_i coefficients in the following form:

$$\psi = B_1\alpha_\theta\theta\cosh\theta\xi + B_2\alpha_\theta\theta\sinh\theta\xi + B_3\alpha_\varphi\varphi\cosh\varphi\xi + B_4\alpha_\varphi\varphi\sinh\varphi\xi \tag{16}$$

$$Q = B_1\beta_\theta\theta\cosh\theta\xi + B_2\beta_\theta\theta\sinh\theta\xi + B_3\beta_\varphi\varphi\cosh\varphi\xi + B_4\beta_\varphi\varphi\sinh\varphi\xi \tag{17}$$

$$M = B_1\delta_\theta\sinh\theta\xi + B_2\delta_\theta\cosh\theta\xi + B_3\delta_\varphi\sinh\varphi\xi + B_4\delta_\varphi\cosh\varphi\xi \tag{18}$$

where:

$$\alpha_\theta = \frac{\theta^2s^2\left(1-s^2p^2\right)}{L\left(1-s^2k_\theta^*\right)} - \frac{s^2}{L}\left[\frac{-s^2k_y^*}{1-s^2k_\theta^*} - p^2 - k_\theta^*\left(1-s^2p^2\right) - \frac{1-s^2p^2}{s^2}\right] \tag{19}$$

$$\alpha_\varphi = \frac{\varphi^2s^2\left(1-s^2p^2\right)}{L\left(1-s^2k_\theta^*\right)} - \frac{s^2}{L}\left[\frac{-s^2k_y^*}{1-s^2k_\theta^*} - p^2 - k_\theta^*\left(1-s^2p^2\right) - \frac{1-s^2p^2}{s^2}\right] \tag{20}$$

$$\beta_\theta = -\frac{EI\theta^2}{L^3}\cdot\frac{1-s^2p^2}{1-s^2k_\theta^*} + \frac{EI}{L^3}\left[\frac{-s^2k_y^*}{1-s^2k_\theta^*} - p^2 - k_\theta^*\left(1-s^2p^2\right)\right] \tag{21}$$

$$\beta_\varphi = -\frac{EI\varphi^2}{L^3}\cdot\frac{1-s^2p^2}{1-s^2k_\theta^*} + \frac{EI}{L^3}\left[\frac{-s^2k_y^*}{1-s^2k_\theta^*} - p^2 - k_\theta^*\left(1-s^2p^2\right)\right] \tag{22}$$

$$\delta_\theta = -\frac{EI\theta^2}{L^2}(1-s^2p^2) - \frac{EI}{L^2}s^2k_y^* \tag{23}$$

$$\delta_\varphi = -\frac{EI\varphi^2}{L^2}(1-s^2p^2) - \frac{EI}{L^2}s^2k_y^* \tag{24}$$

For a pile element and the general boundary conditions at the ends 1, 2, using expressions (8) and (16), the vector of displacements {**D**} can be expressed in terms of the B_i (i=1…4) coefficients as:

$$\{\mathbf{D}\}=[\mathbf{S}]\{\mathbf{B}\} \tag{25}$$

or analytically

$$\begin{bmatrix} v_1 \\ \psi_1 \\ v_2 \\ \psi_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ \theta\cdot\alpha_\theta & 0 & \varphi\cdot\alpha_\varphi & 0 \\ \sinh\theta & \cosh\theta & \sinh\varphi & \cosh\varphi \\ \theta\cdot\cosh\theta\cdot\alpha_\theta & \theta\cdot\sinh\theta\cdot\alpha_\theta & \varphi\cdot\cosh\varphi\cdot\alpha_\varphi & \varphi\cdot\sinh\varphi\cdot\alpha_\varphi \end{bmatrix}\cdot\begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{bmatrix}$$

Similarly, using expressions (14) and (15), the general boundary conditions and the global sign convention at the ends 1, 2 of the element, the force vector {**F**} can be expressed as:

$$\{\mathbf{F}\}=[\mathbf{S^*}]\{\mathbf{B}\} \tag{26}$$

or analytically

$$\begin{bmatrix} Q_1 \\ M_1 \\ Q_2 \\ M_2 \end{bmatrix} = \begin{bmatrix} -\theta\cdot\beta_\varphi & 0 & -\varphi\cdot\beta_\varphi & 0 \\ 0 & \delta_\theta & 0 & \delta_\varphi \\ \theta\cdot\cosh\theta\cdot\beta_\theta & \theta\cdot\sinh\theta\cdot\beta_\varphi & \varphi\cdot\cosh\varphi\cdot\beta_\varphi & \varphi\cdot\sinh\varphi\cdot\beta_\varphi \\ -\sinh\theta\cdot\delta_\theta & -\cosh\theta\cdot\delta_\theta & -\sinh\varphi\cdot\delta_\varphi & -\cosh\varphi\cdot\delta_\varphi \end{bmatrix} \cdot \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{bmatrix}$$

As matrix [**S**] is not singular, it can be inverted and equation (25) yields: {**B**}=[**S**]$^{-1}${**D**}. Substituting in (26) gives: {**F**}=[**S***][**S**]$^{-1}${**D**}. Hence, [**K**]=[**S***][**S**]$^{-1}$ and the coefficients k_{ij} are given by the following relationships:

$$k_{11}=\varphi\cdot\theta\cdot\left(\beta_\theta\cdot\alpha_\varphi-\beta_\varphi\cdot\alpha_\theta\right)\cdot\left(\varphi\cdot\alpha_\varphi\cdot\sinh\varphi\cdot\cosh\theta-\theta\cdot\alpha_\theta\cdot\sinh\theta\cdot\cosh\varphi\right)/\Omega \tag{27}$$

$$k_{12}=\left[\varphi\cdot\theta\cdot\left(\beta_\theta\cdot\alpha_\varphi-\beta_\varphi\cdot\alpha_\theta\right)\cdot\left(\cosh\varphi\cdot\cosh\theta-1\right)-\sinh\varphi\cdot\sinh\theta\cdot\left(\theta^2\cdot\beta_\theta\cdot\alpha_\theta+\varphi^2\cdot\beta_\varphi\cdot\alpha_\varphi\right)\right]/\Omega \tag{28}$$

$$k_{13}=\varphi\cdot\theta\cdot\left(\beta_\theta\cdot\alpha_\varphi-\beta_\varphi\cdot\alpha_\theta\right)\cdot\left(\theta\cdot\alpha_\theta\cdot\sinh\theta-\varphi\cdot\alpha_\varphi\cdot\sinh\varphi\right)/\Omega \tag{29}$$

$$k_{14}=\varphi\cdot\theta\cdot\left(\beta_\varphi\cdot\alpha_\theta-\beta_\theta\cdot\alpha_\varphi\right)\cdot\left(\cosh\theta-\cosh\varphi\right)/\Omega \tag{30}$$

$$k_{22}=\left(\delta_\theta-\delta_\varphi\right)\cdot\left(\varphi\cdot\alpha_\varphi\cdot\sinh\theta\cdot\cosh\varphi-\theta\cdot\alpha_\theta\cdot\sinh\varphi\cdot\cosh\theta\right)/\Omega \tag{31}$$

$$k_{23}=\varphi\cdot\theta\cdot\alpha_\varphi\cdot\alpha_\theta\cdot\left(\delta_\varphi-\delta_\theta\right)\cdot\left(\cosh\varphi-\cosh\theta\right)/\Omega \tag{32}$$

$$k_{24}=\left(\delta_\theta-\delta_\varphi\right)\cdot\left(\theta\cdot\alpha_\theta\cdot\sinh\varphi-\varphi\cdot\alpha_\varphi\cdot\sinh\theta\right)/\Omega \tag{33}$$

$$k_{33}=\varphi\cdot\theta\cdot\left(\beta_\theta\cdot\alpha_\varphi-\beta_\varphi\cdot\alpha_\theta\right)\cdot\left(\varphi\cdot\alpha_\varphi\cdot\sinh\varphi\cdot\cosh\theta-\theta\cdot\alpha_\theta\cdot\sinh\theta\cdot\cosh\varphi\right)/\Omega \tag{34}$$

$$k_{34}=\left[\theta\cdot\varphi\cdot\left(\beta_\theta\cdot\alpha_\varphi+\beta_\varphi\cdot\alpha_\theta\right)\cdot\left(1-\cosh\theta\cdot\cosh\varphi\right)+\sinh\theta\cdot\sinh\varphi\cdot\left(\theta^2\cdot\beta_\theta\cdot\alpha_\theta+\varphi^2\cdot\beta_\varphi\cdot\alpha_\varphi\right)\right]/\Omega \tag{35}$$

$$k_{44}=\left(\delta_\theta-\delta_\varphi\right)\cdot\left(\phi\cdot\alpha_\varphi\cdot\sinh\theta\cdot\cosh\varphi-\theta\cdot\alpha_\theta\cdot\sinh\varphi\cdot\cosh\theta\right)/\Omega \tag{36}$$

where:

$$\Omega=2\cdot\varphi\cdot\theta\cdot\alpha_\varphi\cdot\alpha_\theta\cdot\left(1-\cosh\varphi\cdot\cosh\theta\right)+\sinh\varphi\cdot\sinh\theta\cdot\left(\theta^2\cdot\alpha_\theta^2+\varphi^2\cdot\alpha_\varphi^2\right) \tag{37}$$

Once the component element stiffness matrices have been established, they can be assembled into the global (structure) stiffness matrix. For line structures, as piles, where the element and the global axes coincide, the procedure is quite simple and straightforward. Once the global stiffness matrix is assembled, stability analysis can be performed on the basis of the solution of a transcendental eigenvalue problem. The method of converging with certainty upon the critical buckling load is based on the Wittrick-Williams[11] algorithm. This algorithm states that:

$$J=J_0+s\{\mathbf{K}\} \tag{38}$$

where J represents the number of the buckling loads of the structure exceeded by some trial value of the axial load, P^*, J_0 represents the number of buckling loads which would be still exceeded if constraints were imposed upon the structure so as to make all the nodal displacements, **D**, zero, and $s\{\mathbf{K}\}$ is the sign count of the matrix **K**. $s\{\mathbf{K}\}$ is equal to the number of negative elements of the leading diagonal of the upper (or lower) triangular matrix obtained from **K**, when $P=P^*$ by the Gaussian elimination procedure in the following form. The rows of **K** are taken pivotal in order and suitable multiples of each pivotal row are added to all following rows so as all the elements in the pivotal column, which are below the pivot, become zero. The computation of J_0 is more complicated and it is dealt with at the following paragraph. It is quite straightforward then to assume that if J can

be established for any particular trial value of the axial compressive force, P^*, it is easy to develop a procedure for converging upon any required buckling load to any required accuracy.

As stated in the previous paragraph, where the Wittrick-Williams algorithm is presented, the number of the buckling loads exceeded by some trial value of the axial load, P^*, can be determined by the sum of J_o and $s\{\mathbf{K}\}$. J_o represents the number of buckling loads that would still be exceeded if constraints were imposed upon the structure so as to make all the nodal displacements, **D**, equal to zero. This action has the effect of isolating each individual structural element so that J_o can be calculated from the relationship:

$$J_o = \sum J_m \tag{39}$$

where J_m is the number of buckling loads of a component member, with its ends clamped, which have been exceeded by P^*, and the summation extends all over the component members. This value cannot be determined directly, but a convenient procedure, based again on Wittrick-Williams algorithm, is followed: if each component member is treated as a structure, then the relationship: $J_{pp}=J_m+s\{\mathbf{K}\}$ is again valid. J_{pp} represents the buckling loads exceeded by the trial value P^* if the member was pin supported, $s\{\mathbf{K}\}$ is the sign count of the -pin supported- member stiffness matrix and J_m is the number of buckling loads exceeded by the trial value P^* for the clamped-ended member which is the required value. In other words, for any of the component members, J_m can be calculated by the relationship:

$$J_m = J_{pp} - s\{\mathbf{K}\} \tag{40}$$

J_{pp} for each component member can be determined by the following procedure. Assuming the member is simply supported, the boundary conditions are:

$$v=M=0 \text{ i.e. } v=\frac{d\psi}{d\xi}=0$$

Assuming solutions of the form: $v=C\sin a\xi$, where $a=n\pi$, the boundary conditions are satisfied. Clearly, $v=0$ for $\xi=0$, $\xi=1$ (values of ξ at the end supports) as $\sin 0=\sin\pi=0$. Equations (1) and (4) yield:

$$\psi=\frac{1}{L}\left(1-s^2p^2\right)\cdot\frac{dv}{d\xi}-\frac{Q}{\phi'} \tag{41}$$

$$\frac{dQ}{d\xi}=k_yL\cdot v \tag{42}$$

Differentiating equation (41) with respect to ξ and using equation (42), yields:

$$\frac{d\psi}{d\xi}=\frac{1}{L}\left(1-s^2p^2\right)\cdot\frac{d^2v}{d\xi^2}-\frac{k_yL}{\phi'}\cdot v \tag{43}$$

which is satisfied for the boundary conditions by substituting for v ($=C\sin a\xi$) and for $\frac{d^2v}{d\xi^2}$ ($=-Ca^2\sin a\xi$). Therefore, $v=C\sin a\xi$, is a solution for the general equation. Substituting v and the derivatives of v in the governing differential equation (7) yields:

$$\left(1-s^2p^2\right)Ca^4\sin a\xi+\left[p^2+s^2k_y^*+k_\theta^*\left(1-s^2p^2\right)\right]\left(-Ca^2\sin a\xi\right)-\left[k_y^*\left(1-s^2k_\theta^*\right)\right]C\sin a\xi=0 \tag{44}$$

Solving for p yields:

$$p^2=\frac{a^4-a^2\left(s^2k_y^*+k_\theta^*\right)-k_y^*\left(1-s^2k_\theta^*\right)}{a^2\left[1+s^2\left(a^2-k_\theta^*\right)\right]} \tag{45}$$

Substituting for $p^2=\frac{PL^2}{EI}$ and $a=n\pi$ it gives:

$$P_n = \frac{EI}{L^2}\left\{\frac{n^4\pi^4 - n^2\pi^2\left(s^2 k_y^* + k_\theta^*\right) - k_y^*\left(1 - s^2 k_\theta^*\right)}{n^2\pi^2\left[1 + s^2\left(n^2\pi^2 - k_\theta^*\right)\right]}\right\},\ n=1,\ 2,\ 3,\ldots\ldots \tag{46}$$

Consequently, the first n critical buckling loads can be calculated for each pin ended component member. Hence, they can be stored and whenever a new trial load, P^*, is applied, it becomes easy to check how many pin-pin buckling loads have been exceeded for each member, determining the J_{pp} value. The sign count, $s\{\mathbf{K}\}$, of a pin ended member, represents the number of the negative elements on the leading diagonal of the upper triangular matrix obtained from the pin ended member stiffness matrix (consequently of order 2x2) using Gauss-elimination. Provided that the element stiffness matrix is already formulated, sign count, $s\{\mathbf{K}\}$, can be easily determined. As soon as J_m is calculated for each component member, J_o can be obtained by the algebraic sum of all the J_m.

The method for converging upon the critical (first) buckling load is based on the Wittrick-Williams algorithm and it utilises a bisection iterative procedure. A lower bound and an upper bound are taken as default, setting the initial limits where the P_{cr} lies in. For any trial value of the axial load, P^*, J is determined and its value signifies whether the first (critical) buckling load has been exceeded by the trial load, i.e. if J equals zero, no buckling load have been passed, and if J takes any other value, J buckling loads have been passed. Thus, if J equals zero, P^* is taken as a lower bound, $P_L = P^*$, and if J is greater than zero, P^* becomes an upper bound, $P_U = P^*$. Once a new value for either P_U or P_L has been established, P^* takes the value given by: $P^* = (P_U + P_L)/2$ and the procedure continues with a new iteration until $P_U - P_L \le CV$. CV is a value given as data so as to define the required accuracy. The greatest advantages of this method are the high converging rate by minimising the number of the required iterations and the fact that the result is determined with a predefined accuracy.

For axially, simply supported, compressed elements without lateral support along their length, a single half-wave can always describe the mode shape in the critical buckling condition. This response can be predicted by all the employed analytical and approximate methods of analysis. However, for laterally supported compressed elements, as piles embedded in elastic media and different end support conditions, this mode shape no longer describes rigorously the critical buckling modal response. The method of retrieving the eigenvector that corresponds to the critical buckling condition is based on the fact that, since P_{cr} is only an approximation to the true critical buckling load, an approximation to the true eigenvector can be obtained by determining the response of the pile to a small disturbance when $\mathbf{K}(P)$ is evaluated at $P=P_{cr}$. Hence, the problem reduces to one of calculating the nodal displacement vector, $\mathbf{D}$, from the established stiffness equation: $\mathbf{K}(P)\mathbf{D}=\mathbf{F}$ when $P=P_{cr}$ and $\mathbf{F}$ is an appropriate vector of forces, which depicts the small forcing disturbance. $\mathbf{D}$ can be obtained by reducing $\mathbf{K}(P_{cr})$ to the upper triangular matrix $\mathbf{K}^{\Delta}(P_{cr})$ using Gauss elimination, and then performing a normal back substitution. The resulting eigenvector is finally normalised by dividing throughout by the absolute value of the numerically largest element. The forcing vector should be such as to ensure that the required eigenvalue would be excited in every possible case. In order to maximise this possibility, the elements of the force vector, $\mathbf{P}$, should be chosen all to be the same order of magnitude. By the procedure described to generate the eigenvector in a critical buckling condition, it is likely to obtain meaningless results like in cases when the modal displacements occur between nodes, i.e. $\mathbf{D}=0$ or in cases where the information concerning the eigenvector are not sufficient to establish comprehensively the corresponding mode shape. In such cases, a simple solution is to redefine the nodal mesh, with sufficient extra nodes, and to re-run the problem.

3 CONCLUSIONS

This paper gives the exact stiffness coefficients for an axially compressed pile member embedded in an elastic medium taking into account shear deformations. It gives also the method for converging with certainty upon the critical buckling load and the critical mode shape. The method applies to all types of piles. Accounting for the variable parameters along the length of the pile in terms of the geometric and material properties as well as loading, boundary and subsoil conditions, an element discretisation is necessary in order to obtain an adequate model and consequently a sufficiently accurate and meaningful solution.

The substantial difference between the employed exact stiffness matrix method, compared to the traditional finite element techniques, is that the numerical answers are “exact” because the “exact” differential equation of the elastic line of the buckled element is been solved instead of solutions based on an assumed deflected shape of the buckled element. The validity of the equations presented has been confirmed by comparing calculated buckling loads and the corresponding mode shapes with those obtained from analytical solutions, like those in reference [8]. Profoundly, any computational effort on the basis of the developed theory without the aid of a computer would be rather impossible. But the greatest advantage of the method is its applicability limits while its flexible and adaptable framework is able to accommodate or easily modified to accommodate individual needs concerning the multi-parametric nature of the problem.

ACKNOWLEDGMENTS

The present work was supervised by Dr. W. P. Howson from Cardiff University. The author acknowledges his support.

REFERENCES

[1] BANERJEE, J.R. and WILLIAMS, F.W., (1994), "The effect of shear deformation on the critical buckling of columns", J. Sound and Vibration 174 (5), pp. 607-616.
[2] HETENYI, M. (1946), *Beams on elastic foundations*, University of Michigan Press, Ann Arbor, Michigan.
[3] IYENGAR, K.T.S.R. and ANANTHARAMU, S. (1965), "Influence lines for beams on elastic foundations", J. Struct. Div., Proc. A.S.C.E., 91, pp. 45-56.
[4] MOZINGO, R.R. (1967), "General method for beams on elastic supports", J. Struct. Div., Proc. A.S.C.E., 93, pp. 177-188.
[5] FRASER, D.J. (1969), "Beams on elastic foundations. A computer oriented solution for beams with free ends", Civ. Eng. Trans., Inst. Civ. Eng. Aust., 11 (1), pp. 25-30.
[6] BOWLES, J.E. (1974), *Analytical and computer methods in foundation engineering.* McGraw-Hill, New York.
[7] BOWLES, J.E. (1977), *Foundation analysis and design.* McGraw-Hill, New York.
[8] TIMOSHENKO, S.P. and GERE, J.M. (1961), *Theory of elastic stability.* McGraw-Hill, New York.
[9] SELVADURAI, A.P.S. (1979), *Elastic analysis of soil-foundation interaction.* Elsevier Scientific Publishing Company, Holland.
[10] POULOS, H.G. and DAVIS, E.H. (1980) *Pile foundation analysis and design.* John Wiley Inc, England.
[11] WITTRICK, W.H. and WILLIAMS, F.W. (1971), "A general algorithm for computing natural frequencies of elastic structures", Q. J. Mech. Appl. Math., 24(3), pp. 263-284.
[12] CAPRON, M.D. and WILLIAMS, F.W. (1988), "Exact dynamic stiffnesses for an axially loaded uniform Timoshenko member embedded in an elastic medium", J. Sound and Vibration 124 (3), pp. 453-466.
[13] HOWSON, W.P. (1979), "A compact method for computing the eigenvalues and eigenvectors of plane frames", Advances in Engineering Software, v 1, n 4, pp. 181-190.
[14] HOWSON, W.P. (2001), "Critical buckling of shear sensitive spring supported columns", *International Conference on Enhancement & Promotion of Computational Methods in Engng & Science,* Shanghai, ISBN 7 5426 1598 X, pp. 368-373.
[15] HOWSON, W.P. and WILLIAMS, F.W. (1973), "Natural frequencies of frames with axially loaded Timoshenko members", J. Sound and Vibration, 26(4), pp. 503-515.
[16] WILLIAMS, F.W. and KENNEDY, D. (1987), "Exact dynamic member stiffnesses for a beam on an elastic foundation", Earthquake Engineering and Structural Dynamics 15, pp. 133-136.

FLEXURAL-TORSIONAL BUCKLING ANALYSIS OF BEAMS BY BEM

E.J. Sapountzakis [*†] **and G.C. Tsiatas** [†]

[*] School of Civil Engineering
National Technical University
Zografou Campus, GR-157 80 Athens, Greece
e-mail: cvsapoun@central.ntua.gr

[†] Greek Association of Computational Mechanics

Keywords: Flexural-torsional buckling, beam, boundary integral equation, analog equation method

Abstract. I*n this paper a boundary element method is developed for the general flexural-torsional buckling analysis of Euler Bernoulli beams of arbitrarily shaped cross section. The beam is subjected to a compressive centrally applied load together with arbitrarily transverse and torsional distributed loading, While its edges are restrained by the most general linear boundary conditions. The resulting boundary value problem, described by three coupled ordinary differential equations, is solved employing a boundary integral equation approach. Besides the effectiveness and accuracy of the developed method, a significant advantage is that the displacements as well as the stress resultants are computed at any cross-section of the beam using the respective integral representations as mathematical formulae. The general character of the proposed method is verified from the fact that all basic equations are formulated with respect to an arbitrary coordinate system, which is not restricted to the principal one. Several beams are analysed to illustrate the method and demonstrate its efficiency and wherever possible its accuracy. The range of applicability of the thin-tube theory is also investigated through examples with great practical interest.*

1 INTRODUCTION

Elastic stability of beams is one of the most important criteria in the design of structures subjected to compressive loads. This beam buckling analysis becomes much more complicated in the case the cross section's centroid does not coincide with its shear center (asymmetric beams), leading to the formulation of the flexural-torsional buckling problem. The extensive use of the aforementioned structural elements necessitates a reliable and accurate analysis of the flexural-torsional buckling problem.

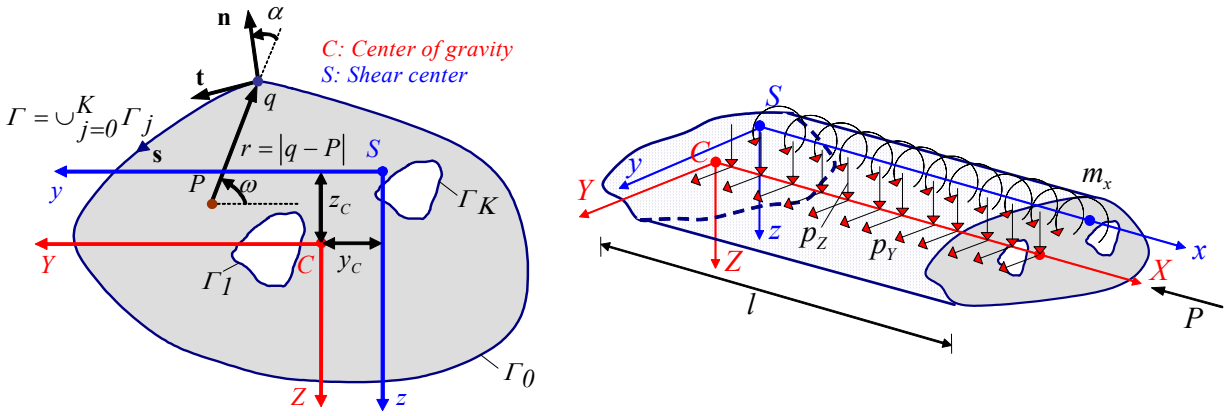
The flexural-torsional buckling problem of thin-walled beams, based on the assumptions of the thin tube theory, has been studied by many researchers[1-5]. To the authors' knowledge publications on the solution to the general problem of arbitrarily shaped cross sections do not exist.

In this investigation, an integral equation technique is developed for the solution of the general flexural-torsional buckling problem of beams of arbitrarily shaped cross-section. The beam is subjected to a compressive centrally applied load together with arbitrarily transverse and torsional distributed loading, while its edges are restrained by the most general linear boundary conditions. The solution method is based on the concept of the analog equation[6]. According to this method, the three coupled fourth order ordinary differential equations are replaced by three uncoupled ones subjected to fictitious load distributions under the same boundary conditions. The general character of the proposed method is verified from the fact that all basic equations are formulated with respect to an arbitrary coordinate system, which is not restricted to the principal one. Several beams are analysed to illustrate the method and demonstrate its efficiency and wherever possible its accuracy. The range of applicability of the thin-tube theory is also investigated through examples with great practical interest.

2 STATEMENT OF THE PROBLEM

Let us consider an initially straight Euler-Bernoulli beam of length l (Fig.1), of constant arbitrary cross-section of area A. The homogeneous isotropic and linearly elastic material of the beam cross-section occupies the region Ω of the y,z plane and is bounded by the $\Gamma_j \left(j = 1,2,...,K\right)$ boundary curves, which are piecewise smooth, i.e. they may have a finite number of corners. In Fig. 1a CYZ and Syz are coordinate systems (not necessarily principal) through the cross section's centroid C and shear center S, respectively. Moreover, y_C, z_C are the coordinates of the centroid C with respect to Syz system of axes. The beam is subjected to a compressive load P, to the combined action of the arbitrarily distributed transverse loading $p_Y = p_Y\left(X\right)$,

$p_Z = p_Z(X)$ acting in the Y and Z directions, respectively and to the arbitrarily distributed twisting moment $m_x = m_x(x)$ (Fig. 1b).



(a) (b)

Figure 1. Prismatic element of an arbitrarily shaped constant cross section occupying region Ω (a) subjected in bending, torsional and buckling loading (b).

The boundary value problem of the beam under consideration subjected to the combined action of flexure, torsion and compression is described by the following three coupled ordinary differential equations

$$EI_Z \frac{d^4v}{dx^4} + EI_{YZ} \frac{d^4w}{dx^4} + P\left(\frac{d^2v}{dx^2} - z_C \frac{d^2\theta}{dx^2} \right) = p_Y \tag{1}$$

$$EI_Y \frac{d^4w}{dx^4} + EI_{YZ} \frac{d^4v}{dx^4} + P\left(\frac{d^2w}{dx^2} + y_C \frac{d^2\theta}{dx^2} \right) = p_Z \tag{2}$$

$$EC_S \frac{d^4\theta}{dx^4} - GI_t \frac{d^2\theta}{dx^2} + P\left(\frac{I_S}{A} \frac{d^2\theta}{dx^2} - z_C \frac{d^2v}{dx^2} + y_C \frac{d^2w}{dx^2} \right) = m_x + p_Z y_C - p_Y z_C \tag{3}$$

inside the beam, subjected to the following boundary conditions

$$\alpha_1 v(x) + \alpha_2 R_Y(x) = \alpha_3 \qquad \bar{\alpha}_1 \frac{dv(x)}{dx} + \bar{\alpha}_2 M_Z(x) = \bar{\alpha}_3 \tag{4a,b}$$

$$\beta_1 w(x) + \beta_2 R_Z(x) = \beta_3 \qquad \bar{\beta}_1 \frac{dw(x)}{dx} + \bar{\beta}_2 M_Y(x) = \bar{\beta}_3 \tag{5a,b}$$

$$\gamma_1 \theta(x) + \gamma_2 M_t(x) = \gamma_3 \qquad \bar{\gamma}_1 \frac{d\theta(x)}{dx} + \bar{\gamma}_2 M_b(x) = \bar{\gamma}_3 \tag{6a,b}$$

at the beam ends $x = 0, l$, where $v = v(x)$, $w = w(x)$ are the deflections of the shear center along y, z axes, respectively; $\theta(x)$ is the angle of twist of the cross-section about the shear center S; E is the modulus of elasticity and I_Y, I_Z, I_{YZ} are the moments and the product of inertia with respect to the centroid C and I_S is the polar moment of inertia with respect to the shear center S. Moreover, C_S, I_t are the warping and torsion constants of the cross section, respectively given as[7]

$$C_S = \int_\Omega \left(\varphi_S^P \right)^2 d\Omega \tag{7}$$

$$I_t = \int_\Omega \left(y^2 + z^2 + y \frac{\partial \varphi_S^P}{\partial z} - z \frac{\partial \varphi_S^P}{\partial y} \right) d\Omega \tag{8}$$

where $\varphi_S^P(y,z)$ is the primary warping function with respect to the shear center S of the cross section of the beam (Fig. 1a). This function can be established by solving independently the Neumann problem

$$\nabla^2 \varphi_S^P = 0 \qquad \text{in } \Omega \tag{9}$$

$$\frac{\partial \varphi_S^P}{\partial n} = \frac{1}{2}\frac{\partial\left(\rho_S^2\right)}{\partial s} \qquad \text{on } \Gamma_j \ (j = 1,2,...,K) \tag{10}$$

where $\nabla^2 = \partial^2/\partial y^2 + \partial^2/\partial z^2$ is the Laplace operator; $\rho_S = \sqrt{y^2+z^2}$ is the distance of a point on the boundary Γ_j from the shear center S; $\partial/\partial n$ denotes the directional derivative normal to the boundary Γ_j and $\partial/\partial s$ denotes differentiation with respect to its arc length s.

In the boundary conditions (4), (5) R_Y, M_Y and R_Z, M_Z are the reactions and bending moments with respect to Y and Z axes, respectively, given as

$$R_Y = -P\frac{dv(x)}{dx} - EI_Z\frac{d^3v(x)}{dx^3} - EI_{YZ}\frac{d^3w(x)}{dx^3} \tag{11}$$

$$M_Y = -EI_Y\frac{d^2w(x)}{dx^2} - EI_{YZ}\frac{d^2v(x)}{dx^2} \tag{12}$$

$$R_Z = -P\frac{dw(x)}{dx} - EI_Y\frac{d^3w(x)}{dx^3} - EI_{YZ}\frac{d^3v(x)}{dx^3} \tag{13}$$

$$M_Z = EI_Z\frac{d^2v(x)}{dx^2} + EI_{YZ}\frac{d^2w(x)}{dx^2} \tag{14}$$

while in eqns. (6) M_t and M_b are the torsional and warping moments, respectively, given as[7]

$$M_t = -EC_S\frac{d^3\theta(x)}{dx^3} + GI_t\frac{d\theta(x)}{dx} \tag{15}$$

$$M_b = -EC_S\frac{d^2\theta(x)}{dx^2} \tag{16}$$

Finally, $\alpha_k, \bar{\alpha}_k, \beta_k, \bar{\beta}_k, \gamma_k, \bar{\gamma}_k$ $(k = 1,2,3)$ are functions specified at the beam ends $x = 0,l$. Eqs. (4-6) describe the most general linear boundary conditions associated with the problem at hand and can include elastic support or restrain. It is apparent that all types of the conventional boundary conditions (clamped, simply supported, free or guided edge) can be derived form these equations by specifying appropriately these functions (e.g. for a clamped edge it is $\alpha_1 = \beta_1 = \gamma_1 = 1$, $\bar{\alpha}_1 = \bar{\beta}_1 = \bar{\gamma}_1 = 1$, $\alpha_2 = \alpha_3 = \beta_2 = \beta_3 = \gamma_2 = \gamma_3 = \bar{\alpha}_2 = \bar{\alpha}_3 = \bar{\beta}_2 = \bar{\beta}_3 = \bar{\gamma}_2 = \bar{\gamma}_3 = 0$).

3 INTEGRAL REPRESENTATIONS - NUMERICAL SOLUTION

According to the precedent analysis, the flexural-torsional buckling problem of a beam reduces in establishing the displacement components $v(x)$, $w(x)$ and $\theta(x)$ having continuous derivatives up to the fourth order satisfying the coupled governing equations (1)-(3) inside the beam and the boundary conditions (4)-(6) at the beam ends $x = 0,l$.

Eqns (1)-(3) are solved using the Analog Equation Method as it is developed for ordinary differential equations in[8]. This method is applied for the problem at hand as follows. Let $v(x)$, $w(x)$ and $\theta(x)$ be the sought solution of the boundary value problem described by eqns. (1)-(3) and (4)-(6). Setting as $u_1(x) = v(x)$, $u_2(x) = w(x)$, $u_3(x) = \theta(x)$ and differentiating these functions four times yields

$$\frac{d^4 u_i}{dx^4} = b_i(x) \quad (i = 1,2,3) \tag{17}$$

Eqns. (17) indicate that the solution of eqns. (1)-(3) can be established by solving eqns. (17) under the same boundary conditions (4)-(6), provided that the fictitious load distributions $b_i(x)$ $(i = 1,2,3)$ are first established. These distributions can be determined using BEM as follows.

The solution of eqns (17) is given in integral form as[8]

$$u_i(x) = \int_0^l b_i u^* dx - \left[u^* \frac{d^3 u_i}{dx^3} - \frac{du^*}{dx}\frac{d^2 u_i}{dx^2} + \frac{d^2 u^*}{dx^2}\frac{du_i}{dx} - \frac{d^3 u^*}{dx^3} u_i \right]_0^l \tag{18}$$

where u^* is the fundamental solution given as

$$u^* = \frac{1}{12} l^3 \left(2 + |\rho|^3 - 3|\rho|^2 \right) \tag{19}$$

with $\rho = r/l$, $r = x - \xi$, x, ξ points of the beam, which is a particular singular solution of the equation

$$\frac{d^4 u^*}{dx^4} = \delta(x, \xi) \tag{20}$$

Employing eqn (19) the integral representation (18) can be written as

$$u_i(x) = \int_0^l b_i \Lambda_4(r) dx - \left[\Lambda_4(r) \frac{d^3 u_i}{dx^3} + \Lambda_3(r) \frac{d^2 u_i}{dx^2} + \Lambda_2(r) \frac{du_i}{dx} + \Lambda_1(r) u_i \right]_0^l \tag{21}$$

where the kernels $\Lambda_j(r), (j = 1,2,3,4)$ are given as

$$\Lambda_1(r) = -\frac{1}{2} sgn\, \rho \qquad \Lambda_2(r) = -\frac{1}{2} l(1 - |\rho|) \tag{22a,b}$$

$$\Lambda_3(r) = -\frac{1}{4} l^2 |\rho| (|\rho| - 2) sgn\, \rho \qquad \Lambda_4(r) = \frac{1}{12} l^3 \left(2 + |\rho|^3 - 3|\rho|^2 \right) \tag{22c,d}$$

Notice that in eqn (21) for the line integral it is $r = x - \xi$, x, ξ points inside the beam, whereas for the rest terms it is $r = x - q$, x inside the beam, q at the beam ends $0, l$.

Differentiating eqn (21) results in the integral representations of the derivatives of u_i as

$$\frac{du_i(x)}{dx} = \int_0^l b_i \Lambda_3(r) dx - \left[\Lambda_3(r) \frac{d^3 u_i}{dx^3} + \Lambda_2(r) \frac{d^2 u_i}{dx^2} + \Lambda_1(r) \frac{du_i}{dx} \right]_0^l \tag{23a}$$

$$\frac{d^2 u_i(x)}{dx^2} = \int_0^l b_i \Lambda_2(r) dx - \left[\Lambda_2(r) \frac{d^3 u_i}{dx^3} + \Lambda_1(r) \frac{d^2 u_i}{dx^2} \right]_0^l \tag{23b}$$

$$\frac{d^3 u_i(x)}{dx^3} = \int_0^l b_i \Lambda_1(r) dx - \left[\Lambda_1(r) \frac{d^3 u_i}{dx^3} \right]_0^l \qquad \frac{d^4 u_i(x)}{dz^4} = b_i(x) \tag{23c,d}$$

The integral representations (21) and (23a), when applied for the beam ends $(0,l)$, together with the boundary conditions (4)-(6) are employed to express the unknown boundary quantities $u_i(q)$, $u_{i,x}(q)$, $u_{i,xx}(q)$ and $u_{i,xxx}(q)$ $(q = 0,l)$ in terms of b_i. This is accomplished numerically as follows.

The interval $(0,l)$ is divided into N equal elements (Fig. 2), on which $b_i(x)$ is assumed to vary according to certain law (constant, linear, parabolic etc). The constant element assumption is employed here as the numerical implementation becomes very simple and the obtained results are very good.

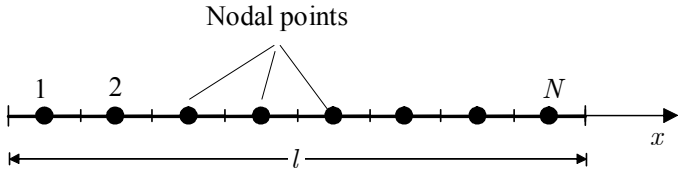


Figure 2. Discretization of the beam interval and distribution of the nodal points.

Employing the aforementioned procedure for the coupled boundary conditions (4), (5) the following set of linear equations is obtained

$$\begin{bmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} & \mathbf{0} & \mathbf{D}_{14} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_{18} \\ \mathbf{0} & \mathbf{D}_{21} & \mathbf{D}_{22} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_{27} & \mathbf{0} \\ \mathbf{E}_{31} & \mathbf{E}_{32} & \mathbf{E}_{33} & \mathbf{E}_{34} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{42} & \mathbf{E}_{43} & \mathbf{E}_{44} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_{53} & \mathbf{0} & \mathbf{D}_{55} & \mathbf{D}_{56} & \mathbf{0} & \mathbf{D}_{58} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_{63} & \mathbf{0} & \mathbf{0} & \mathbf{D}_{66} & \mathbf{D}_{67} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{E}_{31} & \mathbf{E}_{32} & \mathbf{E}_{33} & \mathbf{E}_{34} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{E}_{42} & \mathbf{E}_{43} & \mathbf{E}_{44} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_{1,x} \\ \hat{\mathbf{u}}_{1,xx} \\ \hat{\mathbf{u}}_{1,xxx} \\ \hat{\mathbf{u}}_2 \\ \hat{\mathbf{u}}_{2,x} \\ \hat{\mathbf{u}}_{2,xx} \\ \hat{\mathbf{u}}_{2,xxx} \end{Bmatrix} = \begin{Bmatrix} \boldsymbol{\alpha}_3 \\ \overline{\boldsymbol{\alpha}}_3 \\ \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\beta}_3 \\ \overline{\boldsymbol{\beta}}_3 \\ \mathbf{0} \\ \mathbf{0} \end{Bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{F}_3 \\ \mathbf{F}_4 \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \mathbf{b}_1 + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{F}_3 \\ \mathbf{F}_4 \end{bmatrix} \mathbf{b}_2 \tag{24}$$

while for the boundary conditions (6) we have

$$\begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} & \mathbf{0} & \mathbf{E}_{14} \\ \mathbf{0} & \mathbf{E}_{22} & \mathbf{E}_{23} & \mathbf{0} \\ \mathbf{E}_{31} & \mathbf{E}_{32} & \mathbf{E}_{33} & \mathbf{E}_{34} \\ \mathbf{0} & \mathbf{E}_{42} & \mathbf{E}_{43} & \mathbf{E}_{44} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{u}}_3 \\ \hat{\mathbf{u}}_{3,x} \\ \hat{\mathbf{u}}_{3,xx} \\ \hat{\mathbf{u}}_{3,xxx} \end{Bmatrix} = \begin{Bmatrix} \boldsymbol{\gamma}_3 \\ \overline{\boldsymbol{\gamma}}_3 \\ \mathbf{0} \\ \mathbf{0} \end{Bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{F}_3 \\ \mathbf{F}_4 \end{bmatrix} \mathbf{b}_3 \tag{25}$$

where $\mathbf{D}_{11}$, $\mathbf{D}_{12}$, $\mathbf{D}_{14}$, $\mathbf{D}_{18}$, $\mathbf{D}_{21}$, $\mathbf{D}_{22}$, $\mathbf{D}_{27}$, $\mathbf{D}_{53}$, $\mathbf{D}_{55}$, $\mathbf{D}_{56}$, $\mathbf{D}_{58}$, $\mathbf{D}_{63}$, $\mathbf{D}_{66}$, $\mathbf{D}_{67}$, $\mathbf{E}_{22}$, $\mathbf{E}_{23}$, $\mathbf{E}_{1j}$, ($j=1,2,4$) are 2×2 known square matrices including the values of the functions $a_j, \overline{a}_j, \beta_j, \overline{\beta}_j$ ($j=1,2$) of eqns (4)-(6); $\boldsymbol{\alpha}_3$, $\overline{\boldsymbol{\alpha}}_3$, $\boldsymbol{\beta}_3$, $\overline{\boldsymbol{\beta}}_3$, $\boldsymbol{\gamma}_3$, $\overline{\boldsymbol{\gamma}}_3$ are 2×1 known column matrices including the boundary values of the functions $a_3, \overline{a}_3, \beta_3, \overline{\beta}_3, \gamma_3, \overline{\gamma}_3$ of eqns (4)-(6); $\mathbf{E}_{jk}$, ($j=3,4$, $k=1,2,3,4$) are square 2×2 known coefficient matrices resulting from the values of the kernels $\Lambda_j(r)$ ($j=1,2,3,4$) at the beam ends and $\mathbf{F}_j$ ($j=3,4$) are $2\times N$ rectangular known matrices originating from the integration of the kernels on the axis of the beam. Moreover,

$$\hat{\mathbf{u}}_i = \left\{ u_i(0) \quad u_i(l) \right\}^T \qquad \hat{\mathbf{u}}_{i,x} = \left\{ \frac{du_i(0)}{dx} \quad \frac{du_i(l)}{dx} \right\}^T \tag{26a,b}$$

$$\hat{\mathbf{u}}_{i,xx} = \left\{ \frac{d^2u_i(0)}{dx^2} \quad \frac{d^2u_i(l)}{dx^2} \right\}^T \qquad \hat{\mathbf{u}}_{i,xxx} = \left\{ \frac{d^3u_i(0)}{dx^3} \quad \frac{d^3u_i(l)}{dx^3} \right\}^T \tag{26c,d}$$

are vectors including the two unknown boundary values of the respective boundary quantities and $\mathbf{b}_i = \left\{ b_1^i \; b_2^i \; \ldots \; b_N^i \right\}^T$ ($i=1,2,3$) is the vector including the N unknown nodal values of the fictitious load.

Discretization of eqns (21), (23) and application to the N collocation points yields

$$\mathbf{u}_i = \mathbf{C}_4\mathbf{b}_i - \left(\mathbf{E}_1\hat{\mathbf{u}}_i + \mathbf{E}_2\hat{\mathbf{u}}_{i,x} + \mathbf{E}_3\hat{\mathbf{u}}_{i,xx} + \mathbf{E}_4\hat{\mathbf{u}}_{i,xxx}\right) \tag{27a}$$

$$\mathbf{u}_{i,x} = \mathbf{C}_3\mathbf{b}_i - \left(\mathbf{E}_1\hat{\mathbf{u}}_{i,x} + \mathbf{E}_2\hat{\mathbf{u}}_{i,xx} + \mathbf{E}_3\hat{\mathbf{u}}_{i,xxx}\right) \tag{27b}$$

$$\mathbf{u}_{i,xx} = \mathbf{C}_2\mathbf{b}_i - \left(\mathbf{E}_1\hat{\mathbf{u}}_{i,xx} + \mathbf{E}_2\hat{\mathbf{u}}_{i,xxx}\right) \qquad \mathbf{u}_{i,xxx} = \mathbf{C}_1\mathbf{b}_i - \mathbf{E}_1\hat{\mathbf{u}}_{i,xxx} \qquad \mathbf{u}_{i,xxxx} = \mathbf{b}_i \tag{27c,d,e}$$

where $\mathbf{C}_j$ $(j = 1,2,3,4)$ are $N \times N$ known matrices; $\mathbf{E}_j$ $(j = 1,2,3,4)$ are $N \times 2$ also known matrices and $\mathbf{u}_i$, $\mathbf{u}_{i,x}$, $\mathbf{u}_{i,xx}$, $\mathbf{u}_{i,xxx}$, $\mathbf{u}_{i,xxxx}$ are vectors including the values of $u_i(x)$ and their derivatives at the N nodal points.

The above equations, after eliminating the boundary quantities employing eqns (24), (25), can be written as

$$\mathbf{u}_i = \mathbf{T}_i\mathbf{b}_i + \mathbf{t}_i \qquad \mathbf{u}_{i,x} = \mathbf{T}_{ix}\mathbf{b}_i + \mathbf{t}_{ix} \qquad \mathbf{u}_{i,xx} = \mathbf{T}_{ixx}\mathbf{b}_i + \mathbf{t}_{ixx} \tag{28a,b,c}$$

$$\mathbf{u}_{i,xxx} = \mathbf{T}_{ixxx}\mathbf{b}_i + \mathbf{t}_{ixxx} \qquad \mathbf{u}_{i,xxxx} = \mathbf{b}_i \tag{28d,e}$$

where $\mathbf{T}_i$, $\mathbf{T}_{ix}$, $\mathbf{T}_{ixx}$, $\mathbf{T}_{ixxx}$ are known $N \times N$ matrices and $\mathbf{t}_i$, $\mathbf{t}_{ix}$, $\mathbf{t}_{ixx}$, $\mathbf{t}_{ixxx}$ are known $N \times 1$ matrices. It is worth here noting that for homogeneous boundary conditions ($\alpha_3 = \bar{\alpha}_3 = \beta_3 = \bar{\beta}_3 = \gamma_3 = \bar{\gamma}_3 = 0$) it is $\mathbf{t}_i = \mathbf{t}_{ix} = \mathbf{t}_{ixx} = \mathbf{t}_{ixxx} = \mathbf{0}$.

In the conventional BEM, the load vectors $\mathbf{b}_i$ are known and eqns (28) are used to evaluate u_i and their derivatives at the N nodal points. This, however, can not be done here since $\mathbf{b}_i$ are unknown. For this purpose, $3N$ additional equations are derived, which permit the establishment of $\mathbf{b}_i$. These equations result by applying eqns (1)-(3) to the N collocation points, leading to the formulation of the following set of $3N$ simultaneous equations

$$\left(\mathbf{A} + P\mathbf{B}\right)\begin{Bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \mathbf{b}_3\end{Bmatrix} = \begin{Bmatrix}\mathbf{p}_Y\\ \mathbf{p}_Z\\ \mathbf{m}_x + \mathbf{p}_Z y_C - \mathbf{p}_Y z_C + \mathbf{G}\mathbf{I}_t\mathbf{t}_{3xx}\end{Bmatrix} - P\begin{Bmatrix}\mathbf{t}_{1xx} - z_C\mathbf{t}_{3xx}\\ \mathbf{t}_{2xx} + y_C\mathbf{t}_{3xx}\\ -z_C\mathbf{t}_{1xx} + y_C\mathbf{t}_{2xx} + \dfrac{I_S}{A}\mathbf{t}_{3xx}\end{Bmatrix} \tag{29}$$

In the above set of equations the matrices $\mathbf{A}$ and $\mathbf{B}$ are evaluated from the expressions

$$\mathbf{A} = \begin{bmatrix}\mathbf{E}\mathbf{I}_Z & \mathbf{E}\mathbf{I}_{YZ} & \mathbf{0}\\ \mathbf{E}\mathbf{I}_{YZ} & \mathbf{E}\mathbf{I}_Y & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{E}\mathbf{C}_S - \mathbf{G}\mathbf{I}_t\mathbf{T}_{3xx}\end{bmatrix} \tag{30}$$

$$\mathbf{B} = \begin{bmatrix}\mathbf{T}_{1xx} & \mathbf{0} & -z_C\mathbf{T}_{3xx}\\ \mathbf{0} & \mathbf{T}_{2xx} & y_C\mathbf{T}_{3xx}\\ -z_C\mathbf{T}_{1xx} & y_C\mathbf{T}_{2xx} & \dfrac{I_S}{A}\mathbf{T}_{3xx}\end{bmatrix} \tag{31}$$

where $\mathbf{E}\mathbf{I}_Y$, $\mathbf{E}\mathbf{I}_Z$, $\mathbf{E}\mathbf{I}_{YZ}$, $\mathbf{E}\mathbf{C}_S$, $\mathbf{G}\mathbf{I}_t$ are $N \times N$ diagonal matrices including the values of the EI_Y, EI_Z, EI_{YZ}, EC_S, GI_t quantities, respectively, at the N nodal points. Moreover, $\mathbf{p}_Y$, $\mathbf{p}_Z$ and $\mathbf{m}_x$ are vectors containing the values of the external loading at these points.

Solving the linear system of eqns (29) for the fictitious load distributions $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$, the displacements and their derivatives in the interior of the beam are computed using eqns (28).

Buckling equation

In this case it is $\alpha_3 = \bar{\alpha}_3 = \beta_3 = \bar{\beta}_3 = \gamma_3 = \bar{\gamma}_3 = 0$ (homogeneous boundary conditions) and $\mathbf{p}_X = \mathbf{p}_Y = \mathbf{m}_x = \mathbf{0}$. Thus, eqn (29) becomes

$$\left(\mathbf{A} + P\mathbf{B}\right)\begin{Bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \mathbf{b}_3\end{Bmatrix} = \mathbf{0} \tag{32}$$

The condition that eqn (32) has a non-trivial solution yields the buckling equation

$$det\left(\mathbf{A}+P\mathbf{B}\right)=0 \tag{33}$$

4 NUMERICAL EXAMPLES

On the basis of the analytical and numerical procedures presented in the previous sections, a computer program has been written and representative examples have been studied to demonstrate the efficiency, wherever possible the accuracy and the range of applications of the developed method.

For comparison reasons, a simply supported thin-walled beam of length $l=1.0m$ ($EI_Y=355.6kNm^2$, $EI_Z=1568.0kNm^2$, $EI_{YZ}=297.5kNm^2$, $EC_s=1.6338kNm^4$, $GI_t=1.352kNm$, $A=9.68\times10^{-5}m^2$, $I_S=4.60\times10^{-8}m^4$, $y_C=-9.09\times10^{-3}m$, $z_C=10.43\times10^{-3}m$), as shown in Fig. 3, has been studied. In the case of a simply supported beam eqns. (1)-(3) are simplified to

$$EI_Z\frac{d^2v}{dx^2}+EI_{YZ}\frac{d^2w}{dx^2}+P\left(v-z_C\theta\right)=0 \tag{34}$$

$$EI_Y\frac{d^2w}{dx^2}+EI_{YZ}\frac{d^2v}{dx^2}+P\left(w+y_C\theta\right)=0 \tag{35}$$

$$EC_S\frac{d^2\theta}{dx^2}-GI_t\theta+P\left(\frac{I_S}{A}\theta-z_Cv+y_Cw\right)=0 \tag{36}$$

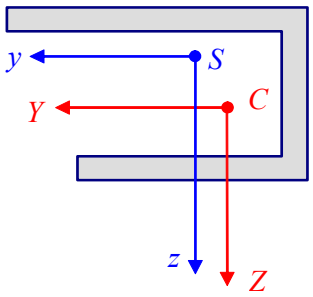


Figure 3. Cross section of the thin-walled beam.

and an analytical solution can be obtained by setting[9]

$$v=A_1\sin\frac{\pi x}{l},\qquad w=A_2\sin\frac{\pi x}{l},\qquad \theta=A_3\sin\frac{\pi x}{l} \tag{37a,b,c}$$

Inserting the above expressions and their derivatives into eqns. (34)-(36) the following homogeneous system with respect to A_1, A_2 and A_3 is obtained

$$\begin{bmatrix} P-EI_Z\frac{\pi^2}{l^2} & -EI_{YZ}\frac{\pi^2}{l^2} & -Pz_C \\ -EI_{YZ}\frac{\pi^2}{l^2} & P-EI_Y\frac{\pi^2}{l^2} & Py_C \\ -Pz_C & Py_C & P\frac{I_S}{A}-EC_S\frac{\pi^2}{l^2}-GI_t \end{bmatrix}\begin{Bmatrix} A_1 \\ A_2 \\ A_3 \end{Bmatrix}=\begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix} \tag{38}$$

For a non-trivial solution, the determinant of the above system must be equal to zero. Thus, a cubic equation is obtained, leading to three positive roots P_y, P_z, P_θ, from which the smallest one is of importance in engineering design.

In Table 1 the computed buckling loads P_y, P_z, P_θ for the aforementioned case are presented as compared with those obtained from the analytical solution, in which the bending stiffness EI_{YZ} is excluded and included in turn. From the obtained results the influence of the inclusion of the product of inertia is remarkable leading to the conclusion that the aforementioned stiffness has to be taken into account. Moreover, the accuracy of the obtained results using the proposed method is also remarkable.

	$EI_{YZ} = 0$		EI_{YZ} included	
	analytical	computed	analytical	computed
P_y	3.447	3.447	2.807	2.807
P_z	13.809	13.809	13.743	13.744
P_θ	70.267	70.269	72.934	72.937

Table 1 : Buckling load ($\times 10^3 kN$) of the beam of example 1.

5 CONCLUDING REMARKS

The main conclusions that can be drawn from this investigation are

a. The numerical technique presented in this investigation is well suited for computer aided analysis for homogeneous beams of arbitrary cross section, subjected to any linear boundary conditions and to an arbitrarily distributed or concentrated loading.
b. Accurate results are obtained using a relatively small number of beam elements.
c. The displacements as well as the stress resultants are computed at any cross-section of the beam using the respective integral representations as mathematical formulae.
d. The general character of the proposed method is verified from the fact that all basic equations are formulated with respect to an arbitrary coordinate system, which is not restricted to the principal one.
e. The developed procedure retains the advantages of a BEM solution over a pure domain discretization method since it requires only boundary discretization.

Acknowledgments

Financial support for this work provided by the "*Pithagoras: Support of Research Groups in Universities*", an EU funded project in the special managing authority of the Operational Program in Education and Initial Vocational Training.

REFERENCES

[1] Vlasov, V.Z. (1961), "*Thin-walled elastic beams*", 2nd ed. Israel program for scientific translation, Jerusalem, Israel.
[2] Chajes, W. and Winter, G. (1965). "Torsional-flexural buckling of thin-walled members", Journal of Structural Enginnering ASCE, 91, pp. 103-124.
[3] Barsoum, R. and Gallagher, R. (1970), "Finite element analysis of torsional and torsional-flexural stability problems", International Journal for Numerical Methods in Engineering, 2, pp. 335-352.
[4] Chen, W.F. and Atsutra, T. (1977). *Theory of beam columns*, McGraw-Hill, New York.
[5] Trahair, N.S. (1993), *Flexural-torsional buckling of structures*, E.N. Spon, London.
[6] Katsikadelis, J.T. (2002), "The analog equation method. A boundary-only integral equation method for nonlinear static and dynamic problems in general bodies", Theoretical and Applied Mechanics 27, pp. 13-38.
[7] Sapountzakis, E.J. and Mokos, V.G. (2003), "Warping Shear Stresses in Nonuniform Torsion by BEM", Computational Mechanics, 30, pp. 131-142.
[8] Sapountzakis, E.J. and Katsikadelis, J.T (2000), "Analysis of Plates Reinforced with Beams", Computational Mechanics, 26, pp. 66-74.
[9] Timoshenko, S.P. and Gere, J.M. (1961), *Theory of Elastic Stability*, McGraw-Hill, Tokyo.

Finite Elements

FINITE ELEMENT MODELS OF STRAIN-GRADIENT ELASTICITY: ACCURACY AND ERROR ESTIMATES

E. Amanatidou[1], A. Giannakopoulos[2] and N. Aravas[1]

[1] Department of Mechanical and Industrial Engineering
University of Thessaly
38334 Volos, GREECE

[2] Department of Civil Engineering
University of Thessaly
38334 Volos, GREECE

Keywords: Strain-gradient elasticity, Mixed FEM, Couple stress, Condition number, Error estimator.

Abstract. *In recent years, the finite element implementation of strain-gradient constitutive models has been the subject of several publications. Such models include strain-gradient elasticity, deformation type of strain-gradient plasticity and gradient plasticity. The finite element strategies depend on the selection of the kinematic variables (which have to do with the solution approach) and with the continuity requirements (which have to do with the element type). It is the purpose of this work to investigate the issue of accuracy of the finite element models of strain-gradient elasticity. Special emphasis will be given to one dimensional problems. The results are expected to be useful in assessing singular stress fields around crack-tips and concentrated forces.*

1 INTRODUCTION

In strain-gradient elasticity, microstructural lengths enter the constitutive equation through the elastic strain energy function which depends not only on the strain tensor but also on the gradients of the rotation and strain tensors. Amanatidou and Aravas (2002) presented a finite element (FE) implementation of linear elastic strain-gradient constitutive models, using a mixed finite element formulation in which the displacements and the displacement gradients are used as independent unknowns and their relationships are enforced in an integral (weak) sense. The present work is based on the work of Amanatidou and Aravas in order to address the issue of accuracy of the related FE methodology. Two particular aspects are examined, namely the condition number of the stiffness matrix and the a posteriori element error estimate. Thus, the FE computations can be assessed and strategies for mesh refinements can be established in order to improve the accuracy and efficiency of the FE models of strain-gradient elasticity. The paper starts with the variational formulation of the Type I description of the problem, Mindlin (1964). The error estimate is local and is based on an energy norm and the super-convergent patch recovery. A particular one-dimensional, 3-noded bar-element is worked-out explicitly, in order to investigate the effect of the excessive refinement on the stiffness condition number and the error distribution inside the mesh.

2 A REVIEW OF STRAIN-GRADIENT ELASTICITY THEORIES

2.1 Kinematic variables

Let $\mathbf{u}$ be the displacement field. The following quantities are defined:

$$\varepsilon_{ij} = u_{i,j} = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right) = strain \tag{1}$$

$$\tilde{\kappa}_{ijk} = u_{k,ij} = \varepsilon_{jk,i} + \varepsilon_{ki,j} - \varepsilon_{ij,k} = \tilde{\kappa}_{jik} = second\ \ gradient\ of\ \ displacement. \tag{2}$$

2.2 Constitutive equations

The strain energy density W is

$$W = \tilde{W}\left(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\kappa}}\right). \tag{3}$$

Mindlin (1964) refers to the description $W = \tilde{W}\left(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\kappa}}\right)$ as "Type I".

Using the above form of the elastic strain energy density, one defines the following quantities:

$$\bar{\sigma}_{ij} = \frac{\partial \tilde{W}}{\partial \varepsilon_{ij}} = \bar{\sigma}_{ji}, \qquad \tilde{\mu}_{ijk} = \frac{\partial \tilde{W}}{\partial \tilde{\kappa}_{ijk}} = \tilde{\mu}_{jik}, \tag{4}$$

It is worthy of note that $\bar{\sigma}_{ij}$ and $\tilde{\mu}_{ijk}$ are introduced as "conjugate" quantities to ε_{ij} and $\tilde{\kappa}_{ijk}$, and their relationship to "true" couple stress is not obvious.

The "external forces" are of the form

$$\tilde{P}_i = n_j\left(\bar{\sigma}_{ji} - \tilde{\mu}_{kji,k} - \Phi_{ji}\right) - \left[D_j - \left(D_p n_p\right) n_j\right]\left(n_k \tilde{\mu}_{kji}\right), \tag{5}$$

$$\tilde{R}_i = n_k n_j \tilde{\mu}_{jki}, \tag{6}$$

$$\tilde{E}_i = [[\ell_j n_k \tilde{\mu}_{kji}]], \tag{7}$$

where Φ_{ij} is the "body double force" per unit volume.

In the above expressions, the double brackets $[\![\ ]\!]$ indicate the jump in the value of the enclosed quantity across an edge C^α, $\mathbf{l} = \mathbf{s} \times \mathbf{n}$, where $\mathbf{s}$ is the unit vector tangent to C^α and $\mathbf{D}f = \nabla f - (Df)\mathbf{n}$ or $D_i f = f_{,i} - f_{,k} n_k n_i$.

2.3 Boundary value problem

The equilibrium equation and the appropriate boundary conditions for Type I formulation is as follows:

$$\left(\bar{\sigma}_{\mathrm{ij}} - \tilde{\mu}_{kji,k} - \Phi_{ji}\right)_{,j} + f_i = 0 \quad in \quad V. \tag{8}$$

where f_i is the "body force" per unit volume.

The true stress $\boldsymbol{\sigma}$ is related to $\bar{\boldsymbol{\sigma}}$ and $\tilde{\boldsymbol{\mu}}$ by the expressions

$$\sigma_{(ij)} = \bar{\sigma}_{ij} - \frac{1}{3}\left(\tilde{\mu}_{ijk,k} + \tilde{\mu}_{jki,k} + \tilde{\mu}_{kij,k}\right) - \Phi_{(ij)} \tag{9}$$

$$\sigma_{[ij]} = -\frac{4}{3}\tilde{\mu}_{k[ij],k} - \Phi_{[ij]} \tag{10}$$

and the couple stress $\boldsymbol{\mu}$ is

$$\mu_{ij} = \frac{4}{3}\tilde{\mu}_{ikp} e_{jkp}. \tag{11}$$

2.4 Variational formulation

A given boundary value problem in strain-gradient elasticity can be formulated by the stationary condition $\delta \Pi = 0$ of the functional (see [1] for details)

$$\begin{aligned} \Pi\left(\mathbf{u}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\sigma}}^{(2)}\right) = & \int_V W\left(\boldsymbol{\varepsilon}(\mathbf{u}), \tilde{\boldsymbol{\kappa}}(\boldsymbol{\alpha})\right) dV + \int_V \left(u_{i,j} - \alpha_{ij}\right)\tilde{\sigma}^{(2)}_{ji} dV + \\ & - \int_V f_i u_i dV - \int_{\Gamma_P} \tilde{P}_i u_i dS - \int_{\Gamma_R} \tilde{R}_i n_j \alpha_{ij} dS - \sum_\alpha \oint_{C_E^\alpha} \tilde{E}_i u_i ds + \int_S \left(D_j u_i - \alpha^t_{ij}\right) n_k \tilde{\mu}_{kji}(\boldsymbol{\alpha}) dS \end{aligned} \tag{12}$$

where $\varepsilon_{ij} = u_{(i,j)}$, $\tilde{\kappa}_{ijk} = \alpha_{k(i,j)}$, $\alpha_{ij} = u_{i,j}$, $\tilde{\mu}_{ijk} = \dfrac{\partial W}{\partial \tilde{\kappa}_{ijk}}$, $\delta u_i = 0$ on $S - \Gamma_P$ and $C^\alpha - C^\alpha_E$, and $n_k \delta \alpha_{ik} = 0$ on $S - \Gamma_R$. The stationary condition $\delta \Pi = 0$ implies the appropriate field equations and boundary conditions, the relationship $\tilde{\sigma}^{(2)}_{ij} = -\tilde{\mu}_{kij,k}$ in V and $D_j u_i = \alpha^t_{ij}$ on S. In the above functional the quantities $\tilde{\sigma}^{(2)}_{ij}$ and $n_k \tilde{\mu}_{kji}$ are Lagrange multipliers that enforce the corresponding constraints in V and on S.

The above functional can form the basis for a finite element solution, in which the nodal unknowns are $\mathbf{u}$, $\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\sigma}}^{(2)}$.

It should be mentioned that the numerical calculation of the true stresses and true couple stresses in a finite element solution is not trivial when the Type I formulation is used. In such a finite element approach one can calculate easily the stress like quantity, $\tilde{\sigma}_{ij} = \bar{\sigma}_{ij} + \tilde{\sigma}_{ij}^{(2)} = \bar{\sigma}_{ij} - \tilde{\mu}_{kij,k}$, which is different from the true stress σ_{ij} in general. In fact the true stress and the true couple stress are given using Eqs. (9)-(11) (for $\Phi_{ij} = 0$).

We conclude this section with the statement of the weak form. Find

$\mathbf{u} \in H^1$ satisfying $\mathbf{u} = \tilde{\mathbf{u}}$ on S_u and $\mathbf{u} = \tilde{\mathbf{u}}^{\alpha}$ on C_u^{α},

$\boldsymbol{\alpha} \in H^1$ satisfying $\boldsymbol{\alpha} \cdot \mathbf{n} = \tilde{\mathbf{d}}$ on $S - \Gamma_P$,

and $\tilde{\boldsymbol{\sigma}}^{(2)} \in L^2$ such that for all $\mathbf{u}^* \in H^1$ satisfying $\mathbf{u}^* = \mathbf{0}$ on S_u and C_u^{α}, for all $\boldsymbol{\alpha}^* \in H^1$ satisfying $\boldsymbol{\alpha}^* \cdot \mathbf{n} = \mathbf{0}$ on $S - \Gamma_P$, and for all $\tilde{\boldsymbol{\sigma}}^{(2)*}$, $\tilde{\mathbf{T}}^* \in L^2$,

$$\int_V \left(\bar{\sigma}_{ji} + \tilde{\sigma}_{ji}^{(2)}\right) u_{i,j}^* \, dV + \int_S \tilde{T}_{ik} \, u_{i,k}^{*t} \, dS = \int_V f_i u_i^* \, dV + \int_{\Gamma_P} \tilde{P}_i u_i^* \, dS + \sum_{\alpha} \oint_{C_E^{\alpha}} \tilde{E}_i u_i^* \, ds \,, \tag{13}$$

$$\int_V \left(-\tilde{\sigma}_{ij}^{(2)} \, \alpha_{ji}^* + \tilde{\mu}_{kij} \, \tilde{\kappa}_{kij}^*\right) dV - \int_S \tilde{T}_{ik} \, \alpha_{ik}^{*t} \, dS = \int_{\Gamma_R} \tilde{R}_i \, n_k \, \alpha_{ik}^* \, dS \,, \tag{14}$$

$$\int_V \left(u_{i,j} - \alpha_{ij}\right) \tilde{\sigma}_{ji}^{(2)*} \, dV = 0 \,, \tag{15}$$

$$\int_S \left(u_{i,k}^t - \alpha_{ik}^t\right) \tilde{T}_{ik}^* \, dS = 0 \,, \tag{16}$$

where

$$\bar{\sigma}_{ij} = \frac{\partial \tilde{W}}{\partial u_{(i,j)}} \,, \qquad \tilde{T}_{ij} = n_k \tilde{\mu}_{kji} \,, \qquad \tilde{\mu}_{ijk} = \frac{\partial \tilde{W}}{\partial \tilde{\kappa}_{ijk}} \,, \tag{17}$$

with $\tilde{W} = \tilde{W}\left(u_{(i,j)}, \tilde{\kappa}_{ijk}\left(\boldsymbol{\alpha}\right)\right)$, and $\tilde{\kappa}_{ijk}\left(\boldsymbol{\alpha}\right) = \alpha_{k(i,j)}$, $\tilde{\kappa}_{ijk}^* = \alpha_{k(i,j)}^*$. (18)

As an example we consider the special case where

$$\tilde{\mu}_{ijk} = \frac{\partial \tilde{W}}{\partial \tilde{\kappa}_{ijk}} = \frac{\ell^2}{2}\left[\lambda\left(\tilde{\kappa}_{ipp}\delta_{jk} + \tilde{\kappa}_{jpp}\delta_{ik}\right) + \mu\left(2\tilde{\kappa}_{ijk} + \tilde{\kappa}_{kji} + \tilde{\kappa}_{kij}\right)\right] \text{ and} \tag{19}$$

$$\bar{\sigma}_{ij} = 2\mu\varepsilon_{ij} + \lambda\varepsilon_{kk}\delta_{ij} \tag{20}$$

where λ and μ are the Lamé's constants and ℓ is the microstructure length.

2.5 The one-dimensional case

The static analysis of a gradient-elastic bar has been examined by Tsepoura et al (2002) and herein we give a brief account of their results. The bar is assumed to be of length L, cross section area A, elastic modulus E and microstructural length ℓ. Let $u(x)$ be the uniaxial displacement along longitudinal axis x $(0 \le x \le L)$. In the absence of body forces, static equilibrium gives

$$u'' - \ell^2 u'''' = 0 \tag{21}$$

where the prime denotes differentiation with respect to x. Assuming $u \in H^1$, the solution of (21) is

$$u(x) = c_1 e^{x/\ell} + c_2 e^{-x/\ell} + c_3 x + c_4 \tag{22}$$

where c_1, c_2, c_3, c_4 are constants to be determined by appropriate boundary conditions. In the case under consideration

$$u(0)=0\,,\quad u'(L)=\alpha(L)=\varepsilon_0 \tag{23}$$

and the generalized forces

$$A\tilde{P}(L)=AE\left(u'(L)-\ell^2u'''(L)\right)=P_0 \tag{24}$$

$$A\tilde{R}(0)=AE\left[\ell u(0)+\ell^2u''(0)\right]=0\,. \tag{25}$$

Conditions (23) and (24)-(25) give

$$c_4=0\,,\ \ c_3=P_0/(AE)\,,\ \ c_1=c_2=\frac{\ell\left(\varepsilon_0-P_0/(AE)\right)}{2\cosh\left(L/\ell\right)}. \tag{26}$$

Note that if $\varepsilon_0-P_0/(AE)$, then $c_1=c_2=0$ and the solution is the same as the classic one. The stress and the double stress are given by

$$\bar{\sigma}=Eu'\,,\ \ \tilde{\mu}=\ell^2Eu'' \tag{27}$$

and the generalized tractions are

$$\tilde{P}=\bar{\sigma}-\tilde{\mu}'\,,\quad \tilde{R}=\tilde{\mu}\,. \tag{28}$$

Finite element analysis is performed for $\varepsilon_0=0.4$ and 0.6, $P_0/(AE)=1$ and $\ell/L=1/3$. The results are shown in Fig. 1c and the relative error of $u(L)$ is 0.877×10^{-3}.

3 THE FINITE ELEMENT MODEL (I3-8)

The problem is solved by using a three-node isoparametric element shown in Figure 1a. The element of size h is of mixed type with independent isoparametric (Lagrange) interpolation of the displacement u_1, its derivative $\alpha_{11}=u_{1,1}$ and of $\tilde{\sigma}_1^{(2)}=-\tilde{\mu}_{111,1}$. The total number of unknowns per element are $8=n_u+n_\alpha+n_{\tilde{\sigma}^{(2)}}$, where $n_u=n_\alpha=3$ (3 displacements and 3 derivatives, one for every node) and $n_{\tilde{\sigma}^{(2)}}=2$ (at the edge nodes only). A three point Gauss integration is used (properly constrained, Fig. 1b) and the element has none zero eigenvalue, thus satisfying the patch test. Note that the necessary condition for convergence, $n_u+n_\alpha\geq n_{\tilde{\sigma}^{(2)}}$, is also satisfied ([7], [8]). The stiffness matrix of the element, $\left[k^e\right]$, together with the element equilibrium equation is given by the following expression.

$$\frac{A}{3}\begin{bmatrix}
7\frac{E}{h} & 0 & -2.5 & -8\frac{E}{h} & 0 & \frac{E}{h} & 0 & -0.5\\
0 & \frac{7}{9}\frac{\ell^2E}{h} & -0.5h & 0 & -\frac{8}{9}\frac{\ell^2E}{h} & 0 & \frac{1}{9}\frac{\ell^2E}{h} & 0\\
-2.5 & -0.5h & 0 & 2 & -h & 0.5 & 0 & 0\\
-8\frac{E}{h} & 0 & 2 & 16\frac{E}{h} & 0 & -8\frac{E}{h} & 0 & -2\\
0 & -\frac{8}{9}\frac{\ell^2E}{h} & -h & 0 & \frac{16}{9}\frac{\ell^2E}{h} & 0 & -\frac{8}{9}\frac{\ell^2E}{h} & -h\\
\frac{E}{h} & 0 & 0.5 & -8\frac{E}{h} & 0 & 7\frac{E}{h} & 0 & 2.5\\
0 & \frac{1}{9}\frac{\ell^2E}{h} & 0 & 0 & -\frac{8}{9}\frac{\ell^2E}{h} & 0 & \frac{7}{9}\frac{\ell^2E}{h} & -0.5h\\
0.5 & 0 & 0 & -2 & -h & 2.5 & -0.5h & 0
\end{bmatrix}
\begin{Bmatrix}
u_1^{\boxed{1}}\\ \alpha_{11}^{\boxed{1}}\\ \tilde{\sigma}_{11}^{(2)\boxed{1}}\\ u_1^{\boxed{2}}\\ \alpha_{11}^{\boxed{2}}\\ u_1^{\boxed{3}}\\ \alpha_{11}^{\boxed{3}}\\ \tilde{\sigma}_{11}^{(2)\boxed{3}}
\end{Bmatrix}=
\begin{Bmatrix}
\tilde{P}_1^{\boxed{1}}\\ \tilde{R}_1^{\boxed{1}}\\ 0\\ \tilde{P}_1^{\boxed{2}}\\ \tilde{R}_1^{\boxed{2}}\\ \tilde{P}_1^{\boxed{3}}\\ \tilde{R}_1^{\boxed{3}}\\ 0
\end{Bmatrix}$$

The resulting finite element requires only C^0 continuity of u_i, α_{ij} and $\tilde{\sigma}_{ij}^{(2)}$.

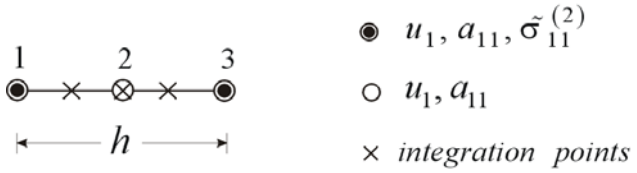


Figure 1a.

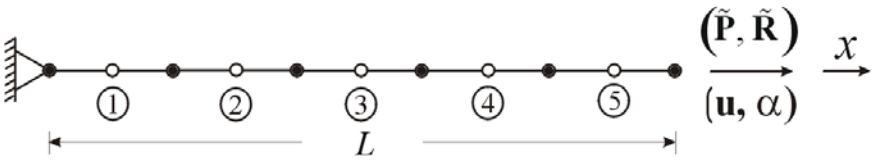


Figure 1b.

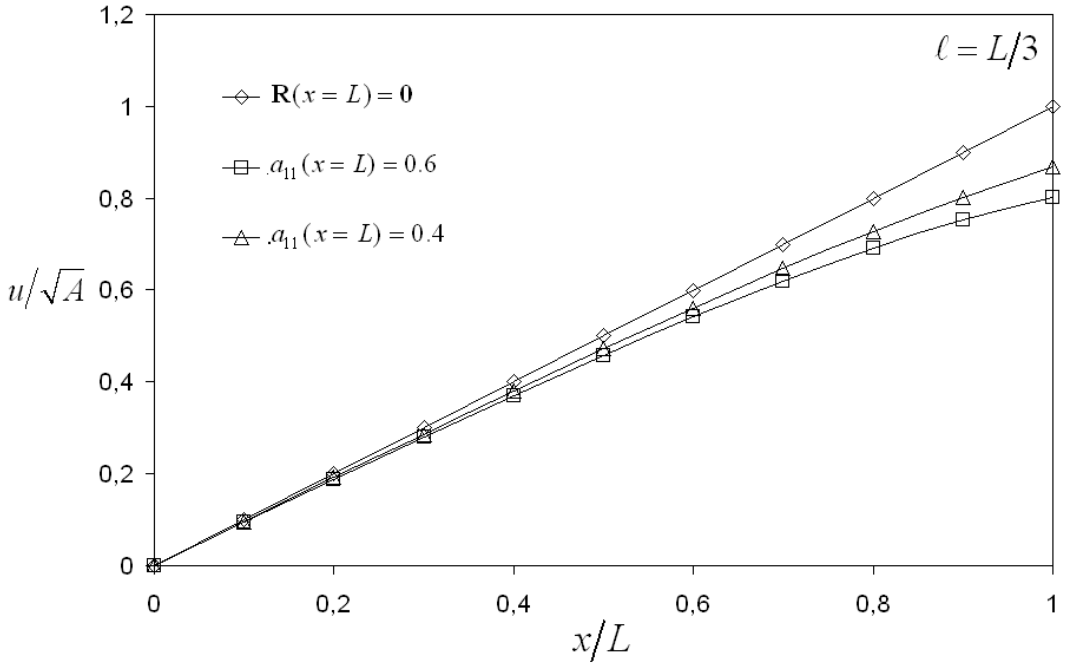


Figure 1c.

4 STIFFNESS CONDITION NUMBER

A sufficient condition for stability of the FEM analysis is the uniform boundedness of the Babuska-Brezzi (BB) constant from below, (e.g. Babuska and Narasimhan, 1977). Recent developments of apriori numerical estimates of the BB stability constant on an element level are based on the solution of the adjoin problem using the basis functions of the discrete trial space (e.g. Tsamasphyros and Markolefas, 2003).

Essentially, the requirement is to have a small condition number for the overall stiffness matrix $[K]$. Since the stiffness matrix is symmetric, the condition number β can be found from the relation

$$\beta = \max\left|\lambda_i\right|_{[K]} \max\left|\lambda_i\right|_{\left[K^{-1}\right]} \tag{29}$$

where $\max\left|\lambda_i\right|_{[K]}$ is the maximum of the absolute values of the eigenvalues of $[K]$ and $\max\left|\lambda_i\right|_{\left[K^{-1}\right]}$ is the maximum of the absolute values of the eigenvalues of $\left[K^{-1}\right]$. Figure 2 shows the development of the condition number with the number of elements. We observe that as the number of elements increases (with uniform refinement), the condition number increases and the rate of increase depends on h/ℓ (the relative size of the element h with respect to the microstructural length ℓ). It appears that the element size should not exceed the order of the microstructural length.

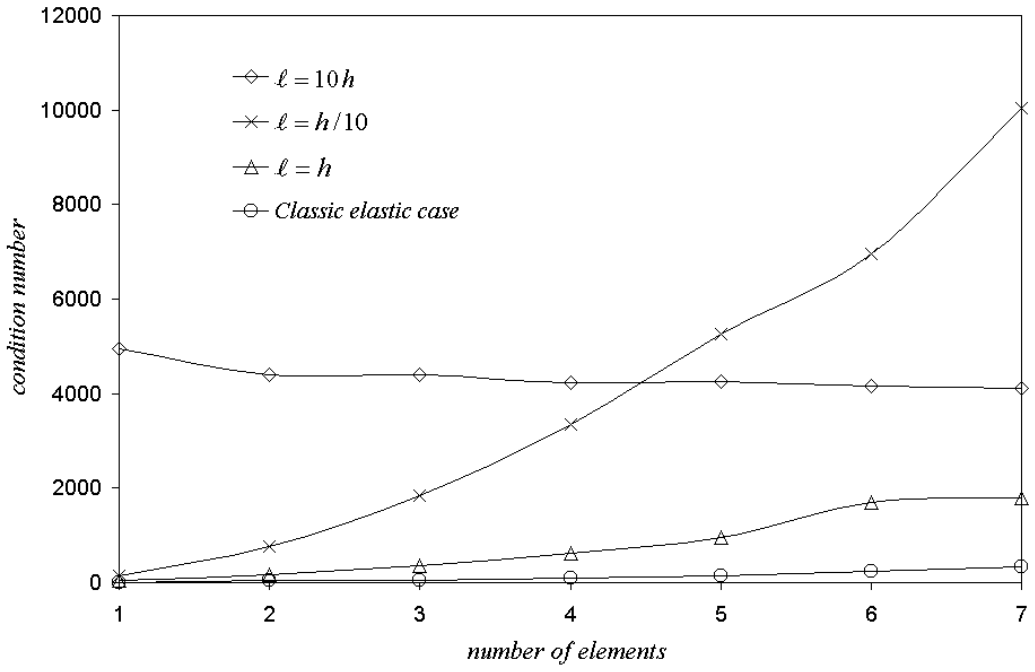


Figure 2.

5 ERROR ESTIMATE

Starting from the expression of the variation of the internal energy

$$\delta W^{\text{int}} = \int_V \left(\bar{\sigma}_{ij}\delta\varepsilon_{ij} + \tilde{\mu}_{ijk}\delta\tilde{\kappa}_{ijk}\right) dV \tag{30}$$

we formulate an error estimate for each element, based on the super-convergent patch recovery established by Zienkiewich and Zhu (1992). The error estimate is based on an (a posteriori) energy norm, assuming that $\bar{\sigma}_{ij}$ and $\tilde{\mu}_{ijk}$ are approximated very well at the integration points of the element. Assuming linear constitutive equations of the type $\tilde{\boldsymbol{\mu}} = [\mathbf{D}]\tilde{\boldsymbol{\kappa}}$ and $\bar{\boldsymbol{\sigma}} = [\mathbf{C}]\boldsymbol{\varepsilon}$, the error estimate for each element Ω^e is

$$e = \left[\int_{\Omega^e} \left\{\bar{\boldsymbol{\sigma}}^* - \bar{\boldsymbol{\sigma}}^d\right\}^T \left[C^{-1}\right]\left\{\bar{\boldsymbol{\sigma}}^* - \bar{\boldsymbol{\sigma}}^d\right\} d\Omega + \left\{\tilde{\boldsymbol{\mu}}^* - \tilde{\boldsymbol{\mu}}^d\right\}^T \left[D^{-1}\right]\left\{\tilde{\boldsymbol{\mu}}^* - \tilde{\boldsymbol{\mu}}^d\right\} d\Omega\right]^{1/2} \tag{31}$$

where $\bar{\boldsymbol{\sigma}}^d$ and $\tilde{\boldsymbol{\mu}}^d$ are obtained from FEM, and $\bar{\boldsymbol{\sigma}}^*$ and $\tilde{\boldsymbol{\mu}}^*$ are second order polynomial interpolations of the computed values at the integration points.
For the 1-D case, Eq. (31) becomes

$$e = \left[\int_0^h \frac{\left(\bar{\sigma}^* - \bar{\sigma}^d \right)^2}{E} dx + \int_0^h \frac{\left(\tilde{\mu}^* - \tilde{\mu}^d \right)^2 \ell^2}{E} dx \right]^{1/2} \tag{32}$$

where $\bar{\sigma}^*$ and $\tilde{\mu}^*$ are given by

$$\begin{aligned} \bar{\sigma}^* &= a_1 + a_2 x + a_3 x^2 \\ \tilde{\mu}^* &= b_1 + b_2 x + b_3 x^2 \end{aligned} \tag{33}$$

where $\lfloor a_1 \quad a_2 \quad a_3 \rfloor$ and $\lfloor b_1 \quad b_2 \quad b_3 \rfloor$ are computed from

$$\begin{aligned} \{a\} &= \left[A^{-1} \right] \{c\} \\ \{b\} &= \left[A^{-1} \right] \{d\} \end{aligned} \tag{34}$$

and

$$[A] = \begin{bmatrix} 3 & \sum_{i=1}^{NINT} x_i & \sum_{i=1}^{NINT} x_i^2 \\ \sum_{i=1}^{NINT} x_i & \sum_{i=1}^{NINT} x_i^2 & \sum_{i=1}^{NINT} x_i^3 \\ \sum_{i=1}^{NINT} x_i^2 & \sum_{i=1}^{NINT} x_i^3 & \sum_{i=1}^{NINT} x_i^4 \end{bmatrix} \quad \{c\} = \begin{Bmatrix} \sum_{i=1}^{NINT} \bar{\sigma}^d \left(x_i \right) \\ \sum_{i=1}^{NINT} x_i \, \bar{\sigma}^d \left(x_i \right) \\ \sum_{i=1}^{NINT} x_i^2 \bar{\sigma}^d \left(x_i \right) \end{Bmatrix}, \quad \{d\} = \begin{Bmatrix} \sum_{i=1}^{NINT} \tilde{\mu}^d \left(x_i \right) \\ \sum_{i=1}^{NINT} x_i \, \tilde{\mu}^d \left(x_i \right) \\ \sum_{i=1}^{NINT} x_i^2 \tilde{\mu}^d \left(x_i \right) \end{Bmatrix} \tag{35}$$

NINT=3 (the integration points).

Obviously, the error estimate can be used for mesh refinement strategies, of particular use for cases of bodies with crack-tips and concentrated forces.
Figure 3 shows the error distribution along bar for different (uniform) mesh densities. For the particular case, the largest error appears at the last element where non-linearity is stronger. Figure 3a shows the error distribution for $\ell = L/3$ and Figure 3b for $\ell = L/9$.

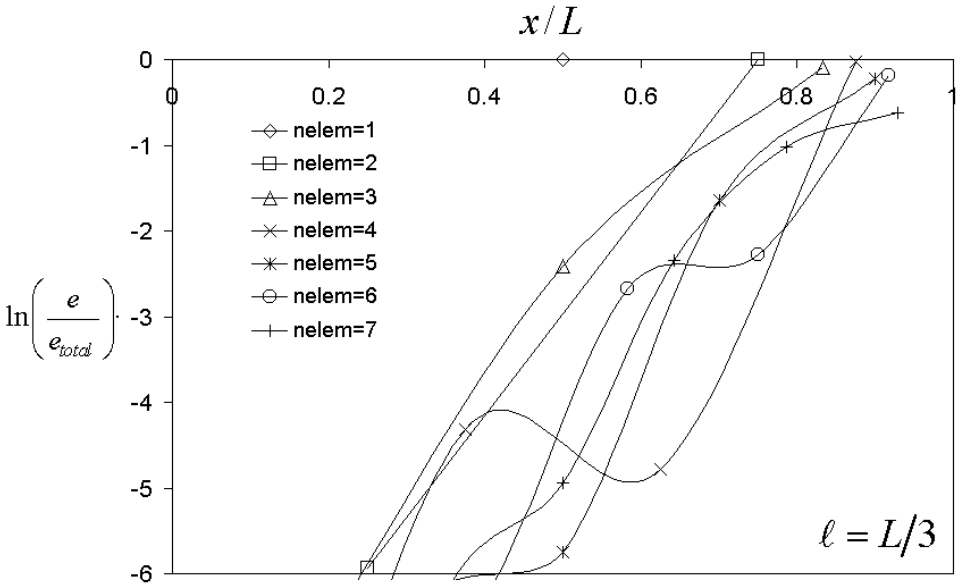


Figure 3a.

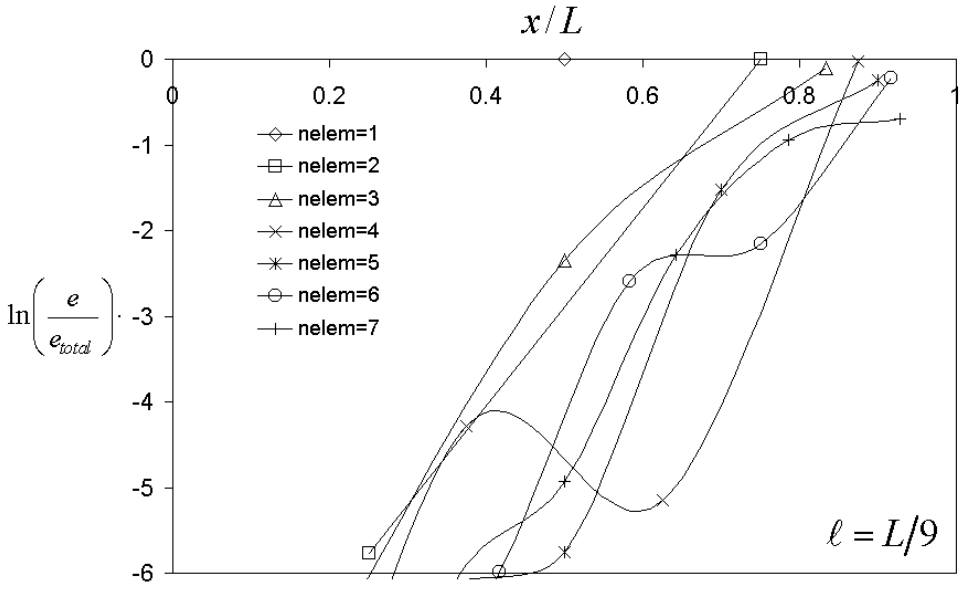


Figure 3b.

6 CONCLUSIONS

The mixed finite element formulation of strain-gradient elasticity was investigated regarding the stiffness condition number and the a posteriori element error estimate. The analysis was particularized to the one-dimensional case. For this purpose, we constructed a novel, Type I, 3-noded element. The size of the element should not be much larger than the microstructural length, in order to keep the condition number low. The element error estimate (energy norm) can serve as an index for regions that would require local mesh refinement, for example regions of high displacement curvature (near crack-tips and concentrated loads).

REFERENCES

[1] Amanatidou, E. and Aravas, N. (2002), "Mixed finite element formulations of strain-gradient elasticity problems", Comp. Meth. Appl. Mech. Engng., vol.191, pp. 1723-1751.

[2] Babuska, I. and Narasimhan, R. (1977), "The Babuska-Brezzi condition and the patch test: an example", Comp. Meth. Appl. Mech. Engng., vol.140, pp. 183-199.

[3] Mindlin, R. D. (1964), "Microstructure in linear elasticity", Arch. Rational Mech. Anal., vol.16, pp. 51-78.

[4] Tsamasphyros, G. and Markolefas, S. (2003), "An estimate of the Babuska-Brezzi inf-sup discrete stability constant to general linear Petrov-galerkin finite element formulations (an estimate of the Babuska-Brezzi stability constant)", Applied Mathematics and Computation, vol. 144, pp. 107-116.

[5] Tsepoura, K. G., Papargyri-Beskou, S., Polyzos, D. and Beskos, D. E. (2002), "Static and dynamic analysis of a gradient-elastic bar in tension", Archive of Applied Mechanics, vol.72, pp. 483-497.

[6] Zienkiewicz, O. C. and Taylor, R. L. (1997), *The Finite Element Method*, Mc Graw-Hill Book Company, London, 1991.

[7] Zienkiewicz, O. C. and Taylor, R. L. (1997), "The finite element patch test revisited. A computer test for convergence, validation and error estimates", Comp. Meth. Appl. Mech. Engng., vol.149, pp. 223-254.

[8] Zienkiewicz, O. C. and Zhu, J. Z. (1992), "The syperconvergent patch recovery and a posteriori error estimates. Part I: The recovery technique", Internat. J. Numer. Methods Engng., vol.33, pp. 1331-1364.

PERFORMANCE OF LARGE GORDON-COONS FINITE ELEMENTS IN 2-D POTENTIAL PROBLEMS

Christopher G. Provatidis

School of Mechanical Engineering
National Technical University of Athens
9 Iroon Polytechniou Str, Zografos Campus, GR-15773 Athens, Greece
e-mail: cprovat@central.ntua.gr, web page: http://users.ntua.gr/cprovat

Keywords: Transfinite elements, Gordon-Coons interpolation, Bivariate interpolation, Potential problems

Abstract. *This paper discusses several mathematical aspects of large transfinite elements that are capable of solving engineering problems. It is shown that Gordon-Coons interpolation, which is well-known in the theory of CAD surfaces, is also capable of constructing large isoparametric finite elements with boundary and/or internal nodes. Serendipity and Lagrangian type elements are distinguished members of this family. Moreover, it is possible to introduce any number of internal nodes not necessarily at the same normalized curvilinear coordinates with the boundary nodes. Although the same concept is applicable to both 2-D and 3-D Finite Element Method as well as 3-D Boundary Element Method, this paper reduces in investigating their performance in 2-D steady state and transient potential problems using the finite element method. The theory is sustained by steady state and transient thermal test cases.*

1 INTRODUCTION

It is well known that early finite elements appeared in two basic formulations, i.e. Lagrangian and Sependipity interpolations. For a unit square uniformly meshed by m segments per side, the corresponding Lagrangian element requires $(m+1)^2$ nodes as shown in Figure 1, while the serendipity one requires $4m$ nodes. For each case, the corresponding monomials in Pascal's triangle are well defined[1]. According to Zienkiewicz[1, pp.155-159], besides the well-known bilinear and quadratic elements, cubic and quartic elements were originally derived by inspection. Ergatoudis et al.[2] derived the shape function for some Lagrangian and Serendipity elements. A systematic way of generating the 'serendipity' shape functions was first introduced by Zienkiewicz et al.[3] and a simpler formulation was reported by Taylor[4]. However, no general theory for Serendipity elements was produced. Dunne[5] showed that two-dimensional shape functions can be complete bivariate polynomials for the nth degree, if the number of element nodes equals to $(n+1)(n+2)/2$. Zlámal[6] has given a geometric method for the derivation of Serendipity type elements, which has been simplified by Ball[7]. In Lagrangian type quadrilateral elements the polynomial space contains all terms that correspond to a certain discretization, while in Serendipity type elements a lot of them are missing. In more detail, let us consider the above-mentioned $(m+1)^2$ symmetrically placed nodes in a unit square (m segments per side), with the reference coordinates denoted by p and q as shown in Figure 1[8]. With respect to Pascal's triangle, Lagrangian type elements include all terms $p^i q^j$ $(i,j=0,1,\ldots,m)$ and occupy a whole square area[2,p.156] (see also Figure 2) while Serendipity type elements include the edges $(\{1,p,\ldots,p^m\}, \{1,q,\ldots,q^m\})$ with a surplus of two (i.e., $\{1,p,\ldots,p^m\}q$, $\{1,q,\ldots,q^m\}p$)[1,p.160] and occupy the legs of a lamda ($\wedge$)-shape with length "m+1" as shown in Figure 3.

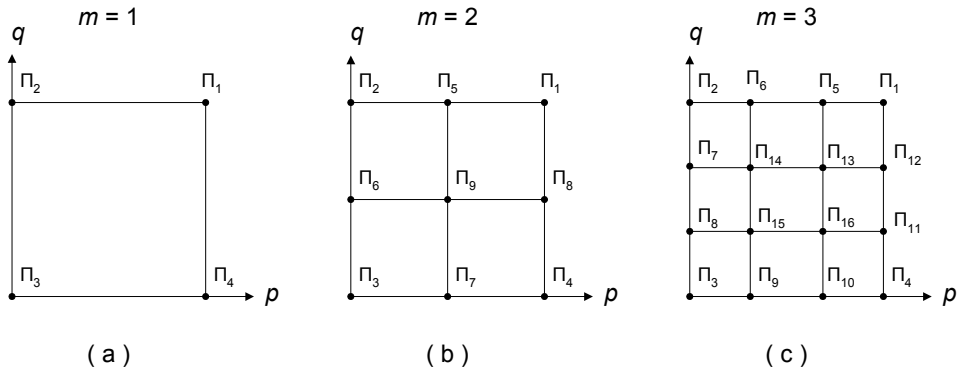


Figure 1: Unit squares of typical low-degree finite elements of Lagrangian type: (a) bilinear: m=1, (b) biquadratic: m=2 and (c) bicubic: m=3.

Figure 2 summarizes the most common Lagrangian and Serendipity type finite elements for m=1 (bilinear), m=2 (biquadratic) and m=3 (bicubic). The monomials included in the shape functions of quintic (20-node), septic (24-node) and still higher-order (4m-node) Serendipity type elements are given in Figure 3. The topic of arbitrary-noded elements has been also discussed first by Irons[9] and later by El-Zafrany[10] while the author has presented closed-form analytical expressions for the so-called Coons-Patch Macroelements (CPM)[11,12]. However, if someone is interested in deriving new elements like Serendipity ones *plus* some arbitrary interior nodes (less or more than those of the corresponding Lagrangian type element), not a general recipe seems to exist. A first approach has been attempted by Gordon and Hall[13] while the author[14] has recently successfully compared with both FEM and BEM for one internal node only.

It is the aim of this paper to present a general theory for dealing with both boundary and internal nodes. Although the theory is general, due to lack of space, here it will be tested only on 2-D elliptic and parabolic potential problems.

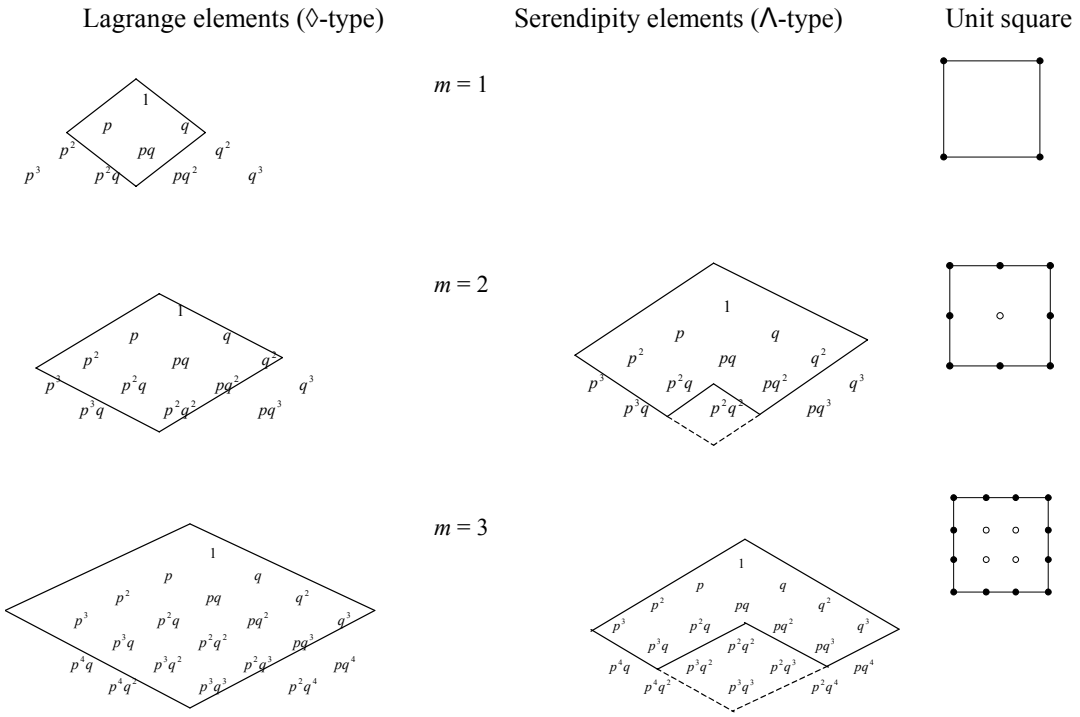


Figure 2: Pascal triangles for several conventional finite elements (m=1: bilinear, m=2: biquadratic, m=3: bicubic). Left columns correspond to Lagrange elements while middle columns to Serendipity ones. The corresponding elements are shown in the right column. Black nodes correspond to both Serendipity and Lagrange elements while white nodes to only Lagrange ones.

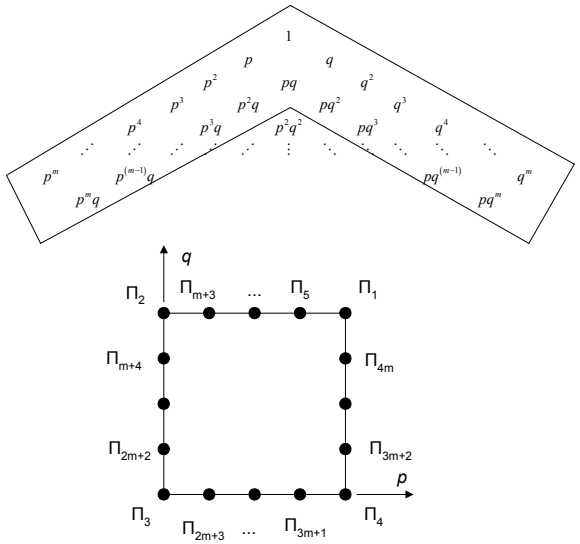


Figure 3: Unit square of a Coons-patch macroelement with (m+1) nodes per side.

2 GORDON-COONS BIVARIATE INTERPOLATION

Let us consider a two-dimensional function $u(p,q)$, defined in a square domain $\Omega=(\Pi_1\Pi_2\Pi_3\Pi_4)$, with p and q denoting normalized coordinates ($0\le p,q\le 1$), known at lines p=const. and q=const. These lines are (n_p+1) p-lines vertical to the p-axis at the points: $[\mathbf{p}]=\left[p_0,p_1,\ldots,p_{n_p}\right]$, and (n_q+1) q-lines vertical to the q-axis at the points: $[\mathbf{q}]=\left[q_0,q_1,\ldots,q_{n_q}\right]$. Then, the following functions are assumed to be known:

$$u\left(p_i,q\right),\ i=0,1,\ldots,n_p \qquad u\left(p,q_j\right),j=0,1,\ldots,n_q \tag{1}$$

Let us, further, define cardinal blending functions $E_i(p)$ for i=0,1,…, n_p (δ_{im}=Kronecker's delta):

$$E_i\left(p_m\right)=\delta_{im} \tag{2}$$

with p_i, rsp., p_m, being elements of $[\mathbf{p}]$. By analogy we define cardinal blending functions $E_j(q)$ for j=0,1,…,n_q.

Based on Eq(1) and Eq(2), we construct the following unidirectional, or lofting, operators $A_p(u)$ and $A_q(u)$:

$$A_p\left(u\right)=\sum_{i=1}^{n_p}u\left(p_i,q\right)\cdot E_i\left(p\right),\quad A_q\left(u\right)=\sum_{j=1}^{n_q}u\left(p,q_j\right)\cdot E_j\left(q\right). \tag{3}$$

and in the sequence a two-dimensional operator $A_{pq}(u)$, given by

$$A_{pq}\left(u\right)=A_pA_q\left(u\right)=\sum_{i=1}^{n_p}\sum_{j=1}^{n_q}u\left(p_i,q_j\right)\cdot E_i\left(p\right)\cdot E_j\left(q\right) \tag{4}$$

Then, the two-dimensional Gordon-Coons function $u(p,q)$ is interpolated by the following formula[13]:

$$u\left(p,q\right)=A_p\left(u\right)+A_q\left(u\right)-A_{pq}\left(u\right) \tag{5}$$

If for the interpolation of Eq(5) the values of u vertical to the p-axis are not available, then the unidirectional operator $A_p(u)$ equals to zero, as well as all the subsequent operators with the p-index. The same holds for the values of u vertical to the q-axis. Based on Eq(5), it is very easy to derive any of the well-known isoparametric elements. As an example, the case of a 5-node element (Figure A1) is examined in the Appendix A. In the context of this paper, it was found that both Serendipity and Lagrangian type elements belong to this family. The use of Lagrangian interpolation along the four sides of the patch $=(\Pi_1\Pi_2\Pi_3\Pi_4)$ constitutes the Model-3 of this study. Alternatively, we can use cubic B-splines (Model-1) or piecewise-linear interpolation (Model-2). Moreover, the same procedure can be applied to arbitrary-nodded elements with any number of internal nodes.

In all cases, Eq(5) leads to the well-known expression $u\left(x,y,t\right)=\sum_{i=1}^{q_e}N_j\left(x,y\right)u_j\left(t\right)$ while for the parabolic problem it holds that $[C]\{\dot{u}(t)\}+[K]\{u(t)\}=f(t)$, where q_e denotes the total number of nodes. For any node apart from the corners, the global shape functions are products of univariate basis functions, B, and polynomial blending functions, E [11,12].

3 APPLICATIONS

In order to demonstrate the efficiency of the Gordon-Coons macroelements, three model problems with known analytical solution are presented. The first refers to a steady state Poisson problem, while the second two refer to parabolic (transient heat problems) in which the Crank-Nicolson (CN) formulation has been used for both the proposed CPM, the conventional FEM and the FDM, for the same number of boundary nodes.

Problem 1: A square domain governed by Poisson's equation

The governing equation and the domain are given as:

$$\frac{\partial^2u}{\partial x^2}+\frac{\partial^2u}{\partial y^2}-2=0\quad\left(-\frac{1}{2}\pi<x,y<\frac{1}{2}\pi\right) \tag{6}$$

subject to

$$u\left(x,\pm\frac{1}{2}\pi\right)=0\quad\left(|x|\le\frac{1}{2}\pi\right)\qquad\text{and}\qquad u\left(\pm\frac{1}{2}\pi,y\right)=0\quad\left(|y|\le\frac{1}{2}\pi\right) \tag{7}$$

The theoretical solution is given by [8, p.85]:

$$u(x,y)=-\frac{\pi^2}{4}+x^2+\frac{8}{\pi}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{(2k-1)^3}\cdot\frac{\cosh(2k-1)y}{\cosh(2k-1)\frac{1}{2}\pi}\cdot\cos(2k-1)x. \tag{8}$$

Use of "bubble" functions. Due to the homogeneous Dirichlet boundary conditions, this problem is first solved using the above-mentioned blending ('bubble") functions, which for one central internal node, becomes a 2nd-degree polynomial: $E_1(p,q)=pq(1-p)(1-q)$. As it can be noticed in Table 1, this solution is more accurate than that obtained using the second alternative: $E_2\left(p,q\right)=\sin\pi p\sin\pi q$, but none of them can be really accepted as a reliable approximate solution. As shown, by progressively increasing the number of (uniformly arranged) internal nodes from 1×1 (1 node) to 3×3 (9 nodes) and 5×5 (25 nodes), the solution dramatically improves.

Table 1: Calculated potential using "bubble" functions at the center of the domain (x=y=0).

"Bubble" function	Error (in %) of the approximate CPM solution Number and arrangement of internal nodes			EXACT SOLUTION (U)
	1×1	3×3	5×5	
$E_1(p,q) = pq(1-p)(1-q)$	+6.0452	+0.0802	+0.0005	-1.454
$E_2(p,q) = \sin\pi p \sin\pi q$	+11.49	-	-	

Introduction of homogeneous Neumann B.C.: Owing to symmetry only one-fourth of the region needs to be analysed ($0 \le x,y \le \pi/2$). So, besides the Dirichlet boundary conditions along $x=\pi/2$ and $y=\pi/2$, the axes of symmetry are subject to Neumann ones $\partial u/\partial x=0$ ($x=0$) and $\partial u/\partial y=0$ ($y=0$). Each side of the square $[0,\pi/2]\times[0,\pi/2]$ is uniformly divided into m segments, with m varying so that it obtains the values m=2, 4 and 8, the last case shown in Figure 4(a). The usual linear blending functions, $(1-p)$ and p, were chosen. The errors of the calculated potential, u, at the eight nodes along the line $y=0$, shown in Figure 4a, are given in Table 2. One can there notice that none of the three Coons-patch models (Models 1, Model-3) converges in a sufficient manner. It is also interesting that in the finest Coons-patch models (m=8) the error is around 7%, which is larger even from 6.5%, which was found in Table 2 using the bubble function $E_1(p,q)$. On the contrary, for m=2, 4 and 8, the conventional FEM appears a mean average error of 5.28%, 1.24% and 0.30%, respectively[8].

Table 2: Solution quality of the calculated variable, u, using one Coons macroelement and alternative interpolations along each side. Each side is uniformly divided into m=2,4 and 8 segments.

NODE	ERRORS in % of the approximate CPM solution									EXACT SOLUTION (U)
	MODEL-1 (Cubic B-splines)			MODEL-2 (Piecewise linear)			MODEL-3 (Lagrange polynomials)			
	m=2	m=4	m=8	m=2	m=4	m=8	m=2	m=4	m=8	
1	-7.68	-7.16	-7.24	0.37	-5.38	-6.79	-8.59	-7.27	-7.31	-1.454
2	–		-3.36	–	–	-3.02	–	–	-3.65	-1.435
3	–	0.05	-0.36	–	1.20	0.04	–	-0.34	-0.20	-1.376
4	–		2.36	–	–	2.69	–	–	2.36	-1.276
5	7.19	4.34	4.75	10.54	6.18	5.12	4.79	4.64	4.48	-1.132
6	–		7.21	–	–	7.50	–	–	7.44	-0.938
7	–	10.77	9.48	–	11.02	9.97	–	9.75	9.82	-0.690
8	–		12.79	–	–	12.64	–	–	11.32	-0.379
9	Given			Given			Given			0.000

Changing the blending functions. Now, using again one macroelement only (Figure 4a), the blending functions along the x (equivalently, p)-direction were modified as $E_0(p) = \cos\pi p/2$ and $E_1(p) = 1-\cos\pi p/2$. Due to the symmetry, the same blending functions were chosen for the q-direction, too. The results are shown in the left part of Table 3 and one can notice a significant improvement in comparison with the linear blending functions.

Table 3: Solution quality of the calculated variable, u, using one or two Coons-patch macroelements, alternative interpolations along each side (Model-1, Model-2 and Model-3) as well as alternative blending functions. Each external side is uniformly divided into eight segments.

NODE (Fig.4a)	ERRORS in % of the approximate solution					EXACT SOLUTION (U)
	One macroelement (Cosine-like blending functions)			**Two-macroelements** MODEL-2 (Alternative blending functions)		
	MODEL-1	MODEL-2	MODEL-3	LINEAR	COSINE-like	
1	-0.83	-0.68	-0.86	-1.15	+0.10	-1.454
2	-0.71	-0.55	-0.92	-1.27	-0.05	-1.435
3	-0.49	-0.25	-0.36	-1.24	-0.55	-1.376
4	-0.01	+0.23	+0.00	-0.80	-1.24	-1.276
5	+0.62	+0.91	+0.39	+0.62	-1.70	-1.132
6	+1.54	+1.80	+1.75	-5.22	+0.99	-0.938
7	+2.56	+2.94	+2.81	-10.15	+2.64	-0.690
8	+4.41	+4.38	+3.26	-14.50	+2.81	-0.379
9	Given					0.000

Domain decomposition. Finally, the influence of subdividing the domain was investigated. The square region was divided in two regions, occupied by two corresponding macroelements. The first macroelement is a square

of dimensions $\pi/4\times\pi/4$ and it consists of 16 nodes. This element is defined by its corners Π_1, Π_2, Π_3 and Π_4, shown in the lower left corner of Figure 4b. The second macroelement is the L-shaped $P_1AP_2P_3BP_4$, with corner points P_1, P_2, P_3 and P_4 (in its corresponding unit square). Since the interface P_3BP_4 ($\equiv\Pi_2\Pi_1\Pi_4$) is a complete side for the second macroelement while it spans two sides of the first one, the easiest interpolation is that of hat-functions (Model-2). In this case, both the aforementioned linear and cosine-like blending functions have been tested. As it is clearly shown in the right part of Table 3, in both cases the errors significantly reduces but the most accurate solution is that obtained using the cosine-like blending function.

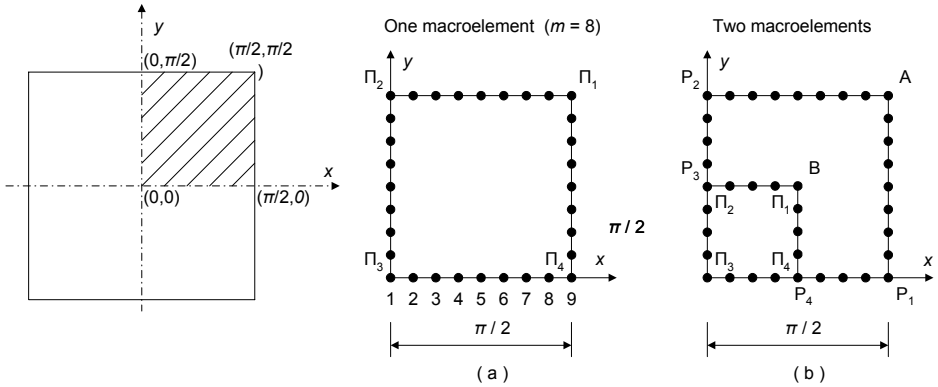


Figure 4: Test problem with homogeneous Dirichlet boundary conditions using (a) one and (b) two macroelements.

Problem 2: Transient heat analysis - Sinusoidal initial condition on a square plate

The problem is defined in a rectangular of dimensions 1.0×1.0. The boundary conditions are specified as

$$u(0,t)=0,\ u(1,t)=0 \tag{9}$$

subject to the initial condition (t=0):

$$u(x,y;0)=\sin\pi x \qquad 0\le x\le 1 \tag{10}$$

For this example, there is not any discontinuity between initial and boundary conditions at x=0,1, because in both cases the potential remains equal to zero.

The analytical solution for this problem is easily shown to be

$$u(x,y;t)=e^{-\pi^2 t}\sin\pi x \tag{11}$$

and obviously behaves well for all values of x and t.

Due to the symmetry, it is only necessary to work with one fourth of the domain shown in Figure 5, that is a square of dimensions $\Delta x\times\Delta y=0.5\times 0.5$ and modified boundary conditions as

$$u(0,y,t)=0\ ,\quad \frac{\partial u}{\partial x}(0.5,y,t)=\frac{\partial u(x,0,t)}{\partial y}=\frac{\partial u(x,0.5,t)}{\partial y}=0 \tag{12}$$

By choosing (Δx=0.1,Δt=0.005), Table 4 compares the CPM solution with the conventional FEM solution as well the Finite Difference Method (FDM). The latter was applied in three alternative formulations according to Reference [15,pp.635-638]. The first, FDM-1, is the Forward Time Centered Space (FTCS) method, a fully explicit scheme. The second, FDM-2, is a fully implicit or Backward time scheme. The third, FDM-3, is the well-known Crank-Nicholson method that consists of the mean average of FDM-1 and FDM-2 as shown in Table 5.

It is noted that the macroelement solution using linear interpolants is *identical* with the conventional FE solution, probably due the fact that this particular problem is essentially one-dimensional. It is interesting that both FEM and FDM-3 based on the Crank-Nicholson differentiation scheme are of the same quality (0.42% versus 0.40%), while the proposed CPM based on B-splines (Model-1) interpolation is superior (only 0.06%).

Table 4: Calculated temperature at time t=0.05 using several numerical techniques for Problem 2 (time step: Δt=0.005)

Coordinate X	0.1	0.2	0.3	0.4	0.5
Exact	0.18865	0.35884	0.49390	0.58062	0.61050
FEM (CN)	0.18787	0.35735	0.49185	0.57820	0.60795
CPM	**0.18843**	**0.35881**	**0.49301**	**0.58112**	**0.60989**
FDM-1	0.18709	0.35586	0.48980	0.57580	0.60543
FDM-2	0.19163	0.36450	0.50169	0.58977	0.62012
FDM-3	0.18940	0.36026	0.49586	0.58291	0.61291

FEM: Finite Element Method using the Crank-Nicholson (CN) scheme
CPM: Coons-Patch Macroelements (present methodology)
FDM-1: Finite Difference Method using a fully explicit scheme
FDM-2: Finite Difference Method using a fully implicit (backward time) scheme
FDM-3: Finite Difference Method using the Crank-Nicholson (CN) scheme.

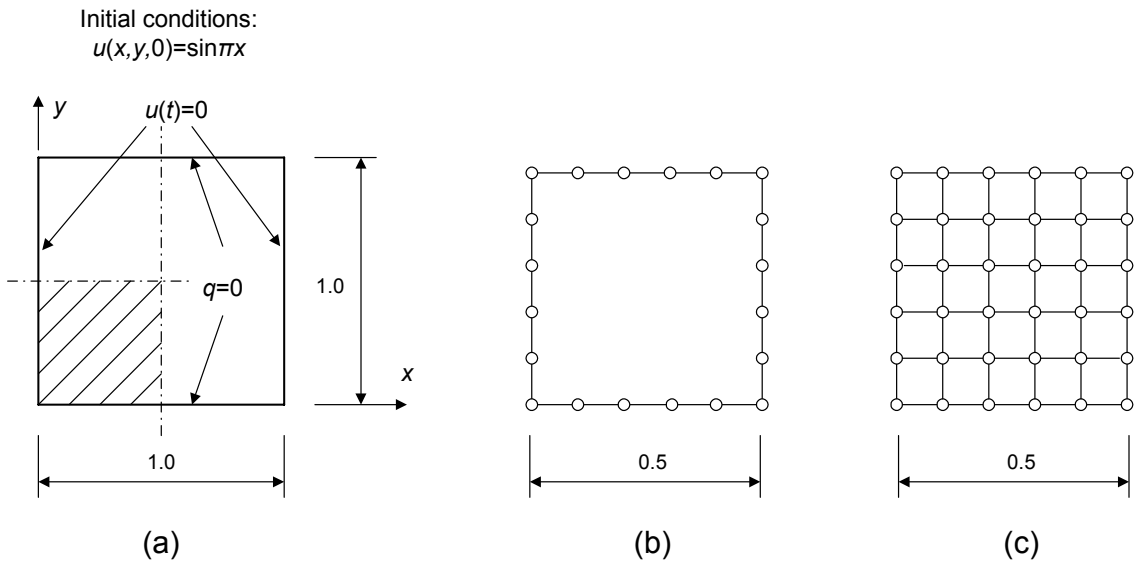


Figure 5: Problem-2: Sinusoidal initial condition on a square plate
(a) Geometry, Boundary conditions and Initial conditions
(b) Coons macroelement using 20 boundary nodes
(c) Finite element mesh (37 nodes)

Table 5: Mean average errors (in %) of the calculated temperature at nodal points along x-axis (Problem 2).

FEM	**CPM (B-splines)**	FDM-1	FDM-2	FDM-3
-0.42	**-0.06**	-0.83	+1.58	+0.40

Problem 3: Transient heat analysis - Discontinuity at the boundary of a square plate

As previously, we work with only one fourth of the domain, as shown in Figure 5, and this time the initial condition changes to

$$u(x,0)=1 \ \left(0 \le x \le 0.5,\ 0 \le y \le 0.5\right) \tag{13}$$

while the boundary conditions remain unaltered [$u(0,t)=u(1,t)=0$].
The analytical solution is given by

$$u(x,y;t)=\frac{4}{\pi}\sum_{n=0}^{\infty}\frac{1}{(2n+1)}e^{-(2n+1)^2\pi^2 t}\sin(2n+1)\pi x \tag{14}$$

Obviously, there is a discontinuity between initial and boundary conditions in this problem, which is anticipated to lead to difficulties in obtaining an accurate numerical solution near to the end points. Results are presented in Tables 6 and 7 where both FEM and macroelement CPM solutions are compared with the FDM. As previously occurred, due to the one-dimensionality of this problem, the macroelement using piecewise-linear interpolants (Model-2) coincides with the conventional FEM solution. Again, the cubic B-splines (Model-1) lead to a numerical solution of better overall quality in comparison to the linear interpolants. The FDM-3 is slightly better.

Table 6: Calculated temperature at time t=0.05 hr using several numerical techniques for Problem 3 (time step: Δt=0.005 hr)

X (m)	0.1	0.2	0.3	0.4	0.5
Exact	0.68269	0.95450	0.99730	0.99994	1.00000
FEM	0.69150	0.98615	1.01070	0.99929	0.99930
CPM	**0.70017**	**0.95650**	**0.99879**	**1.00030**	**0.99843**
FDM-1	0.68703	0.94266	0.99329	0.99947	0.99994
FDM-2	0.68129	0.93078	0.98831	0.99833	0.99959
FDM-3	0.68433	0.93654	0.99078	0.99894	0.99980

FEM: Finite Element Method using the Crank-Nicholson (CN) scheme
CPM: Coons-Patch Macroelements (present methodology): Model-1
FDM-1: Finite Difference Method using a fully explicit scheme
FDM-2: Finite Difference Method using a fully implicit (backward time) scheme
FDM-3: Finite Difference Method using the Crank-Nicholson (CN) scheme.

Table 7: Mean average errors (in %) of the calculated temperature at nodal points along *x*-axis (Problem 3).

FEM (CN)	CPM (B-splines)	FDM-1	FDM-2	FDM-3 (CN)
+1.16	+0.56	-0.21	-0.76	-0.48

CN: Crank-Nicholson
FDM1: Finite Difference Method using a fully explicit scheme
FDM2: Finite Difference Method using a fully implicit (backward time) scheme.
FDM-3: Finite Difference Method using the Crank-Nicholson (CN) scheme.

4 CONCLUSIONS

The generalized Gordons-Coons interpolation, which has been extensively used for describing CAD surfaces in automotive industry, is also capable of developing macroelements that "sweep" the whole spectrum between the well-known Serendipity and Lagrangian type finite elements. Boundary-only macroelements have been previously called Coons Parch Macroelements (CPM) and this name has been also used here. Moreover, any number of internal nodes can be introduced, not necessarily related to the position of the boundary nodes. The functional space spanned by the proposed macroelements has been rigorously defined with respect to Pascal's triangle. It has been shown that the weakness of classical Lagrangian elements due to numerical oscillations for a large number of nodes can be overcome by using piecewise-linear or cubic B-splines. Here, the performance of CPM was tested on a typical elliptic Poisson's problem and two parabolic problems but is has been previously successfully applied in hyperbolic problems[12]. It was found that CPM is applicable to heat transient problems only when used in conjunction with *consistent* capacity ("mass") matrix. The methodology can be extended to three-dimensional domains where two preliminary papers have recently appeared [16,17].

REFERENCES

[1] Zienkiewicz, O.C. (1977), *The Finite Element Method*, McGraw-Hill, London.
[2] Ergatoudis I., Irons B.M., Zienkiewicz O.C. (1968), "Curved isoparametric quadrilateral elements for finite element analysis," Int. J. Solids & Structures 4, pp. 31-42.
[3] Zienkiewicz O.C., Irons B.M., Campell J., Scott F.C. (1970), "Three dimensional stress analysis," Int Un Th Appl Mech Syp High Speed Comput. Elasticity, Liége, 1970.
[4] Taylor, R.L. (1972), "On completeness of shape functions for finite element analysis," Int. J. Numer. Meth. Engng 4, pp. 17-22.
[5] Dunne, P.C. (1968), "Complete polynomial displacement fields for the finite element method," Trans. Roy. Aero. Soc. 72, pp. 245-246.
[6] Zlámal, M. (1973), "A remark on the serendipity family," Int. J. Numer. Meth. Engng 7, pp. 98-100.
[7] Ball, A.A. (1980), "The interpolation function of a general serendipity rectangular element," Int. J. Numer. Meth. Engng 15, pp. 773-778.
[8] Wait, R. and Mitchell, A.R. (1985), *Finite Element Analysis and Applications*, John Wiley & Sons, Chichester.
[9] Irons, B.M. (1966), "Engineering application of numerical integration in stiffness method,", AIAA J 14(11), pp. 2035-2037.
[10] El-Zafrany, A. and Cookson, R.A. (1986), "Derivation of Lagrangian and Hermitian shape functions for quadrilateral elements," Int J Numer Meth Engng 23, pp. 1939-1958.
[11] Provatidis, C.G. (2003), "Analysis of axisymmetric structures using Coons' interpolation," Finite Elements in Analysis and Design 39, pp. 535-558.
[12] Provatidis, C.G. (2004), "Coons-patch macroelements in two-dimensional eigenvalue and scalar wave propagation problems," Computers & Structures 82, pp. 383-395.
[13] Gordon, W.J. and Hall, C.A. (1973), "Transfinite element methods blending function interpolation over arbitrary curved element domains," Numerische Mathematik 21, pp. 109-129.
[14] Provatidis, C.G. (2004), "Solution of two-dimensional Poisson problems in quadrilateral domains using transfinite Coons interpolation," Communications in Numerical Methods in Engineering 20, pp. 521-533.
[15] Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1986), *Numerical Recipes*, Cambridge University Press, Cambridge.
[16] Provatidis, C.G. (2005), "Three-dimensional Coons macroelements in Laplace and acoustic problems," Computers & Structures (in press).
[17] Provatidis, C.G. (2005), "Analysis of boxlike structures using 3-D Coons' interpolation," Communications in Numerical Methods in Engineering (in press).
[18] Bathe, K.J. (1982), *Finite Element Procedures in Engineering Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey.

APPENDIX A: Application of Coons transfinite formula to derive the 5-node isoparametric element

Let us consider a five-node element as shown in Figure A1. The transfinite Coons interpolation becomes:

$$u(p,q) = A_p(u) + A_q(u) - A_{pq}(u) \tag{A1}$$

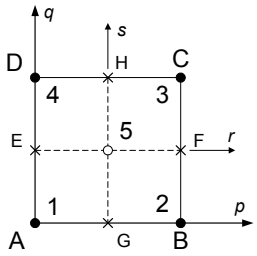


Figure A1: Five-node isoparametric element

In Equation (A1), the three lofting operators are given by

$$\begin{aligned}
A_p(u) &= E_0(p)u(0,q) + E_{1/2}(p)u(1/2,q) + E_1(p)u(1,q) \\
A_q(u) &= E_0(q)u(p,0) + E_{1/2}(q)u(p,1/2) + E_1(q)u(p,1) \\
A_{pq}(u) &= E_0(p)E_0(q)u(0,0) + E_1(p)E_0(q)u(1,0) + E_1(p)E_1(q)u(1,1) \\
&\quad + E_0(p)E_1(q)u(0,1) + E_{1/2}(p)E_0(q)u_G + E_1(p)E_I(q)u_F \\
&\quad + E_{1/2}(p)E_1(q)u_H + E_0(p)E_{1/2}(q)u_E + E_{1/2}(p)E_{1/2}(q)u_I
\end{aligned} \tag{A2}$$

where the blending functions are given as

$$\begin{aligned}
&E_0(p) = (2p-1)(p-1),\ E_{1/2}(p) = 4p(p-1),\ E_1(p) = p(2p-1) \\
&E_0(q) = (2q-1)(q-1),\ E_{1/2}(q) = 4q(q-1),\ E_1(q) = q(2q-1)
\end{aligned} \tag{A3}$$

We consider that along each side (AB, BD, CD and DA) u is interpolated using 1st degree Lagrange polynomials:

$$L_0(p) = 1-p,\ L_1(p) = p \qquad \text{(similarly for } q\text{)} \tag{A4}$$

as follows:

$$\begin{aligned}
u(0,q) &= L_0(q)u_1 + L_1(q)u_4 \\
u(1/2,q) &= E_0(q)u_G + E_I(q)u_5 + E_1(q)u_H \\
u(1,q) &= L_0(q)u_2 + L_1(q)u_3 \\
u(p,0) &= L_0(p)u_1 + L_1(p)u_2 \\
u(p,1/2) &= E_0(p)u_E + E_I(p)u_5 + E_1(p)u_F \\
u(p,1) &= L_0(p)u_4 + L_1(p)u_3
\end{aligned} \tag{A5}$$

By substituting Eq.(A5) into Eq(A2) and further into Eq(A1), one obtains:

$$\begin{aligned}
u(p,q) &= u_1(0,0)\cdot\left[L_0(p)E_0(q) + E_0(p)L_0(q) - E_0(p)E_0(q)\right] \\
&\quad + u_2(1,0)\cdot\left[L_1(p)E_0(q) + E_1(p)L_0(q) - E_1(p)E_0(q)\right] \\
&\quad + u_3(1,1)\cdot\left[L_1(p)E_1(q) + E_1(p)L_1(q) - E_1(p)E_1(q)\right] \\
&\quad + u_4(0,1)\cdot\left[L_0(p)E_1(q) + E_0(p)L_1(q) - E_0(p)E_1(q)\right] \\
&\quad + u_5(1/2,1/2)\cdot\left[E_{1/2}(p)E_{1/2}(q)\right]
\end{aligned} \tag{A6}$$

Obviously, the quantities in the brackets denote the shape functions obtained through transfinite Coons interpolation. According to Bathe[18,p.200], the conventional shape functions are given by the intuitive formulas:

$$\begin{aligned}
N_1(r,s) &= h_1 - \frac{1}{4}h_5 = \frac{(1-r)(1-s)}{4} - \frac{1}{4}(1-r^2)(1-s^2) \\
N_2(r,s) &= h_2 - \frac{1}{4}h_5 = \frac{(1+r)(1-s)}{4} - \frac{1}{4}(1-r^2)(1-s^2) \\
N_3(r,s) &= h_3 - \frac{1}{4}h_5 = \frac{(1+r)(1+s)}{4} - \frac{1}{4}(1-r^2)(1-s^2) \\
N_4(r,s) &= h_4 - \frac{1}{4}h_5 = \frac{(1-r)(1+s)}{4} - \frac{1}{4}(1-r^2)(1-s^2) \\
N_5(r,s) &= h_5 = \frac{1}{4}(1-r^2)(1-s^2)
\end{aligned} \tag{A7}$$

where h_1, h_2, h_3 and h_4 denote the shape functions of the 4-node (bilinear) element and h_5 the "bubble" function ¼(1-r^2)(1-s^2). Using the transformation $r = 2p-1,\ s = 2q-1\ \ (-1 \le r,s \le 1,\ 0 \le p,q \le 1)$, it is trivial to prove that the shape functions derived in (A6) through transfinite Gordon-Coons interpolation coincide with those of the well-known finite element (A7).

SUCTION CAISSONS: FINITE ELEMENT MODELING

Dilip R. Maniar[1], L.F. Gonzalo Vásquez[2], and John L. Tassoulas[3]

[1]Stress Engineering Services, Inc.
[2]Ensoft, Inc.
[3]Department of Civil, Architectural and Environmental Engineering
The University of Texas
Austin, Texas, 78712-1076
U.S.A.

Keywords: suction, caisson, pile, porous medium, finite element, offshore.

Abstract. *This paper presents an overview of efforts at the Offshore Technology Research Center of The University of Texas at Austin toward development and validation of a computational procedure suitable for simulations of suction caisson behavior under axial and lateral loads considering the effects of installation into clayey soil by self weight and suction. The soil is treated as a two-phase medium: a water-filled porous solid. Nonlinear behavior of the solid phase (soil skeleton) is described by means of a bounding-surface plasticity model. Caisson-soil interaction is described by a frictional contact algorithm based on a slide-line formulation that allows large relative displacement between the caisson and the soil. In addition, a remeshing tool eliminates the need for a priori specification of the caisson penetration path. Specifically, as installation of the caisson progresses, the finite-element mesh is adjusted so that the nodes below the caisson tip remain along a straight line in the axial direction. A brief account of the computational procedure along with simulations of caisson installation, reconsolidation of the soil-skeleton and caisson pullout are provided. The computational results are compared with measurements from laboratory tests also conducted at the Offshore Technology Research Center.*

1 INTRODUCTION

A suction caisson is a hollow cylinder capped at the top. It is allowed to penetrate the seafloor bottom sediments under its own weight, and then pushed to the required depth with differential pressure applied by pumping water out of the caisson interior. The use of suction caissons as foundations for deep-water offshore structures and anchors for mooring lines has been increasing in the last decade. Suction caissons are an attractive option with regard to providing anchorage for floating structures in deep water as they offer a number of advantages in that environment. They are easier to install than impact driven piles and can be used in water depths well beyond where pile driving becomes infeasible. Suction caissons have higher load capacities than drag embedment anchors and can be inserted reliably at pre-selected locations and depths with minimum disturbance to the seafloor environment and adjacent facilities[31] (Sparrevik 2001).

Better and reliable understanding of suction caisson behavior has been sought by means of field tests, laboratory tests, and numerical simulations. Extensive field tests on small-scale and full-scale caissons have been carried out to determine their installation characteristics and their axial and lateral load capacities[19, 35, 34], e.g. Hogervorst (1980), Tjelta et al. (1986), and Tjelta (1995). Field tests are valuable in obtaining geotechnical information relevant in the design of future caissons, but they are expensive and time-consuming. On the other hand, laboratory testing of model suction caissons can be employed to investigate performance of the caissons under a variety of conditions. Geotechnical centrifuge tests on model suction caissons have been carried out to simulate the stress conditions and soil response at the field scale[6, 29] (see Clukey et al. 1995, Randolph et al. 1998). These are quite costly and remain subject to various limitations. Model suction caissons have been tested under 1-*g* and controlled laboratory conditions[37, 32, 30, 13, 15, 38, 5] (Wang et al. 1977, Steensen-Bach 1992, Rao et al. 1997, El-Gharbawy and Olson 1999, El-Gharbawy et al. 1999, Whittle et al. 1998, Byrne and Houlsby 2002). The caissons studied were of aspect ratio (length-to-diameter ratio) in the range of 2-12 and were tested under various loading conditions. Laboratory tests on model suction caissons conducted by Wang et al. (1977) were focused on studying caisson efficiency and feasibility and identifying important parameters governing their performance[37]. The recent laboratory tests[30, 13] (Rao et al. 1997, El-Gharbawy and Olson 1999) were focused on improving the design methodology.

Studies of suction caisson behavior involving extensive axisymmetric and three-dimensional numerical simulations (Sukumaran et al. 1999, Erbrich and Tjelta 1999, El-Gharbawy and Olson 2000, Deng and Carter 2002) have been carried out[33, 16, 14, 12] to determine their capacity under different loading and drainage conditions. Sukumaran et al. (1999) and Erbrich and Tjelta (1999) used the commercial finite element code ABAQUS

(www.hks.com), El-Gharbawy and Olson (2000) used the commercial finite element code PLAXIS (http://www.plaxis.nl) developed for geotechnical computations, and Deng and Carter (2002) used the finite element software AFENA developed at the Center for Geotechnical Research at the University of Sydney (http://www.civil.usyd.edu.au/cgr). In all cases, the stress-strain behavior of the soil skeleton was represented by means of plasticity models such as the modified cam-clay model. The suction caisson was wished in place, with no attempt to simulate the installation process. Perfect interface bonding was assumed between the caisson and the surrounding soil skeleton. The initial state of stress in the soil skeleton was typically estimated[12] in terms of the submerged unit weight and the lateral earth pressure coefficient at rest (Deng and Carter 2002).

The computational procedure developed[36, 22] in the course of the study (Vásquez 2000, Maniar 2004) reported herein simulates suction-caisson installation and estimates the axial and lateral capacities. An axisymmetric formulation was implemented in a computer code for analysis of installation and axial-pullout problems. On the other hand, a three-dimensional analysis formulation that utilizes the general-purpose finite-element analysis code ABAQUS (http://www.hks.com) but imports the state of the state of the soil-caisson system from axisymmetric installation computations was used in lateral-pullout analysis. The soil is modeled with water-saturated porous solid finite elements and the caisson is discretized using (impermeable) solid finite elements. Nonlinear soil behavior is taken into account by means of a bounding-surface plasticity model. A frictional-contact algorithm based on a slide-line formulation is used in representing soil-caisson interface behavior. Various remeshing tools are developed to eliminate the need for a priori specification of the caisson penetration path and to avoid use of excessively distorted finite-elements along the caisson-soil interfaces. Using the formulation developed, numerical results are obtained from simulations of caisson installation, reconsolidation of the surrounding soil, and caisson pullout. The computed behavior is compared with measurements from laboratory tests[24, 21, 7, 8] conducted at The University of Texas at Austin (Mecham 2001, Luke 2002, Coffman 2003, Coffman et al. 2004).

2 GOVERNING EQUATIONS

The behavior of the saturated, porous, clayey soil is described using a mixture theory[2, 3, 1, 4, 27, 28] (Biot 1941, 1955, Atkin and Craine 1976, Bowen 1976, Prevost 1980, 1981) that accounts for coupling between the soil-skeleton deformation and the pore-fluid motion. The saturated soil is thus modeled as a two-phase medium composed of solid (soil skeleton) and pore-fluid (water) phases.

Summarized below are the balance laws governing the interaction or coupling between the soil-skeleton deformation and the pore-fluid motion. The differential equations are expressed in terms of solid displacements, Darcy's velocities, and the excess pore-fluid pressure.

2.1 Conservation of Mass

Assuming incompressibility and homogeneity of the soil particles that form the skeleton, the law of conservation of mass of the mixture can be expressed as:

$$div(\mathbf{V}^S) + div(\mathbf{V}^R) - \frac{n_w}{\lambda_w}\frac{\partial(p^w)}{\partial t} - \frac{n_w\gamma_w}{\lambda_w}[\mathbf{1}_z]^T\mathbf{V}^S - \frac{\gamma_w}{\lambda_w}[\mathbf{1}_z]^T\mathbf{V}^R - \frac{1}{\lambda_w}[grad(p^w)]^T\mathbf{V}^R = 0 \tag{1}$$

where: $\mathbf{V}^S$ is the velocity of the solid phase; $\mathbf{V}^R$ is Darcy's velocity, defined as the relative velocity of the fluid phase with respect to the solid phase multiplied by the porosity, n_w; p^w is the excess pore-fluid pressure, λ_w is the bulk modulus of the fluid, γ_w is the specific weight of the fluid, and $\mathbf{1}_z$ is the unit vector in the vertical direction. The spatial divergence and spatial gradient operators are denoted by *div* and *grad,* respectively.

2.2 Conservation of Linear Momentum

The total Cauchy stress tensor, $\boldsymbol{\sigma}$, can be written as the sum of the *effective* stress tensor, $\boldsymbol{\sigma}^{eff}$, and the excess pore-fluid pressure, p^w, as:

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}^{eff} + p^w\mathbf{I} \tag{2}$$

The conservation of linear momentum of the fluid phase can be written as:

$$-\rho_w\mathbf{a}^S - \frac{\rho_w}{n_w}\mathbf{a}^R + grad\left(p^w\right) - \frac{\rho_w}{n_w}\mathbf{k}^{-1}\mathbf{V}^R = \mathbf{0} \tag{3}$$

where: ρ_w is the mass density of the fluid phase; $\mathbf{a}^S$ is the acceleration of the solid phase; $\mathbf{a}^R$ is the relative acceleration between the fluid and solid phases, and $\mathbf{k}$ is the permeability matrix of the soil skeleton. The conservation of linear momentum of the mixture can be written as:

$$\left[\rho_S(1-n_w)+\rho_w n_w\right](\mathbf{b}-\mathbf{a}^S)+\rho_w(\mathbf{b}-\mathbf{a}^R)+div(\boldsymbol{\sigma}^{eff})+grad(p^w) \tag{4}$$

where: ρ_S is the mass density of the solid phase, and $\mathbf{b}$ is the body force per unit volume of the solid phase.

3 FINITE-ELEMENT DISCRETIZATION

The axisymmetric discretization of soil-skeleton is accomplished with eight-node, quadratic, isoparametric, underintegrated finite elements for solid displacements, and Darcy's velocities, with interpolation functions $\mathbf{N}_S$ and $\mathbf{N}_R$, respectively. Spatially continuous discretization of excess pore-fluid pressure is applied using four-node bilinear finite elements with interpolation functions $\mathbf{N}_P$.

Following standard variational arguments, the weak statements corresponding to the governing differential equations can be obtained:

$$\begin{aligned}
&-\int_{\Gamma_q}(\mathbf{N}_P)^T q dA+\int_{\Omega}\left[grad(\mathbf{N}_P)\right]^T\mathbf{V}^R J dV-\int_{\Omega}(\mathbf{N}_P)^T div(\mathbf{V}^S)J dV\\
&+\int_{\Omega}(\mathbf{N}_P)^T\frac{n_w J}{\lambda_w}\frac{\partial(p^w)}{\partial t}dv+\int_{\Omega}(\mathbf{N}_P)^T\frac{J}{\lambda_w}\left[grad(p^w)\right]^T\mathbf{V}^R dV\\
&+\int_{\Omega}(\mathbf{N}_P)^T\frac{n_w\gamma_w J}{\lambda_w}(\mathbf{1}_z)^T\mathbf{V}^S dV+\int_{\Omega}(\mathbf{N}_P)^T\frac{n_w J}{\lambda_w}(\mathbf{1}_z)^T\mathbf{V}^R dV=0
\end{aligned} \tag{5}$$

$$\begin{aligned}
&\int_{\Gamma_R}(\mathbf{N}_R)^T\mathbf{T}^w dA-\int_{\Omega}\left[div(\mathbf{N}_R)\right]^T(p^w)J dV-\int_{\Omega}(\mathbf{N}_R)^T\rho_w J\mathbf{a}^S dV\\
&-\int_{\Omega}(\mathbf{N}_R)^T\frac{\rho_w J}{n_w}\mathbf{a}^R dV-\int_{\Omega}(\mathbf{N}_R)^T\frac{\rho_w J}{n_w}\mathbf{k}^{-1}\mathbf{V}^R dV=\mathbf{0}
\end{aligned} \tag{6}$$

$$\begin{aligned}
&\int_{\Gamma_S}(\mathbf{N}_S)^T\mathbf{T}^{Tot}dA-\int_{\Omega}\left[grad(\mathbf{N}_S\delta\mathbf{U}_S)\right]^T\boldsymbol{\sigma}^{eff}J dV\\
&-\int_{\Omega}\left[div(\mathbf{N}_S)\right]^T p^w J dV-\int_{\Omega}(\mathbf{N}_S)^T(\rho_S-\rho_w)(1-n_0)\mathbf{b}dV\\
&-\int_{\Omega}(\mathbf{N}_S)^T\left[\rho_S(1-n_0)+\rho_w n_w J\right]\mathbf{a}^R dV=\mathbf{0}
\end{aligned} \tag{7}$$

The tangent stiffness matrix required for Newton iterations can be derived using consistent linearization of the weak statements. In Eq. 7, the first term is the virtual work of the surface traction due to total stresses, which, within the finite element framework, leads to the (consistent) definition of the equivalent total force. Similarly, the first term in Eq. 6 is the virtual work of the surface tractions due to excess pore-fluid pressure, giving rise to the equivalent excess pore-fluid force. Therefore, the virtual work of the surface tractions due to effective stresses, or the equivalent effective force, can be obtained as the difference between the former and the later. This way of extracting the forces due to effective traction is possible because of the way the governing equations are crafted and the fact that the interpolation functions adopted for the solid displacements and Darcy's velocities are identical. The contact formulation described below makes use of this fact in order to estimate the equivalent frictional force as a function of the equivalent effective normal force.

The caisson is represented using conventional, axisymmetric, solid finite elements: eight-node, quadratic, isoparametric elements for displacements.

4 SOIL CONSTITUTIVE MODEL

A bounding-surface plasticity model[10, 9, 11, 20] for isotropic cohesive soils (Dafalias and Herrmann 1982, Dafalias 1986, Dafalias and Herrmann 1986, Kaliakin and Herrmann 1991) is used for the description of nonlinear clayey soil behavior in terms of strain and effective stress increments. The bounding surface is a

versatile concept for representation of clay behavior along arbitrary stress and strain paths.

5 SOIL-CAISSON INTERFACES

The interior and exterior soil-caisson interfaces are modeled with a contact algorithm based on a slide-line formulation[17] (Hallquist et al. 1985), which allows for large relative displacements between the caisson and the soil. The slide-line formulation involves nodes on the soil side of the interface and surface elements on the caisson side.

In the contact algorithm, penetration of soil nodes into the caisson is prevented with constraints imposed on the solid displacement, Darcy's velocity, and the excess pore-fluid pressure using Lagrange multipliers. Friction between the soil and the caisson is assumed to obey the classical Coulomb law. "Stick" and "slip" conditions are distinguished on the basis of the level of interface frictional force in comparison with the Coulomb force, which is taken equal to the effective compressive (normal) force multiplied by the soil-caisson interface friction coefficient. Therefore, for a slave node that is in contact with a surface element, contact contributions arising from constraining solid displacement, Darcy's velocity, and excess pore-fluid pressure and contribution from the frictional interface must be added in the tangent stiffness matrix during Newton iterations.

The slide-line contact formulation is developed in terms of effective forces along the interface, which are integrals of the effective traction along the interface. By the arrangement of the governing differential equations and the corresponding weak statements, it is straightforward to extract these effective forces along the interface.

6 POTENTIAL FLOW

During installation of the caisson, by self-weight or suction, water flows out of the caisson interior through holes in the top cap. The size of the holes is considerably smaller than the interior cross section of the caisson. Therefore, water cannot flow freely and some change in pressure is expected in the interior of the caisson. To simulate this phenomenon, a potential flow formulation was used to estimate the pressure generated at the top of the interior soil during various stages of the simulation[36] (Vásquez 2000). The potential flow formulation is developed on the basis of assumptions that the fluid is incompressible and inviscid and its flow is irrotational.

7 REMESHING

A remeshing tool was developed to eliminate the need for a priori specification of the caisson penetration path. As installation of the caisson progresses, the finite-element mesh is adjusted so that the line of nodes below the tip remains straight in the axial direction. By performing this adjustment, it is possible to eliminate overconfinement of the soil in the caisson interior, thus permitting calculation of the path of penetration in the soil domain[23, 22] (Maniar and Tassoulas 2002, Maniar 2004).

Another remeshing tool was developed to adjust the finite-element mesh along the caisson-soil interfaces. This tool is intended for eliminating distortion of the soil elements along the caisson-soil interfaces and is convenient in cases where a high coefficient of friction on the soil-caisson interfaces leads to significant finite-element distortion. This second remeshing tool was not used in the simulations described below, as it turned out to be unnecessary for this case.

Mapping of field variables from the current finite-element mesh to the adjusted one is carried out using least squares estimation and quadratic interpolation functions[18] (Hinton and Campbell 1974) over the selected set of finite-elements along and below the caisson.

8 SIMULATIONS

Computations are carried out in a sequence that closely follows both laboratory and field tests. The sequence of steps is: a) preparation of the soil test bed starting with the original slurry, b) installation of the caisson by self weight and suction, c) reconsolidation of the soil; and d) pullout of the caisson under either drained or undrained conditions. For the first step, the initial state of the soil domain is obtained from the experimental data. For each of the remaining steps, the initial state of the soil domain is obtained from the end of the previous step. The formulation outlined above is applied to the analysis of model suction caissons installed and tested[21, 7] at The University of Texas at Austin (Luke 2002, Coffman 2003). In this section, the computational results are presented and verified by comparison with the experimental data.

8.1 Preparation of Soil Test Bed

The soil test bed was formed by allowing slurry of kaolinite to consolidate under self-weight, resulting in normally consolidated clay. Details of the preparation and consolidation of the test bed[26, 25] are presented elsewhere (Pedersen 2001, Olson et al. 2003). The consolidation of kaolinite slurry was simulated by analyzing a 24-in diameter kaolinite slurry cylinder with 61-in initial height (same as in the test) and frictional contact on

the lateral impermeable surface (container wall). The time required for consolidation was found to be about 6 months, close to the seven-month period recorded in the laboratory, and the computed final slurry height at the axis of the soil cylinder was 44.6 in, close to a measurement of 43.4 in taken at about the same distance from the wall of the steel tank in which the test bed was prepared (Fig. 1). It is worth mentioning that the three-dimensional (axisymmetric) soil state obtained at the end of the initial-consolidation simulation is critical for success in later computations of caisson response during installation and pullout because of the noticeably weaker soil obtained as a result of friction on the container wall[22] (Maniar 2004). The more straightforward one-dimensional consolidation simulation[22] overestimates the soil strength in the vicinity of the axis by about 30% (Maniar 2004).

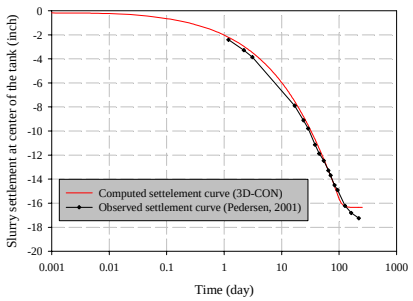


Figure 1: Observed and computed consolidation curves for the kaolinite slurry.

8.2 Self-Weight Installation

In one of the modes of installation examined in the laboratory, the 4-in exterior-diameter model caisson penetrated 32 in under self-weight in about 200 sec. The simulation of self-weight installation was conducted in a similar manner. The computed and measured excess pore-water pressures, recorded at five locations, on interior and exterior wall surfaces of the model caisson during self-weight installation are plotted versus the position of the caisson tip in Fig. 2. Good agreement can be seen, especially at interior locations away from the caisson tip. The so-called "penetration path"[22] (Maniar 2004), i.e., the undeformed surface (line in axisymmetric geometry) on which penetration occurs is shown in Fig. 3. It can be seen that, in the self-weight mode of installation, soil is displaced outwards.

8.3 Suction Installation

In the other mode of installation considered in the laboratory tests, the 4-in exterior-diameter model caisson penetrated 16 in under self-weight in about 69 sec and, subsequently, suction was applied resulting in additional 16 in penetration in about 420 sec. The simulation of this self-weight-followed-by-suction installation, referred to as suction installation below for simplicity, was conducted in a similar manner. The computed and measured excess pore-fluid pressures are in very good at all interior locations and the exterior location away from the tip (O1) but substantial difference was found[22] at the exterior location next to the tip (Maniar 2004). The plot of cumulative soil displaced during penetration[22] shows (Maniar 2004) that, in this mode of installation, a small amount of soil is displaced outwards during the self-weight installation segment but, as expected, soil is drawn into the caisson interior during the suction segment.

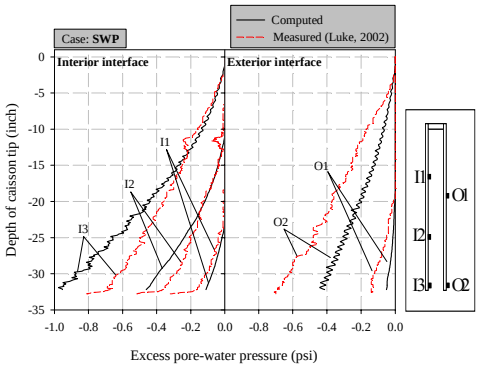


Figure 2: Computed and measured pore-water pressures during self-weight caisson installation.

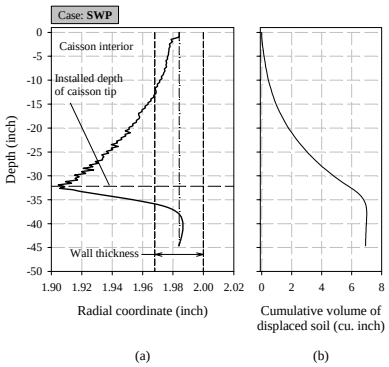


Figure 3: Self-weight penetration path (a) and cumulative volume of displaced soil (b).

8.4 Reconsolidation

In the tests[7] (Coffman 2003) and the simulations alike, the excess pore-water pressures generated during

caisson installation were reduced to negligible levels after about 96 hours of reconsolidation. The computed and measured pore-water pressure time histories are in good agreement as well[22] (Maniar 2004). Fig. 4 depicts the distribution of radial stress after reconsolidation for the case of suction installation. It can be seen that the level of radial stress in the interior of the caisson is higher than in the exterior, especially in the vicinity of the tip. Furthermore, at a depth of about 16 in, where the installation mode was switched from self-weight to suction, there is notably higher radial stress in the caisson exterior, apparently resulting from soil having been "pushed" outwards.

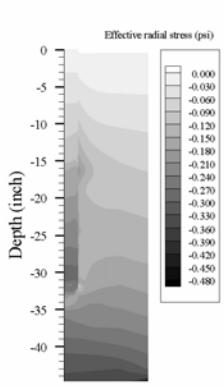


Figure 4: Radial stress distribution after reconsolidation following suction installation.

8.5 Axial Pullout

Starting from each of the two installations, by self-weight penetration (SWP) and suction (SUC), described above, axial-pullout simulations were conducted under a variety of conditions: vented (V) or closed (C) cap, rapid (R), slow (S), or drained (D) pullout (for explanation[22] of these conditions, see Maniar 2004). Not all combinations of installations and conditions have been tested in the laboratory. Table 1 summarizes the computational results regarding axial capacity and the available experimental data along with the computed contributions to resistance from interior friction, exterior friction and suction. The agreement between computations and measurements is good with maximum difference of about 20%. It is worth noting that the simulations as well as the experiments indicate clearly that the capacity of suction caissons installed by self-weight penetration is higher than by suction installation, regardless of pullout speed and independently of whether the cap is vented or closed. The increase in capacity of caissons installed by self-weight appears to be related to the higher soil strength reached in this mode of installation[22] (Maniar 2004).

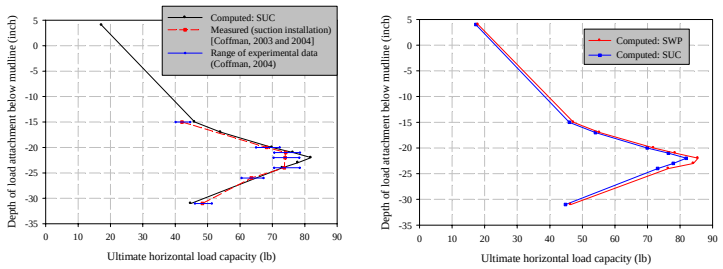
Simulation	Measured Capacity[21] (Luke 2002) (lb)	Computed capacity (lb)	Exterior friction (%)	Interior friction (%)	Total friction (%)	Suction force (%)
SWP-VR	24.0	20.4	52.2	37.4	89.6	-
SWP-VS	-	17.8	58.7	29.4	88.1	-
SWP-VD	-	18.5	53.8	34.8	88.6	-
SWP-CR	28.0	23.4	42.1	7.0	49.1	41.9
SUC-VR	19.2	18.2	40.1	48.3	88.4	-
SUC-VS	-	14.8	48.1	37.6	85.7	-
SUC-VD	20.0	17.3	41.9	45.8	87.7	-
SUC-CR	18.6	22.7	32.2	14.2	46.4	44.3

Table 1: Computed and measured axial capacities (see Section on Axial Pullout for explanation of symbols).

8.6 Lateral Pullout

Abaqus (http://www.hks.com) was used for three-dimensional simulations of caissons subjected to lateral loads, horizontal, or, in general, inclined loads. The soil stresses and state parameters as computed from the installation analysis (self-weight or suction) conducted using the axisymmetric finite-element code developed in the course of this study were imported into Abaqus as initial conditions and, subsequently, the lateral-pullout computations were carried out[22] (see Maniar 2004) for further details. Shown in Fig. 5 (a) are the computed and measured horizontal capacities, in the case of suction installation, considering different locations of the point of load application (pad eye). The agreement is excellent. Computations and measurements indicate that the horizontal capacity is highest for a pad eye located at about 2/3 of the installation depth. Self-weight installation

leads to a small increase in horizontal capacity, as can be seen in Fig. 5 (b). The interaction diagram provided in Fig. 6 shows the capacities computed for inclined loads applied at about the optimal (2/3-point) pad-eye location.



(a) **(b)**

Figure 5: Horizontal capacity: (a) computations and measurements, (b) effect of installation mode.

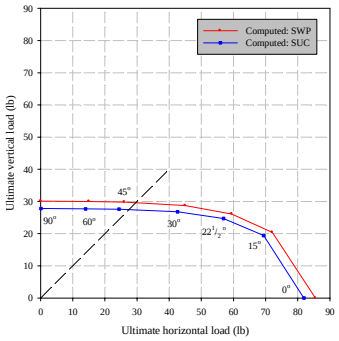


Figure 6: Interaction diagram for inclined loads applied at the optimal pad-eye location

9 CONCLUSIONS

A computational procedure has been developed at the Offshore Technology Research Center for the analysis of suction-caisson behavior under both axial and lateral loads. The procedure has been used in simulations of tests conducted in the course of a concurrent OTRC project at The University of Texas at Austin on caisson models. Computational results and experimental data regarding all facets of the tests have been found to be in good agreement. Installation by self-weight and suction, soil reconsolidation, and axial, horizontal and inclined pullouts have been examined.

Further work, currently underway, involves simulations of centrifuge tests of suction caissons and additional 1-g tests on caisson models as well as documentation of the finite-element code that has been developed.

REFERENCES

[1] Atkin, R.J., and Craine, R.E. (1976), "Continuum theories of mixtures: Basic theory and historical development," Quarterly Journal of Mechanics and Applied Mathematics, Vol. 29, pp. 209-244.

[2] Biot, M.A. (1941), "General theory of three dimensional consolidation," Journal of Applied Physics, Vol. 12, pp. 155-164.

[3] Biot, M.A. (1955), "Theory of elasticity and consolidation of porous anisotropic solid," Journal of Applied Physics, Vol. 26, pp. 182-185.

[4] Bowen, R.M. (1976), "Theory of mixtures," Continuum physics, Ed. A.C. Eringen, Academic Press, New York, pp. 1-127.

[5] Byrne, B.W., and Houlsby, G.T. (2002), "Experimental investigations of response of suction caissons to transient vertical loading," Journal of Geotech. and Geoenv. Engrg., Vol. 128, No. 11, pp. 926-939.

[6] Clukey, E. C., Morrison, M. J., Gariner, J., and Corté, J. F. (1995), "The response of suction caissons in normally consolidated clays to cyclic TLP loading conditions," Proc., Offshore Tech. Conf., pp. 909-918.

[7] Coffman, R.A. (2003), "Horizontal capacity of suction caissons in normally consolidated kaolinite," M.S. Thesis, The University of Texas at Austin.

[8] Coffman, R.A., El-Sherbiny, R.M., Rauch, A.F., and Olson, R.E. (2004), "Measured horizontal capacity of suction caissons," OTC 16161, Proceedings, Offshore Technology Conference, Houston, Texas.

[9] Dafalias, Y.F. (1986), "Bounding surface plasticity I: Mathematical foundation and hypoplasticity," Journal of Engineering Mechanics, ASCE, Vol. 112, No. 9, pp. 966-987.

[10] Dafalias, Y.F., and Hermann, L.R. (1982), "Bounding surface formulation in soil plasticity," Chapter 10 in Soil mechanics – Transient and cyclic loads, Ed., G.N. Pande and O.C. Zienkiewicz, J. Wiley, New York.
[11] Dafalias, Y.F., and Herrmann, L. R., (1986), "Bounding surface plasticity II: Application to isotropic cohesive soils," Journal of Engineering Mechanics, ASCE, Vol. 112, No. 12, pp. 1263-1291.
[12] Deng, W., and Carter, J.P. (2002), "A theoretical study of the vertical uplift capacity of suction caissons," International Journal of Offshore and Polar Engineering, Vol. 12, No. 2, pp. 89-97.
[13] El-Gharbawy, S., and Olson, R. (1999), "Suction caisson foundations in the Gulf of Mexico," Analysis, Design, Construction, and Testing of Deep Foundations, ASCE Geotechnical Sp. Pub. 88, pp. 281-295.
[14] El-Gharbawy, S.L., and Olson, R.E. (2000), "Modeling of suction caisson foundations," Proceedings of the International Offshore and Polar Engineering Conference, Vol. 2, pp. 670-677.
[15] El-Gharbawy, S.L., Olson, R.E., and Scott, S.A. (1999), "Suction anchor installations for deep Gulf of Mexico applications," Proceedings, Offshore Technology Conference, OTC 10992, pp. 747-754.
[16] Erbrich, C.T., and Tjelta, T.I. (1999), "Installation of bucket foundations and suction caissons in sand – Geotechnical performance," Proceedings, Offshore Technology Conference, OTC 10990, pp. 725-735.
[17] Hallquist, J.O., Goudreau, G.L., and Benson, D.J. (1985), "Sliding interfaces with contact-impact in large-scale Lagrangian computations," Computer Methods in Applied Mech. and Engrg., Vol. 51, pp. 107-137.
[18] Hinton, E., and Campbell, J.S. (1974), "Local and global smoothening of discontinuous finite element functions using a least squares method," International Journal for Num. Meth. in Engrg., Vol. 8, pp. 461-480.
[19] Hogervorst, J. R. (1980), "Field trials with large diameter suction piles," Proceedings, Offshore Technology Conference, OTC 3817, pp. 217-224.
[20] Kaliakin, V. N., and Herrmann, L. R. (1991), "Guidelines for implementing the elastoplastic-viscoplastic bounding surface model," Technical Report, Department of Civil Engineering, Univ. of California, Davis.
[21] Luke, A. M. (2002), "Axial capacity of suction caisson in normally consolidated kaolinite," M.S. Thesis, The University of Texas at Austin.
[22] Maniar, D.R. (2004), "A computational procedure for simulation of suction caisson behavior under axial and inclined loads," Ph.D. Dissertation, The University of Texas at Austin, 2000.
[23] Maniar, D. R., and Tassoulas, J. L. (2002), "Nonlinear finite element simulation of suction caisson behavior," Proceedings, Fifteenth ASCE Engineering Mechanics Conference, New York, New York.
[24] Mecham, E.C. (2001), "A laboratory for measuring the axial and lateral capacity of model suction caissons," M.S. Thesis, The University of Texas at Austin.
[25] Olson, R. E., Rauch, A. F., Mecham, E. C., and Luke, A. M. (2003), "Self-weight consolidation of large laboratory deposits of clay," Proc., 12th Pan-Am. Conf. Soil Mech. Geotech. Engrg, Cambridge, Mass.
[26] Pedersen, R.C. (2001), "Model offshore soil deposit: design, preparation and characterization," M.S. Thesis, The University of Texas at Austin.
[27] Prevost, J.H. (1980), "Mechanics of continuous porous media," International Journal of Engineering Science, Vol. 18, pp. 787-800.
[28] Prevost, J.H. (1981), "Consolidation of anelastic porous media," Journal of Engineering Mechanics Division, ASCE 107, EM1, pp 169-186.
[29] Randolph, M. F., O'Neill, M. P., and Stewart, D. P. (1998), "Performance of suction anchors in fine-grained calcareous soils," Proceedings, Offshore Technology Conference, OTC 8831, pp. 521-529.
[30] Rao, S. N, Ravi, R., and Ganapathy, C. (1977), "Pullout behavior of model suction anchors in soft marine clays," Proceedings, Seventh International Offshore and Polar Engineering Conference, Honolulu.
[31] Sparrevik, P. 2001. "Suction pile technology and installation in deep waters," Proceedings, OTRC 2001 Intl. Conf., Geotech., Geol., and Geophy. Props. of Deepwater Sediments, Houston, pp. 182-197.
[32] Steensen-Bach, J. O. (1992), "Recent model tests with suction piles in clay and sand," Proceedings, Offshore Technology Conference, OTC 6844, pp. 323-330.
[33] Sukumaran, B., McCarron, W.O., Jeanjean, P., and Abouseeda, H. (1999), "Efficient finite element techniques for limit analysis of suction caissons under lateral loads," Comp. and Geotech, V. 24, pp. 89-107.
[34] Tjelta, T. I. (1995), "Geotechnical experience from the installation of the Europipe jacket with bucket foundations," Proceedings, Offshore Technology Conference, OTC 7795, pp. 897-908.
[35] Tjelta, T. I., Guttormsen, T. R., and Hermstad, J. (1986), "Large-scale penetration test at a deepwater site," Proceedings, Offshore Technology Conference, OTC 5103, pp. 201-212.
[36] Vásquez, L.F.G. (2000), "Computational procedure for the estimation of pile capacity including simulation of the installation process," Ph.D. Dissertation, The University of Texas at Austin.
[37] Wang, M. C., Demares, K. R., and Nacci, V. A. (1977), "Application of suction anchors in the offshore technology," Proceedings, Offshore Technology Conference, OTC 3203, pp. 1311-1320.
[38] Whittle, A. J., Germaine, J. T., and Cauble, D. F. (1998), "Behavior of miniature suction caissons in clay," Offshore Site Investigation and Foundation Behaviour '98, SUT, pp. 279-300.

IMPROVING THE PERFORMANCE OF THE TRIC SHELL ELEMENT

Andreas G. Gisakis*, Manolis Papadrakakis* and Loutsia Karapitta†

*Institute of Structural Analysis & Seismic Research
National Technical University of Athens
Zografou Campus, 15780 Athens, Greece
e-mail: mpapadra@central.ntua.gr

†Laboratory for Earthquake Engineering
National Technical University of Athens
Zografou Campus, 15780 Athens, Greece

Keywords: shell element, triangle, natural mode, drilling freedoms, membrane, plane stress.

Abstract: *The TRIC element is a simple but sophisticated three-node shear-deformable isotropic and composite facet shell element suitable for large-scale linear and nonlinear engineering computations of thin and moderately thick anisotropic plate and complex shell structures. The element formulation is based on the natural mode finite element method where the deformation or natural modes are separated from the rigid body modes. The element, compared to the conventional isoparametric finite element formulations, has substantial computational advantages such as analyticall expressions for the computation of the stiffness matrix, as well as locking-free properties [1,2]. The proposed improvement concerns the membrane behavior of TRIC which becomes predominant in the detailed simulation of steel structures. The proposed enhancement follows an optimal method in calculating the stiffness matrix regarding the rotational degrees of freedom and is based on an Assumed Natural Deviatoric Strain (ANDES) formulation [3]. The improved element exhibits excellent performance in shell problems with significant membrane stresses and unstructured meshes.*

1 INTRODUCTION

An attempt to device a shell element with robustness, accuracy and efficiency has led to the derivation of the TRIC shell element[4,5], a simple but sophisticated triangular, shear-deformable facet shell element suitable for the analysis of thin and moderately thick isotropic as well as composite plate and shell structures. Its formulation is based on the natural mode finite element method[6], a method introduced by Argyris in the 1950s that separates the pure deformational modes -also called natural modes- from the rigid body movements of the element. The natural mode method in connection with the triangular shape of the element has substantial computational advantages compared to the conventional isoparametric finite element formulations. Appropriate treatment of the element kinematics eliminates automatically locking phenomena while all computations are performed analytically thus avoiding the expensive numerical computation of the stiffness matrix. Furthermore, the inclusion of the transverse shear deformations in the formulation of the TRIC shell element based on a first order shear-deformable beam theory is performed in a way that eliminates the shear locking effect in a physical manner.

The standard TRIC element formulation can be decomposed in two different triangular elements. One responsible for the pure-bending behavior and one responsible for the membrane (plane stress) behavior. The element performance in plane stress problems is identical to the Constant Strain Triangle (CST)[12]. This element encounters great difficulties in determining the exact structural behavior, while it exhibits serious aspect ratio locking.

In the present study, we present an enhancement of the element's membrane behaviour. The enhancement follows an optimal method in calculating the stiffness matrix regarding the rotational or drilling degrees of freedom (corner rotations normal to the plane of a shell element) and is based on an Assumed Natural Deviatoric Strain (ANDES) formulation[3,13]. The enhancement was implemented in two steps. First we improved the azimuth stiffness matrix in the natural system, using the drilling degrees of freedom deformation concept. The drilling degrees of freedom of TRIC were not involved in the derivation of the element's stiffness. These degrees of freedom were retained arbitrarily, solely for computational reasons. Secondly, further enhancement is achieved of the stiffness terms of the drilling degrees of freedom in the local Cartesian system following a constant-stress/hybrid method.

Test examples for membrane and shell problems will de demonstrated. The results will be evaluated with commercial finite element software. We will demonstrate that the proposed element exhibits excellent performance in shell problems with significant membrane stresses and unstructured meshes.

2 THE TRIC ELEMENT

The formulation of the TRIC shell element has been presented thoroughly in a number of papers[1,4,5,10,11]. In the present paper, due to lack of space, only basic theory features of the element will be highlighted and only those features that are essential for the introduced improvement.

2.1 Natural kinematics of the element

The key-point for the formulation of the TRIC shell element is the adoption of the so-called natural coordinate system which has the three axes parallel to the sides of the triangle (α, β, γ). Apart from the natural system (α, β, γ) there are also the local elemental coordinate system (x', y', z') placed at the triangle's centroid, and the global cartesian coordinate system (x, y, z) where global equilibrium refers to. Finally, for each ply of the triangle, a material coordinate system (*1, 2, 3*) is defined with axis 1 being parallel to the direction of the fibers. The use of these different coordinate systems makes TRIC a suitable element in modeling a multilayer anisotropic shell structure that can degenerate, as a special case, to a sandwich or a single-layer configuration.

In the natural mode method the cartesian strains have been replaced by the total natural strains $\boldsymbol{\gamma}_{\mathrm{t}} = \left\{ \gamma_{\mathrm{t}\alpha} \quad \gamma_{\mathrm{t}\beta} \quad \gamma_{\mathrm{t}\gamma} \right\}^{\mathrm{T}}$. These strains are measured directly parallel to the triangle's sides, while by definition straining of one side leaves all other triangular sides unstrained. The total natural axial strains $\boldsymbol{\gamma}_{\mathrm{t}}$ are related to the three in-plane local cartesian strains $\boldsymbol{\gamma}'$ according to the expression

$$\boldsymbol{\gamma}_{\mathrm{t}} = \mathbf{B}^{\mathrm{T}} \boldsymbol{\gamma}' \Longleftrightarrow \begin{Bmatrix} \gamma_{\mathrm{t}\alpha} \\ \gamma_{\mathrm{t}\beta} \\ \gamma_{\mathrm{t}\gamma} \end{Bmatrix} = \begin{bmatrix} c^2_{\alpha x'} & s^2_{\alpha x'} & c_{\alpha x'} s_{\alpha x'} \\ c^2_{\beta x'} & s^2_{\beta x'} & c_{\beta x'} s_{\beta x'} \\ c^2_{\gamma x'} & s^2_{\gamma x'} & c_{\gamma x'} s_{\gamma x'} \end{bmatrix} \begin{Bmatrix} \gamma_{x'x'} \\ \gamma_{y'y'} \\ 2\gamma_{x'y'} \end{Bmatrix} \quad \left| \begin{array}{l} c_{ix'} = cos\langle i, x' \rangle \\ s_{ix'} = sin\langle i, x' \rangle \end{array}, \quad i = \alpha, \beta, \gamma \right. \tag{1}$$

2.2 Constitutive relations (The natural stress-strain matrix $\boldsymbol{\kappa}_{\mathrm{ct}}$)

In the present paragraph no mention will be made for the transverse shear strains, as no modification has been made to them and they present no interest for the proposed improvement.

With respect to the material coordinate system we will define orthotropic properties such as Young's moduli and Poisson's ratios, for every layer *k*. Taking into consideration that in general all material properties are temperature dependent and for complete isotropy ($v_{12} = v_{21} = v$, $E_1 = E_2 = E$) the equation becomes

$$\begin{Bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{12} \end{Bmatrix}_k = \underbrace{\frac{1}{1-v^2} \begin{bmatrix} E & vE & 0 \\ vE & E & 0 \\ 0 & 0 & G(1-v^2) \end{bmatrix}_k}_{\boldsymbol{\kappa}_{12}} \begin{Bmatrix} \gamma_{11} \\ \gamma_{22} \\ \gamma_{12} \end{Bmatrix}_k \quad , \quad G = \frac{E}{2(1+v)} \tag{2}$$

In order to express the constitutive matrix in the natural system the following sequence of material transformations are initiated: Material system → Local system → Natural system

$$\boldsymbol{\kappa}_{\mathrm{ct}}{}^{\mathrm{k}} = \left[\mathbf{B}^{-1} \left[\mathbf{A}^{\mathrm{T}} \boldsymbol{\kappa}_{12} \mathbf{A} \right] \mathbf{B}^{-\mathrm{T}} \right]_k \tag{3}$$

in which $\mathbf{A}$ is the transformation matrix that relates the material with the local system (its form is presented in [4]). In this work $\mathbf{A}$ is a unit vector since the material is taken as isotropic and material and local coordinate systems are identical.

2.3 Natural modes

In principle, the natural stiffness of an element is only based on deformation and not on associated rigid body motions. Thus, to the triangular shell element TRIC correspond 6 × 3 = 18 nodal displacements but only 18 - 6 = 12 independent straining modes can be denned in order to satisfy all kinematic compatibility conditions. The stiffness matrix $\mathbf{k}_{\mathrm{N}}$ corresponding to these deformations is of dimensions 12 × 12 and is denoted as the natural stiffness matrix. A simple congruent transformation leads to the full 18 × 18 cartesian stiffness matrix of the element.

In order to generate these pure deformational modes, which are called natural modes, a projection of the nodal displacements and rotations as well as the corresponding forces and moments on the triangular edges takes place. A decomposition of the rotations and moments into symmetric and antisymmetric components is then initiated. From this process, the axial as well as the symmetric and antisymmetric modes of deformation are generated. Note that the antisymmetric modes comprise the antisymmetric bending and transverse shear deformations, a key-point for the shear-locking elimination[1].

The 6 rigid-body and 12 straining natural modes of the TRIC element, can be grouped in the vector:

$$\underset{(18\times1)}{\boldsymbol{\rho}} = \left\{ \underset{(6\times1)}{\boldsymbol{\rho}_0} \quad \underset{(12\times1)}{\boldsymbol{\rho}_N} \right\}^T \left| \begin{array}{l} \boldsymbol{\rho}_0 = \left[\rho_{01} \quad \rho_{02} \quad \rho_{03} \quad \rho_{04} \quad \rho_{05} \quad \rho_{06} \right]^T \quad and \\ \boldsymbol{\rho}_N = \left[\gamma_{ta}^0 \quad \gamma_{t\beta}^0 \quad \gamma_{t\gamma}^0 \quad \psi_{S\alpha} \quad \psi_{A\alpha} \quad \psi_{S\beta} \quad \psi_{A\beta} \quad \psi_{S\gamma} \quad \psi_{A\gamma} \quad \psi_\alpha \quad \psi_\beta \quad \psi_\gamma \right]^T \end{array} \right. \tag{4}$$

where $\boldsymbol{\rho}_0$ and $\boldsymbol{\rho}_N$ represent the rigid-body and straining modes, respectively.

TRIC, as a shell element, can be regarded as a combination of a membrane and a plate element that is subjected to extension and flexure, respectively. Its membrane behavior is represented by six translational nodal degrees of freedom in the plane of the triangle, while its flexural behavior is described by nine degrees of freedom, one out-of-plane translation and two rotations for each node. The remaining three degrees of freedom, one for each node, are the drilling degrees of freedom that were implemented solely for computational reasons and they were treated separately from the other modes with no coupling terms between them.

2.4 Natural stiffness matrix $\mathbf{k}_N$

The components of the natural stiffness matrix are [1]:

$$\underset{(12\times12)}{\mathbf{k}_N} = \begin{bmatrix} \underset{(6\times6)}{\mathbf{k}_{qc}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \underset{(3\times3)}{\mathbf{k}_{qh}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \underset{(3\times3)}{\mathbf{k}_{az}} \end{bmatrix} \tag{5}$$

where: $\mathbf{k}_{qc}$ are the axial and symmetrical bending stiffness terms, $\mathbf{k}_{qh}$ are the anti-symmetrical bending and shear stiffness terms and $\mathbf{k}_{az}$ are the stiffness terms due to in-plane rotations (azimuth stiffness terms). The Derivation of $\mathbf{k}_{qc}$ and $\mathbf{k}_{qh}$ was already presented in [1].

2.4.1 Azimuth stiffness matrix $\mathbf{k}_{az}$ (Initial formulation)

The drilling degrees of freedom are not involved in the derivation of the element's stiffness. However, these degrees of freedom are retained, solely for computational reasons. In the case of the TRIC element, the assignment of stiffness terms to the drilling degrees of freedom is done in a way that is consistent with the natural mode method. Three rotational springs with the same stiffness $\mathbf{k}_Z$ are considered at the three vertices of the triangle and they are used to simulate the in-plane rotation about z' axis. The corresponding natural modes consist of a unit rotation about z' axis at each one of the three vertices $\mathbf{q}_{az} = \left\{ \psi_\alpha \quad \psi_\beta \quad \psi_\gamma \right\}^T$. The azimuth stiffness matrix $\mathbf{k}_{az}$ is then calculated as:

$$\mathbf{k}_{az} = k_z \begin{bmatrix} 1 & -0.5 & -0.5 \\ -0.5 & 1 & -0.5 \\ -0.5 & -0.5 & 1 \end{bmatrix}, \; k_z = 10^{-6} \Omega \times max \left\{ \frac{1}{l_\alpha^2} \int_{-h/2}^{h/2} z^2 k_{\alpha\alpha} dz \;\; , \;\; \frac{1}{l_\beta^2} \int_{-h/2}^{h/2} z^2 k_{\beta\beta} dz \;\; , \;\; \frac{1}{l_\gamma^2} \int_{-h/2}^{h/2} z^2 k_{\gamma\gamma} dz \right\} \tag{6}$$

with k_z having an arbitrary but small enough value, compared to the rest of the stiffness terms, so that it will produce a negligible effect on the final equilibrium equations. In (6) h is the thickness of the shell element, l_i (i = α, β, γ) are the lengths of the three sides of the element, Ω is the element area and $k_{\alpha\alpha}$, $k_{\alpha\alpha}$, $k_{\alpha\alpha}$ are the diagonal terms of $\boldsymbol{\kappa}_{ct}$. It should be noted that these modes are defined exclusively for computational reasons and this is the reason why the coupling terms (Eq. (5)) connecting these degrees of freedom with the rest of the modes are set equal to zero.

2.5 Local cartesian stiffness matrix $\mathbf{k}$

The cartesian stiffness matrix $\mathbf{k}$ in the local coordinate system can be found using the following transformation

$$\underset{(18\times18)}{\mathbf{k}} = \underset{(18\times12)}{\bar{\boldsymbol{\alpha}}_N^T} \underset{(12\times12)}{\mathbf{k}_N} \underset{(12\times18)}{\bar{\boldsymbol{\alpha}}_N} \tag{7}$$

where $\bar{\boldsymbol{\alpha}}_N$ is the matrix that relates the natural straining modes $\mathbf{p}_N$ to the cartesian nodal displacements $\bar{\mathbf{p}}$

$$\underset{(1\times18)}{\bar{\mathbf{p}}^T} = \left[u' \quad \upsilon' \quad w' \quad \theta' \quad \phi' \quad \psi' \right]_i^T \qquad i = 1,2,3 \tag{8}$$

with u_i, v_i, w_i being the three nodal translations and θ_i, φ_i, ψ_i the three nodal rotations of node i. The calculation of $\bar{\boldsymbol{\alpha}}_N$ is done using strictly geometrical arguments and can be expressed analytically[4].

3 IMPROVEMENT OF THE MEMBRANE BEHAVIOR

3.1 Drilling degrees of freedom

The drilling degrees of freedom are useful for compactly expressing the higher order behavior of the element. Their geometric interpretation is shown in Fig 1.

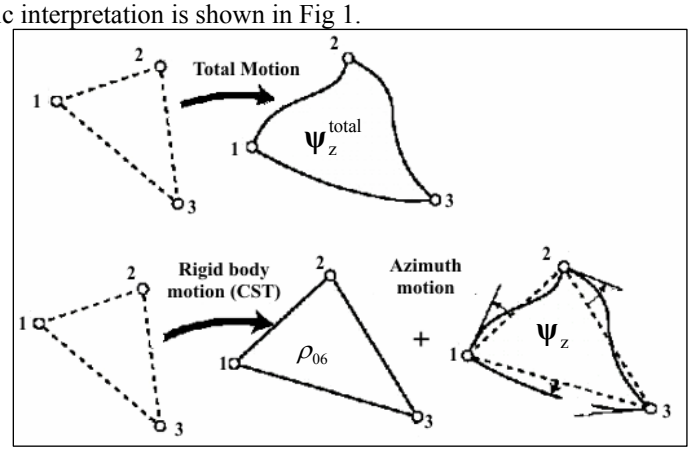


Figure 1. Decomposition of inplane motion into CST (linear displacement) and azimuthian.

To extract the azimuth corner rotations $\boldsymbol{\psi}_z = \{\psi_\alpha \quad \psi_\beta \quad \psi_\gamma\}^T$ from the total corner rotations $\boldsymbol{\psi}_z^{total}$, subtract the mean or CST rotation ρ_{06}:

$$\boldsymbol{\psi}_z = \boldsymbol{\psi}_z^{total} - \rho_{06} \tag{9}$$

where z = α, β, γ and $\rho_{06} = \frac{1}{4\Omega}(x_\alpha u_1 + x_\beta u_2 + x_\gamma u_3 + y_\alpha v_1 + y_\beta v_2 + y_\gamma v_3)$.

3.2 Azimuth stiffness matrix $\mathbf{k}_{az}$ (modification)

The modification of the azimuth stiffness matrix follows the work of Felippa and co workers and is based on an Assumed Natural Deviatoric Strain (ANDES) formulation [3,13].

The modified azimuth stiffness matrix is given by

$$\underset{(3\times3)}{\mathbf{k}_{az}} = \sum_{k=1,2,..,nl} \underset{(3\times3)}{\mathbf{k}_\psi^k} \tag{10}$$

where $\mathbf{k}_\psi^k$ is the azimuth stiffness matrix in each layer (k) of the element in terms of the azimuth corner rotations $\boldsymbol{\psi}_z$ of (9).

To express $\mathbf{k}_\psi^k$ compactly, the following matrices are introdused, which depend on nine free dimensionless parameters, β_1 through β_9

$$\mathbf{Q}_1 = \frac{2\Omega}{3}\begin{bmatrix} \frac{\beta_4}{l_a^2} & \frac{\beta_5}{l_a^2} & \frac{\beta_6}{l_a^2} \\ \frac{\beta_7}{l_\beta^2} & \frac{\beta_8}{l_\beta^2} & \frac{\beta_9}{l_\beta^2} \\ \frac{\beta_1}{l_\gamma^2} & \frac{\beta_2}{l_\gamma^2} & \frac{\beta_3}{l_\gamma^2} \end{bmatrix}, \quad \mathbf{Q}_2 = \frac{2\Omega}{3}\begin{bmatrix} \frac{\beta_3}{l_a^2} & \frac{\beta_1}{l_a^2} & \frac{\beta_2}{l_a^2} \\ \frac{\beta_6}{l_\beta^2} & \frac{\beta_4}{l_\beta^2} & \frac{\beta_5}{l_\beta^2} \\ \frac{\beta_9}{l_\gamma^2} & \frac{\beta_7}{l_\gamma^2} & \frac{\beta_8}{l_\gamma^2} \end{bmatrix}, \quad \mathbf{Q}_3 = \frac{2\Omega}{3}\begin{bmatrix} \frac{\beta_8}{l_a^2} & \frac{\beta_9}{l_a^2} & \frac{\beta_7}{l_a^2} \\ \frac{\beta_3}{l_\beta^2} & \frac{\beta_1}{l_\beta^2} & \frac{\beta_2}{l_\beta^2} \\ \frac{\beta_5}{l_\gamma^2} & \frac{\beta_6}{l_\gamma^2} & \frac{\beta_4}{l_\gamma^2} \end{bmatrix} \tag{11}$$

Matrix $\mathbf{Q}_i$ relates the natural strains $\boldsymbol{\gamma}_t$, at elements corner i, to the azimuth corner rotations $\boldsymbol{\psi}_z$. At a point of triangular coordinates {ζ_1, ζ_2, ζ_3}, $\boldsymbol{\gamma}_t = \mathbf{Q}\boldsymbol{\psi}_z$, where $\mathbf{Q} = \mathbf{Q}_1\zeta_1 + \mathbf{Q}_2\zeta_2 + \mathbf{Q}_3\zeta_3$. Evaluate this at the midpoints $\mathbf{Q}_4 = \frac{1}{2}(\mathbf{Q}_1 + \mathbf{Q}_2)$, $\mathbf{Q}_5 = \frac{1}{2}(\mathbf{Q}_2 + \mathbf{Q}_3)$, $\mathbf{Q}_6 = \frac{1}{2}(\mathbf{Q}_3 + \mathbf{Q}_1)$. Finally:

$$\mathbf{k}_\psi^k = h(\mathbf{Q}_4^T \boldsymbol{\kappa}_{ct} \mathbf{Q}_4 + \mathbf{Q}_5^T \boldsymbol{\kappa}_{ct} \mathbf{Q}_5 + \mathbf{Q}_6^T \boldsymbol{\kappa}_{ct} \mathbf{Q}_6) \tag{12}$$

and $\mathbf{k}_{\text{az}} = \frac{3}{4}\beta_0 \mathbf{k}_{\psi}^{\text{k}}$, where β_0 is an overall scaling coefficient. This coefficient could be absorbed into the β_1 through β_9 but it is left separate, to simplify the incorporation of material behavior into (12). The factor ¾ comes from "historical grandfathering"[13,14]

The free dimensionless parameters β_i are determined from a higher order patch test which tunes up the higher order stiffness of triangular elements[14]. Using such a patch test the optimal parameters are calculated as follows[13]: $\beta_0 = \frac{1}{2}(1 - 4\nu^2)$, $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 1$, $\beta_4 = 0$, $\beta_5 = 1$, $\beta_6 = -1$, $\beta_7 = -1$, $\beta_8 = -1$, $\beta_9 = -2$ ·

Since for $\nu = ½$ the optimal β_0 is 0, the azimuth stiffness would vanish and the element is rank deficient. To maintain stability β_0 is set to a minimum value, for example $\beta_0 = \max(½(1 - 4\nu^2), 0.01)$.

3.3 Cartesian stiffness matrix k (modification)

So far we focused our attention on a more exact method for calculating the stiffness terms of azimuth modes in the natural system. What has not been evaluated yet, is the relation between the azimuth modes in the natural and in the local systems. Furthermore it should be noted that the coupling terms (5) connecting these degrees of freedom with the rest of the modes were set equal to zero. This means, that according to the previous theory of TRIC there is no tuning of the axial, symmetric bending, antisymmetric bending and shearing modes when the element is distorted in the z′ direction.

We've discussed in §2.5 that the cartesian stiffness matrix **k** in the local coordinate system can be found using (7), and $\bar{\boldsymbol{\alpha}}_{\text{N}}$ is the matrix that relates the natural straining modes $\mathbf{p}_{\text{N}}$ to the cartesian nodal displacements $\bar{\mathbf{p}}$. Therefore $\bar{\boldsymbol{\alpha}}_{\text{N}}$ is solely a function of the current geometry of the element and is devised based upon pure geometrical relations between $\mathbf{p}_{\text{N}}$ and $\bar{\mathbf{p}}$. So this matrix can be used to solve all the above mentioned geometrical problems.

3.3.1 Relation between the azimuth modes in the natural and in the local system (1)

By applying (9) and ρ_{06} to the three corners we assemble the transformation:

$$\boldsymbol{\psi}_{\text{Z}} = \begin{Bmatrix} \psi_\alpha \\ \psi_\beta \\ \psi_\gamma \end{Bmatrix} = \frac{1}{4\Omega} \begin{bmatrix} x_a & y_a & 4\Omega & x_\beta & y_\beta & 0 & x_\gamma & y_\gamma & 0 \\ x_a & y_a & 0 & x_\beta & y_\beta & 4\Omega & x_\gamma & y_\gamma & 0 \\ x_a & y_a & 0 & x_\beta & y_\beta & 0 & x_\gamma & y_\gamma & 4\Omega \end{bmatrix} \begin{Bmatrix} u_i' \\ v_i' \\ \psi_i' \end{Bmatrix}_{i=1,2,3} \tag{13}$$

In (13) we present only those elements of $\bar{\mathbf{p}}$ that have a relation with $\boldsymbol{\psi}_{\text{z}}$. All other degrees of freedom (w_i, θ_i, ϕ_i *i=1,2,3*) have columns equal to zero, in the above transformation matrix. This relationship was incorporated in the TRIC's theory[4], but due to the arbitrary and small value of $\mathbf{k}_{\text{N}}$ and subsequently $\mathbf{k}_{\text{az}}$ its influence was negligible.

3.3.2 Relation between the azimuth modes in the natural and in the local system (2)

One more modification remains related to the coupling terms in (5). According to TRIC's theory[4] the natural axial straining modes $\boldsymbol{\gamma}_{\text{t}}$ are related only with the inplane local deformation degrees of freedom (x_i', y_i' for *i=1,2,3*). These modes are also causing a distortion to the element angles. This distortion can be expressed as:

$$\begin{Bmatrix} \gamma_{t\alpha} \\ \gamma_{t\beta} \\ \gamma_{t\gamma} \end{Bmatrix} = \frac{a_b h}{12} \begin{bmatrix} y_a(y_b - y_\gamma) & x_a(x_b - x_\gamma) & 2(y_b x_b - y_\gamma x_\gamma) \\ y_\beta(y_\gamma - y_\alpha) & x_\beta(x_\gamma - x_\alpha) & 2(y_\gamma x_\gamma - y_\alpha x_\alpha) \\ y_\gamma(y_\alpha - y_\beta) & x_\gamma(x_\alpha - x_\beta) & 2(y_\alpha x_\alpha - y_\beta x_\beta) \end{bmatrix} \mathbf{B} \begin{Bmatrix} \psi_1' \\ \psi_2' \\ \psi_3' \end{Bmatrix} \Leftrightarrow \boldsymbol{\gamma}_{\text{t}} = \frac{a_b h}{12} \mathbf{L}\mathbf{B}\boldsymbol{\psi}_i' \tag{14}$$

where h is the elements height and **B** is from (1). **L** is the lumping matrix[14] which for a general three-dimensional element expresses the relationship between the nodal forces $\mathbf{t}_c$ produced by constant-strain modes and an arbitrary constant-stress field ($\mathbf{t}_c = \mathbf{L}\boldsymbol{\sigma}_c$). Finally a_b is a variable which controls the amount of the element angle distortion.

A short examination of a_b shows that, if it is set equal to zero we return to the previous theory, if it is set a lot greater than 1 the element will have very large rotational stiffness. The value of a_b is always in the range of 1 thought 2 and is not sensitive to the choice of material properties. Its value depends only the elements

geometry and on the elements corners (3 different a_b for every corner). The average value $a_b = 1,5$ is recommended for general use in arbitrary grids[14].

Using (14) we get:

$$\frac{a_b h}{12}\mathbf{LB} = \begin{bmatrix} LB_{11} & LB_{12} & LB_{13} \\ LB_{21} & LB_{22} & LB_{23} \\ LB_{31} & LB_{32} & LB_{33} \end{bmatrix} \tag{15}$$

3.3.3 Transformation matrix $\bar{\boldsymbol{\alpha}}_{\mathrm{N}}$

We are now in the position to construct matrix $\bar{\boldsymbol{\alpha}}_{\mathrm{N}}$ (7) which is partitioned as follows

$$\underset{(12x18)}{\bar{\boldsymbol{\alpha}}_{\mathrm{N}}} = \begin{bmatrix} \underset{(6x6)}{\bar{\boldsymbol{\alpha}}_{\mathrm{N}}^{11}} & \underset{(6x6)}{\bar{\boldsymbol{\alpha}}_{\mathrm{N}}^{12}} & \underset{(6x6)}{\bar{\boldsymbol{\alpha}}_{\mathrm{N}}^{13}} \\ \underset{(6x6)}{\bar{\boldsymbol{\alpha}}_{\mathrm{N}}^{21}} & \underset{(6x6)}{\bar{\boldsymbol{\alpha}}_{\mathrm{N}}^{22}} & \underset{(6x6)}{\bar{\boldsymbol{\alpha}}_{\mathrm{N}}^{23}} \end{bmatrix} \tag{16}$$

A typical submatrix of (16) is given by:

$$\underset{(6x6)}{\bar{\boldsymbol{\alpha}}_{\mathrm{N}}^{11}} = \begin{bmatrix} . & . & . & . & . & LB_{11} \\ x_\beta/l_\beta^{\,2} & y_\beta/l_\beta^{\,2} & . & . & . & LB_{21} \\ -x_\gamma/l_\gamma^{\,2} & y_\gamma/l_\gamma^{\,2} & . & . & . & LB_{31} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & -y_\beta/l_\beta & -x_\beta/l_\beta & . \end{bmatrix} \tag{17}$$

the others can be expressed similarly.

4 NUMERICAL EXAMPLES

For the comparisons we used HKS ABAQUS v6.4 (S3R general purpose triangle shell element), MSC NASTRAN 2004 (CTRIA3 general purpose triangle shell element), MSC NASTRAN 2004 (CQUAD4 general purpose quadrilateral shell element) and standard TRIC and the improved TRIC elements.

4.1 Cantilever under end moment (a)

The cantilever beam of Fig. 2 is subjected to an end moment M = 100. The modulus of elasticity is set to E = 768 and the exact tip deflection δ_{tip} = ML/(2EI) is 100. Regular meshes ranging from 32 × 2 to 2 × 2 are used. Each rectanglular mesh unit being composed of 2 equal triangles. For the TRIC elements, 2 isotropic layers with equal height were used for every triangle. The element aspect ratios (γ) vary from 1:1 through 16:1. The root clamping condition is imposed by setting

$$u_1 = u_2 = u_3 = 0, \quad u_2 = 0, \quad y_1 = y_2 = y_3 = 0 \tag{18}$$

where 1, 2, 3 are the root nodes, numbered from the top.

Table 1 reports the computed tip deflections (*y* displacement at C) for several element types and five aspect ratios.

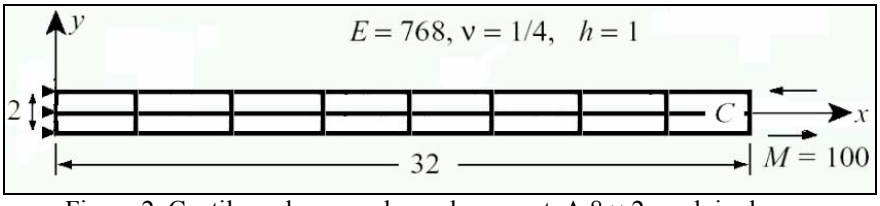


Figure 2. Cantilever beam under end moment. A 8 × 2 mesh is shown.

dy^C	2x2 (γ = 16:1)	4x2 (γ = 8:1)	8x2 (γ = 4:1)	16x2 (γ = 2:1)	32x2 (γ = 2:1)
NASTRAN (Quads)	94.12	96.92	98.31	98.35	98.36
NASTRAN (Triangles)	1.28	4.81	15.80	36.70	54.01
ABAQUS (Triangles)	1.28	4.84	15.84	36.62	54.29
CST	1.28	4.84	15.84	36.62	54.29
TRIC	1.28	4.84	15.84	36.62	54.29
TRIC (improved)	101.32	101.19	100.96	100.54	100.08
Felippa[13]	101.32	101.19	100.96	100.54	100.08

Table 1. Computed tip deflections (*y* displacement at C)

4.2 Cantilever under end moment (b)

The cantilever beam of Fig. 3 is identical to example 4.1. Their only difference is that the total moment is applied by both concentrated moments and forces. Table 2 reports the computed tip deflections (*y* displacement at C) for several element types and five aspect ratios for example 4.2.

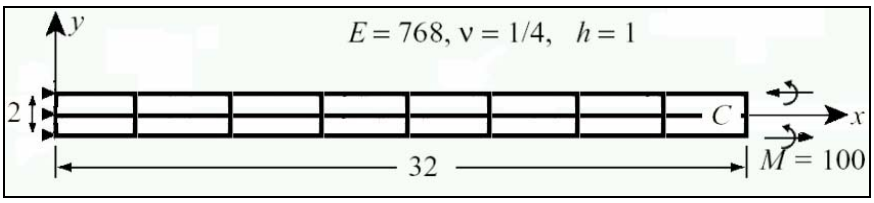


Figure 3. Cantilever beam under end moment. A 8 × 2 mesh is shown.

dy^C	2x2 (γ = 16:1)	4x2 (γ = 8:1)	8x2 (γ = 4:1)	16x2 (γ = 2:1)	32x2 (γ = 2:1)
NASTRAN (Quads)	94.67	97.44	98.16	98.33	98.37
NASTRAN (Triangles)	1.25	4.82	15.83	36.62	54.31
TRIC	0.64	2.42	7.92	18.31	27.15
TRIC (improved)	101.32	101.22	101.0	100.58	100.07
Felippa[13]	99.99	99.99	99.99	99.96	100.07

Table 2. Computed tip deflections (*y* displacement at C)

4.3 Single bay frame

The single bay model frame of Figure 4 consists of isotropic material, I shaped beams for both the columns and the beam. The beam web is perforated in two positions with elliptical holes. The geometry of these structures demands a more detailed simulation with shell finite elements, in order to capture the true deformation and stress patterns.

The structure is fully constrained in the bottom of the two columns and a concentrated force of magnitude 50 along the x direction was applied at the center top node of the right column (Fig. 4). The columns height is 220 and the beam length is 300. For the columns h_w = 24 (web height), h_f = 18 (flange height) and the thickness for both flange and web is 6. For the beam h_w = 20 (web height), h_f = 13 (flange height) and the thickness for both flange and web is 5.

In order to have a better view of the true structural behavior two different analysis were conducted with NASTRAN quadrilateral elements. The first one with 550 elements and 648 nodes and the resulting tip deflection was found 160.49, while the second was simulated with 1982 elements and 2168 nodes and its tip deflections was found 161.81.

Table 3 reports the computed tip deflections (*x* displacement at the node where the force was applied) for several element types and 6 different meshes.

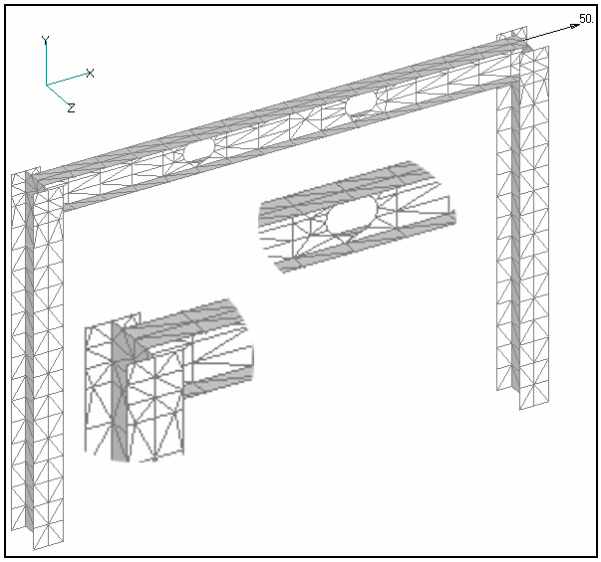


Figure 4. Single bay frame (433 nodes, 718 elemens). A better view of the joint and the hole is presented in the details. The visualization was made by FEMAP. (E = 28, *v* = 0.2)

Nodes	Elements	ABAQUS (Triangles)	NASTRAN (Triangles)	TRIC	TRIC (improved)
253	**408**	45.40	49.63	49.09	146.03
330	**540**	70.70	75.36	75.13	151.40
363	**598**	72.60	76.90	76.79	153.67
433	**718**	93.10	97.44	97.65	155.62
658	**1108**	97.00	100.68	101.09	158.70
1726	**2970**	119.00	120.09	121.40	160.11

Table 3. Computed tip deflections (*x* displacement at the node where the force was applied)

5 CONCLUSIONS

From the above examples a number of concluding remarks can de drawn regarding the TRIC element and the proposed improvement. The standard TRIC element is very stiff and incapable to compute the correct deformation, although finer meshes converge rather slowly, to the correct solution. The improved TRIC has substantial differences with its original version. It is able to compute the correct solution with coarse meshes as well as with distorted elements (aspect ratio 16:1).

The last example was a test bed for the behavior of the improved element in all kinds of modes (membrane, bending and shear). In this case the elements are of arbitrary aspect ratio and the mesh is quite distorted. The improved TRIC exhibits excellent performance.

In all the examples the improved TRIC exhibits fast convergence to the correct solution compared to other general purpose shell elements. Furthermore, the results obtained are very similar to those obtained with quadrilateral shell elements of commercial finate element programs.

6 REFERENCES

[1] Argyris, J.H., Papadrakakis, M., Apostolopoulou, C. and Koutsourelakis, S. (2000), ‘‘The TRIC shell element: theoretical and numerical investigation,’’ Comp. Meth. Appl. Mech. Engrg. 182, pp. 217-245.

[2] Argyris, J.H., Papadrakakis, M. and Karapitta, L. (2002), ‘‘Elasto-plastic analysis of shells with the triangular element TRIC,’’ Comp. Meth. Appl. Mech. Engrg. 191, pp. 3613-3636.

[3] Felippa, C.A. (2003), ‘‘A study of optimal membrane triangles with drilling freedoms,’’ Comp. Meth. Appl. Mech. Engrg. 192, pp. 2125-2168.

[4] Argyris, J.H., Tenek, L. and Olofsson, L. (1997), ‘‘TRIC: a simple but sophisticated 3-node triangular element based on 6 rigid body and 12 straining modes for fast computational simulations of arbitrary isotropic and laminated composite shells,’’ Comp. Meth. Appl. Mech. Engrg. 145, pp. 11-85.

[5] Argyris, J.H., Tenek, L., Papadrakakis, M. and Apostolopoulou, C. (1998), ‘‘Postbuckling performance of the TRIC natural mode triangular element for isotropic and laminated composite shells,’’ Comp. Meth. Appl. Mech. Engrg. 166, pp. 211-231.

[6] Argyris, J.H., Balmer, H., Doltsinis, J.St., Dunne, P.C., Haase, M., Muller, M. and Scharpf, W.D. (1979), ‘‘Finite element method – the natural approach,’’ Comp. Meth. Appl. Mech. Engrg. 17/18, pp. 1-106.

[7] Argyris, J.H., Dunne, P.C., Malejanakis, G.A. and Schekle, E. (1977), ‘‘A simple triangular facet shell element with applications to linear and nonlinear equilibrium and inelastic stability’’, Comp. Meth. Appl. Mech. Engrg. 10, pp. 371-403.

[8] Argyris, J.H., Haase, M. and Mlejnek, H.-P. (1977), ‘‘On an unconventional but natural formation of the stiffness matrix,’’ Comp. Meth. Appl. Mech. Engrg. 22. pp. 371-403.

[9] Argyris, J.H. and Tenek, L. (1994), ‘‘An efficient and locking free flat anisotropic plate and shell triangular element,’’ Comp. Meth. Appl. Mech. Engrg. 118, pp. 63-119.

[10] Stefanou, G. and Papadrakakis, M. (2004), ‘‘Stochastic finite element analysis of shells with combined random material and geometric properties,’’ Comp. Meth. Appl. Mech. Engrg. 193, pp. 139-160.

[11] Argyris, J., Papadrakakis, M. and Mouroutis, Z.S. (2003), ‘‘Nonlinear dynamic analysis of shells with the triangular element TRIC,’’ Comp. Meth. Appl. Mech. Engrg. 192, pp. 3005-3038

[12] Turner, M.J., Clough, R.W., Martin, H.C. and Topp, L.J. (1956), ‘‘Stiffness and deflection analysis of complex structures,’’ J. Aero. Sci. 23, pp. 805–824.

[13] Felippa, C.A. and Militello, C. (1992), ‘‘Membrane triangles with corner drilling freedoms. II. The ANDES element,’’ Finite Elements Anal. Des. 12, pp. 189-201.

[14] Bergan, P.G. and Felippa, C.A. (1985), ‘‘A triangular membrane element with rotational degrees of freedom’’ Comp. Meth. Appl. Mech. Engrg. 50, pp. 25-69

SIMPLE, LOCKING-FREE QUADRILATERAL AND TRIANGULAR BENDING ELEMENTS: LINEAR AND NONLINEAR PROBLEMS

Elias A. Paraskevopoulos and Demosthenes G. Talaslidis

Department of Civil Engineering
Aristotle University Thessaloniki
Mail Stop 502, 54124 Thessaloniki, Greece
e-mail: talaslid@civil.auth.gr

Keywords: bending quadrilaterals, excessive energy/locking, Hu-Washizu functional, shape sensitivity, convergence properties, elastoplastic analysis.

Abstract. *In the present work, various sources of excessive energy in quadrilateral and triangular bending finite elements are presented. Emphasis is placed upon geometric parameters that are inherently related to quadrilaterals and triangles and provide a measure for the distortion of aforesaid forms from the corresponding regular forms. Furthermore, the effect of that distortion on the behavior of the element is investigated. An approach based on a modified version of the functional of Hu-Washizu is proposed for the development of bending elements that are less sensitive to element distortions. The methodology incorporates the rotations of the material lines and decomposes the shear strains into one part, which contains only shear modes and another part that contains flexural as well as shear modes. Appropriate selection of rotations, stresses, and strains leads to simple and reliable elemental formulations, which minimize excessive energy or, in case of triangles, are free of locking. The convergence of the formulation is investigated both analytically and numerically. Furthermore, extension of the proposed approach to elasto-plastic problems is straightforward. In this case, we assume piecewise constant approximations for stress and strain within subregions instead of linear approximations. Finally, results of numerical examples are presented.*

1 INTRODUCTION

Low order triangular and quadrilateral elements are often preferred in structural mechanics. Despite the great variety of alternative schemes that have been proposed, a series of difficulties inherent in these elements exist (see, e.g., [1] to [5]). Besides the coupling between transverse shear and bending deformational modes that leads to excessive energy (locking), there is a coupling between torsion and twist bending mode that causes also locking and that can not be avoided by employing conventional approaches.

In this paper, a modified version of the Hu-Washizu principle is employed that incorporates the rotations of the material lines, i.e., the in plane rotations of a surface at a distance from the midsurface and the rotations of the normal. The kinematical constraints are imposed via nonsymmetric Lagrange multipliers. In order to avoid coupling between flexural and transverse shear modes (and thus additional excessive energy), the shear strains are decomposed into one part, which contains only shear modes and another part, which contains flexural as well as shear modes. The proposed approach offers advantages: We obtain a weak formulation for the rotational equilibrium, which is more consistent than pointwise satisfaction of moment equilibrium. Furthermore, coupling between flexural and transverse shear is avoided. The principle provides a powerful basis for examining - eliminating additional sources of excessive energy and developing means to reduce it. Most important however is the fact that by incorporating the rotations of the normal and by obtaining the corresponding equations for moment equilibrium, different approximations for the shear stresses can be employed. As a result, the aforementioned coupling between torsion and twist bending modes that leads to locking is eliminated. We discuss the selection of the approximations for the various fields and justify the rationale for the choice of these approximations. In this respect, we utilize our previous experience with elemental formulations based on the Hu-Washizu principle (see, e.g., [1] to [5]). Appropriate selection of the approximations for stresses and strains leads to simple and reliable elemental formulations, which minimize excessive energy or are free of locking. Furthermore, extension of the proposed approach to elasto-plastic problems is straightforward. In this case, we assume piecewise constant approximations for stress and strain within subregions instead of linear approximations. This provides a computationally efficient and consistent mechanism for stepwise progression of yielding and reduces significantly the computational effort required for the evaluation of the residual forces. The

convergence characteristics of the proposed formulation are investigated by examining the inf-sup condition and applying the patch test. Finally, results of numerical examples are presented.

Due to space limitations only basic ideas are presented here; details can be found in [5] and in a forthcoming publication [6].

2 GEOMETRIC FEATURES

Geometry, loading, stresses, etc. are referred to an arbitrary reference coordinate system θ^α (α=1, 2) of the surface, while θ^3 is a special coordinate, the distance along the normal. A second, natural coordinate system (ξ^α, ξ^3) is located in every quadrilateral (see Fig. 1). In the subsequent sections, another coordinate system $\left(\bar{\xi}^\alpha, \bar{\xi}^3\right)$ is used with origin through the center of the quadrilateral and coordinate axes parallel to the directions of vectors $\boldsymbol{\rho}_i$. The position vector $\mathbf{r}$ to an arbitrary point is given by the trilinear approximation:

$$\mathbf{r} = {}_0\mathbf{r}_{(\xi^1,\xi^2)} + \xi^3\,\hat{\boldsymbol{\rho}}_{3\,(\xi^1,\xi^2)}, \qquad {}_0\mathbf{r} = \boldsymbol{\rho}_0 + \boldsymbol{\rho}_1\xi^1 + \boldsymbol{\rho}_2\xi^2 + \boldsymbol{\rho}_{12}\xi^1\xi^2, \qquad \hat{\boldsymbol{\rho}}_{3\,(\xi^1,\xi^2)} = \text{const.} \quad (1)$$

Some of the results obtained in [5] concerning the differential geometry of the quadrilateral may be summarized as follows: In each element there is a parallelogram, which forms an intrinsic geometric feature of the quadrilateral (see the parallelogram ABCD illustrated in Fig. 1). The differential geometry of the quadrilateral approaches in the limit that of the aforementioned parallelogram. That parallelogram is a consistent approximation of the quadrilateral. The differential geometry of the quadrilateral is expressed in terms of two geometric quantities (α, β) that provide a measure for the departure of the quadrilateral form from the inherent parallelogram. The vector $\boldsymbol{\rho}_{12}$ of the bilinear term of Eq. (1) represents the departure of the quadrilateral from the shape of the intrinsic parallelogram ABCD; it can be expressed in terms of the constants α and β:

$$\boldsymbol{\rho}_{12} = \beta\,\boldsymbol{\rho}_1 + \alpha\,\boldsymbol{\rho}_2 \quad (2)$$

In order to follow an approach similar to that presented for the quadrilateral and to use the same notations also for the triangle, we do not employ the usual area coordinates to describe the geometry, approximations, etc. but the coordinate system ξ^α shown in Fig. 2. In this case, the vector ${}_0\mathbf{r}$ assumes the form:

$${}_0\mathbf{r} = \mathbf{r}_1 + (\mathbf{r}_2 - \mathbf{r}_1)\xi^1 + (\mathbf{r}_3 - \mathbf{r}_1)\xi^2. \quad (3)$$

Note that by creating the mirror image of the triangle with respect to the side (B-C), we create a parallelogram. In this case, the coordinate system through the center $\left(\bar{\xi}^\alpha, \bar{\xi}^3\right)$ has axes parallel to the coordinate axes (ξ^α, ξ^3) of the triangle.

3 VARIATIONAL FORMULATION

A modified version of the Hu-Washizu principle serves as starting point for the discretization. The modified potential incorporates the internal energy $u_0({}_1\gamma_{\alpha\beta}, {}_0\gamma_{\alpha 3})$ (per unit initial surface) as a function of the flexural ${}_1\gamma_{\alpha\beta}$ and transverse shear ${}_0\gamma_{\alpha 3}$ strains. Since states of strain and stress (e.g., constant states) are defined with respect to the reference coordinate system θ^α, all field variables are initially referred to this system. Furthermore, in keeping with Eq. (1) we introduce two displacement vectors ${}_0\mathbf{V}_{(\xi^1\,\xi^2)}$ and ${}_1\mathbf{V}_{(\xi^1\,\xi^2)}$, which fully determine the position vector $\mathbf{R}$ of a particle after deformation:

$$\mathbf{V} = {}_0\mathbf{V}_{(\xi^1,\xi^2)} + \xi^3\,{}_1\mathbf{V}_{(\xi^1,\xi^2)}, \qquad {}_0\mathbf{V}_{(\xi^1,\xi^2)} = {}_0\mathbf{R}_{(\xi^1,\xi^2)} - {}_0\mathbf{r}, \qquad {}_1\mathbf{V}_{(\xi^1,\xi^2)} = \mathbf{A}_3 - \hat{\boldsymbol{\rho}}_3 \quad (3)$$

Here, ${}_0\mathbf{R}$ denote the position vector to a point on the deformed reference surface and $\mathbf{A}_3$ the base vector tangent to the deformed ξ^3 line. The vector ${}_1\mathbf{V}$ can also be expressed in terms of the rotation of the normal.

To derive all field equations (including moment equilibrium) in discrete form, we also incorporate the in plane rotations $\Omega^1{}_{\alpha\beta}$ of a surface at a distance from the midsurface and the rotations $\Omega^0{}_{\alpha 3}$ of the normal. In order to avoid coupling between flexural and transverse shear modes (and thus additional excessive energy), the kinematical constraints associated with the shear strains are decomposed into one part, which contains only shear modes and another part, which contains flexural as well as shear modes. Those constraints are imposed via Lagrange multipliers $q^{\alpha 3}$ and $q^{\alpha 3}$, respectively. The proposed approach offers advantages: We obtain a weak formulation for the rotational equilibrium, which is more consistent than pointwise satisfaction. Furthermore,

coupling between flexural and transverse shear is avoided. The principle provides a powerful basis for examining and eliminating additional sources of excessive energy and developing means to reduce it. Most important however is the fact that by incorporating the rotations of the normal and by obtaining the corresponding equations for moment equilibrium, different approximations for the shear stresses can be employed. As a result, the aforementioned coupling between torsion and twist bending modes that leads to locking is eliminated.

The kinematical constraints between strains, rotations Ω^{α}_{ij}, and displacement gradients ($\mathbf{q}_i \bullet \mathbf{V}_{,j}$), are imposed via nonsymmetric Lagrange multipliers (tensorial stress components):

$$V = \iint_s \{\, u_0({}_1\gamma_{\alpha\beta}, {}_0\gamma_{\alpha 3}) - m^{\alpha\beta}({}_1\gamma_{\alpha\beta} + \Omega^1_{\alpha\beta} - \mathbf{q}_\alpha \cdot \frac{\partial_1 \mathbf{V}}{\partial \theta^\beta}) \\ - q^{\alpha 3}({}_0\gamma_{\alpha 3} + \Omega^0_{\alpha 3} - \mathbf{q}_\alpha \cdot {}_1\mathbf{V}) - q'^{\alpha 3}({}_0\gamma_{\alpha 3} + \Omega^0_{3\alpha} - \mathbf{q}_3 \cdot \frac{\partial_0 \mathbf{V}}{\partial \theta^\alpha}) + \ldots\} ds + \ldots \tag{4}$$

The functional V is subject to variations of the displacements, strains (${}_1\gamma_{\alpha\beta}$, ${}_0\gamma_{\alpha 3}$), rotations Ω^{α}_{ij}, and the nonsymmetric stresses ($m^{\alpha\beta}$, $q'^{\alpha 3}$, $q^{\alpha 3}$).

4 DISCRETIZATION

Following the isoparametric concept, the two displacement vectors ${}_\alpha\mathbf{V}$ (α=1, 2) of Eq. (4) assume the following form:

$${}_\alpha\mathbf{V} = {}_\alpha\mathbf{V}^0 + {}_\alpha\mathbf{V}^1 \xi^1 + {}_\alpha\mathbf{V}^2 \xi^2 + \underline{{}_\alpha\mathbf{V}^{12}} \xi^1 \xi^2 \tag{5}$$

The underlined bilinear term is not present in case of the triangle. To gain insight into the work expended by the various stress and strain modes, it is useful to inquire as to the physical meaning of the displacement vector. Towards this aim, the displacement vector at a particle Q is expressed in terms of the deformational modes:

$$\mathbf{V} = {}_0\mathbf{V}^0 + \boldsymbol{\rho}^3\left[(\bar{e}^0_{13} - \bar{\omega}^0_{13})\xi^1 + (\bar{e}^0_{23} - \bar{\omega}^0_{23})\xi^2 + \Gamma^1_{33}\xi^1\xi^2\right] + \xi^3 {}_1\mathbf{V} \\ + \xi^3\left\{\boldsymbol{\rho}^1\left[\bar{e}^1_{11}\xi^1 + (\bar{e}^1_{12} + \bar{\omega}^1_{12})\xi^2 + \Gamma^1_{11}\xi^1\xi^2\right] + \boldsymbol{\rho}^2\left[(\bar{e}^1_{12} - \bar{\omega}^1_{12})\xi^1 + \bar{e}^1_{22}\xi^2 + \Gamma^1_{22}\xi^1\xi^2\right]\right\} \tag{6}$$

Thus, the displacement is expressed in terms of the rigid motions of the element, the five constant flexural $(\bar{e}^1_{\alpha\beta})$, and transverse shear $(\bar{e}^0_{\alpha 3})$ strains at the center. Four additional modes are present, which correspond to the bilinear terms of the displacement: Two higher order flexural ($\Gamma^1{}_{11}$, $\Gamma^1{}_{22}$) modes; two additional modes cause transverse shear strain: one mode is warping of the middle surface ($\Gamma^1{}_{33}$) and the other is associated with the linear rotation $\bar{\omega}^1_{12}$. Note that the latter rotation causes the surface ξ^3 = const. to rotate rigidly, relative to the reference surface at $\xi^3 = 0$.

In the following we justify the selection of the approximations for the tensorial stress components (for details see [5], [6]). Recall that the nonsymmetric stress components are defined with respect to the reference system θ^α. It can be shown that independence of affine transformations of the reference system can be achieved without increasing the number of stress parameters, if the approximations are formulated for stress components referred to a special coordinate system and then transformed to the original reference system by employing the usual tensor transformation law for contravariant components[4]. This special coordinate system should be independent of the element orientation with respect to the reference system, or with respect to a coordinate system resulting from an affine transformation of the reference system. Furthermore, this coordinate system should also be an inherent feature of the arbitrary quadrilateral. A coordinate system with the aforesaid properties is the coordinate system $\bar{\xi}^\alpha$ with axes parallel to the sides of the inherent parallelogram through the center of the quadrilateral/triangle. According to our previous experience, the following approximations of stresses are most suitable for parallelogram elements:

$$\bar{s}^{kl} = \bar{\bar{s}}^{kl} + \tilde{\bar{s}}^{kl}(\delta^k_1\delta^l_1\xi^2 + \delta^k_2\delta^l_2\xi^1) \qquad k,l = 1,2 \tag{7}$$

where $\bar{s}^{12} \neq \bar{s}^{21}$. Approximation (7) is valid for components $m^{\alpha\beta}$ *and* $q'^{\alpha 3}$. For the shear components $q^{\alpha 3}$ the following approximation is introduced:

$$\bar{q}^{13} = \bar{\bar{q}}^{13} + \bar{\tilde{q}}\,\xi^2, \qquad \bar{q}^{23} = \bar{\bar{q}}^{23} - \bar{\tilde{q}}\,\xi^1 \tag{8}$$

The form of (8) results from the requirement that the shear forces $q^{\alpha 3}$ should produce no work upon combinations of flexural deformational modes[6]. The components $m^{\alpha\beta}$ *and* $q^{'\alpha 3}$ for the triangle are assumed constant, whereas the form of the approximations for the shear forces $q^{\alpha 3}$ follows from the requirement of zero work upon the corresponding combinations of bending modes.

For the approximation of the tensorial components of strain, a similar approach as that for the stress could be employed. However, in that case the transformation law for the covariant strain components requires evaluation of the inverse expression $\partial\bar{\xi}^k/\partial\vartheta^i$ that can not be readily performed or may introduce complicated expressions. Therefore, we introduce approximations for the tensorial strain components ε_{ij} that are referred to the directions of the reference system. These assumptions must ensure that the element properties remain invariant in case of an affine transformation of our reference system. An approximation that preserves invariance assumes the following form:

$$\varepsilon_{ij} = \varepsilon_{ij}^0 + \varepsilon_{ij}^1\xi^1 + \varepsilon_{ij}^2\xi^2 \quad (\varepsilon_{12} = \varepsilon_{21}) \tag{9}$$

These assumptions contain 15 strain parameters though our displacement approximation allows only 9 independent deformational modes. For this reason, additional conditions must be employed in order to reduce the number of independent parameters. Since appropriate approximations of strain that avoid shear locking in the case of a parallelogram are well known (see, e.g., [1] to [6]), we employ again as a starting point the intrinsic parallelogram. The keys to our strain approximation were: (i) the identification of the constant and higher-order deformational terms of the displacement approximation, (ii) the realization that the higher-order terms reappear in different strain components, and (iii) our approximations of the strains need not retain the higher-order terms in two different strain components; they are needed only to inhibit a mode. Suppressing such terms in one, or the other component, serves to reduce the internal energy and improve convergence. Thus, according to our previous experience, the following approximations are introduced for the strain components $\bar{\varepsilon}_{kl}$ of the parallelogram in the directions of the coordinate axes $\bar{\xi}^\alpha$:

$$\bar{\varepsilon}_{kl} = \bar{\bar{\varepsilon}}_{kl} + \bar{\tilde{\varepsilon}}_{kl}(\delta_1^k\delta_1^l\xi^2 + \delta_2^k\delta_2^l\xi^1) \qquad (k,l=1,2), \quad \bar{\bar{\varepsilon}}_{12} = \bar{\bar{\varepsilon}}_{21} \tag{10}$$

Note these assumptions contain only 9 independent parameters. As next step, we express the 15 parameters ε_{ij}^k of the approximations (9) by the 9 parameters of assumptions (10). Towards this aim, we equate the work done by the stress components $\bar{s}^{ij}$ and strains $\bar{\varepsilon}_{ij}$ upon the parallelogram with the work expended by the same stress components, $\bar{s}^{ij}$, upon the strains $\bar{\varepsilon}_{ij}^{/}$. The latter strains are referred to the directions of the parallelogram sides but are obtained from the strain components ε_{kl} using the transformation law for covariant tensor components:

$$\iint_{s_p} \bar{s}^{ij}\bar{\varepsilon}_{ij}A\,d\xi^1\,d\xi^2 = \iint_{s_p} \bar{s}^{ij}\bar{\varepsilon}'_{ij}A\,d\xi^1\,d\xi^2 \qquad \bar{\varepsilon}'_{ij} = \frac{\partial\theta^k}{\partial\bar{\xi}^i}\frac{\partial\theta^l}{\partial\bar{\xi}^j}\varepsilon_{kl} \tag{11}$$

For the quadrilateral as well as the triangle, the rotations $\Omega^1{}_{\alpha\beta}$ are approximated by constants. The rotation of the normal $\Omega^0{}_{\alpha 3}$ is assumed constant within the triangular element whereas the corresponding rotation for the quadrilateral takes a form similar to that of Eq. (8).

4 DERIVATION OF STIFFNESS MATRIX

With the stress and strain approximations previously presented, functional (4) assumes the form:

$$V = \iint_s \{\, u_0(\overline{{}_1\gamma_{\alpha\beta}}, \overline{{}_0\gamma_{\alpha 3}}) - \overline{m}^{\alpha\beta}(\overline{{}_1\gamma_{\alpha\beta}} + \overline{\Omega}^1_{\alpha\beta} - P^\gamma_\beta \boldsymbol{\rho}_\alpha \cdot \frac{\partial_1 \mathbf{V}}{\partial \xi^\gamma}) \\ - \overline{q}^{\alpha 3}(\overline{{}_0\gamma}_{\alpha 3} + \overline{\Omega}^0_{\alpha 3} - \boldsymbol{\rho}_\alpha \cdot {}_1\mathbf{V}) - \overline{q'}^{\alpha 3}(\overline{{}_0\gamma}_{\alpha 3} + \overline{\Omega}^0_{3\alpha} - P^\gamma_\alpha \boldsymbol{\rho}_3 \cdot \frac{\partial_0 \mathbf{V}}{\partial \xi^\gamma}) + \ldots\} ds + \ldots \tag{12}$$

Furthermore, variation of $\Omega^0_{\alpha 3}$ yields the equations of moment equilibrium:

$$\begin{aligned} &\overline{\overline{q}}^{13} + \frac{1}{3}\beta \widetilde{\overline{q}}^{13} = \overline{\overline{q}}'^{13} + \frac{1}{3}\beta \widetilde{\overline{q}}'^{13} \\ &\overline{\overline{q}}^{23} - \frac{1}{3}\alpha \widetilde{\overline{q}}^{13} = \overline{\overline{q}}'^{23} + \frac{1}{3}\alpha \widetilde{\overline{q}}'^{23} \\ &\beta \overline{\overline{q}}^{13} - \alpha \overline{\overline{q}}^{23} + 2\widetilde{\overline{q}}^{13} = \beta \overline{\overline{q}}'^{13} - \alpha \overline{\overline{q}}'^{23} + \widetilde{\overline{q}}'^{13} - \widetilde{\overline{q}}'^{23} \end{aligned} \tag{13}$$

In general, the constants α and β are nonzero. It follows from the last terms of Eq. (12) that there is a coupling between higher order and constant deformational modes. This coupling contributes to an increase in locking since the constant strain modes also expend work upon the constant stress modes. This locking is due to the presence of the geometrical constants α and β. Since the values of these constants increase with increasing departure from the shape of the parallelogram (or rectangle), there is a gradual increase in the amount of locking due to the shape distortion of the quadrilateral. Unfortunately, these terms can not be omitted, if the element is to pass the classical patch test. Note that, in the limit, the constants α and β approach zero and, therefore, this kind of locking disappears. Furthermore, there is no "parallelogram locking" since the aforesaid terms vanish due to α=β=0 and the usual shear locking is prevented by using the Hu-Washizu principle and appropriate assumptions for stresses and strains. By analytically performing the integrations and varying the independent variables, the corresponding stiffness matrices for the bending plate quadrilateral and triangle are obtained.

The approximations employed preserve invariance under affine transformations, minimize excessive energy, and lead to an element formulation that is characterized by simplicity and reliability

In case of an elastoplastic analysis, the continuous, linear functions are replaced by the Heaviside function. As a result, the strains and stresses have piecewise constant approximations within subregions (Figs. 3 and 4). This provides a consistent mechanism for describing progressive yielding through the element and leads to a significant reduction of computational cost. Furthermore, reevaluation of the stiffness matrix at successive increments requires no additional integrations over the elemental area. Also, evaluation of residual/internal forces is simple. Consequently, the correction of imbalances is readily achieved via the incremental-iterative process

4 CONVERGENCE CHARACTERISTICS

In order to investigate the convergence characteristics of the proposed elemental formulation, we follow Strang's[7] convergence criterium: ``Stability and consistency imply convergence." The convergence characteristics of the proposed formulation are investigated by examining the inf-sup condition and applying the patch test. Employing the ``Rayleigh nesting theorem," we determine the lowest eigenvalue of each individual element instead to determine the lowest eigenvalue of the entire system. Furthermore, to show that the inf-sup condition is satisfied, we also apply the singular-value decomposition. Due to space limitations, we do not present details of the investigations; instead the reader is referred to [5] and [6].

10 NUMERICAL EXAMPLES

The sensitivity of the formulation to distortions of the element geometry is tested using two elements to discretize the beam under bending depicted in Fig. 5 (E=1500, ν=0.25, thickness t=1). The left edge of the arrangement is clamped and at the opposite side two vertical loads of magnitude 1 are applied. The distortion parameter e is a measure for the departure of the trapezoidal form from that of the rectangle. By employing this test, the sensitivity of the formulation to mesh distortions should be investigated. Furthermore, this test should demonstrate the amount of increase of the excessive energy as the element geometry departures from that of the rectangle, i.e., as the parameter e increases. Note that the proposed as well as most of the alternative formulations eliminate shear locking in the case of a rectangle. Therefore, the results presented here are referred to the results obtained for the rectangle, i.e., the results shown in Fig. 6 denote the ratios of the displacements

obtained for various values of the distortion parameter ($e \neq 0$) to the corresponding displacement of the rectangle ($e = 0$).

The skew cantilever plate (L=100, t=4, E=100, v=0.3) depicted in Figure 7 has one clamped side and is subjected to a uniformly distributed transverse load p=1. For a 4x4 and 8x8 mesh and for angles β=30º and β=70º, the ratio of the exact solution to the result (using quadrilaterals) obtained for the vertical displacement at B is given in Table 1. Results for the triangles are presented in [6].

11 CONCLUSIONS

A modified version of the Hu-Washizu principle is employed that incorporates the rotations of the material lines, i.e., the in plane rotations of a surface at a distance from the midsurface and the rotations of the normal. The kinematical constraints are imposed via nonsymmetric Lagrange multipliers. In order to avoid coupling between flexural and transverse shear modes (and thus additional excessive energy), the shear strains are decomposed into one part, which contains only shear modes and another part, which contains flexural as well as shear modes. The proposed approach offers a series of advantages: We obtain a weak formulation for the rotational equilibrium, which is more consistent than pointwise satisfaction. Furthermore, coupling between flexural and transverse shear is avoided. The principle provides a powerful basis for examining - eliminating additional sources of excessive energy and developing means to reduce it. Most important however is the fact that by incorporating the rotations of the normal and by obtaining the corresponding equation for moment equilibrium, different approximations for the shear stresses can be employed. As a result, the aforementioned coupling between torsion and twist bending modes that leads to locking is eliminated.

We discuss the selection of the approximations for the various fields and justify the rationale for the choice of these approximations. Appropriate selection of the approximations for stresses and strains leads to simple and reliable elemental formulations, which minimize excessive energy or are free of locking. Furthermore, extension of the proposed approach to elasto-plastic problems is straightforward. In this case, we may assume piecewise constant approximations for stress and strain within subregions instead of linear approximations. This provides a computationally efficient and consistent mechanism for stepwise progression of yielding and reduces significantly the computational effort required for the evaluation of the residual forces. The convergence characteristics of the proposed formulation are investigated by examining the inf-sup condition and applying the patch test. Finally, results of numerical examples are presented.

ACKNOWLEDGEMENTS

Financial support of the European Union and the General Secretary for Research and Technology/Greece in the framework of the Research Program ``Heraclitus'' (Grant No. 50) is gratefully acknowledged.

REFERENCES

[1] Talaslidis, D. and Wempner, G. (1982), "A simple finite element for elastic-plastic deformations of shells," Comput. Methods Appl. Mech. Engrg. 103, pp. 375-397.

[2] Talaslidis, D. and Wempner, G. (1993), "The linear isoparametric triangular element: Theory and application," Comput. Methods Appl. Mech. Engrg. 103, pp. 375-397.

[3] Karakostas , C., Talaslidis, D. and Wempner, G. (1993), "Triangular C^0 bending elements based on the Hu-Washizu principle and orthogonality conditions," Internat. J. Numer. Methods Engrg. 36, pp. 181-200.

[4] Wempner, G., Talaslidis, D. (2002), *Mechanics of Solids and Shells: Theory and Approximations,* CRC Press, Boca Raton.

[5] Paraskevopoulos, E. and Talaslidis, D. (2005), "Reduction of excessive energy in the four-noded membrane quadrilateral element, Part I: Linear theory-compressible materials," Comput. Methods Appl. Mech. Engrg., in press.

[6] Paraskevopoulos, E. and Talaslidis, D. (2005) "Simple and efficient quadrilateral and triangular plate bending elements based on the Hu-Washizu principle," in preparation.

[7] Strang, G. and Fix, G.J. (1973), An Analysis of the Finite Element Method, Prentice-Hall, Englewood Cliffs.

Angle β	4x4 mesh	8x8 mesh	exact
β = 30º	2113,77	2556,03	2440,63
β = 70º	17063,88	16247,77	16315,63

Table 1: Skew cantilever plate: Displacement at B

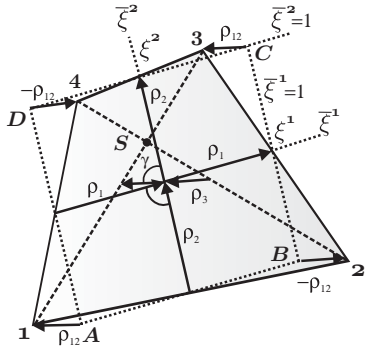


Figure 1. Geometric features of the quadrilateral

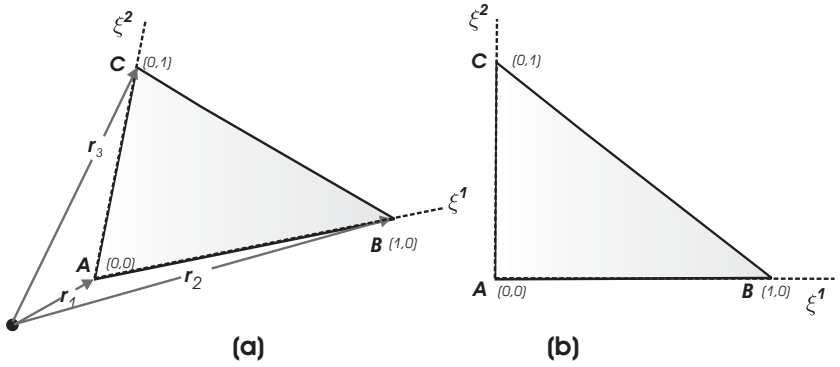


Figure 2. Geometric features of the triangle

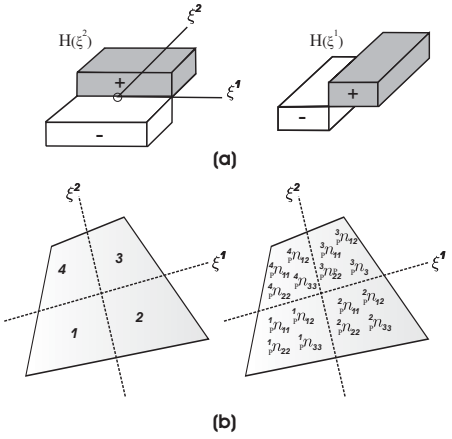


Figure 3. Quadrilateral element: Piecewise constant approximation

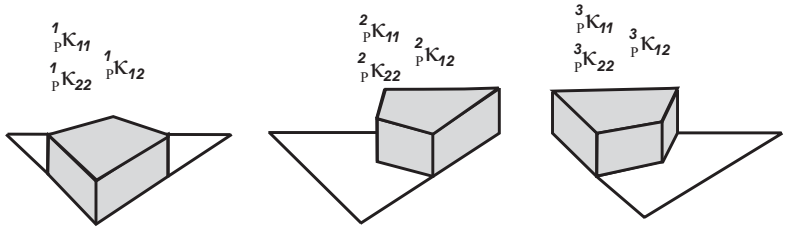


Figure 4. Triangular element: Piecewise constant approximation

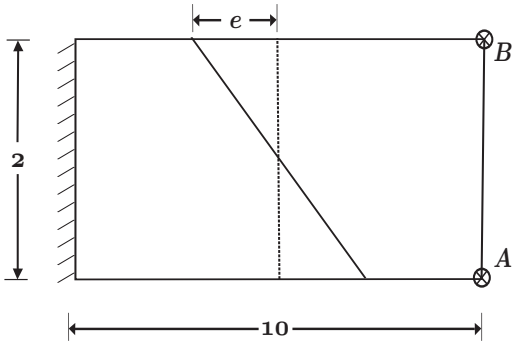


Figure 5. Mesh distortion test (Bending)

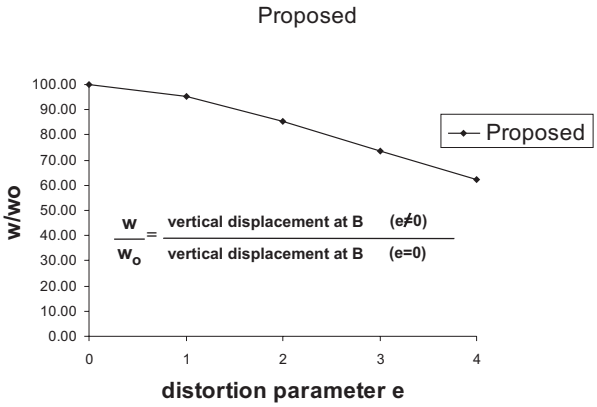


Figure 6. Distortion test: Influence of distortion on results

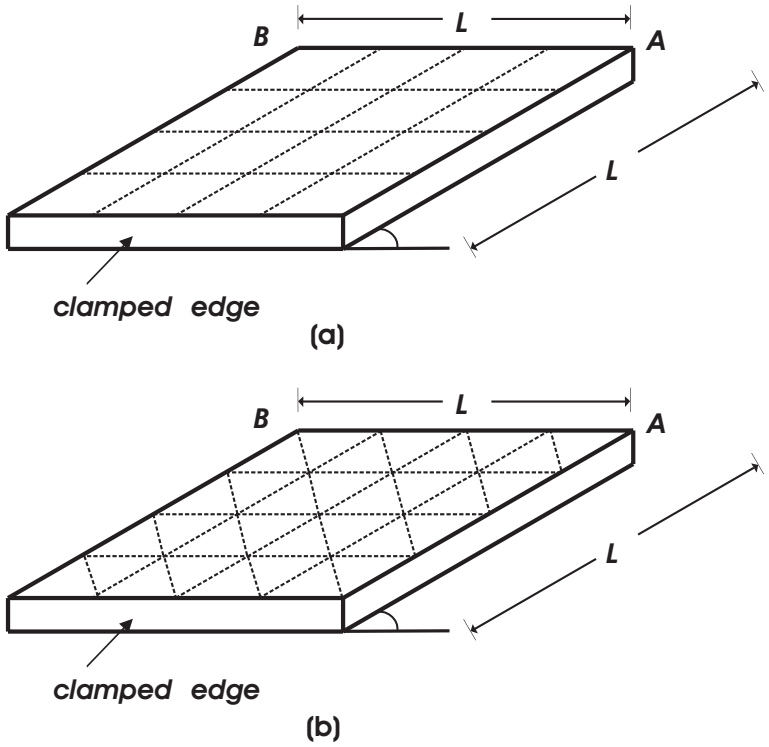


Figure 7. Skew cantilever plate

LOCKING REDUCTION IN 2D QUADRILATERAL ELEMENTS: COROTATIONAL FORMULATION, NEAR INCOMPRESSIBILITY, AND J_2 PLASTICITY

Elias A. Paraskevopoulos and Demosthenes G. Talaslidis

Department of Civil Engineering
Aristotle University Thessaloniki
Mail Stop 502, 54124 Thessaloniki, Greece
e-mail: talaslid@civil.auth.gr

Keywords: membrane quadrilateral, excessive energy, modified Hu-Washizu functional, corotational formulation, near incompressibility, J_2 plasticity.

Abstract. *The methodology presented in [1] is extended to geometrically nonlinear problems, to problems with nearly incompressible materials, and to J_2 plastic flow-problems. A rigorous continuum mechanics basis to the corotational framework is provided and a corotational kinematical model for geometrically nonlinear analysis is developed. In the nearly incompressible case, sources of excessive energy for volumetric as well as deviatoric deformational states are examined and conclusions are drawn concerning the optimal choice of approximations for the field variables. A modified version of the Hu-Washizu principle is employed that incorporates part of the plain strain conditions. In case of J_2 plasticity, the current formulation incorporates an important modification that leads to further simplifications: The continuous, linear functions employed for the approximations are replaced by the Heaviside function. Finally, results of numerical examples and comparisons with other formulations are presented.*

1 INTRODUCTION

In a recent paper[1], reasons for the appearance of excessive energy are examined and a methodology is presented, which minimizes excessive energy and yields simple and effective elements that are less sensitive to distortions of the element shape. The purpose of this contribution is to extend the approach presented in [1] to geometrically nonlinear problems based on a corotational formulation, to problems with nearly incompressible materials, and to J_2 plastic flow-problems. Towards this aim, a rigorous continuum mechanics basis to the corotational framework is provided and a corotational kinematical model for geometrically nonlinear analysis is developed. In case of nearly incompressible materials, conclusions are drawn concerning the optimal choice of approximations for the field variables by examining sources of excessive energy for volumetric and deviatoric deformational states. A modified version of the Hu-Washizu principle is employed that utilizes the orthogonality between deviatoric and volumetric terms and incorporates part of the plain strain conditions. In deriving the weak form of the governing equations for the J_2 plasticity problem, attention is focused on a straightforward extension of the linear problem without explicit reference to additional postulates. In case of J_2 plasticity, the current formulation incorporates an important modification that leads to further simplifications: The continuous, linear functions employed for the approximations are replaced by the Heaviside function. Finally, results of numerical examples and comparisons with other formulations are presented.

In this paper, a consistently derived corotational formulation is presented that avoids ad hoc techniques, unjustified assumptions, and problems inherent in alternative formulations. The concept of decomposition of motion for a finite region emanates from that for an infinitesimal region. The modified principle of Hu-Washizu employed here depends on the position, the engineering strains and associated symmetric and antisymmetric stresses, but also incorporates the rotation of the base vectors. Approximation of this rotation at every point by the same physical rotation yields a series of advantages. Furthermore, differentiation of the variational principle leads to a consistent symmetric tangent stiffness matrix. For the approximation of stress and strain we follow the rationale and the procedure presented in [1]. The approximations employed preserve invariance under affine transformations, minimize excessive energy, and lead to an element formulation that is characterized by simplicity and reliability. Furthermore, reevaluation of the stiffness matrix at successive increments requires no additional integrations over the elemental area. Also, evaluation of residual/internal forces is simple. Consequently, the correction of imbalances is readily achieved via the incremental-iterative process.

A further objective of the present paper addresses to the reliable treatment of nearly incompressible behavior that characterizes a series of materials (e.g., rubber or rubber-like materials) in linear stress analysis or in the

nonlinear regime (e.g., plastic behavior of von Mises elastic-plastic materials). In those cases, low order elements show poor performance, which may be attributed to shortcomings of the formulation and sources of superfluous energy. This part of the paper aims to examine sources of excessive energy in case of volumetric as well as deviatoric deformations of nearly incompressible materials. The results of those investigations serve to develop a methodology for identifying and eliminating sources of excessive energy and to select appropriate approximations for the various field variables. Again, the generalized principle of Hu-Washizu serves as a basis for the discretization. However, departing from the usual pointwise satisfaction of all plane strain conditions, the constraint concerning vanishing of the transverse deformation is taken into account in weak form, whereas vanishing of the transverse shear strain is satisfied pointwise. Such an approach allows the use of concepts presented in [1] and proves most useful in minimizing excessive energy and in leading to simple elemental formulations. Furthermore, utilizing the orthogonality between volumetric and deviatoric parts, the functional is transformed to a form that offers advantages with regard to the optimal choice of stress and strain approximations and yields the most appropriate relations between discrete strain parameters and displacement modes. Departing from common practice, we prove analytically that the patch test is satisfied. The test is also applied numerically to a patch of elements under different loading conditions.

The last part of the present study deals with the extension of the formulation to the analysis of plane strain problems assuming a J_2 flow theory with linear isotropic/kinematic hardening. In deriving the weak form of the governing equations, attention is focused on a straightforward extension of the linear problem without reference to additional postulates. Furthermore, to avoid problems related to the nonconservative character of the formulation, we follow the original approach suggested by Lagrange for handling auxiliary conditions. The methodology employed in the previous sections serves as a basis for the approximation of the displacement, total strain, stress deviator, elastic pressure, and volumetric strain in case of J_2 plasticity. However, the current formulation incorporates an important modification that leads to further simplifications: The continuous, linear functions are now replaced by the Heaviside function. As a result, the strains and stresses have piecewise constant approximations within subregions. This provides a consistent mechanism for describing progressive yielding through the element and leads to a significant reduction of computational cost. Arguments for the choice of the plastic strain approximations are given and the assumptions concerning the plastic strain deviators as well as the consistency or plastic parameter are presented.

Finally, results of numerical examples and comparisons with other formulations are presented. Due to space limitations only the essential concepts and results are presented here; additional details can be found in [1] and [2].

2 GEOMETRIC FEATURES

Some of the results obtained in [1] concerning the differential geometry of the quadrilateral may be summarized as follows: In each element there is a parallelogram, which forms an intrinsic geometric feature of the quadrilateral (see the parallelogram ABCD illustrated in Fig. 1a). The differential geometry of the quadrilateral approaches in the limit that of the aforementioned parallelogram. That parallelogram is a consistent approximation of the quadrilateral. The differential geometry of the quadrilateral is expressed in terms of two geometric quantities (α, β) that provide a measure for the departure of the quadrilateral form from the inherent parallelogram.

3 COROTATIONAL FORMULATION

Three coordinate systems are introduced in order to describe the motion of the element (see Fig. 2): (i) A fixed, global reference coordinate system (X^α) as a common reference basis for the computations. (ii) A convected, global reference coordinate system (θ^α). All governing equations are expressed in terms of field variables referred to this system. (iii) Finally, a convected, local (natural) coordinate system (ξ^α) located within the element. The motion of a continuous medium in the neighborhood of a particle can be decomposed into a rigid motion followed by a deformation (or vice-versa)[3]. The concept of decomposition of motion, which applies to an infinitesimal region can also be applied to a finite region. In particular, the motion of a finite region can be conceived as a finite rigid rotation followed by a small relative deformation (see Fig. 3).

As a starting point for the discretization, a modified version of the Hu-Washizu principle is employed. The principle is dependent on the position $\boldsymbol{R}$, the engineering strains $h_{\alpha\beta}$, and associated tensorial stress components $t^{\alpha\beta}$ but also on the rotation of base vectors $\mathbf{q}'_\beta$ .Variation with respect to the field variables yields:

$$\delta V^* = \iint_s \left\{ \begin{array}{l} \delta h_{\alpha\beta}\left[C^{\alpha\beta\gamma\delta}\, h_{\gamma\delta} - t_s^{\alpha\beta}\right] - \delta t_s^{\alpha\beta}\left[h_{\alpha\beta} - \frac{1}{2}\left(\mathbf{q}'_\alpha \cdot \mathbf{R}_{,\beta} + \mathbf{q}'_\beta \cdot \mathbf{R}_{,\alpha}\right) + \mathbf{q}_{\alpha\beta} \right] \\ + \delta t_\alpha^{\alpha\beta}\left[\frac{1}{2}\left(\mathbf{q}'_\beta \cdot \mathbf{R}_{,\alpha} - \mathbf{q}'_\alpha \cdot \mathbf{R}_{,\beta}\right)\right] + \delta \mathbf{R}_{,\alpha} \cdot \left(t^{\alpha\beta}\mathbf{q}'_\beta\right) + \delta\overline{\omega}^3 e_{3\alpha\gamma}\left(t^{\beta\alpha}\mathbf{q}'^{\gamma} \cdot \mathbf{R}_{,\beta}\right) + .. \end{array} \right\} ds + .. \quad (1)$$

From the last term of Eq. (1), two alternative conditions of rotational equilibrium associated with the rigid rotation are obtained. The modified principle employed here and the incorporation of the rotation of vectors $\mathbf{q}'_\alpha$ as independent variable lead to a consistent, variationally derived corotational formulation, which avoids problems inherent in alternative approaches. Furthermore, the proposed principle serves to eliminate sources of excessive energy. A consistent tangent stiffness matrix is obtained by differentiating (1). By rearranging terms, the following consistent linearization is obtained:

$$d\delta V^* = \iint_s \left\{ \begin{array}{l} \delta h_{\alpha\beta}\left[C^{\alpha\beta\gamma\delta}\, dh_{\gamma\delta} - dt_s^{\alpha\beta}\right] \\ - \delta t_s^{\alpha\beta}\left[dh_{\alpha\beta} - \frac{1}{2}\left(d\mathbf{q}'_\alpha \cdot \mathbf{R}_{,\beta} + d\mathbf{q}'_\beta \cdot \mathbf{R}_{,\alpha} + \mathbf{q}'_\alpha \cdot d\mathbf{R}_{,\beta} + \mathbf{q}'_\beta \cdot d\mathbf{R}_{,\alpha}\right)\right] \\ + \delta t_\alpha^{\alpha\beta}\left[\frac{1}{2}\left(d\mathbf{q}'_\beta \cdot \mathbf{R}_{,\alpha} - d\mathbf{q}'_\alpha \cdot \mathbf{R}_{,\beta} + \mathbf{q}'_\beta \cdot d\mathbf{R}_{,\alpha} - \mathbf{q}'_\alpha \cdot d\mathbf{R}_{,\beta}\right)\right] \\ + \delta \mathbf{R}_{,\alpha} \cdot \left(dt_s^{\alpha\beta}\mathbf{q}'_\beta + dt_\alpha^{\alpha\beta}\mathbf{q}'_\beta + t_s^{\alpha\beta} d\mathbf{q}'_\beta + t_\alpha^{\alpha\beta} d\mathbf{q}'_\beta\right) \\ + \delta \mathbf{q}'_\beta \cdot \left(dt_s^{\alpha\beta}\mathbf{R}_{,\alpha} + dt_\alpha^{\alpha\beta}\mathbf{R}_{,\alpha} + t_s^{\alpha\beta} d\mathbf{R}_{,\alpha} + t_\alpha^{\alpha\beta} d\mathbf{R}_{,\alpha}\right) \\ + d\delta \mathbf{q}'_\beta \left(t_s^{\alpha\beta}\mathbf{R}_{,\alpha} + t_\alpha^{\alpha\beta}\mathbf{R}_{,\alpha}\right) + ... + \end{array} \right\} ds + ... \quad (2)$$

The position vector of a particle before and after deformation, ***r*** and ***R***, respectively, are approximated by bilinear polynomials. Contrary to other alternative corotational formulations, we do not split the total displacement $\boldsymbol{V_{tot}}$ into relative displacements due to deformation and displacements due to rigid body motion. Instead, we use in relationship (1) the vector $\boldsymbol{R_{,\alpha}}$. Such an approach is computationally more efficient and leads to reduced computational effort. In order to satisfy consistency of the formulation and to be able to describe a constant rigid rotation, vectors $\mathbf{g}'_\alpha$ at every particle within the element are assumed to be rotated with the same physical rotation. Since in the limit we approach the center of the element it is more natural to take the rotation associated with the center of the element instead of the rotation of the continuum theory that changes at each point.

To obtain appropriate approximations for the tensorial stress components $t^{\alpha\beta}$ referred to the convected reference system, the procedure presented in [1] can be followed. The stress components $\bar{t}^{\alpha\beta}$ in directions of the coordinate axes ξ^α are approximated by:

$$\bar{t}^{kl} = \bar{\bar{t}}^{kl} + \tilde{\bar{t}}^{kl}(\delta_1^k \delta_1^l \xi^2 + \delta_2^k \delta_2^l \xi^1) \qquad k,l = 1,2 \quad (3)$$

The antisymmetric part of the stress is obtained from the weak form of the rotational equilibrium. For the approximation of the engineering strains we also follow the procedure in [1], Subsection 4.3. Those approximations preserve invariance under affine transformations and minimize excessive energy.

By inserting the aforesaid approximations into the variational principle (1), enforcing rotational equilibrium, and performing the integrations analytically, the surface integral in (1) yields a consistent tangent stiffness matrix. Reevaluation of the stiffness matrix at successive increments of deformation (and load) requires no

additional integrations over the elemental area. In this respect, the element is simpler than alternative comparable formulations. Also, evaluation of residual/internal forces is simple and requires no numerical integrations. Consequently, the correction of imbalances is readily achieved via the incremental-iterative process. Furthermore, invariance of the element formulation is preserved, excessive energy is minimized, and the element formulation is characterized by simplicity and reliability. The reader is referred to the remarks in Section 5 of Ref. [1] concerning elimination of excessive energy and coupling between constant and higher modes in the case of an arbitrary quadrilateral.

4 NEAR INCOMPRESSIBILITY

As a first step, reasons for the appearance of excessive energy are investigated. The results of those investigations serve to develop a methodology for the selection of appropriate approximations for the various field variables and to justify the rationale for choosing these approximations. The conclusions drawn here concern sources of excessive energy in the case of volumetric as well as deviatoric deformations under plane strain conditions. Notice that many earlier publications confine the investigations mainly to volumetric deformations and consider only rectangular elements. For the case of isochoric deformations it is shown[2] that the volumetric strain should have a constant distribution over the element and the approximation for the dilatational part should avoid coupling between constant and higher order modes. Following concepts which have been proved successful in deriving elemental approximations[3], we conclude in case of a deviatoric deformation that coupling, and thus excessive energy, appears not only between the deviatoric parts associated with the in plane bending and shear strains but also between the in plane bending strain components itself. The proposed approach circumvents the aforementioned flaws and minimizes excessive energy.

Again, the generalized principle of Hu-Washizu serves as a basis for the discretization. However, departing from the usual pointwise implementation of the plane strain conditions, $\varepsilon_{33}=\varepsilon_{\alpha 3}=0$, the constraint $\varepsilon_{33}=0$ is taken into account in weak form, whereas $\varepsilon_{\alpha 3}=0$ is satisfied pointwise. The motivation for such an approach is presented in [2]. Subdividing the strain energy into deviatoric and dilatational parts and considering the orthogonality conditions between volumetric and deviatoric parts, the modified Hu-Washizu principle employed here assumes the form:

$$\iint_s \left\{ G\left(\eta_\kappa^\alpha \eta_\alpha^\kappa + \eta_3^3 \eta_3^3\right) + \frac{K}{2} \varepsilon_\alpha^\alpha \varepsilon_\kappa^\kappa + \sigma^{\alpha\beta}\left(\varepsilon_{\alpha\beta} - D_{\alpha\beta}(\mathbf{u})\right) + \sigma^{33}\left(\varepsilon_{33}\right) + p\left(\varepsilon_v - D^v(\mathbf{u})\right) \right\} ds + \ldots \qquad (4)$$

Here, ε_v, ε_{ij} , η_{ij}, and σ^{ij} denote the volumetric strain, the components of the total, deviatoric strains, and deviatoric stresses, respectively. For the approximation of the strain and stress components (total and deviatoric), we follow to some extend the methodology presented in [1] and the results and conclusions of the previous sections. It should be noted that the approximation for the deviatoric stress σ^{33} is not independent from that of the other components. The proposed approach leads to simple stiffness matrices, which are derived in a consistent and straightforward manner without recourse to intermediate procedures or computational schemes.

The convergence properties of the formulation are investigated by examining stability and consistency. Departing from common practice, we prove analytically[2] that the patch test is satisfied. Furthermore, it is concluded that terms, which may-in case of an arbitrary quadrilateral-cause some excessive energy can not be omitted, if we want to pass the classical patch test. Finally, the test is applied numerically to a patch of elements under different loading conditions[2].

5 J_2 PLASTICITY

In deriving the weak formulation, we focus attention on a straightforward extension of the linear problem without explicit reference to additional postulates or principles. Furthermore, to avoid problems related to the nonconservative character of the formulation and the existence of a functional and to keep the derivation general, we follow the original approach suggested by Lagrange for handling auxiliary conditions[5]. The weak formulation takes the form:

$$
\begin{aligned}
&\iint_s \left\{ \begin{array}{l} \dot{\eta}_\kappa^\alpha \, 2G \, \delta\eta_\alpha^\kappa + \dot{\eta}_3^3 \, 2G \, \delta\eta_3^3 + \dot{\varepsilon}_\alpha^\alpha \, \dfrac{K}{2} \delta\varepsilon_\kappa^\kappa \\ -{}_p\dot{\eta}_\kappa^\alpha \, 2G \, \delta\,{}_p\eta_\alpha^\kappa - {}_p\dot{\eta}_3^3 \, 2G \, \delta\,{}_p\eta_3^3 - \dot{\alpha}_i \, \overline{C}^{ij} \, \delta\alpha_j - \dot{\mathbf{f}} \cdot \delta\mathbf{R} \end{array} \right\} ds \\
&- \iint_s \left\{ \dot{\sigma}^{\alpha\beta} \left(\delta\varepsilon_{\alpha\beta} - D_{\alpha\beta}(\delta\mathbf{u}) \right) + \dot{\sigma}^{33} \delta\varepsilon_{33} + \dot{p} \left(\delta\varepsilon_v - D^v(\delta\mathbf{u}) \right) \right\} ds \\
&- \iint_s \left\{ \dot{\lambda} \left[\frac{\partial \mathrm{Y}}{\partial(\eta_{ij} - {}_p\eta_{ij})} \left(\delta\eta_{ij} - \delta\,{}_p\eta_{ij} \right) + \frac{\partial \mathrm{Y}}{\partial(\alpha_i)} \left(\delta\alpha_i \right) \right] \right\} ds \qquad (5) \\
&- \iint_s \left\{ \delta\lambda \left(\mathrm{Y} \right) \right\} ds \\
&- \iint_s \left\{ \delta\sigma^{\alpha\beta} \left(\dot{\varepsilon}_{\alpha\beta} - D_{\alpha\beta}(\dot{\mathbf{u}}) \right) + \delta\sigma^{33} \left(\dot{\varepsilon}_{33} \right) + \delta p \left(\dot{\varepsilon}_v - D^v(\dot{\mathbf{u}}) \right) \right\} ds + \ldots .
\end{aligned}
$$

The methodology employed in the previous sections[1],[2] serves as a basis for the approximation of the displacement, total strain, stress deviator, elastic pressure, and volumetric strain in case of J_2 plasticity. However, the current formulation incorporates an important modification that leads to further simplifications: The continuous, linear functions are now replaced by the Heaviside function $H(\xi^\alpha)$ (see Fig. 4). As a result, the strains and stresses have piecewise constant approximations within subregions. This provides a consistent mechanism for describing progressive yielding through the element and leads to a significant reduction of computational cost. Arguments for the choice of approximations for the plastic strain are given; they are approximated by different values within each subdomain. The hardening parameters and the plastic parameter are also approximated by constants within each subregion.

6 NUMERICAL EXAMPLES

The first numerical example presents a geometrically nonlinear analysis of the cantilever beam (Elasticity Modulus =120000) depicted in Fig. 5. Results for various element formulations and for mesh subdivisions are shown in Table 1 for the horizontal and vertical components of the displacement (mean value between upper and lower displacement) at the free end of the beam. For comparison reasons, results obtained using ten beam elements (denoted as ``Beam'' in Table 1) are also given. For the regular meshes shown in Figs. 5a,b the results are compared with those obtained for 4-node and 9-node elements that employ enhanced gradient displacement modes[6]. The 4-node (9-node) element [denoted as QR4/I4 (QR9/I9)] uses bilinear (biquadratic) shape functions for the interpolation of displacements and rotations. The corresponding elements without incompatible modes are denoted as QR4 and QR9. Results (indicated as ``distorted'') are also presented for the distorted mesh illustrated in Fig. 5c. As expected the element exhibits a performance that is less sensitive to distortions of the element shape. Notice that the proposed formulation does not employ incompatible modes or additional degrees of freedom, such as rotations, but is based solely on bilinear approximations for the nodal displacements.

The swept panel with uniformly distributed load along the right side (known as Cook's membrane problem) of Fig. 6 has been analyzed assuming an elasticity modulus of E=250 and a Poisson's ratio of ν=0.49999995. Results (horizontal and vertical displacement at A) for various meshes and various element formulations are shown in Table 2. Elemental formulations including ``incompatible modes''[7] and elements employing enhanced strain fields[8] yield similar results. Those results are shown in the column entitled ``Enhanced.'' The columns furnished with the caption ``B-bar'' contain results obtained by the B-bar method[9].

The swept panel of Fig. 6 with uniformly distributed load (resultant force: F=1.00) along the right side has also been analyzed assuming J_2 flow theory with linear isotropic/kinematic hardening. The values of the material constants are Elasticity modulus (E=70), Poisson's ratio ν=1/3), yield stress σ_Y=0.243), isotropic (K=0.015) and kinematic hardening modulus (H=0.135). Results (horizontal and vertical displacement at top corner) for various meshes and element formulations are shown in Table 3. With respect to the local integration of the constitutive equations, a return-mapping algorithm in combination with a backward-Euler scheme has been employed[4]. For the global system of equations, a modified Newton-Raphson method has been used.

7 CONCLUSIONS

A consistently derived corotational formulation is presented that avoids ad hoc techniques and unjustified assumptions. The modified principle of Hu-Washizu employed here depends on the position, the engineering strains and associated symmetric and antisymmetric stresses but also incorporates the rotation of the base vectors. The variationally consistent corotational formulation thus derived avoids problems inherent in alternative formulation. Furthermore, by approximating this rotation at every point by the same physical rotation, a series of advantages emerges: (i) Consistency of the formulation and thus convergence of the discrete equations to their differential counterparts is satisfied. (ii) The proposed assumption is the simplest possible and leads to simple elemental formulations. (iii) Considering the behavior of the element in the limit, the proposed assumption for the rotation appears more natural then alternative choices. For the approximation of stress and strain we follow the rationale and the procedure presented in [1]. Those approximations preserve invariance under affine transformations and minimize excessive energy. The element formulation is characterized by simplicity and reliability and the proposed approach leads to symmetric stiffness matrices.

The second part of the paper focuses on the treatment of nearly incompressibility. Towards this goal, sources of excessive energy in case of volumetric as well as deviatoric deformations of nearly incompressible materials are examined. A modified version of the Hu-Washizu principle is employed as a basis for the discretization. Departing from the usual pointwise satisfaction of all plane strain conditions, the constraint concerning vanishing of the transverse deformation is taken into account in weak form, whereas vanishing of transverse shear is satisfied pointwise. The convergence properties of the formulation are investigated by examining stability and consistency. Departing from common practice, satisfaction of the patch test is proved analytically. Furthermore, it is concluded that terms, which may-in case of an arbitrary quadrilateral-cause some excessive energy can not be omitted, if we want to pass the classical patch test. The test is also applied numerically to a patch of elements under different loading conditions.

The third part of the present study deals with the extension of the formulation to the analysis of plane strain problems assuming a J_2 flow theory with linear isotropic/kinematic hardening. To avoid problems related to the nonconservative character of the formulation, we follow the original approach suggested by Lagrange for handling auxiliary conditions. The approach employed in the previous sections serves as a basis for the approximations of the displacement, total strain, stress deviator, elastic pressure, and volumetric strain in the present case. However, the current formulation incorporates an important modification that leads to further simplifications: The continuous, linear functions for the approximations are now replaced by the Heaviside function. This provides a consistent mechanism for describing progressive yielding through the element and leads to a significant reduction of computational cost.

ACKNOWLEDGEMENTS

Financial support of the European Union and the General Secretary for Research and Technology/Greece in the framework of the Research Program ``Heraclitus'' (Grant No. 50) is gratefully acknowledged.

REFERENCES

[1] Paraskevopoulos, E. and Talaslidis, D. (2005), “Reduction of excessive energy in the four-noded membrane quadrilateral element, Part I: Linear theory-compressible materials,” Comput. Methods Appl. Mech. Engrg., in press.

[2] Paraskevopoulos, E. and Talaslidis, D. (2005) “Reduction of excessive energy in the four-noded membrane quadrilateral element, Part II: Corotational formulation, near incompressibility, and J_2 plasticity,” submitted for publication, Comput. Methods Appl. Mech. Engrg..

[3] Wempner, G., Talaslidis, D. (2002), *Mechanics of Solids and Shells: Theory and Approximations,* CRC Press, Boca Raton.

[4] Simo, J.C., Hughes, T.J.R. (1998), *Computational Inelasticity,* Springer-Verlag, New York.

[5] Lanczos, C. (1970), *The Variational Principles of Mechanics,* 4th ed., Cambridge University Press, Cambridge.

[6] Ibrahimbegovic, A. and Frey, F. (1993) “Geometrically non-linear method of incompatible modes in application to finite elasticity with independent rotations,” Internat. J. Numer. Methods Engrg. 36, 4185-4200.

[7] Taylor, R.L., Beresford, P.J., and Wilson, E.L. (1976) “A non-conforming element for stress analysis,” Internat. J. Numer. Methods Engrg. 10, 1211-1219.

[8] Kasper, E.P. and Taylor, R.L. (1997), *A Mixed-Enhanced 3d Strain Method: Linear Problems*, UCB/SEMM-97/02 Report, University of California at Berkeley.

[9] Hughes, T.J.R. (1987), *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*, Prentice-Hall Inc., Englewood Cliffs.

Element Formulation	Mesh NxN	Horizontal Displacement	Vertical Displacement
QR4	10x1	-0.388	0.722
QR4/I4	10x1	-0.555	0.819
QR9	10x1	-0.558	0.822
QR9/I4	10x1	-0.559	0.822
Proposed	5x1	-0.441	0.800
Proposed	10x1	-0.534	0.811
Proposed	10x1(distorted)	-0.511	0.780
Beam	10 elements	-0.555	0.811

Table 1: In plane bending of a beam: Displacements at the free end (N: elements per side)

Reference: NxN	Proposed horizontal	Proposed vertical	Enhanced horizontal	Enhanced vertical	B-bar horizontal	B-bar vertical
4x4	-3.0943	7.2675	-2.8918	6.8192	-2.665	6.3627
10x10	-3.2346	7.3409	-3.1662	7.2277	-3.1229	7.1641
16x16	-3.2352	7.3620	-3.19998	7.3015	-3.1923	7.2836

Table 2: Cook's membrane problem: Nearly incompressible material (ν=0.49999995)

Reference: NxN	Proposed horizontal	Proposed vertical	Enhanced horizontal	Enhanced vertical	B-bar horizontal	B-bar vertical
4x4	-1.135	1.500	-0.836	1.108	-0.447	0.631
10x10	-1.464	1.942	-1.657	2.178	-0.867	1.164
16x16	-1.569	2.070	-1.697	2.234	-1.139	1.506

Table 3: Cook's membrane problem: J_2 plastic flow

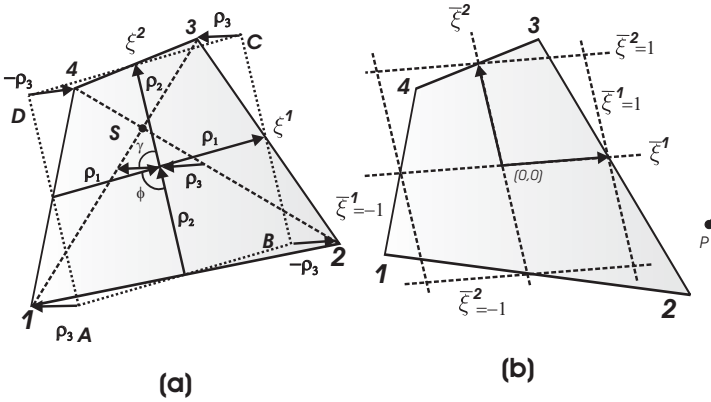


Figure 1. Geometric features of the quadrilateral

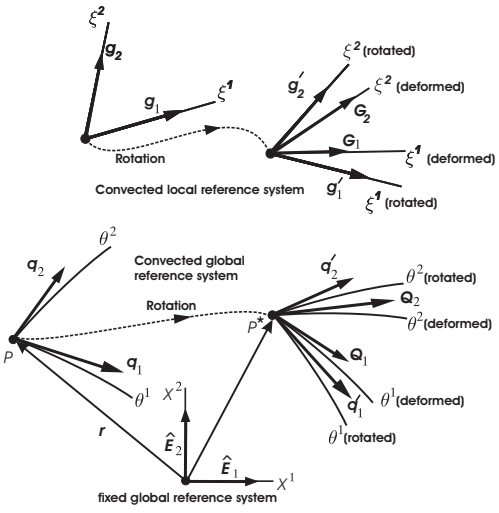


Figure 2. Corotational formulation

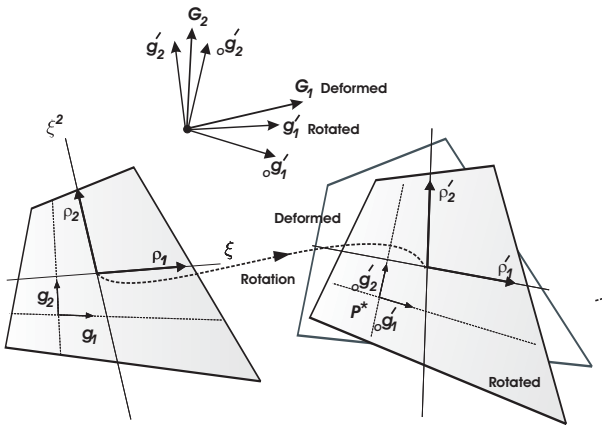


Figure 3. Rotation followed by deformation

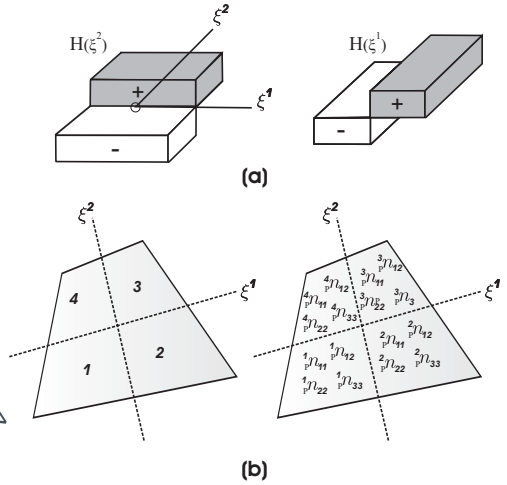


Figure 4. Piecewise constant approximations

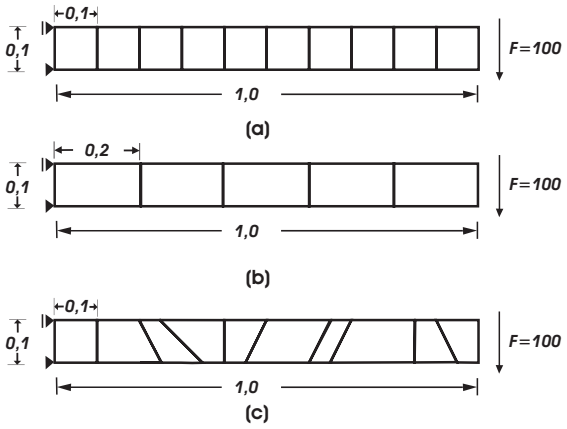


Figure 5. Corotational formulation: In plane bending

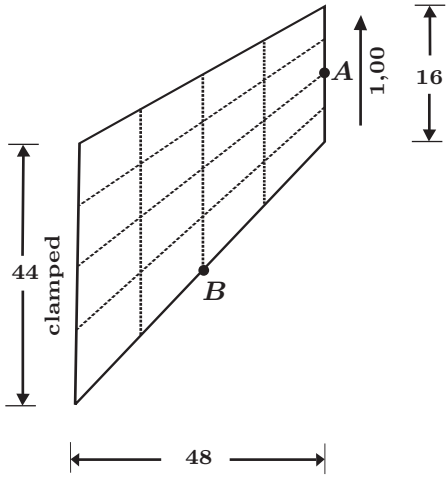


Figure 6. Cook's membrane problem

LARGE FINITE ELEMENTS BASED ON COONS INTERPOLATION FOR THE SOLUTION OF TWO DIMENSIONAL POTENTIAL PROBLEMS

Vasilis M. Dimitriou, Dimitris V. Koulocheris, and Andreas E. Kanarachos

Vehicles Laboratory
School of Mechanical Engineering
National Technical University of Athens
Iroon Politechniou 9, 15780, Zografou Campus, Athens, Greece
e-mail: vidim@central.ntua.gr

Keywords: Large Finite Elements, *C*-elements, Coons Interpolation, Potential Problems.

Abstract. *This paper discusses the construction of "large" quadrilateral and triangular finite elements (C-elements) based on the interpolation method of S.A. Coons. It is shown how triangular elements can be constructed from quadrilateral elements through the degeneration of a selected boundary to one point. The C-elements can be advantageously used for the approximation of high order solutions, as well as for the approximation of singularity or near-singularity type solutions of second order differential equations. Two numerical examples from the solution of the Laplace equation in a quadrilateral and a triangular domain, demonstrate the good performance of the proposed C- elements.*

1 INTRODUCTION

The development of "large" elements aiming at the reduction of mesh generation workload, DOF-number and CPU-time, has been always of high interest in the area of computational mechanics.

"Large" elements with global approximation characteristics were introduced in the past by Jirousek and associates[11,12] and by Brebbia and Dominguez[13] on the basis to Trefftz's method. However, the above "boundary elements" require the knowledge of the fundamental solution of the PDE to be solved, which encounters a number of non-trivial problems. For this reason, "large" finite elements are still of interest.

Historically, it was Irons[1] who generalized the idea of arbitrary nodded finite elements with global interpolation characteristics. But also blending function methods based on ideas put forward by Coons[2] have been used to produce large finite elements: for example Gordon[3], Gordon and Hall[4], Zafrany and Cookson[5], Kanarachos and associates[6,7] and Provatidis and Kanarachos[8].

The present paper is based on the "generalized" formulation of Coons interpolation, as presented originally by Kanarachos and associates[9,10] and deals with quadrilateral and triangular *C*-elements. The reason for introducing triangular *C*-elements is their ability to fit to triangular domains, corners and other geometrical irregularities and to satisfactorily approximate singular or near singular type solutions of PDEs.

In the following, it is shown how triangular *C*-elements can be constructed following reference [10] and numerical results for the solution of Laplace equation in rectangular and triangular solution domains are presented.

2 PROPERTIES OF THE C-ELEMENTS

2.1 Quadrilateral C-elements

The "generalized" transfinite interpolation is characterized by an impressive flexibility, as it allows the use of Lagrange or Hermite-type polynomials, splines, (periodical) trigonometric base functions, or even exponential functions.

Thus, it is easy to interpolate closed-form surfaces such as a helical surface (*u*: trigonometric, *v*=arbitrary), a sphere (*u, v*: trigonometric), *etc*. Also axis-symmetric surfaces, such as a part from a torus or the circular fillet between two coaxial cylinders with different diameters, with the lofting directed circumferentially, can be approximated by the generalized Coons interpolation[9].

Applying now the two-dimensional bivariate interpolation,

$$F(u,v) = (1-u)F(0,v) + uF(1,v) + (1-v)F(u,0) + vF(u,1) - \begin{bmatrix} 1-u & u \end{bmatrix} \begin{bmatrix} F(0,0) & F(0,1) \\ F(1,0) & F(1,1) \end{bmatrix} \begin{bmatrix} 1-v \\ v \end{bmatrix} \quad (1)$$

a boundary nodded *C*-element can be constructed (Figure 1).

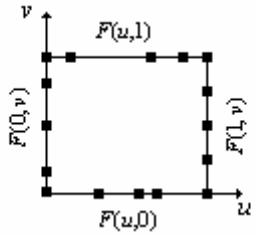


Figure 1. *C*-element with nodes appearing only at the boundary

The geometrical similarity of the above *C*-element to a boundary type element is obvious, as all DOF's appear at the element's boundaries. However, the mathematical properties of the *C*-element are different from the properties of a boundary element as it has been proved that they satisfy the Euler-Lagrange PDE and their performance may be improved by the introduction of "inner" nodes (DOF's) and higher order interpolation functions[10].

The introduction of inner DOF's demands in many cases the construction of "artificial" nodes (e.g. nodes 5, 6, 7 and 8 in Figure 2):

$$F_5 = \frac{F_1 + F_2}{2},\ F_6 = \frac{F_1 + F_3}{2},\ F_7 = \frac{F_2 + F_4}{2},\ F_8 = \frac{F_3 + F_4}{2} \quad (2)$$

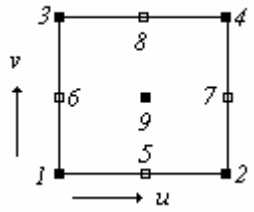


Figure 2. *C*-element with DOF's at the points 1, 2, 3, 4 (4-noded) and an inner DOF at point 9

The example of Figure 2 allows the introduction a quadratic interpolation function, e.g. in *v*-direction:

$$F_+(\frac{1}{2}, v) = F_5(1 - 3v + 2v^2) + F_9(4v - 4v^2) + F_8(-v + 2v^2) \quad (3)$$

which leads to the generation of the interpolating function $F_+(u,v)$:

$$\begin{aligned} F_+(u,v) &= u(1-u)\cdot F_+(\frac{1}{2}, v) \\ &= u(1-u)\cdot[F_5(1-3v+2v^2) + F_9(4v-4v^2) + F_8(-v+2v^2)] \end{aligned} \quad (4)$$

The new function $F_+(u,v)$ has the following properties:

$$\frac{\partial^2 F_+(u,v)}{\partial u^2} = -2[F_5(1-3v+2v^2)+F_9(4v-4v^2)+F_8(-v+2v^2)]$$
$$\frac{\partial^2 F_+(u,v)}{\partial v^2} = u(1-u)\cdot[4F_5-8F_9+4F_8] \tag{5}$$

At u=v=1/2 equation 5 yields:

$$\frac{\partial^2 F_+(\frac{1}{2},\frac{1}{2})}{\partial u^2} = -2F_9, \qquad \frac{\partial^2 F_+(\frac{1}{2},\frac{1}{2})}{\partial v^2} = F_5 - 2F_9 + F_8 \tag{6}$$

meaning, that if $F_9 = (F_5 + F_8)/4$, the Laplace equation $F_{+uu} + F_{+vv} = 0$ is fulfilled at the point u=v=1/2. Therefore, the introduction of inner DOF's and associated artificial nodes increases the performance *C*-elements.

2.2 Triangular *C*-elements

The finite element approximation of singular or near singular type of solutions is a well-researched field with many implementations in engineering practice, especially in the traditional field of structural analysis where singularities mostly arise from geometry features (e.g. sharp corners, stiffeners, etc.) or from boundary conditions (e.g. concentrated loads). The solutions characteristically have high local (i.e. in the vicinity of the singularities) gradients, which have to be properly resolved by numerical approximation.

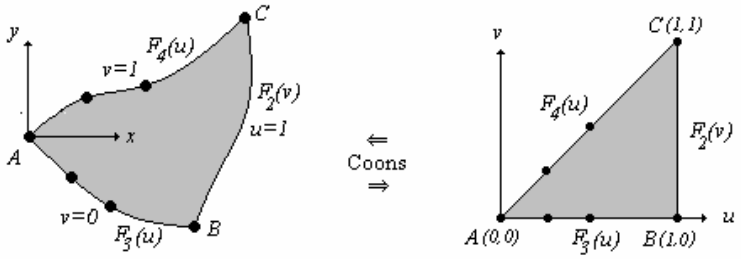


Figure 3. Mapping of the Euclidean space into Ω (u,v)

Based on references [9] and [10], let us first construct high order triangular *C*-elements, assuming that an arbitrary shaped triangle in the x-y Euclidean space is mapped in a "unit" triangle in Ω (coordinates (u,v)). The triangle is characterized by the boundary curves $\mathbf{x}_i$ (i=2,3, 4), while u,v represent arbitrary curvilinear coordinates within an Euclidean x, y–space:

$$\mathbf{x} = \begin{bmatrix} x(u,v) \\ y(u,v) \end{bmatrix} \qquad \mathbf{x}_2 = \begin{bmatrix} x(1,v) \\ y(1,v) \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} x(u,0) \\ y(u,0) \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} x(u,1) \\ y(u,1) \end{bmatrix} \tag{7}$$

The triangle is mapped in Ω using the bivariate interpolation:

$$x(u,v) = x_2(v)\cdot u + x_3(u)\cdot(1-v) + x_4(u)\cdot v - u\cdot[(1-v)\cdot x_B + v\cdot x_C]$$
$$y(u,v) = y_2(v)\cdot u + y_3(u)\cdot(1-v) + y_4(u)\cdot v - u\cdot[(1-v)\cdot y_B + v\cdot y_C] \tag{8}$$

In the case of a boundary nodded *C*-element (Figure 3), the discrete DOF's along the boundaries 2,3 and 4 are interpolated by Lagrange, Splines or other interpolating functions:

$$F_2(v),\ F_3(u)\ F_4(u) \tag{9}$$

Then, $F(u,v)$ is given by the bivariate interpolation:

$$F(u,v) = F_2(v)\cdot u + F_3(u)\cdot(1-v) + F_4(u)\cdot v - u\cdot[(1-v)\cdot F_B + v\cdot F_C] \tag{10}$$

The same holds also for *C*-elements with inner nodes. Assuming, that "*n*" inner and boundary nodes (DOF's) are placed at u_k=constant stations, a Lagrange polynomial ${}^nL(v)$ is used to interpolate the *n F*-values at each u_k:

$${}^nL_k(v) \tag{11}$$

Then, the unknown function *F* can be interpolated using equation 12:

$$F(u,v) = \sum_{k=1}^{K} E_k(u)\cdot {}^nL_k(v) \tag{12}$$

while blending function $E_k(u)$ fulfils the cardinality conditions (equation 13):

$$E_k(u_j) = 1 \;\; for \;\; k = j, \qquad E_k(u_j) = 0 \;\; for \;\; k \neq j \tag{13}$$

3 NUMERICAL SOLUTION OF THE LAPLACE EQUATION

At the subsections 3.1 and 3.2 that follow, rectangular and triangular *C*-elements are used to solve two problems governed by the Laplace equation[14,15]. As it is shown from the numerical results the performance of the *C*-elements is too good, especially when inner nodes are added to them.

3.1 Rectangular domain

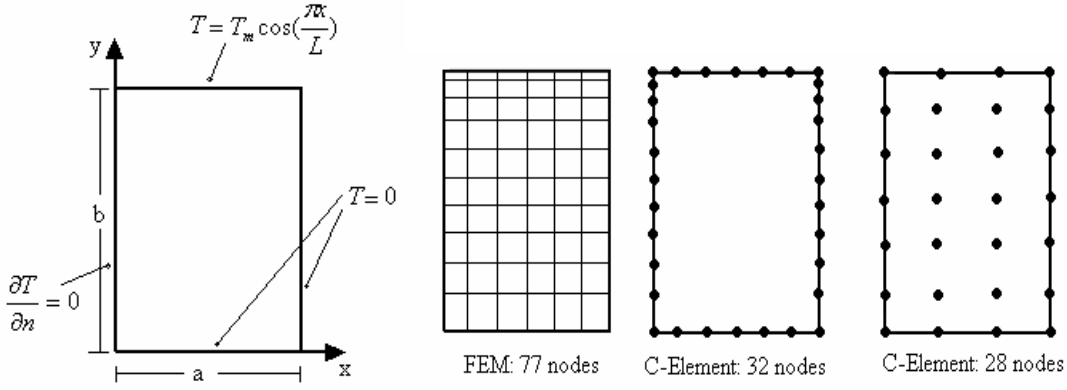


Figure 4. Rectangular adiabatic plate

The first model problem refers to a thin rectangular plate free of heat sources, insulated at the top and bottom surfaces with dimensions L=2a=6 and b=12 (Figure 4). The analytical solution is given as:

$$T(x,y) = T_m\left[\frac{\sinh(\frac{\pi y}{L})}{\sinh(\frac{\pi b}{L})}\right]\cos(\frac{\pi x}{L}) \tag{14}$$

with T_m=100°C. The above exact solution is a very steep exponential function along the *y*-axis of symmetry. Due to the symmetry of the problem with respect to the *y*-axis, only the half of the domain is to be analyzed.

According to reference [15] the problem was analyzed using FEM (77 nodes), BEM (32 nodes) and CP-macroelements (32 nodes) and the same non-uniform mesh on the boundary, adapted to the solution. Instead of the linear blending functions used firstly for the construction of the Coons elements, a second option with linear blending functions in *y*-direction and sinusoidal ones in *x*-direction, presented in equation 15, were proposed.

$$E_0(\xi) = \cos(\frac{\pi\xi}{2}), \quad E_1(\xi) = 1 - \cos(\frac{\pi\xi}{2})$$
$$E_0(\eta) = 1 - \eta, \quad E_1(\eta) = \eta \tag{15}$$

As it can be seen from Table 1 the use of the sinusoidal blending functions reduce the numerical error because they approximate accurately the sinusoidal boundary condition on the top of the plate.

These choices are very interesting but they demand, in a matter, the knowledge of the exact solution. This is the reason why in the present work the problem is solved by the help of a ‘‘Large’’ rectangular Coons element (28 nodes) with inner nodes which are necessary for the improvement of the results, as it was analytically shown at section 2, uniformly distributed to the element as it is presented in Figure 4.

y	Exact Solution	Errors in %			
		FEM 77 nodes	BEM 32 nodes	CP-element 32 nodes	Coons -element 28 nodes
12.00	100.0000	data	data	data	data
11.781	89.1661	0.09	0.32	-0.024	-0.08
11.345	70.9666	0.32	0.40	-0.004	-0.13
10.691	50.3888	0.75	0.49	0.047	-0.02
9.818	31.9013	1.51	0.68	0.093	0.12
8.727	18.0172	2.73	1.04	0.172	-0.17
7.418	9.0759	4.59	1.77	0.363	-0.90
5.891	4.0731	7.29	3.25	0.692	0.41
4.145	1.6148	10.91	6.30	1.237	3.28
2.182	0.5258	14.97	12.95	2.073	-11.60
0	0	data	data	data	data
at the points: x=11.781, 11.345, 10.691 και 9.818		Mean average absolute error (%)			
		0.6675	0.4725	0.051	0.0875

Table 1. Comparison of the four different approximation results

Despite the uniform distribution of the nodes, the accuracy of the proposed solution is very satisfactory. As it can be seen from Figure 5 the approximated numerical solution is identical with the analytical, without the need of the use of an adapted to the solution mesh or ‘‘special’’ blending functions.

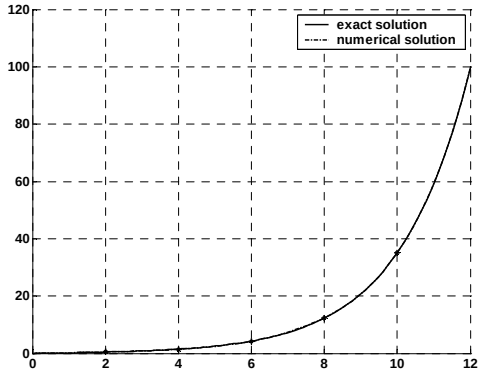
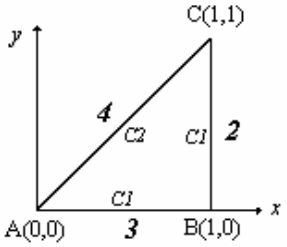


Figure 5. Approximation of the exact solution using a 28 node Coons element

3.2 Triangular domain



$$\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} = 0$$

$$F(C1) = \alpha,\ \nabla F(C2) = \beta$$

Figure 6. Laplace equation

In the following the model case of the Laplace equation (Figure 6), with the exact solution:

$$F_{exact}(x,y) = 6x^2y^2 + (1-x^4) + (1-y^4) \tag{16}$$

will be investigated[14]. The model problem addresses a triangular solution area ABC with the A, B, C-node coordinates (x,y)= (0,0), (0,1), (1,1) and the boundary conditions:

$$C1: F(x,0) = (1-x^4),\ F(1,y) = 6y^2 + (1-y^4) \qquad C2:\ \beta = 0 \tag{17}$$

Introducing equation 18:

$$F(x,y) = \sum_{j=1}^{J} F_i \cdot N_i(x,y) \tag{18}$$

the finite element formulation results in (e=element area):

$$\delta F_i \cdot [\iint_e \sum_{j=1}^{J} (\frac{\partial N_i}{\partial x} \cdot \frac{\partial N_j}{\partial x} + \frac{\partial N_i}{\partial y} \cdot \frac{\partial N_j}{\partial y}) \cdot F_j \cdot de = -\iint_e N_i \cdot q(x,y) \cdot de + \\ + \oint_{C2} N_i \cdot \nabla F(C2) \cdot dC2] \tag{19}$$

The exact solution along AC is equal to:

$$F(x,x) = 6x^4 + 2(1-x^4) \tag{20}$$

For the numerical solution, one triangular *C*-element with $n=n_2=n_3=n_4$ is used. As it can be seen from the results presented at Figure 7, although acceptable, are characterized by a lack of improvement of the accuracy of the numerical solution when *n* is increasing. This behavior of the numerical solution can be explained by the help of Figure 8, which shows the exact *F*-values at the boundaries 2, 3 and 4. Although *F* is satisfactorily approximated at the boundaries by the Lagrange polynomials of rank *n-1*, Coons's linear blending cannot approximate satisfactorily *F* in the whole solution domain. In fact, the real blending at a station *x*=const is according to equation 16 a polynomial of rank 3 equal to $a_0 + a_1y^2 + a_2y^4$.

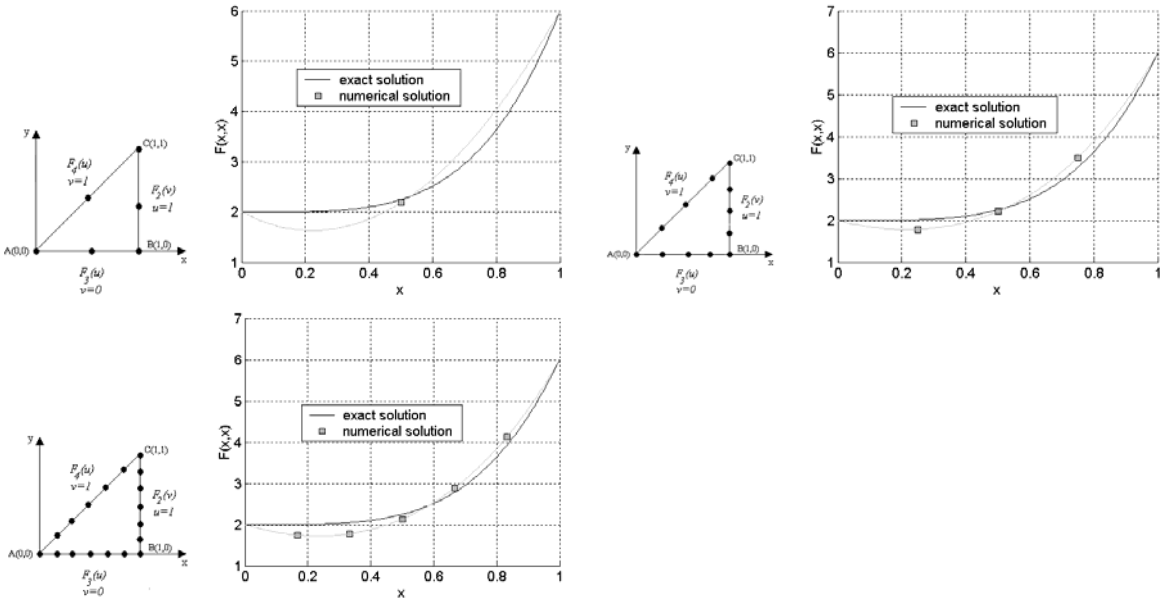


Figure 7. Problem solution for n=3, 5 and 7

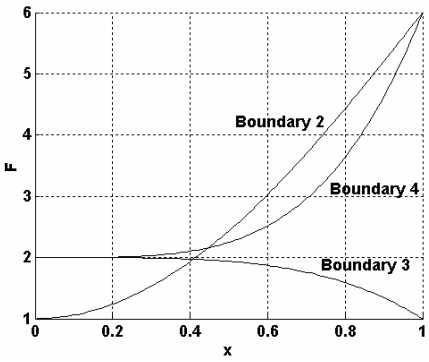


Figure 8. Boundary curves corresponding to the exact F

As it was mentioned in the previous paragraph, an improvement of the numerical performance can therefore be achieved by the introduction of inner nodes.

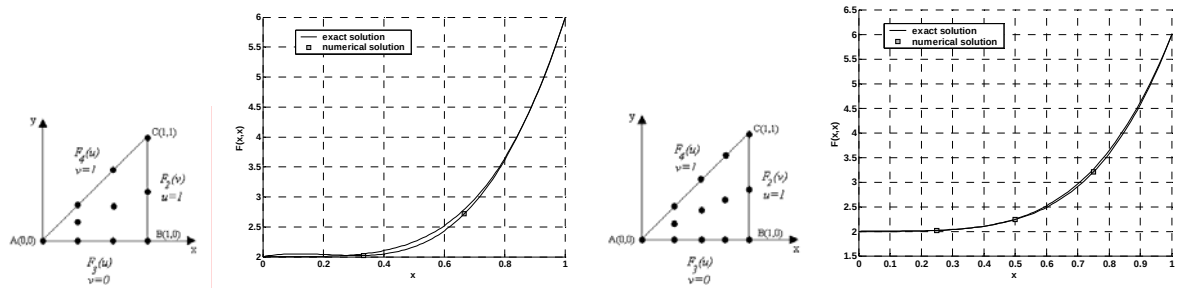


Figure 9. Problem solution: (a) n_3=n_4=4, n_2=3 and 2 inner nodes and (b) n_3=n_4=5, n_2=3 and 3 inner nodes

The numerical results presented at Figure 9 configure that the insertion of inner nodes at the C-elements improves their performance, especially for the solution of the Laplace equation, as it was mentioned before.

4 CONCLUSIONS

In this paper it is shown how "large" high order rectangular and triangular finite elements (C-elements) can be constructed, based on the interpolation method developed by S.A. Coons. The triangular elements can be advantageously used for global approximation of the solution in regions with sharp corners and other

geometrical irregularities and for the approximation of singularity or near-singularity type solutions of PDEs. Numerical results for the Laplace equation sustain the expected performance of the proposed *C*-elements and show no tendency to numerical inaccuracies, which could be feared from using high order Lagrange interpolating polynomials. The insertion of inner nodes to the proposed *C*-elements improves dramatically their performance which is proved analytically and experimentally by the application of the method in two Laplace model case problems. Their performance is that good that as the numerical results show there is no need for an adaptive mesh.

REFERENCES

[1] Irons, B.M. (1966), "Engineering application of numerical integration in stiffness method", J. AIAA 14, 11, pp. 2036-2037.

[2] Coons S.A. (1964), *Surfaces for computer aided design of space forms, Project MAC, MIT, revised for MAC-TR-41*, Springfield, VA. USA.

[3] Gordon, W.J. (1971), "Blending function methods of bivariate multivariate interpolation and approximation", J. SIAM Numer. Anal. 8, pp. 158-177.

[4] Gordon, W.J., Hall, C.A. (1973), "Transfinite element methods blending function interpolation over arbitrary curved element domains", J. Numer. Math. 21, pp. 109-112.

[5] Zafrany, El., Cookson, R.A. (1986), "Derivation of Lagrangian and Hermitian shape functions for quadrilateral elements", Int. J. Numer. Meth. Engng, 23, pp. 1939-1958.

[6] Kanarachos, A., Deriziotis D. (1989), "On the solution of Laplace and wave propagation problems using *C*-elements", J. Finite Element Anal. Des. 5, pp. 97-109.

[7] Kanarachos, A., Provatidis, Ch., Deriziotis, D, Foteas, N.A. (1999), "A new approach of the FEM analysis of two-dimensional elastic structures using global (Coons) interpolation functions" In: J. Wunderlich (editor). *European Conference on Computational Mechanics*, Munich, Germany.

[8] Provatidis, Ch., Kanarachos, A. (2001), "Performance of a macro-FEM approach using global interpolations (Coons') functions in axis-symmetric potential problems", J. Computers and Structures 79, pp. 1769-1779.

[9] Kanarachos, A., Grekas, D., Provatidis, Ch. (1995), *Generalized formulation of Coons's interpolation*, *In: P.D. Kaklis and N.S. Sapidis (editors), Computer aided geometric design; from theory to praxis*, pp. 65-76, NTUA Press, Athens.

[10] Kanarachos, A. (2002), "Large finite elements", *Proceedings of the 4th International Congress on Computational Mechanics (GRACM 2002),* Patras, Greece, 27-29 June, Vol. III, pp. 1055-1062.

[11] Jirousek, J., Leon N. (1977), "A powerful finite element for plate bending", J. Comput. Meth. Engng. 24, pp. 77-96

[12] Jirousek, J. (1978), "Basis for development of large elements locally satisfying all field equations", J. Comput. Meth. Engng. 14, pp. 64-92.

[13] Brebbia, C.A., Dominguez, J. (1992), *Boundary elements: an introductory course*. *Computational Mechanics Publications,* McGraw-Hill Book Company, Southampton.

[14] Dimitriou, V., Pantelelis, N., Kanarachos, S. (2002), "Large triangular finite elements for second order differential equations using the interpolation method of Coons", *Proceedings of the 4th International Congress on Computational Mechanics (GRACM 2002),* Patras, Greece, 27-29 June, Vol. III, pp. 1063-1070.

[15] Provatidis, C. (2002), "A comparative study between Coons-patch macroelements and boundary elements in two-dimensional potential problems", *Proceedings of the 4th International Congress on Computational Mechanics (GRACM 2002),* Patras, Greece, 27-29 June, Vol. I, pp.43-50.

CONVERGENCE ANALYSIS AND COMPARISON OF THE H- AND P-EXTENSIONS WITH MIXED FINITE ELEMENT C^0 CONTINUITY FORMULATIONS, FOR SOME TYPES OF ONE DIMENSIONAL BIHARMONIC EQUATIONS[1]

G. I. Tsamasphyros*, S. Markolefas and D. A. Tsouvalas

**Department of Applied Mechanics*
Faculty of Applied Mathematics and Physics
National Technical University of Athens
9 Iroon Polytechniou, Zografou 157 73, Athens, Greece

Keywords: Biharmonic equation, Finite elements, *h*- and *p*- extension, Mixed Methods, Ciarlet-Raviart Method, Gradient elasticity

Abstract. *A mixed formulation with two main variables, based on the Ciarlet-Raviart technique, with C^0 continuity shape functions is employed for the solution of some types of biharmonic equations in 1-D. The continuous and discrete Babuška-Brezzi inf-sup conditions are established. The formulation is numerically tested for both the h- and p- extensions. The model problems involve the standard biharmonic equation, with variable bending stiffness (with both regular and irregular exact solutions), as well as, a more general biharmonic equation with lower order term, constant coefficients and complicated boundary conditions (with smooth exact solution), resulting from a gradient elasticity problem. The standard, quasi-optimal finite element error rates of convergence are numerically confirmed in all cases. The basic conclusion of the numerical experimentation is that the p-extension provides much better accuracy than the h-extension, for both main variables. For the irregular exact solution, the observed rate of convergence corresponding to the second variable (displacements) is always higher than the error rate for the first variable (bending moment). The latter may be theoretically explained via the particular structure of the bilinear functionals of the given formulation.*

1. INTRODUCTION

A mixed formulation with C^0 continuity conforming finite element basis functions is employed for the solution of some types of one dimensional biharmonic equations. The formulation is theoretically analyzed and numerically tested against both the uniform *h*- and *p- extensions* of the finite element method, with equal interpolation order for both main variables.

The conforming finite element discretizations of the standard *Galerkin* formulations (*G-F*) for the biharmonic equations demand C^1 continuity shape functions [1,2]. In general, however, it is difficult to satisfy the C^1 continuity conditions for shape functions in multi-dimensional problems. One way to circumvent the difficulty of C^1 continuity is to use *non-conforming* elements, like the *DKT* (*Discrete Kirchhoff Triangle*) element [2,chVI,§5]. Another way is to employ *mixed formulations*, in order to reduce the regularity requirements in the bilinear functionals [1-3]. The most commonly used techniques employ the components of the gradient of displacement (∇u), as independent variables (usually referred to as: *rotations*, $\underline{\theta}$), while the *Kirchhoff constraint* $\nabla u = \underline{\theta}$ is enforced via *Lagrange multipliers* or *penalty terms* [2,chVI,§5,§6], [3].

Another possibility is to introduce the second order derivatives of the displacement as new independent variables. The original technique (*Ciarlet-Raviart method*) is analyzed in [1,ch.7]. Various modifications of this method can be found in the literature. For example, the *Herrmann-Miyoshi (H-M) method* [4,sect.4] employs as auxiliary variables all the second order derivatives of the displacement. The method developed in [5] constitutes a generalization of the previous *H-M* technique, for biharmonic equations with variable coefficients. The new variables in [5] are the components of the *bending moment tensor*.

In the current work we attempt a comparison between the *h*- and *p- extensions* (with *Legendre polynomial* based hierarchical shape functions [6 - 8]), for some types of 1-D biharmonic equations. More specifically, the one dimensional analogue of the mixed technique of [5] is employed. Furthermore, a more general biharmonic equation is considered, resulting from a *gradient elasticity* formulation [9]. The complex boundary conditions give rise to modifications on the standard structure of the bilinear forms of the mixed formulation. Note that,

[1] This work was conducted within the framework of the EPEAEK program "Pythagoras II". The Project is co-funded by the European Social Fund (75%) and National Resources (25%).

biharmonic equations with analogous (*Robin type*) boundary conditions are common in the formulation of boundary value problems within the framework of various *gradient elasticity* theories in many dimensions [10-13].

Besides the validation of the theoretical analysis of this work, one of the goals of the numerical experimentation is to provide a basis for the application of the *p-extension*, in conjunction with mixed, C^0 continuity, conforming formulations, in relevant multi-dimensional problems. It is also noted that the bulk of the numerical experience with *Ciarlet-Raviart* type formulations in 2-D is based on *h-extension* discretization schemes [5,14-16].

It is finally noteworthy that, for the specific mixed formulations, the *pollution error* of the displacement variable (that is, the difference between the finite element solution and the *optimal projection*, measured in the solution space H^1 norm) is bounded by lower order norms of the error of the new variable w. This result explains the observed higher order convergence rate of the displacement for the problems with irregular exact solution.

The structure of the work is as follows. Section 2 introduces the nomenclature and the definition of the model problems. The exact weak forms and the continuous Babuška – Brezzi conditions [17,18] are described in section 3. Section 4 contains the mixed finite element approximations and the verification of the respective discrete Babuška – Brezzi conditions. The numerical experimentation is given in Section 5. Finally, section 6 contains a closing discussion and future research directions.

2. NOMENCLATURE AND DEFINITION OF THE MODEL PROBLEMS

The standard notation for Sobolev spaces and norms is employed [1,2,6]. Let $\Omega = (b,c)$, with boundary $\partial\Omega = \{b,c\}$. Let $\Gamma_E \subset \partial\Omega$ and $\Gamma_N \subset \partial\Omega$, such that $\Gamma_N \cup \Gamma_E = \partial\Omega$. The strong form of the model problem is: Find $u \in H^r(\Omega)$, $r \geq 3$, such that,

$$Su = f \tag{1}$$

where S is a fourth order linear differential operator and f is a known forcing function.

The function u satisfies essential boundary conditions on Γ_E and natural (or Robin) boundary conditions on Γ_N. We denote by $\|w\|_r$, $r \in \Re$, the H^r norm of the (real valued) function w, defined on Ω. The Sobolev spaces $H^r(\Omega)$, $H_0^r(\Omega)$, $r \in \Re$, $L^\infty(\Omega)$ and $C_0^\infty(\Omega)$, have the usual definitions [1,2], while $H^1_{0,\Gamma_E}(\Omega)$ is the subspace of functions $w \in H^1(\Omega)$, such that $w = 0$ on Γ_E (in the sense of trace).

Two different types of strong forms are considered:

a) *Standard beam bending with variable bending stiffness*:

$$(J(x)u'')'' = f \ \text{ in } \ \Omega = (0,1), \quad u(0) = u'(0) = u(1) = u'(1) = 0 \tag{2}$$

where $J(x) > 0$ is the variable bending stiffness of the beam.

b) *Gradient elastic bar in tension [9]*:

$$g^2 u'''' - u'' = \frac{\bar{f}}{AE} = f \ \text{ in } \ \Omega = (0,1) \tag{3}$$

where AE and g^2 are real constants. The following boundary conditions are considered:

$$\left.\begin{aligned} &u(0) = 0, \ \ u'(1) = \varepsilon_0 \} \text{ essential conditions} \\ &AE(u'(1) - g^2 u'''(1)) = P, \ \ AE(lu'(0) + g^2 u''(0)) = R(0) = 0 \} \text{ Robin conditions} \end{aligned}\right. \tag{4}$$

where l, P, ε_0 are given constants. It is noteworthy that $0 < l < g^2$ (in order to get a positive definite strain energy functional [9]). Moreover, P is the standard axial force, while R is the so-called *double force* [9-13].

3. MIXED FORMULATIONS - EXACT WEAK FORMS AND CONTINUOUS BABUSKA-BREZZI CONDITIONS

A new variable is introduced as follows,

$$w := Gu'' \tag{5}$$

where $G \coloneqq J(x)$ for (2) and $G \coloneqq g^2$ for (3).

Using (5), (2) and appropriate test functions for both w and u, we get the following weak formulation:

Formulation 3.1: *Find* $w \in H^1(\Omega)$ *and* $u \in H_0^1(\Omega)$, *such that,*

$$\begin{aligned} &\int_0^1 \frac{wr}{G} dx + \int_0^1 r'u' dx = 0, \forall r \in H^1(\Omega) \\ &\int_0^1 w's' dx = -\int_0^1 f(x) s dx, \forall s \in H_0^1(\Omega) \end{aligned} \tag{6}$$

The respective formulation in 2-D may be stated as follows [1,4]: Find $w \in H^1(\Omega)$ and $u \in H_0^1(\Omega)$, such that,

$$\begin{aligned} &\int_0^1 \frac{wr}{G} d\Omega + \int_0^1 \nabla r.\nabla u \; d\Omega = 0, \forall r \in H^1(\Omega) \\ &\int_0^1 \nabla w.\nabla s \; d\Omega = -\int_0^1 f(x) s \; d\Omega, \forall s \in H_0^1(\Omega) \end{aligned} \tag{7}$$

For the problem of gradient elastic bar in tension (3) the following weak formulation is derived:

Formulation 3.2: *Find* $w \in H^1(\Omega)$ *and* $u \in H_{0,\Gamma_E}^1(\Omega)$, *such that,*

$$\begin{aligned} &\int_0^1 \frac{wr}{G} dx - \frac{w(0)r(0)}{l} + \int_0^1 r'u' dx = \varepsilon_0 r(L), \forall r \in H^1(\Omega) \\ &\int_0^1 w's' dx - \int_0^1 u's' dx = -\int_0^1 f(x) s dx - \frac{P}{AE} s(1), \forall s \in H_{0,\Gamma_E}^1(\Omega) \end{aligned} \tag{8}$$

Note that the given *Robin* boundary conditions (4) introduce the boundary relations $w'(1) = \varepsilon_0 - \frac{P}{AE}$, $u'(0) = -\frac{w(0)}{l}$, which are used in the formation of (8).

We observe that both (6) and (8) are special cases of the well known *constraint mixed formulation* [2,18]:

Formulation 3.3: *Find* $(w,u) \in (U \text{x} Q)$, *such that*

$$\begin{aligned} &A(w,r) + B(r,u) = y(r), \forall r \in U \\ &B(w,s) - C(u,s) = g(s), \forall s \in Q \end{aligned} \tag{9}$$

where U, Q are infinite dimensional *Hilbert spaces*, endowed with inner products $(*,*)_U$, $(*,*)_Q$ and associated norms $\|*\|_U$, $\|*\|_Q$ respectively.

The necessary and sufficient conditions for the well posedness of (9), for all admissible forcing functions, are the well known *continuous Babuška-Brezzi* (*BB*) conditions [2,17,18], associated with $A(*,*)$ and $B(*,*)$.

Definition 3.4: *There exist strictly positive constants* ω *and* β *such that:*

$$\inf_{\substack{w \in U_0 \\ w \neq 0}} \sup_{\substack{r \in U_0 \\ r \neq 0}} \frac{A(w,r)}{\|w\|_U \|r\|_U} = \inf_{\substack{r \in U_0 \\ r \neq 0}} \sup_{\substack{w \in U_0 \\ w \neq 0}} \frac{A(w,r)}{\|w\|_U \|r\|_U} = \omega > 0 \tag{10}$$

$$\inf_{\substack{s \in Q \\ s \neq 0}} \sup_{\substack{w \in U \\ w \neq 0}} \frac{B(w,s)}{\|w\|_U \|s\|_Q} = \beta > 0 \tag{11}$$

where $U_0 \coloneqq \{ r \in U : \; B(r,s) = 0 \;\; , \forall s \in Q \}$ is the so-called (continuous) *kernel*.

For (6) and (8), the second *BB* condition (11) may be easily verified, as follows:

$$\beta = \inf_{\substack{s \in Q \\ s \neq 0}} \sup_{\substack{w \in U \\ w \neq 0}} \frac{\int_0^1 w's' dx}{\|w\|_1 \|s\|_1} \geq \inf_{\substack{s \in Q \\ s \neq 0}} \frac{\int_0^1 s's' dx}{\|s\|_1 \|s\|_1} = \inf_{\substack{s \in Q \\ s \neq 0}} \frac{|s|_1^2}{\|s\|_1^2} \tag{12}$$

where $U = H^1(\Omega)$ and $Q = H_{0,\Gamma_E}^1(\Omega) \subset U$.

However, the seminorm $|*|_1$ is equivalent to the norm $\|*\|_1$ on $Q = H^1_{0,\Gamma_E}(\Omega)$, hence $\beta \geq C > 0$, where C is a constant [1,2]. The above argument may be easily extended into multi-dimensional problems. The other *BB* condition (10) is not as trivial as the above (it is not valid in many dimensions, see also section 6). The first step is to define precisely the continuous *kernel*. For the given one dimensional formulations there follows,

$$B(r,s) = 0\ ,\quad \forall s \in Q\ \Rightarrow \int_0^1 r's'dx = 0,\ \forall s \in\ C_0^\infty(\Omega)\ \Rightarrow \int_0^1 r''sdx = 0,\ \forall s \in\ C_0^\infty(\Omega) \tag{13}$$

Therefore, based on the theory of distributions [1,2], the general form of $r \in U_0$ is as follows,

$$r = C_2 + C_1 x \tag{14}$$

where C_1, C_2 are arbitrary real constants.

It can be easily seen that for the gradient elasticity formulation 3.2, $C_1 = 0$, i.e., the continuous *kernel* contains only constant functions. We note that U_0 is finite dimensional. First we consider formulation 3.1.

Theorem 3.5: *Assume that* $J \in L^\infty(\Omega)$. *Then, the continuous BB condition* (10) *is valid for the formulation 3.1.*

Proof:

It follows that,

$$A(r,r) = \int_0^1 J^{-1} r^2 dx \geq \|J\|_\infty^{-1} \|r\|_0^2 \tag{15}$$

It is well known that in a finite dimensional space all norms are equivalent. Thus, $\|r\|_0 \geq C_m \|r\|_1$, $\forall r \in U_0$, for some positive constant C_m which depends only on the dimension of U_0. Thus, from (15),

$$A(r,r) \geq C_m^2 \|J\|_\infty^{-1} \|r\|_1^2\ ,\ \forall r \in U_0 \tag{16}$$

Based on relation (16), it follows that (10) is valid, with $\omega \geq C_m^2 \|J\|_\infty^{-1} > 0$ and the proof is completed **..**

Now we examine formulation 3.2.

Theorem 3.6: *Assume that* $0 < l < g^2$. *Then the continuous BB condition* (10) *is valid for the formulation 3.2.*

Proof:

$$\omega = \inf_{\substack{r \in U_0 \\ r \neq 0}} \sup_{\substack{w \in U_0 \\ w \neq 0}} \frac{A(w,r)}{\|w\|_U \|r\|_U} \geq \inf_{\substack{r \in U_0 \\ r \neq 0}} \sup_{\substack{w \in U_0 \\ w(0) = -r(0) \\ w \neq 0}} \frac{\int_0^1 g^{-2} wrdx + w(0)^2 l^{-1}}{\|w\|_1 \|r\|_1} \tag{17}$$

The test functions are now the following: $r(x) = c, w(x) = -c$, where c is an arbitrary real constant. Simple algebra results in inequality (18). Based on the hypothesis, the proof is completed **..**

$$\omega \geq \left(\frac{1}{l} - \frac{1}{g^2}\right) > 0 \tag{18}$$

It is interesting to mention that the mixed formulation 3.2 is well posed in general, if $l \neq g^2$. This can be easily deduced by selecting appropriately the sign of the product $r(x)w(x)$. Then it follows: $\omega \geq \left|\frac{1}{l} - \frac{1}{g^2}\right| > 0$.

4. FINITE ELEMENT APPROXIMATIONS WITH C^0 CONTINUITY AND DISCRETE BABUŠKA-BREZZI CONDITIONS FOR THE H- AND P- EXTENSIONS

Let $U^d \subset U$ and $Q^d \subset Q$ be finite dimensional spaces, composed of piecewise polynomials (C^0 continuity *basis functions*). The general formulation is stated as follows.

Formulation 4.1: *Find* $(w^d, u^d) \in (U^d \mathrm{x} Q^d)$, *such that*

$$\begin{aligned} A(w^d, r^d) + B(r^d, u^d) &= y(r^d), \forall r^d \in U^d \\ B(w^d, s^d) - C(u^d, s^d) &= g(s^d), \forall s^d \in Q^d \end{aligned} \tag{19}$$

The parameter '*d*' in (19) denotes the *level of refinement*. We consider the two well known types of refinement: a) The (quasi-uniform) *h*-extension, where the polynomial degree p of the shape functions is kept fixed and the accuracy increases by reducing the sizes of the elements ($U^d \equiv U^h$ and $Q^d \equiv Q^h$). b) The *p*-extension, where the finite element mesh is kept fixed and the number of degrees of freedom increases, by adding higher order polynomials, usually in a hierarchical fashion ($U^d \equiv U^p$ and $Q^d \equiv Q^p$).

The *discrete inf-sup* conditions refer to the finite dimensional subspaces.

Definition 4.2: *There exist strictly positive constants* δ *and* ζ *such that:*

$$\inf_{\substack{w^d \in U_0^d \\ w^d \neq 0}} \sup_{\substack{r^d \in U_0^d \\ r^d \neq 0}} \frac{A(w^d, r^d)}{\|w^d\|_U \|r^d\|_U} = \delta^d \geq \delta > 0, \quad \inf_{\substack{s^d \in Q^d \\ s^d \neq 0}} \sup_{\substack{r^d \in U^d \\ r^d \neq 0}} \frac{B(r^d, s^d)}{\|r^d\|_U \|s^d\|_Q} = \zeta^d \geq \zeta > 0 \tag{20}$$

where $U_0^d := \left\{ r^d \in U^d : B(r^d, s^d) = 0 \ , \forall s^d \in Q^d \right\}$ is the *discrete kernel* [2,18].

The most interesting consequence of the *discrete BB conditions* is that (20) secures the *solvability* and *quasi-optimal convergence* of the approximation method. The latter is generally expressed as follows,

$$\|w - w^d\|_U + \|u - u^d\|_Q \leq C\left(\inf_{r^d \in U^d} \|w - r^d\|_U + \inf_{s^d \in Q^d} \|u - s^d\|_Q \right) \tag{21}$$

where (w,u) is the exact solution of (9) and C is a constant

For the given formulations 3.1 and 3.2 the second inf-sup condition (20) is always valid (see proof of the continuous condition (12)). For the first *inf-sup* condition (20), the discrete *kernel* contains $r^d \in U^d$ such that,

$$B(r^d, s^d) = 0 \ \Rightarrow \int_0^1 (r^d)'(s^d)' dx = 0 \Rightarrow -\sum_{e=1}^{N_{el}} \int_{\Omega_e} (r^d)'' s^d dx + \sum_{i=1}^{N} J_i[(r^d)' s^d] = 0, \ \forall s^d \in Q^d \tag{22}$$

where N_{el} is the total number of elements, Ω_e is the domain of the typical 1-D element and $J_i[v^d]$ denotes the *jump* of the function v^d at the *free node* i (which is not subject to essential conditions), $i = 1, 2, .., N$. First we consider the case of *h-extension* with $p = 1$. Then from (22), there follows,

$$\sum_{i=1}^{N} J_i[(r^h)' s^h] = 0, \ \forall s^h \in Q^h \tag{23}$$

By taking sequentially $s^h = \phi_i(x)$, where ϕ_i is the nodal (piecewise linear) basis function associated with a free node i, we get $J_i[(r^h)'] = 0$. Then it follows that the function $r^h(x)$ should be linear over the whole problem domain. Hence, for $p = 1$ the *discrete kernels* are the same as the continuous ones. Now we consider the case $p > 1$. By selecting sequentially s^d to be the higher order *internal shape functions* (*bubble modes* [6]) in each element separately, from (22) we get,

$$\int_{\Omega_e} (r^d)'' s^d dx = 0 \tag{24}$$

However, if r^d has higher order than 1, then we can always select some s^d such that (24) is violated. Therefore $r^d \in U_0^d$ is still linear in each element. But then, by the same argument as previously, we get $J_i[(r^d)'] = 0$ for every free node i. Hence, the *discrete kernel* contains only globally linear functions, see (14). Therefore, the first *discrete inf-sup condition* (20) follows directly from the respective continuous condition.

Inequality (21) secures convergence for all possible exact solutions, however in some cases may be too conservative for some of the variables [2,18]. The next theorem confirms this observation for the displacement variable of the formulation 3.1. For generality purposes we focus on the respective 2-D formulation, see (7).

Theorem 4.3: *The pollution error for the displacement approximation of the formulation 3.1, is bounded by the lower order norm* $\|w - w^d\|_{-1}$*, defined on the dual space of the solution space* $Q = H^1_{0,\Gamma_E}(\Omega)$.

Proof:

From the first equation of group (7) and the first respective discrete equation of group (19) we get,

$$\int_{\Omega} \frac{(w-w^d)r^d}{G} d\Omega + \int_{\Omega} \nabla r^d . \nabla (u-u^d)\, d\Omega = 0, \forall r^d \in U^d \tag{25}$$

The optimal projection $\tilde{u}^d \in Q^d$ for the displacement variable may be defined as follows,

$$\int_{\Omega} \nabla s^d . \nabla (u-\tilde{u}^d)\, d\Omega = 0, \forall s^d \in Q^d \tag{26}$$

From (25) and (26) then easily follows that,

$$\int_{\Omega} \frac{(w-w^d)s^d}{G} d\Omega + \int_{\Omega} \nabla s^d . \nabla (\tilde{u}^d - u^d)\, d\Omega = 0, \forall s^d \in Q^d \tag{27}$$

Dividing by the semi-norm $\left|s^d\right|_1$, taking supremum over Q^d and noting that $(\tilde{u}^d - u^d) \in Q^d$, we get,

$$\left|\tilde{u}^d - u^d\right|_1 \le \sup_{s^d \in Q^d} \frac{\left|\int_{\Omega} \frac{(w-w^d)s^d}{G} d\Omega\right|}{\left|s^d\right|_1} \tag{28}$$

Using the positiveness and boundedness of G^{-1}, the *inclusion* $Q^d \subset Q = H^1_{0,\Gamma_E}(\Omega)$, as well as the equivalency of the norms $\left|*\right|_1$ and $\left\|*\right\|_1$, we finally deduce,

$$\left\|\tilde{u}^d - u^d\right\|_1 \le C \sup_{s \in Q} \frac{\left|\int_{\Omega} (w-w^d) s d\Omega\right|}{\left\|s\right\|_1} \tag{29}$$

The right hand side of (29) is the lower order norm $\left\|w-w^d\right\|_{-1}$, defined on the *dual space* of $Q = H^1_{0,\Gamma_E}(\Omega)$ **..**

Standard application of the well known *Aubin-Nitsche duality argument* [1,2], shows that the rate of convergence with respect to lower order norms is generally greater than the rate with respect to the solution space norm $\left\|*\right\|_1$. In fact, depending on the regularity of the solution of the respective *auxiliary* (*dual*) *problem* [1,2], the rate improvement for $\left\|w-w^d\right\|_{-1}$ may be (at most) 2 orders. The latter conforms to the numerical results regarding the model problems with irregular solutions, see section 5.

5. COMPUTATIONAL RESULTS

For the verification of the above theoretical results, a FORTRAN program was assembled. The results of the code are, among others, the relative error in the H^1 seminorm for *u* and *w*. The following log – log figures show the comparison between *h*- and *p*- extensions for the relative error for both variables. The horizontal axe shows the total number of degrees of freedom, including the hierarchical ones while the vertical axe shows the relative error in the H^1 seminorm . Standard Gauss integration scheme is employed, except for the irregular force vector terms, for which exact integration is performed. Generally, the mixed method converges fast and achieves very low error levels, except for the variable *w* in the singular problems. However, in all cases the mixed finite element solutions are very close to the *optimal projections*. Moreover, the *p*- extension is much more efficient than the *h*-, in the sense of both accuracy and convergence rate, versus the number of degrees of freedom.

For the problem of simple bending of a clamped ends bar, two model displacement fields (solutions) are used. The first solution (IEX 152) is a smooth displacement field equal to $u(x) = x(x-1)sin(\pi x)$ using variable elastic properties $\left(J(x) = e^x\right)$. It is clear, that the *h*- extension has an algebraic asymptotic convergence rate whereas the *p*- extension converges exponentially, as it is theoretically anticipated [6-8], see figure 1.

The second solution (IEX 1541) is an irregular displacement field with a singular force vector and variable elastic properties $\left(J(x) = e^x\right)$. The displacement field has the form $u(x) = \mathrm{O}\left(x^{\lambda+2}\right)$ and the force vector $f(x) = \mathrm{O}\left(x^{\lambda-2}\right)$. The results of IEX 1541 for λ=0.85 are displayed in figure 2. The observed convergence rates are also of the anticipated form. More specifically, the asymptotic rates for the *h*- extension are $\min(p, \lambda - 0.5)$ for variable *w* and $\min(p, \lambda + 1.5)$ for the variable *u* [6-8]. Note that the observed rate of convergence for the

variable *u* is two orders higher than that of *w*, which is theoretically confirmed by theorem 4.3. It should be also mentioned, that the theoretical rates for the *p*- extension are two times the ones of the *h*- extension.

The results for the problem of the gradient elastic bar in tension (IEX 160), are displayed in figure 3. The theoretical convergence rates, which are the same as those for IEX 152, are also observed in the current numerical implementation.

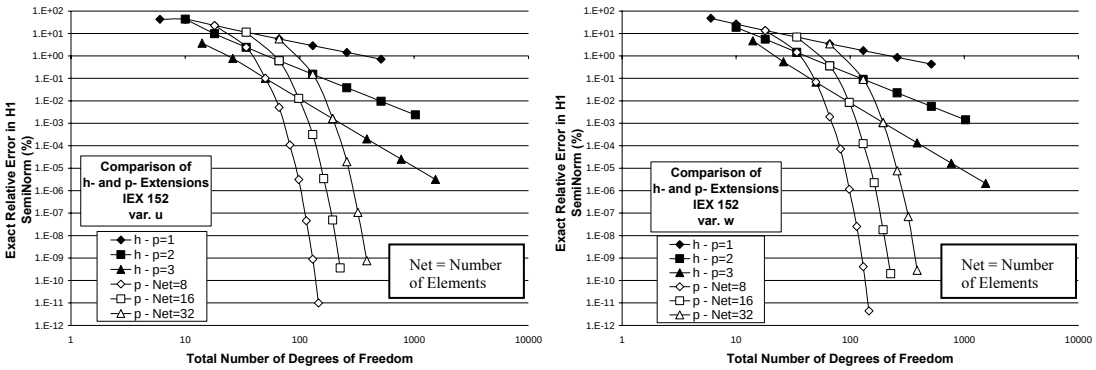


figure 1 : Comparison of the error of h and p extensions for both variables *u* and *w* (solution IEX 152,Clamped Ends Bar)

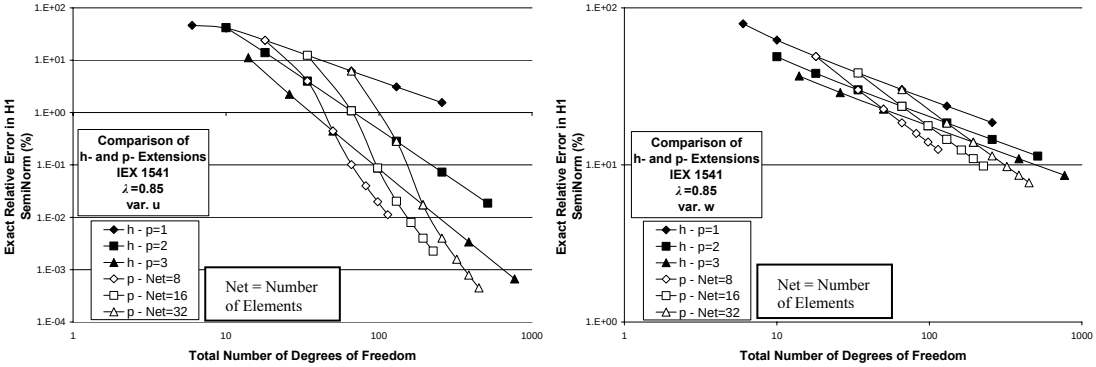


figure 2 : Comparison of the error of h and p extensions for both variables *u* and *w* (solution IEX 1541, λ = 0.85)

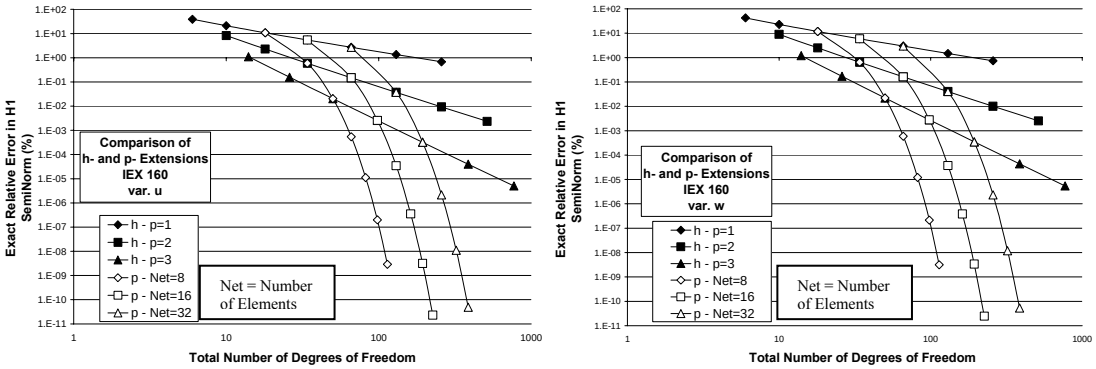


figure 3 : Comparison of the error of h and p extensions for both variables *u* and *w* (solution IEX 160, Gradient Elasticity)

6. CLOSING DISCUSSION AND FUTURE RESEARCH DIRECTIONS

Theoretical analysis of mixed C^0 continuity, conforming finite element formulations, based on the *Ciarlet-Raviart* method, has been performed for some types of one-dimensional biharmonic boundary value problems. The interpolation order is the same for both main variables. Numerical experimentations based on the *h- and p-extensions* confirm the quasi-optimality of the method. The standard asymptotic rates of convergence of the finite element error are generally observed. In terms of the quality of approximation, the performance of the *p-extension* is much higher than that of the *h-extension*.

The basic goal of the numerical results is to provide a basis for the application of similar mixed formulations, with *p-extension*, C^0 continuity, conforming approximations, in relevant multi-dimensional problems. However, as previously noted, the direct extension of *Ciarlet-Raviart* method based techniques in multi-dimensional problems presents a basic difficulty in terms of theoretical analysis. The *continuous kernel* is infinite dimensional. Moreover, it is easy to see that the respective *BB conditions* (10), (20) are not valid [4,5]. The *a priori* error analysis of such mixed methods employs further assumptions regarding the regularity of the

respective exact solutions [1,4,5]. A more general technique is used in [4], where the continuous formulation 3.3 is modified, by enlarging the solution space for u and reducing the solution space for w. In principle one could employ similar techniques to prove the convergence of the *p-extension* in many dimensions.

Another way to go about is to use various *patch test* techniques [19,13], to validate the respective multi-dimensional formulations. Moreover, approximations of the numerical behaviour of the *BB constants*, under certain types of refinements can be employed [18,20]. This can be helpful in conjunction with the standard error estimate (21), which is always valid, to check the convergence of the current refinement process.

The solution process employed in this work, to solve the discrete system of equations resulting from (19), was the standard *Gauss-elimination* with pivoting (direct solver). However, due to the nature of the equations, other more sophisticated techniques may be explored [14-16]. The latter is necessary in large scale problems. It is finally mentioned that, other important research directions are related *to a posteriori error* estimation [21] and *adaptive techniques*[22], in connection with the mixed formulations employed in the current work. The adaptive techniques are of vital importance in large scale problems, where it is necessary to achieve the desired accuracy with as few degrees of freedom as possible.

References

[1] Ciarlet, P. G. (1978), *The Finite Element Method for Elliptic Problems*, North-Holland Publishing Company, Amsterdam,

[2] Braess , D. (1997), *Finite elements*, (Cambridge University Press, United Kingdom,

[3] Xiao-liang Cheng, Weimin Han, Hong-ci Huang (2000), *Some mixed finite element methods for biharmonic equation*, Journal of Computational and Applied Mathematics, 126, pp. 91-109

[4] Babuška, I., Osborn, J., Pitkaranta, J. (1980), *Analysis of mixed methods using mesh dependent norms*, Math. Comput. 35, pp. 1039-1062

[5] Balasundaram, S., Bhattacharyya, P. K. (1984), *A mixed finite element method for fourth order elliptic equations with variable coefficients*, Computers Math. Applic. 10, pp. 245-256

[6] Szabo, B., Babuška, I. (1991), *Finite Element Analysis*, John Wiley & Sons, Inc. New York,

[7] Babuška, I., Szabo, B. A., Katz, I. N. (1981), *The p-version of the finite element method*, SIAM J. Numer. Anal. 18(3), pp. 515-545

[8] Babuška, I., Suri, M. (1987), *The optimal convergence rate of the p-version of the finite element method*, SIAM J. Numer. Anal. 24(4), pp. 750-776.

[9] Tsepoura, K. G., Papargyri-Beskou, S., Polyzos, D., Beskos, D. E. (2002), *Static and dynamic analysis of a gradient elastic bar in tension*, Archive of Applied Mechanics 72, pp. 483-497

[10] Mindlin, R. D., Eshel, N. N. (1968), *On first-gradient theories in linear elasticity*, Int. J. Solids Struct. 4, pp. 109-124

[11] Georgiadis, H.G., (2003), *The mode III crack problem in microstructured solids governed by dipolar gradient elasticity : static and dynamic analysis*, Journal of Applied Mechanics 70, pp. 517-530

[12] Bleustein, J. L. (1967), *A note on the boundary conditions of Toupin's strain-gradient theory*, Int. J. Solids Structures 3, pp. 1053-1057

[13] Amanatidou, E., Aravas, N. (2002), *Mixed finite element formulations of strain-gradient elasticity problems*, Comput. Methods Appl. Mech. Engrg. 191, pp. 1723-1751

[14] Kulshreshtha, Kshitij, Nataraj, Neela, Jung, Michael (2004), *Performance of a parallel mixed finite element implementation for fourth order clamped anisotropic plate bending problems in distributed memory enviroments*, Applied Mathematics and Computation 155, pp. 753-777

[15] M. B. van Gijzen (1995), *Conjugate gradient-like solution algorithms for the mixed finite element approximation of the biharmonic equation, applied to plate bending problems*, Comput. Methods Appl. Mech. Engrg. 121, pp. 121-136

[16] Milan D. Mihajlovic, David J. Silvester (2004), *Efficient parallel solvers for the biharmonic equation*, Parallel Computing 30, pp. 35-55

[17] Babuška, I. (1971), *Error-bounds for finite element method*, Numer. Math. 16, pp. 322-333

[18] Brezzi, F., Bathe, K. J. (1990), *A discourse on the stability conditions for mixed finite element formulations*, Comput. Methods Appl. Mech. Engrg. 82, pp. 27-57

[19] Zienkiewicz, O. C, Qu, S., Taylor, R. L., Nakazawa, S. (1986), *The patch test for mixed formulations*, Int. J. Numer. Methods Engrg. 23 , pp. 1873-1883

[20] Tsamasphyros, G., Markolefas, S. (2003), *An estimate of the Babuška-Brezzi inf-sup discrete stability constant for general linear Petrov-Galerkin finite element formulations*, Applied Mathematics and Computation 144 , pp. 107-116

[21] Ainsworth, M., Oden, J. T. (1997), *A posteriori error estimation in finite element analysis*, Comput. Methods Appl. Mech. Engrg. 142 , pp. 1-88

[22] Fish, J., Markolefas, S. (1993), *Adaptive s-method for linear elastostatics*, Comput. Methods Appl. Mech. Engrg 104 , pp. 363-396

Boundary Elements

5th GRACM International Congress on Computational Mechanics
Limassol, 29 June-1July, 2005

QUADRATURE FORMULA FOR INTEGRALS WITH NEARBY SINGULARITIES

G. Tsamasphyros and E.E. Theotokoglou

Faculty of Applied Sciences, Dept. of Mechanics-Lab. of Strength of Materials
The National Technical Universtiy of Athens, Zoghraphou Campus,
Theocaris Bld., GR-0157 73, Athens, Greece
email: stathis@central.ntua.gr

Key words: Nearby poles, Quadrature formula, Lagrangian interpolation, Elasticity, Boundary element method, Numerical integration.

Abstract. *The purpose of this paper is to propose a new quandrature formula for integrals with nearby singularities. In the Boundary Element Method, the integrands of nearby singular boundary integrals vary drastically with the distances between the field and the source point. Especially field variables and their derivatives at a field point near a boundary cannot be computed accurately . In the present paper a quadrature formulas for ℓ isolated singularities near to the integration interval, based on Langrange interpolatory polynomials, is obtained. Quadrature formulas for regular integrals with conjugate poles are also derived. Numerical examples are given and the proposed quadrature rules present the expected polynomial accuracy.*

1 INTRODUCTION

In the Boundary element method [1-4] if the distance the source point and the nearest field point is small, the integrand takes abruptly a great value. The difficulties arised from the fact that the integrands of integrals along the boundary vary rapidly with the distance when the field point approaches the source point. This strong gradient of the integrand creates difficulties in the accuracy of the standard quadrature formulas. In fact using standard quadrature procedures, which neglect this pathological behaviour of the integrand leads to a computational error which increases as the source point approaches the boundary: this is the so-called **boundary layer effect.** If the distance is very small this error is very important and surprisingly is much more important as the number of integration points increase. We say that these integrals present "**a nearby singularity"** or a quasi-singularity.

Strong, weak or nearby singularities are treated in two ways. The first, which is dominated by the Sladek's work [5], try to avoid singularities by regularizing the integral and the integral equations. In the same lines are procedures trying to avoid singularities [6,7]. Some of these methods concern strong singularities but they may be expanded easily to nearby singularities. The other way to proceed with these integrals is based on quadrature rules for each of these integrals. We can distinguish procedures based on transformation which remedy the singularities [8-9] and on procedures based on general quadrature rules [10-15] (special quadratures rules). In general, there is a prejudice concerning the last procedure. Many authors claim that special quadrature rules are very complicated and not very efficient. Nothing is so wrong. The truth is that special quadrature rules are able to give the exact result with only few integration points. On the other hand, the determination of the weights and the integration points, is, in general, a simple procedure.

The general form of an integral with nearby and strong singularities can be put in the following form

$$I(\varphi)=\int_{-1}^{1}\frac{w(t)\varphi(t)}{\omega(t)}dt \tag{1}$$

with

$$\omega(t)=\prod_{i=1}^{\ell}(t-y_i),\quad y_i=x_i+i\varepsilon_i \tag{2}$$

where ε_i is small or zero and x_i is on [-1,1] or very close to this interval and $w(t)$ the weight function which is fixed positive and integrable on $[-1,1]$, given by

$$w(t)=(1-t)^{\alpha}(1+t)^{\beta}\ , \quad \alpha,\beta>-1\,, \tag{3}$$

The previous form (3) concerns not only line integrals but also surface integrals if the later can be reduced to one-dimensional integrals and then the form (3) can be applied. Obviously any integral with respect the arc length s of a part of the boundary or of the whole boundary can be transformed to an integral along $(-1,1)$.

In the present study, we have expanded the known **Gauss-integration rules** as to confront the above problem where many strong or weak singularities can coexist. For the sake of generality we have considered the most general case by supposing that, for some reasons, we need to evaluate the unknown function at some preassigned nodes.. A "no-Gauss", quadrature formula has been proposed. This quadrature formula is produced from a simple Lagrange interpolation $\varphi(t)$.

Based on the proposed rule, we have also constructed modified quadrature formulas, for regular integrals with conjugate poles $\int_{-1}^{1}\frac{w(t)\varphi(t)}{t^2+c^2}dt\ \ (c<<1)$.

The obtained modified quadrature formulas is successfully tested in numerical examples, where it is proved the necessity of using the modified rule when the pole of the regular integral Cauchy-type integrals, are very close to the integration interval.

2 NO-GAUSS QUADRATURE FORMULA

Let us give some necessary definitions:

We denote by D a simply connected domain containing $(-1,1)$ in its interior.

We denote by $\varphi(z)$ the analytic continuation of $\varphi(t)$ into D.

Let C a simple contour inside D, presenting angular points at -1 and 1 with angles α and β respectively. These angular points are due to the behaviour of $\varphi(z)$.

Let $\{t_j\}_{j=1}^{n}$, $\{z_k\}_{k=1}^{m}$ two sets of points inside D, belonging or not to the integration interval. These points will be used as interpolation points, i.e.

$$\varphi(\tau_j)=\varphi_{n+m}(\tau_j);\quad j=1,2,...,(n+m) \tag{4a}$$

where $\varphi_{n+m}(z)$ is a complex polynomial of degree $(n+m-1)$ that interpolates $\varphi(z)$ at the points

$$\tau_j=\begin{cases} t_j\ ; & 1\le j\le n \\ z_k\ ; & j=n+k\le n+m \end{cases} \tag{4b}$$

We suppose that z_k are preassigned, whereas t_j can be selected arbitrarily.

We denote by

$$\pi_n(t)=\prod_{j=1}^{n}(t-t_j),\ \Omega(t)=\prod_{k=1}^{m}(t-z_k),\ \Pi_{n+m}(t)=\pi_n(t)\Omega(t),\ W(t)=w(t)\Omega(t). \tag{5}$$

Using the interpolation formula (4), and taking into consideration Lagrange procedure [16,17], we have

$$\varphi(t)=\varphi_{n+m}(t)+R_{n+m}(t) \tag{6}$$

where

$$\varphi_{n+m}(t)=\sum_{j=1}^{n}\frac{\Pi_{n+m}(t)}{(t-t_j)\Omega(t)\pi_n'(t_j)}\varphi(t_j)+\sum_{k=1}^{m}\frac{\Pi_{n+m}(t)}{(t-z_k)\Omega'(z_k)\pi_n(z_k)}\varphi(z_k) \tag{7}$$

or in more compact form

$$\varphi_{n+m}(t)=\sum_{j=1}^{n+m}\frac{\Pi_{n+m}(t)}{(t-\tau_j)\Pi_{n+m}'(\tau_j)}\varphi(t_j) \tag{8}$$

and

$$R_{n+m}(t)=\frac{\Pi_{n+m}(t)}{2\pi i}\oint_C\frac{\varphi(z)}{(z-t)\Pi_{n+m}(z)}dz \tag{9}$$

Multiplying both sides of (7) by $w(t)/\omega(t)$ and integrating on $[-1,1]$, we get

$$\int_{-1}^{1}\frac{w(t)}{\omega(t)}\varphi(t)dt=Q_{n+m}(\varphi)+E_{n+m}(\varphi) \tag{10a}$$

where

$$Q_{n+m}(\varphi)=\sum_{j=1}^{n+m}\Lambda_j\varphi(\tau_j) \tag{10b}$$

is the quadrature formula,

$$\Lambda_j=\frac{1}{\Pi_{n+m}'(\tau_j)}\int_{-1}^{1}\frac{w(t)\Pi_{n+m}(t)}{(t-\tau_j)\omega(t)}dt=\frac{1}{\Pi_{n+m}'(\tau_j)}\int_{-1}^{1}\frac{W(t)\pi_n(t)}{(t-\tau_j)\omega(t)}dt \tag{11}$$

are the weights of the quadrature formula, and

$$E_{n+m}(\varphi)=\frac{1}{2\pi i}\oint_C\frac{\varphi(z)}{\Pi_{n+m}(z)}\left(\int_{-1}^{1}\frac{W(t)\pi_n(t)}{(z-t)\omega(t)}dt\right)dz \tag{12}$$

is the error

Taking into consideration that

$$\frac{1}{(t-\tau_j)\omega(t)}=\frac{1}{\omega(\tau_j)(t-\tau_j)}+\sum_{i=1}^{\ell}\frac{1}{\omega'(y_i)(y_i-\tau_j)(t-y_i)} \tag{13}$$

relations (11), (12) can be written as:

$$\Lambda_j = \lambda_j \left[\frac{1}{\omega(\tau_j)} - \frac{1}{\psi_n(\tau_j)} \sum_{i=1}^{\ell} \frac{\psi_n(y_i)}{\omega'(y_i)(\tau_j - y_i)} \right] \tag{14}$$

and

$$E_{n+m}(\varphi) = \frac{1}{\pi i} \oint_C \frac{\varphi(z)}{\Pi_{n+m}(z)} \left(\frac{\psi_n(z)}{\omega(z)} - \sum_{i=1}^{\ell} \frac{\psi_n(y_i)}{\omega'(y_i)(z - y_i)} \right) dz \tag{15}$$

where

$$\lambda_j = \frac{-2\psi_n(\tau_j)}{\Pi'_{n+m}(\tau_j)} \tag{16}$$

is the weight of the standard interpolatory quadrature formula for regular integrals, and

$$\psi_n(z) = \frac{1}{2} \int_{-1}^{1} \frac{w(t)\Pi_{n+m}(t)}{z - t} dt = \frac{1}{2} \int_{-1}^{1} \frac{W(t)\pi_n(t)}{z - t} dt \,; \quad z \notin [-1,1] \tag{17}$$

is the associated (to $\pi_n(t)$) function, with

$$\psi_n(\xi) = \frac{1}{2} [\psi_n(\xi + i0) + \psi_n(\xi - i0)], \quad \xi \in (-1,1) \tag{18}$$

At this point we may have that:

$\pi_n(t)$ is the orthogonal polynomial with respect to $W(t)$ (no-Gauss formula)

Thus

$$\int_{-1}^{1} W(t)\pi_n(t)\pi_k(t) dt = h_n \delta_{nk} \tag{19}$$

or

$$\int_{-1}^{1} W(t)\pi_n(t) t^k dt = 0, \quad k \le n-1. \tag{20}$$

3 INTEGRALS WITH NEARBY POLES

It is suppored that:

i. There are not preassigned nodes, i.e. , $\Omega(t) = 1$.

ii. The weight is of the form, $w(t) = (1-t)^{\alpha}(1+t)^{\beta}$.

iii. The integration points are the roots of the corresponding orthogonal polynomial i.e. the Jacobi polynomial $P_n^{(\alpha,\beta)}(t)$, and

$$\int_{-1}^{1} (1-t)^{\alpha}(1+t)^{\beta} P_n^{(\alpha,\beta)}(t) P_m^{(\alpha,\beta)}(t) dt = h_n^{(\alpha,\beta)} \delta_{nm} \tag{21}$$

$$h_n^{(\alpha,\beta)} = \frac{2^{1-\kappa}\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{(2n-\kappa+1)n!\Gamma(n-\kappa+1)}, \quad \kappa = -(\alpha+\beta), \quad \alpha,\beta > -1$$

or the roots of the orthogonal polynomial $p_n^{(\alpha,\beta)}(t)$ with respect to the weight $w(t)/\omega(t)$.

$$\int_{-1}^{1} \frac{(1-t)^{\alpha}(1+t)^{\beta}}{\omega(t)} p_n^{(\alpha,\beta)}(t) t^k dt = 0\,, \quad k \le n-1 \tag{22}$$

It is considered the regular integrals

$$I_1(\varphi;c) = \int_{-1}^{1} \frac{w(t)\varphi(t)}{t^2+c^2} dt = \int \frac{w(t)\varphi(t)}{(t+ic)(t-ic)} dt\,. \tag{23}$$

with conjugate poles at $\pm ic$, where $c << 1$.

Using formulas (10) and (14), we have

$$Q_1^{nG}(\varphi;c) = \sum_{j=1}^{n} \Lambda_j \frac{\varphi(t_j)}{t_j^2+c^2} \tag{24}$$

and the modified weights Λ_j, are given by

$$\Lambda_j = \lambda_j - \frac{1}{P_n'^{(\alpha,\beta)}(t_j)} \int_{-1}^{1} \frac{w(t)(t+t_j)}{t^2+c^2} P_n^{(a,\beta)}(t) dt \tag{25a}$$

where $\{t_j\}_{j=1}^{n}$ are the zeroes of the Jacobi polynomial $P_n^{(\alpha,\beta)}(t)$ corresponding to $w(t)$ [18], and

$$\lambda_j = -\frac{2\psi_n^{(\alpha,\beta)}(t_j)}{P'^{(\alpha,\beta)}(t_j)}, \qquad \psi_n^{(\alpha,\beta)}(z) = \frac{1}{2}\int_{-1}^{1} \frac{(1-t)^{a}(1+t)^{\beta} P_n^{(\alpha,\beta)}(t)}{z-t} dt \tag{25b}$$

Three differed kinds of polynomials have been considered:

a. Legendre polynomials $(w(t)=1)$

Using formula (25), we have

$$\Lambda_j = \lambda_j \left\{ 1 - \frac{1}{\lambda_j P'(t_j)} \left(\frac{2}{c}\tan^{-1}\frac{1}{c} - \sum_{k=1}^{k} \frac{\lambda_k}{t_k^2+c^2} \right) \left(t_j \,\mathrm{Re}(P_n(ic)) - c\,\mathrm{Im}(P_n(ic)) \right) \right\} \tag{26}$$

where λ_j, λ_k and t_j, t_k $(j,k = 1,2,...,n)$ being the weights and the integration points respectively, of the classical Gauss-Legendre formula [18] arising from the classical Gauss-Jacobi quadrature formula $(a = \beta = 0)$,and P_n (t) being the Legendre polynomial of the second kind .

b. Chebyshev polynomials of the first kind $\left(w(t) = (1-t^2)^{-1/2}\right)$

If the weight $w(t)$ is of the form

$$w(t) = (1-t^2)^{-1/2}\,,$$

we denote by $T_n(t)$ and $U_{n-1}(t)$ the Chebyshev Polynomials of the first and second kind respectively [16,17], which are orthogonal with respect the weights $w(t)=\left(1-t^2\right)^{-1/2}$ and $w(t)=\left(1-t^2\right)^{1/2}$ respectively.

From (25) we get:

$$\Lambda_{j_{Cf}}=\lambda_j\left\{1+\frac{n}{T_n'(t_j)}\frac{\left(c-\sqrt{c^2+1}\right)^n}{\sqrt{c^2+1}}\left(\frac{t_j}{c}\sin\frac{(n-1)\pi}{2}+\cos\frac{(n-1)\pi}{2}\right)\right\} \tag{27}$$

with

$$t_j=\cos\left(\frac{2j-1}{2n}\pi\right),\quad \lambda_j=\frac{\pi}{n};\quad j=1,2,...,n \tag{28}$$

c. Chebyshev polynomials of the second kind $\left(w(t)=\left(1-t^2\right)^{1/2}\right)$

Taking into consideration (25), it is obtained:

$$\Lambda_{j_{Cs}}=\lambda_j\left\{1+\frac{(n+1)\left(c-\sqrt{c^2+1}\right)^{n+1}}{\left(\sin\left(\frac{j\pi}{n+1}\right)\right)^2U_n'(t_j)}\left(\frac{t_j}{c}\sin\frac{(n+1)\pi}{2}+\cos\frac{(n+1)\pi}{2}\right)\right\} \tag{29}$$

with

$$t_j=\cos\frac{j\pi}{n+1},\quad \lambda_j=\frac{\pi}{n+1}\left\{\sin\frac{j\pi}{n+1}\right\}^2;\quad j=1,2,...,n \tag{30}$$

4 NUMERICAL APPLICATIONS

The proposed quadrature formula has been applied to two test functions, in order to have the opportunity to compare the proposed rule with the classical Gauss formula in the case that $c=0.01$.

Test function (i), is the evaluation of the integral

$$I_{1_{Cf}}=\int_{-1}^{1}\frac{dt}{\sqrt{1-t^2}\left(t^2+c^2\right)}=\frac{\pi}{c\sqrt{1+c^2}}\cong 314.143559 \tag{31}$$

$I_{1_{Cf}}(\varphi;0.01)\cong 314.143559$		
n	Classical Gauss-Chebyshev Quadrature rule of the first kind	Modified weight (27)
2	3.140965	314.143559
3	10474.767670	314.143559
4	12.558836	314.143559

Table 1 : Differences between the classical Gauss-Chebyshev Quadrature formula of the first kind, and the modified weight (27)

In Table 1, we have the values obtained by using the classical Gauss-Chebyshev quadrature formula of the first kind [18] and the modified weight quadrature where the integration points are determined from (28). It is observed that the *"classical"* Gauss-Chebyshev rule is impossible to approximate the correct result for a few integration points whereas the modified gives very good results for $n=1$.

Test function (ii), is the evaluation of the integral

$$I_{1_{Cs}}=\int_{-1}^{1}\frac{\sqrt{1-t^2}}{t^2+c^2}dt=\pi\frac{\sqrt{c^2+1}-c}{c}\cong 311.033380 \tag{32}$$

$I_{1_{Cs}}(\varphi;0.01)\cong 311.033380$		
n	Classical Gauss-Chebyshev Quadrature rule of the second kind	Modified weight (29)
1	15707.963270	311.033380
2	6.280673	311.033380
3	7855.552116	311.033380

Table 2 : Differences between the classical Gauss-Chebyshev Quadrature formula of the second kind, and the modified weight (29)

The integration points of the modified weight Gauss Chebyshev quadrature formula of the second kind (29), are determined from relations (30). From Table 2, it is observed that the modified weight Gauss-Chebyshev quadrature formula gives very good results for $n=1$.

5 CONCLUSIONS

In our study a "no-Gauss" quadrature formulas has been proposed. Based on this rule, quadratures formulas that ensure the exact calculation of integrals of the form, $\int_{-1}^{1}\frac{w(t)\varphi(t)}{t^2+c^2}dt\ \ (c<<1)$ and, where $\varphi(t)$ is a regular function, are resulted. These integrals are appeared in a lot of problems in mathematical physics and in engineering.

The proposed, quadrature formulas is derived from a Langrangian interpolatory procedure, where integration points are not taken at the poles $\pm ic\ \ (c<<1)$. This formulas is called "no-Gauss" formulas because it results in a modification of the weights relative to the classical Gauss weights. The "no-Gauss" formulas has the advantage, that it may be used for the solution of integral equations. The disadvantage of this formula, is that a computation effort is needed in order to construct the modified weights.

The proposed modified formula is very effective in the case that the regular integrals, is given by (23), where $\varphi(t)$ is a holomorphic function in the complex z -plane or in the worst case a function with removable singularities at ± 1. The modified formulas have been arised with a procedure based on preassigned nodes, that is similar with the technique for the production of Lobatto or Radau formulas. With our analysis we have proved in very simply way, a quadrature formula applied to regular and also to singular integrals with nearby singularities.

The proposed quadrature formulas may also be used in three-dimensional problems in the case that the surface integrals can be evaluated by repeated one-dimensional integrations.

REFERENCES

[1] Theocaris, P.S., Tsamasphyros, G. (1975), "On the solution of boundary-value problems in the plane elasticity for multiply-connected regions (I: The second boundary-value problem)", Lett. Appl. Eng. Sci. 3, pp. 167-176.

[2] Brebbia C.A., Telles J.C.F., Wrobel L.C. (1984) , *Boundary element techniques-theory and applications in engineering*, Springer, Berlin, Heidelberg New York.

[3] Ioakimidis, N.I. (1982), "Application of finite-part integrals to the singular integral equations of crack problems in plane and three-dimensional elasticity" Acta Mechanica 45, pp. 31-47.

[4] Tsamasphyros G. (1994), "Numerical methods for fracture parameters calculation." In: *Handbook of 1igue Crack Propagation in Metallic Structures*, Elsevier Science B.V., pp. 107-147.

[5] Sladek N., Sladek J., Tanaka M. (1993), "Regularization of hypersingular and nearly singular integrals in the potential theory and elasticity", Int. J. Numer. Methods Eng. 36, pp. 1609-1628.

[6] Krishnasamy G., Schmerr L.W., Rudolphi T.J., Rizzo F.J. (1990), "Hypersingular boundary integral equations: some applications in acoustic and elastic waves scattering", ASME J. Appl. Mech. 57, pp. 404-414.

[7] Yun, B.I. (2003), " A composite transformation for numerical integration of singular integrals in the BEM", International Journal for Numerical Methods in Engineering 57, pp. 1883-1898.

[8] Johnston P.R., Elliott, D. (2000), " Error estimation of quadrature rules for evaluating singular integrals in boundary element problems", Int. J. Numer. Methods Eng. 48, pp. 949-962.

[9] Ma, H., Kamiya, N. (2002), "A general algorithm for the numerical evaluation of nearly singular boundary integrals of various orders for two- and three-dimensional elasticity", Computational Mechanics 29, pp. 277-288.

[10] Paget, D.E.; Elliott, D. (1972) , "An algorithm for the numerical evaluation of certain Cauchy principal value integrals", Numer. Math. 19, pp. 373-385.

[11] Tsamasphyros, G., Theocaris, P.S. (1976a), "Numerical solution of systems of singular integral equations with variable coefficients", Report. National Technical University of Athens, November 1976, also (1980) Appl. Analysis 9, pp. 37-52.

[12] Ioakimidis, N.I. (1983), "A direct mehod for the construction of Gaussian quadrature rules for Cauchy type and finite part integrals" L' Analyse Numer. Et la Theorie de l' Approx. 12, pp.131-140.

[13] Monegato G., (1986), " Quadrature formulas for functions with poles near the interval of integration", Math. Computation, 47, pp. 137-158.

[14] Tsamasphyros, G., (1986), "A study of factors influencing the solution of singular integral equations", Engineering Fracture Mechanics 24, pp. 567-578.

[15] Tsamasphyros, G. and Dimou G. (1990), "Gauss quadrature rules for finite part integrals", Int. J. Numer. Methods Eng. 30, pp. 13-26.

[16] Isaacson, E., Keller, H.B. (1966), *Analysis of numerical methods*, London, Wiley.

[17] Hildebrand, F.B. (1974) , *Introduction to numerical analysis*, 2nd ed., New York, McGraw-Hill.

[18] Stroud, A. and Secret, D. (1966), *Gaussian quadrature formulas*, Prentice-Hall, Englewood Cliffs, N.J.

PERFORMANCE OF SEVERAL RBFS IN DR/BEM EIGENVALUE ANALYSIS OF 2-D STRUCTURES

Christopher G. Provatidis

School of Mechanical Engineering, National Technical University of Athens
9 Iroon Polytechniou Str, Zografos Campus, GR-15773 Athens, Greece
e-mail: cprovat@central.ntua.gr, web page: http://users.ntua.gr/cprovat

Keywords: Dual Reciprocity Boundary Element Method; Radial Basis Functions; Eigenvalues; Elastodynamics

Abstract. *This paper investigates the performance of several radial basis functions (RBFs) for the eigenvalue extraction of 2-D structures using the dual reciprocity boundary element method. The traditional conical RBF, high-order, thin plate splines, multiquadratics and Gaussian (exponential) RBFs are investigated. Moreover, a new functional set based on Pascal's triangle, which was previously applied on low-order serendipity elements, is extended and proposed as an alternative basis. In this paper the numerical investigation restricts to a simple square domain and, therefore, further research needs to be conducted on more complex shapes.*

1 INTRODUCTION

The pioneering work of Nardini and Brebbia[1] opened new horizons in the Boundary Element Method (BEM) and particularly in dynamic analysis of structures. Before that work, the BEM eigenvalue extraction was a non-algebraic problem that required the graph of the determinant and the numerical determination of its roots. In 1982 the goal was to derive a mass matrix [**M**] that could be combined with the static ([**H**] and [**G**]) influence matrices[2], thus leading to an algebraic eigenvalue problem. Later, the method was successfully applied to 3-D solid structures [3], plate bending[4], heat transfer transient analysis[5], acoustics[6] and so on.

The DR/BEM uses global basis functions, called radial basis functions (RBF), to approximate the inertia (generally time-dependent) terms of the partial differential equation to be solved. Usually, the RBFs are closed-form polynomials, logarithms or exponentials, of which dual functions are available. As a result, the mass matrix can be transformed to a boundary integral, usually expressed in terms of [**H**] and [**G**]. However, the conical RBF in DR/BEM has some flaws. In acoustics, the sum of the elements in mass matrix is calculated smaller than the real one but the convergence is rather good [7]. In elasticity, the relevant sum was found to be again smaller than the real surface times the mass density, deviating between 4.3% and 1.7% [8]. Despite this shortcoming, it is accepted that, in general, the DR/BEM behaves well [9-12].

The purpose of this paper is to compare the several RBFs, to investigate their performance, and report on some shortcomings in eigenvalue extraction of 2-D elastic structures. Apart from well-known RBFs (conical, high-order, Gaussian, thin plate splines and multiquadratics), a blending-function interpolation will be tested.

2 GENERAL THEORY

2.1 DR/BEM formulation

In the absence of body forces, the stress equilibrium becomes:

$$\sigma_{ij,j} - \rho \ddot{u}_i = 0 \tag{1}$$

The reciprocity Betti-Maxwell principle, which correlates two independent states, i.e. the actual $\left(u_i, \sigma_{ij}, p_i\right)$ and the virtual $\left(u_i^*, \sigma_{ij}^*, p_i^*\right)$, is written as follows:

$$\int_\Omega \left(u_i^* \sigma_{ij,j} - u_i \sigma_{ij,j}^*\right) d\Omega = \int_\Gamma \left(u_i^* p_i - u_i p_i^*\right) d\Gamma \tag{2}$$

The state $\left(u_i^*, \sigma_{ij}^*, p_i^*\right)$ is chosen so that to fulfill the static part of the elastodynamical partial differential equation (1), and concretely to correspond to a concentrated unit load, that is

$$\sigma_{kij,j}^* + \delta_{ki} \Delta\left(\mathbf{x} - \mathbf{y}\right) = 0 \tag{3}$$

where δ_{ki} denotes Kronecker's delta and Δ is Dirac function, which fulfils the relationships:

$$\begin{aligned} &\Delta\left(\mathbf{x} - \mathbf{y}\right) = 0 \ \text{ for } \ \mathbf{x} \neq \mathbf{y} \\ &\int_\Omega \Delta\left(\mathbf{x} - \mathbf{y}\right) d\Omega = 1 \end{aligned} \tag{4}$$

The solution of Eq(3) is denoted by u_{ki}^*, and the corresponding tractions by p_{ki}^*. Taking into consideration Eq(3)

and (4), Eq(2) leads to

$$\int_{\Omega} \sigma_{ij,j} u_{ki}^{*} d\Omega = -c_{ki} u_i + \int_{\Gamma} u_{ki}^{*} p_i d\Gamma - \int_{\Gamma} p_{ki}^{*} u_i d\Gamma \tag{5}$$

that is known as integral equation in elastostatics. By substituting $\sigma_{ij,j}$ from Eq(1), assuming a constant mass density ρ, Eq(5) becomes

$$\rho \int_{\Omega} \ddot{u}_i u_{ki}^{*} d\Omega = -c_{ki} u_i + \int_{\Gamma} u_{ki}^{*} p_i d\Gamma - \int_{\Gamma} p_{ki}^{*} u_i d\Gamma \tag{6}$$

Following the general DR/BEM concept[1], we assume that the inertial terms are approximated by

$$u_i(\mathbf{y},t) = a_i^j(t) f^j(\mathbf{y}) \tag{7}$$

where Einstein's convention is understood on the repeated index j=1 to m (it denotes summation over j).

Therefore, the corresponding accelerations will be:

$$\ddot{u}_i(\mathbf{y},t) = \ddot{a}_i^j(t) f^j(\mathbf{y}) \tag{8}$$

and, consequently, the domain integral in Eq.(6) becomes:

$$\int_{\Omega} \ddot{u}_i u_{ki}^{*} d\Omega = \ddot{a}_i^j \int_{\Omega} f^j u_{ki}^{*} d\Omega = \ddot{a}_l^j \int_{\Omega} \delta_{lk} f^j u_{ki}^{*} d\Omega \tag{9}$$

The domain integral in Eq(9) transforms to the boundary, if a *particular* solution of a static problem is found, that is a solution of the partial differential equation

$$\bar{\sigma}_{\lim,m} + \delta_{lk} f^j = 0 \tag{10}$$

The solution of Eq(10) is a displacement field denoting by ψ_{li}^j, and the corresponding traction by η_{li}^j. Using the transformation of Eq(5), the domain integral in Eq(9) reduces to a corresponding boundary one, as follows

$$\int_{\Omega} \delta_{lk} f^j u_{ki}^{*} d\Omega = -\int_{A\Omega} \bar{\sigma}_{\lim,m}^j u_{ki}^{*} d\Omega = c_{ki} \psi_{li}^j - \int_{\Gamma} u_{ki}^{*} \eta_{li}^j d\Gamma + \int_{\Gamma} p_{ki}^{*} \psi_{li}^j d\Gamma \tag{11}$$

By substituting Eq(11) into Eq(9) and recasting Eq(6), one obtains:

$$c_{ki} u_i + \int_{\Gamma} p_{ki}^{*} u_i d\Gamma - \int_{\Gamma} u_{ki}^{*} p_i d\Gamma + \rho \left(c_{ki} \psi_{li}^j + \int_{\Gamma} p_{ki}^{*} \psi_{li}^j d\Gamma - \int_{\Gamma} u_{ki}^{*} \eta_{li}^j d\Gamma \right) \ddot{a}_l^j = 0 \tag{12}$$

The above equation is a boundary integral formulation from which one can derive a numerical solution using a proper procedure.

2.2 Matrix formulation

By assuming a certain interpolation of both displacement and tractions (e.g., linear or quadratic elements), Eq(12) becomes:

$$\hat{\mathbf{M}} \cdot \ddot{\boldsymbol{\alpha}} + \mathbf{H} \cdot \mathbf{u} = \mathbf{G} \cdot \mathbf{p} \tag{13}$$

where **H** and **G** are the conventional static matrices, $\mathbf{u}, \mathbf{p}$ the displacement and traction, respectively, and also:

$$\hat{\mathbf{M}} = \rho \cdot (\mathbf{G} \cdot \boldsymbol{\eta} - \mathbf{H} \cdot \boldsymbol{\psi}) \tag{14}$$

Using the relationship between nodal displacements $\mathbf{u}$ and coefficients $\boldsymbol{\alpha}$ through a matrix **F**:

$$\mathbf{u} = \mathbf{F}\boldsymbol{\alpha} \tag{15}$$

and pre-multiplying both members of equation (13) by the inverse of matrix **G** (assuming a smooth boundary), it becomes

$$\bar{\mathbf{M}} \quad \ddot{\mathbf{u}}(t) + \bar{\mathbf{K}}\, \mathbf{u}(t) = \mathbf{p}(t) \tag{16}$$

where

$$\bar{\mathbf{M}} = \mathbf{G}^{-1} \hat{\mathbf{M}} \mathbf{F}^{-1} \tag{17}$$

$$\bar{\mathbf{K}} = \mathbf{G}^{-1} \mathbf{H} \tag{18}$$

For the purpose of this paper, we use a suitable matrix **L**, well-known from the BEM/FEM coupling procedures[13, p.274], which achieves to transform the time-dependent tractions $\mathbf{p}(t)$ into boundary nodal forces $\mathbf{f}(t)$ as follows

$$\mathbf{f}(t) = \mathbf{L}\, \mathbf{p}(t) \tag{19}$$

By left-multiplying both sides of Eq(16) by the matrix **L** and then considering Eq(19), the latter becomes:

$$\mathbf{M} \quad \ddot{\mathbf{u}} + \mathbf{K}\, \mathbf{u} = \mathbf{f}(t) \tag{20}$$

where

$$\mathbf{M} = \mathbf{L}\bar{\mathbf{M}}, \quad \mathbf{K} = \mathbf{L}\bar{\mathbf{K}} \tag{21}$$

Therefore, the DR/BEM leads finally to a matrix formulation similar to that of the FEM. Numerical experience, using equations (20,21) in static and dynamic problems with subregions, has been previously reported[7]. It

should become clear that the matrices **M** and **K** in Eq(13) are nonsymmetric but, nevertheless, most of the usual FEM procedures such as eigenvalue search, modal analysis and time-integration can be performed.

3. Radial Basis Functions (RBFs)

3.1 Conical functions

Nardini and Brebbia[1] proposed the functional set $f^j(\mathbf{y})$, briefly $f(r)$ with $r=r(A_j,\mathbf{y})$, for the approximation of inertial forces. Due to its dependency on the radial distance r between the source and field point, it was named *conical*:

$$f(r)=C-r \tag{22}$$

where C is a "suitably chosen constant" [1]. Unfortunately, the meaning of the *suitable choice* is not that clear. A specialized textbook[2] from the same group gives advantage to the global function $f=1+r$, while in a recent communication the originator claims that the constant C is not given by the user but "…it is found as a part of the solution to the problem"[14]. It is also remarkable that other authors[3] have proposed to choose C as the maximum distance between any two nodal points along the boundary. Also, from the relevant open literature it is not clear if C is globally defined, or if it obtains different values for a given couple of field and source points.

3.2 Gaussian functions

Yamada and Wrobel[15] have proposed the use of RBFs appearing an exponential decay as follows:

$f(r)=\exp(-r^2/c^2)$, where c is a positive constant (23)

3.3 High-degree radial functions

Wendland[16] proposed the following base functions:

Second-degree function:
$$f=\left(1-\frac{R}{d}\right)_+^2=\begin{cases}\left(1-\frac{R}{d}\right)^2, & R<d\\ 0, & \text{otherwise}\end{cases} \tag{24}$$

Fifth-degree function:
$$f=\left(1-\frac{R}{d}\right)_+^4\left(1+4\frac{R}{d}\right) \tag{25}$$

Eighth-degree function:
$$f=\left(1-\frac{R}{d}\right)_+^6\left[3+18\frac{R}{d}+35\left(\frac{R}{d}\right)^2\right] \tag{26}$$

Eleventh-degree function:
$$f=\left(1-\frac{R}{d}\right)_+^8\left[1+8\frac{R}{d}+25\left(\frac{R}{d}\right)^2+32\left(\frac{R}{d}\right)^3\right] \tag{27}$$

In equations (24)-(27), the variable d denotes the radius of compact support. In this paper, d was chosen to be equal to the diagonal (maximum distance between any two points) of the rectangular cavity tested. In other words, these high-degree basis functions were positive within the entire domain.

3.4 Thin Plate Splines (TPS)

Goldberg and Chen[17] proposed the following base functions: $f=r^2\ln r$ (28)

3.5 Multiquadratics (MQ)

Goldberg et al.[18] proposed the following base functions: $f=\sqrt{c^2+r^2}$ (29)

3.6 Polynomial expansion

Provatidis[6] has proposed a bivariate expansion in the form: $u(x,y;t)=\sum_{i=0}^{m}\sum_{j=0}^{n}a_{ij}(t)x^iy^j$,

(30)

In case of a rectangular domain $a\times b$ uniformly discretized into $m\times n$ segments, the proposed basis functions consist of $\{1,x,y,xy\}$, $\{x^2,x^3,\ldots,x^m\}$, $\{y^2,y^3,\ldots,y^n\}$ as well as $\{x^2,x^3,\ldots,x^m\}y$ and $\{y^2,y^3,\ldots,y^n\}x$, which corresponds to the sides of Pascal's triangle with a surplus of two terms. This selection is related to both the serendipity type elements and Coons' bivariate interpolation, which is well known in CAD theory of surface representation[19].

4. NUMERICAL EXAMPLES

The first example refers to in-plane vibrations of a square (a=6.0 m). The material constants are: $E/\rho=10^4$, ν=0·2. The problem was solved using different boundary discretizations of quadratic type (ELQUABE)[13] as shown in Figure 1. Comparison is performed with a fine FEM mesh using 24 subdivisions per side that is totally 625 nodes and 576 quadrilateral finite elements ("exact" solution denoted by the dotted line).

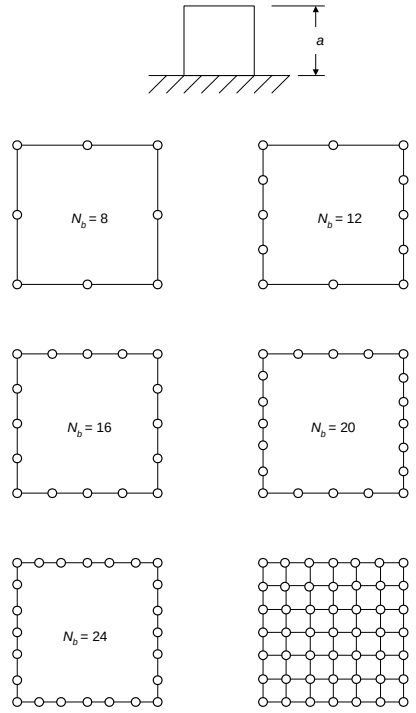
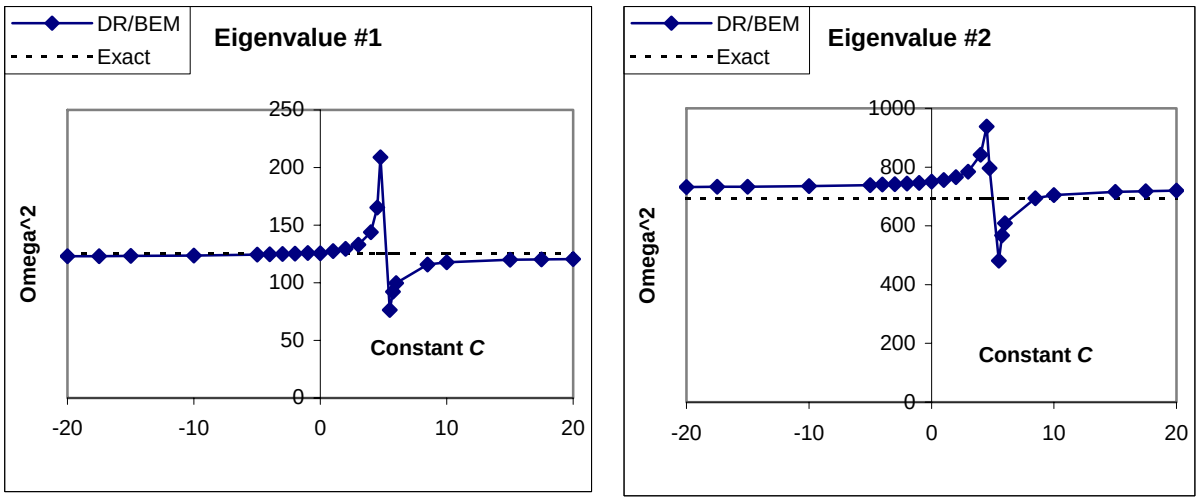


Figure 1: A clamped square of dimensions 6 m×6 m using several BEM and FEM discretizations

With respect to the conical RBF, a BEM model of N_b=24 boundary nodes (12 elements) was tested. It can be noticed in Figure 2 that the quality of the first two calculated eigenvalues highly depends on the chosen constant C. It is remarkable that for C≅5, the DR/BEM solution leads to unacceptable large values and in this case it is noted that intensively complex eigenvalues appear. Not a clear "overall optimum" value of C seems to exist. The most critical values are shown in Table 1. The diagonal of the square is about d=8.49m.



(a) (b)

Figure 2: Influence of the constant "C" [Eq(22)] in the accuracy of the (a) first and (b) second calculated eigenvalues in the square shown in Fig.1, for N_b=24 boundary nodes (Fig.1, bottom left). Results are compared with the "exact" solution obtained from a fine FEM model.

Table 1: Influence of the constant "C" on the quality of the first two calculated eigenvalues

Constant " C " Eq(22)	Active basis function (f)	Errors (in %) of calculated eigenvalues	
		1st eigenvalue	2nd eigenvalue
0.00	r	+0.06	+8.56
-1.00	1+r	+0.28	+8.01
8.49	8.49-r	-7.68	+0.41

With respect to the first eigenvalue, the most accurate value was obtained for C=0 while for negative values of C the first eigenvalue is rather underestimated. Obviously, for C=-1 the conical basis function becomes f=-1-r, but essentially, it is f=1+r; in this case the relative errors were small but not the best ones. The choice f=d-r=8.49-r[3], corresponding to the third symbol ♦ after the lowest singular point shown in Figure 2, was the best one for the second calculated eigenvalue but not for the first one. To have a better idea about the DR/BEM solution quality, it is noted that for the same boundary discretization using 4-node finite elements (Figure 1: bottom right), the corresponding relative error for the first and the second eigenvalues were found to be +1.73% and +0.48%, respectively. Therefore, there are C-values that partially achieve a better DR/BEM solution than FEM.

The singularity reflected to the first two eigenvalues will be now investigated in terms of the equivalent mass matrix described by Eq(21). As shown in Figure 3, the total equivalent mass was calculated different than the exact value that is twice the area of the square times the mass density, which is 72·0 kg. The maximum deviation from the real value was again found for C≅5. At C=-1, 0 and 8.49, in terms of the total mass the error was found equal to –4.5%, -5.4% and 8.6%, respectively.

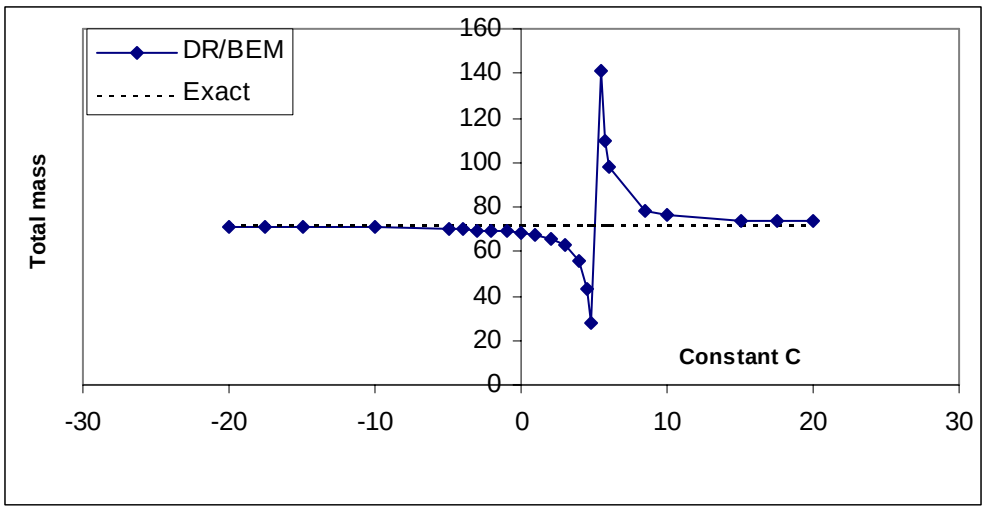


Figure 3: Dependence of equivalent total mass on the constant "C" [Eq(22)].

In the sequence, we investigate the performance of other RBFs. In Figure 4, a comparison is performed with respect to the first eigenvalue, for all the above-mentioned DR/BEM schemes as well as with conventional four-node bilinear FEM (49 nodes, as shown in Figure 1: bottom right). In order to have a common reference, wherever a compact support appears (conical, Gaussian, Wendland, MQ) it was chosen equal to the diagonal of the square. It is also clarified that in this comparison, the TPS has been implemented only in its simplest formulation described by Eq(28); in other words, without adding the {1,x,y} terms.

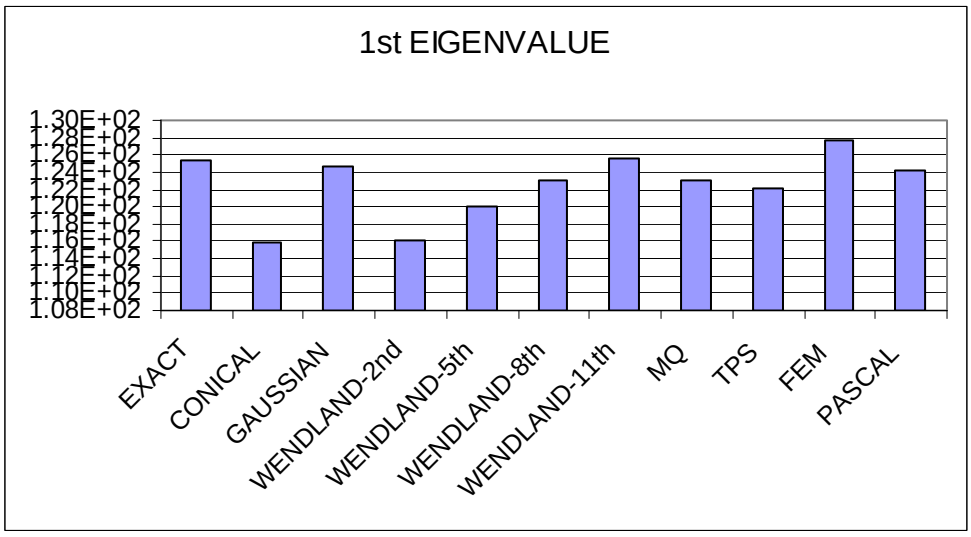


Figure 4: Comparison of the accuracy achieved for the calculation of the first eigenvalue.

The general impression is that conical (1st degree) and Wendland (2nd degree) are very close each other and significantly underestimate the first eigenvalue, while the accuracy increases in accordance to the degree of Wendland's RBF. The Gaussian RBF is rather good and of the same quality are the proposed (Pascal) polynomials.

A similar comparison with respect to the second eigenvalue is presented in Figure 5. It can be noticed that the Gaussian RBF underestimates, Wendland's RBF of 11th degree significantly overestimates this eigenvalue while the proposed (Pascal) polynomial scheme is again adequately accurate.

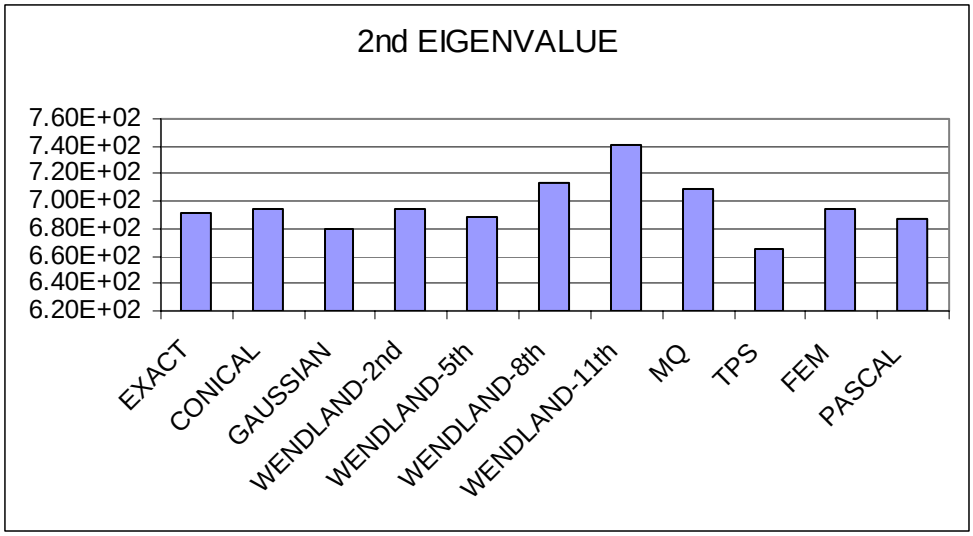


Figure 5: Comparison of the accuracy achieved for the calculation of the second eigenvalue.

A similar comparison with respect to the third eigenvalue is presented in Figure 6. It can be noticed that the Gaussian RBF is good, Wendland's RBF of 11th has a certain deviation while the proposed (Pascal) polynomial scheme is again adequately accurate.

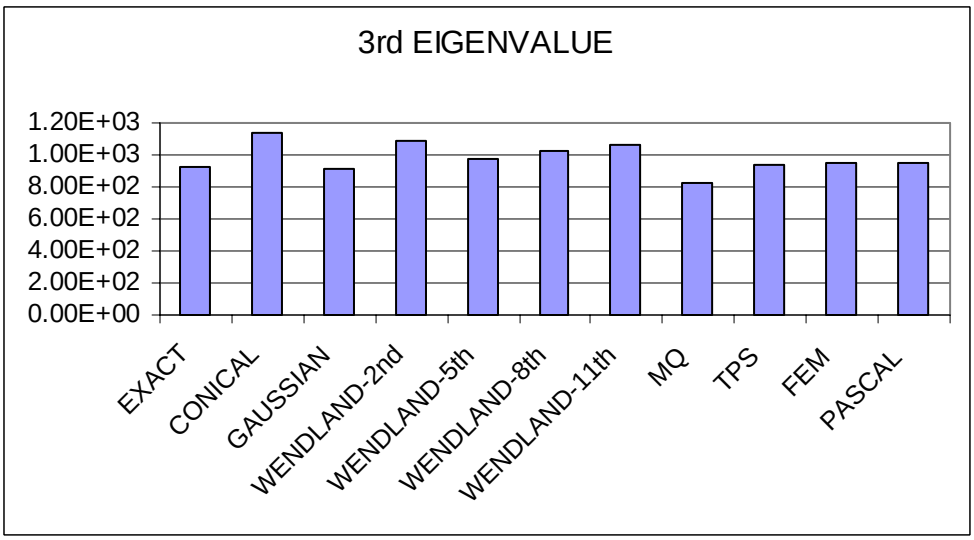


Figure 6: Comparison of the accuracy achieved for the calculation of the third eigenvalue.

Finally, a similar comparison with respect to the fourth eigenvalue is presented in Figure 7. It can be noticed that in this case all methods appear a certain error. The conical and the second degree RBF are characterized by the maximum error, the Gaussian is rather good, Wendland's RBF of 11th degree is not again a very good choice while the proposed (Pascal) polynomial scheme is very similar to the FEM solution (again, for the same number of 24 boundary nodes).

To come to an end, the last matter we investigated is the convergence behaviour of the several RBFs. For the discretizations shown in Figure 1, the convergence of the first eigenvalue is shown in Figure 8, where the proposed (Pascal-based) polynomial set shows an excellent convergence.

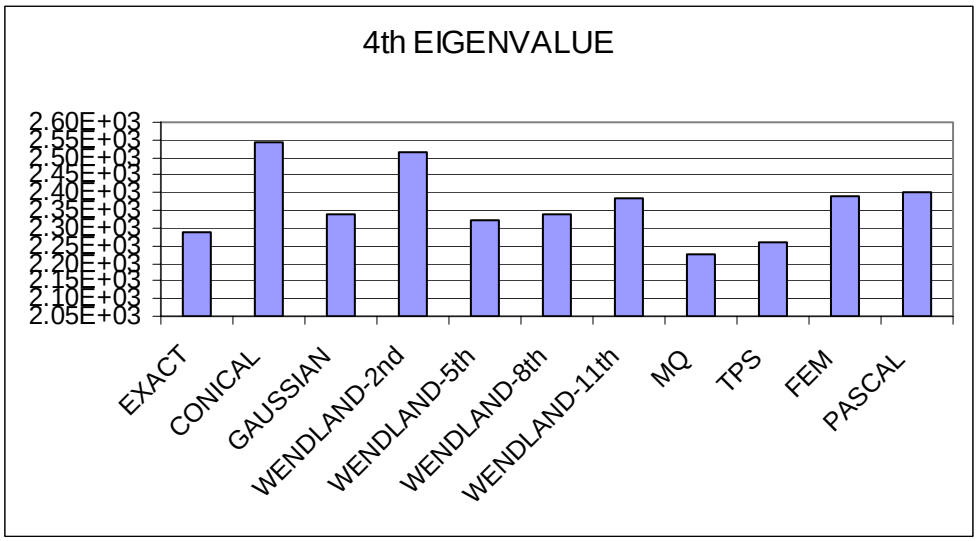


Figure 7: Comparison of the accuracy achieved for the calculation of the fourth eigenvalue.

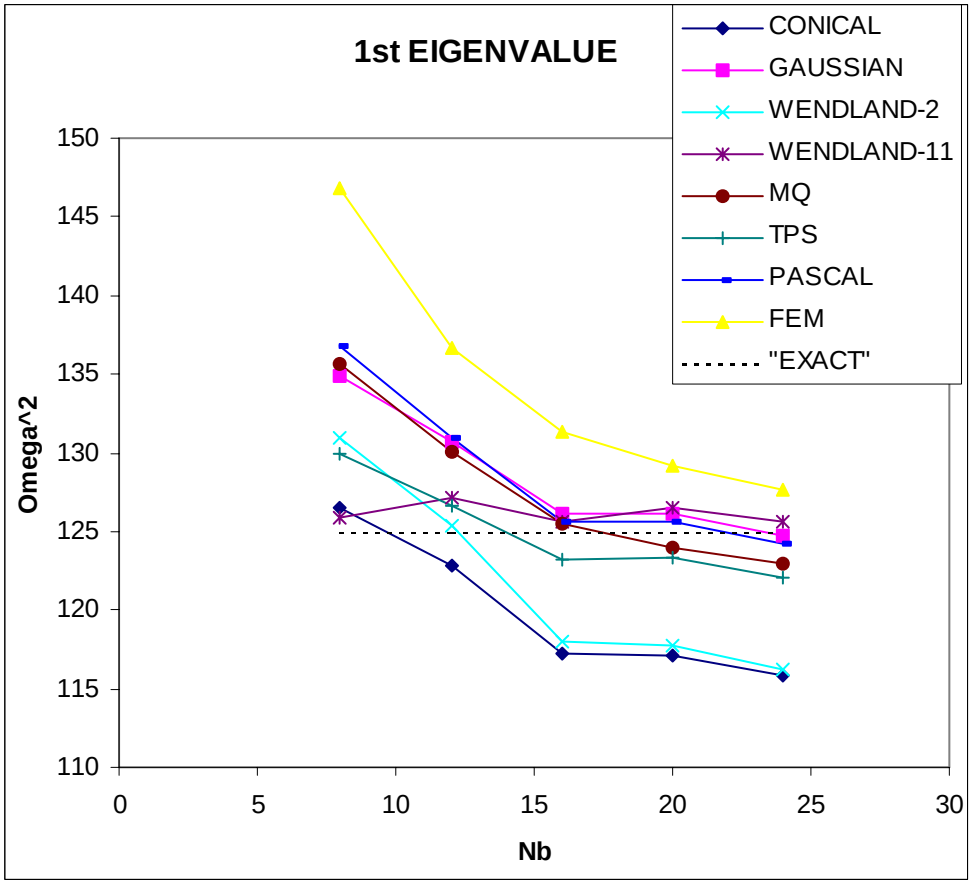


Figure 8: Convergence quality for the first eigenvalue using several RBFs.

5. DISCUSSION & CONCLUSIONS

While the conical RBF is influenced by the choice of the constant "*C*", the same does not seem to occur for the Gaussian, while slight variations were noticed in MultiQuadratics. As high-degree RBFs were tested only with the radius of support equal to the diagonal of the square, the influence of their support cannot be commented.

From the above analysis it becomes obvious that none of the RBFs can be considered to be the overall best choice. The conical RBF has the weakness that it can become dramatically instable when the constant *C* is chosen about 60% of the diagonal and it also overestimates the higher eigenvalues. Concerning Wendland's

RBF, unfortunately, the accuracy does not monotonically increase with the polynomial degree. Although it is not possible to propose guidelines from a single example of a square structure, it seems that the Gaussian RBF works well. Finally, the overall behaviour of the proposed (Pascal) polynomial functional set is encouraging.

The idea behind the proposed Pascal monomials is closely related to Coons' (blending-function) interpolation, which was previously applied for the mathematical representation of CAD surfaces in the three-dimensional space. In the position of 3-D *surfaces* we can think of *eigenmodes*, which should be accurately approximated by any of the existing or proposed RBFs. It has been also proven that the monomials participating in Coons' interpolation are the same with those involved in high-order serendipity elements[20]. In case of existing unrestrained nodes along two vertical sides of a rectangular structure, there is sufficient information to approximate a mode in terms of only boundary data. However, in case of a fully fixed boundary, the use of internal nodes is obligatory (for all RBFs), but the proposed polynomial scheme is again applicable on the basis of Gordon-Coons (transfinite) interpolation[21].

REFERENCES

[1] Nardini, D., and Brebbia, C.A. (1982), "A new approach to free vibration analysis using boundary elements," In: C.A. Brebbia (Ed.), Boundary Element Methods in Engineering, Berlin, Springer-Verlag, pp. 313-336.

[2] Partridge, P.W; Brebbia, C.A.; Wrobel, L. C. (1992), *The Dual Reciprocity Boundary Element Method*, Computational Mechanics Publications, Southampton.

[3] Wilson, R.B., Miller, N.M. and Banerjee, P.M. (1990), "Free-vibration analysis of three-dimensional solids by BEM," Int J Numer Meth Engng 29, pp. 1737-1757.

[4] Kamiya, N. and Sawaki, Y. (1988), "The plate bending analysis by the Dual Reciprocity Boundary Elements," Engineering Analysis 5(1), pp. 36-40.

[5] Wrobel, L.C., Brebbia, C.A. and Nardini, D. (1986), "The Dual Reciprocity Boundary Element Formulation for transient heat conduction," In: *Finite Elements in Water Resources VI*, Computational Mechanics Publications, Southampton and Springer-Verlag, Berlin and New York.

[6] Provatidis, C.G. (2004), "On DR/BEM for eigenvalue analysis of 2-D acoustics," Computational Mechanics 35, pp. 41-53.

[7] Kanarachos, A. and Provatidis, Ch. (1992), "Potential and wave propagation problems using the boundary element method and BEM-subregions," Engineering Analysis with Boundary Elements 9, pp. 117-124.

[8] Provatidis, C.G. (2003), "Free vibrations of two-dimensional structures using Coons-patch macroelements, FEM and BEM," *CD Proceedings of the Int Conf Comput & Exp Engng & Sciences (ICCES 03)*, Corfu, Greece, 24-29 July.

[9] Agnantiaris, J.P., Polyzos, D. and Beskos, D.E. (1996), "Some studies on dual reciprocity BEM for elastodynamic analysis," Computational Mechanics 17, pp. 270-277.

[10] Agnantiaris, J.P., Polyzos, D. and Beskos, D.E. (2001), "Free vibration analysis of non-axisymmetric and axisymmetric structures by the dual reciprocity BEM," Engineering Analysis with Boundary Elements 25, pp. 713-723.

[11] Rashed, Y.F. (2002), "Transient dynamic boundary element analysis using Gaussian-based mass matrix," Engineering Analysis with Boundary Elements 26, pp. 265-279.

[12] Rashed, Y.F. (2002), "BEM for dynamic analysis using compact supported radial basis functions," Computers and Structures 80, pp. 1351-1367.

[13] Brebbia, C.A. and Dominguez, J. (1992), *Boundary Elements: An Introductory Course,* 2nd ed., Computational Mechanics and Springer-Verlag, Berlin.

[14] Private Communication (2003), included in a letter from Prof. D.R.J. Owen (9th December 2003).

[15] Yamada, T. and Wrobel, L.C. (1993), "Properties of Gaussian radial basis functions in the dual reciprocity boundary element method," ZAMP 44, pp. 1054-67.

[16] Wendland, H. (1995), "Piecewise polynomial, positive definite and compactly supported radial basis functions of minimal degree," Advances in Computational Mathematics 4, pp. 389-396.

[17] Goldberg, M.A. and Chen, C.S. (1994), "The theory of radial basis functions applied to the BEM for inhomogeneous partial differential equations," Boundary Element Communications 5, pp. 57-61.

[18] Goldberg, M.A. and Chen, C.S. (1996), "A bibliography on radial basis function approximation," Boundary Element Communications 7, pp. 155-163.

[19] Provatidis, C.G. (2003), "Analysis of axisymmetric structures using Coons' interpolation," Finite Elements in Analysis and Design 39, pp. 535-558.

[20] Provatidis, C.G. (2005), " Performance of large Gordon-Coons finite elements in 2-D potential problems," Proceedings GRACM05, Limassol, Cyprus, 29 June – 1 July, 2005.

[21] Provatidis, C.G. (2004), "Solution of two-dimensional Poisson problems in quadrilateral domains using transfinite Coons interpolation," Communications in Numerical Methods in Engineering 20, pp. 521-533.

THE HIGH ACCURATE COMPOSITE GRIDS METHOD FOR SOLVING LAPLACE'S BOUNDARY VALUE PROBLEMS WITH SINGULARITIES ON L-SHAPED POLYGON AND MOTZ PROBLEM

Mehmet Bozer

Department of Mathematics
Eastern Mediterranean University
Famagusta, Cyprus
e-mail: mehmet.bozer@emu.edu.tr

Keywords: Composite Grids, Matching Operator, Singularities.

Abstract. *High accurate composite grids method on graduated polygons is constructed. $O(h^6)$ order of accuracy, for the number of nodes $O(h^{-2} \ln h^{-1})$ is obtained by using a 9-point scheme on the polar and on the square grids, as well as constructing the sixth order matching operator connecting the subsystems. This estimate being obtained for requirements on the functions given in the boundary conditions which cannot be essentially lowered in $C_{k,\lambda}$. Finally, we illustrate the high accuracy of the method in solving the problem in L-shaped polygon with a corner singularity, and the well known Motz problem which has a boundary singularity due to an abrupt change in the type of boundary conditions.*

1 INTRODUCTION

It is known that the use of classical finite difference or finite element methods to solve the elliptic boundary value problems with singularities becomes ineffective. To solve this problem, different modifications are given by many authors[1],[2],[3],[4],[5],[6] . One of the effective modifications of a finite-difference method for a nonsmooth solutions of elliptic problems is the method of composite grids. This modification with an accuracy $O(h^2)$ in solving the Laplace's equation was proposed and justified in[7],[8], and developed up to $O(h^4)$[9].

In this paper the sixth order accurate composite grids method on graduated polygons is constructed. High, order $O(h^6)$, accuracy is obtained by using 9-point scheme on exponentially compressed polar grids (on sectors, around the vertices of re-entrant angles), and on square grids (on the finite number of rectangles, which cover the given polygonal domain together with the chosen sectors), as well as by using the sixth order matching operator constructed in[10] to connect the subsystems. Moreover, the requirements on the functions given in the boundary conditions cannot be essentially lowered in $C_{k,\lambda}$. The system of difference equations obtained separates into a fixed number of subsystems each of which is adequate for the simplest difference equations on a rectangle, and may be solved by the alternating method of Schwarz.

2 BOUNDARY VALUE PROBLEM AND FINITE DIFFERENCE EQUATIONS

Let G be an open $\gamma = \bigcup_{j=1}^{N} \gamma_j$ simply connected polygon with sides parallel to x and y axis, let γ_j, j=1(1)N, be its sides, including the ends, enumerated counterclockwise ($\gamma_0 \equiv \gamma_N$, $\gamma_1 \equiv \gamma_{N+1}$; $\gamma = \bigcup_{j=1}^{N} \gamma_j$ is the boundary of G, $\alpha_j \pi$, $0<\alpha_j \le 2$, be the interior angle formed by its sides γ_{j-1} and γ_j, $A_j = \gamma_{j-1} \cap \gamma_j$ be the vertex of the j-th angle; r_j, θ_j be a polar system of coordinates with pole in A_j and the angle θ_j taken counterclockwise from the side γ_j, ν_j be a parameter taking the values 0 or 1, and $\overline{\nu}_j = 1 - \nu_j$.

We consider the boundary value problem

$$\Delta u = 0 \text{ on } G, \quad \nu_j u + \overline{\nu}_j u_n^{(1)} = \nu_j \varphi_j + \overline{\nu}_j \psi_j \text{ on } \gamma_j, \; j = 1(1)N, \tag{1}$$

where $\Delta \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$, $u_n^{(1)}$ is the derivative along the inner normal, φ_j and ψ_j are given functions of the arc length s taken along γ, and

$$\nu_j \varphi_j + \overline{\nu}_j \psi_j \in C_{6,\lambda}(\gamma_j), \; 0<\lambda<1, \; 1 \le j \le N, \tag{2}$$

$$1 \le \nu_1 + \nu_2 + \ \ldots \ + \nu_N \le N, \tag{3}$$

at the vertices A_j, ($s=s_j$) for $\alpha_j=1/2$ the conjugation conditions

$$\nu_j \varphi_j^{(2q+\delta_{\tau-2})} + \overline{\nu}_j \psi_j^{(2q+\delta_\tau)} = (-1)^{q+\delta_\tau+\delta_{\tau-1}} (\nu_{j-1} \varphi_{j-1}^{(2q+\delta_{\tau-1})} + \overline{\nu}_{j-1} \psi_{j-1}^{(2q+\delta_\tau)}), \tag{4}$$

except may be for q=3, for $\tau=3$, where $\tau=\nu_{j-1}+2\nu_j$; $\delta_\omega=1$ when $\omega=0$; $\delta_\omega=0$, when $\omega\neq 0$, q=0,1,...Q, $Q=[(6-\delta_{\tau-1}-\delta_{\tau-2})/2]-\delta_\tau$, are satisfied. At the vertices A_j with $\alpha_j\neq 1/2$ no compatibility conditions are required to hold for the boundary functions, in particular, the values of φ_{j-1} and φ_j at A_j might be different.

Let $\widehat{N}$ be the set of j, ($1\le j\le N$), for which $\alpha_j\neq 1/2$ or $\alpha_j=1/2$ but conjugation condition (4) not satisfied. In the neighborhood for each vertex A_j, $j\in E$ we construct two fixed sectors $T_j=T_j(r_{j0})$ and $T_j^*=T_j(r_j^*),\ 0<r_j^*<r_{j0}$ for which $T_l^* \bigcap T_k^* = \varnothing$, $l\neq k$, $l,k\in\widehat{N}$ and also $G^* \bigcap T_j = T_j \setminus T_j^*$, where $G^* = G \setminus (\bigcup_{j\in E} T_j^*)$.

Let $\Pi_k \subset G^*$, k=1(1)M, ($M<\infty$) be certain fixed open rectangles with sides a_{1k} and a_{2k} parallel to x and y axis, with a_{1k}/a_{2k} rational and $G=(\cup_{k=1}^M \Pi_k)\cup(\cup_{j\in\widehat{N}} T_j^*)$. We will assume that any point $P\in G\cap\eta_k$, where η_k is the boundary of the rectangle Π_k, $1\le k\le M$, ($P\in G\cap\upsilon_j$, υ_j is the boundary of the sector $T_j\in\widehat{N}$) falls inside at least one of the figures $\Pi_{k(p)}$, $1\le k(p)\le M$, or $T_{j(p)}$, $j(p)\in\widehat{N}$ depending on P, where at the distance from P to $G\cap\eta_{k(P)}$ or to $G\cap\upsilon_{j(P)}$ is not less than some constant $\nu_0>0$ independent of P. The quantity ν_0 is called a depth of gluing of the figures Π_k, $1\le k\le M$; T_j, $j\in\widehat{N}$.

We introduce the parameter $h>0$ and define a square grid on Π_k, $1\le k\le M$, with maximum possible step $h_k\le\min\{h,\min\{a_{1k},a_{2k},\nu_0\}/6\}$, and such that the boundary η_k lies entirely on the grid lines. Let Π_k^h be the set of grids on Π_k, η_{k0}^h be the set of grids on $G\cap\eta_k$, η_{k1}^h be the set of remainder grids on η_k, and let $\overline{\Pi}_k^h=\Pi_k^h\cup\eta_{k0}^h\cup\eta_{k1}^h$. We construct on each closed sector T_j, $j\in\widehat{N}$, the polar grids of rays with maximum angular step $\beta_j\le\min\{\alpha_j\pi/6, h/r_{j0}\}$, where γ_{j-1} and γ_j are situated on the extreme rays, and of circumferences of circles with centers at A_j and radii $r_{jq}=r_{j0}\exp\{-q\,\beta_j\}$, q=0,1,...$N_j$,

$$N_j = 1 + \left[\max\left\{ 5, \frac{1}{\beta_j}\ln\frac{r_{j0}}{r_j^*}, \frac{(6+\varepsilon_0\sigma_j)\beta_j^{-1}\ln h^{-1}}{\min\left\{1, 1/\alpha_j\left(1+\left|\nu_{j-1}-\nu_j\right|\right)\right\}} - N_{j0} \right\} \right] \tag{5}$$

where $\varepsilon_0>0$, $N_{j0}\ge 0$ are arbitrary fixed numbers, [ ]- indicates the integer part, $\sigma_j=1$ when $(\nu_j-1\}+\nu_j(2\alpha_j-\nu_{j-1}-\nu_j)=0$; $\sigma_j=0$ when $(\nu_{j-1}+\nu_j)(2\alpha_j-\nu_{j-1}-\nu_j)\neq 0$.

From (5) it follows that for any h, $r_{jNj}<r_j^*$, $j\in\widehat{N}$. Let ν_{j0}^h (ν_{j2}^h) be the set of nodes of the j-th polar grids on the circle with radius r_{j0}(r_{jNj}), ν_{j1}^h be the set of remainder grids on ν_j, T_{jh} be the set of nodes of the j-th polar grids on $T_j\setminus\nu_{j2}^h$ and $\overline{T}_j^h=T_j^h\cup\nu_{j0}^h\cup\nu_{j1}^h\cup\nu_{j2}^h$.

Let B,B_m, and $\dot{B}$ are the finite difference operators, defined on Π_k^h, $\eta_{k1}^h\cap\gamma_m$, and $\eta_{k1}^h\cap(\gamma_m\cap\gamma_{m+1})$ by the formulas (5.5), (5.11), and (5.13) from[11], respectively. We define $\varepsilon_{qh}(\varphi_q,\psi_q)\equiv E_{qh}(0,\varphi_q,\psi_q)$, and $\dot{\varepsilon}_{qh}(\varphi_q,\varphi_{q+1},\psi_q,\psi_{q+1})\equiv\dot{E}_{qh}(0,\varphi_q,\varphi_{q+1},\psi_q,\psi_{q+1})$ by applying the formulas (5.12) and (5.14) in[11] respectively. We transform each polar grid T_j^h to a square grid with step β_j by applying new variables $x_j'=\ln r_j$, $y_j'=\theta_j$, $j\in\widehat{N}$.

So, all operators B, B_m, $\dot{B}_m$, ε_{mh} and $\dot{\varepsilon}_{mh}$ will be defined on the obtained square grids for the function $v(x_j',y_j')\equiv u(e^{xj'},y_j')$, and consequently for the function $u(r_j,\theta_j)$ on the polar grids T_j^h, $j\in\widehat{N}$.

Let $\varphi=\{\varphi_j\}_{j=1}^N$ and $\psi=\{\psi_j\}_{j=1}^N$, where φ_j, and ψ_j are given functions in (1). On the set $\omega^h=\left(\cup_{k=1}^M\eta_{k0}^h\right)\cup\left(\cup_{j\in E}\nu_{j0}^h\right)$ we introduce the linear matching operator S^6. The value of $S^6(u_h,\varphi,\psi)$ at the point $P\in\omega^h$ is defined linearly in terms of the values of the function u_h at nodes of the grid constructed on the

rectangle $\Pi_{k(P)} \ni P$ or on the sector $T_{\nu(P)} \ni P$ part f the boundary $\eta_{k(P)}$ or $\upsilon_{\nu(P)}$ which belongs to G of which is the maximum distance away from P, and in terms of the assigned values of φ or ψ at a fixed number of points or in terms of the values of the derivatives $\varphi^{(m)}$, m=0,1,...,5 (or $\psi^{(q)}$, q=0,1,...,4). The pattern of the operator S^6 is located in a neighborhood O(h) of the point P, its norm in the uniform metric for $\varphi \equiv 0$, $\psi \equiv 0$ is not greater than one, and

$$u-S^6(u,\varphi,\psi)=O(h^6)$$

uniformly on ω^h. An operator S^6 with these properties for a square grid $\overline{\Pi}^h_{k(P)}$ was constructed in[10] (for the Dirichlet problem see[7]). If $P \in \overline{T}_{\nu(P)} \cap \omega^h$ then we use the variables $x'_{\nu(P)}=\ln r_{\nu(P)}$, $y'_{\nu(P)}=\theta_{\nu(P)}$ to transform $\overline{T}^h_{\nu(P)}$ to a square grid with step $\beta_{\nu(P)}$ and the point P to some point P′. We choose all necessary nodes with respect to P′, and calculate corresponding coefficients to design S^6 on the obtained square grid. Since the transformation is invertible and $\beta_j \le h/r_{j0}$, $j \in \widehat{N}$, it follows that the operator S^6 is constructed on $\overline{T}^h_{\nu(P)}$, and it is a sixth order matching operator in a polar grids also.

Consider the following system $O(h^2 \ln h^{-1})$ of finite difference equations which has a unique solution for all $h \le \nu_0$, where ν_0 is the depth of gluing,

$$u_h = Bu_h \text{ on } \Pi^h_k,\ u_h = \overline{\nu}_m B_m u_h + \varepsilon_{mh_k}(\varphi_m,\psi_m) \text{ on } \eta^h_{k1} \cap \gamma_m \setminus \left(A_m \cup A_{m+1}\right), \tag{6}$$

$$u_h = \overline{\nu}_m \overline{\nu}_{m+1} B u_h + \varepsilon_{mh_k}\left(\varphi_m, \varphi_{m+1}\psi_m, \psi_{m+1}\right) \text{ on } \eta^h_{k1} \cap A_{m+1},\ k=1,2,...,M,\ 0 \le m \le N-1 \tag{7}$$

$$u_h = S^6\left(u_h,\varphi,\psi\right) \text{ on } \omega^h,\ u_h = Bu_h \text{ on } T^h_j, \tag{8}$$

$$u_h = \overline{\nu}_\tau B_\tau u_h + \varepsilon_{\tau\beta_j}(\varphi_\tau, \rho_\tau\psi_\tau) \text{ on } \nu^h_{j1} \cap \gamma_\tau, \tau = j-1, j, \tag{9}$$

$$\begin{aligned} u_h &= \nu_{j-1}\nu_j\left(\theta_j\varphi^*_{j-1} + \left(\alpha_j\pi - \theta_j\right)\varphi^*_j\right)/\left(\alpha_j\pi\right) + \nu_{j-1}\overline{\nu}_j\varphi^*_{j-1}\overline{\nu}_{j-1}\nu_j\varphi^*_j + \overline{\nu}_{j-1}\overline{\nu}_j B_j u_h, \\ &\text{on } \nu^h_{j2}, j \in \widehat{N}, \end{aligned} \tag{10}$$

where φ^*_μ being the value φ_μ at the point on γ_μ distant from A_j by the distance r_{jNj}, and ρ_τ being the distance from A_j to the present grid on γ_τ.

3 ERROR ANALYSIS

Let v be the solution of problem (1)-(4) when $G \equiv \Pi = \{(x,y): 0<x<a,\ 0<y<b\}$, where a/b being rational, and v_h is the solution of corresponding finite difference problem, i.e., of system (6),(7).

Everywhere below we will denote constants which are independent of h by $c, c_0, c_1, \ldots$.

Lemma1. The estimation $\max_{\overline{\Pi}^h} \left|v_h - v\right| \le c_0 h^6$ holds true.

Lemma 1 is proved in the same way as Theorem 1.1 of[8], except that instead of the function (1.14) of[8] we need to take the function

$$u_j(x,y) = (-1)^j \frac{2\left[\varphi^{(6)}_j(s_{j+1}) + \varphi^{(6)}_{j+1}(s_{j+1})\right]}{\pi 6!} \operatorname{Im}\left\{(z-z_j)^6 \ln(z-z_j)\right\},$$

(making the corresponding changes), where $z=x+iy$, $z_j=x_j+iy_j$, is the complex coordinate of the vertex $\gamma_j \cap \gamma_{j+1}$.

Remark1. The obtained estimation in Lemma 1 is true for the corresponding problems in sectorial domains, i.e., the 9-point analogy of Theorem 2.1 in[8].

Theorem1. Let the conditions (2)-(4) satisfied. Then for $h \le \nu_0$, where ν_0 is the depth of gluing, the inequality

$$\max_{\overline{G}^h} \left|u_h - u\right| \le ch^6 \tag{11}$$

holds true, where u is the bounded solution of problem (1), u_h is the solution of system (6)-(10),

$\overline{G}^h = (\cup_{k=1}^m \overline{\Pi}_k^h) \cup (\cup_{j\in\widehat{N}} \overline{T}_j^h)$.

Remark2. When $\nu_j\varphi_j + \overline{\nu}_j\psi_j \in C_{6,0}(\gamma_j)$, then the function v(x,y) in Lemma1 may not belong to $C_{6,0}(\Pi)$ (see[11],[13]). Consequently, by analogy[14], it can be shown that the requirement (2) in Theorem1 can not be lowered in $C_{k,\lambda}$.

Remark3. A parallelism of the sides of graduated polygon G to the x and y axis is assumed only for simplicity of presentation. The method and results can be extended also to arbitrary polygonal domains, by special arrangement some of rectangles Π_k, k=1(1)M, around the sides γ_j, j=1(1)N.

System (6)-(10) is solved by the alternating method of Schwarz. Suppose we have a zero approximation to u_h. For each k in succession, 1≤k≤M, having calculated u_h on η_{k0}^h by formula (8), we solve system (6) and (7). Further, in a similar way, having calculated u_h on ν_{j0}^h by (8), we solve (9) and (10) successively for every $j\in\widehat{N}$. The next iteration is similar.

Theorem2. System (6)-(10) may be solved by the alternating method of Schwarz with arbitrary precision ε in the uniform metric with the number of iterations $O(\ln\varepsilon^{-1})$, independent of h, by solving the standard difference equations of Laplace on rectangular domains at each iteration.

4 NUMERICAL EXAMPLES

To test the high accuracy of the composite grids method, we have computed numerical examples. In Example1, the polygon G is L-shaped, and the exact solution has the corner singularity at the vertex A_1 with the interior angle $\alpha_\pi=3\pi/2$. The comparisons are made between the exact and the approximate solutions. In Example2, the exact solution has singularity at the vertex A_1, with the interior angle $\alpha_\pi=\pi$, caused by abrupt changes in the type of boundary conditions, and because of boundary conditions on the sides for y≥0 the exact solution is unknown and comparisons are made with the best results obtained in the literature.

Example 1. Let G is the L-shaped, and is defined as follows

$$G = \{(x,y) : -1 < x < 1, \quad -1 < y < 1\} \setminus G_1$$

where $G_1=\{(x,y):0\le x\le 1,-1\le y\le 0\}$. Let γ be the boundary of G. We consider the following problem:

$$\Delta u = 0 \text{ on } G,$$
$$u = v(r,\theta) \text{ on } \gamma,$$

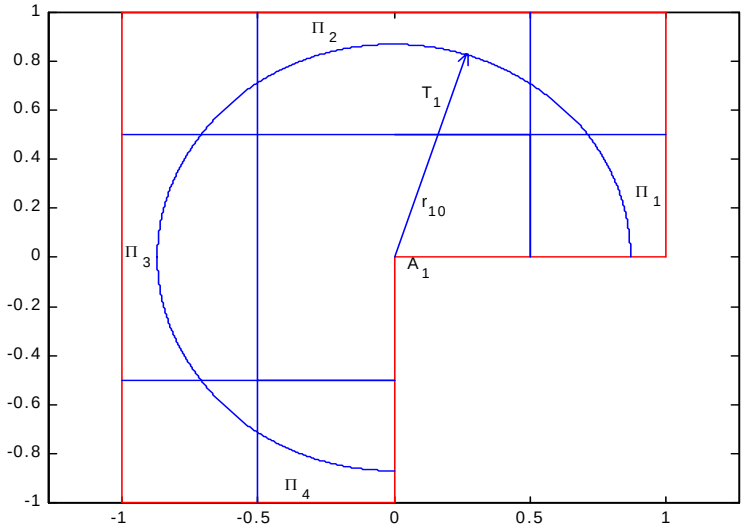


Figure 1. The L-shaped domain with angle singularity

where

$$v(r,\theta) = r^{\frac{2}{3}} \cos\frac{2\theta}{3}$$

is the exact solution of this problem.

$\mathbf{h^{-1}}$	$\varepsilon_h = \max_{\overline{G}^h} \lvert u_h - v \rvert$
16	**1.12D-9**
32	**1.86D-11**

Table1: The maximum error between the composite grids solution u_h and the exact solution u of the problem in Example 1 for different h

Example 2 Let G ={(x,y):-1<x<1,0<y<1}, and γ is its boundary. We consider the following problem:

$$-\Delta u = 0 \text{ in } G,$$
$$u = 0 \text{ on } y = 0, \ -1 \le x \le 0,$$
$$u = 500 \text{ on } x = 1, \ 0 \le y \le 1,$$
$$\frac{\partial u}{\partial n} = 0 \text{ on the other boundary segments of } \gamma.$$

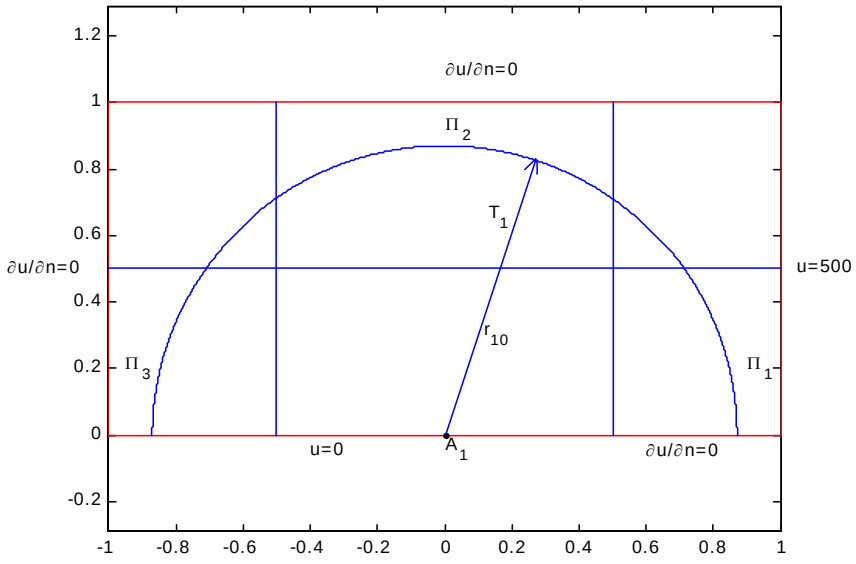


Figure 2. The Motz problem

The Motz problem as a benchmark in many approaches for the singular problems is used (see[1],[2],[5] and references therein). An extremely accurate result is obtained in[15], for which piecewise expansions into particular solutions are used to approximate the boundary conditions in a least square sense. To obtain a high accurate results (for instance, the maximum error on x=1 is 5.47E-9) by this boundary method a large number (34) of particular solutions are needed and this may reduce serious difficulties due to illconditioning of the associated least-squares matrices (the condition number is 3.97E+07). To decrease the condition number (down to 3617) in[15] the given domain is subdivided into three subdomains, and for each of them different numbers of particular solutions are used, but the best combination of these numbers gives only E-06 accuracy. Lucas and Oh[16], for the solution of the Motz problem, the method of auxiliary mapping (MAM) in the context of the h-p version of the finite element method is used, and the best result (maximum error is 2.22E-08) is obtained, with an optimal mesh refinement and p=10. Comparisons in[16] were made with the extremely accurate results obtained in[15].

The following results in Table 2 are obtained for the Motz problem by comparing composite grids solution u_h to the extremely accurate solution v from[15] after correction of the 31-th coefficient (dividing by 10), in the expansion of approximate solution, discovered by Lucas and Oh[16].

$\mathbf{h^{-1}}$	$\varepsilon_h = \max_{\overline{G}^h} \lvert u_h - v \rvert$
16	**7.97D-7**
32	**2.21D-8**

Table 2: The maximum error between composite grids solution u_h
and extremely accurate solution v[10] of the Motz problem for different h

5 CONCLUDING REMARKS

The comparisons in Tables 1 and 2 show that the accuracy of the composite grids method in solving the problem in Example1 is higher than the problem in Example2. This happens because, in the Motz problem (the exact solution unknown) the comparisons are made with the results obtained in[15] (see also[16] , p.337).

The high order composite grids method for the solution of Laplace's boundary value problem on graduated polygons is constructed and justified. $O(h^6)$ order of accuracy for the absolute error in maximum norm is obtained by constructing the sixth order matching operator connecting the subsystems of 9-point approximations on the polar and square grids. Numerical examples show the high accuracy of the proposed modification, and the possibility of effective realization of the obtained system of finite difference equations by using the discrete Fourier solution on rectangle with square grids for each Schwarz's iteration.

The parallelism of the sides of graduated polygon G to x and y axis is assumed only for simplicity of presentation.

The method and the results carry over to multiply connected graduated polygons without change.

REFERENCES

[1] Li Z.C., (1998), "*Combined Methods for Elliptic Equations with Singularities, Interfaces and Infinities*", Kluwer Academic Publishers, Dordrech, Boston and London.

[2] L. Olson, G. Georgiu, W. Schultz (1991) " *An efficient finite element method for treating singularities in Laplace's Equation*", Jour. of Compu. Phys. 96, No.2.

[3] M. Elliotis, G. Georgiou, C. Xenophontos (2002) "*The solution of Laplacian problems over L-shaped domains with a singular function boundary integral method*", Commun. Numer. Meth. Engng.18 pp.213-222.

[4] G. Georgiou, A. Boudouvis, A.Poullikas, (1997), "*Comparison of the two methods for the computation of singular solutions in elliptic problems*", Jour. of Compu. and Appl. Math. 79 pp.277-287.

[5] G. Georgiu, L. Olson, Y. Smyrilis, (1996), "*A singular function boundary integral method for the Laplace equation*" Commun. in Numer. Methds. in Engineer.,12 pp. 127-134.

[6] N. Papamichael, M.S. Soares, (1991), "*An $O(h^6)$ cubic spline interpolating procedure for harmonic functions*", Num. Methds. for Part. Diff. Eqns. 1, pp.43-59.

[7] E.A.Volkov, (1968), "*Method of composite meshes for finite and infinite domain with a piecewise-smooth boundary*",Tr. Mat. Inst. Akad. Nauk SSSR 96, 117-148 (1968); English transl. in Proc. Steklov Inst. Math. 96.

[8] Volkov E.A., (1976), "*On the method of regular composite meshes for Laplace's equation on polygons*", Tr. Mat. Inst. Akad. Nauk SSSR 140, pp. 68-102 (1976); English transl. in Proc. Steklov Inst.Math. 140 (1).

[9] Dosiyev A.A., (2002), "*A fourth order accurate composite grids method for solving Laplace's boundary value problems with singularities*", Zh. Vychisl. Mat. Mat. Fiz. 42, 6, pp.869-886.

[10] Dosiyev A.A. and Bozer M. "*A high order matching operator on square grids for a solution of Laplace's equation*", Submitted to Appl. Math. Letters.

[11] Volkov E.A., (1966). "*Effective error estimates for grid method solutions of boundary-value problems for Laplace's and Poisson's equations on a rectangle and certain triangles*", Tr. Mat. Inst. Akad. Nauk SSSR.74, pp.55-85.

[12] Dosiyev A.A., (1994). "*A block-grid method of increased accuracy for solving Dirichlet's problem for Laplace's equation on polygons*", Comp.Maths Math.Phys. 34(5), pp.591-604.

[13] Volkov E.A., (1965). "*Differentiability properties of solutions of boundary value problems for the Laplace and Poisson equations on a rectangle*", Proc. Steklov Inst.Math. 77, pp.101-126.

[14] Volkov E.A., (1969). "*The differential properties of the solutions of Laplace's equation, and the errors in the method of nets with boundary values in C_2 and $C_{1,1}$*", Comput. Math. and Math. Phys. 9(3), pp. 97-112.

[15] Li Z.C., Mathon R., and Sermer P., (1987) "*Boundary methods for solving elliptic problems with singularities and interfaces*", SIAM J. Numer. Anal. 24, 3, pp.487-498.

[16] Lucas T.R. and Oh H.S., "*The method of auxiliary mapping for the finite element solutions of elliptic problems containing singularities*", J. Comput. Phys. 108, (1993), pp.327-342.

ON SOLVING THE CRACKED BEAM PROBLEM BY A BLOCK METHOD

A. A. Dosiyev and S. Cival Buranay

Department of Mathematics,
Eastern Mediterranean University,
Famagusta, Cyprus.
e-mail: adiguzel.dosiyev@emu.edu.tr
e-mail: suzan.buranay@emu.edu.tr

Keywords: Singularity, cracked beam problem, block method .

Abstract. *A high accurate approximate solution is obtained to the mixed boundary-value problem for the Laplace equation on a rectangle, known as the cracked beam problem by the one-single block method with the use of analytic continuation. Furthermore, to compute the stress intensity factor a simple and high accurate formula is given. Comparisons with other results have been carried out.*

1 INTRODUCTION

The cracked beam problem is the one of which the solution has a singularity at a boundary caused by the abrupt changes in the type of boundary conditions For this reason the problem has become a classical test for approximate methods for solving boundary value problems with singularities. We single out such three important papers[4],[5],[12] . Using the asymptotic behavior of the solution Fix et al.[4], included some of the lower order singular functions of the asymptotic expansion in their space of trial functions for the finite element method. They found that their method was efficient and moderately accurate. The local solution

$$u(r,\theta)=\sum_{i=1}^{\infty}\alpha_i r^{(2i-1)/2}\cos\left[\left(\frac{2i-1}{2}\right)\theta\right] \tag{1}$$

is used by Wigley[12]. He obtained accurate estimates for the leading singular coefficients with an inherently iterative approach. He first generated an approximate numerical solution using finite differences and then estimated the first singular coefficient from this solution. Next, he modified the original problem by subtracting out the first singular term and again generated an approximate numerical solution to this modified problem, which he used to estimate the second singular coefficient, and so on. Olson et al.[5] used the ordinary finite element method, one-and two-zone blending singular basis functions method (BSBFM) and the integrated singular basis function method (ISBFM) which does not require high- order integration and the singular functions are defined over the entire domain. The solution of the cracked beam problem confirmed that both (BSBFM) and (ISBFM) yield improved results compared to the ordinary element method moreover both the estimate of the leading coefficients and the solution to the problem with the (ISBFM) are more accurate than (BSBFM).

An exponentially convergent block method (BM) to solve boundary-value problems for the Laplace equation on an arbitrary polygon was proposed and justified in[6],[7]. In the (BM), the approximate solution is an approximation of the integral representation of the harmonic function in a fixed number of blocks (sectors, semicircles and circles) covering the given domain. Some problems for the Lapace equation involving singularities were solved by the one-block version of BM in[8]-[9]. A block method without analytical continuation was applied in[10] , where the torsion problem was solved for L-profile, including the case of infinite flanges, with the use of four interacting blocks that lie within the L-section. The rate of convergence of the approximate solution of the BM is higher than the results obtained in above mentioned approaches.

In this paper the BM for the problem of torsion of a cracked beam with a square cross section (see [4]) is solved by the one-single block method with the use of analytic continuation. The accuracy of the resulting solution is close to the accuracy of computer representation of numbers, where the computations are carried in extended precision. The system of algebraic equations arising in the case of one block is stable as the number of

unknowns n increases and has a simple structure that solving it does not require complete storage of its entire matrix in a computer. The simplest version of iterative methods (Gauss-Seidel), convergent as a geometric progression with a common ratio independent of n can be applied to find its solution. Furthermore, to compute the stress intensity factor a simple and high accurate formula is given. The comparison with the other above mentioned results are presented.

In[1]-[3] the BM to design a combined method for the solution of more general boundary value problems with singularities is used.

2 CRACKED BEAM PROBLEM

Let

$$G=\{(x,y):-0.5<x<0.5,\ 0<y<0.5\}, \tag{2}$$

and γ be its boundary (Figure 1). We consider the following problem

$$\begin{aligned}
&\Delta u=0 \ \text{ in } G,\\
&u=0 \ \text{ on } y=0,\ \ -0.5\le x\le 0,\\
&u=0.125 \ \text{ on } y=0.5,\ \ -0.5\le x\le 0.5\\
&\frac{\partial u}{\partial n}=0 \ \text{ on the other boundary segments of } \gamma\setminus(1,0)
\end{aligned} \tag{3}$$

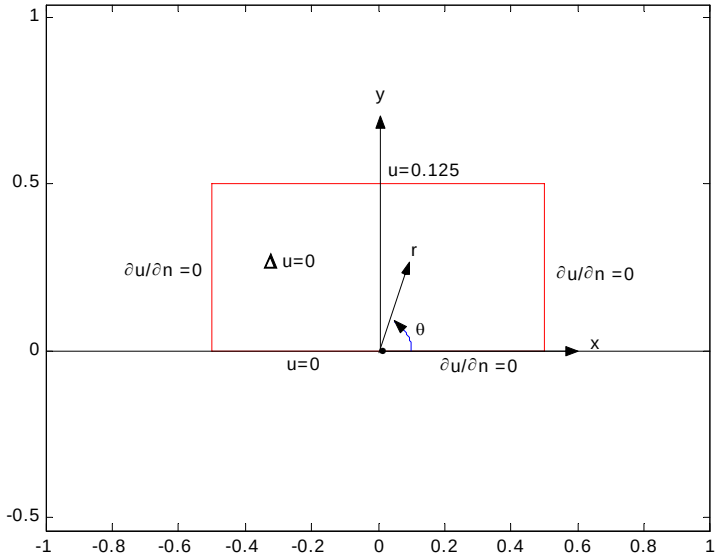


Figure 1. Cracked beam problem.

In the original problem, $\Delta v=-1$ and $v=0$ along $y=0.5$, where $\Delta=\partial^2/\partial x^2+\partial^2/\partial y^2$. By using the transformation $u=v+y^2/2$ we obtain the above boundary value problem.

Let γ_0 be the line segment $[-0.5,0]$ of the x-axis. γ_1 be the line segment $(0,0.5)$ of the x-axis. and γ_2 be its right side belonging to the line $x=0.5$ with the end points included. γ_3 and γ_4 are the line segments $y=0.5$ and $x=-0.5$ respectively, Problem (3) is known to have a unique solution u that is continuous on the closed rectangle $\overline{G}$ and infinitely differentiable everywhere on the closed rectangle $\overline{G}$ except the point $(0,0)$, where according to (1) it has the asymptotic form

$$u = Br^{1/2}\cos(\theta/2) + O(r^{3/2}), \tag{4}$$

with a constant $B > 0$. We use the polar coordinates r and θ defined as follows:

$$r = \sqrt{x^2 + y^2}, \qquad \theta = \pi/2 - \arctan(x/y) \tag{5}$$

The derivatives of u are singular at $(0,0)$, and the leading parts of the singularities are equal to the corresponding derivatives of the first term on the right-hand side of (4).

3 APPROXIMATING THE SOLUTION OF THE CRACKED BEAM PROBLEM BY THE BLOCK METHOD

Let us continue the solution u beyond the rectangle G. First, define the even continuation across the line $x = 0.5$, next define the even continuation across the line $x = -0.5$. Finally, define an odd continuation across the line $y = 0.5$ by the formula

$$u(x,y) = 0.25 - u(x,1-y), \quad -0.5 < y < 0.5, \tag{6}$$

The continued function u is harmonic on the semicircle

$$T = \{(r,\theta): 0 < r < r_0, \ \ 0 < \theta < \pi\}, \qquad \sqrt{0.5} < r_0 < 1, \tag{7}$$

and is continuous on its closure $\overline{T}$. The semicircle contains the rectangle (2). Define

$$U(r,\theta) = u(r\cos\theta, r\sin\theta), \tag{8}$$

semicircle (7) is a sector with the angle at the vertex equal to π. The function u vanishes on one of the radii of this sector, namely on the segment $[-r_0, 0]$ of the x-axis. The normal derivative $\partial u/\partial n$ vanishes on the other radius (except for the point $(0,0)$). Consequently by Theorem 3.1 from[7], the harmonic function u defined on the semicircular block $T \supset G$ can be represented as

$$u(x,y) = \frac{1}{\pi}\int_0^{\pi} U(r_0,\eta) R_0(r,\theta,\eta)\, d\eta, \tag{9}$$

where r and θ are related to x and y by (5),

$$R_0(r,\theta,\eta) = \frac{1}{4}\sum_{\beta=0}^{1}(-1)^{\beta}\sum_{\alpha=0}^{1} R\left[\left(\frac{r}{r_0}\right)^{\frac{1}{2}}, \frac{\theta}{2}, \beta\pi + (-1)^{\alpha+\beta}\frac{\eta}{2}\right], \tag{10}$$

and

$$R(r,\theta,\eta) = \frac{1-r^2}{1-2r\cos(\theta-\eta)+r^2}, \tag{11}$$

is the kernel of the Poisson integral on the unit circle.

Consider the n points $P_k^{\,0}(m = 1,2,...,n)$ uniformly distributed over the arc of the semicircle (7) with the orthogonal coordinates

$$x_k^{\,0} = r_0\cos\frac{\pi(k-1/2)}{n}, \qquad y_k^{\,0} = r_0\sin\frac{\pi(k-1/2)}{n}. \tag{12}$$

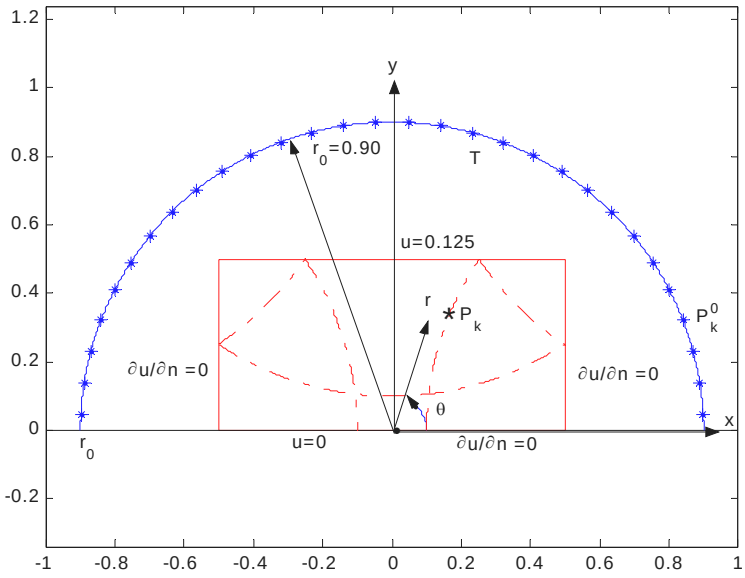


Figure 2. Points uniformly distributed over the arc of the semicircle.

These points lie outside the rectangle (2) (see Figure 2). Let us associate each point $P_k^{\,0}$ with the point P_k inside the rectangle G having the orthogonal coordinates,

$$x_k = \begin{cases} x_k^{\,0}, & \left|x_k^{\,0}\right| \le 0.5, \\ x_k^{\,0}\left(1-\left|x_k^{\,0}\right|\right)/\left|x_k^{\,0}\right|, & \left|x_k^{\,0}\right| > 0.5, \end{cases} \tag{13}$$

$$y_k = \begin{cases} y_k^{\,0}, & y_k^{\,0} \le 0.5, \\ 1-y_k^{\,0}, & y_k^{\,0} > 0.5. \end{cases} \tag{14}$$

If, $\min\{\left|x_k^{\,0}\right|, y_k^{\,0}\} \ge 1/2$, then P_k is symmetric to $P_k^{\,0}$ with respect to the vertex of G closest to $P_k^{\,0}$. Otherwise, P_k is symmetric to $P_k^{\,0}$ with respect to the vertex of G closest to $P_k^{\,0}$. The points P_k lie on the arcs of the circles represented by dashed lines in the Figure 2.

Considering equations (6), and (7), and the evenness of the function u with respect to the lines $x = 0.5$ and $x = -0.5$ we write

$$u(x_k^{\,0}, y_k^{\,0}) = \tau_k + \sigma_k u(x_k, y_k), \qquad k = 1, 2, ..., n, \tag{15}$$

$$\tau_k = \begin{cases} 0.25, & y_k^{\,0} > 0.5, \\ 0 & y_k^{\,0} \le 0.5, \end{cases} \quad \sigma_k = \begin{cases} -1, & y_k^{\,0} > 0.5, \\ 1 & y_k^{\,0} \le 0.5, \end{cases} \tag{16}$$

By (9) and (15) we have

$$u(x_k^{\,0}, y_k^{\,0}) = \tau_k + \frac{\sigma_k}{\pi}\int_0^{\pi} U(r_0, \eta) R_0(r_k, \theta_k, \eta)\, d\eta, \tag{17}$$

where r_k and θ_k are the polar coordinates of $P_k \in G \subset T$, and $u(x_k^{\,0}, y_k^{\,0})$ is the value of u at the point

$P_k^{\,0}$ located at the boundary of $T(k=1,2,...,n).$

Let

$$\eta_m = \pi(m-1/2)/n. \tag{18}$$

The integral on the right-hand side of (17) is approximated by a rectangle quadrature with equally spaced nodes $P_k^{\,0}(m=1,2,...,n).$ We replace the unknown $U(r_0,\eta_m)=u(x_m^{\,0},y_m^{\,0})$ with their desired approximate values at the nodes, u_m and substitute $u(x_k^{\,0},y_k^{\,0})$ with u_k on the left-hand side of (17). Equating the left-and right-hand sides obtained in this way, we arrive at the following linear system with respect to the unknown u_k (see [11])

$$u_k = \tau_k + \frac{\sigma_k}{n}\sum_{m=1}^{n} u_m R_0(r_k,\theta_k,\eta_m), \qquad k=1,2,...,n. \tag{19}$$

Using (10) and (11) the system (19) is transformed as

$$u_k = \tau_k + \sigma_k \sum_{m=1}^{n} u_m \sum_{\beta=0}^{1}(-1)^{\beta}\sum_{\alpha=0}^{1}\frac{d_k}{g_k-(-1)^{\beta}a_k c_m-(-1)^{\alpha}b_k s_m}, \qquad k=1,2,...,n \tag{20}$$

where τ_k and σ_k are defined by (16) and $c_m = 2\cos(\eta_m/2), \qquad s_m = 2\sin(\eta_m/2),$ and

$$a_k=\left(\frac{r_k}{r_0}\right)^{1/2}\cos\frac{\theta_k}{2}, \qquad b_k=\left(\frac{r_k}{r_0}\right)^{1/2}\sin\frac{\theta_k}{2} \quad g_k=1+\frac{r_k}{r_0}, \quad d_k=\frac{1-r_k/r_0}{4n},$$

where (r_k,θ_k) is the polar coordinate of P_k. Linear system (20) has a simple structure and does not require the complete storage of its entire matrix in a computer. Only the arrays $\{c_m,s_m\}_{m=1}^{n}, \quad \{a_k,b_k,g_k,d_k,\tau_k,\sigma_k\}_{k=1}^{n}$ are computed and stored beforehand.

Denote by $u_{kN} \quad (k=1,2,...,n)$ the approximate solution to solve the system (20) obtained after Nth Gauss-Seidel iteration from the zero initial approximation. Consider again representation (9) of the function u of the sector (7). Let (x,y) be an arbitrary point in the closed rectangle $\overline{G}$ and r and θ be its polar coordinates. Replacing the integral on the right-hand side of (9) with the rectangle quadrature and substituting each unknown value $U(r_0,\eta_k)$ with the component u_{kN} of the approximate solution to system (20) at each $P_k^{\,0}$ we obtain the approximate solution $u_N^n(x,y)$ to the boundary value problem (3) corresponding to the Nth iteration

$$u_N^n(x,y)=\frac{1}{n}\sum_{k=1}^{n} u_{kN} R_0(r,\theta,\eta_k), \tag{21}$$

where R_0 is the kernel defined by (10) and $\eta_k=\pi(k-1/2)/n.$

A FORMULA FOR THE STRESS INTENSITY FACTOR

From an engineering standpoint the most interesting quantity is not the values of stress function v but rather the stress intensity factor which is the coefficient B of the leading term in the asymptotic expression (4) for the solution u to the boundary value problem (3).

On the basis of (4), we have

$$B=\lim_{r\to+0}\left.\frac{u}{r^{1/2}}\right|_{\theta=0}.$$

This gives the following approximation of the coefficient B:

$$B_N^n = \lim_{r\to+0} \frac{1}{n}\sum_{k=1}^{n} u_{kN}\,\frac{R_0(r,0,\eta_k)}{r^{1/2}} = \frac{2}{n\sqrt{r_0}}\sum_{k=1}^{n} u_{kN}\cos\frac{\eta_k}{2}. \tag{22}$$

5 NUMERICAL RESULTS

A numerical algorithm based on the Block Method is implemented using extended precision. The linear system of equations when $r_0 = 0.90$ is solved by Gauss-Seidel method with a zero initial approximation. Iteration is controlled by the parameter

$$\varepsilon_N{}^1 = \max_{1\le j\le 50}\left| u_N{}^n\left(-0.5+\frac{j-1}{49}, 0.5\right) - 0.125\right|, \tag{23}$$

which is the maximum magnitude of the deviation of approximate solution (21) to the boundary value problem (3) from a given boundary value taken over 50 equally spaced points located on the line $y = 0.5$ of the rectangle. At any N, the approximate solution $u_N{}^n$ satisfies the boundary condition set in (3) on the bottom side of the rectangle both the left and on the right of the point $(0,0)$. Iteration is terminated at the minimal $N \ge 5$ that satisfies the condition

$$\varepsilon_N{}^1 \ge 0.999\varepsilon_{N-1}{}^1, \tag{24}$$

After the iterations are terminated, the second control parameter

$$\varepsilon_N^2 = \max\left\{\max_{1\le p\le 40} F(-0.5, \frac{p}{80}), \max_{1\le q\le 40} F(0.5, \frac{q}{80})\right\}, \tag{25}$$

is computed, where

$$F(x,y) = \left|\frac{\partial u_N^n(x,y)}{\partial x}\right|.$$

The quantity defined by (25) is the maximum deviation from zero of the normal derivative of the approximated solution over 80 equally spaced points on the left and right sides of the rectangle G

n	N	$v_N{}^n(0,\frac{1}{4})$	$v_N{}^n\left(\frac{-11}{24},\frac{1}{4}\right)$	$v_N{}^n\left(\frac{11}{24},\frac{1}{4}\right)$
20	6	0.027427	0.03287	0.07084
40	8	0.02742789	0.03287788	0.0708435
60	10	0.0027427895	0.032877886	0.070843513
80	12	0.027427895505	0.032877886397	0.07084351322
100	14	0.02742789550524	0.03287788639764	0.0708435132220
120	16	0.0274278955052476	0.0328778863976421	0.070843513222074
160	19	0.0274278955052476912	0.0328778863976422216	0.0708435132220747256
200	18	0.0274278955052476912	0.0328778863976422215	0.0708435132220747256

Table 1: Solution of cracked beam problem at specific points by (BM)

Table 1 lists the present value n (number of points over the arc of the semicircle, iteration numbers N, and the corresponding values of the stress function $v = u - y^2/2$ obtained at three points when $r_0 = 0.90$ is taken. Table 2 gives the stress intensity factor $B_N{}^n$ and the two control parameters $\varepsilon_N{}^1$, $\varepsilon_N{}^2$ with respect to n.

In the values of v_N^n and B_N^n the coinciding decimals grows almost linearly with n, the iteration number

N also grows linearly, the control parameters $\varepsilon_N^{\,1}$, $\varepsilon_N^{\,2}$, shows the exponential convergence of the BM. This conclusion is not valid for $n>160$ because the decimal precision available on the computer becomes insufficient. Here it does not sufficient to increase n to 200 to obtain v_N^n and B_N^n more than 19 decimals the number of decimal places used in a computer must be increased.

n	$B_N^{\,n}$	$\varepsilon_N^{\,1}$	$\varepsilon_N^{\,2}$
20	0.19111	2.115×10^{-5}	3.601×10^{-4}
40	0.1911186	3.389×10^{-8}	2.549×10^{-7}
60	0.1911186319	2.755×10^{-10}	1.997×10^{-9}
80	0.191118631971	2.651×10^{-12}	1.889×10^{-11}
100	0.191118631971867	2.799×10^{-14}	1.973×10^{-13}
120	0.1911186319718720 45	3.105×10^{-16}	2.171×10^{-15}
160	0.1911186319718720 89	3.184×10^{-19}	3.652×10^{-18}
200	0.1911186319718720 89	3.049×10^{-19}	3.381×10^{-18}

Table 2: The control parameters $\varepsilon_N^{\,1}$ and $\varepsilon_N^{\,2}$ and the approximate value $B_N^{\,n}$

(x,y)	OFE	BSBFM	ISBFM	Fix et al.	Wigley
$\left(0,\frac{1}{24}\right)$	0.026192	0.027431	0.027429	0.027425	0.027428
$\left(\frac{-11}{24},\frac{1}{4}\right)$	0.032847	0.032878	0.032879	0.032877	0.032878
$\left(\frac{11}{24},\frac{1}{4}\right)$	0.070657	0.070844	0.070844	0.070844	0.070844
B			0.191119	0.1917	0.19112

Table 3: The solution to the cracked beam problem from the literature

6 CONCLUSION

The rate of convergence of the approximate solution obtained hear is higher (they converge exponentially with respect to the number of quadrature nodes) than the results obtained in the above mentioned approaches. Furthermore, to compute the stress intensity factor a simple and high accurate formula is given. Therefore, the obtained extremely accurate results can be used to test the other approaches for the singular problems, also.

REFERENCES

[1] Dosiyev, A.A., (1992), " A block-grid method for increasing accuracy in the solution of the Laplace equation on polygons," Doklady Mathematics, 45, 2, pp. 396-399.

[2] Dosiyev, A.A., (1994), "A block-grid method of increased accuracy for solving Dirichlet's problem for Laplace's equation on polygons," Comp. Maths Math. Phys., 34, 5, pp. 591-604.

[3] Dosiyev, A.A., (2004), The high accurate block-grid method for solving Laplace's boundary value problem with singularities," SIAM J. Numer.Anal, 42, 1, pp.153-178.

[4] Fix, G.J., Gulati S., and Wakoff G.I., (1973), "On the use of singular functions with finite element approximations," Journal of Computational Physics 13, pp. 209-228.

[5] Olson, L.G, Georgiou, G.C., and Schultz, W.W., (1991), " An efficient finite element method for treating singularities in Laplace's equation, " J. Comput. Phys. 96, pp.391-410.

[6] Volkov, E.A., (1980), "An exponentially converging method for solving Laplace's equation on polygons," Math. USSR Sb. 37, 3, pp.295-325.

[7] Volkov, E.A., (1994), Block method for solving the Laplace equation and constructing conformal mappings, CRC Press, USA.

[8] Volkov, E.A., (1987), "An approximate conformal mapping of some polygons to a strip by the block method", Zh. Vychisl. Mat. Mat. Fiz., vol. 27, no. 8, pp. 1166-1175.

[9] Volkov, E.A., Kornoukhov, A.K., and Yakovleva, E.A., (1998), "Experimental Investigation of the block method for the Laplace equation on polygons", Zh.Vychisl. Mat. Mat. Fiz., vol. 38, no.9, pp.1544-1552.

[10] Volkov, E.A., Kornoukhov, A.K., (2002)," Solving the torsion problem for an L-section rod by the block method", Zh., Vychisl. Mat. Mat. Fiz., vol. 42, no. 8, pp.1207-1216.

[11] Volkov E. A. (2003), "On Solving the Motz Problem by a Block Method," Comput. Math. Math. Phys. 43, no. 9, pp. 1331-1337.

[12] Wigley, N.M. (1988), "An efficient method for subtracting off singularities at corners for Laplace's equation," J. Comput. Phys., 78, pp. 368-376.

5th GRACM International Congress on Computational Mechanics
Limassol, 29 June-1 July, 2005

A MESHLESS BEM FOR THE ELASTOSTATIC PROBLEM IN PLANE BODIES EXHIBITING ANISOTROPY AND INHOMOGENEITY

Maria S. Nerantzaki†, and Joanna G. Platanidi

School of Civil Engineering
National Technical University of Athens
Zografou Campus, Athens GR-15773, Greece
e-mail:majori@central.ntua.gr

†Greek Assosiarion of Computational Mechanics

Keywords: inhomogeneous, anisotropy, meshless technique, boundary elements

Abstract. *A new BEM approach is presented for the plane elastostatic problem for inhomogeneous anisotropic bodies. The incapability of establishing the fundamental solution for equations with variable coefficients is overcome by using the known fundamental solution of the Poisson equation to derive the necessary boundary integral equations. This formulation introduces two additional unknown field functions, which physically represent the two components of a fictitious body force. They are determined from two supplementary domain integral equations. The latter are converted to boundary ones by employing a meshless technique based on global approximation by radial basis functions series. Thus, the presented method maintains the pure boundary character of the BEM. The obtained numerical results demonstrate the effectiveness and accuracy of the method.*

1. INTRODUCTION

The inhomogeneity and anisotropy in plane elastic bodies results from the position dependent material constants. The response of such bodies leads to boundary value problems for partial differential equations with variable coefficients. The conventional BEM can be employed only if the respective fundamental solution is known or can be established. This, however, is out of question for differential equations with variable coefficients. Though the problem is of great interest, very few solutions can be found in the literature. Most of them deal with specific variation laws of the material constants. A more general solution has been reported by Clements et al. [1], where we refer for additional literature. The herein presented method uses the known fundamental solution of the Poisson equation to establish the necessary boundary integral equations. This new formulation, based on the concept of the analog equation [2], introduces two additional unknown field functions, which physically represent the two components of a fictitious body force. They are determined by employing a meshless technique based on global approximation by radial basis functions series of multiquandric type. Then the displacements and the stresses are evaluated from their integral representations based on the known fundamental solution. Thus, the presented method maintains the pure boundary character of the BEM, since the discretization into elements and the integrations are limited only to the boundary. Example problems are studied. The obtained numerical results demonstrate the effectiveness and accuracy of the method. A significant advantage of the proposed method is that the same computer program is utilized to obtain numerical results regardless the specific form of the governing differential operator.

2. STATEMENT OF THE PROBLEM

Consider an inhomogeneous anisotropic elastic plane body with unit thickness, occupying the two-dimensional in general multiply connected region Ω with boundary Γ. Its material constants are position dependent. The stress displacement relations are

$$\sigma_{ij}(\mathbf{x}) = c_{ijkl}\frac{\partial u_k}{\partial x_l} \tag{1}$$

where $i, j, k, l = (1,2)$, $\mathbf{x} = (x, y)$, with u_k denotes the displacement and $c_{ijkl}(\mathbf{x})$ the elastic parameters.

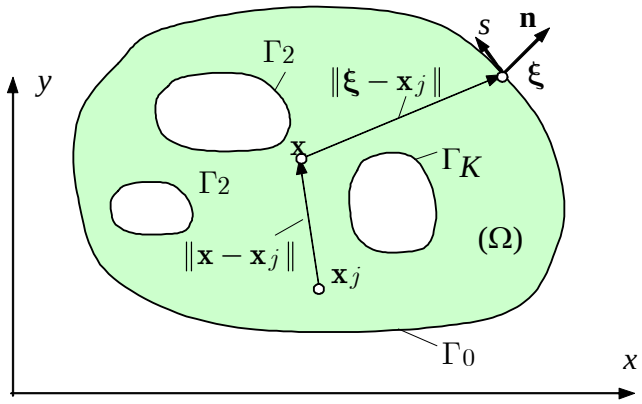


Figure 1. Domain Ω and Boundary $\Gamma = \bigcup_{i=0}^{K} \Gamma_i$

The equilibrium equations for plane stress in the xy plane are

$$\frac{\partial}{\partial x_j}\left(c_{ijkl} \frac{\partial u_k}{\partial x_l} \right) + q_i = 0 \tag{2}$$

in which q_i (i=1,2) are the body force components. The summation convention (summing from 1 to 2) is used for repeated Latin suffices.
The above equations are subjected to the boundary conditions

$$u = \tilde{u}, \quad v = \tilde{v} \qquad \text{on } \Gamma_u \tag{3a}$$

$$t_x = \tilde{t}_x, \quad t_y = \tilde{t}_y \qquad \text{on } \Gamma_t \tag{3b}$$

where $\tilde{u}$, $\tilde{v}$, $\tilde{t}_x$, $\tilde{t}_y$ represent prescribed quantities along the boundary; t_x, t_y are the boundary tractions per unit thickness defined as

$$t_i(\mathbf{x}) = \sigma_{ij} n_j = c_{ijkl} \frac{\partial u_k}{\partial x_l} n_j \tag{4}$$

where $\mathbf{n} = (n_x, n_y)$ denotes the outward pointing normal to the boundary Γ.
For all points the coefficients $c_{ijkl}(\mathbf{x})$ are required to satisfy the usual symmetry conditions

$$c_{ijkl}(\mathbf{x}) = c_{ijlk}(\mathbf{x}) = c_{jikl}(\mathbf{x}) = c_{klij}(\mathbf{x}) \tag{5}$$

This limits the number of the coefficients to 6, i.e. C_{11}, C_{12}, C_{16}, C_{22}, C_{26}, C_{66}. Moreover, the coefficients ensure the requirement that the system of partial differential equations (2) is elliptic throughout Ω.

3. SOLUTION PROCEDURE

Let $u = u(x, y)$, $v = v(x, y)$ be the sought solution of equations (2). These functions are twice differentiable in Ω. Thus, if the Laplacian operator is applied to them, we have

$$\nabla^2 u_i = b_i(x, y) \ (i=1,2) \tag{6}$$

where $u_1 = u,\ u_2 = v$

Equations (6), which henceforth will be referred to as the analog equations of the problem at hand, indicate that the solution of equations (2) could be established by solving these Poisson equations under the boundary conditions (3a) and (3b), if the fictitious sources $b_i(x, y)$ (i=1,2) were known. The fictitious sources are established using BEM, b_i are approximated by

$$b_i = \sum_{j=1}^{M} \mathrm{a}_j^{(i)} f_j \tag{7}$$

where $f_j = f_j(r)$ are M approximation radial basis functions and $a_j^{(i)}$ are 2M coefficients to be determined. Note that $r = r_{jP} = \mid P - P_j \mid$ is the distance between the collocation point $P_j : \{x_j, y_j\}$ and the point $P : \{x, y\} \in \Omega \cup \Gamma$ (see Fig. 1). We look for a solution in the form $\overline{u}_i + u_i^p$ where $\overline{u}_i$ is the homogeneous solution and u_i^p a particular one. The particular solution is obtained as

$$u_i^p = \sum_{j=1}^{M} \mathrm{a}_j^{(i)} \hat{u}_j \tag{8}$$

where $\hat{u}_j = \hat{u}_j(r_{jP}) = \hat{u}_j(x,y)$ is a particular solution of

$$\nabla^2 \hat{u}_j = f_j \tag{9}$$

Thus, writing the solution of the homogeneous equations (6) in integral form, we have

$$c\overline{u}_i(P) = -\int_{\Gamma} [u^*(r_{Pq}) \frac{\partial \overline{u}_i}{\partial n}(q) - \overline{u}_i(q) \frac{\partial u^*(r_{Pq})}{\partial n}] ds_q \tag{10}$$

where $r_{Pq} = \mid q - P \mid$ is the distance between the field point $P : \{x, y\} \in \Omega \cup \Gamma$ and source point $q : \{\xi, \eta\}$ which varies during the integration.

Moreover

$$u^*(r_{Pq}) = \frac{1}{2\pi} \ln(r_{Pq}) \tag{11}$$

is the fundamental solution of the Laplace equation and

$$\frac{\partial u^*(r_{Pq})}{\partial n} = \frac{1}{2\pi} \frac{\cos\varphi}{r_{Pq}} \tag{12}$$

is its normal derivative with respect to point $q \in \Gamma$; $\varphi = \measuredangle r_{Pq}, \mathbf{n}$. Finally $c = 1, \alpha / 2\pi, 0$ depending on whether $P \in \Omega$, $p \in \Gamma$, $P \notin \Omega \cup \Gamma$, respectively; α is the angle between the tangents to the boundary at point P. For points where the boundary is smooth it is $c = 1/2$.

For points P inside Ω ($c = 1$) the solution of equations (6) on the basis of equations (8) and (10), is written as

$$u_i(P) = -\int_{\Gamma} [u^*(r_{Pq}) \frac{\partial \overline{u}_i}{\partial n}(q) - \overline{u}_i(q) \frac{\partial u^*(r_{Pq})}{\partial n}] ds_q + \sum_{j=1}^{M} \mathrm{a}_j^{(i)} \hat{u}_j(r_{Pq}) \tag{13}$$

Using the BEM with N constant boundary elements equations (10) yields

$$\mathbf{H}\bar{\mathbf{u}} - \mathbf{G}\frac{\partial \bar{\mathbf{u}}}{\partial n} = 0 \tag{14}$$

in which $\mathbf{H}$ and $\mathbf{G}$ are known matrices originating from the integration of kernels of the integral equations on the boundary; their dimensions are, $2N \times 2N$ and $2N \times 2N$, respectively; $\bar{\mathbf{u}} = \{\bar{u}_1\, \bar{u}_2 \cdots \bar{u}_N\, \bar{v}_1\, \bar{v}_2 \cdots \bar{v}_N\}^T$, $\frac{\partial \bar{\boldsymbol{u}}}{\partial n} = \{\frac{\partial \bar{u}_1}{\partial n}\frac{\partial \bar{u}_2}{\partial n}\cdots\frac{\partial \bar{u}_N}{\partial n}\frac{\partial \bar{v}_1}{\partial n}\frac{\partial \bar{v}_2}{\partial n}\cdots\frac{\partial \bar{v}_N}{\partial n}\}^T$ are vectors including the $2N$ nodal values of the boundary displacements and normal derivatives of the homogenous solution.

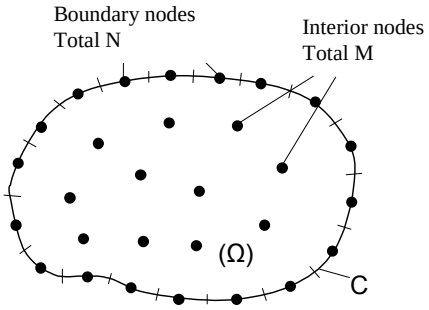


Figure 2. Boundary discretization and domain nodal points

Application of equations (13) at the M interior collocating points $\mathbf{x}_j$ in Ω ($c = 1$) (see Fig. 2) yields

$$\hat{\mathbf{u}} = \hat{\mathbf{B}}\mathbf{a} + \hat{\mathbf{H}}\bar{\mathbf{u}} - \hat{\mathbf{G}}\frac{\partial \bar{\mathbf{u}}}{\partial n} \tag{15}$$

where $\hat{\boldsymbol{u}} = \{\hat{u}_1\, \hat{u}_2 \cdots \hat{u}_N\, \hat{v}_1\, \hat{v}_2 \cdots \hat{v}_N\}^T$ includes the values of the displacements at the internal collocation points. The hat designates quantities referring to the internal collocation points. We differentiate now equations (13) to obtain the first and second derivatives of the displacements at the internal collocation points. Thus, using the same discretization we obtain

$$\hat{\mathbf{u}}_{,x} = \hat{\mathbf{B}}_x\mathbf{a} + \hat{\mathbf{H}}_x\bar{\mathbf{u}} - \hat{\mathbf{G}}_x\frac{\partial \bar{\mathbf{u}}}{\partial n} \tag{16a}$$

$$\hat{\mathbf{u}}_{,y} = \hat{\mathbf{B}}_y\mathbf{a} + \hat{\mathbf{H}}_y\bar{\mathbf{u}} - \hat{\mathbf{G}}_y\frac{\partial \bar{\mathbf{u}}}{\partial n} \tag{16b}$$

$$\hat{\mathbf{u}}_{,xx} = \hat{\mathbf{B}}_{xx}\mathbf{a} + \hat{\mathbf{H}}_{xx}\bar{\mathbf{u}} - \hat{\mathbf{G}}_{xx}\frac{\partial \bar{\mathbf{u}}}{\partial n} \tag{16c}$$

$$\hat{\mathbf{u}}_{,xy} = \hat{\mathbf{B}}_{xy}\mathbf{a} + \hat{\mathbf{H}}_{xy}\bar{\mathbf{u}} - \hat{\mathbf{G}}_{xy}\frac{\partial \bar{\mathbf{u}}}{\partial n} \tag{16d}$$

$$\hat{\mathbf{u}}_{,yy} = \hat{\mathbf{B}}_{yy}\mathbf{a} + \hat{\mathbf{H}}_{yy}\bar{\mathbf{u}} - \hat{\mathbf{G}}_{yy}\frac{\partial \bar{\mathbf{u}}}{\partial n} \tag{16e}$$

where $\hat{\mathbf{B}}, \hat{\mathbf{B}}_x, \hat{\mathbf{H}}_x, \hat{\mathbf{G}}_x, \ldots, \hat{\mathbf{B}}_{yy}, \hat{\mathbf{H}}_{yy}, \hat{\mathbf{G}}_{yy}$ are known matrices originating again from the integration of kernels of the integral equations; $\mathbf{u}_{,x}, \mathbf{u}_{,y}, \ldots, \mathbf{u}_{,yy}$ are the vectors including the derivative of $\hat{\mathbf{u}}$ and $\mathbf{a} = \{a_1^{(1)}\, a_2^{(1)} \cdots a_M^{(1)}\, a_1^{(2)}\, a_2^{(2)} \cdots a_M^{(2)}\}^T$ is the vector including the $2M$ unknown coefficients.

Finally, applying equations (2) at the M internal collocation points and using equations (16) we obtain the following set of $2M$ equations.

$$\mathbf{A}\mathbf{a} + \boldsymbol{C}\bar{\mathbf{u}} + \mathbf{D}\frac{\partial \bar{\mathbf{u}}}{\partial n} = \mathbf{F} \tag{17}$$

Equations (17) and (14) constitute a set of $2N + 2M$ simultaneous equations for $4N + 2M$ unknowns, i.e. $\bar{\mathbf{u}}, \frac{\partial \bar{\mathbf{u}}}{\partial n}, \mathbf{a}$. The additional required $2N$ equations will result from the boundary conditions. Before writing these equations certain manipulations are necessary to express the boundary tractions t_x, t_y prescribed on Γ_t by equations (4). Thus we may write

$$t_i(x) = \sigma_{ij} n_j = c_{ijkl}\frac{\partial u_k}{\partial x_l} n_j = c_{ijkl}\left[\frac{\partial u_k}{\partial n}\cos(\boldsymbol{n}, x_l) + \frac{\partial u_k}{\partial t}\cos(\boldsymbol{t}, x_l)\right] n_j \tag{18}$$

The tangential derivatives $u_{,t}$, $v_{,t}$ can be expressed in terms of the boundary displacements using numerical differentiation. Thus, the boundary conditions on Γ_t will have the form

$$\mathbf{a}_1\bar{\mathbf{u}} + \mathbf{a}_2\frac{\partial \bar{\mathbf{u}}}{\partial n} + \mathbf{a}_3\mathbf{a} = \mathbf{a}_4 \tag{19}$$

while on Γ_u

$$\boldsymbol{b}_1\bar{\mathbf{u}} + \boldsymbol{b}_2\frac{\partial \bar{\mathbf{u}}}{\partial \mathrm{n}} + \boldsymbol{b}_3\mathbf{a} = \boldsymbol{b}_4 \tag{20}$$

Equations (19) and (20) may be combined as

$$\mathbf{A}_1\bar{\mathbf{u}} + \mathbf{A}_2\frac{\partial \bar{\mathbf{u}}}{\partial n} + \mathbf{A}_3\mathbf{a} = \mathbf{A}_4 \tag{21}$$

where

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{a}_1 \\ \boldsymbol{b}_1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} \mathbf{a}_2 \\ \boldsymbol{b}_2 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} \mathbf{a}_3 \\ \boldsymbol{b}_3 \end{bmatrix} \tag{22}$$

Once the quantities $\bar{\mathbf{u}}, \frac{\partial \bar{\mathbf{u}}}{\partial n}, \mathbf{a}$ are established the displacements and their derivatives are evaluated using equations (15) and (16). For points $\mathbf{x}$ not coinciding with the internal collocation points the displacements are evaluated from the discretized counterparts of equations (13).

4. NUMERICAL RESULTS

On the basis of the analytical and numerical procedures developed in previous section, a computer program in FORTRAN has been written to obtain numerical results for the considered problem. Certain example problems have been studied. The obtained numerical results validate the effectiveness and accuracy of the developed method for solving this type of problems. Radial basis functions of multiquadric type were employed to obtain the numerical results (see Appendix).

Example 1. Square plane body

A square plane body, $\Omega: -0.5 \le x, y \le 0.5$, of uniform unit thickness with variable modulus of elasticity subjected to prescribed boundary displacements, tractions and body forces has been analyzed. The employed data are

$E = 110000 + 30(1+x)$, $E_1 = E/\sqrt{\kappa}$, $E_2 = E\sqrt{\kappa}$, $\nu_1 = 0.3$, $\nu_2 = \kappa\nu_1$, $\kappa = 2$, $G_{12} = \frac{E}{2}(1+\nu_1\sqrt{\kappa})$,

$$q_x = -18.2926\sqrt{2}y - 10.976\sqrt{2}x - 0.365(110030+30x)\sqrt{2} - \frac{1.4(110030+30x)}{2+0.6\sqrt{2}},$$

$$q_y = -30\frac{1.4x+1.4y}{2+0.6\sqrt{2}} - \frac{1.4(110030+30x)}{2+0.6\sqrt{2}} - 0.365(110030+30x)\sqrt{2}$$

The body was subjected to the prescribed boundary conditions. The obtained numerical results are presented in Table 1 as compared with those from the exact solution.

$\tilde{u}_1 = -0.5x + 0.05, \quad \tilde{t}_{y1} = 0.1825(110030+30x)\sqrt{2} - 1.2195(110030+30x)\sqrt{2}x$ on $y = -0.5$

$\tilde{u}_2 = 0.5y + 0.2y^2, \quad \tilde{v}_2 = 0.5y + 0.05$ on $x = 0.5$

$\tilde{u}_3 = 0.5x + 0.05, \quad \tilde{t}_{y3} = 0.1825(110030+30x)\sqrt{2} + 1.2195(110030+30x)\sqrt{2}x$ on $y = 0.5$

$\tilde{u}_4 = -0.5y + 0.2y^2, \quad \tilde{v}_4 = -0.5y + 0.05$ on $x = -0.5$

Table 1: Computed interior displacements in a square body

		u		*v*	
x	*y*	Present	Exact	Present	Exact
-0.375	0.375	-0.113	-0.113	-0.112	-0.113
-0.25	0.375	-0.065	-0.066	-0.081	-0.081
-0.125	0.375	-0.019	-0.019	-0.044	-0.044
0.	0.375	0.028	0.028	0	0
0.125	0.375	0.075	0.075	0.049	0.05
0.25	0.375	0.121	0.122	0.106	0.106
0.375	0.375	0.169	0.169	0.169	0.169

Example 2. Deformation of a layered anisotropic elastic slab

In this example a layered slab in frictionless contact with a rigid base along one side under compression along the whole side BC is considered. The material consists of 10 homogeneous anisotropic layers which lie in the intervals $0.1n \le y \le 0.1(n+1)$ for $n = 0,1,2,....,9$

The constant elastic modulus in each of these layers are given by

$c_1 = 6.14[1+\varepsilon\sin(0.1\pi n)], \; c = 1.64[1+\varepsilon\sin(0.1\pi n)], \; c_2 = 5.96[1+\varepsilon\sin(0.1\pi n)], \; c_{12} = 1.64[1+\varepsilon\sin(0.1\pi n)]$

for $n = 0,1,2,....,9$ with ε a small parameter. The boundary conditions for this problem are (see Fig. 3)

$\tilde{t}_x = 0, \; \tilde{v} = 0$ on DA

$\tilde{t}_x = 0, \; \tilde{t}_y = 0$ on AB and CD

$\tilde{t}_x = 0, \; \tilde{t}_y = -\hat{P}$ on BC

It is possible to obtain an analytical solution to this problem by taking displacements in each homogeneous layer in the form $u = a_n x + b_n, \quad v = c_n y + d_n$ for $0.1n \le y \le 0.1(n+1)$ and $n = 0,1,2,....,9$ where a_n, b_n, c_n, d_n are constants. Imposing continuity conditions on the displacements and tractions at each of interfaces $y = 0.1, 0.2, ..., 0.9$ together with the boundary conditions at the boundaries AB, BC, CD and DA leads to 40 linear algebraic equations which may be solved for the 40 unknowns $a_0, b_0, c_0, d_0, a_1, b_1, c_1, d_1, ..., a_9, b_9, c_9, d_9$. Equations of displacements then provides the displacements at all points of the layered slab.

Results of some selected points and selected values of ε are given in Tables 2 and 3 .

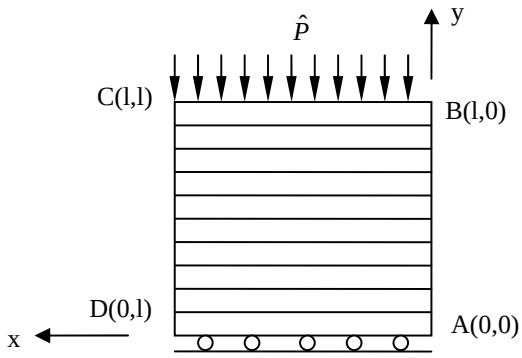


Figure 3. Deformation of a layered slab

Table 2: Displacement v for the layered slab

Position	$\varepsilon = 0.1$		$\varepsilon = 0.2$		$\varepsilon = 0.3$	
(y,x)	BEM	Analytical	BEM	Analytical	BEM	Analytical
(0.25,0.5)	-0.042	-0.042	-0.040	-0.041	-0.038	-0.041
(0.45,0.5)	-0.076	-0.075	-0.071	-0.072	-0.068	-0.069
(0.65,0.5)	-0.110	-0.107	-0.103	-0.101	-0.097	-0.097
(0.85,0.5)	-0.143	-0.140	-0.134	-0.132	-0.127	-0.125

Table 3 : Stresses σ_y for the layered slab

Position	$\varepsilon = 0.1$		$\varepsilon = 0.2$		$\varepsilon = 0.3$	
(y,x)	BEM	Analytical	BEM	Analytical	BEM	Analytical
(0.25,0.5)	-1.030	-1.0	-1.058	-1.0	-1.085	-1.0
(0.45,0.5)	-1.028	-1.0	-1.053	-1.0	-1.077	-1.0
(0.65,0.5)	-1.022	-1.0	-1.042	-1.0	-1.060	-1.0
(0.85,0.5)	-1.012	-1.0	-1.022	-1.0	-1.031	-1.0

6. REFERENCES

[1] Azis, I.M., Clements, D.A. (2001), "A boundary element method for anisotropic inhomogeneous elasticity", International Journal of Solids and Structures, 38, pp. 5747-5763.

[2] Katsikadelis, J.T. (2002), "The analog boundary integral equation method for nonlinear static and dynamic problems in continuum mechanics", Journal of Theoretical and Applied Mechanics, 40, 4,pp. 961-984.

[3] Katsikadelis, J.T. (2002), *Boundary Elements: Theory and Applications*, Elsevier Science, Amsterdam,.

APPENDIX A

A.1. Derivatives of the kernel functions

The derivatives of the distance

$$r = \sqrt{(\xi - x)^2 + (\eta - y)^2} \quad \{x, y\} \in \Omega, \quad \{\xi, \eta\} \in \Gamma \tag{A.1}$$

are evaluated from the following relations [3]

$$r_x = -r_\xi = -\frac{\xi - x}{r}, \quad r_y = -r_\eta = -\frac{\eta - y}{r}$$

$$r_{xx} = \frac{r_y^2}{r}, \quad r_{yy} = \frac{r_x^2}{r}, \quad r_{xy} = -\frac{r_x r_y}{r}$$

$$r_n = -(r_x n_x + r_y n_y), \quad r_t = -(-r_x n_y + r_y n_x) \tag{A.2}$$

n_ξ , n_η are directional cosines of the outward normal vector to the boundary at point $\{\xi, \eta\}$.
Using equations (A.2) the derivatives of the fundamental solution may be expressed as

$$u_x^* = \frac{1}{2\pi}\frac{r_x}{r}, \quad u_y^* = \frac{1}{2\pi}\frac{r_y}{r}$$

$$u_{xx}^* = \frac{1}{2\pi}\frac{r_y^2 - r_x^2}{r^2}, \quad u_{yy}^* = -u_{xx}^*, \quad u_{xy}^* = -\frac{1}{\pi}\frac{r_x r_y}{r^2}$$

$$u_{nxx}^* = -\frac{1}{\pi}\frac{(r_y^2 - r_x^2)r_n + 2r_x r_y r_t}{r^3} \quad u_{nyy}^* = -u_{nxx}^* \quad u_{nxy}^* = -\frac{1}{\pi}\frac{(r_y^2 - r_x^2)r_t + 2r_x r_y r_n}{r^3} \tag{A.3}$$

A.2. Derivatives of the function $\hat{u}_j$

For the approximation radial basis functions of multiquadric type yields after integration equations (9)

$$\hat{u}_j = -\frac{c^3}{3}\ln(c\sqrt{r^2 + c^2} + c^2) + \frac{1}{9}(r^2 + 4c^2)\sqrt{r^2 + c^2} \quad \text{where } r = \sqrt{(x - x_j)^2 + (y - y_j)^2} \tag{A.4}$$

Differentiating equations (A4) gives

$$\hat{u}_{j,x} = \frac{1}{3\sqrt{r^2 + c^2}}(r^2 + 2c^2 - \frac{c^3}{\sqrt{r^2 + c^2} + c}(x - x_j) \quad \hat{u}_{j,y} = \frac{1}{3\sqrt{r^2 + c^2}}(r^2 + 2c^2 - \frac{c^3}{\sqrt{r^2 + c^2} + c}(y - y_j)$$

$$\hat{u}_{j,xx} = \frac{1}{3(r^2 + c^2)^{3/2}}\left[r^2 + c^3\frac{2\sqrt{r^2 + c^2} + c}{(\sqrt{r^2 + c^2} + c)^2}\right](x - x_j) + \frac{1}{3\sqrt{r^2 + c^2}}\left(r^2 + 2c^2 - \frac{c^3}{\sqrt{r^2 + c^2} + c}\right)$$

$$\hat{u}_{j,yy} = \frac{1}{3(r^2 + c^2)^{3/2}}\left[r^2 + c^3\frac{2\sqrt{r^2 + c^2} + c}{(\sqrt{r^2 + c^2} + c)^2}\right](y - y_j) + \frac{1}{3\sqrt{r^2 + c^2}}\left(r^2 + 2c^2 - \frac{c^3}{\sqrt{r^2 + c^2} + c}\right)$$

$$\hat{u}_{j,xy} = \frac{1}{3(r^2 + c^2)^{3/2}}\left[r^2 + c^3\frac{2\sqrt{r^2 + c^2} + c}{(\sqrt{r^2 + c^2} + c)^2}\right](x - x_j)(y - y_j) \tag{A.5}$$

It can be readily proved that

$$\lim_{r\to 0}\hat{u}_{j,x} = 0, \qquad \lim_{r\to 0}\hat{u}_{j,y} = 0 \quad \lim_{r\to 0}u_{j,xx} = \frac{c}{2}, \quad \lim_{r\to 0}u_{j,yy} = \frac{c}{2}, \quad \lim_{r\to 0}u_{j,xy} = 0$$

A BEM BASED MESHLESS VARIATIONAL METHOD FOR THE SECOND ORDER ELLIPTIC PARTIAL DIFFERENTIAL EQUATION USING OPTIMAL MULTIQUADRICS

John T. Katsikadelis †

School of Civil Engineering
National Technical University of Athens
15543 Athens, Greece
e-mail: jkats@central.ntua.gr

†Greek Association of Computational Mechanics

Keywords: Boundary element method, meshless, partial differential equations, analog equation, Ritz method, variational, optimal multiquadrics.

Abstract. *A BEM based meshless variational method is presented for the solution of second order elliptic partial differential equation with variable coefficients on domains of arbitrary shape, which may be multiply connected. Using the concept of the analog equation, the original equation is converted into a Poisson equation under a fictitious source, which is approximated with radial basis function series of multiquadric (MQ) type. The integral representation of the solution of the substitute equation yields shape functions, which satisfy both essential and natural boundary conditions. The functional that yields the considered PDE as the associated Euler-Lagrange equation is constructed. The minimization of this functional yields not only the Ritz coefficients but also permits the evaluation of optimal values for the shape parameters, which may be different for each collocation point of the multiquadrics as well as optimal position of the interior collocation points. Since the arising domain integrals are converted to boundary line integrals, the method is boundary-only and, therefore, it maintains all the advantages of the pure BEM. Example problems are studied, which demonstrate the efficiency and the great accuracy of the developed method.*

1 INTRODUCTION

One the most useful approximate methods for solving differential equations stemming from variational principles is the Ritz method [1]. In this method the solution of the differential equations is approximated by the finite series

$$u_n = \sum_{i=1}^{n} a_i \psi_i(\mathbf{x}), \quad \mathbf{x} : (x, y) \tag{1}$$

where $\psi_i(\mathbf{x})$ are known functions satisfying the essential (kinematic) boundary conditions, and a_i coefficients to be determined from the minimization of the functional, whose the associated Euler-Lagrange equation is the PDE to be solved. The coefficients thus determined are referred to as the Ritz coefficients and the functions $\psi_i(\mathbf{x})$ as shape functions or Ritz coordinates. It can be shown that for $n \to \infty$ in this process the sequence $\{u_n\}$ converges in energy to the exact solution provided that the set of $\psi_i(\mathbf{x})$ is complete [2]. When n is small, we can still reach very good approximations to the exact solution if a judicious selection of $\psi_i(\mathbf{x})$ is made. It is apparent that the application of the Ritz method requires beforehand: (a) The existence of a functional that yields the considered PDE, (b) The possibility to select shape functions satisfying the kinematic boundary conditions

For differential equations of mathematical physics resulting from variational principles, the first requirement is satisfied. For non-Euler functional classes of PDE's, either extremal variational principles can be constructed [3] or recourse to the Galerkin principle can be made [2]. The second requirement appears as the most important because construction of admissible shape function is a very difficult problem. Such functions have been constructed for domains of simple geometry such as rectangles, circles, triangles, and ellipses. However, no method for constructing systematically admissible shape functions for domains of irregular shape has been reported in the literature. Therefore, the applicability of the Ritz method has been limited. Actually, the FEM overcomes this major shortcoming by constructing admissible functions at element level at the great cost of domain discretization and the discontinuity complications.

In this paper a method is presented for constructing admissible shape functions for domains of arbitrary shape, which satisfy not only the kinematic but also the natural boundary conditions. Without restricting its generality, the method is demonstrated for the complete second order elliptic PDE with variable coefficients. The proposed method is based on the concept of the analog equation [4,5], which converts the original equation to a Poisson equation with a fictitious unknown source. The integral representation of the solution of the substitute equation yields admissible shape functions, which satisfy both essential and natural boundary conditions. Hence we may refer to this new method as the *generalized* Ritz method. The unknown source is represented by a radial basis functions series, whose coefficients constitute the Ritz coefficients. Radial basis functions of MQ type are employed in this investigation. The accuracy of the solution depends on a shape parameter of the MQ's and the position of the domain collocation points. Therefore, extended research is ongoing to optimize these parameters [6]. Nevertheless, all these quantities are chosen rather arbitrarily or empirically. A major advantage of the presented method is that it permits the establishment of optimum values for the shape parameters of the MQ's as well as for position of the collocation points, because these quantities can be included in the design parameters of the optimization problem. Moreover, the domain integral of the functional that must be evaluated during the minimization process is converted into a boundary integral. Thus, no domain discretization is required and the method becomes boundary-only maintaining thus all the advantages of the BEM. The obtained numerical results demonstrate the effectiveness and accuracy of the method and bring the Ritz method again into the arena.

2 PROBLEM STATEMENT

We consider the partial differential equation

$$A\,u_{,xx} + 2Bu_{,xy} + Cu_{,yy} + Du_{,x} + Eu_{,y} + Fu = g(\mathbf{x}) \quad \text{in}\, \mathbf{x} \in \Omega \tag{2}$$

subject to the boundary conditions

$$\begin{aligned} u &= \alpha(\mathbf{x}), \quad \mathbf{x} \in \Gamma_a \\ \kappa u + (\mathbf{m}\cdot\mathbf{n})\,u_{,n} + (\mathbf{m}\cdot\mathbf{t})\,u_{,t} &= \gamma(\mathbf{x}), \;\; \mathbf{x} \in \Gamma_b \end{aligned} \tag{3a,b}$$

where $\Gamma = \Gamma_a \cup \Gamma_b$ is the boundary of Ω, which may be multiply-connected; $u = u(\mathbf{x})$ is the unknown field function; $A, B, \ldots, F$ position dependent coefficients satisfying the ellipticity condition $B^2 - AC < 0$ and $\mathbf{m} = (An_x + Bn_y)\mathbf{i} + (Bn_x + Cn_y)\mathbf{j}$ is a vector in the direction of the connormal on the boundary; $\mathbf{n}$ and $\mathbf{t}$ are the normal and the tangential unit vectors on the boundary. Finally, $\kappa(\mathbf{x})$, $\alpha(\mathbf{x})$, $\gamma(\mathbf{x})$ are functions specified on Γ. In writing the boundary condition (3b) the identity $\nabla u \cdot \mathbf{m} = (\mathbf{m}\cdot\mathbf{n})\,u_{,n} + (\mathbf{m}\cdot\mathbf{t})\,u_{,t}$ has been employed, for reasons explained later. A direct BEM solution for the problem at hand is out of question, because, except for special cases, the fundamental solution can not be established. For this purpose we construct the functional

$$J(u) = \int_\Omega \left[\frac{1}{2}\left(A\,u_{,x}^2 + 2Bu_{,x}\,u_{,y} + Cu_{,y}^2 - Fu^2\right) + gu\right] d\Omega + \int_\Gamma \left(\frac{1}{2}\kappa u^2 - \gamma u\right) ds \tag{4}$$

We can readily show that the condition $\delta J(u) = 0$ yields the boundary value (2), (3) provided that

$$A_{,x} + B_{,y} = D\,, \qquad B_{,x} + C_{,y} = E \tag{5a,b}$$

Therefore, the solution of Eqn (2) under the boundary conditions (3a,b) make $J(u) = \min$.

The boundary value problem (2), (3) under the conditions (5) for suitable meaning of the coefficients occurs in many physical problems such as thermostatic, elastostatic, electrostatic and seepage problems, where the involved media exhibit heterogeneous anisotropic properties.

3 THE ANALOG EQUATION SOLUTION

Let $u = u(\mathbf{x})$ be the sought solution to Eqn (2). If the Laplace operator is applied to it, we have

$$\nabla^2 u = b(\mathbf{x}) \tag{6}$$

where $b(\mathbf{x})$ represents an unknown fictitious source. Eqn (6) indicates that the solution of Eqn (2) could be established by solving this equation under the boundary conditions (3), if the $b(\mathbf{x})$ is first established. For this purpose we write the solution of Eqn (6) in integral form [6].

$$\varepsilon u(\mathbf{x}) = \int_\Omega u^* b d\Omega - \int_\Gamma (u^* q - q^* u) ds \quad \mathbf{x} \in \Omega \cup \Gamma \tag{7}$$

in which $q = u_{,n}$; $u^* = \ell nr/2\pi$ is the fundamental solution to Eqn (6) and $q^* = u^*_{,n}$ its derivative normal to the boundary with $r = \|\boldsymbol{\xi} - \mathbf{x}\|$ $\mathbf{x} \in \Omega \cup \Gamma$ and $\boldsymbol{\xi} \in \Gamma$; ε is a constant ($\varepsilon = 1$ if $\mathbf{x} \in \Omega$, $\varepsilon = \alpha/2\pi$ if $\mathbf{x} \in \Gamma$ and $\varepsilon = 0$ if $\mathbf{x} \notin \Omega \cup \Gamma$; α is the interior angle between the tangents of boundary at point $\mathbf{x}$). Note that it is $\varepsilon = 1/2$ for points where the boundary is smooth.

The domain integral in Eqn (7) is converting the to boundary line integral by setting

$$b = \sum_{j=1}^{M} \alpha_j f_j \tag{8}$$

where $f_j = f_j(r)$, $r = \|\mathbf{x} - \mathbf{x}_j\|$, is a set of radial basis approximating functions; $\mathbf{x}_j$ are collocations points in Ω and α_j coefficients to be determined. Using the Green's reciprocal identity [6] we have

$$\begin{aligned} \int_\Omega u^* b d\Omega &= \sum_{j=1}^{M} a_j \int_\Omega u^* f_j d\Omega \\ &= \sum_{j=1}^{M} a_j \left\{ \varepsilon \hat{u}_j + \int_\Gamma \left[u^* \hat{q}_j - q^* \hat{u}_j \right] ds \right\} \end{aligned} \tag{9}$$

in which $\hat{u}_j(\mathbf{x})$ is a particular solution of the equation

$$\nabla^2 \hat{u}_j = f_j \qquad j = 1,2,\ldots,M \tag{10}$$

We can get rid off the tangential derivative $u_{,t}$ in Eqn (3b) applying one of the following methods: (i) Using the integral representation of the tangential derivative on the boundary [6], (ii) Introducing $q = u_{,n}$ from (3b) into (7) and integrating by parts along the boundary to eliminate $u_{,t}$. This is always feasible because it $(\mathbf{m} \cdot \mathbf{n}) \neq 0$ as a result of the ellipticity condition, (iii) Approximating the derivative numerically using the nodal values of u on the discretized boundary. The last method is the computationally simplest and it is used in this investigation. On the base of the aforementioned it is convenient to combine the boundary conditions (3a,b) and write them in the form

$$\beta_1 u + \beta_2 q = \beta_3 \tag{11}$$

where β_i are known quantities.

4 NUMERICAL IMPLEMENTATION

4.1 The admissible functions

The BEM with constant elements is used to approximate the boundary integrals in Eqn (7). In this case it is $\alpha = \pi$, hence $\varepsilon = 1/2$ on the elements. If N is the number of the boundary nodal points (see Fig. 1), we obtain [6]

$$\mathbf{Hu} - \mathbf{Gq} + \mathbf{Ka} = \mathbf{0} \tag{12}$$

where $\mathbf{a}$ is the vector of the M coefficients a_j; $\mathbf{u}$, $\mathbf{q}$ are the vectors of the N boundary nodal values of u and q, respectively, and

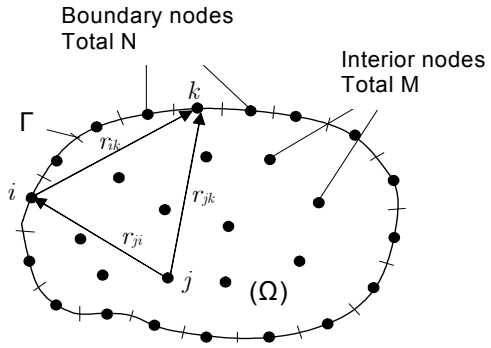


Figure 1: Boundary discretization and domain nodal points.

$$K_{ij} = \frac{1}{2}\hat{u}_j^i - \sum_{k=1}^{N} \tilde{H}_{ik}\hat{u}_j^k + \sum_{k=1}^{N} G_{ik}\hat{q}_j^k\,, \qquad G_{ik} = \int_k u^*(r_{ik})ds \tag{13a,b}$$

$$\tilde{H}_{ik} = \int_k q^*(r_{ik})ds\,, \qquad H_{ik} = \int_k q^*(r_{ik})ds - \delta_{ik}\,, \tag{13c,d}$$

in which $\int_k$ indicates integration on the k element and $\hat{u}_j^i = \hat{u}(r_{ji})$, $\hat{u}_j^k = \hat{u}(r_{jk})$. Moreover, the boundary condition (11), when applied to the N boundary nodal points, yields

$$\boldsymbol{\beta}_1\mathbf{u} + \boldsymbol{\beta}_2\mathbf{q} = \boldsymbol{\beta}_3 \tag{14}$$

where $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ are known $N \times N$ diagonal matrices and $\boldsymbol{\beta}_3$ vector including the values of β_i, $i = 1,2,3$ at the N boundary nodal points. For points $\mathbf{x}$ inside Ω, the discretized counterpart of Eqn (7) gives

$$u(\mathbf{x}) = \sum_{j=1}^{M} K_{\mathbf{x}j}a^j + \sum_{k=1}^{N} \tilde{H}_{\mathbf{x}k}u^k - \sum_{k=1}^{N} G_{\mathbf{x}k}q^k \tag{15}$$

Eqns (12) and (14) can be used to eliminate the boundary quantities $\mathbf{u}, \mathbf{q}$ from Eqn (15). Thus we obtain

$$u(\mathbf{x}) = \sum_{j=1}^{M} a_j S_j(\mathbf{x}) + b(\mathbf{x}) \tag{16}$$

For homogeneous boundary conditions it is $b(\mathbf{x}) = 0$. Thus, we may set

$$\psi_j(\mathbf{x}) = S_j(\mathbf{x})\,, \quad \psi_0(\mathbf{x}) = b(\mathbf{x}) \tag{17}$$

Apparently, the functions defined in Eqn (17) are admissible functions satisfying not only the kinematic boundary conditions but also the natural ones. More, specifically, $\psi_j(\mathbf{x})$ satisfy homogeneous boundary conditions while $\psi_0(\mathbf{x})$ the given nonhomgeneous boundary conditions. The derivatives are of $\psi_j(\mathbf{x})$ for points $\mathbf{x}$ inside Ω are obtained by direct differentiation of Eqn (7) and following similar numerical procedure to eliminate the boundary values.

4.2 Evaluation of the Ritz Coefficients

The coefficients a_j will be evaluated using expression (16) and minimizing the functional (4). The minimization procedure requires the numerical evaluation of $J(u)$. Since the expression (16) satisfies the boundary conditions, both kinematic and natural, the boundary integral vanishes. In order to maintain the pure boundary character of the method, the domain integral in (4) is converted to line boundary integral using the following procedure. Let

$$R(\mathbf{x}) = \frac{1}{2}\left(Au_{,x}^2 + 2Bu_{,x}\,u_{,y} + Cu_{,y}^2 - Fu^2\right) + gu \tag{19}$$

represent the integrand of the domain integral. We approximate it by

$$R(\mathbf{x}) \simeq \sum_{j=1}^{M} \overline{a}_j f_j(r) \tag{20}$$

where $f_j(r)$ radial functions, not necessarily the same used in (8). Application of (19) at the collocation points yields

$$\mathbf{\Phi}\overline{\mathbf{a}} = \mathbf{R} \tag{20}$$

where

$$\mathbf{\Phi} = [f(r_{ji})], \quad \mathbf{R} = \{R(x_i)\}, \quad r_{ji} = \left\|\mathbf{x}_i - \mathbf{x}_j\right\|, \quad i, j = 1,2,\ldots,M \tag{21}$$

Thus the domain integral is written as

$$\int_\Omega R(\mathbf{x})d\Omega \simeq \sum_{j=1}^{M} \overline{a}_j \int_\Omega f_j(r)d\Omega \tag{22}$$

Subsequently, applying the Green' reciprocal identity [6]

$$\int_\Omega \left(v\nabla^2 u - u\nabla^2 v\right)d\Omega = \int_\Gamma \left(vu_{,n} - uv_{,n}\right)ds \tag{23}$$

for $v = 1$ and $u = \hat{u}_j$ we obtain

$$\int_\Omega f_j d\Omega = \int_\Gamma \hat{q}(r_{j\mathbf{x}})ds \simeq \sum_{k=1}^{N} \int_{\Gamma_k} \hat{q}(r_{j\mathbf{x}})ds\,, \quad \mathbf{x} \in \Gamma_k \tag{24}$$

and the domain integral (22) can be written as

$$\int_\Omega R(\mathbf{x})d\Omega \simeq \sum_{j=1}^{M}\sum_{k=1}^{N} \overline{a}_j \hat{Q}_{jk} = \mathbf{1}^T \hat{\mathbf{Q}}^T \overline{\mathbf{a}} \tag{25}$$

where $\hat{Q}_{jk} = \int_{\Gamma_k} \hat{q}_j(r_{jk})ds$ and $\mathbf{1}^T = \{1\ 1\ \cdots\ 1\}$. Finally, introducing the vector $\overline{\mathbf{a}}$ from Eqn (20) into (25) we obtain

$$\int_\Omega R(\mathbf{x})d\Omega \simeq \mathbf{1}^T \hat{\mathbf{Q}} \mathbf{\Phi}^{-1} \mathbf{R} \tag{26}$$

Special care is taken to avoid possible ill-conditioning of the matrix $\mathbf{\Phi}$ [8]

4.3 Minimization of the functional $J(u)$

The herein employed radial basis functions f_j are the multiquadrics (MQ's), which are defined as

$$f_j = \sqrt{r^2 + c_j^2} \tag{27}$$

where c_j are the shape parameters, in general different for each collocation point. The particular solution of Eqn (10) for f_j given by Eqn (27) is obtained as

$$\hat{u}_j = -\frac{c^3}{3}\ell u(c_j\sqrt{r^2 + c_j^2} + c_j^2) + \frac{1}{9}(r^2 + 4c_j^2)\sqrt{r^2 + c_j^2} \tag{28}$$

It is apparent that the functional $J(u)$ depends on the following sets of parameters.

(i) The M coefficients a_j

(ii) The M shape parameters c_j of MQ's.

(iii) The $2M$ coordinates x_j, y_j of the collocation points.

Therefore, we can search for the minimum using various levels of optimization depending on the design parameters that we wish to be involved in the optimization procedure. Although, the functional $J(u)$ is quadratic with respect to a_j, and their values could be established from the solution of a linear algebraic system [9], the inclusion however of c_j and x_j, y_j requires direct minimization methods for nonlinear objective functions. In this investigation the optimization problem is solved using the successive quadratic programming (SQP) with numerical evaluation of the gradient of the objective functions. The results have been improved using the Courant's device, that is by minimizing the functional

$$J(u) = \int_\Omega \left[\frac{1}{2}\left(A u_{,x}^2 + 2B u_{,x}\, u_{,y} + C u_{,y}^2 - F u^2 \right) + gu + \left[L(u) - g \right]^2 \right] d\Omega \tag{29}$$

where $L(u)$ represents the differential operator of Eqn (2). This formulation minimizes also the residual $\|L(u) - g\|$, which implies also minimization of the error $\|u - u_{exact}\|$ [10].

4 NUMERICAL RESULTS

The thermal distribution in a plane body with inhomogeneous material properties and the irregular shape of Fig. 2 is studied. The thermal conductivity is taken to vary according to the law

$$k(x, y) = (2x + y + 2)^2 \tag{30}$$

The flux $q_n = -k u_{,n}$ hence the normal derivative $u_{,n}$, is prescribed along the sides AB and CD as

$$u_{,n}(x, 0) = -\frac{17x^2 + 20x - 15}{2(1 + x)^2}, \qquad u_{,n}(0.5, y) = \frac{-45 + 46y + 32y^2}{(3 + y)^2} \tag{31a,b}$$

while the temperature u is prescribed on the remaining part of the boundary as

$$u_b(x, y) = \frac{6x^2 - 6y^2 + 20xy + 30}{2x + y + 2} + 100 \tag{31c,d}$$

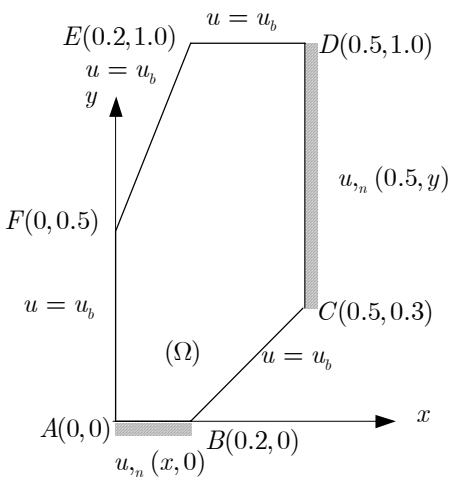


Figure 2: Irregularly shaped plane body.

The temperature $u(x,y)$ will be obtained as a solution of the equation

$$k\left(u_{,xx}+u_{,yy}\right)+k_{,x}\,u_{,x}+k_{,y}\,u_{,y}=0\,,\quad \text{in } \Omega \tag{32}$$

subject to boundary conditions (31a,b,c,d). Apparently, the coefficients of Eqn (32) satisfy the conditions (5a,b) as well as the ellipticity condition. The exact solution is

$$u(x,y)=\frac{6x^2-6y^2+20xy+30}{2x+y+2}+100 \tag{33}$$

x	y	u	q_x	q_y
0.25	0.25	111.365 111.364	40.113 40.500	25.407 25.750
	0.3125	111.151 111.147	36.331 36.687	25.584 27.836
	0.375	110.930 110.924	32.328 32.625	29.842 29.968
	0.4375	110.700 110.694	28.100 28.312	32.131 32.148
	0.500	110.464 110.458	23.628 23.750	34.431 34.375
	0.5625	110.222 110.217	18.889 18.937	36.744 36.648
	0.625	109.975 109.970	13.864 13.875	39.088 38.968
	0.6875	109.722 109.718	8.553 8.562	41.485 41.336
	0.750	109.464 109.462	2.966 3.000	43.941 43.750
	0.8125	109.202 109.200	-2.869 -2.812	46.439 46.210
	0.875	108.936 108.935	-8.928 -8.875	48.934 48.720

Table 1. Results for temperature and derivatives ($N=130$, $M=83$) Upper value: computed; Lower value exact

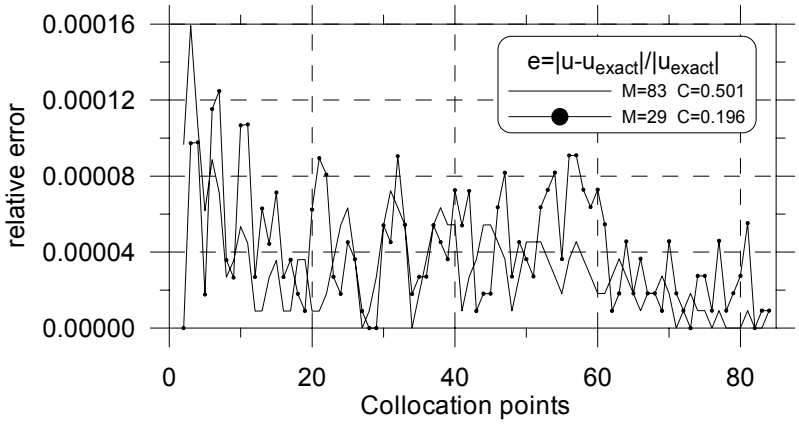


Figure 3. Relative error at collocation points

The obtained results for the temperature and the fluxes $q_x = -kT_{,x}$, $q_y = -kT_{,y}$ are given in Table 1 as compared with the exact ones. Moreover, the relative error for the computed solution at the collocation points is shown in Fig. 3 as obtained for optimal c, and a_j. The results have been obtained using $N = 130$ and $M = 29$ and $M = 83$. It is apparent that a $M = 29$ collocation points yields accurate result. The RMS (Relative Mean Square) error is given in Table 2.

M	RMS	$J(u)$	$J(u)_{exact}$
29	8.139×10^{-6}	56.776	56.962
83	4.724×10^{-6}	56.939	56.962

Table 2. Relative Mean Square error RMS ($N = 130$)

5 CONCLUSIONS

A new variational method, to which we refer as generalized Ritz method, was presented for solving the general second order elliptic PDE on domains of arbitrary shape. The admissible functions are established using a BEM technique based on the concept of the analog equation. The main conclusions that can be drawn are:

(a) A systematic procedure is developed to construct admissible functions for domains of arbitrary shape, which have continuous derivatives on the whole domain

(b) Since the admissible functions satisfy not only the kinematic boundary conditions, as in the conventional Ritz method, but also the natural ones, the boundary integral term in the functional vanishes.

(c). The domain integral of the functional is converted to a boundary line integral. Thus, the method becomes boundary-only and it maintains all the advantages of the pure BEM.

(d) The error is minimized by including the shape parameters of the MQ's and the position of the domain collocation points in the design variable in the minimization procedure. This procedure yields optimal MQ's

(e) Accurate numerical results are obtained not only for the solution but also for its derivatives. The accuracy is highly increased by including the residual of the equation in the functional.

Generally, we can say that the presented method eliminates the drawbacks of the conventional Ritz method and makes it operative and efficient again for solution of physical problems described by 2nd order PDE's.

6 REFERENCES

[1] Ritz, W. (1908), "Ueber eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik, *J. Reine Angew. Math.* 135, pp. 1-61.

[2] Sokolnikoff, I.S. (1956), *Mathematical theory of elasticity*, McGraw-Hill, New York

[3]. Filippov V.M. (2000) "Construction of extremal variational principles in non-Euler functional classes for PDE's of second order with nonpotential operators", *физика элементарных частиц, и атомново ядра* 31 вып 7а 9, pp. 5-96.

[4]. Katsikadelis, J. T. (1994), "The analog equation method. A powerful BEM–based solution technique for solving linear and nonlinear engineering problems", In: C.A. Brebbia (ed.) *Boundary Element Method XVI*, Computational Mechanics Publications pp. 167-182

[5] Katsikadelis, J.T. (2002), "The analog equation method. A boundary-only integral equation method for nonlinear static and dynamic problems in general bodies, *Theoretical and Applied Mechanics*, 27, pp. 13-38.

[6] Kansa, E.J. (2005), "Highly accurate methods for solving elliptic and partial differential equations:, In: Brebbia, C.A., Divo, E. and Poljak, D. (eds.), *Boundary Elements XXVII*, WITPRESS, Southampton, pp. 5-15.

[7]. Katsikadelis, J. T (2002), *Boundary Elements: Theory and Applications*, Elsevier Amsterdam-London

[8] Kansa, E.J. and Hon, Y.C. (2005), "Circumventing the ill-conditioning problem with multiquadric radial basis functions: Applications to elliptic partial differential equations", *Computers and mathematics with Applications* 39, pp. 123-137.

[9] Katsikadelis, J. T (2005), "The BEM for non-homogeneous bodies", *Archive of Applied Mechanics*, (to appear).

[10] Schaback, R. (2005),"Convergence analysis of methods for solving general equations", In: Brebbia, C.A., Divo, E. and Poljak, D. (eds.), *Boundary Elements XXVII*, WITPRESS, Southampton, pp. 17-24.

B SPLINE BOUNDARY ELEMENT METHOD FOR WAVE PROPAGATION IN FLUID MEDIA AND FLUID-STRUCTURE INTERACTION ANALYSIS

Dimitris C. Rizos[1], and Saiying Zhou[2]

[1] A. Professor, Department of Civil and Environmental Engineering, University of South Carolina, Columbia, SC 29208 USA, e-mail: rizos@engr.sc.edu

[2] Graduate Research Assistant Department of Civil and Environmental Engineering, University of South Carolina, Columbia, SC 29208 USA, e-mail: zhou9@engr.sc.edu

Keywords: Scalar Wave Propagation, Boundary Element Method, Fluid-Structure Interaction, B-Spline.

Abstract. *This work presents a direct time domain methodology for Fluid-Structure Interaction analysis between linearized fluids and rigid bodies. To this end, a direct time domain Boundary Element formulation pertaining to 3-D wave propagation in acoustic media is developed. The method is based on higher order B Spline fundamental solutions developed for scalar wave propagation in 3-D infinite media and uses higher order spatial discretization. The method is used within the B Spline Impulse Response framework for the calculation of the time history of the response of the acoustic medium and the calculation of the hydrodynamic forces on the wetted surface of the fluid-structure interface. The proposed B Spline BEM is stable as compared to the relevant Retarded Potential formulations reported in the literature. Although the method is non-local in time, only a relatively small number of time steps are required in the B-Spline Response functions, making the method very efficient especially for prolonged external excitations. Fluid Structure interaction models are proposed for the dynamics of rigid bodies submerged in, or floating on, semi-infinite fluid domains. The applicability of the proposed models are also demonstrated in an application example.*

1 INTRODUCTION

The physical model of a marine vessel traveling over open water is a typical multi-phase problem involving distinct media, such as solids and fluids, and their dynamic interaction. Development of analytical and mathematical models for the representation of such dynamic coupled media draw expertise from such diverse areas as structural and fluid mechanics, wave propagation in fluids and solids, vibrations, hydrodynamics and structural dynamics. Commonly used models consist of three interacting components, i.e., the vessel, the propulsion system, and the surrounding fluid medium, as shown in Figure 1. As the propulsion system puts the vessel into motion, the latter interacts with the surrounding fluid medium that resists the motion, applying, thus, additional demands on the propulsion system. Typically, a number of assumptions related to the behavior of the physical system lead to idealization of the physical problem and yield mathematical models that can be solved through analytical or computational procedures. Analytical solutions are generally restricted to extremely simple geometries and adopt oversimplifying assumptions. Computational procedures, such as the Boundary Element Method (BEM) and the Finite Element Method (FEM), are better suited for the solution of such involved problems. The FEM can be used for the modeling of the coupled fluid-vessel-propulsion media. However, treatment of infinite media requires the development of special FEM procedures[1]. Alternatively, the BEM can be used for the modeling of the fluid media because of its ability to implicitly satisfy the boundary condition at infinity[2], while the FEM models the remaining components of the system. Coupling of the two methods is performed at the interface between the two domains[3-5]. Models for real time ship maneuvering simulation establish a relationship between the ship propulsion, the corresponding hydrodynamic forces and models of marine vessels. Such models integrate the dynamic and kinetic models of the vessel with various hydrodynamic models for the fluid, and are reported in the literature[6-8]. Effective models of real time ship maneuvering simulation require fast and efficient models of the components of the system. Kinetic and dynamic models of the vessel reported in the literature meet these criteria, in general. However, FEM or BEM dynamic

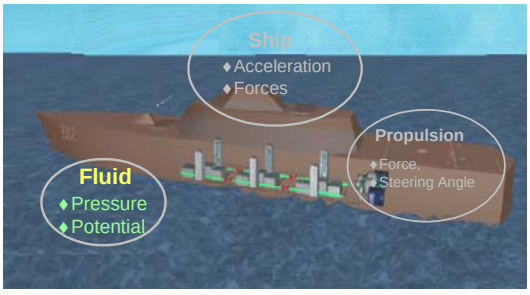


Figure 1 Typical Interacting Components of a Marine Vessel Model

models for the fluid typically require solution of large system of equations, limiting, thus, the ability to simulate the vessel motion in real time.

The objective of the present work addresses the development of an efficient model for ship maneuvering simulation in real time. The proposed model consists of a rigid vessel of arbitrary geometry, and the linearized fluid while the propulsion force is assumed as an external force to the system. The propulsion force is considered as the net driving propeller force of the system required to accelerate the vessel. This implies that hydrodynamic forces that develop due to constant velocity of the vessel are not accounted for; however, it is the focus of current research. A direct time domain B-Spline BEM is proposed for the modeling of fluid media. The method is based on higher order B-Spline fundamental solutions developed for scalar wave propagation in 3-D infinite media. It is used within the B Spline Impulse Response framework [9,10] for the calculation of: (i) the time history of the response of the surface of the acoustic medium, (ii) the calculation of the hydrodynamic forces on the wetted surface of the vessel, and (iii) the development of a simplified model of a rigid vessel that calculates the hydrodynamic forces and moments at a reference point of the rigid vessel, due to an arbitrary applied acceleration.

The following sections describe briefly the proposed formulations and models and present a demonstration example.

2 BEM FORMULATION

2.1 Governing Equation and Integral Representation

The differential governing equation for wave propagation through an inviscid, compressible acoustic medium of volume V is expressed as,

$$\nabla^2 p(\mathbf{x},t) - \frac{1}{c^2}\ddot{p}(\mathbf{x},t) + f(\mathbf{x},t) = 0 \tag{1}$$

where, p is the pressure, $\mathbf{x}$ is the position vector, t represents time, f is the body source, ∇^2 is the Laplacian operator, and c is the wave propagation velocity in the medium. In view of the second derivative appearing in the governing Eq. (1), $f(\mathbf{x},t)$ has to be at least twice continuously differentiable and vanishing for $t<0$. In the proposed formulation, the body source is assumed to be a point source applied at $\mathbf{x}$ with time varying intensity of a 4th order B Spline polynomial, i.e.,

$$f(\mathbf{x},t) = \delta(\mathbf{y}-\mathbf{x})B_o^4(t) \tag{2}$$

B Spline functions are piecewise smooth polynomials of degree k-1 with k-2 continuous derivatives and are defined based on a knot sequence $t_n = n\Delta t/k, \quad n = 0,1,2,\ldots,$. They are expressed in a compact form in the normalized interval $[0,1]$ as,

$$B_n^k(t) = B_o^k(\tau) = \begin{cases} \sum_{m=0}^{k-1} a_{l,m}\tau^m & for\ \tau \in [0,1] \\ 0 & elsewhere \end{cases} \qquad l = 1,2,\ldots,k \tag{3}$$

where $\tau = (t - t_n)/\Delta t$ represents the normalized time, and $\alpha_{l,m}$ are the polynomial coefficients associated to the term τ^m of polynomial l that defines the B-Spline[11]. B-Spline polynomials have the properties

$$\sum B_n^k(t) = 1 \quad \forall t, \qquad B_n^k(t) = 0 \quad t \notin [t_n, t_{n+k}] \tag{4}$$

2.2 Integral Representation and the B-Spline Fundamental Solutions

Following well-established procedures[2], the governing equation can be expressed in a Boundary Integral Equation (BIE) form as,

$$c(\xi)p(\xi,t) = \int_S \left[G(\mathbf{x},t,\xi|q(\mathbf{x},t)) - T(\mathbf{x},t,\xi|p(\mathbf{x},t))\right]dS \tag{5}$$

where $\mathbf{x}$ and ξ are points in the domain or its boundary, the integration is over the boundary, S, of the

domain, and *G* and *T* are the associated fundamental solutions. The field variables $p(\mathbf{x},t)$ and $q(\mathbf{x},t)=\dfrac{\partial p(\mathbf{x},t)}{\partial \mathbf{n}}$ correspond to the pressure and its derivative in the direction of the outward normal, **n**, to the boundary, respectively. The term $c(\boldsymbol{\xi})$ is known as the "jump term" and depends on the location of point $\boldsymbol{\xi}$ and the smoothness of the boundary in its neighbourhood. The fundamental solutions *G* and *T* developed in this work are associated to the application of a point source at point **x** (source point) of B-Spline modulation, Eq. (2). They satisfy the continuity requirements of the governing equation and are expressed as,

$$G^B=\frac{1}{4\pi r}B_0^4\left(\frac{t-\frac{r}{c}}{\Delta t}\right) \tag{6a}$$

$$T^B=\frac{-1}{4\pi r^2}\left\{B_0^k\left(\frac{t-\frac{r}{c}}{\Delta t}\right)+\frac{r}{c}\dot{B}_0^k\left(\frac{t-\frac{r}{c}}{\Delta t}\right)\right\}\frac{\partial r}{\partial \mathbf{n}} \tag{6b}$$

where G^B and T^B indicate the B-Spline fundamental solutions, $r=|\mathbf{x}-\boldsymbol{\xi}|$ represents the distance between the source point, **x**, and a field point, $\boldsymbol{\xi}$ over the 3-D solution domain. It can be shown that the fundamental solutions are singular only when the source point coincides with the receiver point at times $0<t<\Delta T$.

2.3 BEM and the B-Spline Impulse Response Function (BIRF)

Following the procedures established in [11] Eq. [5] can be expressed in a discrete form in space and time on the boundary of the domain, at time instant, t_N, as

$$\frac{1}{2}\mathbf{p}^N=\sum_{n=1}^{N+1}\left[\mathbf{G}^N\mathbf{q}^{N-n+2}-\mathbf{T}^N\mathbf{p}^{N-n+2}\right] \tag{7}$$

where superscripts indicate the time step at which quantities are evaluated at. The discrete form of Eq (4) implies that:

(i) The surface S of the acoustic medium is discretized in a total of NE surface elements for a total of NN boundary nodes. The present work implements higher order 3-D surface elements in an isoparametric formulation.

(ii) The temporal discretization of the field variables is based on B-Spline function approximation schemes of the same order as the B-Spline fundamental solutions. Eq (7) is evaluated at time knot t_N and the associated time knot sequence is

$$t_n=n\frac{\Delta t}{4}\qquad n=0,1,2,\cdots,N+2 \tag{8}$$

(iii) B-Spline fundamental solutions of the 4th order, Eq. (6), are developed and adopted.

(iv) The coefficient matrices $\mathbf{G}^N$ and $\mathbf{T}^N$ are of size *NN* x *NN* and represent the influence of a source boundary node on a receiver boundary node. These matrices are evaluated based on integrals of the B-Spline fundamental solutions over the surface of each element for all boundary nodes. Standard Gaussian quadrature is used for the non-singular integration of the kernels. Singularities are also integrated numerically using triangle coordinates and a successive mapping technique. In both cases, integrations are performed based on a subelement division approach[11]. Vectors **p** and **q** are of size NN x 1 and represent the value of the pressure and its normal derivative, respectively, at every boundary node due to a single B-Spline impulse excitation applied at a time step indicated by the superscript.

Equations (7) indicate that a solution can be obtained for the B-Spline Impulse Response of the system. To this end, a transient B-Spline impulse excitation, $\bar{\mathbf{q}}_j$, of duration Δt is applied in the direction of each degree of freedom, j, i.e.,

$$\bar{\mathbf{q}}_j=\mathbf{d}B_o^k(t),\quad i,j,=1,\ldots,NN \tag{9}$$

where **d** is defined as $d_i=\delta_{ij}$ with δ_{ij} being the Kronecker delta. Eqs (7) are solved in a time marching

scheme for the nodal pressure, $\mathbf{p}\Big|_{\bar{\mathbf{q}}_j} = \mathbf{b}_j^N$, as

$$\mathbf{b}_j^N = \left(\frac{1}{2}\mathbf{I} + 2\mathbf{T}^1 + \mathbf{T}^2\right)^{-1}\left\{\mathbf{g}_j^N - \sum_{n=3}^{N+1}\mathbf{T}^n\mathbf{b}_j^{N-n+2} + \mathbf{T}^1\mathbf{b}_j^{N-1}\right\} \tag{10}$$

$\mathbf{b}_j^N$ represents the B-Spline Impulse Response Function (BIRF) at time step t_N of all nodes of the system when node j is excited and $\mathbf{g}_j^N$ is the j^{th} column of $\mathbf{G}^N$ matrix. These impulse responses can be collected in matrix form as

$$\mathbf{B}(t_N) = \mathbf{B}^N = \left[\mathbf{b}_1^N, \mathbf{b}_2^N, \ldots, \mathbf{b}_j^N, \ldots \mathbf{b}_{NN}^N\right] \tag{11}$$

The dimension of matrix $B(t_N)$ is NN x NN , in general. This matrix is a characteristic of the system and needs to be computed only once for the specific geometry of the free surface of the acoustic domain. In the general case, the B-Spline impulse response is computed for all degrees of freedom of the problem and the impulse response matrix, B^N, is square. Once this characteristic response of the system is known, the system can be analyzed for any arbitrary transient excitation, f(t), as discussed in the following section.

2.4 Response to Arbitrary Excitation

Provided that the BIRF of the system is known, the transient scattered pressure field, p, at step N due to an arbitrary incident flux, q, can be computed as a mere superposition of the BIRFs as

$$\mathbf{p}^N = \sum_{n=1}^{N+1}\mathbf{B}^n\mathbf{q}^{N-n+2} = \mathbf{F}\mathbf{q}^N + \mathbf{H}^N \tag{12}$$

where

$$\mathbf{F} = 2\mathbf{B}^1 + \mathbf{B}^2 \tag{13}$$

$$\mathbf{H}^N = \sum_{n=3}^{N+1}\mathbf{B}^n\mathbf{q}^{N-n+2} - \mathbf{B}^1\mathbf{q}^{N-1} \tag{14}$$

and it has been assumed that

$$\ddot{\mathbf{q}}^N = 0 \Rightarrow \mathbf{q}^{N+1} \cong 2\mathbf{q}^N - \mathbf{q}^{N-1} \tag{15}$$

The proposed approach is very efficient especially when multiple load cases of prolonged duration are considered, since the BIRF functions are independent of the external excitation and are typically of much shorter duration than the external excitation. The proposed method has been validated and has shown superior accuracy and stability[12,13]. Mixed boundary value problems can also be addressed using the proposed methodology. However, it is beyond the scope of the present work and is currently work in progress.

3 BIRF OF RIGID BODIES AND THE KINEMATIC INTERACTION MODEL

The motion of a rigid body is represented by a reference point (RP) that has 3 translational and 3 rotational degrees of freedom, as shown in Figure 2. In order to compute the BIRF of the rigid body it is assumed that the BIRF functions of its wetted surface, Eq. (11) have been calculated. Subsequently, unit acceleration, $\hat{\mathbf{a}}_{RP} = \delta_{ij}B_0^4(t)$, $i,j = 1,2\cdots,6$ with B-Spline time modulation is applied at each of the six degrees of freedom, j, one at a time. The corresponding acceleration normal to a wetted surface node, k, can be computed based on the rigid body conditions, T_k[14], as,

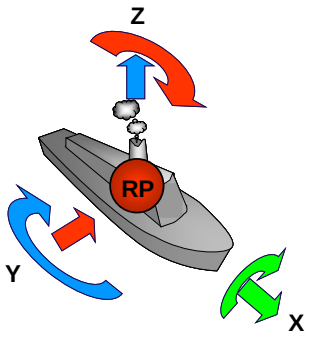


Figure 2 Definition of Rigid Body

$$\hat{a}_{w_k\mathbf{n}} = \left(\mathbf{T}_k\hat{\mathbf{a}}_{CG}\right)\cdot\mathbf{e}_{k\mathbf{n}} \tag{16}$$

where, $\mathbf{e}_{k\mathbf{n}}$ is the unit outward vector normal to the wetted surface. The corresponding hydrodynamic forces, $\hat{\mathbf{F}}_{w\mathbf{n}}^N$, normal to the wetted surface at time step N can be computed as

$$\hat{\mathbf{F}}_{w\mathbf{n}}^{N} = -\rho \mathbf{L} \sum_{n=1}^{N+1} \mathbf{B}^{N-n+2} \hat{\mathbf{a}}_{w\mathbf{n}}^{n} \tag{17}$$

where **L** is a pressure-to-force transformation matrix[14], ρ is the mass density of the fluid medium, and **B** is the BIRF of the wetted surface. Subsequently, the vector $\mathbf{b}_{RP}^{N}$ of forces and moments at the reference point of the vessel due to $\hat{\mathbf{a}}_{RP}$ can be computed as the resultant of all $\hat{\mathbf{F}}_{w\mathbf{n}}^{N}$ vectors and their corresponding moments

$$\mathbf{b}_{RP}^{N} = \left[\sum_{k} \hat{\mathbf{F}}_{w\mathbf{n}}^{N} \quad \sum_{k} \mathbf{r}_k \times \hat{\mathbf{F}}_{w\mathbf{n}}^{N} \right] \tag{18}$$

Where $\mathbf{r}_k$ is the position vector of node k with respect to RP and the summation indicates vectorial summation over all boundary nodes. This response represents column j in the BIRF matrix, $\mathbf{B}_{RP}^{n}, n = 1,2,\cdots N$, of the rigid body that corresponds to the degree of freedom, j, excited by the unit acceleration. It is evident that the BIRF matrix of the rigid body is of size 6x6 and accounts for kinematic fluid-structure interaction effects.

A kinematic hydrodynamic interaction model can de derived based on the BIRF functions, $\mathbf{B}_{RP}^{n}, n = 1,2,\cdots N$, of the rigid body and are obtained in view of Eq. (12). This model relates the arbitrary driving acceleration, $\mathbf{a}_{RP}$, applied at the RP, to the resultant of the hydrodynamic forces and moments, $\mathbf{F}_{RP}$, applied at CG as,

$$\mathbf{F}_{RP}^{N} = \sum_{n=1}^{N+1} \mathbf{B}_{RP}^{N-n+2} \mathbf{a}_{RP}^{n} \tag{19}$$

4 DYNAMIC MODEL FOR SHIP MANEUVERING

The proposed dynamic model is a 3-D hydrodynamic fluid-structure interaction mathematical model that accounts for both inertia and kinematic interaction effects between a rigid ship and the surrounding water. The model is based on the presented methodologies and the characteristic BIRF of rigid bodies floating on, or submerged in, fluid media. The dynamic model is expressed in general as,

$$\mathbf{F}_{net} = \mathbf{m}\mathbf{a}_{RP} \tag{20}$$

where **m** is the mass matrix in a FEM sense associated to the rigid body, and $\mathbf{a}_{RP}$ is the acceleration vector applied at the RP of the rigid body. The force vector $\mathbf{F}_{net}$ represents the net forces applied at the RP of the vessel and accounts for generalized propulsion forces and moments, $\mathbf{F}_{prop}$, and hydrodynamic effects, $\mathbf{F}_{hyd}$, and is expressed as,

$$\mathbf{F}_{net} = \mathbf{F}_{prop} - \mathbf{F}_{hyd} \tag{21}$$

The driving force of the system is considered as the generalized propulsion force and is known either explicitly, or through existing propulsion models. It should be noted that the propulsion forces are considered as the net driving propeller force of the system required to accelerate the vessel. This implies that hydrodynamic forces that develop due to constant velocity of the vessel are not accounted for. However, both the hydrodynamic forces, $\mathbf{F}_{hyd}$ and the acceleration of the vessel, $\mathbf{a}_{RP}$, are not known apriori. Consequently, Equation (20) cannot be solved for the vessel accelerations. In view of Equation (19), and provided that the BIRF of the rigid body is already computed, the following systems of Equations need to be solved:

$$\mathbf{F}_{prop} - \mathbf{F}_{hyd} = \mathbf{m}\mathbf{a}_{RP} \tag{22a}$$

$$\mathbf{F}_{hyd}^{N} = \sum_{n=1}^{N+1} \mathbf{B}_{RP}^{n} \mathbf{a}_{RP}^{N-n+2} \tag{22b}$$

The solution is obtained in a time marching scheme, implying that the time axis is discretized in a series of time steps i=1,2,…,*N*,…,*M*, where *M* is the total number of time steps that the solution is sought after. In order to solve Equations (22a and b) both need to be written for time step *N*. By separating known from unknown acceleration terms in Equation (22b), the latter is written as,

$$\mathbf{F}_{hyd}^{N} = \mathbf{B}_{RP}^{1}\mathbf{a}_{RP}^{N+1} + \mathbf{B}_{RP}^{2}\mathbf{a}_{RP}^{N} + \sum_{n=3}^{N+1}\mathbf{B}_{RP}^{n}\mathbf{a}_{RP}^{N-n+2} \tag{23}$$

where $\mathbf{a}_{RP}^{N}$ and $\mathbf{a}_{RP}^{N+1}$ are the unknown accelerations at steps *N*, and *N+1*, respectively. Following an assumption similar to Equation (15), the acceleration at the forward step is related to the acceleration at previous steps as,

$$\mathbf{a}_{RP}^{N+1} = 2\mathbf{a}_{RP}^{N} + \mathbf{a}_{RP}^{N-1} \tag{24}$$

Introducing Equation (24) into (23) and factoring the unknown acceleration $\mathbf{a}_{RP}^{N}$, the latter is expressed as,

$$\mathbf{F}_{hyd}^{N} = \left(2\mathbf{B}_{RP}^{1} + \mathbf{B}_{RP}^{2}\right)\mathbf{a}_{RP}^{N} + \sum_{n=3}^{N+1}\mathbf{B}_{RP}^{n}\mathbf{a}_{RP}^{N-n+2} - \mathbf{B}_{RP}^{1}\mathbf{a}_{RP}^{N-1} = \mathbf{m}_{hyd}\mathbf{a}_{RP}^{N} + \mathbf{HF}_{hyd}^{N} \tag{25}$$

$\mathbf{m}_{hyd} = \left(2\mathbf{B}_{RP}^{1} + \mathbf{B}_{RP}^{2}\right)$ can be viewed as an equivalent hydrodynamic mass matrix and vector $\mathbf{HF}_{hyd}^{N} = \sum_{n=3}^{N+1}\mathbf{B}_{RP}^{n}\mathbf{a}_{RP}^{N-n+2} - \mathbf{B}_{RP}^{1}\mathbf{a}_{RP}^{N-1}$ is an equivalent force vector representing the effects of the history of the hydrodynamic forces on the current step *N* and is always known. Substituting Equation (25) into Equation (22a), the latter is expressed as,

$$\mathbf{F}_{prop}^{N} = \mathbf{m}\mathbf{a}_{RP}^{N} + \mathbf{m}_{hyd}\mathbf{a}_{RP}^{N} + \mathbf{HF}_{hyd}^{N} \tag{26}$$

Therefore Equations (22a) and (32b) are combined into a single system of Equations as,

$$\widetilde{\mathbf{F}}^{N} = \widetilde{\mathbf{m}}\mathbf{a}_{RP}^{N} \tag{27}$$

where $\widetilde{\mathbf{F}}^{N} = \mathbf{F}_{prop}^{N} - \mathbf{HF}_{hyd}^{N}$ and $\widetilde{\mathbf{m}} = \mathbf{m} + \mathbf{m}_{hyd}$. The system of Equations (27) can be solved for the acceleration of the vessel in a time marching scheme using conventional solvers.

If the coefficient matrix, $\hat{\mathbf{m}}$, is diagonal, indicating that the degrees of freedom are uncoupled, the proposed approach becomes very efficient. For example, such a case is met when only the surge, sway and yaw modes are considered for a vessel with two vertical planes of symmetry and uniformly distributed mass lumped at the vessel's center of gravity. However, if a coupled system is considered the system of equations (27) is still of a small size (6 degrees of freedom) the solution of which can be performed in an efficient manner. The direct coupling of Equations (32a) and (32b), as proposed herein, implies that the time step of the time marching procedure for solving Equation (37) should be the one used for the calculation of the BIRF function of the rigid vessel. The effects of this restriction are currently under investigation and beyond the scope of the presented work.

5 DEMONSTRATION EXAMPLE

The wetted surface of a marine vessel is idealized for the purposes of this example as a rigid rectangular box of length *L*=500 ft, width (beam) *B*=50 ft and depth (draft) D=50 ft. The mass of the vessel is m=10x10^6 lb and it is assumed lumped at a reference point, RP, located at the free surface level at the intersection of the two vertical planes of symmetry of the vessel. As shown in Figure 3. The wave velocity of the water is taken as 4,862 ft/s. The water free surface and the wetted surface of the vessel are discretized into 8-node quadratic quadrilateral elements as shown in Figure 3a. First, the BIRF of the wetted surface, equation (11) is calculated. To this end, a B-Spline impulse normal acceleration of duration Δt=0.01 s is applied at each wetted surface node and the response of the system is calculated as indicated in Equation (10). Subsequently, the BIRF of the reference point, $\mathbf{B}_{RP}$, pertaining to hydrodynamic forces and moments of the rigid vessel is computed as described in section 3. Figure 3 shows the BIRF of the hydrodynamic forces in the longitudinal direction applied at the RP. The time history of the pressure distribution on the water free surface and the wetted surface of the vessel due to the this excitation are also computed based on Equation (12), where the arbitrary acceleration normal to the wetted surface nodes is computed by Equation (16). Figure 4 shows the pressure distribution plotted normal to the surface of the domain at selected time steps. The propagation and attenuation

of the pressure waves emanated at the bow and stern of the vessel is evident.

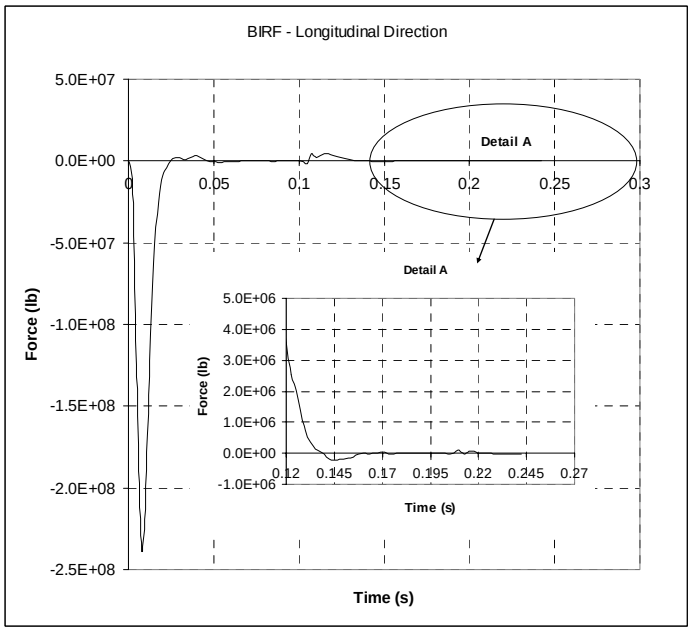


Figure 3 BIRF of Hydrodynamic Forces in the Longitudinal Direction due to Impulse Acceleration Applied at Reference Point in the Longitudinal Direction

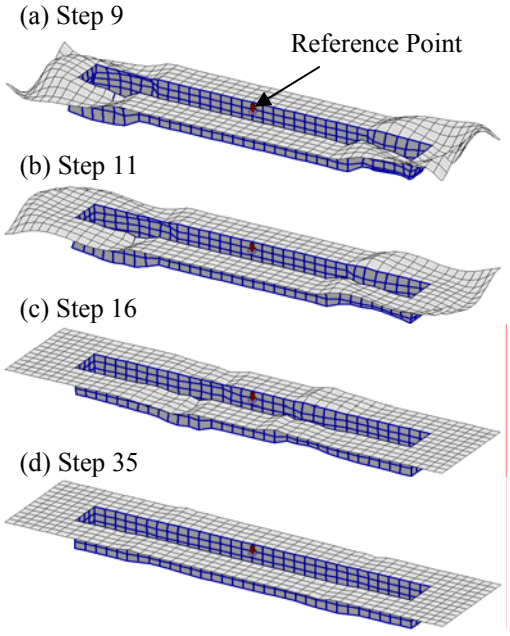


Figure 4 BIRF of Boundary due to Impulse Acceleration Applied at Reference Point in Longitudinal Direction

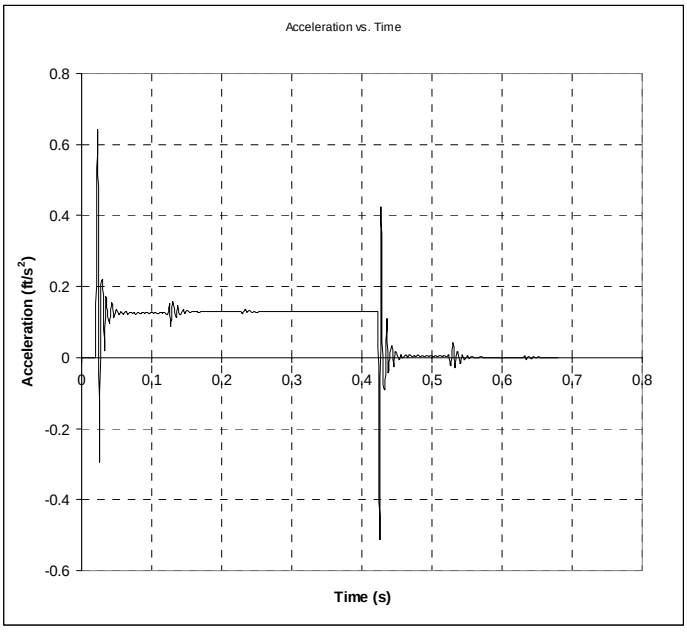


Figure 5 Vessel Acceleration due to a Transient Propulsion Force

Subsequently, the effects of propulsion forces on the acceleration of the vessel are computed based on the procedures of section 4. A step impulse propulsion force of duration Δt_p= 0.4 s, and amplitude F_{po}= 100x10^6 lb is applied at the RP. The force is applied at time t=0.0225 s in order to demonstrate that the condition of quiescent past is satisfied. Figure 5 shows the time history of the acceleration of the vessel due to the application of this load. The transient effects are evident. Due to the small size of the proposed simplified vessel model, computation of the system response is performed at faster than real time. Such models are ideally suited for real time simulation platforms like the VTB simulation software for analysis and virtual testing of integrated systems.

6 CONCLUSIONS

This work presented a direct time domain BEM for 3D scalar wave propagation. The method is based on

B-Spline fundamental solutions for the infinite space. The concept of B-Spline Impulse response functions is introduced for transient Fluid-Structure Interaction analysis. Simplified models for vessel maneuvering are introduced. The proposed methods and models are accurate, efficient and stable. The application example demonstrates the applicability of the method in real time simulation models of ship maneuvering.

7 ACKNOWLEDGEMENTS

This work is sponsored by the VTB program at the University of South Carolina and the U.S. Navy.

REFERENCES

[1] Rizos D.C., Karabalis D.L. (1999), "Soil-Fluid-Structure Interaction", Chapter 9 in: E. Kausel and G.D. Manolis (eds) *Wave Motion Problems in Earthquake Engineering,* WIT Press Southampton.

[2] Bonnet M. (1995), *Boundary Integral Equation Methods for Solids and Fluids*, John Wiley & Sons, New York.

[3] Gaul L., Wenzel W. (2002), "A coupled symmetric BE–FE method for acoustic fluid–structure interaction," Engineering Analysis with Boundary Elements, 26 (7), pp. 629-636.

[4] Czygan O., von Estorff O. (2002), "Fluid-structure interaction by coupling BEM and nonlinear FEM," Engineering Analysis with Boundary Elements, 26 (9), pp. 773-779.

[5] Liu X.G., Sakai S. (1997), "BE-FE combined method for analysis on the dynamic response of large-scale floating structure to random waves", *Proceedings of the16th International Conference on Offshore Mechanics and Arctic Engineering*, Yokohama, Japan, Vol. 6, pp. 203-209.

[6] Sutulo S., Moreira L., Guedes S.C. (2002), "Mathematical Models for Ship Path Prediction in Maneuvering Simulation System", *Ocean Engineering*, 29 (1), pp. 1-19.

[7] Wu J.S., Sheu J. (1996), "An exact solution for a simplified model of the heave and pitch motions of a ship hull due to a moving load and a comparison with some experimental results", *Journal of Sound and Vibration*, 192 (2), pp. 495-520.

[8] Fossen T.I. (1994). *Guidance and Control of Ocean Vehicle*. John Wiley & Sons, Chichester.

[9] Rizos D.C., Loya K.G., (2002), "Dynamic and seismic analysis of foundations based on free field B-Spline characteristic response histories", *Journal of Engineering Mechanics*, 128, pp, 438-448.

[10] Rizos, D.C., Wang, J. (2002) "Coupled BEM-FEM Solutions for Direct Time Domain Soil Structure Interaction Analysis", *Engineering Analysis with Boundary Elements*, 26 (10) pp. 877-888.

[11] Rizos, D.C., Karabalis, D.L. (1998), "A time domain BEM for 3-D elastodynamic analysis using the B-Spline fundamental solutions", *Computational Mechanics*, 15, pp. 249-269.

[12] Rizos, D.C., Zhou, S. (2005) "An advanced direct time domain BEM for 3-D wave propagation in acoustic Media", *Journal of Sound and Vibration* (under review)

[13] Zhou, S. Rizos, D.C. (2005) "An sdvanced BEM for 3D scalar wave propagation and fluid-structure interaction analysis", in: BEM/MRM 27, Eds. A. Kassab, C.A. Brebbia, D. Poljak, Orlando Florida, pp 395-404

[14] Rizos, D.C. (2000) "A Rigid Surface Boundary Element for 3-D Soil-Structure Interaction Analysis in the Direct Time Domain", *Computational Mechanics*, 26 (6) pp. 582-591.

THE SINGULAR FUNCTION BOUNDARY INTEGRAL METHOD FOR A TWO-DIMENSIONAL FRACTURE PROBLEM

M. Elliotis, G. Georgiou, and C. Xenophontos

Department of Mathematics and Statistics
University of Cyprus
P.O.BOX 20537, 1678 Nicosia, Cyprus
e-mail: eliotis@ucy.ac.cy , georgios@ucy.ac.cy , xenophontos@ucy.ac.cy
web pages: http://www.ucy.ac.cy/~georgios , http://www.ucy.ac.cy/~xenophon

Keywords: Biharmonic equation, Boundary singularity, Stress intensity factors, Boundary integral method, Lagrange multipliers.

Abstract. *The Singular Function Boundary Integral Method (SFBIM) originally developed for Laplacian problems with boundary singularities is extended for solving two-dimensional fracture problems formulated as a biharmonic problem in terms of the Airy stress function. Our goal is the direct computation of the associated stress intensity factors, which appear as coefficients in the asymptotic expansion of the solution near the crack tip. In the SFBIM the leading terms of the asymptotic expansion are used to approximate the solution and to weight the governing biharmonic equation in the Galerkin sense. The discretized equations are reduced to boundary integrals by means of Green's theorem and the Dirichlet boundary conditions are weakly enforced by means of Lagrange multipliers. The numerical results on a model problem show that the method converges extremely fast and yields accurate estimates of the leading stress intensity factors.*

1 INTRODUCTION

The elastic field near the tip of a fracture in an elastic body is characterized by the *stress intensity factors* (SIFs). These are the coefficients, α_j, that appear in the asymptotic expansion of the Airy stress function u near the crack tip, which is of the general form

$$u(r,\theta) = \sum_{j=1}^{\infty} \alpha_j r^{\beta_j} f_j(\theta)\,, \tag{1}$$

where (r, θ) denote polar coordinates centered at the crack tip. The eigenvalues $\beta_j \in \mathbb{R}$ and the corresponding eigenfunctions $f_j(\theta)$ are known, whereas the SIFs are unknown, with the values depending on the global problem. The first SIF, α_1, plays a crucial role in the mathematical description of fracture, since

$$K = -\sqrt{2\pi}\alpha_1 \tag{2}$$

is the *opening mode* SIF [1].

In the last few decades there has been a plethora of work aimed at reliably computing the SIFs. The methods used include the finite element method (FEM) with post-processing [2–6], the FEM with local mesh refinement [7], enriched and generalized finite elements [8, 9], the method of fundamental solutions [10], as well as certain versions of the Trefftz method [11–14]. It should be noted that in most of the methods mentioned above, the SIFs are calculated as a post-solution operation, i.e. the solution u is approximated first and the SIFs are then calculated using the approximation to u. If the calculation of the SIFs is the main goal of the computation, then it may be beneficial to use a method in which the SIFs are calculated directly. The method of fundamental solutions, the Trefftz method and the SFBIM presented in this article fall in this category of "direct" methods.

The objective of the present work is to extend the SFBIM to two-dimensional fracture problems. The SFBIM was originally developed in [15] to solve Laplacian problems with boundary singularities aiming at resolving the convergence difficulties encountered with standard numerical methods in the vicinity of singular points. In this method the solution is approximated by the leading terms of the local asymptotic solution expansion, which are also employed to weight the governing equation in the Galerkin sense. Furthermore, the discretized equations are reduced to boundary integrals by means of the divergence theorem, and the Dirichlet boundary conditions are weakly enforced by means of Lagrange multipliers. In addition to reducing the dimension of the problem by one, another important feature of the method is that the singular coefficients α_j are calculated directly (i.e. no post-processing is required) together with the discrete Lagrange multipliers. The

SFBIM has been applied to various problems with singularities, such as the Motz problem [16], the cracked-beam problem [17], and to Laplacian problems over L-shaped domains [18, 19], exhibiting fast convergence and yielding very accurate results, especially for the leading singular coefficients. Since it yields direct estimates of the SIFs, the SFBIM appears to be an excellent candidate for solving fracture problems, which can be expressed as a biharmonic equation in terms of the Airy stress function. To illustrate the extension of the method to such problems, we have chosen a two-dimensional fracture problem, originally studied by Schiff et al. [7].

2 THE MODEL PROBLEM AND THE ASYMPOTIC SOLUTION

We consider here the model problem studied by Schiff et al. [7], which deals with a two-dimensional solid elastic plate containing a single edge crack, subjected to a uniform inplane load normal to the two edges parallel to the crack, while the remaining edges are stress free. Using symmetry, the problem is formulated on $\Omega = (-1, 1)\times (0, 1)$ as a biharmonic equation for of the Airy stress function $u(x, y)$ and is depicted graphically in Figure 1. For simplicity the load in the original problem from [7] has been taken to be 1.

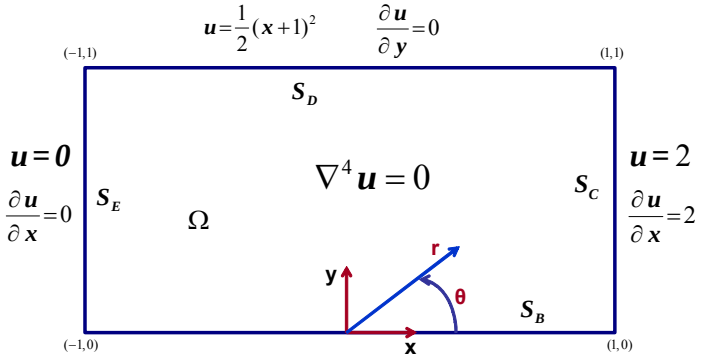


Figure 1. The model fracture problem.

The asymptotic expansion for u in the neighborhood of the singular point (0, 0) can be expressed in terms of an eigenfunction expansion of the form

$$u(r,\theta) = \sum_{j=1}^{\infty} \left[c_j W_1^j(r,\theta) + d_j W_2^j(r,\theta) \right], \tag{3}$$

where (r, θ) are the polar coordinates centered at (0, 0) and c_j, d_j correspond to the even and odd SIFs, respectively. Using this notation, we have $\alpha_1 = d_1$ in (2). In expansion (3) the two sets of the so-called singular functions $W_k^j, k = 1, 2$ are given by

$$W_k^j \equiv r^{\mu_j+1} f_k(\theta, \mu_j), \quad k = 1, 2, \tag{4}$$

where

$$f_1(\theta,\mu_j) = \cos(\mu_j - 1)\theta - \cos(\mu_j + 1)\theta, \quad \mu_j = j, \quad j = 1, 2, \ldots \tag{5}$$

and

$$f_2(\theta,\mu_j) = \cos(\mu_j - 1)\theta - \frac{\mu_j - 1}{\mu_j + 1}\cos(\mu_j + 1)\theta, \quad \mu_j = j - \frac{1}{2}, \quad j = 1, 2, \ldots \tag{6}$$

We note that the singular functions $W_k^j, k = 1, 2$ satisfy the PDE as well as the boundary conditions on S_A and S_B.

3 THE SINGULAR FUNCTION BOUNDARY INTEGRAL METHOD

In the SFBIM the solution u is approximated by the leading terms of the asymptotic expansion. By employing the first N_α terms in (3) the approximate solution $\overline{u}$, is given by

$$\overline{u} = \sum_{i=1}^{N_\alpha} \overline{c}_i W_1^i + \sum_{i=1}^{N_\alpha} \overline{d}_i W_2^i, \tag{7}$$

where $\overline{c}_i$ and $\overline{d}_i$ are the approximations to the SIFs. Obviously, the total number of singular functions involved in the approximation is $2N_\alpha$. It should be pointed out that the method is restricted to fracture problems with only

one crack for which the asymptotic solution is available. Moreover, the proposed approximation (7) is valid only if the domain of the problem is a subset of the domain of convergence of the asymptotic solution. Otherwise, the
domain may be partitioned into subdomains over which separate approximations obeying appropriate compatibility conditions along the interfaces could be used.

By applying Galerkin's principle, the governing equation is weighted by the singular functions, which yields the following set of discretized equations:

$$\int_\Omega \nabla^4 \overline{u} W_k^i dV = 0 \, , \, i = 1,2,...,N_\alpha \, , \quad k = 1,2. \tag{8}$$

Next, applying Green's theorem twice and taking into account that the singular functions satisfy the governing biharmonic equation, the above integrals are reduced to boundary ones:

$$\int_{\partial\Omega} \left(\frac{\partial \overline{u}}{\partial n} \nabla^2 W_k^i - \overline{u} \frac{\partial \left(\nabla^2 W_k^i \right)}{\partial n} \right) dS + \int_{\partial\Omega} \left(\frac{\partial \left(\nabla^2 \overline{u} \right)}{\partial n} W_k^i - \nabla^2 \overline{u} \frac{\partial \left(W_k^i \right)}{\partial n} \right) dS = 0 \, , \quad i = 1,2,...,N_\alpha \, , k = 1,2. \tag{9}$$

The Dirichlet boundary conditions are imposed by means of Lagrange multipliers. In the case of Laplacian problems, the Lagrange multipliers replace the normal derivative $\partial \overline{u} / \partial n$. In the case of biharmonic problems, another option is for the Lagrange multipliers to replace $\partial \left(\nabla^2 \overline{u} \right) / \partial n$. In the current problem, Dirichlet boundary conditions appear along the three boundary parts of interest, i.e. S_C , S_D and S_E , where the normal derivative of the solution is also specified. Therefore, Lagrange multipliers have been chosen to replace $\partial \left(\nabla^2 \overline{u} \right) / \partial n$ at boundary parts S_C, S_D and S_E. These are partitioned into three-node elements and the corresponding Lagrange multipliers, denoted respectively by λ_C , λ_D and λ_E , are expanded in terms of quadratic basis functions M^j :

$$\lambda_C = \frac{\partial \left(\nabla^2 \overline{u} \right)}{\partial x} = \sum_{j=1}^{N_{\lambda_C}} \lambda_C^j M^j \quad \text{on } S_C \, , \tag{10}$$

$$\lambda_D = \frac{\partial \left(\nabla^2 \overline{u} \right)}{\partial y} = \sum_{j=1}^{N_{\lambda_D}} \lambda_D^j M^j \quad \text{on } S_D \tag{11}$$

and

$$\lambda_E = \frac{\partial \left(\nabla^2 \overline{u} \right)}{\partial x} = \sum_{j=1}^{N_{\lambda_E}} \lambda_E^j M^j \quad \text{on } S_E \, , \tag{12}$$

where $N_{\lambda_C}, N_{\lambda_D}$ and N_{λ_E} are the numbers of the discrete Lagrange multipliers λ_C^j, λ_D^j and λ_E^j along the corresponding boundaries. The discrete Lagrange multipliers appear as additional unknowns in the problem. The required $N_{\lambda_C}, N_{\lambda_D}$, N_{λ_E} additional equations are obtained by weighting the Dirichlet boundary conditions along S_C, S_D and S_E by the quadratic basis functions M^i in the Galerkin sense. The following linear system of $2N_\alpha + N_{\lambda_C} + N_{\lambda_D} + N_{\lambda_E}$ discretized equations is thus obtained:

$$\int_{S_C} \left(\lambda_C W_k^i - \overline{u} \frac{\partial \left(\nabla^2 W_k^i \right)}{\partial x} - \nabla^2 \overline{u} \frac{\partial \left(W_k^i \right)}{\partial x} \right) dy + \int_{S_D} \left(\lambda_D W_k^i - \overline{u} \frac{\partial \left(\nabla^2 W_k^i \right)}{\partial y} - \nabla^2 \overline{u} \frac{\partial \left(W_k^i \right)}{\partial y} \right) dx +$$

$$+ \int_{S_E} \left(-\lambda_E W_k^i - \overline{u} \frac{\partial \left(\nabla^2 W_k^i \right)}{\partial x} + \nabla^2 \overline{u} \frac{\partial \left(W_k^i \right)}{\partial x} \right) dy = -\int_{S_C} 2 \nabla^2 W_k^i dy, \quad i = 1,...,N_\alpha \, , k = 1,2 \, , \tag{13}$$

$$\int_{S_C} \overline{u} M^i dy = \int_{S_C} 2 M^i dy, \quad i = 1,...,N_{\lambda_C} \, , \tag{14}$$

$$\int_{S_D} \overline{u} M^i dx = \int_{S_D} \left[\frac{1}{2} (x+1)^2 \right] M^i dx, \quad i = 1,...,N_{\lambda_D} \, , \tag{15}$$

$$-\int_{S_E} \overline{u} M^i dy = 0, \quad i = 1,...,N_{\lambda_E} \, . \tag{16}$$

The above linear system can be written in block form as follows:

$$\begin{bmatrix} K & K_C & K_D & K_E \\ K_C^T & 0 & 0 & 0 \\ K_D^T & 0 & 0 & 0 \\ K_E^T & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_{\bar{c},\bar{d}} \\ \Lambda_C \\ \Lambda_D \\ \Lambda_E \end{bmatrix} = \begin{bmatrix} A \\ C \\ D \\ E \end{bmatrix}, \tag{17}$$

where

$$X_{\bar{c},\bar{d}} = \left[\bar{c}_1, ..., \bar{c}_{N_\alpha}, \bar{d}_1, ..., \bar{d}_{N_\alpha} \right]^T, \Lambda_C = \left[\lambda_C^1, ..., \lambda_C^{N_{\lambda_C}} \right]^T, \Lambda_D = \left[\lambda_D^1, ..., \lambda_D^{N_{\lambda_D}} \right]^T, \Lambda_E = \left[\lambda_E^1, ..., \lambda_E^{N_{\lambda_E}} \right]^T,$$

are the vectors of unknowns, and the entries in the coefficient matrix and right hand side can be read from equations (13) – (16). It should be noted that the integrands in equations (13) – (16) are nonsingular and all integrations are carried out far from the boundaries causing the singularity. Also, the stiffness matrix in (17) is symmetric but becomes singular if $N_\lambda > 2N_\alpha$, where $N_\lambda = N_{\lambda_C} + N_{\lambda_D} + N_{\lambda_E}$. This last fact will be taken into consideration when choosing specific values for these parameters.

4 NUMERICAL EXPERIMENTS

In order to to implement the SFBIM, the boundary parts S_C, S_D and S_E, i.e. the boundary parts away from the singularity, are subdivided into quadratic elements. In particular, we use N_C elements for each of the boundaries S_C and S_E, and N_D elements for boundary S_D, which makes the total number of Lagrange multipliers $N_\lambda = N_{\lambda_C} + N_{\lambda_D} + N_{\lambda_E} = 2N_{\lambda_C} + N_{\lambda_D}$, where $N_{\lambda_C} = 2N_C + 1$ and $N_{\lambda D} = 2N_D + 1$. All integrals are calculated numerically by subdividing each quadratic element into 10 subintervals and using a 15 point Gauss-Legendre quadrature over each subinterval [18, 19].

As mentioned above, the number of the singular functions N_C should be greater than the number of Lagrange multipliers N_λ, because otherwise the stiffness matrix becomes ill-conditioned or singular. On the other hand, large values of N_C should be avoided because the contributions of the high-order singular functions become either negligible (for $r < 1$) or very large (for $r > 1$) beyond the limits double precision can handle. Since, at the moment, no a-priori way of choosing the "optimal" values for N_λ and N_C exists, we have carried out systematic runs in order to study the effects the variation of these parameters would have on the numerical results.

The effect of $2N_\alpha$ on the leading SIFs can be observed in Tables 1 and 2 which show results obtained with $N_\lambda = 39$. Initially, we observe fast convergence as $2N_\alpha$ is increased, but at very high values of the latter (i.e. above $2N_\alpha = 94$) slow divergence is observed due to the inaccuracies introduced by the high-order singular functions. Tables 3 and 4 show the effect of varying $N_\lambda = N_{\lambda_C} + N_{\lambda_D} + N_{\lambda_E}$, when $2N_\alpha = 94$. Again, fast convergence is observed initially but as N_λ approaches the value of $2N_\alpha$, the results start diverging slowly, which is attributed to the fact that the stiffness matrix becomes ill-conditioned.

$2N_\alpha$	d_1	d_2	d_3	d_4	d_5	d_{10}
70	2.12751291	– 1.03669169	0.0371710	0.1177493	– 0.1227288	– 0.01108
80	2.12751343	– 1.03669221	0.0371701	0.1177510	– 0.1227319	– 0.01103
88	2.12751347	– 1.03669218	0.0371701	0.1177511	– 0.1227313	– 0.01103
90	2.12751342	– 1.03669217	0.0371701	0.1177510	– 0.1227316	– 0.01103
92	2.12751342	– 1.03669217	0.0371701	0.1177509	– 0.1227316	– 0.01103
94	**2.12751343**	**– 1.03669217**	**0.0371702**	**0.1177509**	**– 0.1227315**	**– 0.01103**
96	2.12751343	– 1.03669217	0.0371702	0.1177509	– 0.1227314	– 0.01103
100	2.12751343	– 1.03669219	0.0371702	0.1177509	– 0.1227315	– 0.01103
110	2.12751347	– 1.03669237	0.0371705	0.1177508	– 0.1227315	– 0.01102
120	2.12751343	– 1.03669229	0.0371705	0.1177508	– 0.1227314	– 0.01103

Table 1 : Convergence of the leading *odd* SIFs d_i with $2N_\alpha$; $N_\lambda = 39$.

These computations suggest that the "optimal" values for the numbers of singular functions and Lagrange multipliers are $2N_\alpha = 94$ and $N_\lambda = 39$, respectively. For higher values of $2N_\alpha$ (e.g., $2N_\alpha = 120$) satisfactory values of the SIFs are still obtained, but the quality of the global solution is not very good. When comparing the performance of the method with that in the case of Laplacian problems [16–19], we note that convergence is slower in the case of the biharmonic equation, which is reasonable since the latter is more complicated than the Laplace equation. If the smoothness of the calculated Lagrange multiplier functions is used as an indication of the

quality of the solution, then for the combination $2N_\alpha = 94$ and $N_\lambda = 39$, the calculated Lagrange multiplier functions along boundary parts S_C, S_D and S_E are the smoothest possible (see Figure 2). We note that for a slightly different value of N_λ the estimated values of the SIFs are essentially unaffected, while the calculated Lagrange multipliers exhibit oscillations.

$2N_\alpha$	c_1	c_2	c_3	c_4	c_5	c_{10}
70	0.16676222	0.0624426	– 0.1324729	– 0.0102230	0.1058502	0.004334
80	0.16676181	0.0624440	– 0.1324747	– 0.0102203	0.1058466	0.004262
88	0.16676181	0.0624439	– 0.1324746	– 0.0102211	0.1057471	0.004263
90	0.16676182	0.0624439	– 0.1324745	– 0.0102208	0.1057474	0.004264
92	0.16676184	0.0624439	– 0.1324745	– 0.0102208	0.1057474	0.004264
94	**0.16676184**	**0.0624439**	**– 0.1324745**	**– 0.0102209**	**0.1057472**	**0.004264**
96	0.16676184	0.0624439	– 0.1324745	– 0.0102208	0.1057471	0.004264
100	0.16676184	0.0624439	– 0.1324745	– 0.0102207	0.1057470	0.004264
110	0.16676179	0.0624441	– 0.1324753	– 0.0102196	0.1057450	0.004262
120	0.16676181	0.0624440	– 0.1324751	– 0.0102200	0.1057457	0.004265

Table 2 : Convergence of the leading *even* SIFs c_i with $2N_\alpha$; $N_\lambda = 39$.

N_λ	d_1	d_2	d_3	d_4	d_5	d_{10}
7+13+7	2.12751309	– 1.03669185	0.0371707	0.1177516	– 0.1227324	– 0.01103
7+17+7	2.12751334	– 1.03669214	0.0371702	0.1177509	– 0.1227318	– 0.01103
7+21+7	2.12751338	– 1.03669217	0.0371702	0.1177507	– 0.1227319	– 0.01103
7+23+7	2.12751343	– 1.03669217	0.0371702	0.1177510	– 0.1227314	– 0.01103
7+25+7	**2.12751343**	**– 1.03669217**	**0.0371702**	**0.1177509**	**– 0.1227315**	**– 0.01103**
7+27+7	2.12751342	– 1.03669217	0.0371702	0.1177509	– 0.1227315	– 0.01103
7+29+7	2.12751347	– 1.03669213	0.0371703	0.1177513	– 0.1227309	– 0.01103
7+31+7	2.12751346	– 1.03669213	0.0371703	0.1177513	– 0.1227310	– 0.01103
7+33+7	2.12751335	– 1.03669221	0.0371701	0.1177506	– 0.1227324	– 0.01103

Table 3 : Convergence of the leading *odd* SIFs d_i with N_λ ; $2N_\alpha = 94$.

$2N_\alpha$	c_1	c_2	c_3	c_4	c_5	c_{10}
7+13+7	0.16676176	0.0624436	– 0.1324753	– 0.0102209	0.1058486	0.004267
7+17+7	0.16676184	0.0624439	– 0.1324745	– 0.0102207	0.1058479	0.004264
7+21+7	0.16676185	0.0624439	– 0.1324744	– 0.0102204	0.1057480	0.004266
7+23+7	0.16676185	0.0624439	– 0.1324745	– 0.0102209	0.1057471	0.004264
7+25+7	**0.16676184**	**0.0624439**	**– 0.1324745**	**– 0.0102209**	**0.1057472**	**0.004264**
7+27+7	0.16676184	0.0624439	– 0.1324745	– 0.0102208	0.1057470	0.004263
7+29+7	0.16676180	0.0624438	– 0.1324748	– 0.0102213	0.1057461	0.004263
7+31+7	0.16676181	0.0624438	– 0.1324748	– 0.0102212	0.1057462	0.004262
7+33+7	0.16676187	0.0624440	– 0.1324743	– 0.0102201	0.1057485	0.004262

Table 4 : Convergence of the leading *even* SIFs c_i with N_λ ; $2N_\alpha = 94$.

In Table 5 the converged values of coefficients d_i and c_i , $i = 1, \ldots, 10$ obtained with the SFBIM are compared with the most accurate values obtained by the collocation Trefftz method of Li et al. [14], who reported that the leading SIF d_1 is converged up to the seventh significant digit. The SFBIM appears to be more accurate as it achieves convergence up to the eighth significant digit. Since Li et al. [14] do not provide information about the convergence of the other SIFs, in Table 5 we tabulate their computed values with one additional digit than the converged values of the SFBIM. Nevertheless, there is excellent agreement between the results of the two methods. Finally, Figures 3 and 4 show the surface plots of the approximate solution $\overline{u}$ and its partial derivatives $\overline{u}_x, \overline{u}_y, \overline{u}_{xx}$ and $\overline{u}_{yy}$. The effect of the singularity at (0, 0) is clearly visible in these profiles.

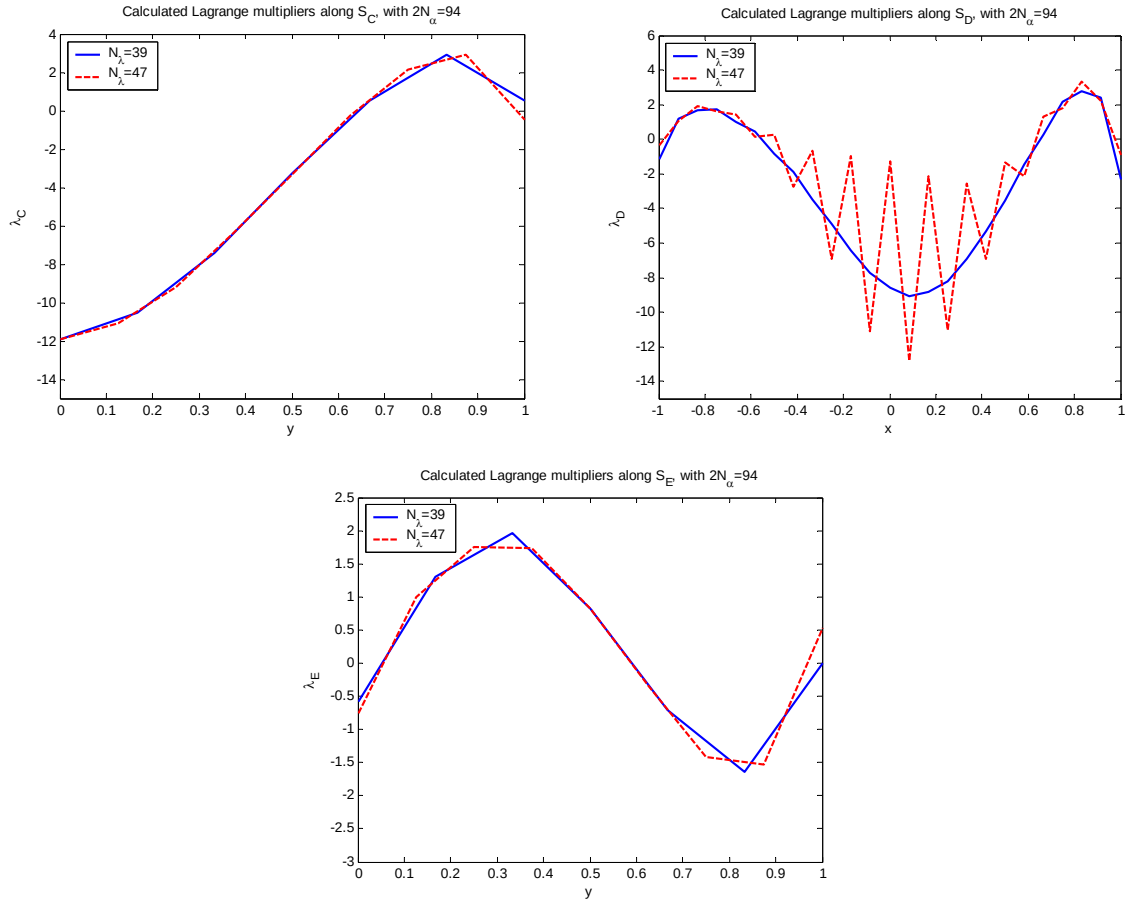


Figure 2. Calculated Lagrange multipliers along S_C, S_D, S_E with $2N_\alpha = 94$; $N_\lambda = 39$ (solid, $N_{\lambda_C} = N_{\lambda_E} = 7$, $N_{\lambda_D} = 25$) and $N_\lambda = 43$ (dashed, $N_{\lambda_C} = N_{\lambda_E} = 9$, $N_{\lambda_D} = 25$).

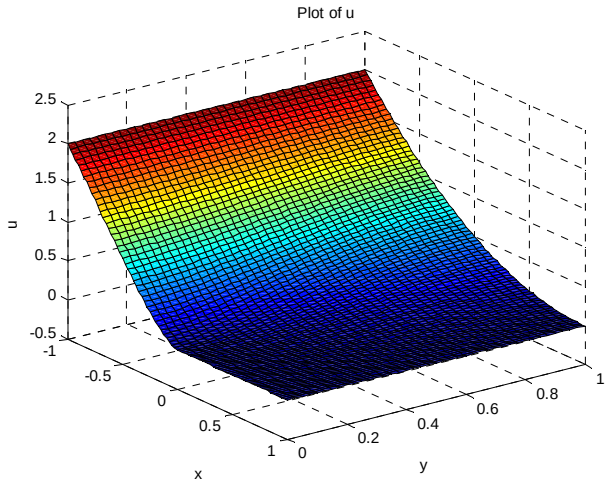


Figure 3. Plot of the converged solution $\bar{u}$.

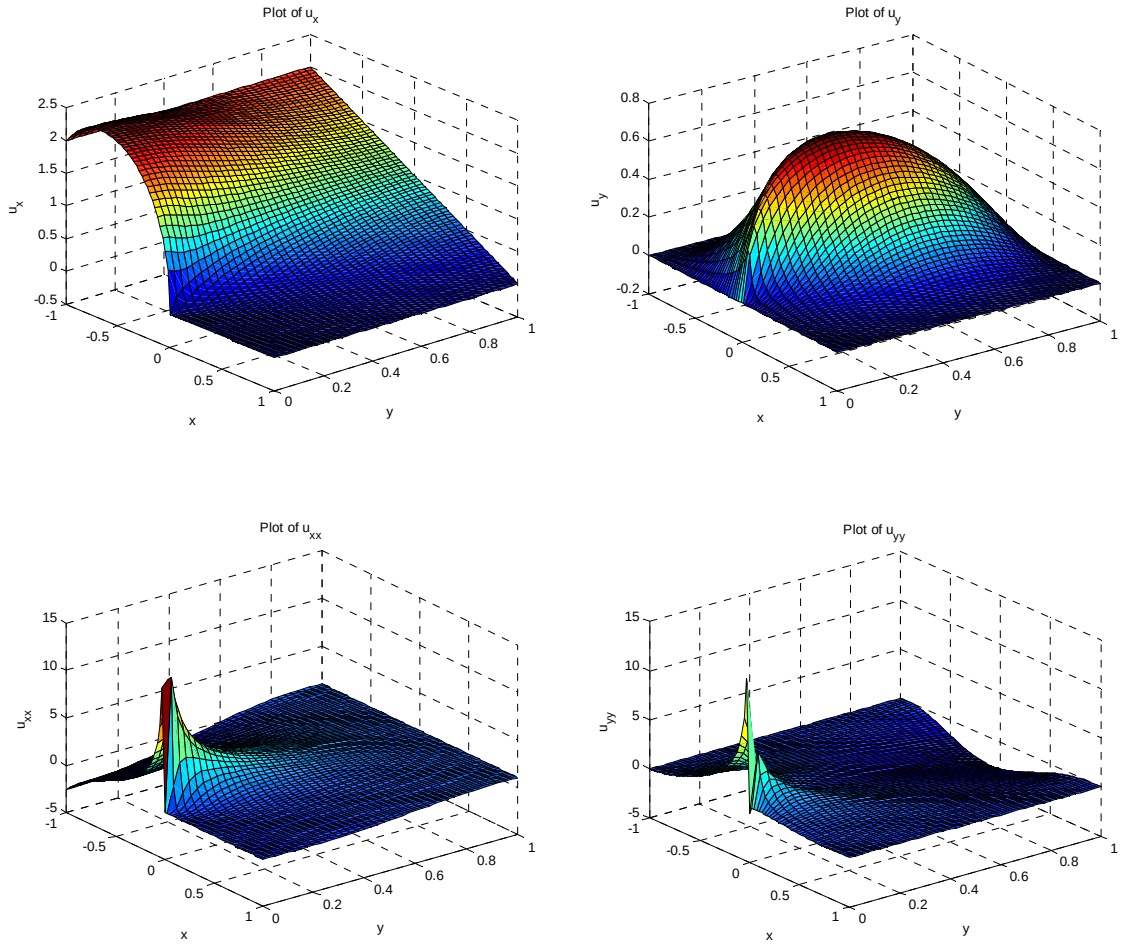


Figure 4. Plots of the first derivatives $\overline{u}_x$ and $\overline{u}_y$ (top) and second derivatives $\overline{u}_{xx}$ and $\overline{u}_{yy}$ (bottom).

5 CONCLUSIONS

The singular function boundary integral method (SFBIM) has been developed for solving two-dimensional fracture problems in terms of the Airy stress function. In this method the leading terms of the asymptotic solution are used to approximate the solution and thus the SIFs are calculated directly (i.e. no post-processing of the numerical solution is required). The governing biharmonic equation is weighted by the singular functions in the Galerkin sense, and the discretized equations are then reduced to boundary integrals by means of a double application of the divergence theorem, which leads to a significant reduction in the computational cost. Another attractive feature of the method is that integration is necessary only along boundary parts that are away from the crack tip. The Dirichlet boundary conditions are weakly enforced by means of Lagrange multipliers which, depending on the type of the boundary conditions, may replace either $\partial u/\partial n$ or $\partial\left(\nabla^2 u\right)/\partial n$ in the integrands of the discretized equations. The Lagrange multipliers are calculated together with the SIFs. The SFBIM has been applied to a model problem proposed by Schiff et al. [7]. The numerical calculations showed that the method converges very fast with the number of singular functions and the number of Lagrange multipliers, and yields accurate estimates of the leading SIFs. The value of the leading SIF, in particular, is converged up to eight significant digits. Our results agree well with the values obtained by Li et al. [14] using the collocation Trefftz method.

REFERENCES

[1] Irwin, G.R. (1957), "Analysis of stresses and strains near the end of a crack traversing a plate", Trans. ASME J. Appl Mech. 24, pp. 361–364 .

[2] Babuška, I., Miller, A. (1984), "The post-processing approach in the finite element method – Part 1: Calculation of displacements, stresses and other higher derivatives of the displacements", Int. J. Numer. Meth. Eng. 20, pp. 1085–1109.

[3] Babuška, I., Miller, A. (1984), "The post-processing approach in the finite element method – Part 2: The calculation of stress intensity factors", Int. J. Numer. Meth. Eng. 20, pp. 1111–1129.

[4] Szabó, B.A., Yosibash, Z. (1995), "Numerical analysis of singularities in two dimensions. Part 1: Computation of the eigenpairs", Int. J. Numer. Meth. Eng. 38, pp. 2055–2082.

[5] Szabó, B.A., Yosibash, Z. (1996), "Numerical analysis of singularities in two dimensions. Part 2: Computation of the generalized flux/stress intensity factors", Int. J. Numer. Meth. Eng. 39, pp. 409–434.

[6] Tracey, D.M. (1971), "Finite elements for determination of crack tip elastic stress intensity factors", J. Eng. Fract. Mech. 3, pp. 255–265.

[7] Schiff, B.D., Fishelov, D., Whiteman, J.R. (1979), "Determination of a stress intensity factor using local mesh refinement", *Proceedings of the Mathematics of Finite Elements and Applications III* (MAFELAP III), London, Academic Press, pp. 55–64.

[8] Rahulkumar, P., Saigal, S., Yunus, S. (1997), "Singular *p*-version finite elements for stress intensity factor computations", Int. J. Numer. Meth. Eng. 40, pp. 1091–1114.

[9] Schnack, E. (1985), "Singularities of cracks with generalized finite elements", Lecture Notes in Math. 1121, pp. 258–277.

[10] Karageorghis, A. (1992), "Modified methods of fundamental solutions for harmonic and biharmonic problems with boundary singularities", J. Num. Meth. Partial Diff. Eqns 8, pp. 1–19.

[11] Herrera, I., Diaz, M. (1999), "Indirect methods of collocation: Trefftz-Herrera collocation", ", J. Num. Meth. Partial Diff. Eqns 15, pp. 709–738.

[12] Jin, W.G., Cheung, Y.K., Zienkiewicz, O.C. (1993), "Trefftz method for Kirchoff plate bending problems", Int. J. Numer. Meth. Eng. 36, pp. 765–781.

[13] Lefeber, D. (1984), *Solving problems with singularities using boundary elements*, Computational Mechanics Publications, Southampton.

[14] Li, Z.C., Lu, T.T., Hu, H.Y. (2004), "The collocation Trefftz method for biharmonic equations with crack singularities", Eng. Anal. Bound. Elem. 28, pp. 79–96.

[15] Olson, L.G., Georgiou, G.C., Schultz, W.W. (1991), "An efficient finite element method for treating singularities in Laplace's equation", J. Comp. Phys. 96, pp. 391–410.

[16] Georgiou, G.C., Olson, L., Smyrlis, G. (1996), "A singular function boundary integral method for the Laplace equation", Commun. Numer. Meth. Eng. 12, pp. 127–134.

[17] Georgiou, G.C., Boudouvis, A., Poullikkas, A. (1997), "Comparison of two methods for the computation of singular solutions in elliptic problems", J. Comp. Appl. Math. 79, pp. 277–290.

[18] Elliotis, M., Georgiou, G.C., Xenophontos, C. (2002), "The solution of a Laplacian problem over an L-shaped domain with a singular function boundary integral method", Commun. Numer. Meth. Eng. 18, pp. 213–222.

[19] Elliotis, M., Georgiou, G.C., Xenophontos, C. (2005), "Solving Laplacian problems with boundary singularities: A comparison of a singular function boundary integral method with the *p/hp* version of the finite element method", in press in Appl. Math. Comp.

SOLUTION OF THE STICK-SLIP PROBLEM WITH THE SINGULAR FUNCTION BOUNDARY INTEGRAL METHOD

Miltiades Elliotis[*], Georgios Georgiou[*†], and Christos Xenophontos[*†]

[*]Department of Mathematics and Statistics
University of Cyprus
P.O. BOX 20537, 1678 Nicosia, Cyprus
e-mail: eliotis@ucy.ac.cy

[†]Greek Association of Computational Mechanics

Keywords: Singular Function Boundary Integral Method, Newtonian stick-slip problem, biharmonic equation, Lagrange multipliers, singular coefficients.

Abstract. *A singular function boundary integral method (SFBIM) is proposed for solving biharmonic problems with boundary singularities. The method is applied to the Newtonian stick-slip flow problem. The stream-function is approximated by the leading terms of the local asymptotic solution expansion which are also used to weight the governing biharmonic equation in the Galerkin sense. By means of the divergence theorem the discretized equations are reduced to boundary integrals. The Dirichlet boundary conditions are weakly enforced by means of Lagrange multipliers, the values of which are calculated together with the singular coefficients. The method converges very fast with the number of singular functions and the number of Lagrange multipliers. Accurate estimates of the leading singular coefficients are obtained. Comparisons with the analytical solution and results obtained with other numerical methods are also made.*

1 INTRODUCTION

In the past few decades, many different numerical methods have been proposed for the treatment of boundary singularities in plane elliptic boundary value problems, in order to improve the solution accuracy and resolve the convergence difficulties occurring in the vicinity of such singular points. These methods range from special mesh-refinement schemes to sophisticated techniques that incorporate, directly or indirectly, the form of the local asymptotic expansion, which is known in many occasions. An exhaustive survey of treatment of singularities in elliptic boundary value problems is provided in the recent articles by Li and Lu[13], by Dosiyev[2] and Shi et al.[20]. Knowledge of the coefficients appearing in the local solution expansion is often desired by many engineering applications. These coefficients, referred to as *singular coefficients* or *generalized stress intensity factors*[21], are calculated either directly[3-5] or by post-processing the numerical solution[1, 22].

In the past few years, we have developed the Singular Function Boundary Integral Method (SFBIM) for Laplacian problems with boundary singularities[3,4,5,9,10], in which the unknown singular coefficients are calculated directly. The solution is approximated by the leading terms of the local asymptotic solution expansion which are also used to weight the governing equation in the Galerkin sense. With a double application of Green's theorem, the discretized equations are reduced to boundary integrals over those parts of the boundary that do not involve the singular point. The Dirichlet boundary conditions are weakly enforced by means of Lagrange multipliers, which are calculated simultaneously with the singular coefficients. The method has been tested on standard Laplacian problems, yielding extremely accurate estimates of the leading singular coefficients and exhibiting exponential convergence with respect to the number of singular functions[3,4,5,9,10].

The objective of the present paper is to extend the SFBIM to biharmonic problems with boundary singularities. For that purpose we have chosen to solve the Newtonian planar stick-slip problem, which is a benchmark Stokes flow problem used to test various numerical methods proposed in the literature for the solution of viscous and non-Newtonian flows, such as the extrudate-swell flow. This concerns the extrusion of a fluid from a slit or an axi-symmetric die into the atmosphere. Due to the relaxation of stresses, the fluid swells as it exits the die. Swelling is particularly pronounced in the case of elastic fluids, but it is also observed in the Newtonian case, provided that the Reynolds number is sufficiently low. The stick-slip problem is a special case of the extrudate-swell problem: in the limit of infinite surface tension, no swelling occurs and free surface becomes flat (in the case of slit die). A boundary inverse-square-root stress singularity appears at the exit of the die due to the sudden change of the boundary conditions from the wall to the flat free surface.

Direct estimates of the leading singular coefficients in the case of the planar stick-slip problem have been reported by various researchers who employed a variety of numerical methods and techniques to incorporate the

leading terms of the local asymptotic solution (which is equivalent to subtracting the leading terms of the singularity). Kelmanson employed a direct modified boundary integral equation method (BIEM) incorporating a subtraction of the singular terms technique that accelerated the rate of convergence and reported estimates of the leading four coefficients[12]. Estimates for these coefficients have also been reported by Georgiou et al.[8] who solved the problem by using the integrated singular basis function method (ISBFM). In this method, the singular functions are directly subtracted from the original problem formulation which leads to a modified problem with the regular part of the solution and the singular coefficients as unknowns. The smooth problem is then solved using finite elements. The integrals involving singular contributions are reduced to boundary ones by means of a double integration by parts and the original essential boundary conditions are enforced by means of Lagrange multipliers. These two features are encountered also with the SFBIM that we propose in the present work.

Georgiou et al.[7] developed a singular finite element method (SFEM), in which special elements incorporating the radial form of the local singularity expansion are employed in a small region around the singular point, in order to resolve the convergence difficulties and improve the accuracy of the global solution. They obtained more accurate results than those achieved with ordinary elements and calculated the leading singular coefficients by post-processing the finite element solution. A similar post-processing technique has been employed by Salamon et al.[19] who obtained accurate results near the singularity using high resolution finite elements, with quasi-orthogonal mesh generation and local, adaptive mesh refinement with irregular imbendded elements. Ngamaramvaranggul and Webster[14] developed a semi-implicit Taylor-Galerkin/pressure-correction finite element method (STGFEM) for free surface flows and applied it to various Newtonian flows including the plane and axi-symmetric stick-slip and extrudate-swell problems. More recently, Normadin et al.[15] solved the Newtonian stick-slip problem using a finite element Galerkin technique associated with stream-tube analysis.

2 GOVERNING EQUATIONS AND ASYMPTOTIC SOLUTION

The planar stick-slip problem is the idealization of the extrusion of a Newtonian fluid between parallel plates at infinite surface tension. The geometry of the flow is depicted in Figure 1. Due to symmetry, only the upper half of the flow domain is considered, i.e. boundary part S_D denotes the plane of symmetry. Boundary parts S_A and S_B represent the wall and the flat free surface, respectively. The latter is flat in the limit of infinite surface tension. Finally, S_C and S_E are, respectively, the artificial inlet and outlet boundaries.

In the creeping case, the flow is governed by the biharmonic equation

$$\nabla^4 u = 0 \qquad in \qquad \Omega, \tag{1}$$

where u is the stream-function defined by $u_x \equiv \partial u/\partial y$ and $u_y \equiv -\partial u/\partial x$, u_x and u_y being the velocity components in the $-x$ and $-y$ directions, respectively. The boundary conditions of the flow and the domain Ω are also depicted in Figure 1. Along the wall S_A there is no slip and no penetration (i.e. the two velocity components are zero). Along the free surface S_B, both u_y and the xy-stress component are zero which leads to $\Delta u = 0$. The inflow and outflow planes are taken at a distance L before and after the die exit. This distance is assumed to be sufficiently large so that the flow corresponds to the fully developed Poiseuille flow at the inflow plane and to a plug (i.e. uniform) flow at the outflow plane. Finally, along the symmetry plane (boundary S_D), the vertical velocity component and the shear stress are zero, i.e. the centerline is a slip surface. The stick-slip flow is characterized by the presence of a stress singularity at the exit caused by the sudden change in the boundary conditions, from no slip(stick) along the wall S_A to full slip along the flat free surface S_B.

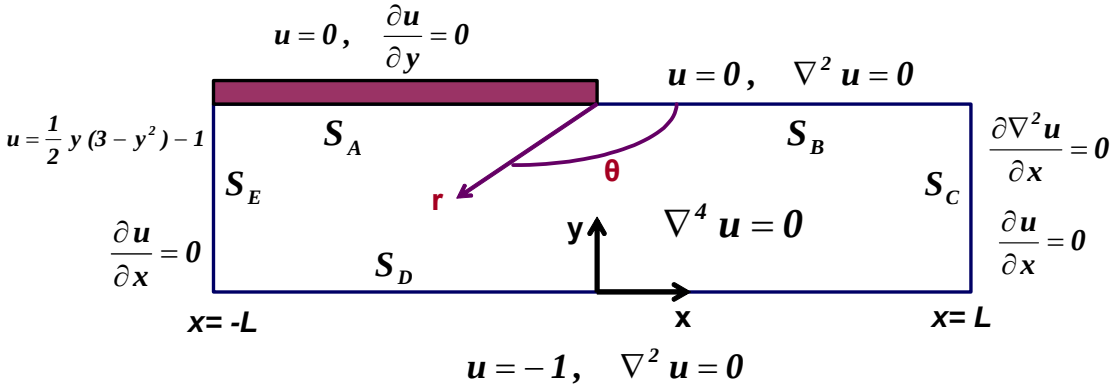


Figure 1. The planar stick-slip problem in terms of the stream-function u

The problem of Figure 1 has the following mathematical expression:

$$\nabla^4 u = 0 \qquad in \qquad \Omega, \tag{2}$$

with

$$\begin{aligned} u &= 0, & \frac{\partial u}{\partial y} &= 0 \quad on \quad S_A, \\ u &= 0, & \nabla^2 u &= 0 \quad on \quad S_B, \\ \frac{\partial \nabla^2 u}{\partial x} &= 0, & \frac{\partial u}{\partial x} &= 0 \quad on \quad S_C, \\ u &= -1, & \nabla^2 u &= 0 \quad on \quad S_D, \\ u &= \frac{1}{2} y(3-y^2) - 1, & \frac{\partial u}{\partial x} &= 0 \quad on \quad S_E. \end{aligned} \tag{3}$$

Note that the weak condition $\partial(\Delta u)/\partial x=0$ along boundary S_C can be replaced by the stronger Dirichlet condition $u=y-1$, which leads to a different formulation, since with the SFBIM, imposing Dirichlet conditions requires the introduction of (unknown) Lagrange multipliers. The asymptotic solution in the neighbourhood of the singularity can be expressed in terms of an eigen-function expansion of the form[8,12,18] :

$$u(r,\theta) = \sum_{j=1}^{\infty} \alpha_j r^{\mu_j+1} f(\theta, \mu_j), \qquad (r,\theta) \in \Omega, \tag{4}$$

where (r,θ) are the polar co-ordinates centered at the singular point , μ_j, with j = 1,2,…, are the singularity powers arranged in ascending order, the functions $f(\theta,\mu_j)$, represent the θ-dependence of the eigensolution, and α_j are the unknown singular coefficients determined by the global flow.

The functions $W^j \equiv r^{\mu_j+1} f(\theta,\mu_j)$ are referred to as singular functions. The local solution (4) consists of even and odd solutions, the corresponding singular functions of which will be denoted by W_1^j and W_2^j, respectively. In the case of even solutions[18], we have

$$W_1^j = r^{\mu_j+1} f_1(\theta,\mu_j), \tag{5}$$

with

$$f_1(\theta,\mu_j) = \cos(\mu_j+1)\theta - \cos(\mu_j-1)\theta, \qquad \mu_j = j - \frac{1}{2}, \qquad j = 1,2,\ldots, \tag{6}$$

whereas in the case of odd solutions we have

$$W_2^j = r^{\mu_j+1} f_2(\theta,\mu_j), \tag{7}$$

with

$$f_2(\theta,\mu_j) = (\mu_j-1)\sin(\mu_j+1)\theta - (\mu_j+1)\sin(\mu_j-1)\theta, \qquad \mu_j = j+1, \quad j = 1,2,\ldots, \tag{8}$$

Thus the first singular function is $W_1^1 = r^{3/2}$ [cos(3θ/2) - cos(θ/2)], which indicates that the velocity gradients and the stresses vary as the inverse square root of the radial distance from the singular point changes.

In what follows we will be using the symbols α_j and β_j the singular coefficients corresponding to the even and odd singular functions, respectively. Thus, the local solution is written as follows:

$$u = \sum_{j=1}^{\infty} \alpha_j W_1^j + \sum_{j=1}^{\infty} \beta_j W_2^j . \tag{9}$$

3 THE SINGULAR FUNCTION BOUNDARY INTEGRAL METHOD

In the SFBIM[3,4,5] the solution of the problem (2)-(3) is approximated by the leading terms of the local asymptotic solution expansion (9). By employing the first N terms in both sums of (9) the approximate solution is

$$\overline{u} = \sum_{j=1}^{N} \overline{\alpha}_j W_1^j + \sum_{j=1}^{N} \overline{\beta}_j W_2^j . \tag{10}$$

where $\overline{\alpha}_j$ and $\overline{\beta}_j$ are the approximations of the singular coefficients. Obviously, the total number of singular functions involved in the approximation (10) is 2*N*. By applying Galerkin's principle, the governing equation is weighted by the singular functions used in the approximation of the solution. Hence, the following set of discretized equations is obtained:

$$\int_\Omega \overline{u} W_k^i dV = 0, \qquad i = 1,2,\cdots,N, \quad k = 1,2. \tag{11}$$

By applying Green's theorem twice and taking into account that the singular functions W_k^i are biharmonic, the above volume integrals are reduced to boundary ones:

$$\int_{\partial\Omega}\left(\frac{\partial\overline{u}}{\partial n}\nabla^2 W_k^i - \overline{u}\frac{\partial\nabla^2 W_k^i}{\partial n}\right)dS + \int_{\partial\Omega}\left(\frac{\partial\nabla^2\overline{u}}{\partial n}W_k^i - \nabla^2\overline{u}\frac{\partial W_k^i}{\partial n}\right)dS \quad i = 1,2,\cdots,N \quad k = 1,2 \tag{12}$$

where $\partial\Omega = S_A \cup S_B \cup S_C \cup S_D \cup S_E$. The dimension of the problem is thus reduced by one, which leads to a considerable reduction of the computational cost. Since W_k^i satisfy exactly the boundary conditions along S_A and S_B, the above integral along these boundary segments, is identically zero. Therefore,

$$\int_S\left(\frac{\partial\overline{u}}{\partial n}\nabla^2 W_k^i - \overline{u}\frac{\partial\nabla^2 W_k^i}{\partial n}\right)dS + \int_S\left(\frac{\partial\nabla^2\overline{u}}{\partial n}W_k^i - \nabla^2\overline{u}\frac{\partial W_k^i}{\partial n}\right)dS \quad i = 1,2,\cdots,N \quad k = 1,2 \tag{13}$$

where $S = S_C \cup S_D \cup S_E$. In the SFBIM the Dirichlet boundary conditions are imposed by means of Lagrange multipliers which replace the normal derivative of the solution *u*. In the problem under study Dirichlet boundary conditions appear only along boundary parts S_D and S_E. Since along S_E the normal derivative $\partial u/\partial x$ vanishes, Lagrange multipliers are chosen to replace $\partial(\Delta u)/\partial x$ in the boundary integrals of Eq. (13). Boundary parts S_D and S_E are partitioned into three-node elements and the corresponding Lagrange multipliers, denoted respectively by λ_D and λ_E, are expanded in terms of quadratic basis functions M^j:

$$\lambda_D = \frac{\partial\overline{u}}{\partial y} = \sum_{j=1}^{N_{\lambda_D}} \lambda_D^j M^j \quad on \quad S_D, \qquad \lambda_E = \frac{\partial\nabla^2\overline{u}}{\partial x} = \sum_{j=1}^{N_{\lambda_E}} \lambda_E^j M^j \quad on \quad S_E, \tag{14}$$

where N_λ is the number of discrete Lagrange multipliers λ_D^j and λ_E^j along the corresponding boundaries. The nodal values of λ_D and λ_E are additional unknowns of the problem in study. Furthermore, the required additional equations ($N_\lambda = N_{\lambda_D} + N_{\lambda_E}$) are obtained by weighting the Dirichlet boundary conditions along S_D and S_E by the quadratic basis functions M^j in the Galerkin sense. The following linear system of $2N+N_\lambda$ discretized equations is obtained:

$$\int_{S_C}\left(-\overline{u}\frac{\partial\nabla^2 W_k^i}{\partial x} - \nabla^2\overline{u}\frac{\partial W_k^i}{\partial x}\right)dy + \int_{S_D}\left(-\lambda_D\nabla^2 W_k^i + \overline{u}\frac{\partial\nabla^2 W_k^i}{\partial y} - \frac{\partial\nabla^2\overline{u}}{\partial y}W_k^i\right)dx +$$

$$\int_{S_E}\left(-\lambda_E W_k^i + \overline{u}\frac{\partial\nabla^2 W_k^i}{\partial x} + \nabla^2\overline{u}\frac{\partial W_k^i}{\partial x}\right)dy = 0, \qquad i = 1,2,\cdots,N, \qquad k = 1,2, \tag{15}$$

$$\int_{S_D}\overline{u}M^i\,dx = -\int_{S_D}M^i\,dx, \qquad i = 1,2,\cdots,N_{\lambda_D}, \tag{16}$$

$$\int_{S_E}\overline{u}M^i\,dy = \int_{S_E}\left[\frac{1}{2}y\left(3-y^2\right)-1\right]M^i\,dy, \qquad i = 1,2,\cdots,N_{\lambda_E}. \tag{17}$$

The above linear system is not symmetric. It should be noted that the integrands in the above equations are non-singular and all integrations are carried out far from the boundaries causing the singularity. Note that the stiffness matrix is not symmetric and that it becomes singular if $N_\lambda > 2N$. We should emphasize that we may also create three alternative formulations which are based on the choice of λ_C and λ_D in terms of $\partial(\Delta u)/\partial x$ and $\partial u/\partial x$.

4 NUMERICAL RESULTS

Calculations have been carried out with all four formulations presented in Section 3. In order to implement

the SFBIM, the boundary parts S_C, S_D and S_E (i.e. the boundary parts away from the singularity) are subdivided into quadratic elements. Specifically, we employ N_E elements for each one of boundaries S_E and S_C and N_D elements over boundary S_D. The total number of Lagrange multipliers is $N_\lambda=(2N_D+1)+(2N_E+1)$. The integrals are calculated numerically by subdividing each quadratic element into 10 subintervals and using a 15-point Gauss-Legendre quadrature over each subinterval[3]. Unless otherwise indicated, the semi-length L of the domain has been taken equal to 3.

As already mentioned, the number of singular functions $2N$ should be much greater than the number of Lagrange multipliers N_λ, since otherwise the stiffness matrix is ill-conditioned or singular. On the other hand, large values of $2N$ should be avoided because the contributions of the high-order singular functions become either negligible or (for r < 1) or very large (for r > 1) beyond the limits double precision can handle.

$2N$	α_1	α_2	α_3	α_4	α_5	α_{10}
70	0.6909892	0.2645003	0.030364	-0.021405	-0.002845	0.00024
80	0.6909881	0.2645007	0.030376	-0.021407	-0.002900	0.00022
86	0.6909882	0.2645004	0.030374	-0.021405	-0.002892	0.00021
88	**0.6909882**	**0.2645004**	**0.030374**	**-0.021404**	**-0.002891**	**0.00021**
90	0.6909882	0.2645002	0.030375	-0.021403	-0.002895	0.00021
92	0.6909885	0.2645045	0.030371	-0.021436	-0.002875	0.00034

Table 1. Convergence of the singular coefficients α_i with $2N$; $N_\lambda=32$

$2N$	β_1	β_2	β_3	β_4	β_5	β_{10}
70	-0.0808635	-0.017115	0.001726	0.001231	-0.000282	-0.000001
80	-0.0808617	-0.017119	0.001720	0.001240	-0.000270	-0.000006
86	-0.0808619	-0.017119	0.001720	0.001238	-0.000271	-0.000005
88	**-0.0808619**	**-0.017119**	**0.001720**	**0.001238**	**-0.000271**	**-0.000005**
90	-0.0808617	-0.017119	0.001720	0.001238	-0.000270	-0.000005
92	-0.0808645	-0.017122	0.001729	0.001245	-0.000287	-0.000009

Table 2. Convergence of the singular coefficients β_i with $2N$; $N_\lambda= 32$

Systematic runs have been carried out in order to study the effects of both N and N_λ , on the numerical results. The effect of $2N$ on the leading singular coefficients can be observed in Tables 1 and 2, which show results obtained with N_λ =32. Fast convergence is observed as $2N$ is increased and accurate estimates of the leading singular coefficients are obtained. However, at very high values of $2N$ (i.e. above $2N=88$) slow divergence is observed due to the inaccuracies introduced by the high-order singular functions.

The convergence of the method with the number of Lagrange multipliers is shown in Tables 3 and 4 in which the values of the leading singular coefficients, calculated with $2N=88$ and various values of N_λ, are tabulated. Again, fast convergence is observed initially but as N_λ approaches the value of $2N$, the results start diverging slowly, which is attributed to the fact that the stiffness matrix becomes ill-conditioned. Our computations showed that the optimal values are $N_\lambda=32$ and $2N=88$. For higher values of $2N$ (e.g. $2N=120$) satisfactory values of the singular coefficients are still obtained, but the quality of the global solution is not good.

N_λ	α_1	α_2	α_3	α_4	α_5	α_{10}
21+5	0.6909883	0.2645002	0.030374	-0.021403	-0.002890	0.00021
21+7	0.6909882	0.2645007	0.030375	-0.021407	-0.002893	0.00022
25+5	0.6909883	0.2645002	0.030373	-0.021403	-0.002888	0.00021
25+7	**0.6909882**	**0.2645004**	**0.030374**	**-0.021404**	**-0.002891**	**0.00021**
25+9	0.6909882	0.2645005	0.030375	-0.021405	-0.002895	0.00021
29+7	0.6909883	0.2645002	0.030374	-0.021403	-0.002892	0.00021

Table 3. Convergence of the singular coefficients α_i with N_λ; $2N=88$

N_λ	β_1	β_2	β_3	β_4	β_5	β_{10}
21+5	-0.0808618	-0.017118	0.001720	0.001237	-0.000271	-0.000005
21+7	-0.0808620	-0.017119	0.001721	0.001239	-0.000272	-0.000006
25+5	-0.0808619	-0.017118	0.001720	0.001237	-0.000271	-0.000005
25+7	**-0.0808619**	**-0.017119**	**0.001720**	**0.001238**	**-0.000271**	**-0.000005**
25+9	-0.0808618	-0.017119	0.001720	0.001239	-0.000271	-0.000006
29+7	-0.0808618	-0.017118	0.001720	0.001237	-0.000271	-0.000005

Table 4. Convergence of the singular coefficients β_i with N_λ; $2N=88$.

An indication of the quality of the solution is given by the smoothness of the calculated Lagrange multipliers. For the optimal combination $N_\lambda=32$ and $2N=88$, the calculated Lagrange multiplier function along boundary S_D is smooth. As shown in Figure 2, for a slightly different value of N_λ (i.e. $N_\lambda=36$), the calculated Lagrange multiplier function exhibits oscillations, while the values of the singular coefficients are essentially the same.

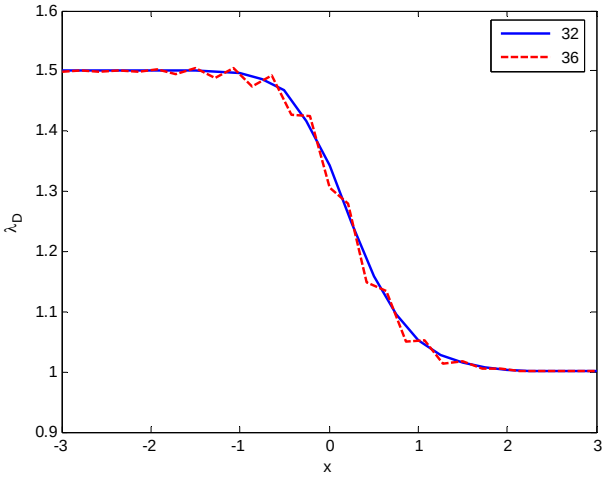


Figure 2. Plot of Lagrange multiplier function along boundary S_D; $N_\lambda=32$ (solid) and 36 (dashed); $2N=88$.

In Table 5, the values of α_1, α_2, α_3 and β_1 calculated with the SFBIM are compared with values obtained with other methods reported in literature. To our knowledge, there are no reports in the literature for the values of the higher order coefficients. The value 0.690988 for α_1 agrees with the analytical solution up to the 6th significant digit.

Method	α_1	α_2	α_3	β_1
Modified BIEM[17]	0.69108	0.26435	0.04962	-0.07990
Singular FEM[7]	0.69173	0.27168	0.05013	
ISBFM[8]	0.69104	0.26140	-0.01263	
Modified MFS[11]	0.690984	0.274807	-0.022104	-0.043983
J-integral method[23]	0.6910			
Spectral DDM[16]	0.69035	0.26404	0.03069	-0.08051
High-resolution FEM[19]	0.69160	0.27183	0.05232	
Modified MFS[17]	0.69019			
SFBIM (present work)	0.690988	0.264500	0.03037	-0.080862
Analytical solution[18]	0.6909883			

Table 5. Comparison of computed singular coefficients with the results of other methods

In Figure 3, the pressure along the wall and the slip surface ($y=1$) is plotted. This is in good agreement with the results of Salamon et al.[19]. The pressure goes to infinity as the singularity is approached from the right,

while it remains finite for negative x. This discontinuity cannot be captured by standard numerical methods in which a continuous approximation is used for the pressure.

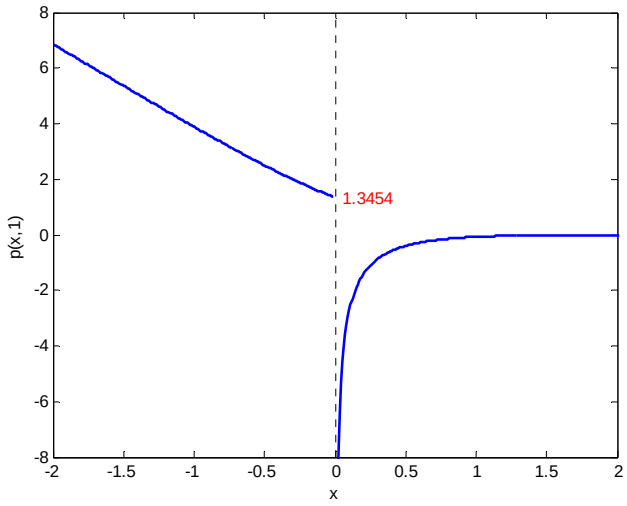


Figure 3. Calculated pressure along $y=1$ (SFBIM with $2N=88$ and $N_\lambda=32$)

The effect of the length of the domain on the computations has also been studied. As expected, the values of singular coefficients change dramatically with L for small values of the latter, since the assumptions for fully developed and uniform flow along the inlet and the outlet planes, respectively, are not valid when the two planes are taken close to the die exit. This effect is illustrated in Figure 4, where the calculated values of α_1, are plotted versus L. We observe that the value $L=3$ is sufficiently high to assure the validity of the imposed inlet and outlet boundary conditions. At higher values of L, the accuracy of the computed solutions starts deteriorating, due to the fact that the number of the corresponding Lagrange multipliers along S_D is kept fixed, while the length of the boundary S_D increases. As already noted, increasing the number of Lagrange multipliers along S_D will not improve the accuracy, since it leads to ill-conditioning of the stiffness matrix.

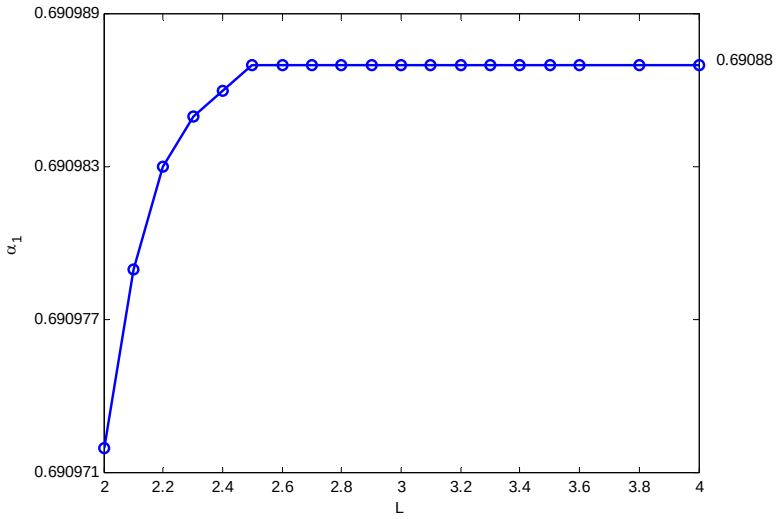


Figure 4. Convergence of α_1 with the semi-length L of the domain ($N_\lambda = 32$).

5 CONCLUSIONS

The SFBIM has been developed for solving a biharmonic problem with a boundary singularity, i.e. the Newtonian planar stick-slip problem in terms of the stream-function. This method converges very fast with the number of singular functions and the number of Lagrange multipliers. Accurate estimates of the leading singular coefficients are obtained. In particular, the value 0.690988 for α_1 agrees very well with the analytical solution

and is much more accurate than previously reported values.

REFERENCES

[1] Babuska, I. and Miller, A. (1984), ‘‘The post-processing approach in the finite element method – Part 2: The calculation of stress-intensity factors,’’ Int. J. Numer. Meth. Eng. 20, pp. 1111-1129.
[2] Dosiyev, A.A. (2004), ‘‘The high accurate block-rigid method for solving Laplace’s boundary value problem with singularities,’’ SIAM J. Numer. Anal. 42, pp. 153-178.
[3] Elliotis, M., Georgiou, G. and Xenophontos, C. (2002), ‘‘The solution of a Laplacian problem over an L-shaped domain with a singular function boundary integral method,’’ J. Commun. Numer. Meth. Eng. 18, 213-222.
[4] Elliotis, M., Georgiou, G. and Xenophontos, C. (2005), ‘‘Solving Laplacian problems with boundary singularities: A comparison of a singular function boundary integral method with the p/hp version of the finite element method,’’ J. of Appl. Math. Comput., in press.
[5] Elliotis, M., Georgiou, G., and Xenophontos, C. (2005), ‘‘Solution of the planar Newtonian stick-slip problem with a singular function boundary integral method,’’ Int. J. Numer. Methods Fluids, in press.
[6] Elliotis, M., Georgiou, G. and Xenophontos, C. (2005), ‘‘The singular function boundary integral method for a two-dimensional fracture problem,’’ submitted.
[7] Georgiou, G.C., Olson, L.G., Schultz, W.W. and Sagan, S. (1989), ‘‘A singular finite element for Stoke’s flow: the stick-slip problem,’’ Int. J. Numer. Methods Fluids 9, pp. 1353-1367.
[8] Georgiou, G.C., Olson, L.G. and Schultz, W.W. (1991), ‘‘The integrated singular basis function method for the stick-slip and the die-swell problems,’’ Int. J. Numer. Methods Fluids 13, pp. 1251-1265.
[9] Georgiou, G.C., Olson, L.G. and Smyrlis, Y. (1996), ‘‘A singular function boundary integral method for the Laplace equation,’’ J. Commun. Numer. Meth. Eng. 12, pp. 127-134.
[10] Georgiou, G.C., Boudouvis, A. and Poullikkas, A. (1997), ‘‘Comparison of two methods for the computation of singular solutions in elliptic problems,’’ J. Comput. Appl. Math. 79, pp. 277-290.
[11] Karageorghis, A. (1992), ‘‘Modified methods of fundamental solutions for harmonic and biharmonic problems with boundary singularities,’’ J. Num. Meth. Partial Diff. Eqns 8, pp. 1-19.
[12] Kelmanson, M.A. (1983), ‘‘An integral equation method for the solution of singular and slow flow problems,’’ J. Comp. Physics 51, pp. 139-158.
[13] Li, Z.C. and Lu, T.T. (2000), ‘‘Singularities and treatments of boundary value problems,’’ J. Math. Comp. Model. 31, pp. 97-145.
[14] Ngamaramvaranggul, V. and Webster, M.F. (2000), ‘‘Computation of free surface flows with a Taylor-Galerkin/pressure-correction algorithm,’’ J. Numer. Methods Fluids 33, pp. 993-1026.
[15] Normandin, M., Radu, D.G., Mahmoud, A. and Clermont, J.R. (2001), ‘‘Finite-element and stream-tube formulations for flow-computations – two-dimensional applications,’’ J. Math. Comp. in Simul. 60, pp. 129-134.
[16] Owens, R.G. and Phillips, T.N. (1991), ‘‘A spectral domain decomposition method for the planar non-Newtonian stick-slip problem,’’ J. Comp. Appl. Math. 41, pp. 43-79.
[17] Poullikkas, A., Karageorghis, A. and Georgiou, G.C. (1998), ‘‘Methods of fundamental solutions for harmonic and biharmonic boundary value problems,’’ J. Comp. Mech. 21, pp. 416-423.
[18] Richardson, S. (1970), A stick.-slip problem related to the motion of a free jet at low Reynolds numbers,’’ Proc. Camb. Phil. Soc. 67, pp. 477-489.
[19] Salamon, T.R., Bornside, D.E., Armstrong, R.C. and Brown, R.A. (1995), ‘‘The role of surface tension in the dominant balance in the die-swell singularity,’’ J. Phys. Fluids 7, pp. 2328-2344.
[20] Shi, J.M., Breurer, M. and Durst, F. (2004), ‘‘A combined analytical-numerical method for treating corner singularities in viscous flow predictions,’’ Int. J. Numer. Methods Fluids 45, pp. 659-688.
[21] Szabo, B. and Babuska, I. (1991), Finite Element Analysis, John Wiley & Sons, Inc., NewYork.
[22] Szabo, B. and Yosibash, Z. (1996), ‘‘Numerical analysis of singularities in two dimensions. Part 2: Computation of generalized flux/stress intensity factors,’’ Int. J. Numer. Meth. Eng. 39, pp. 409-434.
[23] Tanner, R.I. and Huang, X. (1993), ‘‘Stress singularities in non-Newtonian stick-slip and edge flows,’’ J. Non-newtonian Fluid Mech. 50, pp. 135-160.

GAUSS QUADRATURE RULE FOR HYPERSINGULAR INTEGRALS WITH MANY POLES

G. Tsamasphyros

School of Applied Mathematical and Physical Sciences, Dept. of Mechanics-Lab. of Computational Mechanics
The National Technical University of Athens, Zoghraphou Campus,
Theocaris Bld., GR-0157 73, Athens, Greece
e-mail: tsamasph@central.ntua.gr

Keywords: Strong Singularity, Nearby Singularity, Orthogonal Polynomials, Finite part integral, Gauss quadrature.

Abstract: *This paper is a generalization of the author's previous paper where a set of polynomials* $\varphi_n(x)$ *are constructed. These polynomials are orthogonal with respect to the weight* $w(x)/R(x)$, $\left(R(x)=(x-t)^k r(x), r(x)=\prod_{i=1}^{n}(x-y_i)\right)$, *where* $w(x)$ *is a weight function,* k *is an integer and* y_i *belongs or is very close to the integration interval. These polynomials are used to extract Gauss quadrature rules for the hypersingular integrals*

$$I(f)= \oint_{-1}^{1} \frac{w(x)f(x)}{R(x)}dx$$

The process is exactly the same as the one used for the extraction of the classical Gauss formula for the Riemann integrals.

1. INTRODUCTION

It is well known, since 1969 [1], that one experiences serious difficulties when trying to numerically evaluate variables of interest (stresses or displacements) at a point inside a three or two dimensional body S that is *"close"* to its boundary ∂S. In the sequel we limit ourselves in the particular case of two dimensional bodies.

The source of the abovementioned difficulty is the fact that, the traction kernel (in the standard Singular Integral Equation Method) becomes *"nearby singular"* or in other words the traction Kernel has a *"nearby singularity"* $O(1/r)$ where r is the distance between internal source point p *"close"* to ∂S and the nearest point Q on ∂S. In the hypersingular Integral Equation Method the traction Kernel has a *"nearby hypersingularity"*. Obviously the displacement kernel in the same situations is *"nearly weak singular"* $0(\ell nr)$ or *"nearly singular"* $0(1/r)$.

On the other hand, strong or nearby singularities appear not in the Kernel, but in the unknown distribution of displacements along ∂S, in the case of a concentrated force applied at a point p *"close"* to ∂S or on the boundary ∂S. These singularities can superimpose on the previous singularities of the Kernel. Therefore, in many cases more than one singularity coexists.

The use of standard quadrature procedures leads to a computational error which increases as the distance r decreases: this is the so-called *"boundary layer effect"*. BEM researchers have proposed various numerical methods for accurately computing such nearly singular, singular or hypersingular integrals.

Strong, weak or nearby singularities are treated in two ways. The first, which is dominated by the Sladek's work [2-3] i.e., tried to avoid singularities by regularizing the integral and the integral equations. Procedures trying to avoid singularities [4-8] move in the same lines. Some of these methods concern strong singularities but they may be expanded easily to nearby singularities. The one way to proceed with these integrals is based on quadrature rules for such integrals. We can distinguish procedures based on transformation which remedy the singularities [9-12] and on general quadrature rules [13-21] (special quadratures rules). In general, there is a prejudice concerning the last procedure. Many authors claim that special quadrature rules are very complicated and not very efficient. Nothing could be more wrong. The true is that special quadrature rules are able to give the exact result with only few integration points. On the other hand, the determination of the weights and the integration points is, in general, a simple procedure.

Mapping the integration interval of the boundary to the interval $[-1,1]$ the general form of an integral with nearby and strong singularities can be put in the following form

$$I(\varphi)=\int_{-1}^{1}\frac{w(x)f(x)}{R(x)}dx \tag{1}$$

with $R(x)$ a polynomial of degree $\ell=k+n$:

$$R(x)=(x-t)^k r(x),\quad r(x)=\prod_{i=1}^{n}(x-y_i),\quad y_i=x+i\varepsilon_i,\quad t\in(-1,1) \tag{2}$$

where k is an integer ε_i is small or zero and x_i is on $[-1,1]$ or very close to this interval. Thus, the integrand has apart from the poles at t, poles at the points y_i which could or could not belong to the integration interval. The weight function, $w(t)$, which is fixed positive and integrable on $[-1,1]$, is given by

$$w(x)=(1-x)^{\alpha}(1+x)^{\beta},\quad \alpha,\beta>-1, \tag{3}$$

The previous form (1) concerns not only line integrals but also surface integrals if the late can be reduced to one-dimensional integrals and then the form (1) can be applied.

Strong, weak and nearby singularities have already been confronted in the early '80s by our team [27-29]. More specifically, nearby singularities are confronted in two master theses [18, 19] in collaboration with the author and the results are incorporated in our codes for the solution of boundary integral equations.

In the present study, we have expanded the known **Gauss-integration rules** as to confront the above problem where many strong or weak singularities can coexist. For the sake of generality we have considered the most general case by supposing that, for some reasons, we need to evaluate the function at some pre-assigned nodes. For question of generality pre-assigned nodes are considered to coincide with one or both ends of integration so Radau, Lobatto and Kronrod formulas [18, 20, 21] are derived.

For the calculation of stresses defined by finite part integrals, integrals of the form (1) must be evaluated. In this paper we construct the polynomials $\varphi_n(x)$ which are orthogonal with respect to the weight $w(x)/R(x)$. Using these polynomials it is very easy to construct a Gauss type quadrature formula for integrals of the form of the relation (1). The paper is a generalization of my previous papers [20,21].

Sometimes, it is very useful to include in the quadrature formula a number of 'pre-assigned nodes' $[z_k]_{k=1}^{m}$, i.e. nodes which are given beforehand. Obviously there are also nodes which are free and may be chosen by any criterion we may desire. A special case of quadrature with pre-assigned nodes is the Radau quadrature where $m=1$ and τ_1 coincides with either of the ends of the integration interval, or the Lobatto quadrature where $m=2$ and τ_2 coincide with the ends a and b of the integration interval, etc. In order to confront this problem we have generalized our proof and we have included an arbitrary number of pre-assigned nodes.

2. SOME BASIC EQUATIONS

Let

$$\pi_n(x)=k_n x^n+k_n' x^{n-1}+\ldots \tag{3}$$

be a set of orthogonal polynomials [22] such that

$$\int_a^b w(x)\pi_n(x)\pi_m(x)dx=h_n\delta_{nm} \tag{4}$$

Obviously then

$$\int_a^b w(x)x^m\pi_n(x)dx=0 \tag{5}$$

These polynomials satisfy the three term recurrence formula

$$\pi_{n+1}(x)=(A_n x+B_n)\pi_n(x)-C_n\pi_{n-1}(x)\quad x>0 \tag{6}$$

$$A_n=\frac{k_{n+1}}{k_n},\quad C_n=\frac{A_n h_n}{A_{n-1}h_{n-1}},\quad C_0=0$$

$$B_n=A_n\left(\frac{k_{n+1}'k_n'}{k_{n+1}k_0}\right),\quad n>0 \tag{7}$$

with

$$\pi_{-1}=0,\quad \pi_0=1 \tag{8}$$

Definition 1. The functions $\psi_n(z)$, $n=0,1,2,\ldots$ *are defined cut along the real line from* -1 *to* 1 *by* [42]

$$\psi_n(z)=\frac{1}{2}\int_{-1}^{1}\frac{w(x)\pi_n(x)}{z-x}dx\,;\quad z\in[a,b],\quad n=0,1,2,... \tag{9}$$

The 'associated functions' $\psi_n(z)$ are continuous at (a,b) from the left and from the right, i.e. the boundary values

$$\psi_n^{\pm}(x)=\underset{\substack{y\to 0\\ y>0}}{\ell im}\,\psi_n(x\pm iy)$$

exist.

The following relations across the real line $(-1,1)$ are easily established by Plemelj formulas:

$$\psi_n^{+}(x)-\psi_n^{-}(x)=\pi i w(x)\pi_n(x),\quad -1<x<1 \tag{10a}$$

$$\psi_n^{+}(x)+\psi_n^{-}(x)=\ \oint_{-1}^{1}\frac{w(t)\pi_n(t)}{x-t}dt\ ,\quad -1<x<1 \tag{10b}$$

Definition 2. The functions $\psi_n(x),\ n=0,1,2,...$ *are defined in* $(-1,1)$ *by*

$$\psi_n(x)=\frac{1}{2}\int_{-1}^{1}\frac{w(t)\pi_n(t)}{x-t}dt\ ,\quad x\in(a,b),\quad n=0,1,2,... \tag{11}$$

From these definitions it is readily verified that the functions $\psi_n(z)$ satisfy the same recurrence relation (12) as $\pi_n(x)$:

$$\psi_{n+1}(z)=(A_n z+B_n)\psi_n(z)-C_n\psi_{n-1}(z),\quad n>0 \tag{12}$$

For $n=0$ the recurrence formula is slightly different, i.e. it takes the form

$$\psi_1(z)=(A_0 z+B_0)\psi_0(z)-A_0 h_0/2 \tag{13}$$

we denote by $\{z_j\}_{j=1}^{m}$ a set of pre-assigned points and by $\omega(x)$ the polynomial

$$\omega(x)=\prod_{j=1}^{m}(x-z_j) \tag{14}$$

and by

$$W(x)=w(x)\omega(x) \tag{15}$$

a composed weight function.

We propose now to find the polynomials $\varphi_n(x)$ of degree n in x such that

$$\oint_{-1}^{1}\frac{W(x)}{R(x)}\varphi_n(x)x^s dx=0\ ,\quad s<n\quad \text{or}\quad \oint_{-1}^{1}\frac{W(x)}{R(x)}\varphi_n(x)p_s(x)dx=0 \tag{16,17}$$

for every polynomial $p_s(x)$ of degree $s<n$. Taking

$$p_s(x)=R(x)x^{s-\ell} \tag{18}$$

we have

$$\oint_{-1}^{1}W(x)\varphi_n(x)x^{s-\ell}dx=0 \tag{19}$$

From the last relation we obtain

$$\oint_{-1}^{1}W(x)\varphi_n(x)\pi_{s-\ell}(x)dx=0\,.\ \text{ or }\ \int_{-1}^{1}w(x)[\omega(x)\varphi_n(x)]\pi_{s-\ell}(x)dx=0\,. \tag{20,21}$$

By expressing $[\omega(x)\varphi_n(x)]$ as a series of polynomials $\pi_i(x)$

$$\omega(x)\varphi_n(x)=\sum_{i=0}^{n+m}d_i\pi_i(x) \tag{22}$$

and substituting in (21) we obtain.

$$\omega(x)\varphi_n(x)=d_{n+m}\pi_{n+m}(x)+...+d_{n-\ell}\pi_{n-\ell}(x) \tag{23}$$

Taking into consideration z_j are roots of $\omega(x)$ we have

$$d_{n-\ell}\pi_{n-\ell}(z_j)+...+d_{n+m}\pi_{n+m}(z_j)=0\,;\quad j=1,2,...,m \tag{24}$$

Taking into consideration the above relation and the orthogonal condition we can prove the following theorem.

Theorem 1: The orthogonal polynomials $\varphi_n(x)$ over $[a,b]$ with respect to $W(x)/R(x)$ are given by

$$\varphi_n(x)=\frac{c_n}{\omega(x)}\begin{vmatrix} \psi_{n-\ell}(y_1) & \cdots & \psi_{n+m}(y_1) \\ \cdots & \cdots & \cdots \\ \psi_{n-\ell}(y_n) & \cdots & \psi_{n+m}(y_n) \\ \psi_{n-\ell}(t) & \cdots & \psi_{n+m}(t) \\ \psi^{(1)}_{n-\ell}(t) & \cdots & \psi^{(1)}_{n+m}(t) \\ \cdots & \cdots & \cdots \\ \psi^{(k-1)}_{n-\ell}(t) & \cdots & \psi^{(k-1)}_{n+m}(t) \\ \pi_{n-\ell}(z_1) & \cdots & \pi_{n+m}(z_1) \\ \cdots & \cdots & \cdots \\ \pi_{n-\ell}(z_m) & \cdots & \pi_{n+m}(z_m) \\ \pi_{n-\ell}(x) & \cdots & \pi_{n+m}(x) \end{vmatrix} \tag{25}$$

Proof

Taking into consideration (17) we shall prove the correctness of (25).

As a first step we select as $p_s(x)$ the following m polynomials of degree $(\ell-1)$

$$p_{\ell-1}(x)=(x-t)^k r_j(x), \quad j=1,2,...,m \tag{26}$$

with

$$r_j(x)=r(x)/(x-y_j) \tag{27}$$

We put these polynomials in (17) and we substitute φ_n by the relation (25), and we obtain.

$$\int_a^b \frac{W(x)}{R(x)}\varphi_n(x)p_{\ell-1}(x)dx=\frac{c_n}{\omega(x)}\begin{vmatrix} \psi_{n-\ell}(y_1) & \cdots & \psi_{n+m}(y_1) \\ \cdots & \cdots & \cdots \\ \psi_{n-\ell}(y_n) & \cdots & \psi_{n+m}(y_n) \\ \psi_{n-\ell}(t) & \cdots & \psi_{n+m}(t) \\ \psi^{(1)}_{n-\ell}(t) & \cdots & \psi^{(1)}_{n+m}(t) \\ \cdots & \cdots & \cdots \\ \psi^{(k-1)}_{n-\ell}(t) & \cdots & \psi^{(k-1)}_{n+m}(t) \\ \pi_{n-\ell}(z_1) & \cdots & \pi_{n+m}(z_1) \\ \cdots & \cdots & \cdots \\ \pi_{n-\ell}(z_m) & \cdots & \pi_{n+m}(z_m) \\ \psi_{n-\ell}(y_j) & \cdots & \psi_{n+m}(y_j) \end{vmatrix} \tag{28}$$

Taking into consideration that the last line of the determinant in the right hand side coincides with the j-th line, the right hand side is equal to zero.

Let now select $p_s(x)$ as follows.

$$p_{\ell-q}(x)=(x-t)^{k-q}r(x), \quad q\le k \tag{29}$$

We follow the previous procedure, and we obtain

$$\int_a^b \frac{W(x)}{R(x)}\varphi_n(x)p_{\ell-x}(x)dx \frac{c_n(-1)^{r-1}(r-1)!}{\omega(x)} \begin{vmatrix} \psi_{n-\ell}(y_1) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \psi_{n+m}(y_1) \\ \cdot & & & & & & & & & \cdot \\ \psi_{n-\ell}(y_n) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \psi_{n+m}(y_n) \\ \psi_{n-\ell}(t) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \psi_{n+m}(t) \\ \psi'_{n-\ell}(t) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \psi'_{n+m}(t) \\ \cdot & & & & & & & & & \cdot \\ \psi^{(k-1)}_{n-\ell}(t) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \psi^{(k-1)}_{n+m}(t) \\ \pi_{n-\ell}(z_1) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \pi_{n+m}(z_1) \\ \cdot & & & & & & & & & \cdot \\ \pi_{n-\ell}(z_m) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \pi_{n+m}(z_m) \\ \psi^{(r-1)}_{n-\ell}(t) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \psi^{(r)}_{(n+m)}(t) \end{vmatrix} \tag{30}$$

The right hand of (18) is for the same reason is equal to zero. Finally $\varphi_{n+m}(x)$ must also have roots at the points z_j, $j=1,2,...,k$, which is obviously satisfied. So the theorem is proven.

Obviously this result can be generalized in order to obtain the corresponding orthogonal polynomials in the case where

$$r(x)=\prod_{i=1}^{n}(x-y_i)^{k_i} \tag{31}$$

3. THE GENERALIZED QUADRATURE FORMULA

Let us give some necessary definitions:

We denote by D a simply connected domain containing $(-1,1)$ in its interior.

We denote by $f(z)$ the analytic continuation of $f(t)$ into D.

Let C a simple contour inside D, presenting angular points at -1 and 1 with angles α and β respectively. These angular points are due to the behavior of $\varphi(z)$.

Let $\{t_j\}_{j=1}^{n}$ the roots of $\varphi_n(x)$. These roots belong or not to the integration interval. These points as well the pre-assigned points $\{z_k\}_{k=1}^{m}$ will be used as interpolation points, i.e.

$$f(\tau_j)=f_{n+m}(\tau_j), \quad j=1,2,...,(n+m) \tag{32}$$

where $f_{n+m}(z)$ is a complex polynomial of degree $(n+m-1)$ that interpolates $f(z)$ at the points

$$\tau_j=\begin{cases} t_j\ ; & 1\le j\le n \\ z_k\ ; & j=n+k\le n+m \end{cases} \tag{33}$$

We denote by

$$\varphi_n(x)=\prod_{j=1}^{n}(x-t_j), \quad \omega(x)=\prod_{k=1}^{m}(x-z_k), \quad \Pi_{n+m}(x)=\varphi_n(x)\omega(x), \quad W(x)=w(x)\omega(x). \tag{34}$$

Using the interpolation formula (32), and taking into consideration Lagrange procedure, we have

$$f(t)=f_{n+m}(t)+R_{n+m}(t) \tag{35}$$

where

$$f_{n+m}(x)=\sum_{j=1}^{n}\frac{\Pi_{n+m}(x)}{(x-t_j)\omega(t)\varphi'_n(t_j)}\varphi(t_j)+\sum_{k=1}^{m}\frac{\Pi_{n+m}(x)}{(x-z_k)\omega'(z_k)\varphi_n(z_k)}\varphi(z_k) \tag{36}$$

or in more compact form

$$f_{n+m}(x)=\sum_{j=1}^{n+m}\frac{\Pi_{n+m}(x)}{(x-\tau_j)\Pi'_{n+m}(\tau_j)}\varphi(\tau_j) \tag{37}$$

and

$$R_{n+m}(x)=\frac{\Pi_{n+m}(x)}{2\pi i}\oint_C \frac{f(z)}{(z-x)\Pi_{n+m}(z)}dz \tag{38}$$

Multiplying both sides of (7) by $w(x)/R(x)$ and integrating on $[-1,1]$ we get

$$\oint_{-1}^{1}\frac{w(x)}{R(x)}f(x)dx = Q_{n+m}(f)+E_{n+m}(f) \tag{39a}$$

where

$$Q_{n+m}(f)=\sum_{j=1}^{n+m}\Lambda_j f(\tau_j) \tag{39b}$$

is the quadrature formula,

$$\Lambda_j=\frac{1}{\Pi'_{n+m}(\tau_j)}\int_{-1}^{1}\frac{w(x)\Pi_{n+m}(x)}{(x-\tau_j)R(x)}dx=\frac{1}{\Pi'_{n+m}(\tau_j)}\int_{-1}^{1}\frac{W(x)\varphi_n(x)}{(x-\tau_j)R(x)}dx \tag{40}$$

are the weights of the quadrature formula, and

$$E_{n+m}(f)=\frac{1}{2\pi i}\oint_C\frac{f(z)}{\Pi_{n+m}(z)}\left(\int_{-1}^{1}\frac{W(x)\varphi_n(x)}{(z-x)R(x)}dx\right)dz \tag{41}$$

is the error

Taking into consideration are

$$\Psi_n(z)=\frac{1}{2}\oint_{-1}^{1}\frac{w(t)\Pi_{n+m}(t)}{z-t}dt=\frac{1}{2}\int_{-1}^{1}\frac{W(t)\pi_n(t)}{z-t}dt\,;\quad z\in[-1,1] \tag{42}$$

and using asymptotic expansion we can prove that $E_{2n+m}(f)=0$

From (27), we conclude that $E_{n+m}(\varphi)=0$ for every $f(t)\in\wp_{n+m-1}$, where $\wp_{2n+m-1}$ is the class of polynomials of degree $\le n+m-1$.

In the particular case where the pre-assigned nodes coincides with both ends of the integration interval i.e. $\Omega(t)=t^2-1$, the quadrature formula (39b) taking also into consideration Plemelj formula becomes

$$Q_{n+2}=\frac{\alpha}{2\pi}\Lambda_{j+1}f(-1)+\frac{\beta}{2\pi}\Lambda_{j+2}f(1)+\sum_{j=1}^{n}\Lambda_j f(t_j) \tag{50}$$

where Λ_{j+1} and Λ_{j+2} can be obtained from relations (40) for $\tau_{j+1}=-1$ and $\tau_{j+2}=1$ respectively. If C is smooth (contour C without angular points) then $\alpha/2\pi=\beta/2\pi=1/2$ and we found the most known form of Lobatto formula. Obviously the pre-assigned node could coincide with one of the ends, and then one of the first terms in (50) is equal to zero. The last formula goes with the name of Radau

REFERENCES

[1] Cruse TA (1969): Numerical solutions in three dimensional elastostatics: *Int. J. Solids Structures* **5**, 1259-1274.

[2] Sladek N., Sladek J., Tanaka M. (1993): Regularization of hypersingular and nearly singular integrals in the potential theory and elasticity. *Int. J. Numer. Methods Eng.* **36**, 1609-1628.

[3] Liu Y.J., Rudolphi T.J. (1999): New identities for fundamental solutions and their applications to non-singular boundary element formulations. *Comput. Mech.* **24**, 286-292.

[4] Mukherjee S., Chati M.K., Shi X.L. (2000): Evaluation of nearly singular integrals in boundary element contour and node methods for three-dimensional linear elasticity. *Int. J. Solids Struct.*, **37**, 7633-7654.

[5] Telles J.C.F. (1987): A self-adaptive co-ordinate transformation for efficient numerical evaluation of general boundary element integrals. *Int. J. Numer. Methods Eng.* **24**, 959-973.

[6] Krishnasamy G., Schmerr L.W., Rudolphi T.J., Rizzo F.J. (1990): Hypersingular boundary integral equations: some applications in acoustic and elastic waves scattering. *ASME J. Appl. Mech.* **57**, 404-414.

[7] Dumont N.A. (1994): On the efficient numerical evaluation of integrals with complex singularity poles. *Eng. Anal. Boundary Elements* **13**, 155-168.

[8] Yun, B.I., (2003): A composite transformation for numerical integration of singular integrals in the BEM, *International Journal for Numerical Methods in Engineering*, **57**, 1883-1898.

[9] Kim, P. and Choi, U.J. (2003): Two trigonometric quadrature formulas for evaluating hypersingular integrals, *Int. J. Numer. Methods Eng.*, **56**, 469-486.

[10] Huang, Q. and Cruse, T.A. (1993): Some notes as singular integral techniques in Boundary Element Analysis. *Int. J. Numer. Methods Eng.* **36**, 2643-2651.
[11] Johnston P.R. (1999): Application of sigmoidal transformations to weakly singular and nearly-singular boundary element integrals. *Int. J. Numer. Methods Eng.* **45**, 1333-1348.
[12] Johnston P.R., Elliott, D., (2000): Error estimation of quadrature rules for evaluating singular integrals in boundary element problems. *Int. J. Numer. Methods Eng.,* **48**, 949-962.
[13] Ma, H., Kamiya, N., (2002): A general algorithm for the numerical evaluation of nearly singular boundary integrals of various orders for two- and three-dimensional elasticity, *Computational Mechanics* **29**, 277-288.
[14] Korneichuk, A.A. (1964): Quadrature formulae for singular integrals. *Zh Vychisl. Mat. I Mat. Fiz.* 4/4, suppl. 64-74 (in Russian).
[15] Paget, D.E.; Elliott, D. (1972) : An algorithm for the numerical evaluation of certain Cauchy principal value integrals. *Numer. Math.* **19**, 373-385.
[16] Chawla, M.M., Ramakrishnan, T.R. (1974): Modified Gauss-Jacobi quadrature formulas for the numerical evaluation of Cauchy type singular integrals. *BIT* **14**, 14-21.
[17] Tsamasphyros, G., Theocaris, P.S. (1976b): Sur une methode generale de quadrature pour des integrals du type Cauchy. Symp. Appl. Mathem., Salonica August 1976, also (1980) *Rev. Roum. Sci. Techn., Ser. Mec. Appl.* **25**, 839-856.
[18] Ninis, T. (1980) : Factors influencing the solution of singular integral equations, *Master Thesis* (in Greek), National Technical University of Athens.
[19] Strouboulis, T. (1980): Problems in the solution of Singular Integral Equations, *Master Thesis* (in Greek), National Technical University of Athens.
[20] Tsamasphyros, G., Dimou, G. (1989) : A Gauss quadrature formula for Cauchy type integrals, *Computational Mechanics* **4**, 137-148.
[21] Tsamasphyros, G. and Dimou G. (1990), Gauss quadrature rules for finite part integrals, *Int. J. Numer. Methods Eng.* **30**, 13-26.
[22] Szego, G. (1959): *Orthogonal polynomials*. New York, American Mathematical Soc.

General papers

APPLICATIONS OF COMPUTATIONAL GEOMECHANICS IN PETROLEUM ENGINEERING

P. Papanastasiou[*] and A. Zervos[†]

[*]Department of Civil and Environmental Engineering
University of Cyprus
P.O.BOX 20537, 1678 Nicosia, Cyprus
e-mail: panospap@ucy.ac.cy, web page: http://www.eng.ucy.ac.cy/CEE/CEEHome.htm

[†] School of Civil Engineering and the Environment, University of Southampton, Southampton, UK

Keywords: Petroleum geomechanics, Finite element analysis, Wellbore stability, Sand prediction

Abstract. *This paper presents examples on the use of computational mechanics in petroleum geomechanics for the solution of research and advanced practical problems. We present finite element solutions for the problems of wellbore stability, stability of multilateral junctions, sand prediction analysis and in situ stress determination.*

1 INTRODUCTION

In the last two decades the petroleum industry has witnessed what can be called the 'geomechanics revolution'. Geomechanics has become a regular consideration in oil field development from oil exploration to production. Evidence of the significance of geomechanics is drawn from field cases in the areas of drilling, production and reservoir management regarding the commercial value of geomechanics for the operator's perspective. Within the service industry, geomechanics became the fastest growing commercial area for technical investment. Among the main drivers for the promotion and advances of petroleum geomechanics are the high rig daily rates in deep water ($500,000/day in the deep water in the Gulf of Mexico and in offshore West Africa), the drilling in harsh environments such as tectonic fields, salt-domes, high-pressure high-temperature fields, and the drilling of more horizontal and extended reach wells spanned with multilateral junctions. Environmental concern and restrictions for the disposal of contaminated cuttings and produced water promoted new applications for reinjection and hydraulic fracturing. The demand for higher hydrocarbon recovery, leading to extreme reservoir depletions, causes other problems such as surface subsidence and wellbore collapse, and opened a new area called 'reservoir geomechanics'. Like in any other field, advances in the information technology such as software, visualization tools, and the internet have their share in the promotion of petroleum geomechanics.

In this paper we will focus on the contribution of computational geomechanics in the solution of practical problems in petroleum engineering. We will demonstrate this contribution through solution of problems in major areas of petroleum geomechanics. In section 1 we will present how the finite element method is used for calculating the optimum mud density for drilling deviated wellbores. In section 2 we will extend the same analysis for examining the stability of multilateral junctions. In section 3 we will use a 3-D finite element model for determining the optimum wellbore pressure for sand-free production of hydrocarbons. For the solution of problems in petroleum geomechanics the knowledge of insitu stresses is needed as input parameters. We will show in section 4 examples of finite element analysis for the determination of insitu stresses near complex geological structures. We recognize the work done and published in numerous papers on petroleum geomechanics by many researchers and practitioners over the last two decades. A good collection of such papers can be found in the EUROCK/SPE 1998 and SPE/ISRM 2002 proceedings and in the 2004 special issue of International Journal of Geomechanics on petroleum geomechanics.

2 ELASTOPLASTIC FINITE ELEMENT ANALYSIS OF INCLINED WELLBORES

Modern wells, with higher deviations and greater step-outs, are increasingly susceptible to wellbore instability problems. The development of extended reach and horizontal drilling in the last decades, followed recently by another drilling technique, the multilateral, makes the well trajectory more complex. Mud-pressure modelling for complex well trajectories required less conservative solutions. Two-dimensional plane strain solutions can be applied to analyse the stability of a wellbore when the wellbore axis is nearly parallel to the direction of one of the insitu stresses. For example, this is very often the case for a vertical or a horizontal wellbore. In the general case of an inclined hole, however, a more elaborate analysis is required, since the problem is 3D and all components of the stress tensor are involved. Nevertheless, the construction of

engineering models can be simplified and computational time can be reduced significantly by simplifying the geometry taking into account the following two points:

(a) The analysis can be carried out in a local coordinate system, whose z-axis is parallel to the wellbore axis, the x-axis is chosen to be parallel to the lower most radial direction of the wellbore, and the y-axis is horizontal. Then, the insitu stresses in this local frame consist of three normal and three shear components. To model wellbores of different orientations, only the decomposition of the insitu stress field in the local frame needs to be changed. Thus, only one finite element mesh is required, with a change of the applied stresses according to wellbore orientation. The transformed insitu stresses are applied in the form of an initial stress condition.

(b) It is expected that there is no significant variation in the deformation of the cross-sections perpendicular to the wellbores axis caused by the insitu stresses. The differences in the deformation of the cross-sections along the z-axis are mainly due to the changes of the in plane geometries. The nodal displacements of the different cross-sections can be de-coupled resulting in local displacements independent of the position of the node along the wellbore axis. Thus, no significant variation of the displacement field, (u_x, u_y, u_z), is expected parallel to the hole z-axis.

These conditions can be easily implemented in a 2-D finite element code where the third equilibrium equation (i.e. in the z-direction) replaces the plane-strain condition. Thus, the solution-domain is a plane perpendicular to the hole axis with the three components of the displacement being the primary unknowns. This formulation allows out-of-plane warping of a thin layer without change of its thickness. To meet these requirements we developed a finite element code for elastoplastic rocks which can be used for mud-pressure design of a general inclined wellbore. The details of this implementation were presented for an in house developed code[1] and in an existing commercial code[2,3] .

Inputs required by the model are the rock mechanical parameters and insitu stresses. The developed code was tested by carrying out optimum mud-pressure calculations for drilling a wellbore in an oil field offshore of East Canada. The model parameters are derived from triaxial test data performed on core samples. In cases where core data are not available, a default model can be used based on sonic data and calibration functions for calibrating the elastoplastic model. In the following we will present examples of mud density calculations in a silty shale in the overburden.

The material parameters and functions for Mohr-Coulomb elastoplastic model are obtained from triaxial compression tests. The tested cores were recovered from different depths and correspond to different lithologies. In this study we used only the data for the lithology characterized as a silty shale tested perpendicular to the bedding planes. The elastic constants are approximately: E = 8833 MPa and ν= 0.073. The experimental data were fitted by the hyperbolic law:

$$\sigma_e = \sigma_e^0 + \frac{\varepsilon^p}{C_1 + C_2 \varepsilon^p} \tag{1}$$

where $\sigma_e^o = 41.25\ MPa$ is the value of the effective stress at initial yield and the calibration constants are C_1 = 1.225 10^{-05}/MPa and C_2 = 1.96825 10^{-05}/MPa. The angle of internal friction was found to be $\varphi = 27.23^o$. The data supported a non-associative flow rule with dilation angle $\psi = 0^o$. In practical applications failure in geomechanics is defined when the state of stress or the plastic strain reach a critical value which is usually derived from triaxial compression tests.

The computations were carried out at the fixed depth of 8150 ft. At this depth the in situ effective stresses are approximately, P_p = 25.67 MPa, S_h = 18.8 MPa S_H = 23.4 MPa and S_v = 37 MPa. Other parameters required by the FEM model are the wellbore orientation with respect to the insitu stress field. The wellbore is drilled with inclination $i = 31^o$ from vertical and azimuth $a = 250^o$ measured counter-clockwise from North. The maximum horizontal stress is acting along the East-West direction. The angle between S_H and wellbore azimuth is 20^o.

Figure 1 shows the developed dimensionless hole-closure with decreasing mud-pressure. In the same graph we marked the level of mud-pressure for which failure takes place assuming (a) an elastic-brittle model and (b) an elastoplastic model. The elastic-brittle model predicts failure at p_{mud} = 26 MPa whereas the elastoplastic model is less conservative predicting failure at p_{mud} = 10.6 MPa. The hole is deformed like an ellipse with maximum hole closure at 90^o from the breakout direction. In order to appreciate the warping of the inclined wellbore we plotted in Figure 2a the hole cross-section distorted with the amplified calculated displacement parallel to the hole axis, u_z. Figure 2b shows the developed plastic strain. As already mentioned, failure is assumed to take place at p_{mud} = 10.6 MPa when the plastic strain at the borehole wall reaches the critical value ε_p = 0.003.

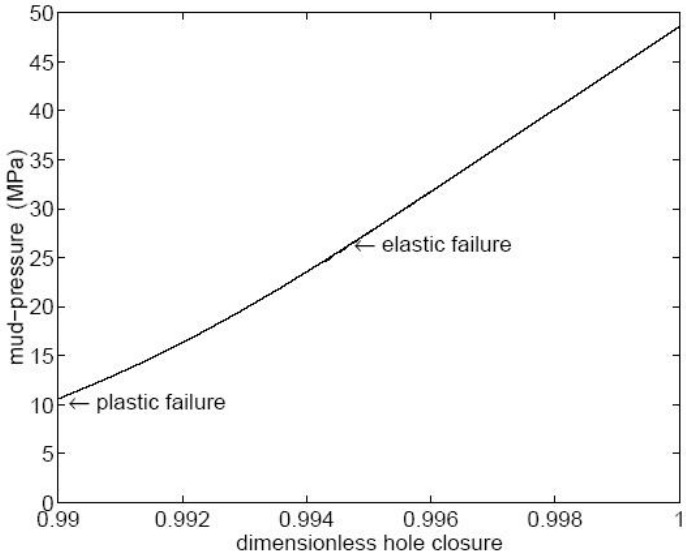


Figure 1. Mud-pressure vs dimension-less hole closure

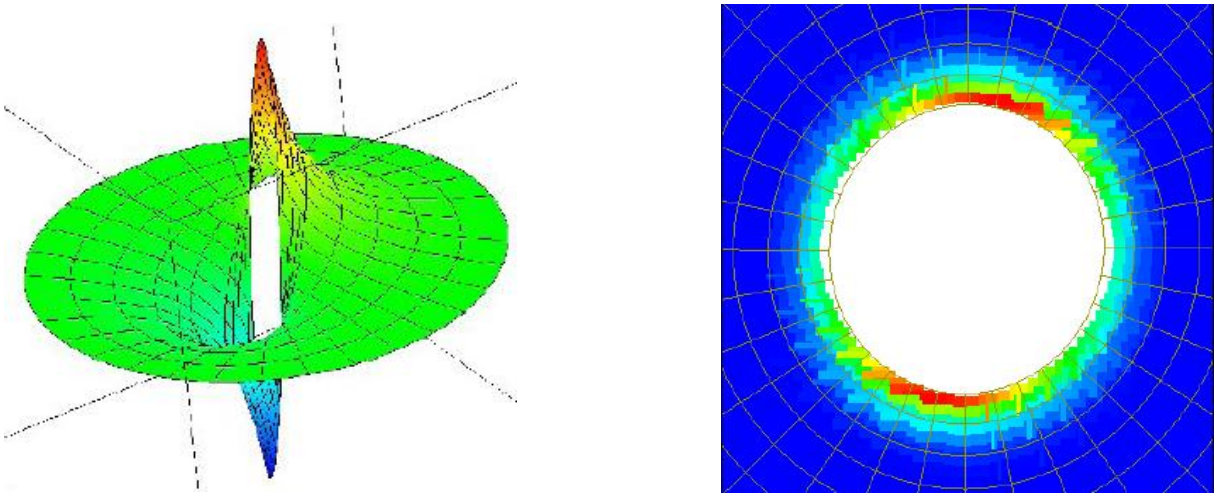


Figure 2. (a) Contours of the axial displacement showing hole warping and (b) contours of the plastic strain.

3 STABILITY OF MULTILATERAL JUNCTION

Drilling inclined wells through producing strata can greatly improve reservoir drainage and hydrocarbon recovery. The horizontal sections are accessed through multiple inclined wells drilled from a relatively small area in many or all directions, something that allows better exploitation of offshore platforms and land rigs that are under economic and environmental restrictions. Drilling inclined and horizontal wells, though, is more difficult and more expensive, due to wellbore instabilities. A particular area of concern is the integrity of the rock near a multilateral (M-L) junction. The junction is the region where a second wellbore (lateral) takes off from the main wellbore (parent). In M-L levels 1 and 2 the rock at the junction is not supported mechanically with cemented casing, so the integrity of the rock around the area of two intersecting tubes becomes very important in terms of stability.

A software tool was developed in Schlumberger for assessing the integrity of rock surrounding a multilateral junction. The tool was developed using finite element analysis and a graphical-user interface for providing the input data and visualizing the results. The build angle between the lateral hole and parent hole is usually small, so, the axis of the lateral is nearly close to the z-axis. Therefore, as explained in the section before, the analysis can be based on a generalized plane strain formulation that is carried out in cross-sections in succession, perpendicular to the parent-hole axis. The use of a 2D analysis is not valid in the area very close to the junction due to strain concentrations. However, the drilling tool will damage the rock in that area at the span off of the

lateral. Therefore, for practical reasons the analysis can be simplified into 2D with a good accuracy.

The data are given through a graphical interface, shown in Figure 4 for a particular example. The data required to define the geometry are the radii of the holes, the inclination and azimuth of the parent hole, build angle between lateral and parent holes and the position of the lateral with respect to the lowest point on the parent hole. The rock parameters are defined by the elastic modulus and Poisson's ratio, the friction angle and the rock strength (UCS). The input stresses in the model are the vertical stress and the two horizontal stresses, the reservoir pressure and the well pressure. The direction of the maximum horizontal stress with respect to North is also needed.

Finally, a Biot constant is needed to define the relation between total and effective insitu stresses. The example presented here is based on elastic/brittle analysis carried out for testing the model by comparing its predictions with the experimental data from true triaxial tests performed on large blocks at Lille University[4]. For the rock samples it was determined from calibration tests that the elastic modulus is E=22500 MPa and the Poisson´s ratio is ν=0.2. The estimated Mohr-Coulomb parameters were 8.5 MPa for the cohesion and 28.5 degrees for the friction angle. During testing, the three applied stresses on the sides of the blocks were increased equally. Failure around the holes in tests was observed after the applied isotropic stresses reached the value of 27 MPa. Figure 5 shows that the model predicts maximum compressive stress around the junction area (Fig.5b).

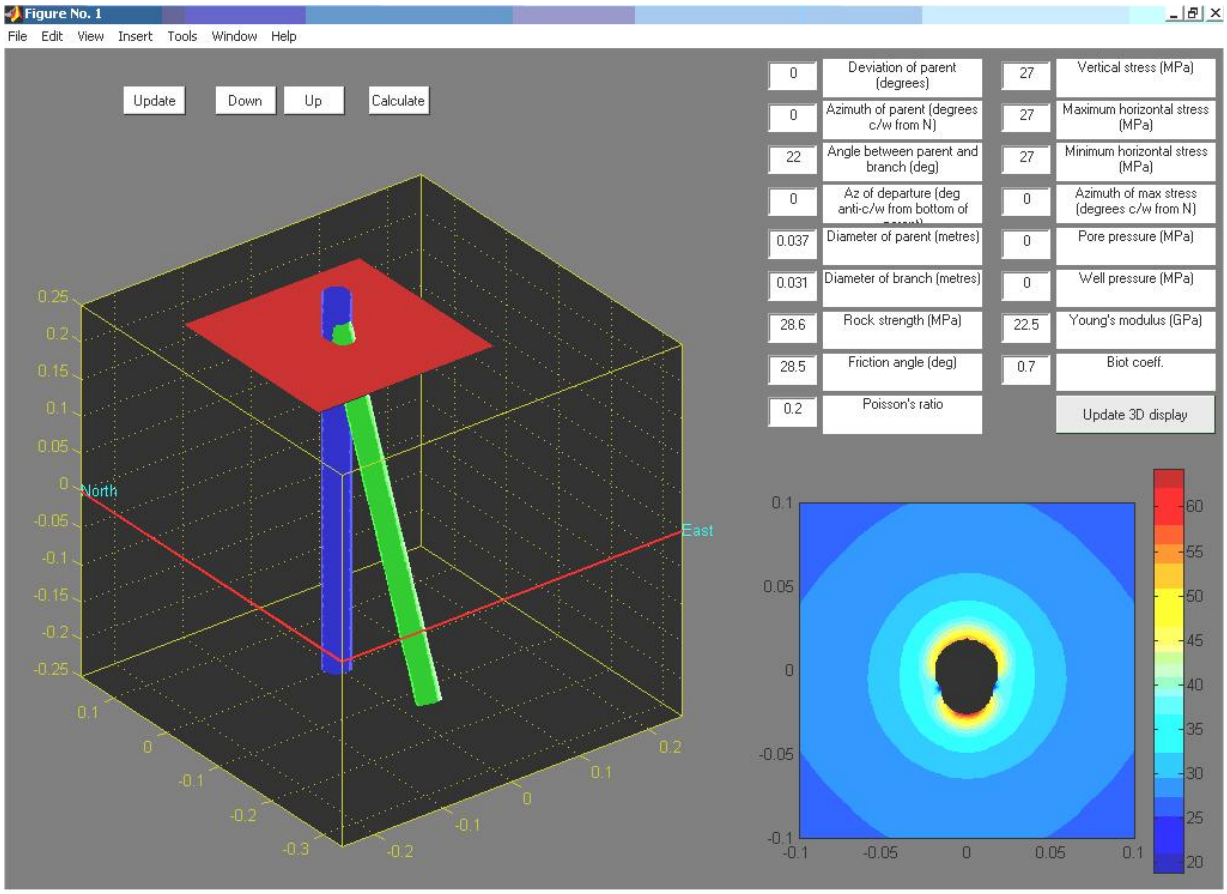


Figure 4. Graphical interface for the analysis of a multilateral junction

The developed tool is useful to drilling engineers to decide in which formation, in which azimuth and which deviation to drill a stable lateral. The completion engineer can use the tool to decide, a) where to place a stable junction, e.g. in the reservoir or in the overburden, b) what level of M-L junction is needed (mechanically supported or not) and c) if the junction is drilled in the reservoir section, at which draw-down and depletion pressures it may become unstable.

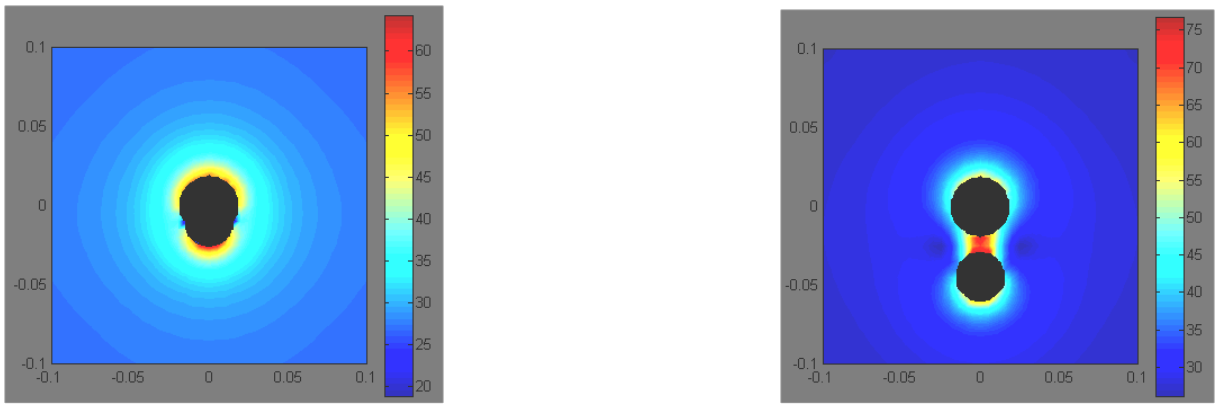


Figure 5. Model predictions of maximum compressive stress at 27 MPa stresses at two different points along the junction

4 SAND PREDICTION ANALYSIS

A major problem of concern to the oil industry is sand production. Production of hydrocarbons at high flow-rates induces formation failure and subsequent sand production. In order to understand the physical mechanisms involved in the sanding problem sanding experiments were carried out on thick-walled cylinders under loading and flow conditions[5] . It was found that in weakly consolidated rocks sanding takes place in two stages: a) the applied stresses fail the rock near the hole and b) if the flow rates are high enough, the loose sand grains are transported. The sand production experiments show clearly that stress induced shear failure occurs before sand production. During field production failure takes place mainly due to increase of effective stresses after depletion and secondary due to increase of draw-down pressure. Petroleum engineers need sanding predictor models that can be used in optimizing production rates from wells in formations that are mechanically weak and for which rock failure and production of solids (mainly sand) are common problems.

Emphasis in sanding modelling was put on predicting and avoiding rock failure; If the conditions for rock failure are avoided no sanding is expected to occur. In the example below we will describe a sand prediction model based on a 3-D poroelastoplastic finite element analysis of a perforated wellbore. All the involved processes: wellbore drilling, casing and cementing, perforation creation, production and reservoir depletion can be simulated by the model. The criterion for sand production is based on predicting rock failure when the plastic strain, which is calculated from a steady state coupled stress analysis of production simulation, reaches a critical threshold. The model can provide directly the critical drawdown for avoiding rock failure near the perforations at different orientations. For this FE modelling, we used the commercial code Abaqus[6]; a similar model was also used in reference [7].

In this example the computations were carried out at the reservoir depth of 3000 m. At this depth the pore pressure and the total in situ stresses are approximately: P_p = 30 MPa, S_h = 37.5 MPa, S_H = 48 MPa and S_v = 60 MPa. The rock parameters required by Mohr-Coulomb elastoplastic model were measured or estimated from core tests. For the elastic constants we assumed the following values: E = 5000 MPa, ν = 0.3. For simplicity or sometimes constrained by the availability of triaxial rock deformation data, we assume an associated elastic-perfectly plastic material behaviour. That means the dilation angle is equal to the friction angle $\psi = \varphi$ and that no hardening takes place with plastic deformation. Under these assumptions, we used the following plastic material parameters representative of a weak rock: $\varphi = 30^o$, $\psi = 30^o$, c = 3 MPa, where c is the rock cohesion. The equivalent uniaxial compressive strength is UCS = 10.4 MPa. A steady state poroelastoplastic stress analysis will be carried out to simulate the production of hydrocarbons from a perforated wellbore. This analysis requires the permeability of the formation k to be specified. Abaqus requires the effective permeability coefficient in m/sec which incorporates viscosity. In addition, the fluid density, ρ_f and fluid bulk modulus, K_f are explicitly required. Typical values for these data are: $k = 2.539\ 10^{-5}$ m/sec, $\rho_f = 786.58$ Kg/m^3, $K_f = 861.85$ MPa.

The diameter of the wellbore is D=20 cm and the diameter of perforations is d_b = 2.5 cm near the base and d_b = 1.0 cm near the tip. The perforation penetrates 20 cm into the rock formation. Due to symmetry, we modelled only a quarter of the wellbore assuming also that there is no perforations interaction in the direction parallel to the wellbore. Figure 6 shows the finite element meshes (a) after drilling of the wellbore and (b) after completion of the perforations. Both processes were modeled by removal of the elements in each stage. Production of hydrocarbon from perforations is simulated by reducing the pressure in the perforation channels below the formation pressure in different production steps. After every decrease of production pressure a steady-state analysis is carried out to reach equilibrium stress conditions. Figure 7 shows the development of plastic strain on two perforations with 90^o phasing, vs decreasing production pressure in the wellbore. The red line (bottom)

corresponds to the perforation which is parallel to the minimum insitu stress and the green line (top) corresponds to the perforation which is parallel to the maximum insitu stress. The results show that failure will take place first near the perforation which is parallel to the maximum insitu stress (green line). For these particular data, the maximum value of the plastic strain is approximately 13% on one perforation and 8% on the other perforation after their creation and before the start of the production. The maximum value increases to 24% on one perforation and to 21.5% on the other perforation when the production pressure in the wellbore drops to zero. The critical wellbore pressure for a sand free production is calculated as follows: If we assume that the critical plastic strain for rock failure is defined to be 18% (e.g from triaxial calibration tests) then the production pressure should be kept above 12 MPa for avoiding sanding (Figure 7).

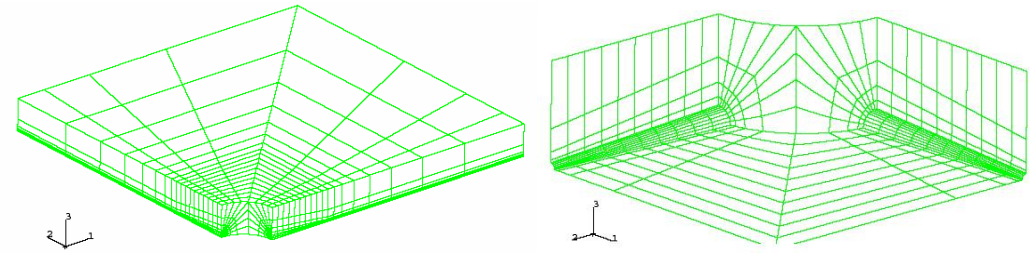


Figure 6. (a) Finite element meshes (a) after wellbore drilling and (b) after perforation completion

Figure 8a shows the contours of the plastic strain near the perforations. The maximum plastic strain occurs within one wellbore radius distance. Similar results are also shown by the flow rate distribution on the perforation faces (Figure 8b). The largest pressure gradients are near the tip of the perforations and the highest flow rates are calculated near the tip of the perforations.

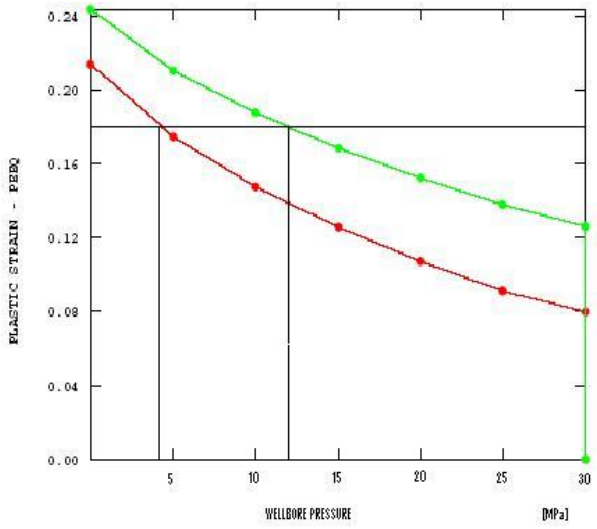


Figure 7: The development of plastic strain at perforations with the decrease of wellbore pressure

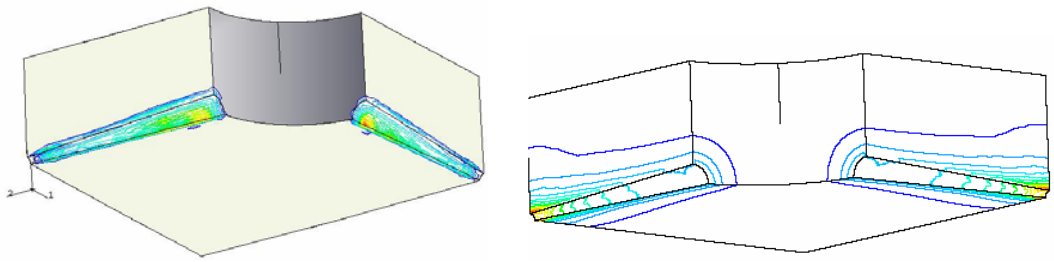


Figure 8. (a) Contours of plastic strain for zero wellbore pressure; failure will take place near the red areas (b) Flow rates near the perforations; maximum flow rates are shown near the perforations tip.

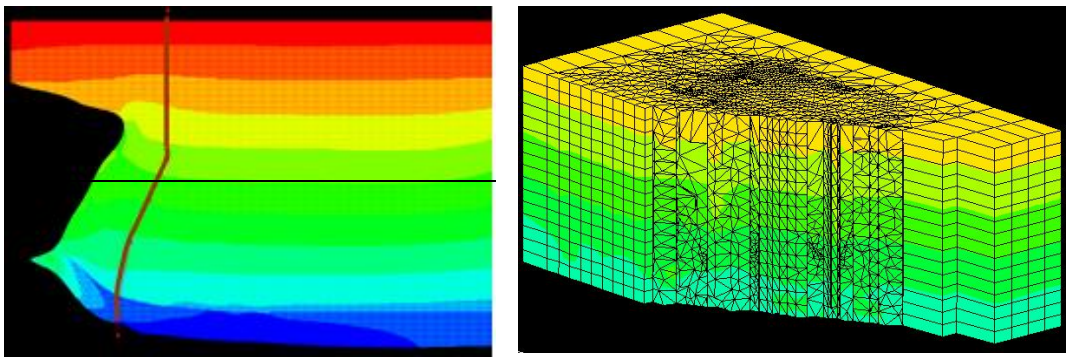
5 INSITU STRESS DETERMINATION NEAR COMPLEX GEOLOGICAL STRUCTURES

From the examples presented above it is made clear that the knowledge of the insitu stresses is needed in the geomechanical prediction models. Other applications in petroleum engineering where the knowledge of the insitu stresses is needed is in the design of Hydraulic Fracturing. The hydraulic fractures are aligned perpendicular to the minimum insitu stress and the hydraulic pressure needed for creating these fractures, called the net-pressure, must exceed the minimum insitu stress. Another area where the knowledge of the insitu stresses is required is in the Reservoir Geomechanics. In this area the reservoir simulators are coupled with stress analysis to account for problems such as reservoir compaction and surface subsidence, casing collapse, fault activation and other.

Engineers can get reasonable estimates of the insitu stresses in cases where the rock layers are nearly horizontal and are extended to large distances. In such cases, the vertical stress is determined from the overburden weight. For the determination of the horizontal stresses, which is less accurate, one has to assume that no lateral deformation has taken place and the rock was deformed according to a constitutive relation such as elasticity.

In this section we present two examples where the finite element analysis was used to obtain an estimate of the insitu stresses near complex geological structures. In Figure 9a the analysis was carried out for determining the insitu stresses near a salt diapir in the N. Sea. From the seismic data the problem was defined to be approximately axisymmetric. For such analysis it is necessary to have measurements of the insitu stresses (usually the minimum insitu stress measured with hydraulic fracturing test) at some points in order to constrain the problem. By varying parameters related to rock properties, tectonic loads or boundary conditions the process attempts to match the predictions with the available measurements in some points in order to get good estimates of the insitu stresses elsewhere. The knowledge of the insitu stresses near a salt diapir is valuable for optimizing the position of the offshore platforms and for choosing the well trajectories (Figure 9a).

Figure 9b shows the finite element mesh of a geomechanical model which was used for estimating the insitu stresses in a highly fractured reservoir and in the overburden. The model of figure 9b contains 20 major faults inserted as discontinuities. Other minor faults were ignored in the analysis. The faults may activate during reservoir depletion and production causing casing collapse and severe damage in the wells. With the assistance of such modelling different scenarios on how to develop an oil-field can be examined in order to minimize the problems that may be encountered later.



(a) (b)

Figure 9. Finite element calculations of the vertical insitu stress (a) near a salt diapir and (b) in a fractured reservoir

For both problems static analyses were carried out. The elements used were linear strain triangles (6 nodes) for 2D, and tri-linear tetrahedra for 3D (4 nodes). The Constitutive laws used were elastoplastic Mohr-Coulomb for the salt diapir and the rock strata (2D and 3D). The material parameters came from a combination of logs/seismic surveys. The loading and boundary conditions varied examining different scenarios applying either the far-field insitu stresses at the far boundaries, or prescribing fixed displacements at the far boundaries to create the stress field. This was necessary, as there were very few points at which the model could be validated (using leak-off tests at 2-3 points for a whole region km x km in size). Also, the way the salt diapir was modelled varied. In some scenarios was modelled as a Mohr-Coulomb solid, in others it was left out and its influence substituted with a pressure as if it were a fluid

6 CONCLUSIONS

Meeting unpredicted problems in petroleum geomechanics such as wellbore instability and sanding is almost inevitable, attributed mainly to the uncertainties on the values of dominant parameters mainly the insitu stresses and rock strength. Recently, this realization motivated the development of real time techniques for facing these problems. The main idea behind the new approach is to use real time measurements in drilling, to evaluate the drilling job and to update the design in order to resolve the encountered problems. In the heart of the design phase an accurate module for estimating the optimum mud-pressure is always needed.

We presented a sand prediction model based on a 3-D poroelastoplastic finite element analysis of a perforated wellbore. All the involved processes – wellbore drilling, casing and cementing, perforation creation, production and reservoir depletion - were simulated by the model. The criterion for sand production was based on predicting rock failure when the plastic strain, which reaches a critical threshold. The model can provide directly the critical drawdown for avoiding rock failure near the perforations at different orientations. For weakly consolidated formations a constitutive model to account for a compactant and cohesion-weakening plasticity response is needed. The production from perforations involves the stability of small holes where the size of the hole (scale effect) is important[8]. Sanding in unconsolidated formations, the transport of sanding from the reservoir to surface facilities and the estimation of the volumes of the produced sand are still outstanding issues in petroleum engineering.

Finally, a new area in petroleum engineering, called reservoir geomechanics has developed in the recent years. The reservoir simulators are coupled with stress analysis in order to give more accurate predictions and to account for problems such as reservoir compaction and surface subsidence, casing collapse, fault activation and other. In all these problems the knowledge of the insitu stresses is required. The finite element analysis can be combined with local measurements for obtaining the insitu stresses near complex geological structures.

ACKNOWLEDGEMENT

The authors are grateful to J. Cook and to other ex-colleagues in Schlumberger for their contribution.

REFERENCES

[1] Papanastasiou P., Zervos A. (2004), "Wellbore stability analysis: from linear elasticity to postbifurcation modelling", Int. J. Geomechanics, 4(1), pp.2-12.

[2] Ewy R., (1993), "Yield and closure of directional and horizontal wells", Int. J. Rock Mech. Min. Sci. & Geomech. Abstr. Vol. 30, No 7, pp.1061-1067.

[3] Zervos A., Papanastasiou P. (1998), "Elastoplastic finite element analysis of inclined wellbores", *Proceedings SPE/ISRM Rock Mechanics in petroleum engineering*, SPE/ISRM 47322, pp.535-544.

[4] Papanastasiou P., Sibai M., Heiland J., Shao J-F, Cook J., Fourmaintraux D., Onaisi A., Jeffryes B., Charlez Ph. (2002), "Stability of a Multilateral Junction: Experimental Results and Numerical Modelling", *SPE/ISRM Rock Mechanics Conference held in Irving, Texas*, SPE/ISRM 78212.

[5] Nicholson, E.D., Goldsmith, G. and Cook, J. (1998), "Direct observation and modelling of sand production processes in weak sandstone", *SPE 47328 presented at the SPE/SRM Eurock'98* in Trondheim Norway.

[6] Hibbitt, Karlsson & Sorensen Inc., (1998), *ABAQUS/Standard User's Manual*, Version 5.7.

[7] Viken R., Kjorholt H., (1998) "State-of-the-Art Sand Prediction Based on an Intranet "Push the Button" Interface", *Proceedings SPE/ISRM Rock Mechanics in petroleum engineering,* SPE/ISRM 47361, pp.271-276.

[8] Papamichos, E. and van den Hoek, P. (1995), "Size dependency of Castlegate and Berea sandstone hollow-cylinder strength on the basis of bifurcation theory". *Proc. 35th US Symp. Rock Mechanics,* pp. 301-306.

THE USE OF COMPUTER-AIDED ANALYSIS IN THE DESIGN OF CVD REACTORS FOR THE GROWTH OF TIN OXIDE FILMS

Theodora C. Xenidou, Andreas G. Boudouvis and Nicolas C. Markatos

School of Chemical Engineering
National Technical University of Athens, Athens 15780, Greece
e-mail: boudouvi@chemeng.ntua.gr, web page: http://www.chemeng.ntua.gr/dep/boudouvis/

Keywords: reactor design, evolutionary algorithm, chemical vapor deposition

Abstract. *A computer-aided analysis is applied to the design of a chemical vapor deposition (CVD) reactor for tin oxide deposition. A three-dimensional computational fluid dynamics (CFD) model for the gas flow, mass and energy transfer in a multicomponent gas mixture is coupled to a heterogeneuous kinetic model for the film growth. A SIMPLE-type finite volume method is used on a non-orthogonal grid to solve the steady-state reaction-transport equations and obtain the growth rate profile across the wafer. The quantitative effects of various operating conditions on the CVD outcome are investigated through parameter continuation. Finally, an evolutionary algorithm is incorporated in the CFD/kinetic model to determine the optimal process parameters, under the objective of minimizing the non-uniformity of the growth rate across the wafer; preliminary optimization analysis results are reported.*

1 INTRODUCTION

Chemical Vapor Deposition (CVD) has been an important technology in microelectronics industry for producing thin solid films on heated substrates (wafers). The essential mechanisms underlying CVD processes are chemical reactions in the gas phase and on surfaces, combined with mass, momentum and energy transport. The interactions of chemical and transport processes, which are determined by reactor design and process parameters, influence the quality of the deposited films.

Due to the continuous reduction of the size of microelectronic components, improved CVD reactor designs are needed to meet the imposed requirements in the uniformity of the deposited films. The modifications of the reactor configurations and the manual control of the operating parameters are time consuming, considerably expensive and probably inefficient. Moreover, experimental data are usually available in a small range of operating conditions. On the other hand, mathematical modeling and computer simulation provide an excellent economic alternative to trial-and-error based experimental techniques. In addition, simulation models allow the evaluation of numerous reactor configurations and process parameters and finally the optimization of a CVD reactor in terms of high degree of spatial uniformity of the growth rate across the wafer.

Among the different mathematical approaches used so far in CVD reactors design, the most common used are based on coupled transport/kinetic models, which are used to investigate the quantitative effects of various operating parameters to the properties of the deposited films[1-3]. In particular, parametric analysis of CVD reactors is performed by varying one process parameter at a time. Although significant conclusions can be obtained from each parameter continuation run, there will be no "global" solution to the design problem of the CVD reactor. Obviously, it would be more efficient to investigate the effect of more process parameters at a time on the CVD reactor performance. To improve the efficiency of the parametric analysis of a CVD reactor, an Evolutionary Algorithm (EA) is incorporated in a detailed Computational Fluid Dynamics (CFD) model for the gas flow, heat and mass transfer inside CVD reactors. Namely, the issue of the spatial uniformity of the growth rate across the wafer is defined as an optimization problem and an objective function is defined to serve the optimization procedure. This approach has been used previously in the optimization of various CVD systems[4-6]. For example, Grujicic and co-workers performed the optimization of carbon nanotubes fabrication coupling a genetic algorithm with a detailed CFD model[5].

In this work, the efficiency of the particular design approach is investigated in a horizontal, cold-wall reactor where tin oxide (SnO_2) thin films are grown on silicon wafers at atmospheric pressure conditions (APCVD). A previous combined experimental/computational analysis of the CVD reactor has resulted in the preliminary detection of the origins of the spatial variation of the growth rate across the wafer[7]. In the meantime, a new set of experimental data was obtained and an improved heterogeneous kinetic mechanism for tin oxide deposition was extracted. The improved kinetic model is incorporated in the CFD model and used in the simulations performed in this study.

2 REACTOR MODEL

2.1 The case studied

Tin oxide is deposited on silicon substrates through the following oxidation reaction:

$$SnCl_4 + O_2 \rightarrow SnO_2 + 2Cl_2 \quad (1)$$

The horizontal cold-wall CVD reactor is sketched in Figure 1. A detailed description of the CVD reactor and the deposition procedure can be found elsewere[7].

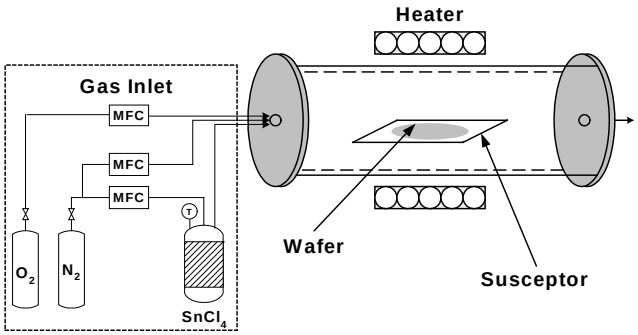


Figure 1. Schematic view of the CVD reactor

The liquid precursor ($SnCl_4$) is delivered from the bubbler to the reactor chamber using nitrogen as carrier gas. The reactant gases are injected through five different inlet holes in the reactor, which contains a silicon substrate placed on a graphite susceptor. The wafer is heated radiatively by lamp arrays, placed outside the quartz reactor, at a constant temperature. Finally, the reactant and the product gases leave the reactor through the outlet, which is fixed at atmospheric pressure. The exterior surfaces of the reactor are cooled with air fans.

2.2 CFD Model

Previous works have focused their attention on the development of three-dimensional simulation models of horizontal CVD reactors. The first of such studies were performed by Moffat and Jensen[8] for the simulation of horizontal epitaxial CVD reactors. Kleijn and Hoogendoorn[9] as well as Evans and Greif[10] have also developed three-dimensional CFD models of horizontal CVD reactors. They demonstrated the relative importance of convection and diffusion (ordinary and thermal) and found that three-dimensional models can accurately predict the experimental data under wide range of operating conditions.

The governing equations summarized in this section have been implemented in the PHOENICS code. More complete description and capabilities can be found in reference[11]. The conservation equations for mass, momentum and heat transfer are applied to a three-dimensional reacting flow in a multicomponent gas mixture. The equations for fluid flow consist of the incompressible Navier-Stokes equations for a variable-density fluid and the continuity equation. The species mass balance equations are solved for N_g-1 gas-phase species. The multicomponent diffusion is modeled through the Stefan-Maxwell equations, and the thermal diffusion (Soret effect) is taken to affect the mass transport. The energy balance equation is solved for temperature.

The boundary conditions include prescribed velocities, temperatures and species concentrations at each inlet hole. The operating pressure is specified at the outlet, where zero normal derivatives are assumed for all other variables. On the solid surfaces of the reactor, the gas velocity components are taken equal to zero and the flux of each chemical species, reactant or product, is calculated by the kinetic rate of the heterogeneous reactions.

The transport and thermochemical properties of the individual gas species and the binary diffusion coefficients are estimated using the kinetic gas theory. Mixing rules are used for calculating the properties of the gas mixture in terms of pressure, temperature and composition.

2.3 Kinetic Model

The numerical simulation of SnO_2 deposition from $SnCl_4/O_2$ mixtures requires a kinetic equation of the deposition rate, R_d. The oxidation reaction in equation (1) represents the global deposition process, which is complex and involves elementary steps, such as adsorption, surface reaction and desorption. To account for these elementary reactions, a "monorhoic" mechanism[12] is used, as an improved alternative to the Langmuir-Hineshelwood type kinetics, previously employed in tin oxide CVD[7]. The "monorhoic" model of tin oxide

deposition consists of the heterogeneous reactions shown in Table 1. A species A(a) is adsorbed at a surface site, a species A(g) denotes a gas phase species and a species A(s) denotes a solid species. The symbol S denotes a free active surface site.

RS1(k_1)	$SnCl_4(g) + 2\cdot[S] \rightarrow SnCl_2(a) + Cl_2(a)$
RS2(k_2)	$SnCl_2(a) + O_2(g) \rightarrow SnO_2(s) + Cl_2(a)$
RS3(k_3)	$Cl_2(a) \rightarrow Cl_2(g) + [S]$

Table 1 : The "monorhoic" model of tin oxide CVD

At equilibrium, the deposition rate of SnO_2 is given by

$$R_d = k_2 P_{O_2} [SnCl_2(a)] \tag{2}$$

In equation (2), k_2 is the rate constant of the surface reaction RS2, P_{O_2} is the partial pressure of oxygen and $[SnCl_2(a)]$ is the surface concentartion of the adsorbed $SnCl_2$ as a function of the total surface covered by it.

The growth rate of SnO_2 is calculated as:

$$GR = f \cdot R_d \tag{3}$$

The proper conversion from mol/m^{2}s (deposition rate, R_d) to nm/min (growth rate, GR) is included in factor f, which is equal to $6\cdot 10^{10} M_s / \rho_s$. M_s and ρ_s are the molecular weight and the density of the deposited film, respectively.

3 OPTIMIZATION METHOD

The optimization study is performed using the optimization software EASY (Evolutionary Algorithms SYstem). EASY was developed and brought to market by the Lab. of Thermal Turbomachines of the National Technical University of Athens. A detailed description of the optimization software can be found in reference[13]. Here we briefly outline the optimization procedure of the CVD reactor. EASY utilizes, keeps and evolves three population sets, namely the set of parents S^{μ}, that of offspring S^{λ} and the archival elite set S^{e}; subscripts denote the population size and, in general, $\mu \neq \lambda$. The population includes the single current optimum in the case of a single objective function. Thinning processes can optionally be utilized to control the number of individuals kept in S^{e}. Computations with $e > 1$ are not only possible but often perform very well even in single-objective problems.

Due to the high number of calls to the computationally expensive CFD code, surrogate evaluation models are employed. In particular, EASY trains local artificial neural networks using information collected during the previous generations[14]. In each new generation, all population members are approximately evaluated using the trained surrogate models and, then, only the most promising among them are re-evaluated using the exact and costly CFD code. The combined use of the local, low-cost approximate and the exact, computationally expensive CFD code leads to a considerable economy in the overall cost of the optimization study.

To measure the non-uniformity of the growth rate across the wafer two different metrics are calculated as a post-processing step of the CFD simulation. The first metric is the non-uniformity, ΔG, defined through the maximum, minimum and average growth rate, i.e., by:

$$\Delta G = \frac{GR_{max} - GR_{min}}{\overline{GR}} \tag{4}$$

This is a L^{∞} norm and is usually reported in percent. The second metric is the L^2 norm of the growth rate profile about its average and is used as the objective function:

$$f(\mathbf{P}) = \left(\frac{1}{\overline{GR}^2 A_w} \int \left(GR(x,z) - \overline{GR} \right)^2 dA \right)^{1/2} \tag{5}$$

In equation (5) A_w is the wafer area, GR(x,z) is the deposition rate profile over the wafer, and $\overline{GR}$ is the average growth rate, pre-calculated by integrating GR(x,z). **P** is the vector of the process parameters to be estimated, including, for example, gas inlet flow rate, temperature, etc.

4 PARAMETER CONTINUATION RESULTS

In this section the quantitative effects of various operating conditions on the quality of the deposited films are investigated through parameter continuation. The solution at base operating conditions is followed by the results of different parameter studies about the base point. Among the probable process variables influencing the tin oxide growth rate and its variation across the wafer, the study is restricted to three variables, namely purge nitrogen flow rate, oxygen flow rate and the deposition temperature.

The system of the coupled nonlinear partial differential equations is solved on a non-orthogonal grid of 49.140 nodes, which corresponds to nearly 400.000 unknowns. A Pentium IV/2.4GHz is used for each parameter continuation run, which consumed 2h CPU time to obtain a CVD reactor state.

4.1 CFD/kinetic model predictions at base conditions

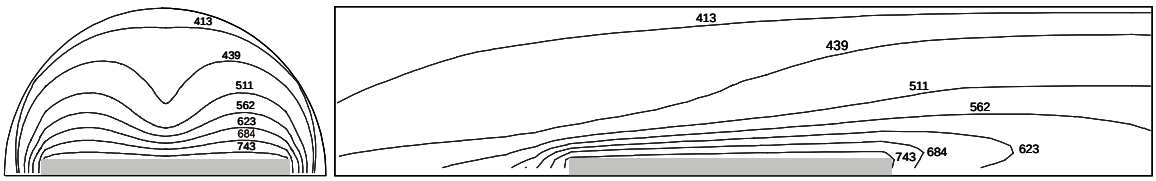
The CFD/kinetic model is first solved at the base operating conditions for the process parameters summarized in Table 2.

Parameter	Description [units]	Base value
F_d	Nitrogen purge flow rate [mL/min][1]	400
F_c	Nitrogen carrier flow rate [mL/min][1]	200
F_o	Oxygen flow rate [mL/min][1]	150
T_b	Bubbler temperature [^{o}C]	28.3
T_{wafer}	Wafer temperature [^{o}C]	470
T_{inlet}	Inlet temperature [^{o}C]	60
T_{wall}	Wall temperature [^{o}C]	140

[1] values refer to standard conditions

Table 2 : Base operating conditions considered for tin oxide CVD

In Figure 2a and b are shown sample isotherms, at base operating conditions, in the transverse and axial direction, respectively. The distribution of the growth rate at base operating conditions, in the transverse and axial direction, are presented in Figure 3a and b, respectively. The simulation results predict a relatively high variation of the growth rate in both directions. The non-uniformity of the growth rate, ΔG, is calculated to be 4.4%, which is above the desired value of 1-2%. In the transverse direction (Fig. 3a), the local minima of the growth rate are closely connected to the large temperature gradients. However, the uniformity of the growth rate is improved in the flow direction, namely, with increase of the distance from the leading edge of the wafer. In the axial direction, the growth rate decreases in the axial direction, as shown in Fig. 3b; again the growth rate distribution along the central axis of the reactor is different compared to the edges of the wafer.



(a) (b)

Figure 2. Temperature contours along the (a) transverse and (b) axial center line of the wafer.

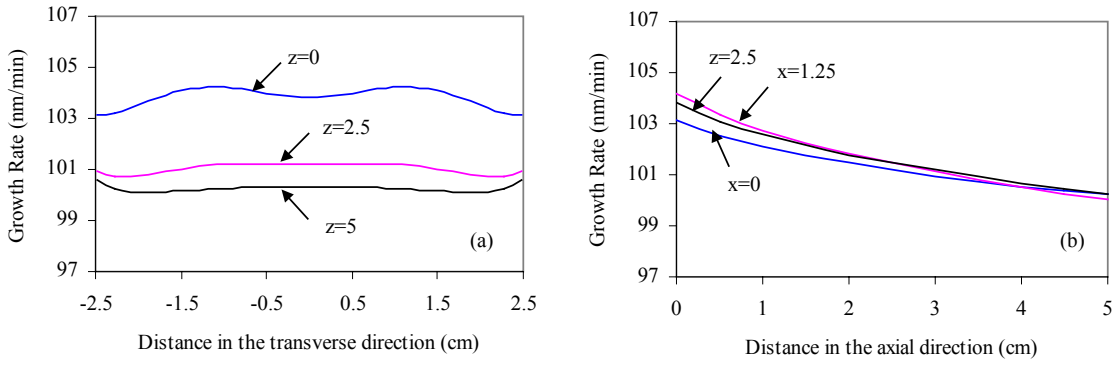


Figure 3. Growth rate profile in (a) the transverse x-direction (at different distances from the leading edge of the wafer) and in (b) the axial z-direction (at different distances from the center line of the wafer).

4.2 Purge Gas Flow Rate

Parameter continuation on the purge gas flow rate, F_d, is performed and growth rate profiles are calculated for flow rates from 200 to 400 mL/min in steps of 40. As shown in Figure 4a and b, the decrease of the purge flow rate is followed by an increase of the growth rate, in both directions. It appears that the improved mixing of the reactants, caused by decrease in the purge flow rate, yields higher reactant concentrations and therefore higher growth rates. This is followed by an increase of the non-uniformity across the wafer, as depicted in Figure 5a.

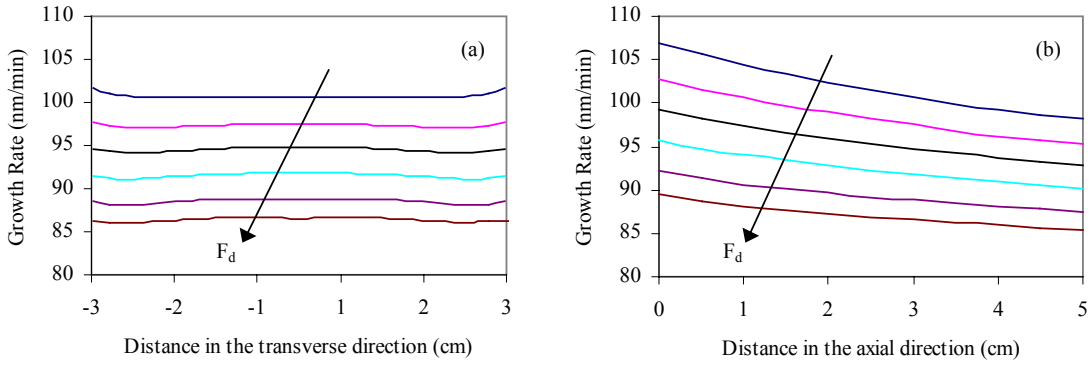


Figure 4. Growth rate profile at different values of purge gas flow rate, F_d (in steps of 40 mL/min) in the (a) transverse and (b) axial direction along the center line of the wafer.

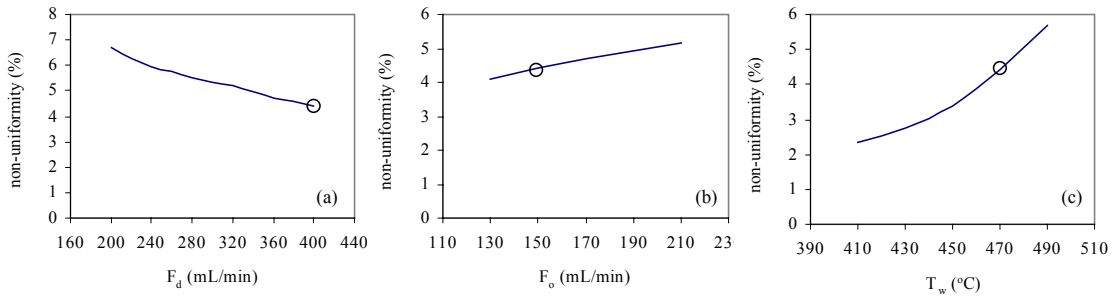


Figure 5. Non-uniformity of the growth rate as a function of (a) purge gas flow rate (F_d), (b) oxygen flow rate (F_o) and (c) wafer temperature (T_w). The circular symbol denotes base conditions.

4.3 Oxygen Flow Rate

Parameter continuation on the oxygen flow rate, F_o, is performed and growth rate profiles are calculated for flow rates from 130 to 210 mL/min in steps of 20. The results in Figure 6a and b show that the increase of the oxygen flow rate produces an increase of the growth rate in both directions. This trend is followed by a slight increase of the non-uniformity across the wafer, as shown in Figure 5b.

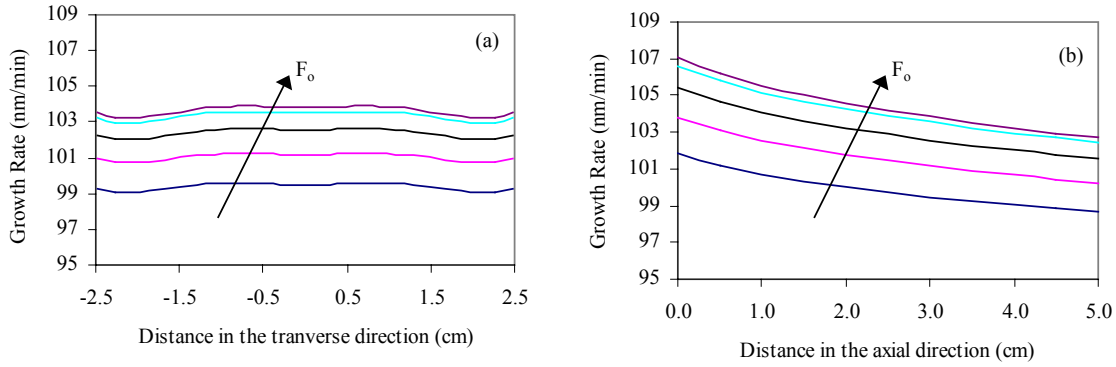


Figure 6. Growth rate profile at different values of oxygen flow rate F_o (in steps of 20 mL/min) in the (a) transverse and (b) axial direction along the center line of the wafer.

4.4 Deposition Temperature

Parameter continuation on the wafer temperature, T_w, is performed and growth rate profiles are calculated for temperatures from 410 to 490 ^oC in steps of 20. The results in Figure 7a and b indicate that any increase of the wafer temperature causes an increase of the growth rate in both directions. When the wafer temperature becomes higher (while the walls temperature is constant), the increase in thermal diffusion causes a significant removal of the reactants from the deposition surface towards the cold walls of the reactor. The non-uniformity of the growth rate is seen to be a relatively strong increasing function of the wafer temperature, as shown in Figure 5c.

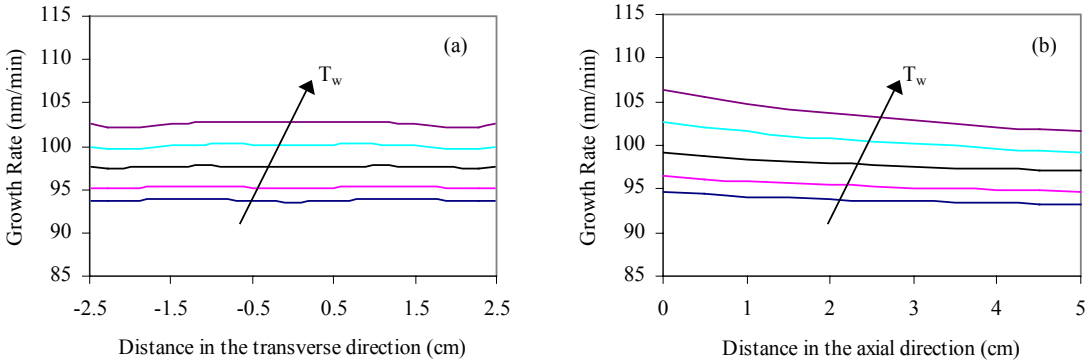


Figure 7. Growth rate profile at different values of wafer temperature T_w (in steps of 20 mL/min) in (a) the transverse and (b) axial direction along the center line of the wafer.

5 OPTIMIZATION RESULTS

The parameter studies presented in the previous section show that the non-uniformity of tin oxide growth rate is affected by the three parameters examined. The lowest value of the non-uniformity achieved by varying one parameter at a time was 2.36% (for $T_w = 410^oC$). The results obtained in the parameter studies are re-plotted in Figure 8, in terms of the L^2 norm of the growth rate variation about its average. This norm is used in two optimization runs. In both cases the parameters not being optimized are held at the base values (Table 1).

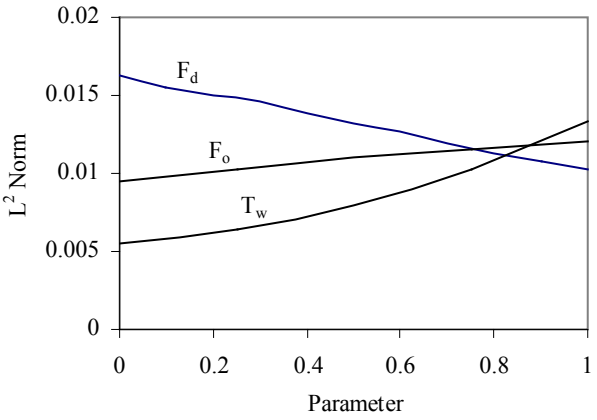


Figure 8. L^2 norm of the variation of the growth rate as a function of purge gas flow rate (F_d), oxygen flow rate (F_o) and wafer temperature (T_w).

The first run is a verification run, so that the result can be compared with the continuation runs. In this run, the two flow parameters are varied, namely the purge flow rate, F_d and the oxygen flow rate, F_o. As shown in Table 2, the result of the first optimization run is consistent with the results of the continuation runs, as read from curves F_d and F_o in Figure 8.

The second run includes all three parameters and achieves the largest reduction in the objective function corresponding to 1.85% non-uniformity. However, the lack of multiple local minima inside the range of the parameters examined in this study restricts the efficiency of the optimization approach. Further improvement in the growth rate uniformity will consist of more optimization parameters (e.g. reactor pressure, precursor inlet flow rate, etc).

Parameter	Optimal Value	L^∞ Norm	L^2 Norm
F_d	400.0	4.40%	0.010283
F_o	130.0	4.08%	0.009545
T_w	410.0	2.36%	0.005523
F_d, F_o	399.8, 130.0	4.26%	0.009437
F_d, F_o, T_w	398.4, 130.0, 411.7	1.85%	0.004252

Table 2 : The results of two optimization runs

6 CONCLUSIONS

A computer-aided analysis was applied to the computational study of tin oxide deposition in an APCVD reactor. The essential tools employed in the design procedure of a single-wafer reactor were a computational fluid dynamics model and an evolutionary algorithm. In particular, a three-dimensional reactor-scale model for the gas flow, mass and energy transfer is coupled to a heterogeneous kinetic model for the film growth. The individual effects of three process parameters on the spatial uniformity of deposition were investigated with parameter continuation runs. Finally, a framework was developed for enabling the CFD/kinetic code to perform optimization studies, under the objective of minimizing the non-uniformity of the growth rate across the wafer.

ACKNOWLEDGEMENTS

Financial support was provided by the European Community Fund (75%) and Public Resources (25%), through the PYTHAGORAS program. The authors would like to thank Prof. K.C. Giannakoglou (NTUA) for providing the optimization software and Prof. D.M. Tsamakis (NTUA) for providing the experimental data of tin oxide CVD.

REFERENCES

[1] Chae, Y.K., Egashira, Y., Shimogaki, Y., Sugawara, K. and Komiyama, H. (1999), "Chemical vapor deposition reactor design using small-scale diagnostic experiments combined with computational fluid dynamics simulations", J. Electrochem. Soc. 146(5), pp. 1780-1788.

[2] Kommu, S.K., Wilson, G.M. and Khomammi B. (2000), "A theoretical/experimental study of silicon epitaxy in horizontal single-wafer chemical vapor deposition reactors", J. Electrochem. Soc. 147(5), pp. 1538-1550.

[3] Barathieu, P., Cassaut, B., Schied, E. and Couderc, J.P. (2001), "Low-pressure chemical vapor deposition of semi-insulating polycrystalline silicon thin films. II. Theoretical local analysis of the process", J. Electrochem. Soc. 148(3), pp. C156-C161.

[4] Cho, W.K. and Choi, D.H. (2000), "Optimization of a horizontal MOCVD reactor for uniform epitaxial layer growth", Int. J. Heat Mass Transfer 43, pp.1851-1858.

[5] Grujicic, M., Cao, G. and Gersten, B. (2002), "Optimization of the chemical vapor deposition process for carbon nanotubes fabrication", Appl. Surf. Sci. 191, pp. 223-239.

[6] Itle, G.C., Salinger, A.G., Pawlowski, R.P., Shadid, J.N. and Biegler, L.T. (2004), "A tailored optimization strategy for PDE-based design: application to a CVD reactor", Comp. Chem. Eng. 28(3), pp. 291-302.

[7] Xenidou, T.C., Boudouvis, A.G., Tsamakis, D.M. and Markatos, N.C. (2004), "An experimentally assisted computational analysis of tin oxide deposition in a cold-wall APCVD reactor", J. Electrochem. Soc. 151(12), pp. C757-C764.

[8] Moffat, H.K and Jensen, K.F. (1986), "Complex flow phenomena in MOCVD reactors: I. Horizontal reactors", J. Cryst. Growth 77(1-3), pp 108-119.

[9] Kleijn, C.R. and Hoogendoorn, C.J. (1991), "A study of 2- and 3-D transport phenomena in horizontal chemical vapor deposition reactors ",Chem. Eng. Sci. 46(1), pp. 321-334.

[10] Evans, G.H. and Greif, R. (1991), "Unsteady three-dimensional mixed convection in a heated horizontal channel with applications to chemical vapor deposition", Int. J. Heat Mass Transfer 34(8), pp. 2039-2051.

[11] CHAM, Ltd. (2001), PHOENICS Documentation, v.3.4.0, London.

[12] ter Heerdt, M., L.H., van der Put, P.J., and Schoonman J. (2001), "Surface kinetics in copper CVD" Chem. Vapor Depos. 7(5), pp.199-203.

[13] Giannakoglou, K.C. (2002), "Design of optimal aerodynamic shapes using stochastic optimization methods and computational intelligence", Prog. Aerosp. Sci., 38(1), pp. 43-76.

[14] Karakasis, M.K., Giotis, A.P. and Giannakoglou, K.C. (2003), "Efficient genetic optimization using inexact information and sensitivity analysis. Application in shape optimization", Int. J. Numer. Meth. Fluids 43, pp. 1149-1157.

SIMULATION OF SAMPLE INJECTION IN MICROCHIP ELECTROPHORESIS FOR DNA SEQUENCING

A. Srivastava[1,2], A. C. Metaxas[3], P. So[1], P. Matsudaira[2,4], D. Ehrlich[2], G. E. Georghiou[5]

[1] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[2] Whitehead Institute for Biomedical Research, Cambridge, MA, USA
[3] St John's College, University of Cambridge, Cambridge UK
[4] Department of Biology and Bioengineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[5] Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K
e-mail: geg@ecs.soton.ac.uk

Keywords: DNA sequencing, microchip electrophoresis, simulation, stacking

Abstract. *A numerical model has been developed for the accurate prediction of pre-concentration and transport of DNA during microcapillary electrophoresis. The model incorporates conservation laws for the different buffer ions, salt ions and DNA sample, coupled through the solution of Poisson's equation to account for the field modifications that cause electromigration. Results are presented for the prediction of DNA sample pre-concentration and transport during microcapillary electrophoresis in a double-T injector microdevice. Experimental results have also been obtained and compared with the corresponding numerical results under similar conditions and excellent agreement has been obtained. Finally, the validity of the results demonstrates the potential of the numerical model for future optimization of such microchip devices with respect to significantly enhanced resolution of sample separation.*

1 INTRODUCTION

Microfluidic devices for chemical and biochemical analyses have attracted significant interest due to large-scale genome sequencing initiatives and recent decisions to generate large genotype databases for human forensics. There is a growing need for improved technology at a lower cost for high-throughput and fast nominally real-time DNA assays. Using micromachining technology, a series of microchannels can be fabricated on a single substrate of glass, quartz, or plastics in which multiple procedures, e.g., sample loading, pre-treatment, chemical reaction, separation, etc., can be performed [1,2]. The primary process of separation performed in microdevices is similar in principle to that of conventional capillary electrophoresis, though highly miniaturized and improved. To a large degree, this improvement is attributable to the very small volume of injected sample plugs, which is often two orders of magnitude lower than that in conventional capillary electrophoresis, yielding faster separation times and offering process integration [3]. Furthermore, microdevice electrophoresis is finding increasing applications in potential lab-on-a-chip formats [4].

Despite significant advances being made in many aspects of DNA microchip electrophoresis, repeatability, and indeed improvements in the DNA injection process have lagged other developments within the field. Currently, there is a critical need for optimizing microdevice systems for improved separations with minimum use of sample and better resolution and sensitivity [5]. A reduction in the amount of sample required, results in a lower consumption of fluorescent dyes, which are significant contributors to the total cost of sequencing. DNA sample pre-concentration or stacking observed in the early periods of electrophoresis, namely sample injection, could potentially achieve these improvements [6].

A numerical model is presented for the accurate and efficient prediction of pre-concentration and transport of DNA during sample introduction and injection in microcapillary electrophoresis. The model incorporates conservation laws for the different buffer ions, salt ions, and DNA sample, coupled through a Gaussian electric field to account for the field modifications that cause electromigration. The accuracy and efficiency required to capture the physics associated with such a complex transient problem are realized by the use of the finite element-flux corrected transport (FE-FCT) algorithm in two dimensions [7]. The model has been employed for the prediction of DNA sample pre-concentration and transport during electrophoresis in a double-T injector microdevice. To test its validity, the numerical results have been compared to the corresponding experimental data under similar conditions, and excellent agreement has been found. Finally, detailed results from a

simulation of DNA sample pre-concentration in electrophoretic microdevices are also presented. These results demonstrate the great potential offered by the model for future optimization of such microchip devices with respect to significantly enhancing the speed and resolution of sample separation.

2 MATERIALS AND METHODS

2.1 Details of Microfabricated device used

A micro-fabricated device based on the double-T injector type design used at the Whitehead Institute for Biomedical Research (WIBR) and Massachusetts Institute of Technology (MIT) BioMEMS Laboratory has been used [8]. A schematic diagram is shown in figure (1). The device consists of the cathode, anode, sample and waste reservoirs together with the three main channel sections, namely the separation channel, cross-injector channel (injector off-set) and channel tail. Photolithographic etching techniques on fused-silica wafers that are 12 cm long, 1 cm wide and 1.1 mm thick are used for the micro-fabrication. The channels are approximately 100 μm wide, 40 μm deep, whilst the separation channel is typically 11.5 cm long. The sample and waste arms of the injector are approximately 5.0 mm in length, and horizontally offset by a distance of 300 μm. Glass reservoirs of 50 ml volume are affixed around the laser-drilled holes that access the electrophoretic channel in order to contain the appropriate volumes of sample and buffer solutions.

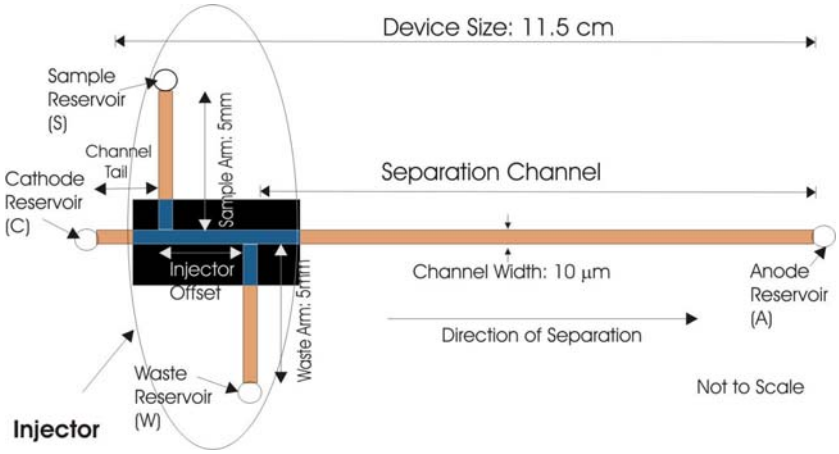


Figure 1. Microdevice used for the experiments. The computational domain used for the simulations is shown in the black rectangle.

Conventional separations are performed using four consecutive steps: pre-electrophoresis, sample introduction, injection and separation. During pre-electrophoresis, a large potential gradient, typically several hundred volts per centimeter, is imposed between the cathode and anode ports of the device and later between the sample and waste reservoirs. This is used to remove any stray contaminants in the injector prior to the loading of DNA. Next, the DNA sample which includes salt ions is added to the sample reservoir and TTE (TRIS-TAPS-EDTA) buffer is added to the waste reservoir. During sample introduction, a different potential gradient is applied (E_l) along the sample and waste reservoirs of the injector in order to draw DNA molecules into the micro-device and create a uniformly distributed DNA sample within the injector offset. The DNA and chloride anions enter from the sample reservoir, whereas the TRIS cations enter from the waste reservoir. Finally, during the separation stage, a different potential gradient, called the run gradient, E_r, is applied between the cathode and anode ports of the device in order to attract the DNA towards the positively charged anode. The process of pre-concentration or stacking of DNA molecules into the channel and their subsequent separations take place during this step.

2.2 Experimental Conditions

The microdevice channels are filled with a solution of the polyacrylamide sieving matrix and TTE buffer by using a syringe, whilst ensuring that the channels are bubble free. The DNA sample is introduced from the sample reservoir, and the TTE buffer is added to the waste reservoir. DNA sequencing-reaction products are introduced into the injector offset by applying a negative potential gradient of 150 V/cm between the sample and waste reservoirs (sample reservoir at positive potential and waste reservoir is kept to ground) for a time span of 180 s. There is no applied voltage at the anode and cathode reservoirs during sample introduction. A run

potential gradient of 200 V/cm (2300 V at the anode and 0 at the cathode in this particular device) is then applied for 15 s during injection. Leakage of excess sample from the injector offset into the separation channel is prevented by the application of a small electric pull back potential gradient of 40 V/cm between the sample and waste reservoirs 15 s after injection throughout the separation stage.

2.3 Theoretical Model

The transport of each ionic species (buffer ions TRIS+ and TAPS-, salt ions namely chloride), as well as the DNA species, is represented by a conservation law and the effect of the reaction and generation terms is assumed to be negligible [9]. In this model the governing equations for the three species is given by:

$$\frac{\partial c_i}{\partial t} + \nabla.(\mathbf{F_i} + \mathbf{G_i}) = 0 \tag{1}$$

where c_i is the concentration of the ionic species i and $\mathbf{F_i}$ and $\mathbf{G_i}$ the convective and diffusive fluxes respectively. The species present in this study are those of TAPS, Cl and TRIS ions. In turn the convective and diffusive fluxes are given by:

$$\mathbf{F_i} = z_i \mu_i c_i \mathbf{E} \tag{2}$$

$$\mathbf{G_i} = D_i \nabla c_i \tag{3}$$

where z_i and μ_i are the valence and mobility of the ionic species i respectively, **E** is the electric field intensity and D_i is the diffusion coefficient of ionic species i.

The three conservation laws are coupled by the electric field, **E**, which is expressed by Poisson's equation in order to account for the effect of the charges on the field:

$$\nabla.\mathbf{E} = -\frac{\rho}{\varepsilon} \tag{4}$$

with ε being the permittivity of the medium and ρ the charge density. Relating the charge density to the concentration of the charged ions yields

$$\rho = \sum_i z_i c_i F \tag{5}$$

where F is the Faraday constant.

The dynamics of the buffer electrolyte and salt ions determine the electric field distribution since their charges influence the total electric field through equation (4). Since the concentration of DNA is orders of magnitude smaller than the buffer and salt, it does not influence the electric field. In fact, the electric field in the electrolyte system drives the sample DNA species of concentration c_{DNA}, mobility μ_{DNA}, diffusion D_{DNA} and valence –1. The dynamics of the DNA sample is therefore governed by equation (6), which completes the model:

$$\frac{\partial c_{DNA}}{\partial t} + \nabla.(-\mu_{DNA} c_{DNA} \mathbf{E} - D_{DNA} \nabla c_{DNA}) = 0 \tag{6}$$

In numerical studies, difficulties arise from the solution of the conservation equations, such as the ones encountered in this model, due to the very steep, shock-like gradients that appear in convection dominated regions, in particular when simulating moving boundaries during the stacking process. Consequently, a very accurate numerical technique is required to capture the development of sharp gradients without introducing unrealistic effects such as excess diffusion or oscillations. In this study we employ the finite-element flux-corrected transport (FE-FCT) method recently reported for the characterization of the shock-like motion of charged particles in the presence of high electric fields in gas discharge problems [10]. The FE-FCT method

maintains the ability to capture steep gradients, through the FCT, and allows the use of unstructured triangular grids via the FE. This significantly reduces the number of unknowns and hence computing time, as fine resolution is only used where necessary.

2.4 Computational Conditions

The aim of this study is to predict the transport of DNA in the initial part of electrophoresis and more specifically during the pre-concentration process. From the experimental images, it is evident that the dynamics are focused in or near the injector region during this time frame. As a result the injector system shown in figure (1) (represented in a black rectangle) is used as the computational domain, in order to capture the behaviour of the system during the introduction and injection stages. This approximation is necessary in our attempt to make the calculations more amenable without compromising the efficiency. The tail, separation channel, sample and waste arms are truncated, while maintaining all the other dimensions which are crucial for the correct characterization of the physics behind the pre-concentration process. The voltages are then adjusted to match the potential gradients applied in the experimental case so that the experimental and numerical results can be fairly compared. These approximations are justified given the fact that the truncation occurs far enough from the injector and hence it is safe to assume that the potential gradients remain constant in the remainder of the device. The initial and boundary conditions are also adjusted accordingly to emulate the experimental conditions [9].

3 RESULTS AND DISCUSSION

3.1 Numerical vs experimental results

The electrophoresis micro-devices studied by this model are from WIBR/MIT, however the methods used are general and can be applied to other devices as well. Results for the sample injection stage are presented after sample introduction has taken place as shown in figure (2).

Figure 2. Sample introduced into the injector off-set (sample introduction stage). Experimental result.

At the end of sample introduction and beginning of sample injection, the DNA occupies the sample arm, the injector, and the waste arm. During the early stages of injection, DNA molecules migrate rapidly within the sample of the injector offset, but experience an abrupt drop in velocity upon reaching the lower field within the high-conductivity electrolyte buffer [8]. The subsequent decrease in velocity creates a narrow, concentrated zone of DNA molecules at the interface between the sample and separation buffer, called the stacked sample. This stacking mechanism is a unique, physical process which is caused by a difference in potential gradient between the sample and the buffer solution. It increases the sample concentration throughout the electrophoretic injection and has generated high-resolution data in numerous subsequent electrophoretograms. The numerical and experimental results for two time instances during sample injection are shown in figures (3) and (4). Sample pre-concentration is evident in both figures, however as time progresses it becomes more pronounced and the peak amplitude of the DNA sample increases. It is also very important to observe the agreement between the experimental and theoretical results, demonstrating the validity of the numerical model to characterize the processes of transport and pre-concentration in microchip electrophoresis.

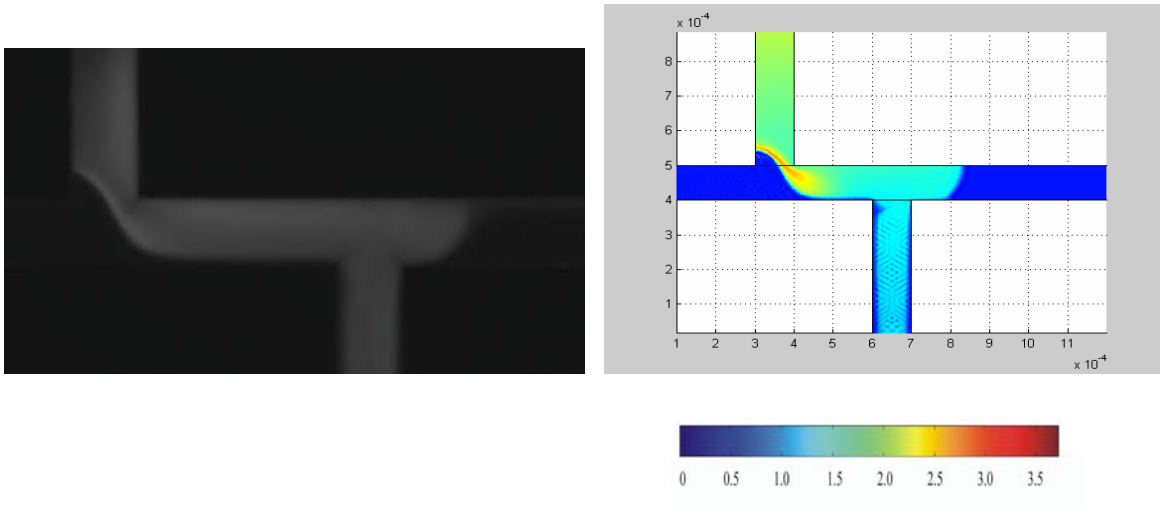


Figure 3. Comparison of experimental results (left) with corresponding numerical simulation (right) during sample injection (t = 1.875 s after sample introduction).

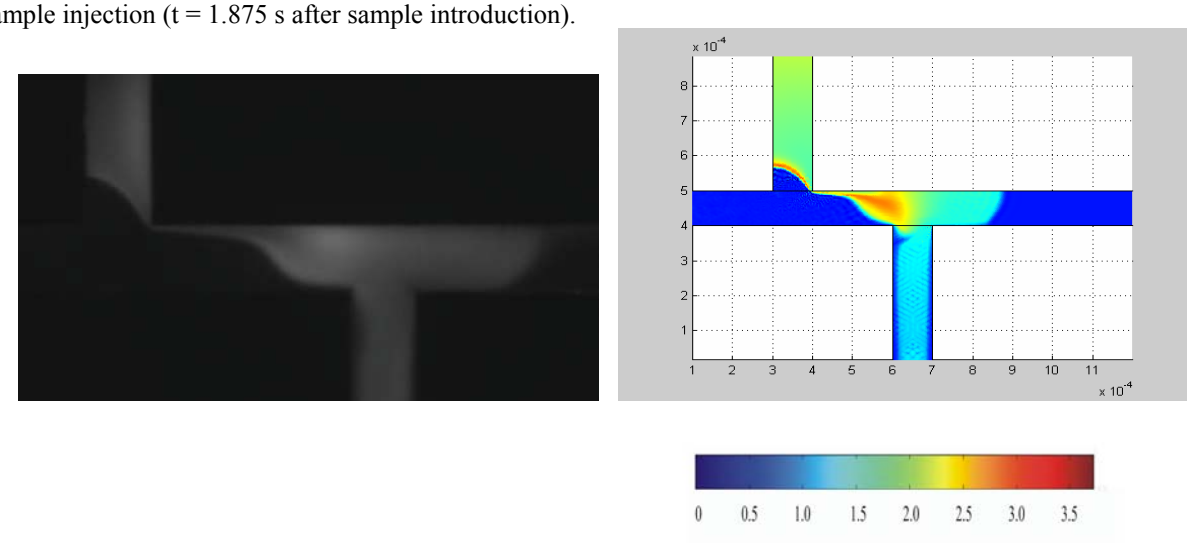


Figure 4. Comparison of experimental results (left) with corresponding numerical simulation (right) during sample injection (t = 4.375 s after sample introduction).

Finally, figure (5) presents the DNA sample evolution along the centre line of the injector during sample injection. Pre-concentration is evident once more as time progresses with the DNA sample reaching its peak value 13.125 s after sample introduction.

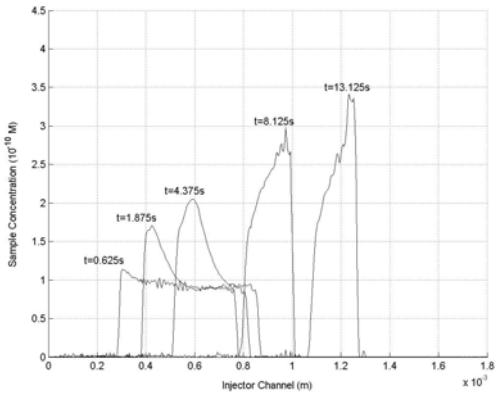


Figure 5. DNA sample pre-concentration at five different time instances. Pre-concentration is evident.

4 CONCLUSIONS

A numerical model based on the incorporation of conservation laws for different buffer ions, salt ions and DNA sample, coupled through Poisson's equation to account for the field modification that causes electromigration, has been applied for the accurate prediction of pre-concentration and transport of DNA during sample introduction and injection in micro-capillary electrophoresis.

The model has been employed for the prediction of DNA sample pre-concentration and transport in a double T injector based microdevice and the numerical results obtained have been compared with the corresponding experimental data under the same conditions, thus demonstrating good agreement and confirming the validity and the predictive power of the numerical model. Finally, good insight into the theory of sample pre-concentration, which is a key step in microdevice DNA separations, has also been presented.

REFERENCES

[1] Mitnik, L., Novotny, M., Felten, C., Buonocore, S., Koutny, L., Schmalzing, D. (2001), "Recent advances in DNA sequencing by capillary and microdevice electrophoresis", Electrophoresis 22, p4104-4117.
[2] Jacobson, S. C., Culbertson, C. T., Daler, J. E., Ramsey, J. M. (1998), "Microchip structures for submillisecond electrophoresis", Anal. Chem. 70, p3476-3480.
[3] Mueller, O., Hahnenberger, K., Dittmann, M., Yee, H., Dubrow, R., Nagle, R., Ilsley, D. (2000), "A microfluidic system for high-speed reproducible DNA sizing and quantitation", Electrophoresis 21, p128-134.
[4] Manz, A., Harrison, D. J., Verpoorte, E. M. J., Fettinger, J. C., Paulus, A., Ludi, H., Widmer, H. M. (1992), "Planar chips technology for miniaturization and integration of separation techniques into monitoring systems: capillary electrophoresis on a chip", J. Chromatogr. 593, p253-258.
[5] Culbertson, C. T., Jacobson, S. C., Ramsey, J. M. (2000), "Microchip devices for high-efficiency separations", Anal. Chem. 72, p5814-5819.
[6] Beckers, J. L., Bocek, P. (2000), "Sample stacking in capillary zone electrophoresis: Principles, advantages and limitations", Electrophoresis 21, p2747-2767.
[7] Georghiou, G. E., Morrow, R., Metaxas, A. C. (2000), "A two-dimensional finite element – flux corrected transport algorithm for the solution of gas discharge problems", J. Phys. D 33, p2453-2466.
[8] Vazquez, M., McKinley, G., Mitnik, L., Desmarais, S., Matsudaira, P., Ehrlich, D. (2002), "Electrophoresis using ultra-high voltages", J. Chromatogr. B 779, p163-171.
[9] Srivastava, A., Metaxas, A. C., So, P., Matsudaira, P., Ehrlich, D., Georghiou, G. E. (2005), "Numerical simulation of DNA sample pre-concentration in microdevice electrophoresis", Accepted, to be published in Electrophoresis 26.
[10] Georghiou, G. E., Morrow, R., Metaxas, A. C. (1999), "An improved finite element – flux corrected transport (FE-FCT) algorithm", J. Comput. Phys. 148, p605-620.

MODELLING THE WIND FLOW IN COMPLEX URBAN TOPOGRAPHIES: A COMPUTATIONAL-FLUID-DYNAMICS SIMULATION OF THE CENTRAL LONDON AREA

Marina K-A Neophytou[1] **and Rex E Britter**[2]

[1] Department of Civil and Environmental Engineering, University of Cyprus,
P.O.BOX 20537, 1678 Nicosia, Cyprus
e-mail: neophytou@ucy.ac.cy,

[2] Department of Engineering, University of Cambridge,
Trumpington Street, Cambridge CB2 1PZ, UK

Keywords: pollution dispersion, atmospheric turbulence, CFD, modeling.

Abstract. *With the advent of Computational Fluid Dynamics (CFD), a lot of progress has been seen in the areas of atmospheric pollution dispersion and urban air quality, particularly at the local scale. This work invetsigates the flow of a turbulent boundary layer over a realistic urban canopy representing a scale model of the central London area (UK), using CFD simulations. CFD is used in this work as a fit-to-purpose application in order to identify the salient features of the wind flow and to provide appropriate parameterizations for use in operational modelling. A qualitative examination of the flow has shown that the flow is dominated by large vortical structures whose form varies with the particular building obstacle area. These vortical structures form systematic air flow exchanges between the in-canopy and above-canopy flows, which can be modelled using the concept of an exchange velocity. Finally this parametrisation the definition of a "breathability" property of the canopy that can be used as merit index in sustainable urban design.*

1 INTRODUCTION

The flow of the wind through and above an urban area, and the dispersion of material in this flow are often addressed by considering the problem at different scales. It has been customary to use spatial scales, as they can describe the major urban flow features; however the spatial scales can also be related to time scales through an approrpiate wind advection velocity. The range of spatial scales consists of: the regional scale (of the order of hundreds of kilometres), the city scale (up to 10 or 20 km), the neighbourhood scale (up to 1 or 2km) and the street canyon scale (less than 100-200m). Within the city the wind flow directly affects air quality, pedestrian mobility and comfort. The same wind flow, but on a larger scale, represents the wind environment within which new buildings are to be placed; this is of concern both for wind-loading problems and for the provision of clean air to the buildings, as well as the removal of exhaust air from the urban canopy[2]. This wind environment and the building construction affect some of the exchange processes between the building interior and exterior, and consequently, building air quality and energy use. Additionally, hazardous materials are normally prohibited from heavily populated areas, but where this is not the case there is a need for emergency authorities and civil defense personnel to have operational tools available to determine what action to take in case of an accident[2]. Very recently there has been increased concern about the non-accidental release of hazardous meterials in urban areas.

This study addresses the wind flow through and above an urban area at the neighbourhood scale over a realistic urban canopy. It forms part of a larger research effort under the UK consortium DAPPLE[4] which involves field experiments, wind-tunnel experiments, computational modelling, meteorological measurements, and model development and evaluation. In this paper we present and discuss qualitatively the results from a numerical study of the wind flow over a central London area. The qualitative examination of the flow aims at the identification of the salient features of the flow and the use of this information to enhance operational modelling. A parallel aim of this work is to identidy appropriate properties of urban canopies for use as merit index in sustainable design.

2 THEORY

2.1 Boundary layer meteorology

Urban obstacles exert a relatively large drag force on the atmosphere. This can be treated by standard atmospheric formulas[14], as long as the mean building height is small compared to the surface boundary-layer depth (usually aproximately 100-200m) and the surface has some statistical homogeneity. Figure 1 depicts a typical urban area, showing the three major sublayers: the urban canopy sublayer, the roughness sublayer and the inertial sublayer. The inertial sublayer is the area where the boundary layer has adapted to the integrated effect of the underlying urban surface, and it is treated by standard boundary-layer formulas. In the urban canopy sublayer the flow at a specific point is directly affected by the local obstacles, and in the roughness sublayer the flow is still adjusting to the effects of many obstacles.

The surface shear stress (averaged over the urban surface) defines a friction velocity, u_*, that can be used to derive wind and turbulence profiles. It is assumed that, regardless of the underlying surface, the wind speed at the top of the boundary layer (i.e., at a height of ~500 to 1000m) is approximately equal to the equilibrium wind speed defined by the geostrophic wind speed, which is based on the synoptic pressure gradient. The wind-speed profile conforms to Monin-Obukhov similarity theory, with friction velocity (u^*) as the key scaling velocity, and two additional scaling lengths, the surface roughness length (z_0) and the surface displacement length (d). For neutral or adiabatic conditions the wind-speed profile can be described by

$$U = \frac{u^*}{\kappa}\ln\left(\frac{z-d}{z_0}\right) \qquad (1)$$

where κ is the von Karman's constant taken to be 0.4.

2.2 Urban morphology

Estimates of the surface roughness length and the displacement length can be made using information about urban morphology, e.g. building sizes and spacing. The total building plan area, A_p, and the total building frontal area, A_f, in a building lot area, A_T, can be used to define suitable parameters for use in many empirical urban boundary-layer formulas: the dimensionless frontal area, λ_f, and the dimensionless plan area, λ_p. The frontal area parameter is more important to drag because it represents the surface facing the wind flow. Due to a different λ_f parameter, different urban flows can be observed. For example, at small λ_f the buildings act in isolation; at larger λ_f the building wakes interfere with each other,and at even larger λ_f a skimming flow over the buildings has limited direct penetration into the spaces between the buildings[8]. Typical values of λ_f are ~0.1 for areas with moderate density of buildings and ~0.3 for downtown areas[2].

Several field and laboratory data sets as well as theoretical and empirical formulas in the literature have been used to derive relations between the non-dimensional frontal area parameter λ_f, and the non-dimensional surface

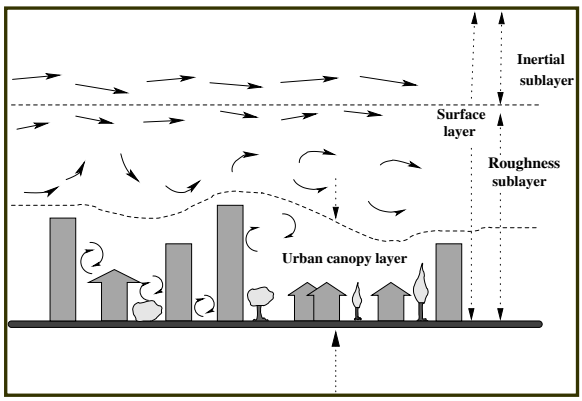


Figure 1: Schematic of the flow through and over an urban area

roughness and dispalcement lengths, as normalised over the average building height[9].For example the surface roughness and the displacement lengths (z_0 and d respectively)., for typical downtown areas (λ_f ~ 0.3) with average building height H_r can be described by:

$$\frac{z_0}{H_r} = \lambda_f \qquad \text{and} \qquad \frac{d}{H_r} = 0.7 + 0.35\left(\lambda_f - 0.15\right) \qquad (2)$$

Although urban areas might be better represented as regions (neighbourhoods) of gradually varying surface

roughness length, an approach that treats an urban area as one of a uniform surface roughness length and an equilibrium boundary layer profile is often adopted; this is mainly because of the difficulties involved in processing geometrical data from multiple individual buildings that are needed to apply some morphological methods in the deduction of the surface roughness and displacement lengths[10] .

3 METHODOLOGY

3.1 The full-scale urban topography and the CFD geometry

The urban area under investigation centers on the intersection of Marylebone Road and Gloucester Place in Central London, UK and spans an area of 250m-radius. Marylebone Road is a busy dual carriageway, in places up to 7 lanes, and forms the northern boundary of the London Congestion-Charging Zone; it is 20˚ North of due East. From topographical maps of the area, the average building height of the area is estimated to 22m; from the specific building arrangement the urban morphometry parameters are estimated as λ_f ~0.25 and λ_p~0.5. In the CFD simulation, a 1:200 scale model of the urban area is represented and the resolved built area includes 42 buildings, with an average model building height H of 0.11m. The resolved built area is depicted in Figure 3. The computational domain however extends a further distance both in the upstream and downstream directions of the outermost features of the built-area model, according to published recommendations[3]. The domain size is 6.2m (x-direction) x3m (y-direction) x0.55m (z-direction) corresponding to 56Hx27Hx5H. The geometry was prepared using the pre-processor GAMBIT[7].

3.2 Governing equations

The flow simulation is produced by solving the 3-D Navier-Stokes equations coupled with a seven-equation Reynolds-Stress turbulence model (RSM)[11], implemented in the CFD code FLUENT[7]. RSM provides the necessary functionality and allows for anisotropic turbulence (which is known to be a characteristic feature behind bluff body obstacles), despite its inherent complexity and a tendency to overestimate the wake region behind a single cube obstacle[12]. Steady simulations were carried out, giving a time-averaged view of the flow, neglecting possible phenomena such as unsteady vortex shedding and occasional sweeps of air from above the canopy[6]. The focus of this work is on novel ways of examining the mean features of urban canopy flow and achieving fit-for-purpose accuracy for the above aims rather than precise prediction accuracy when comparing with experimental profiles.

3.3 The numerical implementation

The finite-volume approach using a high-resolution unstructured, tetrahedral mesh is used to descritise the computational domain. The cell edge size is typically 0.03H, where H is the average model building height within the canopy, and near the ground, with a growth rate of 1.1 applied. Grid refinement tests led to a total number of cells near 4 million in order to resolve sufficiently the computational domain.

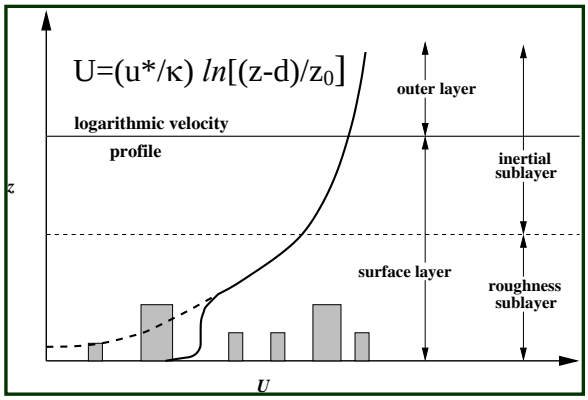


Figure 2. The spatially averaged mean velocity profile near an urban area.

3.4 Boundary conditions

An inflow condition was applied at the leftmost (upwind) side (y-z plane) of the domain (with reference to Figure 3). An inlet wind profile is specified according to equation (1); the values of the surface roughness and displacement lengths were calculated from equation (2) using the morphology parameters of the modeled urban area). For the turbulence parameters, profiles of the turbulent kinetic energy and the turbulent dissipation rate were specified according to design specification rules[13] and measurements from wind tunnel experiments using the same model geometry[4].

A pressure outlet condition was applied at the rightmost (downwind) side (y-z plane) of the domain. Symmetry boundary conditions (zero mean flow and zero normal gradient of all quantities at the plane) were applied at the northern and southern vertical sides of the domain (y-direction pointing to the North in Figure 3). A symmetry condition was also applied at the horizontal top of the domain (x-y plane). A wall boundary condition (with standard wall functions) was applied at the ground and building surfaces.

4 RESULTS

4.1 Overall remarks

General observations of the flow characteristics within the modelled urban area were made by releasing mass-less tracer particles from selected buildings and other surfaces of interest. Contours of velocity, turbulence and pressure were also examined. For the purpose of this paper the discussion will concentrate on the velocity field in a qualitative manner. The main purpose is to illustrate and identify salient features of the flow that can be used and then modeled in operational models, as well as used as merit indices in sustainable urban planning.

Figure 4 shows pathlines followed by different particles released at the inlet plane; the different colours denote different particle identity. This visualization shows that the flow is dominated by large vortical structures whose form varies with the particular building obstacle area. These vortical structures mix air from the top of the urban canopy with air within the canopy and vice-versa. Some minor vortical structures also seem to mix air across the width of canopy; from within but across smaller distances. The vortex patterns seem to originate in the short canyons behind buildings and are different from the archetypal 2-D street canyon flows. It must be recognized that the formations observed here are time-averaged simulated results, and the flow will in fact be unsteady in nature. Similar patterns have been observed however both in the unsteady and the corresponding steady simulations of regular cube arrays[9]. This visualization suggests that the form of these vortices may influence the effectiveness of upward mixing (and air replenishing) within the canopy. Thus the concept of an exchange velocity, as defined by the average velocity of transfer of mass into or out of the canopy can be used. This exchange velocity would be directly relevant to the removal of polluted air and heat from urban canopies. This would point as well towards appropriate design rules for greener and sustainable urban planning.

4.2 Velocity field

The wind velocity vectors and contours at different horizontal slices are shown in Figures 5-6. Wind velocity vectors and contours on a vertical slice are shown in Figure 7.

Figure 5 depicts the wind velocity vectors on a horizontal surface slice at a vertical height z=H/2. The incoming flow is in the direction from left to right. The cross-sections of buildings projected on this surface are shown in grey and act as walls (zero velocity). A strong flow channeling along large roads (e.g. Marylebone and the outermost parallels) is observed. Strong preferential channeling is also observed in the perpendicular direction to the flow, in the case of Gloucester Road -from south to north. Cross-wind channeling however appears less strong than the wind-wise direction. No significant reversed or recirculation flows are observed on this surface, despite the observed vortical structures in the 3-D domain. The oval surface on the northeast of Marylebone Road (with corresponding velocity vectors that appear) is Regent Park, which is below the height of the selected slide; its projected surface appears in this cut for reference.

Figure 6 depicts the contours of wind speed on three different horizontal surfaces, at heights z=H/2 (corresponding to the presented case of wind velocity vectors), H, and 2H - from top to bottom respectively. The colourbar is clipped to the global maximum and minimum values of speed. A key flow feature is the presence of wakes of buildings, formed as the air flows over and around building obstacles. As we move closer to the ground, more buildings are present inside the wakes of neighbouring buildings. Another flow feature observed is the presence of shear layers, e.g. near the edge of the wedge-like building in the contour plot at z=H; overall, higher velocities are also seen throughout the domain. In the contour plot at height z=2H, only the highest buildings (the three towers) are met by the air flow; the wakes of these buildings are the main disturbances to a bulk flow of seemingly higher velocities than those observed at lower heights.

Figure 7 depicts the contours of wind speed and the velocity vectors on a vertical slide at y=4.5H. The incoming flow is in the direction from left to right, and the colourbar is clipped to the global maximum and minimum values. The speed contour plot shows how air movement within a street canyon is more, or less, enhanced, depending mostly on the upwind buildings but sometimes as well on the downwind buildings. In the last downwind canyon for example, depicted in the figure, air movement is induced from a small recirculation region formed by the main wind flow facing the tower (last building) just downwind the canyon.

4.3 Pressure and turbulence fields

Overall pressure variations are consistent with changes in momentum fluxes and the drag exerted from the buildings to the flow. A higher pressure on the upwind side of a building and lower pressure on the downwind side is generally observed. The turbulence field is subject of a more substantial on-going investigation and will be addressed extensively in future work. Specifically Reynolds stresses can be related to the momentum fluxes across the canopy top, and hence to air flow exchange between the in-canopy and above-canopy flows; this flow exchange is a key feature in the characterisation of the *breathability* of an urban canopy. For the scope of this paper, it can be noted that the overall variations of turbulence kinetic energy and turbulent dissipation rate are consistent with known features of flows behind bluff bodies[5].

4 CONCLUSIONS

A numerical investigation of the wind flow over and around a realistic urban canopy was performed using a CFD simulation with the RSM turbulence model and a high-resolution mesh. A qualitative examination of the results shows that the flow is dominated by large vortical structures whose form varies with the particular building obstacle area. These vortical structures form systematic airflow exchanges between the in-canopy and above-canopy flows, as well as cross-canopy exchanges. The vortex patterns appear to originate in the short canyons behind buildings and are different from the archetypal 2-D street canyon flows. The concept of a bulk air-exchange velocity is appropriate to define a dynamic property of an urban area - its *breathability* - that will also be useful as merit index in the design of sustainable urban development.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the EPSRC -UK for generous funding, and the DAPPLE Consortium (University of Bristol, University of Cambridge, Imperial College, University of Leeds, University of Reading and University of Surrey) for enriching discussions and collaboration.

.

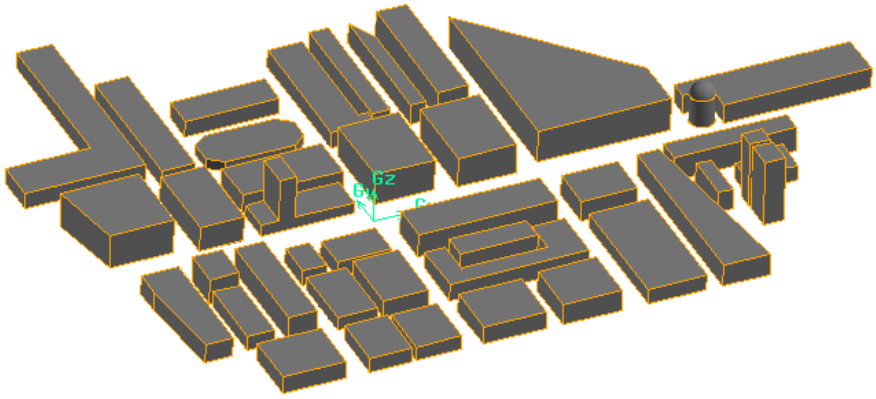


Figure 3. The 1:200 scale model of the Marylebone area topography that was represented in the CFD simulation. The computational domain extends however to a further distance upwind and downwind from the outer most obstacles as flat terrain described in section 3.

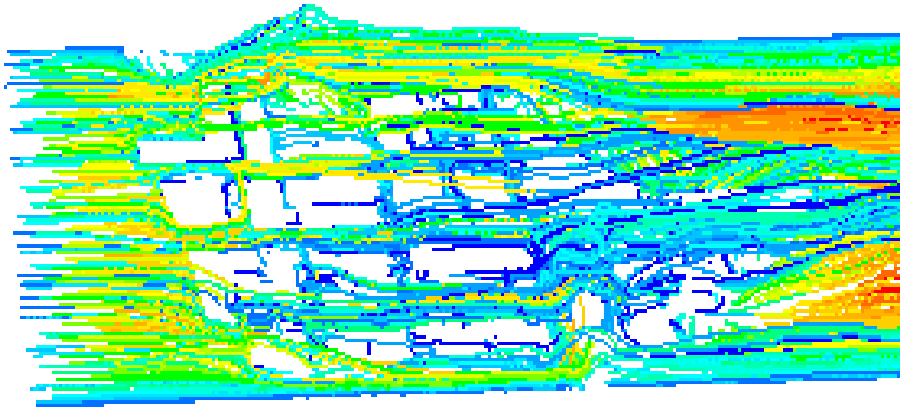


Figure 4: Path lines of mass-less tracer particles to visualize the 3-D flow.

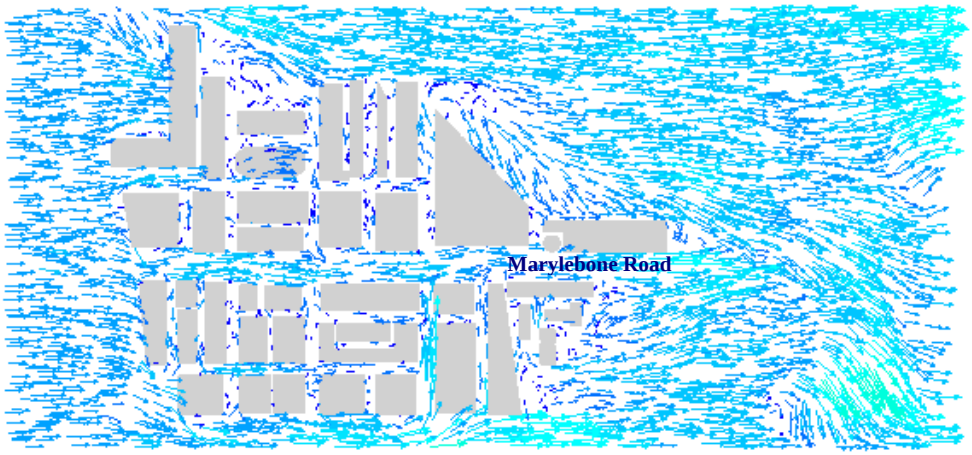


Figure 5. Wind velocity vectors on a horizontal surface at a height z = H/2

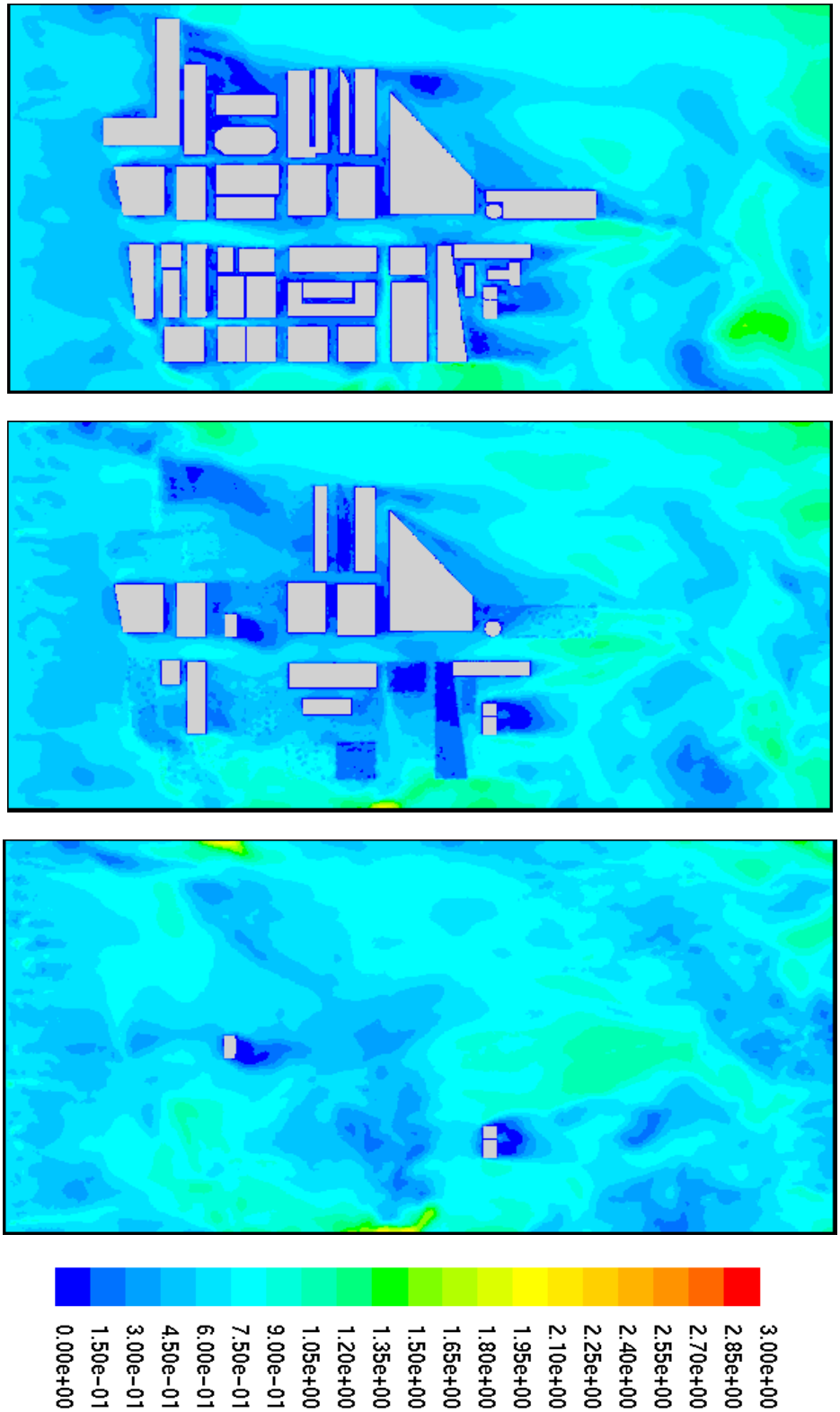


Figure 6: Wind speed contours *(m/s)* on horizontal surfaces at different heights: z =H/2, H, 2H (top to bottom respectively); the colourbar is clipped to the global maximum and minimum values.

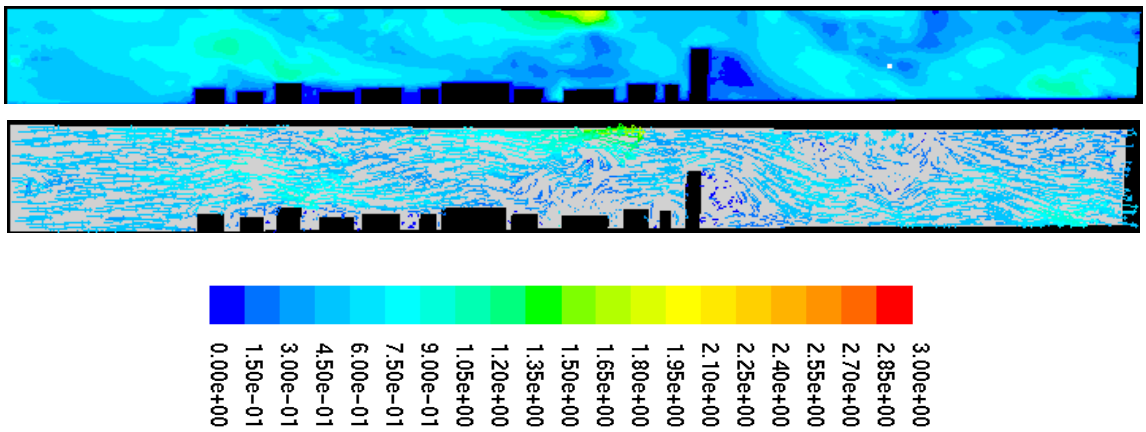


Figure 7: Wind speed contours (top) and velocity vectors (bottom) (*m*/s) on a vertical surface at y=+4.5H; the colourbar is clipped to the global maximum and minimum values.

REFERENCES

[1] Bentham, T. & Britter R.E. (2003). "Spatially averaged flow within obstacle arrays". Atmospheric Environment, 37, 2037-2043.

[2] Britter R.E. & Hanna S.R. (2003). "Flow and dispersion in urban areas". Annu. Rev. Fluid Mech. 35:469-96.

[3] Cowan, I.R., Castro, I.P., Robins, A.G. (1997). "Numerical considerations for Simulation of Flow and Dispersion around Buildings", J. Wind Engrg. Ind. Aero 67&68, 535-545.

[4] DAPPLE: Dispersion of Air Pollution and Penetration to the local environment (2003). Official website: http://www.dapple.org.uk

[5] Durbin, P.A. (1993). "A Reynolds stress model for near-wall turbulence". J. Fluid Mechanics, 249, 465-498.

[6] Finnigan, J.J. (2000). "Turbulence in Plant Canopies". Annu. Rev. Fluid. Mech. 32, 519-571.

[7] FLUENT 6.1(2003). *User Manual*, Fluent.Inc.,

[8] Grimmond, C.S.B.& Oke T.R. (1999). "Aerodynamic properties of urban areas derived from analysis of surface form". J. Appl. Meteorol. 38:1261-92.

[9] Hamlyn, D. & Britter R.E. (2005). "A numerical study of the flow field and exchange processes within a canopy and urban-type roughness". Atmospheric Environment. In press.

[10] Hanna S.R. & Britter R.E. (2002). *Wind flow and Vapour Cloud Dispersion at Industrial Sites*. New York: Am. Inst. Chem. Eng.

[11] Launder, B.E., Reece, G.J., Rodi, W. (1975). "Progress in the development of a Reynolds-Stress turbulence closure". J. Fluid Mech. Vol. 68, pp. 537-566.

[12] Murakami, S. (1998). "Overview of turbulence models applied in CWE-1997". J. Wind Engrg. Ind. Aero 74-76, 1-24.

[13] Richards, P.J.&Hoxey, R.P. (1993). "Appropriate boundary conditions for computational wind engineering models using the k-ε turbulence model". J. Wind Engrg. Ind. Aero. 46&47, 145-153.

[14] Stull R.B. (1997). *An Introduction to Boundary Layer Meteorology*. Dordrecht:Kluwer Atmos. Publ. 670 pp.

Author index

E

F

G

H

I

J

K

L

M

N

P

R

S

T

U

V

W

X

Y

Z