

Studies in Applied Mechanics

47

**Advances in Adaptive
Computational
Methods in Mechanics**

edited by

P. Ladevèze

J.T. Oden

Elsevier

STUDIES IN APPLIED MECHANICS 47

Advances in Adaptive Computational Methods in Mechanics

STUDIES IN APPLIED MECHANICS

20. Micromechanics of Granular Materials (Satake and Jenkins, Editors)
21. Plasticity. Theory and Engineering Applications (Kaliszky)
22. Stability in the Dynamics of Metal Cutting (Chiriacescu)
23. Stress Analysis by Boundary Element Methods (Balas, Sládek and Sládek)
24. Advances in the Theory of Plates and Shells (Voyiadjis and Karamanlidis, Editors)
25. Convex Models of Uncertainty in Applied Mechanics (Ben-Haim and Elishakoff)
28. Foundations of Mechanics (Zorski, Editor)
29. Mechanics of Composite Materials - A Unified Micromechanical Approach (Aboudi)
31. Advances in Micromechanics of Granular Materials (Shen, Satake, Mehrabadi, Chang and Campbell, Editors)
32. New Advances in Computational Structural Mechanics (Ladevèze and Zienkiewicz, Editors)
33. Numerical Methods for Problems in Infinite Domains (Givoli)
34. Damage in Composite Materials (Voyiadjis, Editor)
35. Mechanics of Materials and Structures (Voyiadjis, Bank and Jacobs, Editors)
36. Advanced Theories of Hypoid Gears (Wang and Ghosh)
- 37A. Constitutive Equations for Engineering Materials
Volume 1: Elasticity and Modeling (Chen and Saleeb)
- 37B. Constitutive Equations for Engineering Materials
Volume 2: Plasticity and Modeling (Chen)
38. Problems of Technological Plasticity (Druyanov and Nepershin)
39. Probabilistic and Convex Modelling of Acoustically Excited Structures
(Elishakoff, Lin and Zhu)
40. Stability of Structures by Finite Element Methods (Waszczyszyn, Cichoń and Radwańska)
41. Inelasticity and Micromechanics of Metal Matrix Composites (Voyiadjis and Ju, Editors)
42. Mechanics of Geomaterial Interfaces (Selvadurai and Boulon, Editors)
43. Materials Processing Defects (Ghosh and Predeleanu, Editors)
44. Damage and Interfacial Debonding in Composites (Voyiadjis and Allen, Editors)
45. Advanced Methods in Materials Processing Defects (Predeleanu and Gilormini, Editors)
46. Damage Mechanics in Engineering Materials (Voyiadjis, Ju and Chaboche, Editors)
47. Advances in Adaptive Computational Methods in Mechanics (Ladevèze and Oden, Editors)

General Advisory Editor to this Series:

Professor Isaac Elishakoff, Center for Applied Stochastics Research, Department of Mechanical Engineering, Florida Atlantic University, Boca Raton, FL, U.S.A.

STUDIES IN APPLIED MECHANICS 47

Advances in Adaptive Computational Methods in Mechanics

Edited by

P. Ladevèze

*LMT - Cachan
ENS de Cachan, CNRS, Université Paris 6
61, avenue du Président Wilson
94235 Cachan Cedex, France*

J.T. Oden

*TICAM
The University of Texas at Austin
3500 West Balcones Center Drive
Austin, Texas 78759, USA*



1998

ELSEVIER

Amsterdam - Lausanne - New York - Oxford - Shannon - Singapore - Tokyo

ELSEVIER SCIENCE Ltd
The Boulevard, Langford Lane
Kidlington, Oxford OX5 1GB, U.K.

Copyright © 1998 Elsevier Science Ltd

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, electrostatic, magnetic tape, mechanical, photocopying, recording or otherwise, without permission in writing from the publishers.

First edition 1998

Library of Congress Cataloging-in-Publication Data

Advances in adaptive computational methods in mechanics / edited by P. Ladevèze and J.T. Oden.

p. cm. -- (Studies in applied mechanics ; 47)

Papers selected from a workshop held in Cachan, France, Sept. 17-19, 1997.

Includes index.

ISBN 0-08-043327-8

1. Mechanics, applied--Mathematics. 2. Estimation theory.

I. Ladevèze, Pierre, 1945- II. Oden, J. Tinsley (John Tinsley), 1936- . III. Series.

TA350.A38 1998

620.1--dc21

98-15962
CIP

ISBN: 0-080-43327-8

∞ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper)

Transferred to digital printing 2006

Preface

Mastering modelling, and in particular numerical models, is becoming a crucial and central question in modern computational mechanics. Various tools, able to quantify the quality of a model with regard to another one taken as the reference, have been derived. Applied to computational strategies, these tools lead to new computational methods which are called "adaptive". The present book is concerned with outlining the state of art and the latest advances in both these important areas. Contributors belong to two communities : Applied Mathematics and Mechanics.

Papers are selected from a Workshop (Cachan 17-19 September 1997) which is the third of a serie devoted to Error Estimators and Adaptivity in Computational Mechanics and started at Austin (16-18 October 1989) and Cracow (14-16 October 1991). Cachan-Workshop dealt with latest advances in adaptive computational methods in mechanics and their impacts on solving engineering problems. It was centered too on providing answers to simple questions as :

- What is being used or can be used at present to solve engineering problems ?
- What should be the state of art in the year 2000 ?
- What are the new questions involving error estimators and their applications ?

Six distinct chapters define the main book topics :

- Error estimators and adaptive computational methods for linear problems
- Modelling error estimators and adaptive modelling strategies
- Local error estimators for linear problems
- Error estimators for non linear time-dependent problems and adaptive computational methods
- Adaptive computational methods for 3D problems
- Error estimators and mesh adaptivity for vibration, acoustics and electromagnetics problems

We hope that these contributions provide answers to the basic questions addressed in Cachan-Workshop.

P. Ladevèze and J.T. Oden

This Page Intentionally Left Blank

CONTENTS

Preface	v
---------------	---

PART 1: ERROR ESTIMATORS AND ADAPTIVE COMPUTATIONAL METHODS FOR LINEAR PROBLEMS

Recovery procedures in error estimation and adaptivity: Adaptivity in linear problems O.C. Zienkiewicz, B. Boroomand and J.Z. Zhu	3
The relationship of some a posteriori error estimators J.Z. Zhu and Z. Zhang	25
A technique for a posteriori error estimation of h - p approximations of the Stokes equations J.T. Oden and S. Prudhomme	43
A mathematical framework for the P. Ladevèze a posteriori error bounds in finite element methods P. Destuynder, M. Collot and M. Salaün	65

PART 2: MODELLING ERROR ESTIMATORS AND ADAPTIVE MODELLING STRATEGIES

Adaptive finite element in elastoplasticity with mechanical error indicators and Neumann- type estimators E. Stein, F.J. Barthold, S. Ohnimus and M. Schmidt	81
A reliable a posteriori error estimator for adaptive hierarchic modelling M. Ainsworth and A. Arnold	101
A two-scale strategy and a posteriori error estimation for modeling heterogeneous structures N. Moës, J.T. Oden and K. Vemaganti	115

A modelling error estimator for dynamic structural model updating P. Ladeveze	135
--	-----

PART 3: LOCAL ERROR ESTIMATORS FOR LINEAR PROBLEMS

A posteriori estimation of the error in the error estimate I. Babuska, T. Strouboulis, S.K. Gangaraj, K. Copps and D.K. Datta	155
Bounds for linear-functional outputs of coercive partial differential equations: Local indicators and adaptive refinement J. Peraire and A.T. Patera	199
On adaptivity and error criteria in meshfree methods T. Belytschko, W.K. Liu and M. Singer	217

PART 4: ERROR ESTIMATORS FOR NON LINEAR TIME-DEPENDENT PROBLEMS AND ADAPTIVE COMPUTATIONAL METHODS

A posteriori constitutive relation error estimators for nonlinear finite element analysis and adaptive control P. Ladevèze and N. Moës	231
A review of a posteriori error estimation techniques for elasticity problems R. Verfürth	257
A posteriori error control and mesh adaptation for F.E. models in elasticity and elastoplasticity R. Rannacher and F.T. Suttmeier	275
An adaptive finite element approach in associated and non-associated plasticity considering localization phenomena W. Wunderlich, H. Cramer and G. Steinl	293
An α -adaptivity approach for advective-diffusive and fluid flow problems E. Onate, J. Garcia and S. Idelsohn	309
Adaptive strategy for transient/coupled problems. Applications to thermoelasticity and elastodynamics D. Aubry, D. Lucas and B. Tie	325
Error estimation and adaptive finite element analysis of softening solids A. Huerta, P. Diez, A. Rodriguez-Ferran and G. Pijaudier-Cabot	333

Aspects of adaptive strategies for large deformation problems at finite inelastic strains D. Perić, M. Dutko and D.R.J. Owen	349
Adaptive solutions in industrial forming process simulation T. Coupez, L. Fourment and J.L. Chenot	365
Recovery procedures in error estimation and adaptivity: Adaptivity in non-linear problems of elasto-plasticity behaviour B. Boroomand and O.C. Zienkiewicz	383
 PART 5: ADAPTIVE COMPUTATIONAL METHODS FOR 3D LINEAR PROBLEMS	
3-D error estimation and mesh adaptation using improved R.E.P. method P. Beckers and E. Dede	413
Adaptive methods and related issues from the viewpoint of hybrid equilibrium finite element models O.J.B. Almeida Pereira, J.P. Moitinho De Almeida and E.A.W. Maunder	427
Error estimator and adaptivity for three-dimensional finite element analyses P. Coorevits, J.P. Dumeau and J.P. Pelle	443
 PART 6: ERROR ESTIMATORS AND MESH ADAPTIVITY FOR VIBRATION, ACOUSTICS AND ELECTROMAGNETICS PROBLEMS	
Error estimation and adaptivity for h-version eigenfrequency analysis N.E. Wiberg and P. Hager	461
Error estimation and adaptivity for the finite element method in acoustics Ph. Bouillard and F. Ihlenburg	477
Error through the constitutive relation for beam or C^0 plate finite element: Static and vibration analyses Ph. Boisse and G. Coffignal	493
A posteriori error analysis for steady-state Maxwell's equation L. Demkowicz	513
 AUTHOR INDEX	 527

This Page Intentionally Left Blank

PART 1

ERROR ESTIMATORS AND ADAPTIVE COMPUTATIONAL METHODS FOR LINEAR PROBLEMS

This Page Intentionally Left Blank

Recovery Procedures in Error Estimation and Adaptivity : Adaptivity in Linear Problems

O.C. Zienkiewicz^a, B. Boroomand^a and J.Z. Zhu^b

^a University of Wales, Swansea, U.K.

^b U.E.S. Inc. Annapolis, U.S.A.

ABSTRACT

The procedures of error estimation using stress or gradient recovery were introduced by Zienkiewicz and Zhu in 1987 [3] and with the improvement of recovery procedures (introduction of Superconvergent Patch Recovery SPR) since that date the authors have succeeded in making these error estimators the most robust of those currently available. Very recently introduced methods of recovery such as REP (Recovery by Equilibrium in Patches) allow general application.

In this paper the general idea of using recovery based error estimation in adaptive procedure of linear problems is explained through which the latest developments in recovery techniques is described. The fundamental basis of a 'patch test' introduced by Babuška *et al* [11]- [14] is explained and applied to the recovery based error estimators using both SPR and REP.

1. INTRODUCTION

Two types of procedures are currently available for deriving error estimators. They are either *Residual based* or *Recovery based*.

The residual based error estimators were first introduced by Babuška and Rheinboldt in 1978 [1] and have been since used very effectively and further developed by many others. Here substantial progress was made as recently as 1993 with the introduction of so called residual equilibration by Ainsworth and Oden [2].

The recovery based error estimators are, on the other hand, more recent having been first introduced by Zienkiewicz and Zhu in 1987 [3]. Again these were extensively improved by the introduction of new recovery processes. Here in particular the , so called, SPR (or Superconvergent Patch Recovery) method introduced in 1992 by the

same authors [4]- [5] has produced a very significant improvement of performance of the Recovery based methods. Many others attempted further improvement [7]- [9] but the simple procedure originally introduced remains still most effective.

In this paper we shall concentrate entirely on the *Recovery based method* of error estimation. The reasons of this are straightforward:

- (i) the concept is simple to grasp as the approximation of the error is identified as the difference between the recovered solution u^* and the numerical solution u_h ; thus the estimate

$$\|e^*\| = \|u^* - u_h\| \tag{1}$$

in any norm is achieved simply by assuming that the exact solution u can be replaced by the recovered one.

- (ii) as some recovery process is invariably attached to numerical codes to present more accurate and plausible solutions, little additional computation is involved;
- (iii) if the recovery process itself is superconvergent, it can be shown [5] that the estimator will always be asymptotically exact (we shall repeat the proof of this important theorem in the paper);
- (iv) numerical comparisons on bench mark problems and more recently by a 'patch test' procedure introduced by Babuška *et al* [11]- [14] have shown that the recovery procedures are extremely accurate and robust. In all cases they appear to give a superior accuracy of estimation than that achievable by *Residual based methods*.

It is of interest to remark that in many cases it is possible to devise a Residual method which has an identical performance to a particular recovery process. This indeed are first noted by Rank and Zienkiewicz in 1987 [15] but later Ainsworth and Oden [16] observed that this occurs quite frequently. In a recent separate paper Zhu [17] shows that:

- (v) for every *Residual based* estimator there exists a corresponding *Recovery based* process. However the reverse is not true. Indeed the Recovery based methods with optimal performance appear not to have an equivalent Residual process and hence, of course, the possibilities offered by Recovery methods are greater.

In this paper we shall describe in detail the SPR based recovery as well as a new alternative REP process which appears to be comparable in performance.

With error estimation achieved the question of adaptive refinement needs to be addressed. Here we discuss some procedures of arriving at optimal mesh size distribution necessary to achieve prescribed error.

2. SOLUTION RECOVERY AND ERROR ESTIMATION

In what follows we shall be in general concerned with the numerical solution of problems in which a differential equation of the form given as:

$$\mathbf{S}^T \mathbf{D} \mathbf{S} \mathbf{u} + \mathbf{b} = \mathbf{0} \quad (2)$$

has to be solved in a domain Ω with suitable boundary condition on:

$$\partial\Omega = \Gamma \quad (3)$$

In above \mathbf{S} is a differential operator usually defining stresses or fluxes as

$$\boldsymbol{\sigma} = \mathbf{D} \mathbf{S} \mathbf{u} \quad (4)$$

where \mathbf{D} is a matrix of physical parameters.

We shall not discuss here the detail of the finite element approximation which can be found in texts [18]. In there the unknown function \mathbf{u} is approximated as:

$$\mathbf{u} \approx \mathbf{u}_h = \mathbf{N} \bar{\mathbf{u}} \quad (5)$$

which results in approximate stresses being:

$$\boldsymbol{\sigma}_h = \mathbf{D} \mathbf{S} \mathbf{u}_h = \mathbf{D} \mathbf{B} \bar{\mathbf{u}} \quad (6)$$

In above

$$\mathbf{N} = \mathbf{N}(x_i) \quad i = 1 - 3 \quad \text{and} \quad \mathbf{B} = \mathbf{S} \mathbf{N} \quad (7)$$

are the spatially defined shape-functions.

The solution error is defined as the difference between the exact solution and the numerical one. Thus for instance the displacement error is:

$$\mathbf{e}_u = \mathbf{u} - \mathbf{u}_h \quad (8)$$

and the stress error is

$$\mathbf{e}_\sigma = \boldsymbol{\sigma} - \boldsymbol{\sigma}_h \quad (9)$$

at all points of the domain. It is, however, usual to define the error in terms of a suitable norm which can be written as a scalar value

$$e = \|e_u\| = \|u - u_h\| \quad (10)$$

for any specific domain Ω . The norm itself specifies the nature of the quantity defined. The well known energy norm is given, for instance, as

$$\|e\|^E = \left[\int_{\Omega} (\sigma - \sigma_h)^T D^{-1} (\sigma - \sigma_h) d\Omega \right]^{\frac{1}{2}} \quad (11)$$

With the Recovery process we devise a procedure which gives, by suitable post-processing of u_h and σ_h , the values of u^* and / or σ^* which are (hopefully) more accurate and we estimate the norm of the error as:

$$\|e\| \approx \|e^*\| = \|u^* - u_h\| \quad (12)$$

In the case of energy norm we have:

$$\|e^*\|^E = \left[\int_{\Omega} (\sigma^* - \sigma_h)^T D^{-1} (\sigma^* - \sigma_h) d\Omega \right]^{\frac{1}{2}} \quad (13)$$

The effectivity index of any error estimator is defined as:

$$\theta = \frac{\|e\|_{est}}{\|e\|} \quad (14)$$

or in the case of recovery based estimators:

$$\theta^* = \frac{\|e^*\|}{\|e\|} \quad (15)$$

A theorem proposed by Zienkiewicz and Zhu [5] shows that for all estimators based on recovery we can establish the following bound for the effectivity:

$$1 - \frac{\|\tilde{e}\|}{\|e\|} < \theta^* < 1 + \frac{\|\tilde{e}\|}{\|e\|} \quad (16)$$

In above e is the actual error (viz Equation (10) and (11)) and \tilde{e} is the error of the recovered solution i.e.

$$\|\tilde{e}\| = \|u - u^*\| \quad (17)$$

The proof of the above theorem is straight forward, if we rewrite Equation (12) as:

$$\|e^*\| = \|u^* - u_h\| = \|(u - u_h) - (u - u^*)\| = \|e - \tilde{e}\| \quad (18)$$

Using now the triangle inequality we have:

$$\|e\| - \|\tilde{e}\| \leq \|e^*\| \leq \|e\| + \|\tilde{e}\| \quad (19)$$

from which the inequality (16) follows after division by $\|e\|$. Two important conclusions follow:

- (1) that *any recovery process* which result in reduced error will give a reasonable error estimator and, more importantly,
- (2) if the recovered solution converges at a higher rate than the finite element solution we shall always have asymptotically exact estimation.

To prove the second point we consider a typical finite element solution with shape-functions of order p where we know that the error (in energy norm) is:

$$\|e\| = O(h^p) \quad (20)$$

If the recovered solution gives an error of a higher order eg.

$$\|\tilde{e}\| = O(h^{p+\alpha}) \quad |\alpha| > 0 \quad (21)$$

then the bounds of the effectivity index are:

$$1 - O(h^\alpha) \leq \theta^* \leq 1 + O(h^\alpha) \quad (22)$$

and the error estimator is asymptotically exact i.e ;

$$\theta \rightarrow 1 \quad h \rightarrow 0 \quad (23)$$

This is a very important property of estimation based on recovery not generally shared by estimation of the residual type.

3. THE SPR RECOVERY PROCEDURE

The SPR recovery is based on the assumption that in all element points fixed in space exist for which superconvergence occurs.

For instance it can be shown that for many elements the unknown u is superconvergent at nodal points [19] and that the derivatives σ are superconvergent on Gauss integration points [18, 20, 21]. These statements are not in general true for distorted, isoparametric, elements but hold well for simple, regular, one or two dimensional elements. Fuller discussion is available in [10].

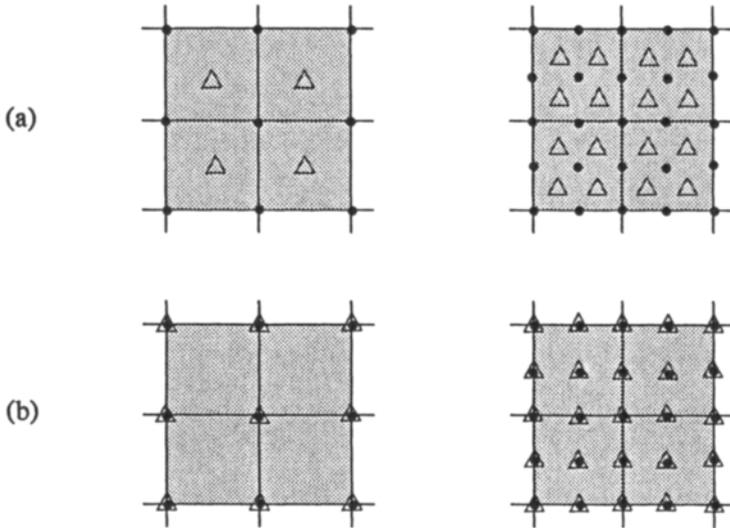


Figure 1: Superconvergent recovery of (a) σ and (b) u for bilinear and biquadratic elements in regular patches

The superconvergent recovery first establishes finite element patches at fully connected, corner nodes and specifies in each element the superconvergent points.

Figure 1 shows an example of two dimensional bi-linear and bi-quadratic elements on which superconvergent points for gradients σ (stresses), and displacement u are shown.

In the SPR procedure we approximate within each patch, each component of the gradient, by a polynomial of order $p + 1$ for u and p for σ and make this approximation fit the superconvergent values in a least square sense.

For the bi-linear elements we have *four* superconvergent points in two dimensions and a least square fit has to be used to find *three* constants of the linear expansion for σ .

For the bi-quadratic elements we have similarly *sixteen* superconvergent points in two dimensions and for σ fit, in a least square sense, the *six* constants of a quadratic expansion for each component are necessary.

However we can proceed differently assuming superconvergence of the function u itself at nodes. Now we have to pass the $p + 1$ polynomial which again is available for both bi-linear and bi-quadratic elements. Differentiation of the superconvergent u expansion will, presumably, yields again superconvergent expansion of σ and such methods has been used by Tabbara em et al [25] for elasticity problems. However, if the energy norm itself includes u values as well as derivatives, it is necessary to use the u expansion directly.

This type of problems is typical of dynamics where acceleration has to be modified and such approximation have been used by Samuelsson *et al* [22]

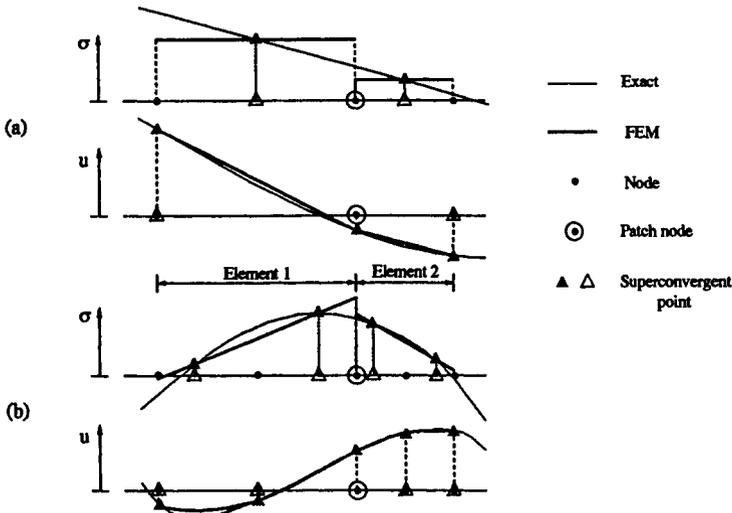


Figure 2: A one dimensional problem. Response to two element patches to exact solution of σ and u one order higher than guaranteed by the polynomial p used (a) $p = 1$ linear elements (b) $p = 2$ quadratic elements

It is important to note that if the points which are being matched are exactly superconvergent then the approximation will again be superconvergent and hence solution giving gradients one order higher be automatically satisfied. This, as we shall see later, gives the fully optimal effectivity index as $\theta^* = 1$ asymptotically.

We note this superconvergence effect in one dimensional problems of Figure 2. Here we note that the solution for u is nodally exact (as shown by Pin Tong in 1969 [23]) and hence both direct interpolation of σ from superconvergent points or use of nodal u values gives identical answers. This example also shows why SPR gives much better answers near boundaries than previously used recovery by averaging of L_2 projections [24].

The simple and direct SPR procedures here described have been variously augmented in attempts to improve their performance. Both Wiberg and Abdulwahab [7, 8] and Blacker and Belytschko [9] for instance added the constraint of satisfying overall equilibrium and/or of stress boundary condition which may be known when using the least square fitting minimization. Such additional constraints will generally counter the requirement of superconvergence as the exact higher order polynomial solution will not fit the solution *exactly*. Thus although some improvement of accuracy can be achieved (usually near the boundaries where some of the exact stresses are known) the performance of recovery process from the point of view of error estimation deteriorates as shown by Babuška *et al* [14]. We thus strongly oppose the use of such devices. However many other recovery methods with sounder foundations may still exist; in the next section we describe one of these.

4. THE REP RECOVERY

Although the SPR recovery has proved to work well generally, its logic for some application where superconvergent points do not in fact exist is doubtful. We have however sought to determine viable recovery alternatives. One of these, known by the acronym REP (Recovery by Equilibrium of Patches) will be here described. This procedure is presented in references [26, 27] with the second one describing a better variant.

To some extent the motivation follows that of Ladevèze [28, 29] who sought to establish (for somewhat different reasons) a fully equilibrating stress field which can replace that of the finite element approximation. However we believe that the process derived in [27] is simpler though equilibration is only approximate.

The starting point is the governing equation, Equation (2), written with the substitution of Equation (4) and in the finite element, Galerkin, approximation sense as:

$$\mathbf{S}^T \boldsymbol{\sigma} + \mathbf{b} = \mathbf{0} \quad (24)$$

This in the finite element approximation is given as;

$$\int_{\Omega} \mathbf{B}^T \boldsymbol{\sigma}_h d\Omega - \int_{\Omega} \mathbf{N}^T \mathbf{b} d\Omega - \int_{\Gamma} \mathbf{N}^T \mathbf{t} d\Omega = \mathbf{0} \quad (25)$$

where the last term comes from the tractions on the boundary of the domain Ω which can, of course, represent the whole of the problem, an element patch or only a single element.

As is well known the stresses $\boldsymbol{\sigma}_h$ which result from the finite element analysis will in general be discontinuous and we shall seek to replace them in *every element patch* by a recovered system which is smooth and continuous.

To achieve the recovery we proceed in exactly an analogous way to that used in the SPR procedure, *first* approximating the stress in each patch by a polynomial of appropriate order $\boldsymbol{\sigma}^*$, *second* using these to obtain nodal values of $\bar{\boldsymbol{\sigma}}^*$ and finally interpolating these by standard shape-functions.

The stress or gradient $\boldsymbol{\sigma}$ is always a vector of appropriate components, which for convenience we write as:

$$\boldsymbol{\sigma} = \begin{Bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{Bmatrix} \quad (26)$$

with $\sigma_1 = \sigma_x$, $\sigma_2 = \sigma_y$ and $\sigma_3 = \tau_{xy}$ in elastic two dimensional analysis for instance.

We shall write each component of above as a polynomial expansion of the form:

$$\sigma_i^* = [1, x, y, \dots] \mathbf{a}_i \quad (27)$$

where \mathbf{a}_i are unknown coefficients.

For equivalence we shall attempt always to ensure that the total smoothed stress $\boldsymbol{\sigma}^*$ satisfies in the least square sense the same patch equilibrium conditions as the finite element solution; i.e that

$$\int_{\Omega_p} \mathbf{B}^T \boldsymbol{\sigma}_h d\Omega \approx \int_{\Omega_p} \mathbf{B}^T \boldsymbol{\sigma}^* d\Omega \quad (28)$$

where for convenient we write

$$\boldsymbol{\sigma}^* = \boldsymbol{\sigma}_1^* + \boldsymbol{\sigma}_2^* + \boldsymbol{\sigma}_3^* \quad (29)$$

$$\boldsymbol{\sigma}_i^* = (\mathbf{1}^i)^T \boldsymbol{\sigma}_i^* \quad \mathbf{1}^2 = [0, 1, 0] \quad \text{etc} \quad (30)$$

and

$$\boldsymbol{\sigma}_i^* = \mathbf{P} \mathbf{a}^i \quad (31)$$

where \mathbf{P} is polynomial terms and \mathbf{a}^i is the set of unknowns coefficients for describing the i th stress component. A similar decomposition is used for $\boldsymbol{\sigma}^h$.

It has been found in practice [26, 27] that the constraint condition provided by Equation (28) are often not sufficient to ensure a non singular minimization to obtain the full set of coefficients \mathbf{a}^i we therefore proposed in [27] that the Equation (28) be replaced by

$$\int_{\Omega_p} \mathbf{B}^T \boldsymbol{\sigma}_i^h d\Omega \approx \int_{\Omega_p} \mathbf{B}^T \boldsymbol{\sigma}_i^* = \left(\int_{\Omega_p} \mathbf{B}^T (\mathbf{1}^i)^T \mathbf{P} d\Omega \right) \mathbf{a}^i \quad (32)$$

As above ensures also satisfaction of Equation (28) and merely enlarges the constrains set ensuring that each component of $\boldsymbol{\sigma}_h$ is equilibrated by the corresponding component of $\boldsymbol{\sigma}^*$. Further the imposition of the approximate equation of (32) allows each set of coefficients \mathbf{a}^i to be solved independently reducing considerably the algorithm cost (and here repeating the procedure used with success in SPR.)

We minimize thus:

$$\Pi = (\mathbf{H}^i \mathbf{a}^i - \mathbf{F}_p^i)^T (\mathbf{H}^i \mathbf{a}^i - \mathbf{F}_p^i) \quad (33)$$

where

$$\mathbf{H}^i = \int_{\Omega_p} \mathbf{B}^T (\mathbf{1}^i)^T \mathbf{P} d\Omega \quad (34)$$

and

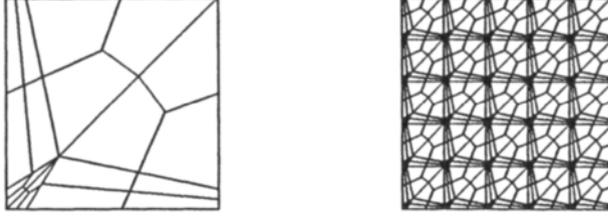


Figure 3: A repeatable patch of arbitrary quadrilateral elements

$$\mathbf{F}_p^i = \int_{\Omega_p} \mathbf{B}^T \boldsymbol{\sigma}_i^h d\Omega = \left(\int_{\Omega_p} \mathbf{B}^T \mathbf{1}^i \otimes \mathbf{1}^i \mathbf{B} d\Omega \right) \bar{\mathbf{u}} \quad (35)$$

resulting in

$$\mathbf{a}^i = [\mathbf{H}^{iT} \mathbf{H}^i]^{-1} \mathbf{H}^{iT} \mathbf{F}_p^i \quad (36)$$

The REP procedure follows precisely the detail of SPR near boundaries and gives overall an approximation which does not require the knowledge of any superconvergent points. The accuracy of both processes is comparable but we quote the new method here as it is our opinion that many other alternative recovery procedures are still possible.

5. ASYMPTOTIC BEHAVIOR AND ROBUSTNESS OF ERROR ESTIMATORS

It is well known that elements in which polynomials of order p are used to represent the unknown will reproduce exactly any problem for which the exact solution is also defined by such a polynomial. Indeed the verification of this behavior is the essential part of the ‘patch test’ which has to be satisfied by all elements to ensure convergence [18].

Thus if we are attempting to determine error in a general smooth solution we will find that this is dominated by terms of the order $(p + 1)$. The response of the patch to exact solution of this order will therefore determine the asymptotic behaviour when the size of the patch and at all the elements tends to zero. If that patch is assumed to be one of repeatable kind, the behaviour of the patch when subjected to exact solution of that order will show the exact finite element solution and the corresponding numerical solution and establish the asymptotic error exactly. Thus any estimator can be compared with this exact value. and the asymptotic effectivity index can be established. Figure 3 show such repeatable patch of quadrilateral elements which evaluates the performance for quite irregular meshes.

In Figure 2 we have indeed shown how true superconvergent behavior reproduces exactly such higher order solution and leads thus to an effectivity index of unity in the asymptotic limit.

In the papers presented recently by Babuška *et al* [11]- [14] the procedure of dealing with such repeatable patches for various patterns of elements is developed. We shall not describe the details of the procedure here in which certain simplification are introduced to avoid some computational labour. However the basic process consists simply of

- (1) subjecting the patches to individual terms of the exact solution of $(p + 1)$ order
- (2) computing the corresponding approximate, finite element, solutions and evaluating the *exact* error of these solution (by comparing it with the exact polynomial)
- (3) applying any error estimator we wish to test to the finite element solution and evaluating its effectivity index θ
- (4) determining the lowest (θ_L) and highest (θ_U) values of the effectivity *for all possible combinations of the expansion terms used in the exact solution*. This is simply accomplished by specifying a suitable eigenvalue problem.

These bounds of the effectivity index are very useful for comparing various error estimators and their behaviour for different mesh and element patterns. However, a single parameter called the *robustness index* has also been devised [11]

$$R = \max \left(|1 + \theta_L| + |1 - \theta_U|, \left| 1 - \frac{1}{\theta_L} \right| + \left| 1 + \frac{1}{\theta_U} \right| \right) \quad (37)$$

A large value of this index obviously indicates a poor performance. Conversely the best behavior is that in which

$$\theta_L = \theta_U = 1 \quad (38)$$

and this gives

$$R = 0 \quad (39)$$

In the series of tests reported in reference [11]- [14] various estimators have been compared. Table 1 below shows the performances of an Equilibrium residual based error estimator with the SPR recovery error estimator for a set of particular patches of triangular elements [13]

Table 1. Robustness index for the Equilibrated Residuals (ERpB) and SPR (ZZ-discrete) estimators for a variety of anisotropic situations and element patterns $p = 2$.

Estimator	Robustness Index
ERpB	10.21
SPR (ZZ-discrete)	0.02

This performance comparison is quite remarkable and it seems that in all the tests quoted by Babuška *et al* [11]- [14] the SPR recovery estimator performs best. Indeed we observe that in many cases of regular subdivision when full superconvergence occur the ideal, asymptotically exact solution characterized by the parameter of Equation (38) is available.

In Table 2 we show some results obtained for regular meshes of triangles and rectangles. In the rectangular elements used for problems of heat conduction type superconvergent points are exact and the ideal result is expected for both linear and quadratic elements. It is surprising that this also occur in elasticity where the proof of superconvergent points is lacking. Further, the REP procedure also seem to yield superconvergence (except for elasticity with quadratic elements which we believe is due to round-off errors encountered.)

For regular meshes of quadratic triangles generally superconvergence is not expected and it does not occur for neither heat conduction nor elasticity. However the robustness index has very small values ($R < 0.09$ for SPR $R < 0.1$ for REP) and the estimators are therefore very good.

In Figure 4 very irregular patterns of triangular and quadrilateral elements are analysed in repeatable figures. It is of course not possible to present here all tests conducted by the effectivity patch test. The results shown are however typical - others are given in [27]. It is interesting to observe that the performance of quadrilateral elements is always superior to that of triangles.

In the recent paper of Babuška *et al* [14] the authors show that the alternative versions of SPR such as [7]- [9], such as those in [8], give a generally much worse performance than the original version especially on irregular elements assembled near boundaries.

6. ADAPTIVE REFINEMENT

Most error estimations can yield reasonable evaluation of the global error and of the error contribution of individual elements. However, once again, the local estimations are substantially improved by recovery procedures and the knowledge of such local error is particularly important during adaptive refinement (and derefinement) if the desired result is to be reached in a small number of solutions.

Table 2. Effectivity bounds and robustness of SPR and REP recovery estimator for regular patterns of triangles or rectangles with linear and quadratic shape function (applied to heat conduction and elasticity problems).

LINEAR TRIANGLES AND RECTANGLES (BOTH HEAT CONDUCTION / ELASTICITY)						
Aspect Ratio	SPR			REP		
	θ_L	θ_U	R	θ_L	θ_U	R
1/1	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
1/2	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
1/4	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
1/8	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
1/16	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
1/32	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
1/64	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
QUADRATIC RECTANGLES (HEAT CONDUCTION)						
	θ_L	θ_U	R	θ_L	θ_U	R
1/1	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
1/2	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
1/4	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
1/8	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
1/16	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
1/32	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
1/64	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
QUADRATIC RECTANGLES (ELASTICITY)						
	θ_L	θ_U	R	θ_L	θ_U	R
1/1	1.0000	1.0000	0.0000	0.9991	1.0102	0.0111
1/2	1.0000	1.0000	0.0000	0.9991	1.0181	0.0189
1/4	1.0000	1.0000	0.0000	0.9991	1.0136	0.0145
1/8	1.0000	1.0000	0.0000	0.9991	1.0030	0.0039
1/16	1.0000	1.0000	0.0000	0.9968	1.0001	0.0033
1/32	1.0000	1.0000	0.0000	0.9950	1.0000	0.0050
1/64	1.0000	1.0000	0.0000	0.9945	1.0000	0.0055
QUADRATIC TRIANGLES (ELASTICITY)						
	θ_L	θ_U	R	θ_L	θ_U	R
1/1	0.9966	1.0929	0.0963	0.9562	1.0503	0.0940
1/2	0.9966	1.0931	0.0965	0.9559	1.0481	0.0923
1/4	0.9967	1.0937	0.0970	0.9535	1.0455	0.0924
1/8	0.9967	1.0943	0.0976	0.9522	1.0603	0.1081
1/16	0.9966	1.0946	0.0980	0.9518	1.0666	0.1148
1/32	0.9966	1.0947	0.0981	0.9517	1.0684	0.1167
1/64	0.9965	1.0947	0.0982	0.9516	1.0688	0.1172

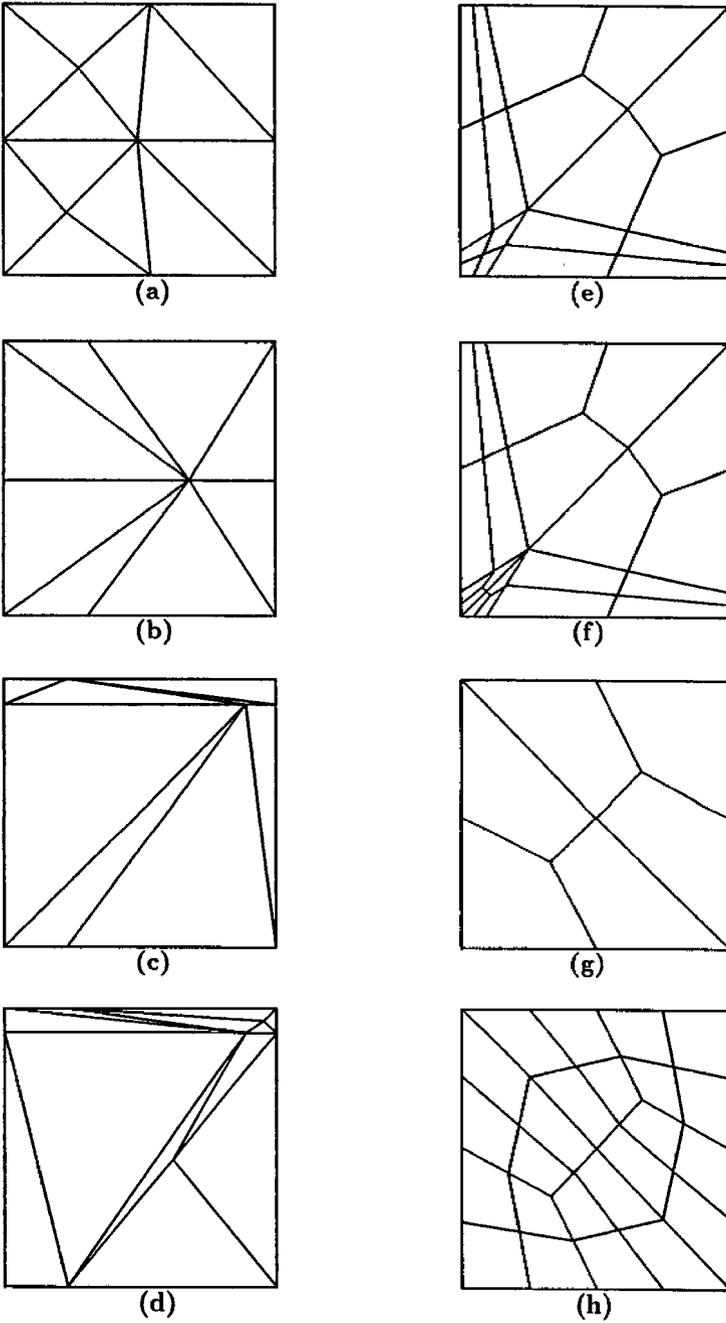


Figure 4: Irregular patterns of triangular and quadrilateral elements, repeatable subregions.

Table 3. Effectivity bounds and robustness indices for various irregular meshes of triangles (a, b, c, d) and quadrilaterals (e, f, g, h).

LINEAR ELEMENT (HEAT CONDUCTION)						
Aspect Ratio	SPR			REP		
	θ_L	θ_U	R	θ_L	θ_U	R
a	0.9626	1.0054	0.0442	0.9709	1.0145	0.0443
b	0.9715	1.0156	0.0447	0.9838	1.0167	0.0329
c	0.9228	1.4417	0.5189	0.8938	1.8235	0.9297
d	0.8341	1.2027	0.3685	0.9463	1.9272	0.9810
e	0.9943	1.0175	0.0232	0.9800	1.0589	0.0789
f	0.9969	1.0152	0.0183	0.9849	1.0582	0.0733
g	0.9987	1.0175	0.0188	0.9987	1.0175	0.0188
h	0.9991	1.0068	0.0077	0.9979	1.0062	0.0083
LINEAR ELEMENTS (ELASTICITY)						
	θ_L	θ_U	R	θ_L	θ_U	R
a	0.9404	1.0109	0.0741	0.9468	1.0148	0.0707
b	0.8869	1.0250	0.1520	0.9392	1.0275	0.0915
c	0.8550	1.6966	0.8415	0.8037	2.0522	1.2486
d	0.7945	1.2734	0.4788	0.7576	1.9416	1.1840
e	0.9946	1.0247	0.0301	0.9579	1.0508	0.0928
f	1.0038	1.0281	0.0318	0.9612	1.0467	0.0855
g	0.9959	1.0300	0.0341	0.9960	1.0298	0.0338
h	0.9972	1.0139	0.0168	0.9965	1.0122	0.0157
QUADRATIC ELEMENTS (HEAT CONDUCTION)						
	θ_L	θ_U	R	θ_L	θ_U	R
a	0.9443	1.0295	0.0877	0.9339	1.0098	0.0805
b	0.8146	1.0037	0.2313	0.9256	1.0028	0.0832
c	0.7640	1.0486	0.3553	0.9559	1.2229	0.2670
d	0.8140	1.0141	0.2423	0.9091	1.2808	0.3717
e	0.9762	1.0053	0.0296	0.9901	1.0177	0.0276
f	0.9691	1.0045	0.0363	0.9901	1.0322	0.0421
g	0.9692	1.0004	0.0322	0.9833	1.0024	0.0195
h	0.9906	1.0113	0.0207	1.0045	1.0261	0.0307
QUADRATIC ELEMENTS (ELASTICITY)						
	θ_L	θ_U	R	θ_L	θ_U	R
a	0.9144	1.0353	0.1277	0.9197	1.0244	0.1111
b	0.7302	1.0355	0.4038	0.8643	1.0346	0.1905
c	0.7556	1.1024	0.4163	0.8387	1.2422	0.4035
d	0.7624	1.0323	0.3430	0.8244	1.2632	0.4388
e	0.9702	1.0102	0.0408	0.9682	1.0058	0.0386
f	0.9651	1.0085	0.0446	0.9749	1.0286	0.0537
g	0.9457	1.0115	0.0688	0.9807	1.0125	0.0321
h	0.9852	1.0141	0.0290	0.9996	1.0522	0.0526

The degree and indeed the nature of the refinement will depend much on the criteria we wish to satisfy. The most commonly used procedure aims at achieving a specific percentage error η in a global norm such as the well known energy norm and here we shall seek to achieve this in a manner we outlined originally in 1987 [3]. Many of the examples studied by the authors and others aim at satisfying;

$$\eta = \frac{\|e\|}{\|u\|} \leq \bar{\eta} \quad (40)$$

where $\bar{\eta}$ is the specified permissible error.

The physical meaning of such a specified error level is not very clear, though approximately it aims at the *RMS* value of stress error being some specified fraction of the *RMS* value of stresses.

A more logical aim perhaps would be to seek a specific value of error density:

$$\sigma \equiv \left(\frac{\|e\|_{\Omega_e}^2}{\Omega_e} \right)^{\frac{1}{2}} \quad (41)$$

where Ω_e is the element area.

This gives in effect a specified value of the *RMS stress error* at all elements. This type of criterion has been discussed extensively by Oñate and Castro [30] and generally requires a high degree of refinement.

With specified total energy norm error we can at any stage evaluate both the element error:

$$\|e\|_{\Omega_e} \quad (42)$$

and the permissible total error:

$$\bar{\eta} \|u_h\| \quad (43)$$

As in general we shall aim at equal distribution of error between all elements and if n is the number of elements in a current analysis the permissible error in each element is:

$$\bar{\eta} \|u_h\| / n^{\frac{1}{2}} \quad (44)$$

Now we can find the approximate size of the new element vis a vis the old element by assuming $O(h^p)$ convergence and thus:

$$\frac{h_{new}}{h_{old}} = \left(\frac{\bar{\eta} \|u_h\| / n^{\frac{1}{2}}}{\|e\|_{\Omega_e}} \right)^{\frac{1}{p}} \quad (45)$$

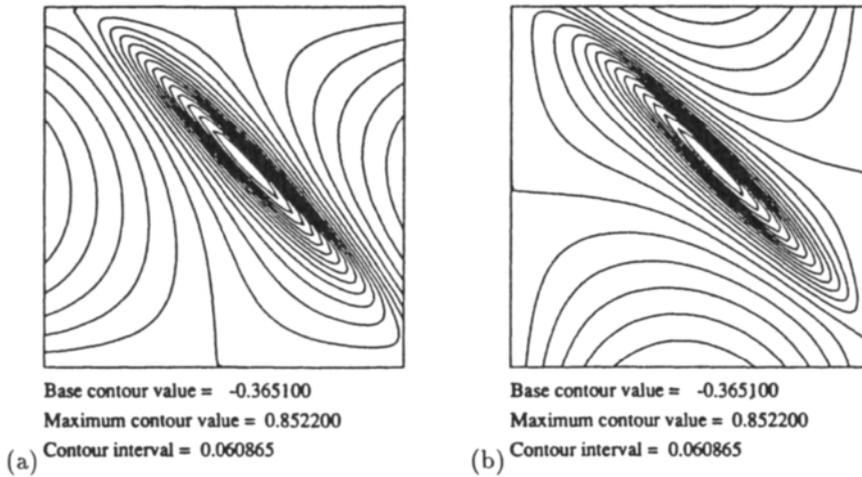


Figure 5: A Poisson equation in a square domain (viz Zienkiewicz, Zhu 1992), contours of $\frac{\partial u}{\partial x}$ (a) and $\frac{\partial u}{\partial y}$ (b)

Although the procedure is very crude (as the subdivisions of course changes the element number) in convergence, it is generally rapid and only one or two resolutions are generally needed to achieve final results.

Two main procedures exist to obtain the refined mesh; these are either complete remeshing or subdivision. The former is more useful generally and in Figure 5 we show a linear example from reference [5]. Figure 6 show the convergence plot for various types of elements used in solving a Poisson type problem in a square domain. The details of the problem and of the refinement are given in reference [5] and we only show the progressive steps of adaptive refinements. In Figure 7 the convergence of the error is shown versus the degree of freedoms. The reader will observe the better performance of quadrilaterals in both linear and quadratic versions.

7. CONCLUDING AND REMARKS

In this paper we have summarized the achievements of an error estimation procedure based on recovery and have demonstrated on linear examples its effectiveness in adaptive refinement. We believe it is generally more efficient method than that based on residual computation.

In Part II of this paper, we shall describe the application of the same procedure to non linear, plasticity, problems.

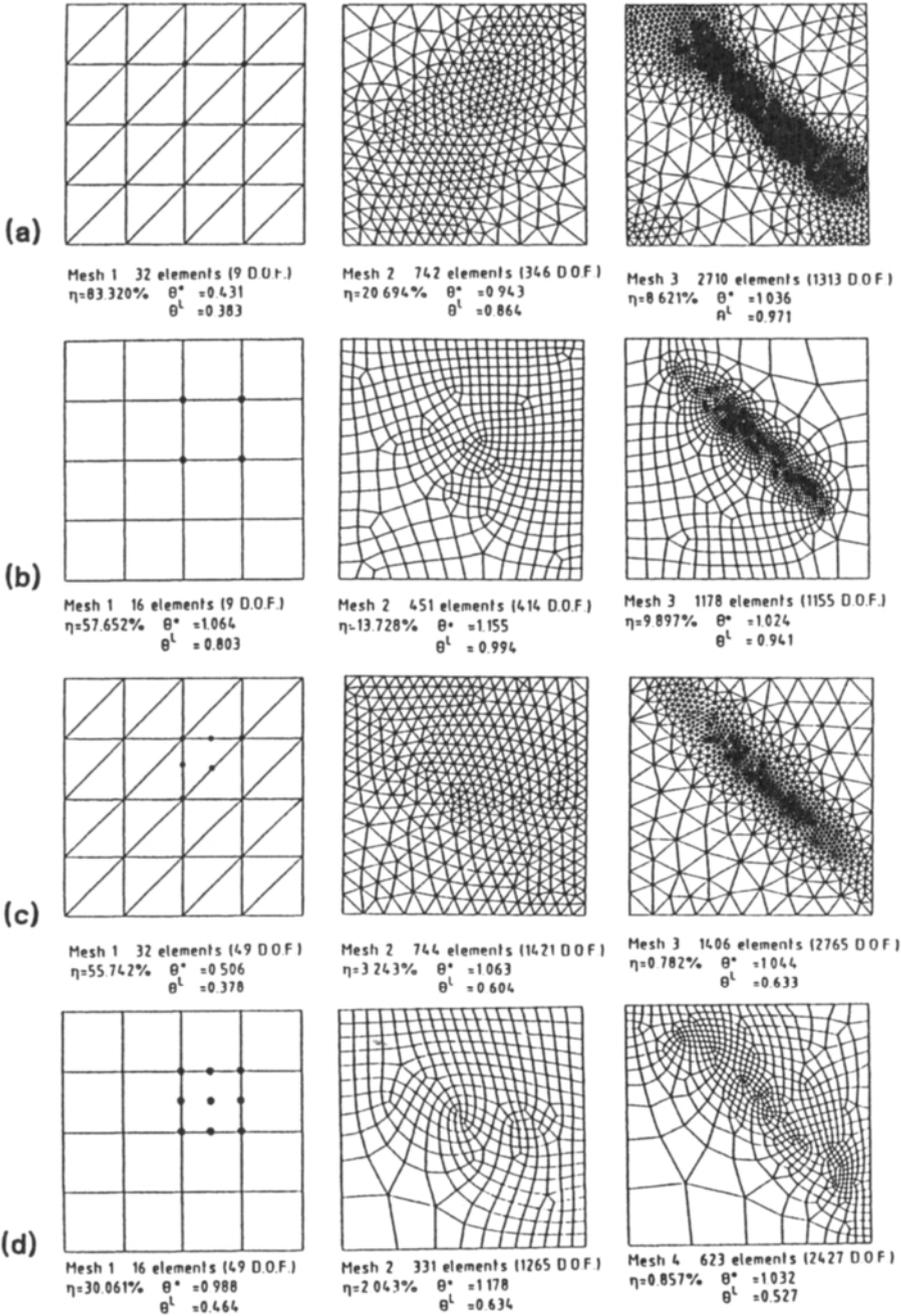


Figure 6: Adaptive mesh refinement for example of Figure 5, triangular and quadrilateral elements (a)(b) linear (c)(d) quadratic. Aim 10% error and 1% error respectively.

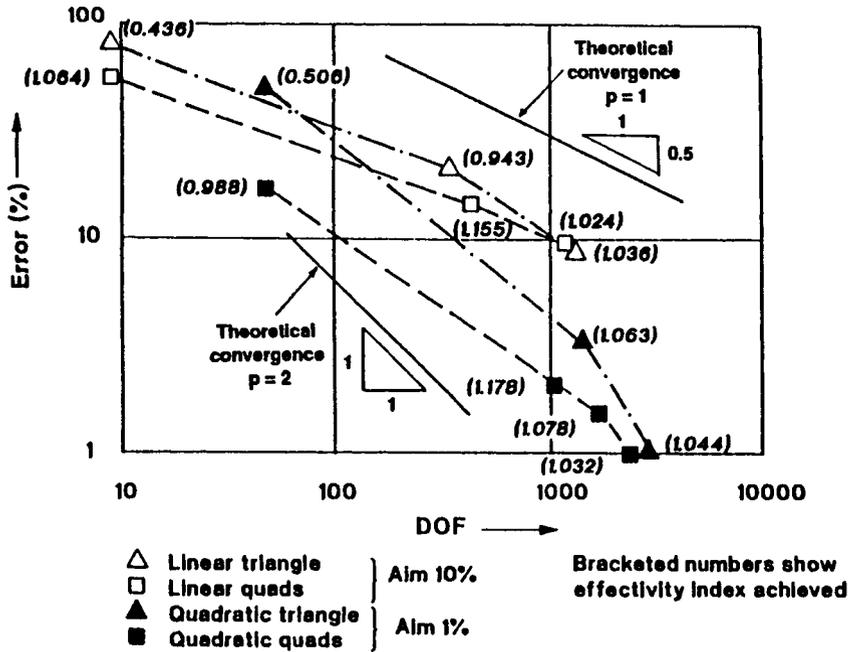


Figure 7: Convergence of adaptive refinement in example of Figure 5

REFERENCES

- [1] I. Babuška, W.C. Rheinboldt, A posteriori error estimates for the finite element method, *Int. J. Num. Meth. Eng.*, 12, (1978), 1597-1615.
- [2] M. Ainsworth, J.T. Oden, A unified approach to a-posteriori error estimation using element residual methods, *Numer. math.*, 65, (1993), 23-50.
- [3] O.C. Zienkiewicz, J.Z. Zhu, A simple error estimator and adaptive procedure for practical engineering analysis, *Int. J. Num. Meth. Eng.*, 24, (1987), 337-357.
- [4] O.C. Zienkiewicz, J.Z. Zhu, The superconvergent patch recovery and *a posteriori* error estimates. Part 1: The recovery technique, *Int. J. Num. Meth. Eng.*, 33 (1992), 1331-1364.
- [5] O.C. Zienkiewicz, J.Z. Zhu, The superconvergent patch recovery and *a posteriori* error estimates. Part 2: Error estimates and adaptivity, *Int. J. Num. Meth. Eng.*, 33, (1992), 1365-1382.
- [6] O.C. Zienkiewicz, J.Z. Zhu, Superconvergent patch recovery (SPR) and adaptive finite element refinement, *Com. Meth. Appl. Mech. Eng.*, 101, (1992), 207-224.

- [7] N-E. Wiberg, F. Abdulwahab, Patch recovery based on superconvergent derivatives and equilibrium, *Int. J. Num. Meth. Eng.*, 36, (1993), 2703-2724.
- [8] N-E. Wiberg, F. Abdulwahab, S. Ziukas, Enhanced superconvergent patch recovery incorporating equilibrium and boundary conditions, *Int. J. Num. Meth. Eng.*, 37, (1994), 3417-3440.
- [9] T. Blacker, T. Belytschko, Superconvergent patch recovery with equilibrium and conjoint interpolation enhancements, *Int. J. Num. Meth. Eng.*, 37, (1994), 517-536.
- [10] O.C. Zienkiewicz, J.Z. Zhu, Superconvergence and the superconvergent patch recovery, *Finite elem. Anal. Design.*, 19, (1995), 11-23.
- [11] I. Babuška, T. Strouboulis and C.S. Upadhyay, A model study of the quality of a posteriori error estimators for linear elliptic problems. Error estimation in the interior of patchwise uniform grids of triangular, *Com. Meth. Appl. Mech. Eng.* 114, (1994), 307-378.
- [12] I. Babuška, T. Strouboulis, C.S. Upadhyay, S.K. Gangaraj and K. Copps, Validation of a posteriori error estimators by numerical approach, *Int. J. Num. Meth. Eng.*, 37, (1994), 1073-1123.
- [13] I. Babuška, T. Strouboulis, C.S. Upadhyay, S.K. Gangaraj and K. Copps, An Objective criterion for assessing the reliability of A-Posteriori error estimators in finite element computations, *U.S.A.C.M Bulletin*, No. 7 (1994), 4-16.
- [14] I. Babuška, T. Strouboulis and C.S. Upadhyay, A model study of the quality of a posteriori error estimators for finite element solutions of linear elliptic problems, with particular reference to the behavior near the boundary, *Int. J. Num. Meth. Eng.*, 40, (1997), 2521-2577.
- [15] E. Rank, O.C. Zienkiewicz, A simple error estimator in finite element method, *Com. Meth. Appl. Mech. Eng.*, 3, (1987), 243-250.
- [16] M. Ainsworth, J.T. Oden, A unified approach to a-posteriori error estimation using element residual methods, *Numer. math.*, 65, (1993), 23-50.
- [17] J.Z. Zhu, A posteriori error estimation- The relationship between different procedures, *Com. Meth. Appl. Mech. Eng.* To be published
- [18] O.C. Zienkiewicz, R.L. Taylor, *The Finite Element Method*, IV Edition, Vol. I (1989), Vol. II (1991)
- [19] C.M. Chen, Superconvergence of finite element solutions and their derivatives, *Numer. math. J. Chinese Univ.*, 3, (1981), 118-125.
- [20] M. Zlamal, Superconvergence and reduced integration in the finite element method, *Math. Comput.*, 32, (1980), 663-685.

- [21] R.J. Mackinnon, G.F. Carey, Superconvergent derivatives: A Taylor series analysis, *Int. J. Num. Meth. Eng.*, 28, (1989), 489-509.
- [22] A. Samuelsson, N.E. Wiberg, Finite element adaptivity in dynamics and elasto-plasticity, In *The finite element method in the 1990's*, O.C Zienkiewicz's anniversary volume, (edited by E. Onate et al.), 152-162 CIMNE and Springer-Verlag, Barcelona (1991)
- [23] P. Tong, Exact solution of certain problem by the finite element method, *J. A.I.A.A.*, 1, (1969), 179-180.
- [24] J.T. Oden, H.J. Brauchli, On the calculation of constant stress distribution in finite element applications, *Int. J. Num. Meth. Eng.*, 3, (1971), 317-325.
- [25] M. Tabbara, T. Blacker, T. Belytschko, Finite element derivative recovery by moving least square interpolants, *Com. Meth. Appl. Mech. Eng.*, 117, (1994), 211-223.
- [26] B. Boroomand, O.C. Zienkiewicz, Recovery by equilibrium in patches (REP), *Int. J. Num. Meth. Eng.*, 40, (1997), 137-154.
- [27] B. Boroomand, O.C. Zienkiewicz, An improved REP recovery and the effectivity robustness test, *Int. J. Num. Meth. Eng.*, 40, (1997), 3247-3277.
- [28] P. Ladevèze, D. Leguillon, Error estimate procedure in the finite element method and applications, *SIAM J. Num. Anal.*, 20, No. 3, (1983), 485-509.
- [29] P. Ladevèze, G. Coffignal, J.P. Pelle, Accuracy of elastoplastic and dynamic analysis, *Accuracy Estimates and Adaptive Refinements in Finite Element Computations I*. Babuška, O.C. Zienkiewicz, J. Gago and E.R. de A. Oliveria. Chapter 11, (1986)
- [30] E. Oñate, J. Castro, Adaptive mesh refinement techniques for structural problems, *The Finite Element Method in the 1990's*, Editors E. Oñate, J. Perieux and A. Samuelsson, Springer, CIMNE, Barcelona (1991)

This Page Intentionally Left Blank

The relationship of some a posteriori error estimators

J. Z. Zhu^a and Zhimin Zhang^b

^a UES, Inc. 175 Admiral Cochrane Drive, Suite 110, Annapolis, MD 21401, USA

^b Department of Mathematics, Texas Tech University, Lubbock, TX 79409, USA

ABSTRACT

We present analytical and numerical investigation in the relationship of the recovery error estimator and the implicit residual error estimator. It is shown that analytically both error estimators are equivalent for one dimensional problems. Numerical study indicates that such equivalence also exist for two dimensional problems.

1. INTRODUCTION

The majority of the error estimators used in practical applications are either the residual type error estimators[1-13] which is computed by using the residual of the finite element solution explicitly (explicit residual error estimator) or implicitly (implicit residual error estimator), or the recovery type error estimator[14-20] which is computed by locally constructing an improved solution from the finite element approximation. The residual type of error estimator and the recovery type error estimator have always been derived by different methodologies in their original forms. The formation of the error estimators and their computational implementation are very much different. Researchers working on these error estimators have always been following different paths in the construction and the analysis of these error estimators, although some efforts were made to find the relationship between them[21-22].

Recently we have demonstrated that the recovery error estimator and the residual error estimator are related. We have shown that for both one dimensional and two dimensional model problems, the explicit residual type error estimators can be derived from recovery type error estimators by using specific recovery techniques [23]. In particular, when a recovery procedure involving local residual is used, the first residual type error estimator for one dimensional problems introduced by Babuska and Rheinboldt [1] was derived from the recovery error estimator. The popular explicit residual error estimator of Babuska and Rheinboldt for two dimensional problem [2] can also be derived from the recovery error estimator using average as recovery technique,. It was also shown that some

recovery techniques used in the computation of the recovery error estimator also forms the foundation of the implicit residual type error estimator.

In this work we further investigate the relationship of the recovery error estimator and the residual error estimator, particularly the relationship of the recovery error estimator and the implicit residual error estimator. We will show analytically that for one dimensional problem the recovery error estimator involving local residual is equivalent to the implicit residual error estimator. We will also provide numerical evidence to demonstrate that such equivalence also exist for two dimensional problems.

2. A RECOVERY TYPE ERROR ESTIMATOR

We shall first introduce a recovery type error estimator. Considering the following two-point boundary value problem

$$-\frac{d}{dx}\left(a(x)\frac{du}{dx}\right) + b(x)u = f \quad x \in (0,1) \quad (2.1a)$$

with boundary conditions

$$u(0) = 0, \quad u(1) = 0 \quad (2.1b)$$

We assume that a , b and f are sufficiently smooth and that

$$\begin{aligned} 0 < a_0 \leq a(x) \leq a_1 < \infty \\ 0 \leq b(x) \leq b_1 < \infty \end{aligned} \quad x \in [0,1] \quad (2.1c)$$

A subdivision of domain I

$$0 = x_0 < x_1 < x_2 < \dots < x_N = 1 \quad (2.2)$$

divides I into N elements $I_i = (x_{i-1}, x_i)$ with element size

$$h_i = x_i - x_{i-1} \quad (2.3)$$

The finite element solution

$$u_h = \sum_{m=1}^M \bar{u}_m N_m \quad (2.4)$$

$$M = pN + 1 \quad (2.5)$$

can be obtained by standard Galerkin procedure [24], where p is the order of the element.

The discretization error of the finite element approximation is defined as

$$e = u - u_h \quad (2.6)$$

The error in the energy norm is written as

$$\|e\| = \left[\int_I \left(a(x)(u' - u_h')^2 + b(x)(u - u_h)^2 \right) dx \right]^{1/2} \quad (2.7)$$

where (\cdot) represents the derivative.

The recovery type error estimator for the two-point boundary problem is defined as,

$$\|\bar{e}\| = \left(\int_I a(x) \left(u^{*\prime} - u_h' \right)^2 dx \right)^{1/2} \quad (2.8)$$

where u^* is the recovered solution. We note that the higher order term in (2.7) is omitted in the error estimator.

We shall now construct a recovery procedure to compute $u^{*\prime}$. Considering element $I_i = [x_{i-1}, x_i]$

$$\text{Let } u_{(i)}^{*\prime} = u_{h(i)}' + \alpha_i Z_p \quad (2.9)$$

where

$$Z_p(x) = Z_p\left(x_i - \frac{1-\xi}{2}h_i\right) = L_p(\xi) \quad \xi \in [-1,1] \quad (2.10)$$

and $L_p(\xi)$ is the Legendre function of order p (same order as N_m).

α_i is determined by minimizing the following functional

$$F(\alpha_i) = \int_{I_i} (r^*)^2 dx = \int_{I_i} \left(-\frac{d}{dx} \left(a(x)u_{(i)}^{*\prime} \right) + b(x)u_{h(i)} - f \right)^2 dx \quad (2.11)$$

The minimization condition of $F(\alpha_i)$ gives

$$\alpha_i = \left\{ \int_{I_i} \left(\frac{d}{dx} (a(x)Z_p) \right)^2 dx \right\}^{-1} \int_{I_i} \frac{d}{dx} (a(x)Z_p) r dx \quad (2.12)$$

where

$$r = -\frac{d}{dx} \left(a(x)u'_{h(i)} \right) + b(x)u_{h(i)} - f \quad x \in (x_{i-1}, x_i) \quad (2.13)$$

is the residual of u_h in the element I_i .

Once α_i is known, u^* can be readily computed from (2.9).

The global error estimator is obtained by substituting u^* into $\|\bar{e}\|$,

$$\|\bar{e}\| = \left(\sum_{i=1}^N \|\bar{e}\|_i^2 \right)^{\frac{1}{2}} \quad (2.14)$$

The local, element error estimator has the form of

$$\begin{aligned} \|\bar{e}\|_{(i)}^2 &= \int_{I_i} a(x) \left(u_{(i)}^* - u'_{h(i)} \right)^2 dx \\ &= \int_{I_i} a(x) \left(\alpha_i Z_p \right)^2 dx \\ &= \alpha_i^2 \int_{I_i} a(x) Z_p^2 dx \end{aligned} \quad (2.15)$$

Remark 2.1. Starting from the recovery type error estimator (2.15), the explicit residual type error estimator of Babuska and Rheinboldt can be derived [15,23].

3. THE RELATIONSHIP OF THE ERROR ESTIMATORS

We shall, in the following, establish the asymptotic equivalence between the recovery type error estimator $\|\bar{e}\|_{(i)R}$ presented in Section 2 and the implicit residual error estimator discussed in [13] in the one-dimensional setting. We refer to reference [13] for additional details about the implicit residual type error estimator.

By asymptotic equivalence we mean that there exist positive constants C_1 and C_2 , independent of h , such that

$$C_1(1+O(h))\|\bar{e}\|_{(i)R} \leq \|\bar{e}\|_{(i)} \leq C_2\|\bar{e}\|_{(i)R}(1+O(h)). \quad (3.1)$$

We first consider a simple case of problem (2.1) with $a(x) = 1$, $b(x) = 0$, and the exact solution is a polynomial of degree $p + 1$ (one order higher than the finite element space). We write the Legendre expansion of the exact solution on I_i as:

$$u(x) = u(x_{i-1})N_{i-1}(x) + u(x_i)N_i(x) + \sum_{k=2}^{p+1} c_{ik} \phi_k(x), \quad x \in I_i, \quad (3.2)$$

where

$$N_i(x) = \begin{cases} 1 + (x - x_i)/h_i, & x \in I_i, \\ 1 + (x_i - x)/h_{i+1}, & x \in I_{i+1}, \\ 0, & \text{otherwise,} \end{cases}$$

$$\phi_k(x) = \phi_k(x_i - \frac{1-\xi}{2}h_i) = \phi_k(\xi), \quad \xi \in (-1,1),$$

with

$$\phi_k(\xi) = \sqrt{\frac{2k-1}{2}} \int_{-1}^{\xi} L_{k-1}(t) dt,$$

and L_{k-1} , the Legendre polynomial of degree $k-1$. By orthogonality, we see that $c_{ik} = (f, \phi_{ik}) / (\phi'_{ik}, \phi'_{ik})$. It has been shown in [25] that in this special case, the finite element solution differs from the exact solution only by the last term, i.e.,

$$u_h(x) = u(x_{i-1})N_{i-1}(x) + u(x_i)N_i(x) + \sum_{k=2}^p c_{ik} \phi_{ik}(x), \quad x \in I_i. \quad (3.3)$$

We now examine the recovery element error estimator introduced in Section 2. Under the assumption $\alpha(x) = 1$ and $b(x) = 0$ when the exact solution is given by (3.2), we have

$$\|\bar{e}\|_{(i)}^2 = \alpha_i^2 \int_{I_i} Z_p^2 dx, \quad \alpha_i = (\int_{I_i} (Z'_p)^2 dx)^{-1} \int_{I_i} Z'_p r dx.$$

We see that

$$-f = u'' = u''_h + c_{i,p+1} \phi''_{i,p+1} = u''_h + c_{i,p+1} \sqrt{\frac{2p+1}{2}} Z'_p,$$

therefore,

$$r = -u''_h - f = c_{i,p+1} \sqrt{\frac{2p+1}{2}} Z'_p, \quad \alpha_i = c_{i,p+1} \sqrt{\frac{2p+1}{2}}, \quad (3.4)$$

and

$$\|\bar{e}\|_{(i)}^2 = \alpha_i^2 \int_{I_i} Z_p^2 dx = c_{i,p+1}^2 \int_{I_i} (\phi'_{i,p+1})^2 dx = \int_{I_i} (u' - u'_h)^2 dx. \quad (3.5)$$

It demonstrates that the recovery error estimator gives the exact error in this simple case.

Next, we consider the residual type error estimator discussed in [13] with the flux splitting of [7]. By splitting the flux at each node, the following local problem is solved.

$$-\phi'' = r \text{ in } I_i, \quad \phi'(x_{i-1}) = \int_{I_i} N_{i-1} r dx, \quad \phi'(x_i) = -\int_{I_i} N_i r dx. \quad (3.6)$$

Note that the consistent condition

$$\int_{I_i} N_{i-1} r dx + \int_{I_i} N_i r dx = \int_{I_i} r dx$$

is satisfied. Therefore, the local problem is well post which has a unique solution up to a constant and has a unique solution for ϕ' . The residual element error estimator is defined by

$$\|\bar{e}\|_{(I)R}^2 = \int_{I_i} (\phi')^2 dx. \quad (3.7)$$

When $r = c_{i,p+1} \phi_{i,p+1}''$ as we demonstrated earlier, it is easy to verify that $\phi' = -c_{i,p+1} \phi_{i,p+1}'$. Hence, the residual error estimator also yields the exact error in this case, i.e.,

$$\|\bar{e}\|_{(I)R}^2 = \int_{I_i} (-c_{i,p+1} \phi_{i,p+1}')^2 dx = \int_{I_i} (u' - u_h')^2 dx. \quad (3.8)$$

We summarize the result in the following theorem.

Theorem 3.1. Let the exact solution be a polynomial of one degree higher than the finite element space, and let $\alpha(x) = 1$, $b(x) = 0$. Then both the recovery and the residual (with the splitting) error estimators are exact.

We now consider the general case. Again, we use the Legendre expansion of u :

$$u(x) = u(x_{i-1})N_{i-1}(x) + u(x_i)N_i(x) + \sum_{k=2}^{p+1} c_{ik} \phi_{ik}(x) + q(x), \quad x \in I_i, \quad (3.9)$$

with

$$\|q'\|_0 \leq Ch^{p+1} |u|_{p+2}. \quad (3.10)$$

Where $\|\bullet\|_0$ is the L_2 norm and $|\bullet|_{p+2}$ is the order $p+2$ seminorm. We denote u_I , the Legendre projection of u by

$$(u' - u_I', v') = 0, \quad \forall v \in V_p^h, \quad (3.11)$$

here V_p^h is the finite element space which consists continuous piecewise polynomials of degree p . From [25],

$$u_I(x) = u(x_{i-1})N_{i-1}(x) + u(x_i)N_i(x) + \sum_{k=2}^p c_{ik} \phi_k(x), \quad x \in I_i, \quad (3.12)$$

with $c_{ik} = (u', \phi_k')$. It is known that the difference between the finite element solution u_h and u_I is superconvergent in the following sense [26].

$$\|u'_h - u'_I\|_{L^\infty(I)} \leq Ch^{p+1} \|u\|_{W_2^{p+1}(I)}. \quad (3.13)$$

Where $\|\bullet\|_{L^\infty}$ is the maximum norm and $\|\bullet\|_{W_2^{p+1}}$ is the standard maximum norm up to $p+1$ order of derivative. We now examine the residual term.

$$\begin{aligned} r &= -(a(x)u'_h)' + b(x)u_h - f \\ &= -(a(x)u'_I)' + b(x)u_I - f + [a(x)(u_I - u_h)']' - b(x)(u_I - u_h) \\ &= [a(x)(u - u_I)']' - b(x)(u - u_I) + [a(x)(u_I - u_h)']' - b(x)(u_I - u_h) \\ &= r_0 + r_1, \end{aligned} \quad (3.14)$$

where

$$r_0 = c_{i,p+1} (a(x)\phi'_{i,p+1})' = c_{i,p+1} \sqrt{\frac{2p+1}{2}} (a(x)Z_p)',$$

and

$$r_1 = (a(x)q')' - b(x)(u - u_I) + [a(x)(u_I - u_h)']' - b(x)(u_I - u_h).$$

We see that r_1 is a higher order term compared with r_0 and more precisely, we have

$$r = r_0(1 + O(h)). \quad (3.15)$$

Therefore, the major part of the coefficient α_i is

$$\frac{\int_{I_i} (a(x)Z_p)' r_0 dx}{\int_{I_i} [(a(x)Z_p)']^2 dx} = c_{i,p+1} \sqrt{\frac{2p+1}{2}}$$

More detailed analysis shows that

$$\alpha_i = c_{i,p+1} \sqrt{\frac{2p+1}{2}} (1 + O(h)). \quad (3.16)$$

Hence,

$$\begin{aligned} \|\bar{e}\|_{(i)}^2 &= \alpha_i^2 \int_{I_i} a(x) Z_p^2 dx \\ &= c_{i,p+1}^2 \int_{I_i} a(x) (\phi'_{i,p+1})^2 dx (1+O(h))^2 = \int_{I_i} a(x) c_{i,p+1}^2 (\phi'_{i,p+1})^2 dx (1+O(h))^2 \end{aligned} \quad (3.17)$$

We see that the recovery type error estimator catches the leading error term $c_{i,p+1} \phi'_{i,p+1}$.

In the case of residual error estimator, the local problem (3.6) can be decomposed to two problems,

$$-\phi_0'' = r_0, \quad \text{in } I_i, \quad \phi_0'(x_{i-1}) = \int_{I_i} N_{i-1} r_0 dx, \quad \phi_0'(x_i) = -\int_{I_i} N_i r_0 dx; \quad (3.18)$$

and

$$-\phi_1'' = r_1, \quad \text{in } I_i, \quad \phi_1'(x_{i-1}) = \int_{I_i} N_{i-1} r_1 dx, \quad \phi_1'(x_i) = -\int_{I_i} N_i r_1 dx; \quad (3.19)$$

Based on (3.15), we have

$$\phi = \phi_0 (1+O(h)). \quad (3.20)$$

It is easy to verify that

$$\begin{aligned} \phi_0'(x) &= -c_{i,p+1} \sqrt{\frac{2p+1}{2}} \left(\int_{x_{i-1}}^x (a(x) Z_p)' - \int_{I_i} N_i (a(x) Z_p)' dx \right) \\ &= -c_{i,p+1} \sqrt{\frac{2p+1}{2}} (a(x) Z_p(x) + \int_{I_i} N_i' a(x) Z_p dx) \end{aligned} \quad (3.21)$$

is the solution of (3.18). The first term in (3.21) is the leading error for $u' - u'_i$. The second term is of higher order which can be seen from

$$\int_{I_i} N_{i-1}' a(x) Z_p dx = \int_{I_i} N_{i-1}' (a(x) - \bar{a}) Z_p dx, \quad \bar{a} = \frac{1}{h_i} \int_{I_i} a(x) dx.$$

Note that Z_p is orthogonal to N_{i-1}' . Therefore,

$$\phi_0' = -c_{i,p+1} \sqrt{\frac{2p+1}{2}} a(x) Z_p (1+O(h)). \quad (3.22)$$

Taking into account of (3.20), we have

$$\phi' = -c_{i,p+1} \sqrt{\frac{2p+1}{2}} a(x) Z_p (1 + O(h)) = a(x) c_{i,p+1} \phi'_{i,p+1} (1 + O(h)). \quad (3.23)$$

Therefore, the residual element error estimator is

$$\|\bar{e}\|_{(i),R}^2 = \int_{I_i} \frac{1}{a(x)} (\phi')^2 dx = \int_{I_i} a(x) (c_{i,p+1} \phi'_{i,p+1})^2 dx (1 + O(h))^2 \quad (3.24)$$

We now have the following theorem for the general case.

Theorem 3.2. For the two-point boundary problem (2.1), when the exact solution has a Legendre expansion (3.9), the recovery error estimator and the residual error estimator are asymptotically equivalent in the sense that

$$\|\bar{e}\|_{(i)} = \left(\int_{I_i} a(c_{i,p+1} \phi'_{i,p+1})^2 dx \right)^{1/2} (1 + O(h)), \quad (3.25)$$

$$\|\bar{e}\|_{(i),R} = \left(\int_{I_i} a(c_{i,p+1} \phi'_{i,p+1})^2 dx \right)^{1/2} (1 + O(h)). \quad (3.26)$$

4. RECOVERY TYPE ERROR ESTIMATOR AND IMPLICIT RESIDUAL TYPE ERROR ESTIMATOR

In this section, the relationship between the recovery error estimator and the implicit residual type error estimator, in particular the implicit residual error estimator proposed by Ainsworth and Oden [13], is investigated for two dimensional problems.

Let $\Omega \in R^2$ be a bounded polygonal domain and let $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$ be its boundary. We shall consider the following problem

$$-\nabla \cdot (a(x, y) \nabla u) + c(x, y)u = f \quad x, y \in \Omega \quad (4.1a)$$

with boundary condition

$$u = 0 \quad \text{on } \partial\Omega_D$$

$$a \frac{\partial u}{\partial n} = g \quad \text{on } \partial\Omega_N \quad (4.1b)$$

We shall assume that a, c, f and g are sufficiently smooth for our analysis and

$$\begin{aligned} 0 < a_1 \leq a(x, y) \leq a_2 \\ 0 < c_1 \leq c(x, y) \leq c_2 \end{aligned} \quad x, y \in \bar{\Omega} \quad (4.1c)$$

Let u_h denotes the finite element approximation of u . It was shown [13] that the discretization error e can be bounded from above as

$$\|e\|^2 \leq \sum_{i=1}^M \eta_i^2(q_i) \quad (4.2)$$

where M is the number of the elements and for each element i

$$\eta_i^2(q_i) = \varepsilon_i^2(q_i) + \Lambda_i^2(q_i) \quad (4.3)$$

with

$$\varepsilon_i^2(q_i) = \int_{\Omega_i} \frac{1}{a} q_i \cdot q_i d\Omega \quad (4.4a)$$

$$\Lambda_i^2(q_i) = \int_{\Omega_i} \frac{1}{c} (\nabla \cdot q_i + r_i)^2 d\Omega \quad (4.4b)$$

Here

$$r_i = f + \nabla \cdot (a \nabla u_h) - c u_h \quad (4.5)$$

is the element residual.

q_i is chosen such that $\eta_i^2(q_i)$ is minimized.

Assuming $q_i = \nabla \varphi_i$, rather than pursuing the minimization of $\eta_i^2(q_i)$ directly to compute q_i , the minimization of $\eta_i^2(q_i)$ is transferred into the boundary value problem: find $\varphi_i \in H^1(\Omega_i)$ such that

$$a_i(\varphi_i, \omega) = L_i(\omega) - a_i(u_h, \omega) + \int_{\partial\Omega_i} \omega \langle n_i \cdot a \nabla u_h \rangle ds \quad \omega \in H^1(\Omega_i) \quad (4.6)$$

where $a_i(\cdot, \cdot)$ is the bilinear form, $L_i(\cdot)$ is the linear form, $\partial\Omega_i$ is the boundary of element Ω_i and

$$\langle n_i \cdot a \nabla u_h \rangle = n_i \cdot \left[(1 - \alpha_y) a \nabla u_h|_i + \alpha_y a \nabla u_h|_j \right] \quad (4.7)$$

Here n_i is the outward norm to $\partial\Omega_i$ and α_y is a linear function obtained by flux splitting. Obviously, in solving the boundary problem (4.6), the residuals are used as data. For details on how to perform flux splitting, we refer to [6] and [13].

The implicit residual error estimator then takes the form of

$$\|\bar{e}\|_R = \left(\sum_{i=1}^N \|\bar{e}\|_{(i),R}^2 \right)^{\frac{1}{2}} = \left(\sum_{i=1}^N \eta_i^2(q_i) \right)^{\frac{1}{2}} \quad (4.8)$$

In the following, we shall construct the recovery type error estimators. Assume that $a\nabla u^*$ is the recovered solution, and

$$q_i = a\nabla u^* - a\nabla u_h \quad (4.9)$$

The recovery type of error estimator is written as

$$\|\bar{e}\| = \left(\sum_{i=1}^N \|\bar{e}\|_{(i)}^2 \right)^{\frac{1}{2}} \quad (4.10)$$

with

$$\begin{aligned} \|\bar{e}\|_{(i)}^2 &= \int_{\Omega_i} a(\nabla u_{(i)}^* - \nabla u_{h(i)})^T (\nabla u_{(i)}^* - \nabla u_{h(i)}) d\Omega \\ &= \int_{\Omega_i} \frac{1}{a} (a\nabla u_{(i)}^* - a\nabla u_{h(i)})^T (a\nabla u_{(i)}^* - a\nabla u_{h(i)}) d\Omega \\ &= \int_{\Omega_i} \frac{1}{a} q_i \cdot q_i d\Omega \end{aligned} \quad (4.11)$$

The following recovery technique is used in the computation of the recovered solution $a\nabla u^*$.

RECOVERY PROCEDURE: The recovered solution $a\nabla u_{(i)}^*$ is determined by the minimization of

$$\begin{aligned} F &= \int_{\Omega_i} a(\nabla u_{(i)}^* - \nabla u_{h(i)})^T (\nabla u_{(i)}^* - \nabla u_{h(i)}) d\Omega + \int_{\Omega_i} r^{*2} d\Omega \\ &= \int_{\Omega_i} a(\nabla u_{(i)}^* - \nabla u_{h(i)})^T (\nabla u_{(i)}^* - \nabla u_{h(i)}) d\Omega + \int_{\Omega_i} (-\nabla \cdot (a\nabla u^*) + cu_h - f)^2 d\Omega \end{aligned} \quad (4.12)$$

Immediately we observe that

$$F = \int_{\Omega_i} \frac{1}{a} q_i \cdot q_i d\Omega + \int_{\Omega_i} (\nabla \cdot q_i + r_i)^2 d\Omega \quad (4.13)$$

Compare with

$$\eta_i^2(q_i) = \varepsilon_i^2(q_i) + \Lambda_i^2(q_i)$$

$$= \int_{\Omega} \frac{1}{a} q_i \cdot q_i d\Omega + \int_{\Omega} \frac{1}{c} (\nabla \cdot q_i + r_i)^2 d\Omega \quad (4.14)$$

The equivalence of F and η_i^2 is evident. Therefore, the basic formulations used in the computation of residual type error estimator and the recovery type error estimator are equivalent. The methodology in deriving these error estimators and the computational implementation of these formulations are, however, completely different.

The minimization condition of F is imposed over a patch of elements, as described in [16], in computing the recovered solution. The flux splitting is not required in the computation of the error estimator, which makes the recovery type error estimator computationally more efficient. The recovery procedure described here is similar, but not identical, to the one used by Wiberg *et al.* [19] and Belytschko *et al.* [20], which is a modification of the superconvergent patch recovery (SPR) technique originally proposed by Zienkiewicz and Zhu [16-18].

5. NUMERICAL STUDY OF THE EQUIVALENCE OF THE ERROR ESTIMATORS

We have, in Section 4, established the equivalence relationship of η_i^2 and F . The equivalence of the recovery error estimator $\|\bar{e}\|$ and the residual error estimator $\|\bar{e}\|_R$ is however, not immediately apparent, because the procedures used in searching q_i to satisfy the minimization condition of η_i^2 and F are different. In this section we shall numerically study the equivalence of the recovery error estimator and the residual error estimator discussed in Section 4.

We define θ as the effectivity index of the error estimator, which is the ratio of the estimated error and the true error. Effectivity is used as a measure of the quality of the error estimator. The error estimator is said to be reliable if θ is close to one while the finite solution converges to the true solution.

From (3.1) we know that the equivalence of the recovery error estimator $\|\bar{e}\|$ and the residual error estimator $\|\bar{e}\|_R$ means that

$$C_1(1+O(h)) \leq \frac{\|\bar{e}\|}{\|\bar{e}\|_R} \leq C_2(1+O(h)). \quad (5.1)$$

In the numerical studies, we are interested in determining the value of C_1 and C_2 asymptotically. In other words, we are interested in the asymptotic equivalence of the error estimators. The numerical test will be performed over a patch of elements.

Note that the value of C_1 and C_2 are different dependent on the solution type and the mesh pattern. Further, for a mesh patch in the interior of the domain the values will be different from that at the boundary. Following assumptions are made in the numerical study:

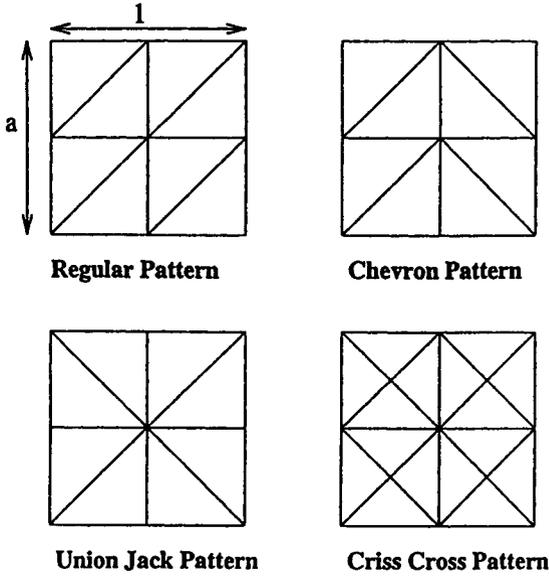


Figure 1. Mesh patterns used in the numerical study.

Laplace Equation, $\Delta u = 0$								
Aspect-ratio	Regular Pattern		Chevron Pattern		Union Jack Pattern		Criss Cross Pattern	
	C_L	C_U	C_L	C_U	C_L	C_U	C_L	C_U
Linear Elements; Full Residual Estimator								
1/1	1.225	1.732	1.225	1.732	1.225	1.732	1.225	1.732
1/2	1.225	1.732	1.225	1.732	1.225	1.732	1.225	1.732
1/4	1.225	1.732	1.225	1.732	1.225	1.732	1.225	1.732
1/8	1.225	1.732	1.225	1.732	1.225	1.732	1.225	1.732
1/16	1.225	1.732	1.225	1.732	1.225	1.732	1.225	1.732
Linear Elements; Bubble Residual Estimator								
1/1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1/2	1.000	1.000	1.000	1.000	1.000	1.008	0.906	1.009
1/4	1.000	1.000	1.000	1.000	1.000	1.008	0.781	1.046
1/8	1.000	1.000	1.000	1.000	1.000	1.003	0.728	1.089
1/16	1.000	1.000	1.000	1.000	1.000	1.001	0.713	1.109
Linear Elements; Recovery Estimator								
1/1	0.885	0.902	0.868	0.902	0.890	0.902	0.920	1.042
1/2	0.890	0.904	0.861	0.914	0.895	0.915	0.920	1.049
1/4	0.893	0.904	0.860	0.919	0.895	0.917	0.920	1.058
1/8	0.894	0.905	0.860	0.921	0.896	0.921	0.920	1.062
1/16	0.894	0.905	0.860	0.921	0.896	0.921	0.920	1.063

Table 1. Range of the effectivity index as a function of the aspect ratio for the four mesh patterns. Laplace Equation. Linear element.

Laplace Equation, $\Delta u = 0$								
Aspect-ratio	Regular Pattern		Chevron Pattern		Union Jack Pattern		Criss Cross Pattern	
	C_L	C_U	C_L	C_U	C_L	C_U	C_L	C_U
Quadratic Elements; Full Residual Estimator								
1/1	1.281	1.541	1.409	1.417	1.417	1.417	1.417	1.417
1/2	1.339	1.487	1.407	1.419	1.407	1.423	1.271	1.724
1/4	1.339	1.508	1.342	1.508	1.341	1.508	1.235	1.958
1/8	1.306	1.558	1.306	1.559	1.306	1.558	1.227	2.047
1/16	1.294	1.575	1.295	1.575	1.295	1.575	1.225	2.078
Quadratic Elements; Bubble Residual Estimator								
1/1	1.000	1.000	1.000	1.008	1.000	1.000	1.000	1.000
1/2	1.000	1.000	1.000	1.009	1.000	1.000	0.922	0.977
1/4	1.000	1.000	1.000	1.004	1.000	1.000	0.796	0.985
1/8	1.000	1.000	1.000	1.001	1.000	1.000	0.733	0.995
1/16	1.000	1.000	1.000	1.000	1.000	1.000	0.714	0.999
Quadratic Elements; Recovery Estimator								
1/1	0.983	0.999	0.989	0.991	0.999	1.001	1.001	1.001
1/2	0.988	0.997	0.984	0.998	0.998	1.002	1.005	1.013
1/4	0.989	0.994	0.981	0.995	0.997	1.003	1.005	1.029
1/8	0.989	0.993	0.980	0.993	0.996	1.004	1.005	1.034
1/16	0.990	0.993	0.980	0.992	0.996	1.004	1.005	1.035

Table 2. Range of the effectivity index as a function of the aspect ratio for the four mesh patterns. Laplace Equation. Quadratic element.

Laplace Equation, $\Delta u = 0$				
Criss Cross Pattern				
Linear Elements; Equivalence Constants				
Aspect-Ratio	Recovery/Full Residual		Recovery/Bubble Residual	
	C_1	C_2	C_1	C_2
1/1	1.272	1.817	0.919	1.042
1/2	1.304	1.883	0.912	1.159
1/4	1.317	1.977	0.879	1.354
1/8	1.313	2.019	0.844	1.458
1/16	1.313	2.032	0.829	1.491
Quadratic Elements; Equivalence Constants				
Aspect-Ratio	Recovery/Full Residual		Recovery/Bubble Residual	
	C_1	C_2	C_1	C_2
1/1	1.407	1.407	1.001	1.001
1/2	1.268	1.687	1.029	1.099
1/4	1.237	2.015	1.021	1.293
1/8	1.236	2.296	1.010	1.409
1/16	1.238	2.439	1.006	1.449

Table 3. Equivalence constant for recovery type error estimator and residual type error estimator. Laplace equation. Criss cross pattern. Linear and quadratic elements.

Orthotropic Heat Conduction, $\nabla \cdot Ku = 0$				
Quadratic Elements; Principal Material Constants $K_1 = 1000, K_2 = 1$				
Angle of Orientation	Chevron Pattern		Criss Cross Pattern	
	C_L	C_U	C_L	C_U
Full Residual Estimator				
00	1.292	1.579	1.225	2.079
15	3.657	12.285	1.349	15.062
30	5.611	17.407	3.808	14.229
45	1.002	25.903	1.292	1.579
60	2.119	25.989	3.808	14.229
75	4.607	16.134	1.349	15.062
90	1.292	1.579	1.225	2.079
Bubble Residual Estimator				
00	1.000	1.000	0.709	1.000
15	0.885	1.000	0.805	0.989
30	0.839	1.000	0.941	0.997
45	0.742	1.000	1.000	1.000
60	0.934	1.000	0.941	0.997
75	0.989	1.000	0.805	0.989
90	1.000	1.000	0.709	1.000
Recovery Estimator				
00	0.980	0.998	1.005	1.038
15	0.989	0.999	0.997	1.050
30	0.997	1.001	0.997	1.026
45	0.993	0.994	0.998	1.004
60	0.991	0.998	0.997	1.026
75	0.993	0.998	0.997	1.050
90	0.996	0.997	1.005	1.038

Table 4. Range of the effectivity index as a function of the mesh-material orientation for the orthotropic heat conduction problem. Chevron pattern and criss cross pattern. Quadratic element.

Orthotropic Heat Conduction, $\nabla \cdot Ku = 0$				
Criss Cross Pattern				
Quadratic Elements; Equivalence Constants				
Angle of Orientation	Recovery/Full Residual		Recovery/Bubble Residual	
	C_1	C_2	C_1	C_2
00	0.498	0.821	1.005	1.459
15	0.069	0.740	1.012	1.300
30	0.072	0.263	1.001	1.089
45	0.635	0.773	0.999	1.004
60	0.072	0.263	1.001	1.089
75	0.069	0.740	1.012	1.300
90	0.498	0.821	1.005	1.459

Table 5. Equivalence constant for recovery type error estimator and residual type error estimator. Orthotropic heat conduction problem. Criss cross pattern. Quadratic elements.

- (1) The solution of the test problem is smooth.
- (2) The mesh patch considered is in the interior of the domain.
- (3) The mesh is locally periodic.

Essentially, it means that the asymptotic solution corresponds to the local $(p+1)$ Taylor series expansion of the exact solution (see, e.g. [27] for details). Thus, following the procedure proposed by Babuska *et al* [27] we will be able to compute the asymptotic value of C_1 and C_2 . We note that, as observed in [28], the conclusions obtained based on the periodic meshes also apply to the general meshes.

The following problems are considered in the numerical study:

(a) $\Delta u = 0$ (Laplace equation)

and

(b) $\nabla \cdot (K\nabla u) = 0$ (Orthotropic heat conduction equation)

as reported in [27-28]. Where K is the 2×2 matrix with principal value K_1, K_2 and the principal axes oriented at φ° to the coordinates axes.

The residual error estimator is computed in two different finite element space [27-28]. The corresponding error estimators are termed as full residual error estimator $\|\bar{e}\|_{R,full}$ and bubble residual error estimator $\|\bar{e}\|_{R,bubble}$ respectively.

We study the performance of the error estimators by, as reported in Table 1 and 2, the values of the bounds C_L and C_U of the effectivity index ($C_L \leq \theta \leq C_U$) for recovery error estimator, full residual error estimator and bubble residual error estimator for problem (a). Here all the four mesh patterns shown in Figure 1 are tested. The value of the equivalence constant C_1 and C_2 are reported in Table 3 for Criss Cross pattern. We observe that both recovery error estimator and bubble residual error estimator are robust and C_1, C_2 are very close to one for ratio of the recovery type error estimator and the bubble residual error estimator $\|\bar{e}\| / \|\bar{e}\|_{R,bubble}$. For ratio $\|\bar{e}\| / \|\bar{e}\|_{R,full}$, C_1 and C_2 are also asymptotically constants, although the full residual error estimator is less robust..

We present the results of problem (b) from similar studies for various value of the orientation of the principal material axes in Table 4 and Table 5. Here $K_1 = 1000$ and $K_2 = 1$. The recovery type error estimator is found to be robust and the bubble residual error estimator is more robust than the full residual error estimator. Again it is observed that C_1 and C_2 are very close to one for $\|\bar{e}\| / \|\bar{e}\|_{R,bubble}$. However, For ratio $\|\bar{e}\| / \|\bar{e}\|_{R,full}$, C_1 and C_2 are functions of the orientation of the principal material axes.

We conclude that the bubble residual error estimator is numerically equivalent to the recovery error estimator and it's performance is also similar to the recovery error estimator. The full residual error estimator, on the other hand, although numerically equivalent to the recovery error estimator, but their numerical performances are not close.

Remark 5.1. We refer to [27-29] for the tests of the numerical performance of various error estimators. It was concluded that, in general, the recovery type error estimators are more robust than the residual type error estimators.

ACKNOWLEDGMENT

The authors would like to thank Dr. C.S. Upadhyay for providing the numerical results presented in this paper. The work of J. Z. Zhu was supported by UES, Inc. The work of Zhimin Zhang was supported by the National Science Foundation Grant DMS-9626193, DMS-9622690 and INT-9605050.

REFERENCES

1. I. Babuska and W. C. Rheinboldt, 'A posteriori error estimates for the finite element method', *Int. J. Numer. Meth. Engrg.*, **12**, 1597-1615, (1978).
2. I. Babuska and W. C. Rheinboldt, 'Reliable error estimation and mesh adaptation for the finite element method', *Computational method in nonlinear mechanics*, (J. T. Oden, ed.) North Holland, Amsterdam, 67-108, (1980).
3. I. Babuska and W. C. Rheinboldt, 'A posteriori error analysis of finite element solutions for one-dimensional problems', *SIAM J. Numer. Anal.*, **18**, 565-589, (1981).
4. O. C. Zienkiewicz, J. P. Gago and D. W. Kelly, 'The hierarchical concept in the finite element method', *Comput. Struct.*, **16**, 53-65, (1983).
5. D. W. Kelly, J. R. Gago, O. C. Zienkiewicz and I. Babuska, 'A posteriori error analysis and adaptive processes in the finite element method, Part I: error analysis', *Int. J. Numer. Meth. Engrg.*, **19**, 1593-1619, (1983).
6. P. Ladeveze and D. Leguillon, 'Error estimate procedure in the finite element method and applications', *SIAM J. Numer. Anal.*, **20**, 485-509, (1983).
7. D. W. Kelly, 'The self-equilibration of residuals and complementary a posteriori error estimates in the finite element method.', *Int. J. Numer. Meth. Engrg.*, **20**, 1491-1506, (1984).
8. R. E. Bank and A. Weiser, 'Some a posteriori error estimators for elliptic partial differential equations', *Math. Comput.*, **44**, 283-301, (1985).
9. I. Babuska and A. Miller, 'A feedback finite element method with a posteriori error estimation: Part I. The finite element method and some basic properties of the a posteriori error estimator', *Comput. Methods Appl. Mech. Engrg.*, **61**, 1-40, (1987).
10. J. T. Oden, L. Demkowicz, W. Rachowicz and T. A. Westermann, 'Toward a universal h-p adaptive finite element strategy. Part 2: a posteriori error estimation', *Comput. Methods Appl. Mech. Engrg.*, **77**, 113-180, (1989).
11. R. Verfurth, 'A posteriori error estimators for the Stokes equations', *Numer. Math.*, **55**, 09-325, (1989).

12. C. Johnson and P. Hansbo, 'Adaptive finite element methods in computational mechanics', *Comput. Methods Appl. Mech. Engrg.*, **101**, 143-181, (1992).
13. M. Ainsworth and J. T. Oden, 'A unified approach to a posteriori estimation using element residual methods', *Numer. Math.*, **65**, 23-50, (1993).
14. O. C. Zienkiewicz and J. Z. Zhu, 'A simple error estimator and adaptive procedure for practical engineering analysis', *Int. J. Numer. Meth. Eng.*, **24**, 337-357, (1987).
15. J. Z. Zhu and O. C. Zienkiewicz, 'Superconvergent recovery technique and a posteriori error estimators', *Int. J. Numer. Meth. Eng.*, **30**, 1321-1340, (1990).
16. O. C. Zienkiewicz and J. Z. Zhu, 'Superconvergent patch recovery and a posteriori error estimates. Part I: The recovery technique', *Int. J. Numer. Meth. Eng.*, **33**, 1331-1364, (1992).
17. O. C. Zienkiewicz and J. Z. Zhu, 'Superconvergent patch recovery and a posteriori error estimates. Part II: Error estimates and adaptivity', *Int. J. Numer. Meth. Eng.*, **33**, 1365-1382, (1992).
18. O. C. Zienkiewicz and J. Z. Zhu, 'Superconvergent patch recovery (SPR) and adaptive finite element refinement', *Comp. Meth. Appl. Meth. Eng.*, **101**, 207-224, (1992).
19. N. E. Wiberg and F. Abdulwahab, 'Patch recovery based on superconvergent derivatives and equilibrium', *Int. J. Numer. Meth. Eng.*, **36**, 2703-2724, (1993).
20. T. Blacker and T. Belychko, 'Superconvergent patch recovery with equilibrium and conjoint interpolant enhancements', *Int. J. Numer. Meth. Eng.*, **37**, 517-536, (1994).
21. E. Rank and O. C. Zienkiewicz, 'A simple error estimator in the finite element method', *Comm. Appl. Numer. Meth.*, **3**, 243-249, (1987).
22. R. Rodriguez, 'Some remarks on Zienkiewicz-Zhu estimator', *Int. J. Numer. Meth. in PDE.*, **10**, 625-635, (1994).
23. J.Z. Zhu, 'A posteriori error estimation - the relationship between different procedures', *Comput. Methods Appl. Mech. Engrg.*, To be published.
24. O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method*, 4th Edition, Vol. 1: Basic Formulation and Linear Problems, McGraw-Hill, London, (1989).
25. Z. Zhang, 'Ultraconvergence of the patch recovery technique', *Math. Comp.* **65**, 1431-1437, (1996).
26. L.B. Wahlbin, 'Local behavior in finite element methods', in: P.G. Ciarlete and J.L. Lions, eds., *Handbook of Numerical Analysis*, Vol. II (North-Holland, Amsterdam, 1991) 357-522.
27. I. Babuska, T. Strouboulis and C. S. Upadhyay, 'A model study of the quality of a posteriori error estimators for linear elliptic problems. Error estimation in the interior of patchwise uniform grids of triangles', *Comp. Meth. Appl. Meth. Eng.* **114**, 307-378, (1994).
28. I. Babuska, T. Strouboulis, C. S. Upadhyay, S. K. Gangaraj and K. Copps, 'Validation of a posteriori error estimators by numerical approach', *Int. J. Numer. Methods Engrg.* **37**, 1073-1123, (1994).
29. I. Babuska, T. Strouboulis and C. S. Upadhyay, 'A model study of a posteriori error estimators for linear elliptic problems. Part II: Error estimation at the boundary of patchwise uniform grids of triangles', to be published.

A technique for a *posteriori* error estimation of h - p approximations of the Stokes equations

J.T. Oden * and S. Prudhomme †

Texas Institute for Computational and Applied Mathematics,
The University of Texas at Austin, Taylor Hall TAY 2.400, Austin TX 78712, USA

This paper deals with a *posteriori* error estimation for finite element approximations of the Stokes problem. The ultimate objective is to analyse how given error quantities of interest are influenced by the residuals, which are viewed as the sources of the numerical error. First, a global error estimator is proposed in terms of the norms of the residuals. Then, techniques to efficiently estimate these norms are advocated. This is followed by the investigation of a general approach to evaluate the numerical error in pointwise, local or global quantities of interest.

1. INTRODUCTION

Error estimation in computational processes has been a subject of interest for more than two decades since the pioneering work of Babuška and Rheinboldt [3]. We refer to Ainsworth and Oden [2] and Verfürth [19] for an extended account of the subject. In the particular case of the Stokes problem, several approaches have been investigated in [18,8,6,16,2].

The method we propose here belongs to the family of *Implicit Error Residual Methods*. The errors in the velocity and pressure variables are driven by the residuals \mathcal{R}_h^m and \mathcal{R}_h^c in the momentum equation and the continuity equation respectively. The residuals represent the sources of error in the finite element approximations, and as such, are post-processed to provide meaningful error estimates in global “energy” norms. The evaluation of \mathcal{R}_h^c is shown to be exact, local and cheap. The calculation of the error measure \mathcal{R}_h^m is however more demanding. A new technique has been developed which provides accurate approximations of \mathcal{R}_h^m through a global but inexpensive iterative process. First, error estimates are obtained using spaces of low-order bubble functions as perturbations. These are then corrected by using enriched spaces constructed through an adaptive procedure. This approach avoids the major difficulty of prescribing proper boundary conditions for each subproblem in the Element Residual method [11,7,2]. The effectiveness of the methodology is demonstrated on various test cases of the Stokes problem with smooth and unsmooth solutions.

In recent years, the success of a *posteriori* error estimation has prompted users to demand error estimates in quantities of interest other than the classical energy norm.

*Director, TICAM, Cockrell Family Regents Chair in Engineering.

†Graduate Research Assistant.

Works in this field have been undertaken by Babuška and Strouboulis [4] and Becker and Rannacher [10,9] in order to estimate and control the error by adapting the mesh parameters with respect to these quantities of interest. We extend the ideas to the Stokes problem and perform a preliminary investigation of the performance of such techniques to evaluate the error in pointwise, local or global quantities of interest.

2. PRELIMINARIES

Let Ω denote an open bounded Lipschitz domain in \mathbb{R}^n , $n = 2$ or 3 , with boundary $\partial\Omega$. We consider the Stokes equations:

$$\begin{aligned} -\Delta \mathbf{u} + \nabla p &= \mathbf{f} & \text{in } \Omega \\ \nabla \cdot \mathbf{u} &= 0 & \text{in } \Omega \end{aligned} \quad (1)$$

with Dirichlet boundary condition:

$$\mathbf{u} = \mathbf{g}, \quad \text{on } \partial\Omega, \quad (2)$$

where $\mathbf{u} = \mathbf{u}(\mathbf{x})$ and $p = p(\mathbf{x})$ are respectively a vector-valued and a scalar-valued function defined at point $\mathbf{x} = (x_1, x_2, \dots, x_n)$ in Ω . Since the Stokes equations can be derived as a linearization of the steady-state Navier-Stokes equations, the variable \mathbf{u} and p are referred to as the velocity and pressure. The source term $\mathbf{f} = \mathbf{f}(\mathbf{x})$ is a prescribed body force and \mathbf{g} is a function defined on $\partial\Omega$ which must satisfy the compatibility condition,

$$\oint_{\partial\Omega} \mathbf{g} \cdot \mathbf{n} \, ds = 0. \quad (3)$$

In what follows, we restrict ourselves to the case of homogeneous boundary conditions, i.e. $\mathbf{g} = \mathbf{0}$ on $\partial\Omega$. This simplifies somewhat the theoretical analysis of the Stokes equations, while retaining all their interesting features. We shall use standard notations for various Sobolev spaces of functions defined on Ω . We begin by introducing the trial spaces of velocities \mathbf{V} and pressures Q defined by:

$$\mathbf{V} = \mathbf{H}_0^1(\Omega) = (\mathbf{H}_0^1(\Omega))^n, \quad Q = \{q \in L^2(\Omega) : \int_{\Omega} q \, dx = 0\},$$

with corresponding norms:

$$|\mathbf{v}|_1^2 = \int_{\Omega} \nabla \mathbf{v} : \nabla \mathbf{v} \, dx, \quad \|q\|_0^2 = \int_{\Omega} q^2 \, dx.$$

We also introduce the bilinear forms a and b ,

$$a : \mathbf{H}^1(\Omega) \times \mathbf{H}^1(\Omega) \longrightarrow \mathbb{R}; \quad a(\mathbf{u}, \mathbf{v}) = \int_{\Omega} \nabla \mathbf{u} : \nabla \mathbf{v} \, dx$$

$$b : \mathbf{H}^1(\Omega) \times L^2(\Omega) \longrightarrow \mathbb{R}; \quad b(\mathbf{v}, q) = - \int_{\Omega} q \nabla \cdot \mathbf{v} \, dx.$$

The bilinear form a is the inner product associated to the norm $|\cdot|_1$ in \mathbf{V} , so that

$$a(\mathbf{v}, \mathbf{v}) = |\mathbf{v}|_1^2, \quad \forall \mathbf{v} \in \mathbf{V}.$$

Lemma 1 *The bilinear form b is continuous on $V \times Q$, and in particular, there exists a positive constant M_b such that*

$$b(\mathbf{v}, q) \leq M_b |\mathbf{v}|_1 \|q\|_0, \quad \forall \mathbf{v} \in V, \forall q \in Q,$$

where $M_b = \sqrt{n}$, where n is the geometrical dimension of the problem.

Proof Let \mathbf{v} and q be arbitrary functions in V and Q . Since $\nabla \cdot \mathbf{v} \in Q$, we have:

$$b(\mathbf{v}, q) \leq \|\nabla \cdot \mathbf{v}\|_0 \|q\|_0.$$

It is now sufficient to show there exists M_b such that $\|\nabla \cdot \mathbf{v}\|_0 \leq M_b |\mathbf{v}|_1$. Indeed,

$$\|\nabla \cdot \mathbf{v}\|_0^2 = \int_{\Omega} \left(\sum_{i=1}^n \frac{\partial v_i}{\partial x_i} \right)^2 dx \leq \int_{\Omega} n \sum_{i=1}^n \left(\frac{\partial v_i}{\partial x_i} \right)^2 dx \leq n \int_{\Omega} \nabla \mathbf{v} : \nabla \mathbf{v} dx = n |\mathbf{v}|_1^2. \quad (4)$$

so that $\|\nabla \cdot \mathbf{v}\|_0 \leq \sqrt{n} |\mathbf{v}|_1$ and $M_b = \sqrt{n}$.

Moreover, it can be shown that b satisfies the standard LBB condition (see Girault and Raviart [12]), i.e. there exists a constant $\beta > 0$ such that

$$\sup_{\mathbf{v} \in V \setminus \{0\}} \frac{|b(\mathbf{v}, q)|}{|\mathbf{v}|_1} \geq \beta \|q\|_0, \quad \forall q \in Q. \quad (5)$$

The Stokes problem is now reformulated in the equivalent weak form:

<p>For $\mathbf{f} \in V'$ given, find $(\mathbf{u}, p) \in V \times Q$, such that</p> $\begin{aligned} a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, p) &= \langle \mathbf{f}, \mathbf{v} \rangle, \quad \forall \mathbf{v} \in V \\ b(\mathbf{u}, q) &= 0, \quad \forall q \in Q \end{aligned} \quad (6)$

Extensive results concerning the existence, uniqueness and regularity of solutions (\mathbf{u}, p) in $V \times Q$ of the Stokes problem can be found in Girault and Raviart [12].

Let $V^h \subset V$ and $Q^h \subset Q$ denote finite element spaces, possibly h - p finite element spaces [14], of the spaces V and Q . Approximate solutions to problem (6) are then obtained by solving the following system of discrete equations:

<p>For $\mathbf{f} \in V'$ given, find $(\mathbf{u}_h, p_h) \in V^h \times Q^h$, such that</p> $\begin{aligned} a(\mathbf{u}_h, \mathbf{v}) + b(\mathbf{v}, p_h) &= \langle \mathbf{f}, \mathbf{v} \rangle, \quad \forall \mathbf{v} \in V^h \\ b(\mathbf{u}_h, q) &= 0, \quad \forall q \in Q^h \end{aligned} \quad (7)$

Let us consider a pair $(\mathbf{u}_h, p_h) \in V^h \times Q^h$, not necessarily a solution of (7). The numerical error $(\mathbf{e}, E) \in V \times Q$ in (\mathbf{u}_h, p_h) is defined as

$$(\mathbf{e}, E) = (\mathbf{u}, p) - (\mathbf{u}_h, p_h). \quad (8)$$

Then, substituting \mathbf{u} and p in (6) by $\mathbf{u}_h + \mathbf{e}$ and $p_h + E$ respectively, we show that the distribution of (\mathbf{e}, E) is governed by the system of equations:

$\begin{aligned} a(\mathbf{e}, \mathbf{v}) + b(\mathbf{v}, E) &= \mathcal{R}_h^a(\mathbf{v}), \quad \forall \mathbf{v} \in V \\ b(\mathbf{e}, q) &= \mathcal{R}_h^b(q), \quad \forall q \in Q \end{aligned} \quad (9)$
--

where the linear functionals $\mathcal{R}_h^m : V \rightarrow \mathbb{R}$ and $\mathcal{R}_h^c : Q \rightarrow \mathbb{R}$

$$\begin{aligned}\mathcal{R}_h^m(\mathbf{v}) &= \langle \mathcal{R}_h^m, \mathbf{v} \rangle \equiv \langle \mathbf{f}, \mathbf{v} \rangle - a(\mathbf{u}_h, \mathbf{v}) - b(\mathbf{v}, p_h) \\ \mathcal{R}_h^c(q) &= \langle \mathcal{R}_h^c, q \rangle \equiv -b(\mathbf{u}_h, q)\end{aligned}$$

are respectively the residual in the *momentum equation* and the residual in the *continuity equation*. The residuals \mathcal{R}_h^m and \mathcal{R}_h^c are viewed as the *sources of error*. Indeed, whenever they are zero, the numerical error is zero as well. We show in the next section how the residuals can provide reliable global error estimates in specific norms.

3. GLOBAL ERROR ESTIMATION

The objective here is to relate the residuals to the numerical errors expressed in some global norms. We recall (see [12, p.22]) that the space $V = H_0^1(\Omega)$ can be decomposed into the direct sum $V = J \oplus J^\perp$, where the spaces J and J^\perp are defined as:

$$\begin{aligned}J &= \{\mathbf{v} \in V; b(\mathbf{v}, q) = 0, \quad \forall q \in Q\}, \\ J^\perp &= \{\mathbf{v} \in V; a(\mathbf{v}, \mathbf{w}) = 0, \quad \forall \mathbf{w} \in J\}.\end{aligned}$$

The space J^\perp is the orthogonal complement of J with respect to $H_0^1(\Omega)$ for the inner product $a(\cdot, \cdot)$. It follows that the error in the velocity variable $\mathbf{e} \in V$ can be uniquely decomposed into a sum of two vectors, $\mathbf{e}_d \in J$ and $\mathbf{e}_\perp \in J^\perp$, such that $\mathbf{e} = \mathbf{e}_d + \mathbf{e}_\perp$.

We naturally measure the error (\mathbf{e}, E) in the usual global norm

$$\|(\mathbf{e}, E)\|^2 = |\mathbf{e}|_1^2 + \|E\|_0^2 = |\mathbf{e}_d|_1^2 + |\mathbf{e}_\perp|_1^2 + \|E\|_0^2. \quad (10)$$

and consider the following norms for the residuals \mathcal{R}_h^m and \mathcal{R}_h^c

$$\|\mathcal{R}_h^m\|_* = \sup_{\mathbf{v} \in V \setminus \{0\}} \frac{|\mathcal{R}_h^m(\mathbf{v})|}{|\mathbf{v}|_1}, \quad \|\mathcal{R}_h^c\|_* = \sup_{q \in Q \setminus \{0\}} \frac{|\mathcal{R}_h^c(q)|}{\|q\|_0}. \quad (11)$$

Lemma 2 *With the above definitions and assumptions:*

$$\beta |\mathbf{e}_\perp|_1 \leq \|\mathcal{R}_h^c\|_* \leq M_b |\mathbf{e}_\perp|_1. \quad (12)$$

Proof From the definition of the residual \mathcal{R}_h^c , we have, for all $q \in Q$

$$\mathcal{R}_h^c(q) = b(\mathbf{e}, q) = b(\mathbf{e}_\perp, q) \leq M_b |\mathbf{e}_\perp|_1 \|q\|_0.$$

The upper bound in (12) follows as

$$\|\mathcal{R}_h^c\|_* = \sup_{q \in Q} \frac{\mathcal{R}_h^c(q)}{\|q\|_0} \leq \sup_{q \in Q} \frac{M_b |\mathbf{e}_\perp|_1 \|q\|_0}{\|q\|_0} \leq M_b |\mathbf{e}_\perp|_1.$$

The divergence operator is an isomorphism of J^\perp onto Q , so the function $\nabla \cdot \mathbf{e}_\perp$ belongs to Q and therefore

$$\|\nabla \cdot \mathbf{e}_\perp\|_0^2 = -\mathcal{R}_h^c(\nabla \cdot \mathbf{e}_\perp) \leq \|\mathcal{R}_h^c\|_* \|\nabla \cdot \mathbf{e}_\perp\|_0$$

which yields $\|\nabla \cdot \mathbf{e}_\perp\|_0 \leq \|\mathcal{R}_h^c\|_*$. From Girault and Raviart [12, p.24, p.81], we obtain the lower bound in (12) as

$$\beta |\mathbf{e}_\perp|_1 \leq \|\nabla \cdot \mathbf{e}_\perp\|_0 \leq \|\mathcal{R}_h^c\|_*.$$

Lemma 3 *With the above definitions and assumptions, we have:*

$$(1) \quad |e_d|_1 \leq \|\mathcal{R}_h^m\|_* . \quad (13)$$

$$(2) \quad \|\mathcal{R}_h^m\|_* \leq |e|_1 + M_b \|E\|_0 . \quad (14)$$

$$(3) \quad \beta \|E\|_0 \leq \|\mathcal{R}_h^m\|_* + |e_\perp|_1 , \quad (15)$$

Proof

1. By simple algebra, we get:

$$|e_d|_1^2 = a(e_d, e_d) = a(e, e_d) = \mathcal{R}_h^m(e_d) - b(e_d, E) = \mathcal{R}_h^m(e_d) \leq \|\mathcal{R}_h^m\|_* |e_d|_1 ,$$

which shows (13).

2. We also have

$$\begin{aligned} \|\mathcal{R}_h^m\|_* &= \sup_{\mathbf{v} \in \mathbf{V}} \frac{\mathcal{R}_h^m(\mathbf{v})}{|\mathbf{v}|_1} \leq \sup_{\mathbf{v} \in \mathbf{V}} \frac{a(\mathbf{e}, \mathbf{v}) + b(\mathbf{v}, E)}{|\mathbf{v}|_1} \\ &\leq \sup_{\mathbf{v} \in \mathbf{V}} \frac{a(\mathbf{e}, \mathbf{v})}{|\mathbf{v}|_1} + \sup_{\mathbf{v} \in \mathbf{V}} \frac{b(\mathbf{v}, E)}{|\mathbf{v}|_1} \leq |e|_1 + M_b \|E\|_0 . \end{aligned}$$

3. The upper bound for the error in the pressure variable follows from the *inf-sup* condition on the bilinear form b , i.e.

$$\begin{aligned} \beta \|E\|_0 &\leq \sup_{\mathbf{v} \in \mathbf{V}} \frac{b(\mathbf{v}, E)}{|\mathbf{v}|_1} \leq \sup_{\mathbf{v} \in \mathbf{J}^\perp} \frac{b(\mathbf{v}, E)}{|\mathbf{v}|_1} \leq \sup_{\mathbf{v} \in \mathbf{J}^\perp} \frac{\mathcal{R}_h^m(\mathbf{v}) - a(\mathbf{e}, \mathbf{v})}{|\mathbf{v}|_1} \\ &\leq \sup_{\mathbf{v} \in \mathbf{J}^\perp} \frac{\mathcal{R}_h^m(\mathbf{v})}{|\mathbf{v}|_1} + \sup_{\mathbf{v} \in \mathbf{J}^\perp} \frac{a(\mathbf{e}, \mathbf{v})}{|\mathbf{v}|_1} \leq \sup_{\mathbf{v} \in \mathbf{V}} \frac{\mathcal{R}_h^m(\mathbf{v})}{|\mathbf{v}|_1} + \sup_{\mathbf{v} \in \mathbf{J}^\perp} \frac{a(\mathbf{e}_\perp, \mathbf{v})}{|\mathbf{v}|_1} \\ &\leq \|\mathcal{R}_h^m\|_* + |e_\perp|_1 . \end{aligned}$$

Theorem 1 *Let (\mathbf{e}, E) be the numerical error in an approximation (\mathbf{u}_h, p_h) of the Stokes problem. Then there exist positive constants C_1 and C_2 such that*

$$C_1 \|(\mathbf{e}, E)\|^2 \leq \|\mathcal{R}_h^c\|_*^2 + \|\mathcal{R}_h^m\|_*^2 \leq C_2 \|(\mathbf{e}, E)\|^2 \quad (16)$$

where C_1 and C_2 depend only on the constants β and M_b respectively:

$$C_1 = 1/2 \min(1, \beta^4), \quad C_2 = \max(2 + M_b^2, 2M_b^2). \quad (17)$$

Proof The proof of this theorem directly follows from Lemma 2 and Lemma 3. In particular, the lower bound is obtained from inequalities (12), (13) and (15), while the upper bound is derived from the inequalities (12) and (14).

This theorem is similar to the Theorem 6.1 introduced by Ainsworth and Oden in [2]. It shows that the global quantity $\|\mathcal{R}_h^c\|_*^2 + \|\mathcal{R}_h^m\|_*^2$ is equivalent to the global error measure $\|(e, E)\|^2$. We emphasize that it represents a meaningful quantitative global estimate as long as the constants C_1 and C_2 do not take values far from one. Here the constant C_2 takes the value 4 in two-dimensions and 6 in three-dimensions since $M_b = \sqrt{n}$. Therefore, the accuracy and robustness of this error estimate essentially depends on the value of C_1 , a fortiori, β , which depends itself on the problem in consideration. Moreover, we emphasize that such a global result does not provide any reliable information about the local (elementwise or pointwise) error, as we know that the error can propagate far away from their sources \mathcal{R}_h^m and \mathcal{R}_h^c . This actually motivated the work undertaken by Babuška, Strouboulis and co-workers [4] and Oden and Feng [15] on *pollution error* in order to estimate the contributions of the residuals to the error outside the region of interest. However, the quantity $\|\mathcal{R}_h^m\|_*^2 + \|\mathcal{R}_h^c\|_*^2$ provides information about the location and intensity of the sources of errors, and as such, should be used to determine local refinement indicators.

4. EVALUATING NORMS OF RESIDUALS

This section is devoted to the evaluation of $\|\mathcal{R}_h^m\|_*$ and $\|\mathcal{R}_h^c\|_*$. The objectives are, on one hand, to minimize the cost of the computations while retaining a certain accuracy, and, on the other hand, to obtain elementwise contributions of these quantities.

4.1. Residual in the continuity equation

The residual \mathcal{R}_h^c gives us information about whether the discrete velocity \mathbf{u}_h does or does not satisfy the incompressibility constraint. This is stated in the next lemma, where the norm of the residual \mathcal{R}_h^c is expressed in terms of the divergence of \mathbf{u}_h .

Lemma 4 *Let $\mathbf{u}_h \in V^h$ be the discrete velocity of the Stokes problem. Then,*

$$\|\mathcal{R}_h^c\|_* = \|\nabla \cdot \mathbf{u}_h\|_0. \quad (18)$$

Proof From the definition of the residual \mathcal{R}_h^c , we have, for all q in Q :

$$\mathcal{R}_h^c(q) = -b(\mathbf{u}_h, q) = \int_{\Omega} q \nabla \cdot \mathbf{u}_h \, dx \leq \|\nabla \cdot \mathbf{u}_h\|_0 \|q\|_0.$$

It follows that

$$\|\mathcal{R}_h^c\|_* = \sup_{q \in Q} \frac{\mathcal{R}_h^c(q)}{\|q\|_0} \leq \sup_{q \in Q} \frac{\|\nabla \cdot \mathbf{u}_h\|_0 \|q\|_0}{\|q\|_0} \leq \|\nabla \cdot \mathbf{u}_h\|_0. \quad (19)$$

We also have

$$\|\nabla \cdot \mathbf{u}_h\|_0^2 = \mathcal{R}_h^c(\nabla \cdot \mathbf{u}_h) \leq \|\mathcal{R}_h^c\|_* \|\nabla \cdot \mathbf{u}_h\|_0,$$

that is, $\|\nabla \cdot \mathbf{u}_h\|_0 \leq \|\mathcal{R}_h^c\|_*$, and the equality (18) has just been proven.

The result of Lemma 4 is remarkable in the sense that the evaluation of $\|\mathcal{R}_h^c\|_*$ is both exact and cheap, as it is equivalent to compute the $L^2(\Omega)$ -norm of the function $\nabla \cdot \mathbf{u}_h$. In addition, it is straightforward to decompose $\|\mathcal{R}_h^c\|_*$ into a sum of elementwise contributions as:

$$\|\mathcal{R}_h^c\|_*^2 = \|\nabla \cdot \mathbf{u}_h\|_0^2 = \sum_K \|\nabla \cdot \mathbf{u}_h\|_{0,K}^2 = \sum_K \int_{\Omega_K} |\nabla \cdot \mathbf{u}_h|^2 dx$$

where \sum_K represents the sum over all the elements in a finite element partition.

4.2. Residual in the momentum equation

Simple considerations reveal that the evaluation of $\|\mathcal{R}_h^m\|_*$ is neither cheap nor exact. At best, one obtains approximations of it. There exist basically two approaches to compute such approximations. The first one always delivers lower bounds on $\|\mathcal{R}_h^m\|_*$ while the other provides upper bounds. Here we only consider the former, which is referred to as the *conforming method*.

4.2.1. Exact evaluation

Since V is a Hilbert space, the *Riesz Representation Theorem* tells us there exists a unique element φ in V which satisfies:

$$a(\varphi, \mathbf{v}) = \mathcal{R}_h^m(\mathbf{v}), \quad \forall \mathbf{v} \in V, \quad (20)$$

as well as:

$$|\varphi|_1 = \sqrt{a(\varphi, \varphi)} = \|\mathcal{R}_h^m\|_*. \quad (21)$$

In other words, the residual $\mathcal{R}_h^m \in V'$ is identified with an element $\varphi \in V$ of equal norm. Moreover, the function φ gives information about the local intensity of the residual \mathcal{R}_h^m , such as:

$$\|\mathcal{R}_h^m\|_*^2 = |\varphi|_1^2 = \sum_K |\varphi|_{1,K}^2 = \sum_K \int_{\Omega_K} \nabla \varphi : \nabla \varphi dx.$$

However, the problem (20) cannot be solved numerically due to the infinite dimension of the space V .

4.2.2. Conforming finite element approximations

The objective here is to construct a finite element subspace $\tilde{V}^h \subset V$, hence the label *conforming method*, in which to approximate problem (20), that is, to find $\varphi_h \in \tilde{V}^h$ such that:

$$a(\varphi_h, \mathbf{v}) = \mathcal{R}_h^m(\mathbf{v}), \quad \forall \mathbf{v} \in \tilde{V}^h. \quad (22)$$

We show that the norm of φ_h is always smaller than the norm of \mathcal{R}_h^m and that both quantities are indeed equivalent. First, we observe from (20) and (22) that:

$$a(\varphi - \varphi_h, \mathbf{v}) = 0, \quad \forall \mathbf{v} \in \tilde{V}^h. \quad (23)$$

In consequence, φ_h is simply the orthogonal projection of φ onto the space \widetilde{V}^h with respect to the inner product $a(\cdot, \cdot)$ associated to $|\cdot|_1$. It follows that:

$$|\varphi_h|_1^2 + |\varphi - \varphi_h|_1^2 = |\varphi|_1^2, \quad (24)$$

which yields the inequality $|\varphi - \varphi_h|_1 \leq |\varphi|_1$. Such a result can be sharpened, as stated in the next lemma.

Lemma 5 *Let $\varphi \in V$ and $\varphi_h \in \widetilde{V}^h$ satisfy (20) and (22) respectively. If $\varphi \neq 0$, let \widetilde{V}^h be such that $\varphi_h \neq 0$. Then, there exists a constant σ , $0 \leq \sigma < 1$ such that:*

$$|\varphi - \varphi_h|_1 \leq \sigma |\varphi|_1. \quad (25)$$

Proof Since $|\varphi_h|_1 \neq 0$, $|\varphi - \varphi_h|_1$ is strictly less than $|\varphi|_1$ by (24).

Theorem 2 *Let $\varphi \in V$, $\varphi_h \in \widetilde{V}^h$ and \widetilde{V}^h satisfy the conditions of Lemma 5. Then, there exists a constant σ , $0 \leq \sigma < 1$, such that:*

$$\sqrt{1 - \sigma^2} \|\mathcal{R}_h^m\|_* \leq |\varphi_h|_1 \leq \|\mathcal{R}_h^m\|_*. \quad (26)$$

Proof The upper bound follows from equality (24). Indeed, $|\varphi_h|_1 \leq |\varphi|_1 = \|\mathcal{R}_h^m\|_*$. Now, making use of (24) and (25), we have:

$$|\varphi|_1^2 - |\varphi_h|_1^2 = |\varphi - \varphi_h|_1^2 \leq \sigma^2 |\varphi|_1^2,$$

that is

$$(1 - \sigma^2) |\varphi|_1^2 \leq |\varphi_h|_1^2, \quad (27)$$

from which the lower bound follows.

Theorem 2 shows that the quantities $|\varphi_h|_1$ and $\|\mathcal{R}_h^m\|_*$ are equivalent. However, $|\varphi_h|_1$ provides a valuable approximation of $\|\mathcal{R}_h^m\|_*$ as long as $\sqrt{1 - \sigma^2}$ is close to one. The constant σ clearly depends on how rich the finite element space \widetilde{V}^h is, but, at the same time, we point out that even a crude approximation φ_h of φ may be sufficient to obtain an acceptable estimate of $\|\mathcal{R}_h^m\|_*$. For instance, by committing a relative error $\sigma = 40\%$ in approximating φ , we still approximate $\|\mathcal{R}_h^m\|_*$ with an effectivity of 91.6%.

The finite element space \widetilde{V}^h has to be chosen so that the action of the residual \mathcal{R}_h^m is different from zero for at least one element of \widetilde{V}^h . In the case where the pair (\mathbf{u}_h, p_h) is the solution of problem (7), we observe that the residual vanishes on V^h , that is:

$$\mathcal{R}_h^m(\mathbf{v}) = 0, \quad \forall \mathbf{v} \in V^h. \quad (28)$$

It follows that the space \widetilde{V}^h should be larger than V^h itself. It is then constructed by enriching V^h with elements of a space W^h , called *space of perturbations*, which satisfies

$$W^h \neq \{0\}, \quad W^h \subset V, \quad V^h \cap W^h = \{0\}. \quad (29)$$

so that $V^h \subset \widetilde{V}^h = V^h + W^h \subset V$. Since $V^h \subset \widetilde{V}^h$, problem (22) is more expensive to solve than the original one. At this point, we want to reduce the cost of the computations by taking advantage of the fact that the residual vanishes on the space V^h . In other words, we would like to approximate $\|\mathcal{R}_h^m\|_*$ by the norm of the function $\psi_h \in W^h$ satisfying

$$a(\psi_h, v) = \mathcal{R}_h^m(v), \quad \forall v \in W^h. \quad (30)$$

Following Bank [5], we suppose that a *Strengthened Cauchy-Schwartz Inequality* holds with respect to the spaces V^h and W^h , in the sense that there exists a positive constant $\gamma < 1$ such that for all $v_h \in V^h$ and for all $w_h \in W^h$,

$$a(v_h, w_h) \leq \gamma |v_h|_1 |w_h|_1, \quad (31)$$

which implies, using Young's inequality, that

$$\begin{aligned} |v_h + w_h|_1^2 &= |v_h|_1^2 + 2a(v_h, w_h) + |w_h|_1^2 \geq |v_h|_1^2 - 2\gamma |v_h|_1 |w_h|_1 + |w_h|_1^2 \\ &\geq |v_h|_1^2 - |v_h|_1^2 - \gamma^2 |w_h|_1^2 + |w_h|_1^2 = (1 - \gamma^2) |w_h|_1^2. \end{aligned}$$

Thus,

$$|v_h + w_h|_1 \geq \sqrt{1 - \gamma^2} |w_h|_1. \quad (32)$$

The *Strengthened Cauchy-Schwartz Inequality* allows us to relate the accuracy of the approximation $|\psi_h|_1 \approx |\varphi_h|_1$ to γ , the cosine of the angle between the spaces V^h and W^h :

Theorem 3 *Let $\varphi \in V$, $\varphi_h \in \widetilde{V}^h$ and \widetilde{V}^h satisfy the conditions of Lemma 5. Let the strengthened Cauchy-Schwartz inequality (31) hold for the spaces V^h and W^h , where $\widetilde{V}^h = V^h + W^h$ and $\mathcal{R}_h^m = 0$ on V^h . Let $\psi_h \in W^h$ satisfy (30). Then*

$$\sqrt{(1 - \sigma^2)(1 - \gamma^2)} \|\mathcal{R}_h^m\|_* \leq |\psi_h|_1 \leq \|\mathcal{R}_h^m\|_*. \quad (33)$$

Proof The upper bound is readily obtained as $W^h \subset V$. In order to prove the lower bound, we write $\varphi_h = \varphi_h^* + \psi_h^*$, where $\varphi_h^* \in V^h$ and $\psi_h^* \in W^h$. Then, from equation (22), we have, since $\varphi_h^* \in V^h$:

$$a(\varphi_h, \varphi_h^*) = \mathcal{R}_h^m(\varphi_h^*) = 0, \quad (34)$$

and from (22) and (30), we get:

$$a(\varphi_h, \psi_h^*) = \mathcal{R}_h^m(\psi_h^*) = a(\psi_h, \psi_h^*). \quad (35)$$

Hence,

$$\begin{aligned} |\varphi_h|_1^2 &= a(\varphi_h, \varphi_h) = a(\varphi_h, \varphi_h^* + \psi_h^*) = a(\varphi_h, \psi_h^*) = a(\psi_h, \psi_h^*) \\ &\leq |\psi_h|_1 |\psi_h^*|_1. \end{aligned} \quad (36)$$

Applying the strengthened Cauchy-schwartz inequality to solution φ_h yields:

$$\sqrt{1 - \gamma^2} |\psi_h^*|_1 \leq |\varphi_h^* + \psi_h^*|_1 = |\varphi_h|_1, \quad (37)$$

which, combined to (36), gives

$$|\varphi_h|_1^2 \leq |\psi_h|_1 |\psi_h^*|_1 \leq \frac{1}{\sqrt{1 - \gamma^2}} |\psi_h|_1 |\varphi_h|_1,$$

that is,

$$|\varphi_h|_1 \leq \frac{1}{\sqrt{1 - \gamma^2}} |\psi_h|_1. \quad (38)$$

Then, using inequality (27) shown in Theorem 2, allows us to write:

$$\sqrt{1 - \sigma^2} |\varphi|_1 \leq |\varphi_h|_1 \leq \frac{1}{\sqrt{1 - \gamma^2}} |\psi_h|_1. \quad (39)$$

The lower bound is proved.

The above theorem shows that the accuracy of the approximation $|\psi_h|_1 \approx |\varphi|_1$ only depends on the constants σ and γ .

We recall that σ depends on the richness of the space \widetilde{V}^h , or W^h . The main difficulty is to construct W^h so that the approximations $\psi_h \in W^h$ deliver reliable estimates of $\|\mathcal{R}_h^m\|_*$ at the lowest cost. In h - p finite element methods, the perturbations can be conveniently constructed from layers of piecewise polynomial basis functions involving monomials of degree between $p+1$ and $p+q$, $q \geq 1$, where p is defined as the maximal degree of the basis functions in V^h . Obviously, in two-dimensional problems, such basis functions consist of *edge* and/or *interior* bubble functions. In three-dimensional problems, they would consist of *edge*, *face* and/or *interior* bubble functions. One question one has to answer is "what is the best value for q ?" Actually, this is problem dependent, and one should consider a method in which the value of q is increased as needed.

We recall that γ represents the angle between V^h and W^h . Its value is directly related to the choice of the shape functions used to construct the basis functions of V^h and W^h . Here we use hierarchic shape functions based on integrated Legendre polynomials (see Szabó and Babuška [17, chapt. 6]), which satisfy the orthogonality property with respect to the inner product $a(\cdot, \cdot)$.

Meanwhile, the cost in solving (30) is controlled by the dimension of W^h . However, the bilinear form $a(\cdot, \cdot)$ is symmetric and positive definite so that the finite system (30) can be cheaply solved using the iterative *Conjugate Gradient method* (CG). Moreover, we emphasize that the solution does not have to be highly accurate as shown by Theorem 2, so that it is usually sufficient to perform a few iterations.

Finally, conforming methods always deliver lower bounds on \mathcal{R}_h^m . In order to construct upper bounds, one has to approximate problem (20) in a space larger than V . This lead to the *equilibrated residual method* developed by Ainsworth and Oden [1,2] or Ladevèze and Leguillon [13]. In this approach, the global problem (20) is decoupled into a collection of local problems, usually constructed on each of the elements in the partition. Our present conforming method avoids the major difficulty of prescribing boundary conditions for each subproblem typical of the equilibrium methods at comparable cost.

5. ESTIMATION AND CONTROL OF ERROR QUANTITIES OF INTEREST

Let L denote a linear functional defined on the product space $V \times Q$. We suppose that we are interested in the *error quantity* $L(e, E) \in \mathbb{R}$, which we want to estimate and control. For example, $L(\cdot, \cdot)$ may represent something more localized than the global estimate (16), such as the error in \mathbf{u} and p or their gradients at a point $\mathbf{x}_0 \in \Omega$ or along a curve $\Gamma \subset \Omega$; e.g. $L(e, E) = E(\mathbf{x}_0)$ or $L(e, E) = \int_{\Gamma} \mathbf{e} \cdot \mathbf{n} \, ds$, etc. We cite more examples later. The main objective is then to relate $L(e, E)$ to the sources of error \mathcal{R}_h^m and \mathcal{R}_h^c ; in other words, we would like to find linear functionals ω^m and ω^c , if they exist, such that

$$L(e, E) = \omega^m(\mathcal{R}_h^m) + \omega^c(\mathcal{R}_h^c). \quad (40)$$

These functions are viewed as *influence functions* as they indicate the influence of the residuals on the quantity $L(e, E)$. They are defined on the bidual of V and Q , and since these spaces are reflexive, we have:

$$\omega^m \in V \quad \text{and} \quad \omega^c \in Q. \quad (41)$$

This is understood in the sense that for each $\omega^m \in V''$ and each $\omega^c \in Q''$, there exist $\hat{\omega}^m \in V$ and $\hat{\omega}^c \in Q$ respectively such that

$$\langle \omega^m, \mathcal{R}_h^m \rangle_{V'' \times V'} = \langle \mathcal{R}_h^m, \hat{\omega}^m \rangle_{V' \times V} \quad \text{and} \quad \langle \omega^c, \mathcal{R}_h^c \rangle_{Q'' \times Q'} = \langle \mathcal{R}_h^c, \hat{\omega}^c \rangle_{Q' \times Q}.$$

By identifying $\hat{\omega}^m$ and $\hat{\omega}^c$ with ω^m and ω^c respectively, the relation (40) becomes

$$L(e, E) = \mathcal{R}_h^m(\omega^m) + \mathcal{R}_h^c(\omega^c). \quad (42)$$

Substituting for the terms $\mathcal{R}_h^m(\omega^m)$ and $\mathcal{R}_h^c(\omega^c)$ in (42) using (9), rearranging and assuming a symmetric, we finally get

$$L(e, E) = a(e, \omega^m) + b(\omega^m, E) + b(e, \omega^c) \quad (43)$$

$$= a(\omega^m, e) + b(e, \omega^c) + b(\omega^m, E). \quad (44)$$

The influence functions can thus be obtained as solutions of the global *dual problem*:

Find $(\omega^m, \omega^c) \in V \times Q$, such that $a(\omega^m, \mathbf{v}) + b(\mathbf{v}, \omega^c) + b(\omega^m, q) = L(\mathbf{v}, q), \quad \forall (\mathbf{v}, q) \in V \times Q$	(45)
---	------

which is equivalent to the *generalized Stokes problem*:

Find $(\omega^m, \omega^c) \in V \times Q$, such that $a(\omega^m, \mathbf{v}) + b(\mathbf{v}, \omega^c) = L(\mathbf{v}, 0), \quad \forall \mathbf{v} \in V$ $b(\omega^m, q) = L(0, q), \quad \forall q \in Q$	(46)
---	------

It readily follows that the functions ω^m and ω^c do exist and are unique as $L(\mathbf{v}, 0) \in V'$ and $L(0, q) \in Q'$ (see Girault and Raviart [12]). Obviously, (46) cannot be solved exactly for the functions ω^m and ω^c in the general case. At best, one seeks numerical approximations of ω^m and ω^c . This raises two questions, which should be simulatenously investigated:

1. How do we effectively calculate finite element approximations of ω^m and ω^c ? In particular, is it necessary to solve a global problem ?
2. How do we utilize the relation (42) to derive elementwise refinement indicators in order to control the error quantity $L(e, E)$?

From a computational point of view, we observe that the cost involved in approximating ω^m and ω^c in the spaces V^h and Q^h is almost negligible, as the resulting finite system has already been factorized once to calculate the solution (\mathbf{u}_h, p_h) . The cost therefore reduces to perform one backward and one forward substitutions. However, we recall that the residuals are identically zero on $V^h \times Q^h$ whenever the pair (\mathbf{u}_h, p_h) is the solution of the finite system (7). Therefore, if we approximate ω^m and ω^c by $\omega_h^m \in V^h$ and $\omega_h^c \in Q^h$, we obtain that

$$\mathcal{R}_h^m(\omega_h^m) = 0 \quad \text{and} \quad \mathcal{R}_h^c(\omega_h^c) = 0, \quad (47)$$

which implies

$$L(e, E) = \mathcal{R}_h^m(\omega^m) + \mathcal{R}_h^c(\omega^c) \approx \mathcal{R}_h^m(\omega_h^m) + \mathcal{R}_h^c(\omega_h^c) = 0.$$

In this case, the quantity $L(e, E)$ would be estimated by a quantity which proves to be 0. This reveals that ω^m and ω^c should be approximated in spaces of greater dimensions than V^h and Q^h .

One strategy, proposed by Becker and Rannacher [10], consists in using (47) such that

$$\begin{aligned} L(e, E) &= \mathcal{R}_h^m(\omega^m) + \mathcal{R}_h^c(\omega^c) - \mathcal{R}_h^m(\omega_h^m) - \mathcal{R}_h^c(\omega_h^c) \\ &= \mathcal{R}_h^m(\omega^m - \omega_h^m) + \mathcal{R}_h^c(\omega^c - \omega_h^c). \end{aligned}$$

Then, an upper bound can be derived as:

$$\begin{aligned} |L(e, E)| &= |\mathcal{R}_h^m(\omega^m - \omega_h^m) + \mathcal{R}_h^c(\omega^c - \omega_h^c)| = |a(\varphi, \omega^m - \omega_h^m) - b(\mathbf{u}_h, \omega^c - \omega_h^c)| \\ &= \left| \sum_K a_K(\varphi, \omega^m - \omega_h^m) - \sum_K b_K(\omega^c - \omega_h^c) \right| \\ &\leq \sum_K |a_K(\varphi, \omega^m - \omega_h^m)| + \sum_K |b_K(\omega^c - \omega_h^c)| \\ &\leq \sum_K \left(|\varphi|_{1,K} |\omega^m - \omega_h^m|_{1,K} + \|\nabla \cdot \mathbf{u}_h\|_{0,K} \|\omega^c - \omega_h^c\|_{0,K} \right) \end{aligned} \quad (48)$$

$$\leq \sum_K \sqrt{|\varphi|_{1,K}^2 + \|\nabla \cdot \mathbf{u}_h\|_{0,K}^2} \sqrt{|\omega^m - \omega_h^m|_{1,K}^2 + \|\omega^c - \omega_h^c\|_{0,K}^2}, \quad (49)$$

where φ is the function in V satisfying (20). Becker and Rannacher [10] use (48) to estimate the quantities $L(e, E)$ using local interpolation properties to evaluate the error quantity $|\omega^m - \omega_h^m|_{1,K}$ and $\|\omega^c - \omega_h^c\|_{0,K}$. They thus introduce an interpolation constant they arbitrarily choose to be 1. Here, we could estimate $L(e, E)$ by introducing the global

error estimates developed in section 3 into (49). However, those estimates are not local and do not reflect the error propagation in ω_h^m and ω_h^c . In general, such a procedure overestimates $|L(e, E)|$ by several orders of magnitude. If one is interested in accurate values of $L(e, E)$, one has to consider another approach.

The strategy we propose is to solve a global problem defined in some finite element spaces \tilde{V}^h and \tilde{Q}^h satisfying

$$V^h \subset \tilde{V}^h \subset V, \quad Q^h \subset \tilde{Q}^h \subset Q.$$

Obviously it is more expensive to solve this global problem than the solution problem in $V^h \times Q^h$. In order to optimize the selection of \tilde{V}^h and \tilde{Q}^h , we propose to construct them by adapting V^h and Q^h according to a combination of global error estimates on (ω_h^m, ω_h^c) as well as error estimates on $L(e, E)$ using (49). Thus, we can utilize the global error estimates (very cheap) developed in the previous sections. Let $(\tilde{\omega}_h^m, \tilde{\omega}_h^c)$ denote the solution in $\tilde{V}^h \times \tilde{Q}^h$. Then we have

$$L(e, E) \approx \mathcal{R}_h^m(\tilde{\omega}_h^m) + \mathcal{R}_h^c(\tilde{\omega}_h^c) \tag{50}$$

$$\leq \sum_K \sqrt{|\varphi|_{1,K}^2 + \|\nabla \cdot \mathbf{u}_h\|_{0,K}^2} \sqrt{|\tilde{\omega}_h^m - \omega_h^m|_{1,K}^2 + \|\tilde{\omega}_h^c - \omega_h^c\|_{0,K}^2}. \tag{51}$$

Therefore, the quantity $L(e, E)$ is estimated using (50) and controlled by adapting the mesh according to the elementwise indicators produced by (51).

We now provide some examples for which we would be interested in calculating the quantity $L(e, E) \in \mathbb{R}$. The error measure $L(e, E)$ obviously represents the numerical error we do in computing $L(\mathbf{u}_h, p_h)$ instead of $L(\mathbf{u}, p)$. Indeed,

$$L(e, E) = L(\mathbf{u}, p) - L(\mathbf{u}_h, p_h).$$

Linear quantities of potential interest in the variable (\mathbf{u}, p) are pointwise values of the velocity component u , $L(\mathbf{u}, p) = u(\mathbf{x}_0)$, $\mathbf{x}_0 \in \Omega$, volume flow rates through a surface Γ , $L(\mathbf{u}, p) = \int_{\Gamma} \mathbf{u} \cdot \mathbf{n} ds$, where \mathbf{n} is the unit normal to Γ , or directional derivatives of the velocity averaged over one element Ω_K , $L(\mathbf{u}, p) = \int_{\Omega_K} \nabla \mathbf{u} \cdot \mathbf{l} dx$, where \mathbf{l} is a given unit vector. In the case one is interested in nonlinear quantities $N(\mathbf{u}, p) \in \mathbb{R}$, one performs the expansion:

$$N(\mathbf{u}, p) = N(\mathbf{u}_h, p_h) + N'(\mathbf{u}_h, p_h) \cdot (\mathbf{u} - \mathbf{u}_h, p - p_h) + \dots$$

so that the error quantity $N(\mathbf{u}, p) - N(\mathbf{u}_h, p_h)$ may be approximated by

$$N(\mathbf{u}, p) - N(\mathbf{u}_h, p_h) \approx N'(\mathbf{u}_h, p_h) \cdot (e, E).$$

Then, we denote the linear quantity $N'(\mathbf{u}_h, p_h) \cdot (e, E)$ by $L(e, E)$ and all the analysis described above applies. For example, let us suppose that one is interested in the kinetic energy K_e in some subregion Ω_K of Ω . We employ

$$N(\mathbf{u}, p) = K_e(\mathbf{u}) = \frac{1}{2} \int_{\Omega_K} \mathbf{u}^2 dx.$$

It results that the error in the kinetic energy may be approximated by

$$K_e(\mathbf{u}) - K_e(\mathbf{u}_h) \approx L(e, E) = \int_{\Omega_K} \mathbf{u}_h \cdot \mathbf{e} dx.$$

6. NUMERICAL EXPERIMENTS

6.1. Global error estimation

In the first example, the Stokes problem is solved on the unit square $\Omega = [0, 1] \times [0, 1]$ with the data f chosen such that the exact solution is given by the velocity profile $\mathbf{u} = (u, v)$ shown in Figure 2 and by the pressure $p = 0$. We consider uniform meshes formed of squared elements with size $h \times h$, $h = 1/2, 1/4, 1/8, 1/16, 1/32$. The polynomial degree p of the basis functions in \mathbf{V}^h is uniformly set to 2 or 3.

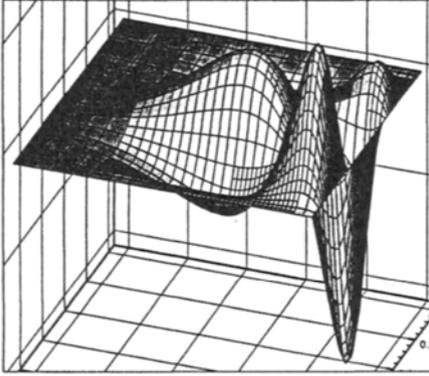


Figure 1. Exact velocity component u .

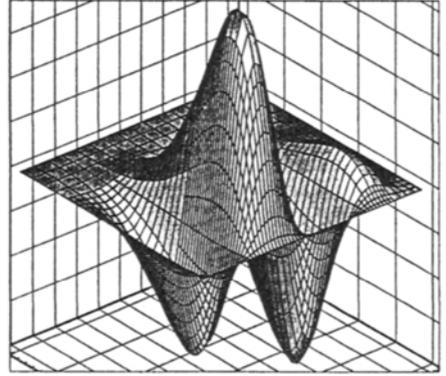


Figure 2. Exact velocity component v .

We consider the global error estimator η defined as:

$$\eta^2 = \|\mathcal{R}_h^m\|_*^2 + \|\mathcal{R}_h^i\|_*^2 = |\varphi|_1^2 + \|\nabla \cdot \mathbf{u}_h\|_0^2$$

where $\varphi \in \mathbf{V}$ is approximated by a function $\varphi_h \in \widetilde{\mathbf{V}}^h$ or a function $\psi_h \in \mathbf{W}^h$ satisfying (22) or (30). We characterize the extra degree (with respect to p) of the perturbations in \mathbf{W}^h by the pair $\mathbf{q} = (q_{\text{edge}}, q_{\text{int}})$, q_{edge} being the extra degree for the edge bubbles and q_{int} the extra degree for the interior bubbles. We recall that the cost of this error estimator is controlled by the cost in solving (22) or (30), that is, by the size of \mathbf{W}^h and the number of iterations performed using the Conjugate Gradient method. What is sought is a trade-off between accuracy in the approximations ψ_h and cost (time) spent to solve for them. Values of approximations of φ and corresponding times are given in Table 1 and Table 2 for various values of h , p and \mathbf{q} . The approximations are normalized with respect to overkills using $\mathbf{q} = (2, 2)$ while the times are normalized with respect to the times spent to solve for the solutions (\mathbf{u}_h, p_h) . The number of iterations in the CG algorithm are determined according to a preset tolerance in solution accuracy. We first observe that the performances greatly increases as the size of the problem increases. For example, for $h = 1/16$, $p = 2$, the residual is estimated within a 1% accuracy while spending less than 7% of the solution time. We also observe that it is approximated with

Table 1
Normalized values of approximations of $\|\mathcal{R}_h^m\|_*$.

h	p	W^h			\widetilde{V}^h
		$q = (0, 1)$	$q = (1, 0)$	$q = (1, 1)$	$q = (1, 1)$
1/2	2	0.7620	0.5235	0.8520	0.8571
1/4		0.7300	0.8282	0.8983	0.9040
1/8		0.8154	0.9684	0.9797	0.9805
1/16		0.8184	0.9941	0.9947	0.9948
1/4	3	0.8406	0.8952	0.9679	0.9693
1/8		0.8815	0.9781	0.9895	0.9896

Table 2
Normalized times and number of iterations () performed using the CG method.

h	p	W^h			\widetilde{V}^h
		$q = (0, 1)$	$q = (1, 0)$	$q = (1, 1)$	$q = (1, 1)$
1/2	2	0.169 (2)	0.196 (2)	0.695 (6)	3.454 (8)
1/4		0.065 (2)	0.117 (4)	0.443 (7)	2.925 (9)
1/8		0.015 (2)	0.042 (4)	0.241 (8)	1.309 (9)
1/16		0.006 (2)	0.018 (3)	0.069 (5)	0.474 (8)
1/4	3	0.071 (2)	0.013 (4)	0.123 (9)	0.921 (9)
1/8		0.004 (2)	0.004 (4)	0.088 (9)	0.339 (9)

less than one percent of error using W^h instead of \widetilde{V}^h with $q = (1, 1)$. Moreover the number of iterations is kept very small while requiring reasonable accuracy of the order 10^{-2} . We also compute effectivity indices, defined as:

$$\gamma^{u,p} = \frac{\eta}{\|(e, E)\|}$$

as well as elementwise effectivity indices. Results are collected in Table 3 for the case $\psi_h \in W^h$, $q = (1, 1)$. We also show in Figure 3 and Figure 4 the elementwise distribution of the exact errors and of the effectivity indices for $h = 1/8$ and $p = 2$. The global effectivity indices are excellent as they are all close to unity. On the other hand, the local effectivity indices, as it is expected for the case of a global error estimator, may behave poorly, and in some cases drop to a value as low as 0.012. However, these are always near unity in elements with large error, and diverge from unity only in elements with rather small error.

6.2. Backward facing step Stokes problem

The second example is the classical backward facing step Stokes problem. The problem is slightly different than in Bank and Welfert [8], as we consider a homogeneous Neumann

Table 3
Global and elementwise effectivity indices.

h	p	$\gamma^{u,p}$	$\min \gamma_K^{u,p}$	$\max \gamma_K^{u,p}$
1/2	2	0.760	0.717	0.921
1/4		0.927	0.274	1.103
1/8		1.038	0.069	1.321
1/16		1.066	0.033	1.388
1/32		1.073	0.012	1.408
1/4	3	1.028	0.519	1.051
1/8		1.084	0.167	1.339
1/16		1.096	0.025	1.381

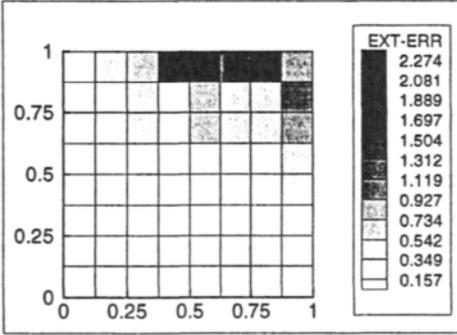


Figure 3. Elementwise distribution of the errors.

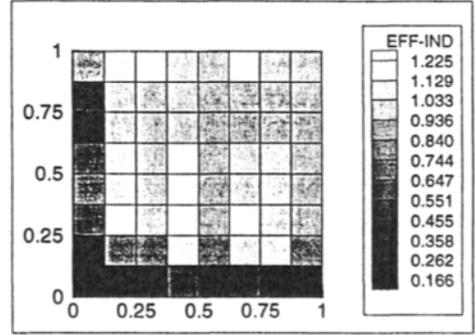


Figure 4. Elementwise distribution of the effectivity indices.

boundary condition downstream of the domain. The solution develops a discontinuity in the pressure variable due to the discontinuity in the geometry of the boundary. The norm of the residual \mathcal{R}_h^m is then expected to be very large in the near region of the reentrant corner and small in the farfield, as illustrated in Figure 5, where the solution has been computed on a pre-adapted mesh of 120 elements.

We are then motivated to construct \mathbf{W}^h in an adaptive manner in order to minimize its dimension, that is, the number of degrees of freedom associated to problem (30). As a first guess, the residual is approximated by $\psi_h \in \mathbf{W}^h$ with $\mathbf{q} = (1, 0)$. Then the space is successively enriched with interior and edge bubble functions of higher degrees only in the elements for which the contribution to $|\psi_h|_1$ is large. Results are shown in Table 4 where the approximations of $\|\mathcal{R}_h^m\|_*$, computed either in the complete space $\mathbf{W}_{\text{cmp}}^h$, with uniform $\mathbf{q} = (q_{\text{edge}}, q_{\text{int}})$, or in the adapted space $\mathbf{W}_{\text{adp}}^h$ with \mathbf{q} varying between $(0, 0)$

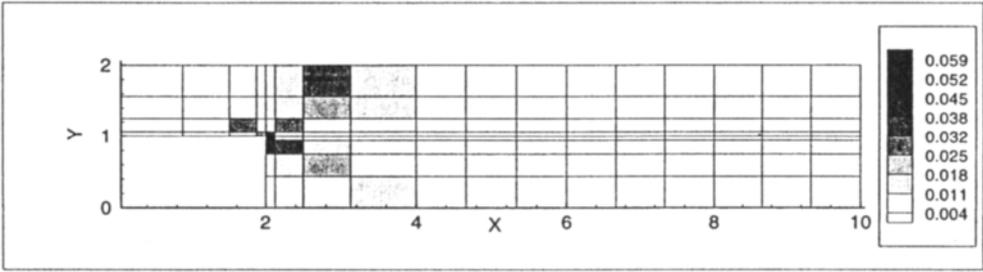


Figure 5. Elementwise distribution of the residual $\|\mathcal{R}_h^m\|_*$.

to $(q_{\text{edge}}, q_{\text{int}})$ depending on the element, are compared. It is clear that the number of degrees of freedom, and a fortiori the cost of the error estimator, is greatly reduced for comparable accuracy. The global effectivity index for this problem has been computed as $\gamma^{u,p} = 0.884$ and 0.915 for $p = 2$ and 3 respectively, noting that the exact solutions are actually approximated by overkills of degree $p + 3$.

Table 4

Approximations of $\|\mathcal{R}_h^m\|_*$ in complete or adapted spaces of perturbations with corresponding numbers of degrees of freedom (\cdot).

p	$q = (1, 0)$		$q = (1, 1)$		$q = (2, 1)$		$q = (2, 2)$	
	W_{cmp}^h	W_{adp}^h	W_{cmp}^h	W_{adp}^h	W_{cmp}^h	W_{adp}^h	W_{cmp}^h	W_{adp}^h
2	0.1078	0.1247	0.1275	0.1271	0.1301	0.1344	0.1361	
	(530)	(650)	(1250)	(716)	(1780)	(922)	(2980)	
3	0.0291	0.0527	0.0531	0.0535	0.0539	0.0607	0.0612	
	(530)	(650)	(1730)	(686)	(2260)	(854)	(3940)	

In Figures 6 and 7, we respectively show an adapted mesh and the corresponding solution (\mathbf{u}_h, p_h) . Such a mesh has been automatically adapted from a given coarse mesh in order to achieve a preset tolerance of 1.2% in the global relative error.

6.3. Error estimation in the quantity of interest

The following results constitute a preliminary study of error estimation in local quantities of interest. We consider the test case described in section 6.1 and suppose we are interested in evaluating the quantity

$$L(\mathbf{u}, p) = \frac{1}{|\Omega_K|} \int_{\Omega_K} \frac{\partial u}{\partial y} dx \in \mathbb{R}$$

where $|\Omega_K|$ represents the area of a given subdomain $\Omega_K \subset \Omega$. Then, the error in the computed quantity $L(\mathbf{u}_h, p_h)$, where (\mathbf{u}_h, p_h) is the finite element solution in $V^h \times Q^h$, is

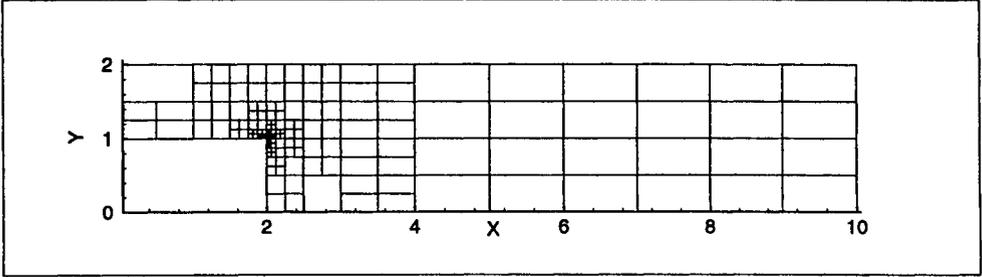


Figure 6. Example of an adapted mesh (201 elements).

given by $L(e, E)$. We consider a uniform mesh with $h = 1/8$ and $p = 2$ and choose the subdomain Ω_K to be either element 1 located in the lower left corner of the domain, or element 55 the next to last element in the upper right corner. Then, $L(e, E)$ is estimated by three quantities in \mathbb{R} denoted η_1 , η_2 and η_3 . The first one, η_1 , is obtained as:

$$\eta_1 = \mathcal{R}_h^m(\tilde{\omega}_h^m) + \mathcal{R}_h^c(\tilde{\omega}_h^c)$$

where the discrete influence functions $\tilde{\omega}_h^m$ and $\tilde{\omega}_h^c$ belong to the spaces \tilde{V}^h and \tilde{Q}^h defined using basis functions up to the degree $p + 1$ and are solutions of the generalized Stokes problem (46). The second one, η_2 , is calculated as

$$\eta_2 = \sum_K \sqrt{|\psi_h|_{1,K}^2 + \|\nabla \cdot \mathbf{u}_h\|_{0,K}^2} \sqrt{|\omega^m - \omega_h^m|_{1,K}^2 + \|\omega^c - \omega_h^c\|_{0,K}^2}$$

where $|\omega^m - \omega_h^m|_{1,K}$ and $\|\omega^c - \omega_h^c\|_{0,K}$ are approximated using the global error estimator. We also note that $\psi_h \in \mathbf{W}^h$, $\mathbf{q} = (1, 1)$, is an approximation of $\varphi \in \mathbf{V}$, that is, of the residual \mathcal{R}_h^m . Finally the last approximation, η_3 , is computed as

$$\eta_3 = \sum_K \sqrt{|\psi_h|_{1,K}^2 + \|\nabla \cdot \mathbf{u}_h\|_{0,K}^2} \sqrt{|\tilde{\omega}_h^m - \omega_h^m|_{1,K}^2 + \|\tilde{\omega}_h^c - \omega_h^c\|_{0,K}^2}$$

where $\tilde{\omega}_h^m$ and $\tilde{\omega}_h^c$ have been defined earlier. The results are displayed in Table 5. The elementwise distribution of η_2 and η_3 are shown in Figures 8-9 and Figures 10-11 for element 1 and element 55 respectively.

Table 5

Estimates η_1 , η_2 and η_3 of the error quantity $L(e, E)$ for elements 1 and 55.

Element	$L(\mathbf{u}, p)$	$L(e, E)$	η_1	η_2	η_3
1	-0.000364	0.000088	0.000081	0.003304	0.015205
55	1.6699	0.0258	0.0271	0.3733	0.4783

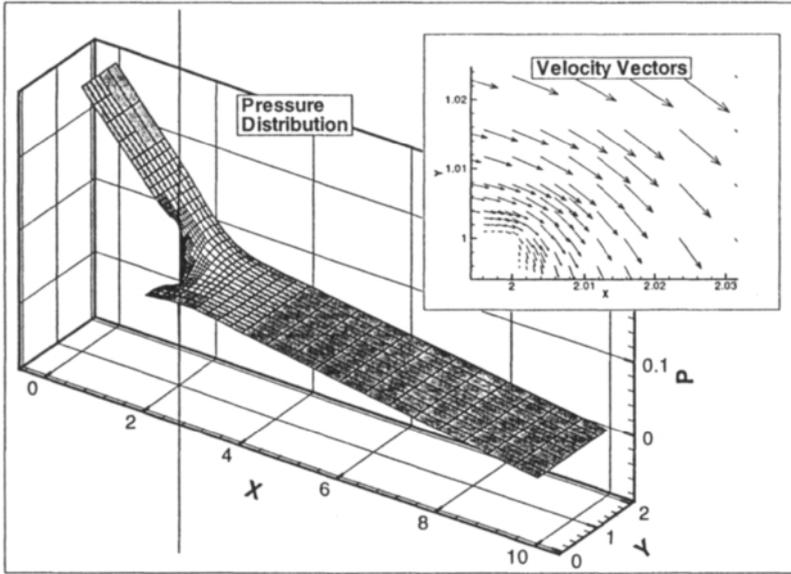


Figure 7. Velocity and pressure on adapted mesh with relative estimated error of less than 1.2%.

The conclusions of this preliminary study are twofold. First, we observe that only the quantity η_1 represents an accurate approximation of $L(e, E)$. The other two quantities η_2 and η_3 provide very pessimistic estimates as expected, but provide local refinement indicators as they reflect the intensity of the elementwise contribution to the error quantity $L(e, E)$. Secondly, we remark that the distribution η_2 , relative to the case involving element 1 (see Figures 8 and 9) is very different from the one of η_3 . This can be explained by the fact that the quantities $|\omega^m - \omega_h^m|_{1,K}$ and $\|\omega^c - \omega_h^c\|_{0,K}$ are approximated using a global error estimator, which does not take in account the propagation of the errors away from their sources. On the other hand, the discrepancy is less noticeable for element 55, because the region of large numerical errors in the influence functions superpose with the regions of large sources of errors \mathcal{R}_h^m and \mathcal{R}_h^c .

ACKNOWLEDGEMENT The support of this work by the Office of Naval Research under contract N00014-95-1-0401 is gratefully acknowledged.

REFERENCES

1. M. AINSWORTH AND J. T. ODEN, *A unified approach to a posteriori error estimation using element residual methods*, Numer. Math., 65 (1993), pp. 23–50.
2. —, *A posteriori error estimation in finite element analysis*, Comp. Meth. in Appl. Mech. and Eng., 142 (1997), pp. 1–88.

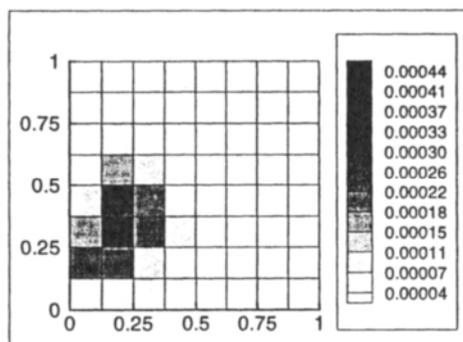


Figure 8. Elementwise contributions of η_2 for element 1.

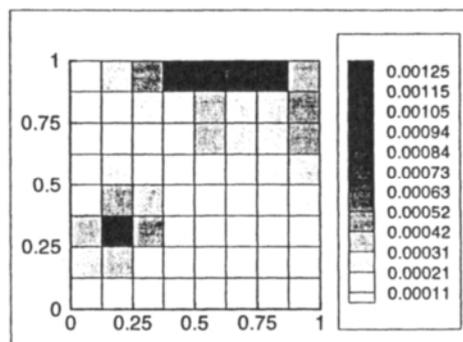


Figure 9. Elementwise contributions of η_3 for element 1.

3. I. BABUŠKA AND W. RHEINBOLDT, *A posteriori error estimates for the finite element method*, Int. J. for Num. Meth. in Eng., 12 (1978), pp. 1597–1615.
4. I. BABUŠKA, T. STROUBOULIS, C. S. UPADHYAY, AND S. K. GANGARAJ, *A posteriori estimation and adaptive control of the pollution error in the h-version of the finite element method*, Int. J. for Num. Meth. in Eng., 38 (1995), pp. 4207–4235.
5. R. E. BANK, *Hierarchical bases and the finite element method*. To Appear in Acta Numerica.
6. R. E. BANK AND R. K. SMITH, *A posteriori error estimates based on hierarchical bases*, SIAM J. Numer. Anal., 30 (1993), pp. 921–935.
7. R. E. BANK AND A. WEISER, *Some a posteriori error estimates for elliptic partial differential equations*, Math. Comp., 44 (1985), pp. 283–301.
8. R. E. BANK AND B. D. WELFERT, *A posteriori error estimates for the Stokes problem*, SIAM J. Numer. Anal., 28 (1991), pp. 591–623.
9. R. BECKER AND R. RANNACHER, *A feedback approach to error control in finite elements methods: Basic analysis and examples*, Preprint 96-52, Institut für Angewandte Mathematik, Universität Heidelberg, (1996).
10. —, *Weighted a posteriori error control in FE method*, in ENUMATH-95, Paris, Sept. 1995.
11. L. DEMKOWICZ, J. T. ODEN, AND T. STROUBOULIS, *Adaptive finite elements for flow problems with moving boundaries. Part 1: Variational principles and a posteriori error estimates*, Comp. Meth. in Appl. Mech. and Eng., 46 (1984), pp. 217–251.
12. V. GIRAULT AND P.-A. RAVIART, *Finite Element Methods for Navier-Stokes Equations*, Springer-Verlag, 1986.
13. P. LADEVÈZE AND D. LEGUILLON, *Error estimate procedure in the finite element method and applications*, SIAM J. Numer. Anal., 20 (1983), pp. 485–509.
14. J. T. ODEN, L. DEMKOWICZ, W. RACHOWICZ, AND O. HARDY, *Toward a universal h-p adaptive finite element strategy. Part 1: Constrained approximation and data structure*, Comp. Meth. in Appl. Mech. and Eng., 77 (1989), pp. 113–180.

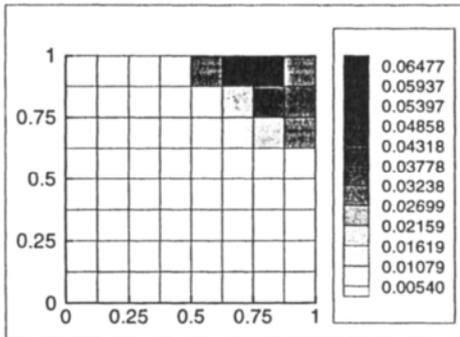


Figure 10. Elementwise contributions of η_2 for element 55.

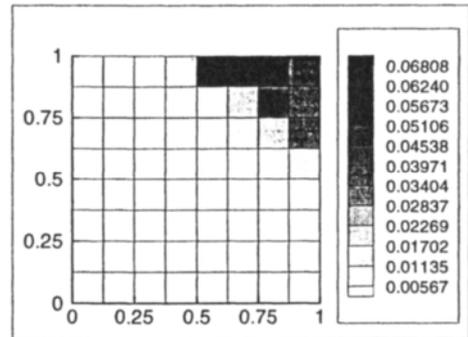


Figure 11. Elementwise contributions of η_3 for element 55.

15. J. T. ODEN AND Y. FENG, *Local and pollution error estimation for finite element approximations of elliptic boundary value problems*, J. Comput. Appl. Math., 74 (1996), pp. 245–293.
16. J. T. ODEN, W. WU, AND M. AINSWORTH, *An a posteriori error estimate for finite element approximations of the Navier-Stokes equations*, Comp. Meth. in Appl. Mech. and Eng., 111 (1993), pp. 185–202.
17. B. SZABÓ AND I. BABUŠKA, *Finite Element Analysis*, John Wiley and Sons, 1991.
18. R. VERFÜRTH, *A posteriori error estimators for the Stokes equations*, Numer. Math., 55 (1989), pp. 309–325.
19. R. VERFÜRTH, *A Review of A Posteriori Error Estimation and Adaptive Mesh-refinement Techniques*, Wiley-Teubner, 1996.

This Page Intentionally Left Blank

A mathematical framework for the P. Ladevèze a posteriori error bounds in finite element methods

Ph. Destuynder, M. Collot and M. Salaün

CNAM-IAT, 15 rue Marat, 78210 Saint-Cyr l'École

Using the Poisson model as an example, we analyze the a posteriori error bound suggested by P. Ladevèze in his thesis. The method is applied to a non-conforming finite element method which seems to be particularly well adapted to the ideas of P. Ladevèze. Few numerical results are given. Finally a comparison with a conforming method is mentioned.

1. INTRODUCTION. Let Ω be an open set in \mathbb{R}^2 with a smooth enough boundary denoted by Γ . We consider the Poisson model defined by :

$$\left\{ \begin{array}{l} \text{find } u \in H_0^1(\Omega) \text{ such that :} \\ -\Delta u = f \quad \text{in } \Omega, \end{array} \right. \quad (1)$$

where f is a given function in the space $L^2(\Omega)$ (at least). Let \mathcal{T}^h be a triangulation of Ω which is assumed to belong to a uniformly regular family (see P.G. Ciarlet [1978]). Then we define an approximation space of $H_0^1(\Omega)$ by :

$$V^h = \left\{ v \in H_0^1(\Omega), \forall K \in \mathcal{T}^h, v|_K \in P_1(K) \right\} \quad (2)$$

where $P_1(K)$ denotes the set of polynomials the degree of which are less or equal to one. The classical finite element approximation of the solution u of (1) consists in finding $u^h \in V^h$ such that :

$$\forall v \in V^h, \quad \int_{\Omega} \nabla u^h \cdot \nabla v = \int_{\Omega} f v. \quad (3)$$

The existence and uniqueness of u^h is very classical and one has the well-known asymptotic error bounds (see G. Strang - G.J. Fix [1973]):

$$|u - u^h|_{1,\Omega} \leq ch |u|_{2,\Omega} \quad (4)$$

where c is a constant which is independent of both h and u and $| \cdot |_{m,2,\Omega}$ denotes the semi-norm of the higher order derivatives in the space $H^m(\Omega)$. Obviously a regularity assumption is necessary in order to make sense to (4).

Unfortunately, (4) is neither an error bound nor an error indicator. Hence several publications have suggested a posteriori estimates. Let us point out the pioneer work of P. Ladevèze [1975] who used the Prager-Synge relationship. Another way was opened by I. Babuska - W.C. Rheinbolt [1978] using estimates in dual spaces. This last method was revisited and more conveniently formulated in view of applications (practical computations) by R. Verfürth C. Bernardi - B. Métivet - R. Verfürth [1992]. More recently B. Métivet extended it to Stokes (and Navier-Stokes) problem. Up to now this last version seems to be one the most well founded and useful for mesh refinement algorithm. But its main drawbacks is that it only gives an error indicator, not an error bound.

On the contrary, the method of P. Ladevèze [1975] which has been widely studied by him and his co-workers (see P. Ladevèze - J.P. Pelle - Ph. Rougeot [1991]), gives an explicit error bound. But the way to use of the method is not unique and several choices must be made. Furthermore few difficulties (from the theoretical point of view) arise for the elasticity model.

Our goal in this paper is to revisit the Ladevèze's method on a simple equation and to focus on a particular case for which this method is almost perfect. It corresponds to a non conforming finite element formulation. Then we shortly compare the results to a conforming case and to the error indicator strategy of C. Bernardi - B. Métivet and R. Verfürth. Let us also point out that M. Ainsworth and J.T. Oden [1993] have suggested a mathematical analysis of Ladevèze's method based on an hybrid-primal variational formulation.

2. A NON CONFORMING APPROXIMATION OF THE POISSON MODEL.

Let us define for each triangulation T^h of the open set Ω , a finite dimensional space by:

$$X^h = \left\{ v \mid \forall K \in \mathcal{T}^h, v|_K \in P_1(K); \forall \gamma \in \mathcal{A}^h, \int_{\gamma} [v] = 0 \right\}, \quad (5)$$

where \mathcal{A}^h is the set of all the sides of \mathcal{T}^h and the symbol $[\bullet]$ denotes the jump of a function across a side γ . If γ is on the boundary of Ω , then the outside value of v is zero. Then we introduce the following bilinear form on X^h by :

$$\forall u, v \in X^h, \quad a^h(u, v) = \sum_{K \in \mathcal{T}^h} \int_K \nabla u^h \cdot \nabla v^h$$

and we define the approximate model of (1) by :

$$\begin{cases} \text{find } u^h \in X^h, \\ \forall v \in X^h, \quad a^h(u^h, v) = \int_{\Omega} f v. \end{cases} \quad (6)$$

The existence and uniqueness of a solution to (6) is classical and furthermore, there is an asymptotic error bound between u and u^h (for details, see Ciarlet [1978]). Setting

$$\|v\| = \sqrt{a^h(v, v)}, \quad (7)$$

one has :

$$\|u - u^h\| \leq ch |u|_{2, \Omega}$$

where c is a constant which is independent of both h and u . An interesting property of the approximate solution u^h of (6) is formulated in the next result.

THEOREM 1. *Let u^h be the solution of (6) and γ an arbitrary side inside the mesh \mathcal{T}^h . Then if K_1 and K_2 are the two elements adjacent to γ , then one has :*

$$\frac{\partial u^h}{\partial v_1} + \frac{\partial u^h}{\partial v_2} = \frac{1}{|\gamma|} \left[\int_{K_1} f \lambda_\gamma + \int_{K_2} f \lambda_\gamma \right]$$

where $\frac{\partial u^h}{\partial v_1}$ (respectively $\frac{\partial u^h}{\partial v_2}$) denotes the normal derivative of $u^h|_{K_1}$ (respectively $u^h|_{K_2}$) in the direction outwards K_1 (respectively K_2). Finally, λ_γ is the function of X^h equal to one at the middle of γ and to 0 at all the other middles of sides. and $|\gamma|$ is the measure of γ

The proof is very simple and obtained from (6) by choosing $v = \lambda_\gamma$. It is worth noticing that $u^h \notin H_0^1(\Omega)$. Hence the method is called non-conforming. In order to furnish a conforming solution from u^h we define a local projection – say u^{hh} – of u^h onto the conforming space V^h defined by :

$$V^h = \left\{ v \mid \forall K \in \mathcal{T}^h \quad v|_K \in P_1(K), \quad v \in H_0^1(\Omega) \right\}. \quad (8)$$

It is characterized by the values of u^{hh} at the vertices of \mathcal{T}^h . One has:

$\forall S$, inner vertex of \mathcal{T}^h :

$$u^{hh}(S) = \frac{\sum_{K \in \mathcal{C}_i^h} u|_K^h(S) |K|}{\sum_{K \in \mathcal{C}_i^h} |K|} \quad (9)$$

where \mathcal{C}_i^h is the set (cluster) of all the elements which have S as a vertex and $|K|$ is the measure of the triangle K . We proved in Ph. Destuynder - B. Métivet ([1996] a), that there exists a constant c such that :

$$\|u - u^{hh}\| = |u - u^{hh}|_{1,\Omega} \leq ch |u|_{2,\Omega} . \quad (10)$$

3. THE PRAGER-SYNGE RELATIONSHIP [1947] AND THE ERROR BOUND.

Let us introduce the functional set for the dual problem associated to (1) :

$$H_f(\text{div}, \Omega) = \left\{ p \in \left(L^2(\Omega) \right)^2, \text{div } p + f = 0 \text{ in } \Omega \right\} .$$

Then one has the following basic result which is a particular version of Pythagore's Theorem.

THEOREM 2 (Prager-Syngé). *Let u be the solution of the Poisson model (1). Then :*

$$\forall p \in H_f(\text{div}, \Omega) ; \forall v \in H_0^1(\Omega),$$

$$|u - v|_{1,\Omega}^2 + \|p - \nabla u\|_{0,\Omega}^2 = \|p - \nabla v\|_{0,\Omega}^2$$

A consequence of this Theorem is the inequality used by P. Ladevèze and called "constitutive error bound" by this author :

$$\forall p \in H_f(\text{div}, \Omega) ; \forall v \in H_0^1(\Omega), |u - v|_{1,\Omega} \leq \|p - \nabla v\|_{0,\Omega} .$$

Let us apply this inequality to the non-conforming finite element described previously. We set :

$$v = u^{hh} \text{ (the projection of } u^h \text{ on } V^h \text{ defined in (9)).}$$

Then :

$$\forall p \in H_f(\text{div}, \Omega) \quad \|u - u^{hh}\|_{1, \Omega} \leq \|p - \nabla u^{hh}\|_{0, \Omega}. \quad (11)$$

The definition of a particular element p is based on the particular property satisfied by u^h (non-conforming solution) and which has been formulated in Theorem 1. On each triangle K of the mesh T^h , we set :

$$p \cdot v = \frac{\partial u^h}{\partial v} - \frac{1}{|\gamma|} \int_K f \lambda_\gamma \quad \text{on } \gamma \subset \partial K \quad (12)$$

where v is the outwards unit normal to γ and λ_γ the linear function equal to one at the middle of γ and zero at the middle of the two other sides of K . Furthermore p must satisfied in K :

$$\text{div } p + f = 0. \quad (13)$$

The solution of (12) can be obtained as follows. We look for a particular solution of (12) - say p^h - using Raviart-Thomas element (see P.A. Raviart - J.M. Thomas [1975]). Then p^h is such that on each element K of \mathcal{T}^h .

$$p^h = \begin{cases} a_K + b_K x_1 \\ c_K + b_K x_2 \end{cases}$$

and satisfies (the 3×3 linear system is well-posed) :

$$p^h \cdot v = \frac{\partial u^h}{\partial v} - \frac{1}{|\gamma|} \int_K f \lambda_\gamma \quad \forall \gamma \subset \partial K. \quad (14)$$

Then from $\sum_{\gamma \subset \partial K} \lambda_\gamma = 1$:

$$\int_{\partial K} p^h \cdot v = \sum_{\gamma \subset \partial K} \int_\gamma p^h \cdot v = \int_{\partial K} \frac{\partial u^h}{\partial v} - \int_K f = - \int_K f \quad (15)$$

and therefore ($\text{div } p^h = 2 b_K = \text{constant in } K$) :

$$\text{div } p^h = - \frac{1}{|K|} \int_K f = 0. \quad (16)$$

In order to define p so that (13) would be satisfied, we introduce a local term denoted by r and solution of the following Neumann model.

$$\begin{cases} -\Delta r_K = f - \frac{1}{|K|} \int_K f & \text{in } K \\ \frac{\partial r_K}{\partial \nu} = 0 & \text{on } \partial K \\ \int_K r_K = 0. \end{cases} \quad (17)$$

The existence and uniqueness of r_K is classical. Furthermore we can prove that ∇r_K is negligible if f is smooth enough. Then setting on each triangle K of \mathcal{T}^h :

$$p = p^h + \nabla r_K,$$

we define an element p which satisfies both the relations (12) and (13). We proved in (Ph. Destuynder - B. Métivet ([1996] a)) that there exists a constant independent of both h and u such that :

$$\|p - \nabla u^{hh}\|_{0,\Omega} \leq c h \|u\|_{2,\Omega} \quad (18)$$

which ensures that $\|p - \nabla u^{hh}\|_{0,\Omega}$ is a coherent error bound.

Remark. The definition of p is not unique. Let us introduce a new term on each K of \mathcal{T}^h satisfying.

$$\begin{cases} q \cdot \nu = 0 & \text{on } \partial K, \\ \operatorname{div} q = 0 & \text{in } K. \end{cases} \quad (19)$$

Such an element is characterized as the rotational of a function defined up to an additive constant (in K) – say Ψ_K –. The boundary condition on ∂K implies that Ψ_K can be chosen such that.

$$q = \operatorname{rot} \Psi_K, \quad \Psi_K \in H_0^1(K).$$

Hence the best choice for Ψ_K is the one which minimizes the error bound (the term Ψ_K is local) :

$$\min_{\Psi_K \in H_0^1(K)} \left\| p + \operatorname{rot} \Psi_K - \nabla u^{hh} \right\|_{0,K}^2. \quad (20)$$

Therefore (because $\operatorname{rot}(p^h + \nabla u^{hh})$) :

$$-\Delta \Psi_K = 0 \quad \Psi_K \in H_0^1(K)$$

which implies :

$$\Psi_K = 0 .$$

Finally we prove that p defined previously is the best choice satisfying (12) and (13). The result would be slightly different for higher order element as we show it in Ph. Destuynder [1997] .

4. THE NUMERICAL TESTS FOR A NON-CONFORMING F.E.M.

In order to compare the exact error and the error bound suggested in this paper, we have considered a simple square open set and a constant function f . The side of the square being L , the analytical solution of (1) is given by the sum of a series :

$$u = \sum_{n,m \geq 1} A_{nm} \sin\left(\frac{n\pi x_1}{L}\right) \sin\left(\frac{m\pi x_2}{L}\right)$$

where :

$$A_{nm} = \frac{4 \int_0^L \int_0^L f(x_1, x_2) \sin\left(\frac{n\pi x_1}{L}\right) \sin\left(\frac{m\pi x_2}{L}\right)}{\pi^2 (n^2 + m^2)} = \frac{4 f (1 - (-1)^n) (1 - (-1)^m)}{\pi^2 (n^2 + m^2)} .$$

Two kind of grids have been used. One is regular and shown on figure 1 (seven sizes of meshes were used). The second one is obtained by a mesh generator and one of them is represented on figure 2. Again seven different sizes were used. Then we have represented on figures 3 and 4 the three following quantities.

1		$\ p - \nabla u^{hh}\ _{0,\Omega}$	versus h (7 values)
2		$ u - u^{hh} _{1,\Omega}$	versus h (7 values)
3		$\ u - u^h\ $	versus h (7 values)

Let us make some comments on the obtained convergence curves.

- 1) The slopes of the three curves for any kind of grids is 1. This is in agreement with the theoretical results.
- 2) The relative positions of curves **1** and **2** are in accordance with the error bound (11). Furthermore the ratio between **1** and **2** is constant equal to: 1.3 (1 is the lower bound which can not be reached) for the regular meshes and to: 2 for an arbitrary mesh.

- 3) The projection of u^h onto a conforming subspace leads to a slight improvement for a regular mesh.
- 4) Because of Prager-Synge equality the numerical results given on figures 3 and 4 indicate that the error between p and ∇u is smaller than the one between ∇u and ∇u^{hh} (in $L^2(\Omega)$ - norm). The ratio is about .83.

Figure 1. Example of a regular mesh

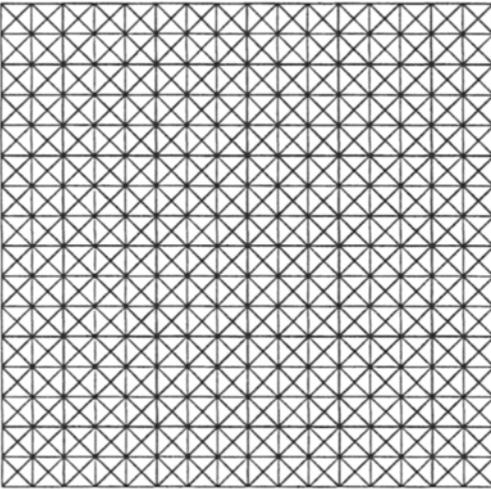


Figure 2. Example of a mesh obtained by

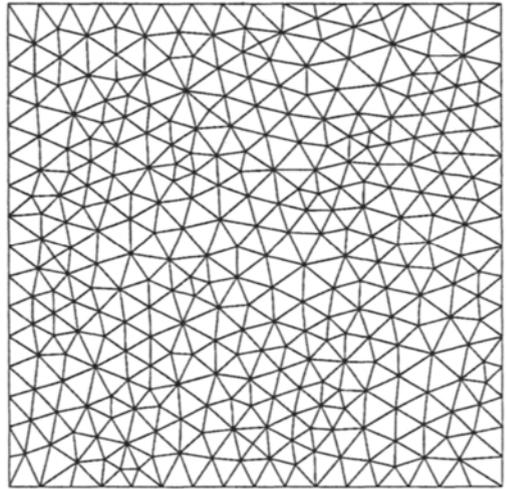


Figure 3. The error bound and the exact error for a regular mesh versus the mesh size

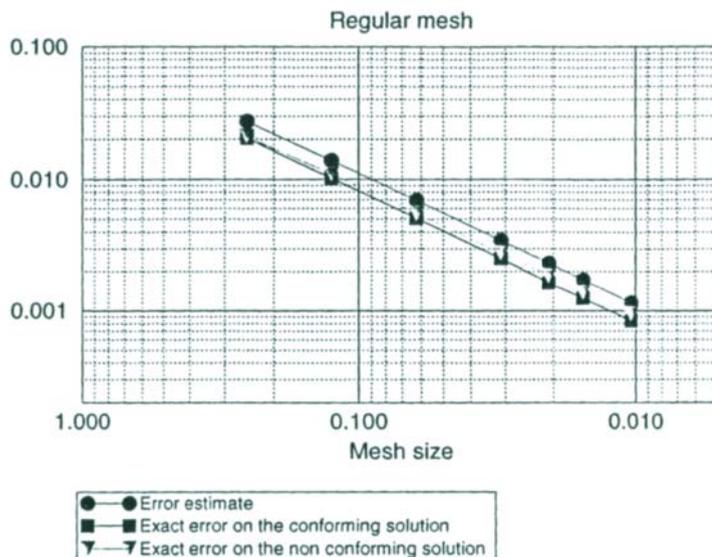
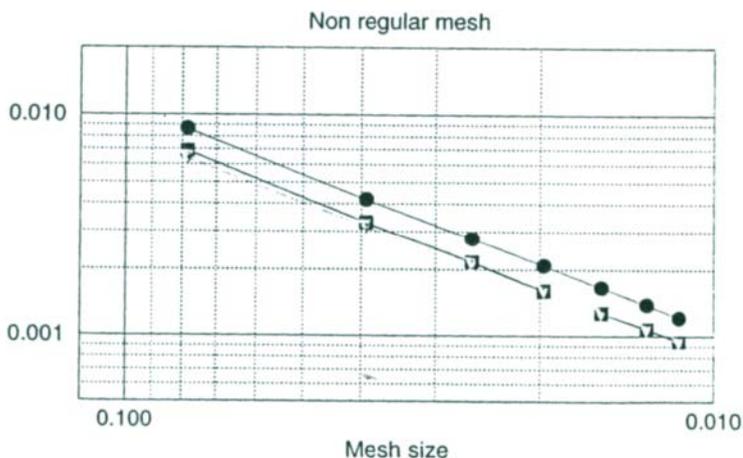


Figure 4. The error bound and the exact error for an irregular mesh versus the mesh size.



5. THE CONFORMING APPROXIMATION

Let us now consider that u^h is solution of the classical conforming finite element method defined at (3). In order to apply the Prager-Synge relationship, one has to construct a vector field p lying in the set $H_f(\text{div}, \Omega)$. Here again the Raviart-Thomas finite element can be used. A possibility is the following. At each node of the mesh T^h – say S_i – we define (following the Ladevèze's idea) a cluster of elements – say C_i^h – which is the union of elements having S_i as a vertex.

Then for each S_i (even on the boundary of Ω), we set (λ_i is the continuous piecewise linear function, equal to one at S_i and 0 at all the other nodes).

$$\begin{cases} p_i^h \in \text{RT1}(\mathcal{C}_i^h), \\ \text{div } p_i^h = \frac{1}{|K|} \left\{ \int_K \nabla u^h \cdot \nabla \lambda_i - \int_K f \lambda_i \right\}, \\ \forall K \in \mathcal{C}_i^h \end{cases} \quad (21)$$

where:

$$\text{RT1}(\mathcal{C}_i^h) = \left\{ p \in H(\text{div}, \mathcal{C}_i^h), p \cdot \nu = 0 \text{ on } \partial \mathcal{C}_i^h \text{ and } \forall K \in \mathcal{C}_i^h p|_K = \begin{Bmatrix} a_K + b_K x_1 \\ c_K + b_K x_2 \end{Bmatrix} \right\}.$$

The existence of a solution to (21) has been proved in Ph. Destuynder - B. Métivet ([1996] b). But the solution is not unique. The general solution is :

$$p_i^h = \tilde{p}_i^h + \alpha_i \text{rot } \lambda_i \quad (22)$$

where \tilde{p}_i^h is a particular solution of (21) and α_i an arbitrary constant. Then we set :

$$p^h = \sum_{i \in S^h} p_i^h = \sum_{i \in S^h} \tilde{p}_i^h + \sum_{i \in S^h} \alpha_i \text{rot } \lambda_i \quad (23)$$

S^h denoting the set of all the nodes of T^h . As in the non-conforming case the coefficients α_i should be chosen in order to minimize the error bound (assuming that $\text{div } p^h + f = 0$ or else that f is piecewise constant) :

$$\alpha \in \mathbb{R}^L \longrightarrow \left\| p^h + \sum_{i \in S^h} \alpha_i \text{rot } \lambda_i - \nabla u^h \right\|_{0,\Omega} \quad (24)$$

where: $L = \text{card}(S^h)$.

Two strategies can be discussed. One consists in replacing (24) by a local minimization (one iteration of Jacobi algorithm). The second one consists in adding one iteration of SSOR algorithm. These two strategies have been checked on the test model presented in section 5. One can observe that the second one is more reliable for irregular mesh (see figures 5 and 6). As the additional cost is neglectible, it has to be recommended for general applications.

Figure 5. The error bounds, the exact errors and the error indicators for regular meshes.

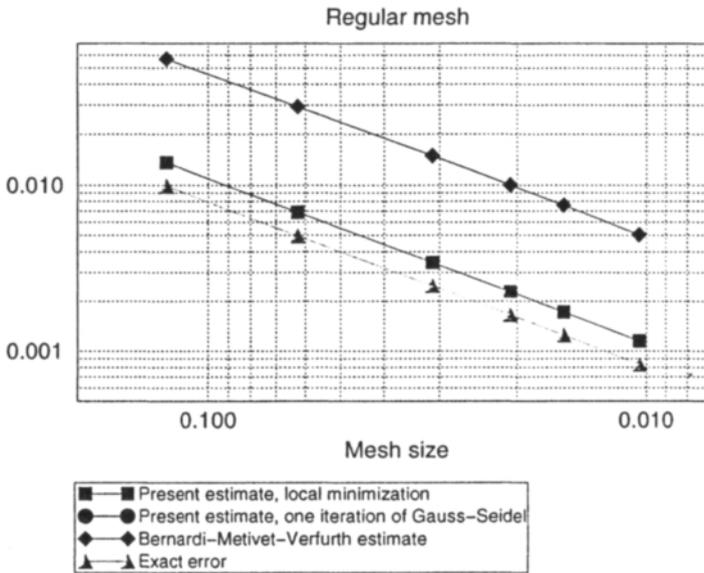
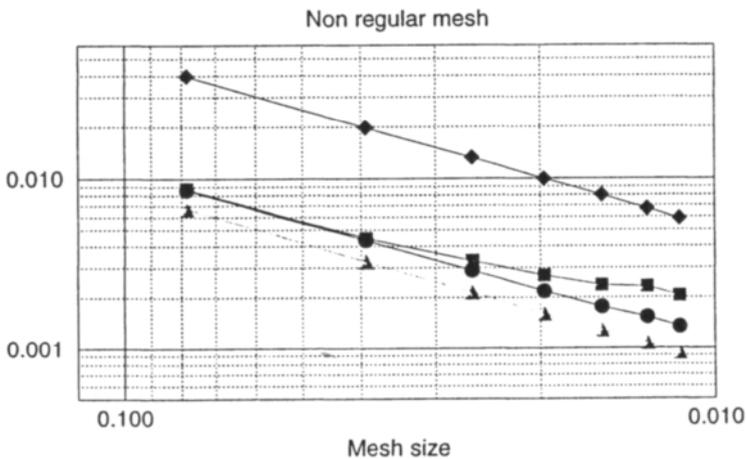


Figure 6. The error bounds, the exact errors and the error indicators for irregular meshes.



Remark. As a matter of fact one only has :

$$\operatorname{div} p^h = -\frac{1}{|K|} \int_K f.$$

Hence it is necessary to add a local term in order to satisfy exactly the divergence condition. But if f is smooth enough, this term is neglectible (the discussion is the same as in

the non-conforming case) with another respect, we proved in Ph. Destuynder - B. Métivet ([1996] b) that the error bound $\|p^h - \nabla u^h\|_{0,\Omega}$ is $O(h)$.

In order to compare the methods of P. Ladevèze up-dated as mentioned previously and the error indicator strategy of C. Bernardi - B. Métivet - R. Verfürth [1992], we have plotted this quantity (denoted by B^h) on the figures 5 and 6. Let us recalled that it is defined by :

$$B^h = \left[\sum_{K \in \mathcal{T}^h} \left\{ h_K \|f + \Delta u^h\|_{0,K} + \frac{1}{2} \sum_{\gamma \subset \partial K} |\gamma|^{1/2} \left\| \left[\frac{\partial u^h}{\partial \nu} \right] \right\|_{0,\gamma} \right\}^2 \right]^{1/2} \quad (25)$$

One can see, that the results are similar to those of the non-conforming model, but the error is larger for the same mesh. The B^h indicator is larger (ratio of 6 with the exact error instead of 1.3 for the Ladevèze based method). Another huge advantage of the error bound suggested by P. Ladevèze is that it is possible to improve the approximation of u by a local minimization problem. For instance by adding degrees of freedom on the sides between elements (second degree polynomials) (see Ph. Destuynder - M. Collot - M. Salaün [1997]).

6. CONCLUSION

The explicit error bound based on the idea of P. Ladevèze, coupled with the use of mixed finite element method, lead to very accurate results which are compatible with the theoretical results known in finite element methods. The extension to elasticity is also possible but the $H(\text{div}, \Omega)$ finite element are then more complex (21 d.o.f. in 2D). This finite element has been developed by D.N. Arnold- J. Douglas and C.P.Gupta [1984].

7. REFERENCES

- [1] M. Ainsworth - J.T. Oden [1993]. A unified approach to a posteriori error estimation using element residual methods. Numer. Math., 65, pp 23-50 .
- [2] D. Arnold, J. Douglas, C. P. Gupta [1984]. A family of higher order mixed finite element methods for plane elasticity. Numer. Math. ,45, pp 1-22.
- [3] I. Babuska - W.C. Rheinbolt [1978]. Error estimates for adaptive finite element computation . SIAM J. Numer. Anal., 15, n° 4,pp. 736-754.

- [4] C. Bernardi - B. Métivet - R. Verfürth [1993]. Analyse Numérique d'indicateurs d'erreur, note EDF n° HI72/93062.
- [5] P.G. Ciarlet [1978]. The finite element method for elliptic problems, North Holland.
- [6] Ph. Destuynder - B. Métivet ([1996] a). Explicit error bounds for a non-conforming finite element method. Internal report I.A.T. CNAM n°261/96 to appear in SIAM. J. Numer. Anal..
- [7] Ph. Destuynder - B. Métivet ([1996] b). Explicit error bounds in a conforming finite element method ; note IAT CNAM n° 265/96.
- [8] Ph. Destuynder - M. Collot - M. Salaün [1997]. On the optimality of error bounds in finite element method ; to appear.
- [9] Ph. Destuynder [1997]. Estimations d'erreur explicites pour des méthodes d'éléments finis quadratiques. Note interne I.A.T.-C.N.A.M. n° 290/97.
- [10] L. Gallimard - P. Ladevèze - J.P. Pelle [1995]. La méthode des erreurs en relation de comportement appliquée au contrôle des calculs non linéaires. École CEA - EDF - INRIA.
- [11] P. Ladevèze - J.P. Pelle - Ph. Rougeot [1991]. Error estimation and mesh optimization for classical finite elements. Eng. Comp., Vol. 8, pp 69-80.
- [12] P. Ladevèze [1975]. Comparaison de modèles de mécanique des milieux continus. Thèse d'état ; Université Paris 6.
- [13] P. Ladevèze - D. Leguillon [1983]. Error estimate procedure in the finite element method and applications. SIAM J. Numer. Anal. Vol. 20, n° 5, pp 485-504.
- [14] W. Prager - J.L. Synge [1949]. Approximations in elasticity based on the concept of function space. Quarterly of applied mathematics vol. V, n° 3, p. 241-269.
- [15] P.A. Raviart - J.M. Thomas [1975]. A mixed finite element method for second order elliptic problems. In the Proceeding of the symposiums on the mathematical aspects of the finite element methods ; Rome.

- [16] G. Strang - G.J. Fix [1973]. An analysis of the finite element method. Prentice Hall, series in automatic computation ; London.
- [17] R. Verfürth [1994]. A posteriori error estimation and adaptive mesh refinement technique. J. Comput. Appl. Math.,50, pp 67-83.

PART 2

MODELLING ERROR ESTIMATORS AND ADAPTIVE MODELLING STRATEGIES

This Page Intentionally Left Blank

Adaptive finite elements in elastoplasticity with mechanical error indicators and Neumann-type estimators

E. Stein, F.J. Barthold, S. Ohnibus and M. Schmidt

Institute for Structural Mechanics and Computational Mechanics (IBNM),
University of Hannover, Germany

Abstract

Many interesting tasks in technology need the solution of complex boundary value problems modeled by the mathematical theory of elasticity and elastoplasticity. Error controlled adaptive strategies should be used in order to achieve a prescribed accuracy of the computed solutions at minimum cost.

In this paper, locally computed residual error indicators in the primal form of the finite-element-method for Elasticity as well as Hencky- and Prandtl-Reuß-plasticity without and with nonlinear hardening are presented, controlling global errors of equilibrium, plastic strain rates, the yield condition and the numerical integration of the flow rule. Furthermore, an error estimator based on local Neumann problems is extended to elastoplasticity, based on improved boundary tractions. This recovery technique is called PEM (Posterior Equilibrium Method).

1. INTRODUCTION

In the adaptive Finite-Element process for elastoplastic deformations, aside from the step-wise optimal choice of the spatial discretization we have to pay attention to the implicit time (load) dependency of the deformation process.

As the usual decomposition of FEM in space and FDM in time (method of lines) does not permit a rigorous coupled a posteriori error analysis in space and time, an heuristic assumption based on the comparison of Prandtl-Reuß elastoplasticity with the so-called Hencky plasticity, (nonlinear elasticity with a yield limit), is used to split the total error into spatial and time discretization errors [6, 8].

In this paper three separate local spatial error indicators of the primal FEM are derived from global errors, namely for equilibrium, direction of plastic strains (using the plastic dissipation) and the yield condition. Each of them yields a scaled local indicator for mesh refinement. The error within the integration of the flow rule by the backward-Euler scheme is estimated by the maximal change of the normal (with respect to the yield function) between two time (load) steps.

In section 2, the basic equations for elastoplastic deformations with isotropic hardening are listed in a canonical form. Section 3 treats the differences between global and local error estimations. Section 4 describes the benchmark example used for the numerical tests. Section 5 presents numerical results of error indicators and refined meshes for all relevant errors. Section 6 gives an overview of the posterior equilibrium method (PEM) with anisotropic error estimation for elastic deformations and the extension to elastoplasticity, [7, 22, 11, 23]. In the last section, some numerical results are given, showing the efficiency of the adaptive process of computation.

2. THE BASIC EQUATIONS OF CLASSICAL ELASTOPLASTICITY

Consider the solid body $\Omega \subset \mathbb{R}^3$ with boundary $\partial\Omega = \Gamma = \Gamma_u \cup \Gamma_t$. The variational form reads

$$\int_{\Omega} \varepsilon(v(x)) : \sigma(u(x)) d\Omega = \int_{\Omega} v(x) f d\Omega + \int_{\Gamma_t} v(x) \bar{t} d\Gamma \quad (1)$$

with test functions $v \in V$; $V = \{v(x) \in [H_0^1(\Omega)]^3, \Omega \subset \mathbb{R}^3, v = 0 \text{ on } \Gamma_u\}$ and mixed boundary conditions for the displacements $u = 0, v = 0$ on Γ_u and given surface tractions $t = \sigma \cdot n = \bar{t}$ on Γ_t . Herein, \bar{t} are prescribed traction forces and f are body forces. σ and ε denote the macro stress tensor and macro strain tensor, respectively.

Among the theories of elastoplastic deforming materials we consider the deformation theory or Hencky-plasticity [13] and the flow theory by Prandtl-Reuß [17]. The latter has the advantages that it can not only describe changes of the material parameters during the loading process but also ‘plastic’ unloading with resulting plastic deformations because of the existence of an ‘internal time’ which is represented by means of a loading history. Consequently, Hencky-plasticity is often referred to as ‘non-linear elasticity’ with a yield condition.

Prandtl-Reuß elastoplasticity for small strains with nonlinear isotropic hardening can be described in the following canonical form, Table. 1:

Elastic strains within linear kinematics	$\varepsilon^e := \varepsilon - \varepsilon^p$
Internal variable for isotropic hardening (micro strain)	α
Free energy function	$\Psi = \hat{\Psi}(\varepsilon^e, \alpha)$
Macro stress	$\sigma = \partial_{\varepsilon^e} \hat{\Psi}$
Micro stress	$\beta = -\partial_{\alpha} \hat{\Psi}$
Plastic dissipation	$\mathcal{D}^p = \sigma : \dot{\varepsilon}^p - \beta \cdot \dot{\alpha} \geq 0$
Yield function	$\Phi = \hat{\Phi}(\sigma, \beta)$
Flow rule	$\dot{\varepsilon}^p = \dot{\gamma} \partial_{\sigma} \hat{\Phi}$
Evolution of α	$\dot{\alpha} = -\dot{\gamma} \partial_{\beta} \hat{\Phi}$
Loading and unloading (Kuhn-Tucker) conditions	$\dot{\gamma} \geq 0, \hat{\Phi} \leq 0, \dot{\gamma} \cdot \hat{\Phi} = 0$

Table 1: The basic equations for Prandtl-Reuß elastoplasticity $\forall x \in \Omega \subset \mathbb{R}^3$

The free energy function $\Psi = \hat{\Psi}(\varepsilon^e, \alpha)$ can be split into a macroscopic part

$$\hat{\Psi}_{Macro}(\varepsilon^e) = \frac{1}{2} \kappa (\text{tr } \varepsilon^e)^2 + \mu \text{tr } (\varepsilon^e)^2 \quad (2)$$

with bulk modulus κ and shear modulus μ , and a microscopic part

$$\hat{\Psi}_{Micro}(\alpha) = \frac{1}{2} h \alpha^2 + (y_{\infty} - y_0) \left(\alpha + \frac{1}{\omega} (\exp(-\omega\alpha) - 1) \right) \quad (3)$$

which contains combined linear and exponential isotropic hardening with stress saturation.

The yield function $\Phi = \hat{\Phi}(\sigma, \beta)$ can also be split into a macro- and a micro-part

$$\Phi = \hat{\Phi}(\sigma, \beta) = \|\text{dev } \sigma\| - \sqrt{\frac{2}{3}} \hat{y}(\beta) \quad (4)$$

with the macroscopic stress tensor σ and the microscopic stress β . The saturation law for the stresses which combines linear and exponential hardening reads

$$\hat{y}(\beta) = y_0 + h\alpha + (y_\infty - y_0)(1 - \exp(-\omega\alpha)). \quad (5)$$

The material model has 6 parameters and the internal variable β . The plastic dissipation function for the macro- and the micro-terms

$$\mathcal{D}^p = \sigma : \dot{\epsilon}^p - \beta \cdot \dot{\alpha} \quad (6)$$

measures the rate of energy dissipation due to inelastic deformation.

3. GLOBAL OR LOCAL A POSTERIORI ERROR ESTIMATORS AND INDICATORS

Since the end of the seventies, a posteriori global estimates of relative errors for elastic problems are available, especially the energy norm of the integrated local residual of equilibrium which are evaluated locally.

The goal there is to obtain equidistributed error contributions throughout the system, i.e. to get the same desired accuracy of the approximated solution in each point of a system.

Recently, local error estimators and indicators were developed by e.g. Ainsworth [2], Fish and Markolefas [12], Babuška and Strouboulis [4], Rannacher and Suttmeier [18], and Ohnimus and Stein [21], which search for an optimal adaptive process with respect to highest possible accuracy of a local quantity or a function of it. The shortcomings of this strategy are evident for non-linear constitutive equations because, e.g., the position of maximum stress can vary considerably during the loading process. Therefore, the adaptive process would have to adapt to these conditions. Furthermore, the pollution error (the influence from outside onto the point in question) has to be computed as a global problem. On the other hand, it may be sufficient for complex systems with an easy load history to compute a solution of guaranteed accuracy at only some distinguished points.

As a consequence, the choice of the optimal strategy for error estimation and adaptivity depends on the goal of the computation and the complexity of the problem.

4. A NUMERICAL EXAMPLE

We chose a 'rectangular plate with a circular hole under plane strain constraint' and unidirectional monotonous loading for numerical tests of the different error sources and their indicators, Fig. 1a.

Three material parameters describe linear isotropic elastic and perfect plastic deformations, while mixed linear and nonlinear isotropic hardening with saturation need additionally three parameters, Table 2.

From a priori error considerations the adaptive computation of the above example starts with a sequence of graded meshes, ranging from 64 bilinear elements with 25 nodes to 8192 bilinear elements with 8385 nodes, Figs. 2b. Additionally, a reference solution with 24200 bilinear elements, 24642 nodes and 49062 d.o.f. was computed. In order to obtain comparable results for all meshes and material equations, some selected points were chosen

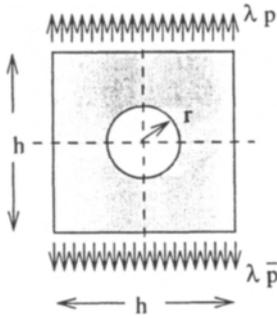


Figure 1a: Benchmark system: Stretched square plate with a hole in plane strain, $h=200$, $r=10$

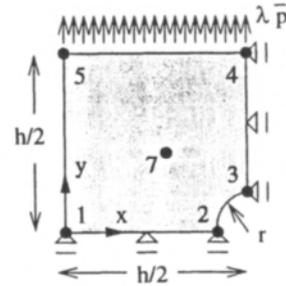


Figure 1b: Upper left quarter, using symmetry conditions, with point numbers of computed data

1. Young's modulus	$E = 206900.00 \text{ N/mm}^2$
2. Poisson's ratio	$\nu = 0.29$
3. Initial yield stress	$y_0 = 450.00 \text{ N/mm}^2$
4. Saturation stress	$y_{\infty} = 750.00 \text{ N/mm}^2$
5. Linear hardening modulus	$h = 129.00 \text{ N/mm}^2$
6. Hardening exponent	$\omega = 16.93$

Table 2: Material parameters

for data evaluation, Figs. 1a, 2b and Table 3. In the following sections, spatial distributions of the different error indicators are shown for a load factor of $\lambda = 4.5$ where the critical load for perfect plasticity is $\lambda_{crit} = 4.66$ and $\lambda_{crit} = 7.89$ for nonlinear hardening material. The load was applied in 18 load steps which renders the time discretization error negligible against the spatial discretization errors. All given results hold for hardening material, Table 2.



Figure 2a: v. Mises stresses at $\lambda = 4.5$ for hardening material

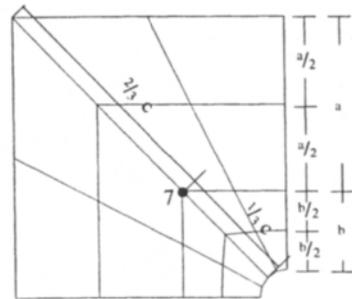


Figure 2b: Smallest mesh with 16 elements $a=61.953$, $b=28.047$, $c=131.421$

Physical quantity	node No.
Displacement \mathbf{u}_x	2
Displacement \mathbf{u}_y	4
Displacement \mathbf{u}_z	5
Stress σ_{xx}	7

Table 3: Selected points and data of system in Fig. 1a,b

5. GLOBAL ERRORS IN ELASTOPLASTICITY

5.1. Spatial discretization errors

As elastoplastic deformations are load-history dependent, the displacement fields $\mathbf{u}(\mathbf{x}, t)$ determine the actual stress and strains states for all points $\mathbf{x} \in \Omega$ at all implicit times $0 \leq t \leq T$. Thus, the spatial discretization error e_u measures the difference between the exact displacement field \mathbf{u} and its FE-approximation \mathbf{u}_h ; $\mathbf{e} = \mathbf{u} - \mathbf{u}_h$.

A physically meaningful error measurement is gained by considering the point-wise incremental total stress power $\dot{\mathcal{P}}$, where $\dot{\gamma} \geq 0$ assures the yield constraint $\Phi \leq 0$

$$\dot{\mathcal{P}} := \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} + \dot{\gamma}\Phi = \underbrace{\boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^e + \beta\dot{\boldsymbol{\alpha}}}_{\dot{\Psi}} + \underbrace{\boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^p - \beta\dot{\boldsymbol{\alpha}}}_{\mathcal{D}^p} + \dot{\gamma}\Phi = \dot{\Psi} + \mathcal{D}^p + \dot{\gamma}\Phi, \quad (7)$$

where all terms have the physical unity of power, i.e. work per unit time.

Time integration over the interval $[t_0, t_{n+1}]$ yields

$$\int_0^t \dot{\mathcal{P}} d\tau = \int_0^t (\dot{\Psi} + \mathcal{D}^p + \dot{\gamma}\Phi) d\tau \quad (8a)$$

$$= \Psi + \int_0^t (\mathcal{D}^p + \dot{\gamma}\Phi) d\tau = \Psi + \sum_{i=0}^n \int_{t_i}^{t_{i+1}} (\mathcal{D}^p + \dot{\gamma}\Phi) d\tau, \quad (8b)$$

with $\dot{\mathcal{P}} = \dot{\Psi}$ holds in the limit case of vanishing plasticity.

The occurring errors can be characterized as follows, see eqns. 8a, 8b:

1. The error of equilibrium related to the elastic strains:

$$\eta_{eq}^2 = \|(\boldsymbol{\sigma}^e - \boldsymbol{\sigma}_h^e) : (\boldsymbol{\varepsilon}^e - \boldsymbol{\varepsilon}_h^e)\|_{L_2(\Omega)}^2 + \|(\boldsymbol{\alpha} - \boldsymbol{\alpha}_h)H(\boldsymbol{\alpha} - \boldsymbol{\alpha}_h)\|_{L_2(\Omega)}^2 \quad (9)$$

2. The error increment of the directions of plastic strain rates, related to plastic dissipation:

$$\eta_{\dot{\boldsymbol{\varepsilon}}^p}^2 = \|\boldsymbol{\sigma} : (\dot{\boldsymbol{\varepsilon}}^p - \dot{\boldsymbol{\varepsilon}}_h^p)\|_{L_2(\Omega)}^2 + \|\beta(\dot{\boldsymbol{\alpha}} - \dot{\boldsymbol{\alpha}}_h)\|_{L_2(\Omega)}^2 \quad (10)$$

3. The error increment in the Kuhn-Tucker conditions for yielding, loading and unloading of a load factor λ :

$$\eta_{KT}(\lambda) = \|\dot{\gamma}_h \Phi_h\|_{L_2(\Omega)}. \quad (11)$$

5.2 Spatial error indicator

The error in the global equilibrium can be controlled by means of the residual error estimator η_{eq} by Babuška, Rheinboldt & Miller [3] and Johnson, Hansbo [14]. It reads

$$\eta_{eq}^2 = C_R h^2 \|\mathbf{R}_h\|_{L_2(\Omega)}^2 + C_j h \|\mathbf{J}_h\|_{L_2(\Gamma)}^2, \quad [\eta_{eq}] = \sqrt{Nm} \quad (12)$$

with the residuals $\mathbf{R}_h = \text{div } \boldsymbol{\sigma}_h - \mathbf{f}_h$, the jumps of the stresses on inter-element boundaries \mathbf{J}_h and the interpolations constants C_R and C_j which remain unknown in case of relative errors. Fig. 3a shows the spatial distribution of η_{eq} for the load $\lambda = 4.5$.



Figure 3a: η_{eq} for load $\lambda = 4.5$

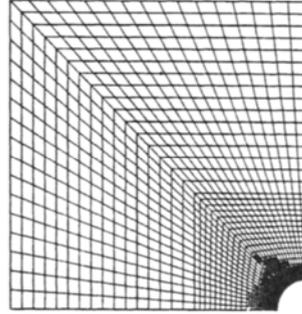


Figure 3b: Adapted mesh for η_{eq} with 1168 elements and 1242 nodes

Figs. 3a, 3b illustrate that the areas of high relative errors are concentrated at the edge of the hole close to the singularity. It is obvious that this error indicator is not sufficient to describe the effects of elastoplastic deformations.

The error increment of the directions of plastic strain rates $\dot{\boldsymbol{\epsilon}}^p$ at the end of a load-step is given by the weighted L_2 -norm

$$\eta_{\dot{\boldsymbol{\epsilon}}^p}^2 = \|\boldsymbol{\sigma} : (\dot{\boldsymbol{\epsilon}}^p - \dot{\boldsymbol{\epsilon}}_h^p)\|_{L_2(\Omega)}^2 + \|\beta \cdot (\dot{\alpha} - \dot{\alpha}_h)\|_{L_2(\Omega)}^2, \quad [\eta_{\dot{\boldsymbol{\epsilon}}^p}] = \sqrt{Nm}, \quad (13)$$

introduced in [6]. As the exact solutions $\boldsymbol{\sigma}$, $\dot{\boldsymbol{\epsilon}}^p$, β and $\dot{\alpha}$ are unknown, we use smoothed discretized solutions $\boldsymbol{\sigma}_*$, $\dot{\boldsymbol{\epsilon}}_*^p$, β_* and $\dot{\alpha}_*$, gained by L_2 -projections according to [23]. Thus $\eta_{\dot{\boldsymbol{\epsilon}}^p}^2$ reads

$$\eta_{\dot{\boldsymbol{\epsilon}}^p}^2 \approx \|\boldsymbol{\sigma}_* : (\dot{\boldsymbol{\epsilon}}_*^p - \dot{\boldsymbol{\epsilon}}_h^p)\|_{L_2(\Omega)}^2 + \|\beta_* \cdot (\dot{\alpha}_* - \dot{\alpha}_h)\|_{L_2(\Omega)}^2 \quad (14)$$

Figs. 4a, 4b show the distribution of $\eta_{\dot{\boldsymbol{\epsilon}}^p}$ at $\lambda = 4.5$ and the related adaptive mesh. This indicator acts inside the plastic domain.

A third error occurs in the flow condition Φ outside the Gauss integration points and is controlled by the error indicator of the Kuhn-Tucker conditions for yielding, loading and unloading

$$\Phi \leq 0, \quad \dot{\gamma} \geq 0, \quad \dot{\gamma} \Phi = 0. \quad (15)$$

The related global error indicator η_{KT} by Barthold and Stein [8] reads with (11)

$$\eta_{KT}^2(\lambda) = \|\dot{\gamma} \Phi - \dot{\gamma}_h \Phi_h\|_{L_2(\Omega)} = \|\dot{\gamma}_h \Phi_h\|_{L_2(\Omega)}; \quad [\eta_{KT}] = \sqrt{Nm}. \quad (16)$$

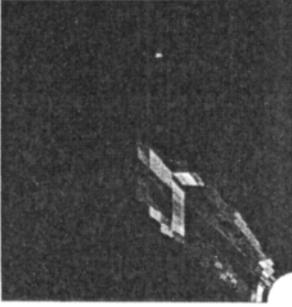


Figure 4a: $\eta_{\dot{\epsilon}^P}$ at a load of $\lambda = 4.5$

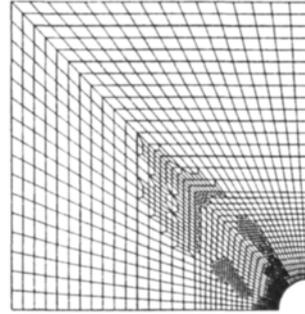


Figure 4b: Adapted mesh from $\eta_{\dot{\epsilon}^P}$ with 1285 elements, 1368 nodes

The yield condition is fulfilled at all quadrature points but not elsewhere in the elements. The idea for η_{KT} is consequently an extrapolation of the given Gauss-point values to integration points of higher polynomial order.

Figs. 5a,b show the error distribution of η_{KT} at a load of $\lambda = 4.5$ and the resulting mesh refinement. This estimator concentrates at the transition zone between the elastic and the plastic parts of the domain.

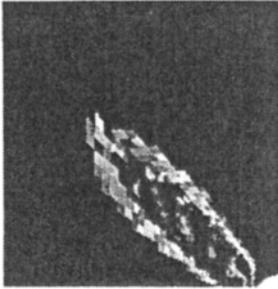


Figure 5a: η_{KT} at a load $\lambda = 4.5$

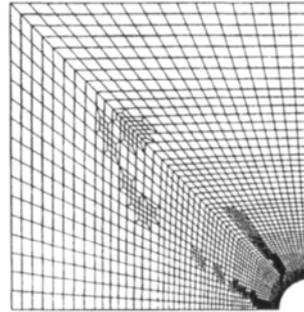


Figure 5b: Adapted mesh for η_{KT}

As mentioned before, the adaptive computations start with a sequence of graded meshes, Figs. 8a–8c. From there the error-controlled adaptive process starts according to the three developed indicators η_{eq} , $\eta_{\dot{\epsilon}^P}$ and η_{KT} , see section 7.2, Figs. 10a–10d and 11a–11d.

5.3 Time error indicator

Using numerical integration schemes such as the backward Euler rule, an additional error is present for each implicit pseudo time interval $[t_i; t_{i+1}]$, which has an accumulation effect.

The basic idea consists in detecting points within the plastic domain with maximal deviation of Hencky-plasticity from Prandtl-Reuß-plasticity. In the latter theory, the plastic strain rates $\dot{\epsilon}^P$ (instead of the plastic strains ϵ^P) are proportional to the stress deviator

in the flow rule. Therefore Hencky-plasticity only describes a non-linear elastic material with an additional yield condition but without a flow rule. The error $\eta_{\Delta t}$ in the numerical integration of Hill's flow rule is defined by

$$\eta_{\Delta t} = \|\sigma\|_{L_2(\Omega)} \left\| \int_{t=t_i}^{t_{i+1}} \dot{\epsilon}^p dt - \Delta \epsilon^p \right\|_{L_2(\Omega)} \quad (17)$$

with the scaling factor $\|\sigma\|_{L_2(\Omega)}$ in order to characterize the influence on the stress power. The physical unit of the square of this error indicator is also Nm.

For numerical computations we assume, that the space discretization does not affect the time discretization error, i.e. the quantities in the definition above can be replaced by FE-solutions. The error can then be bounded for each time step by

$$\left\| \int_{t_i}^{t_{i+1}} \dot{\epsilon}_h^p dt - \Delta \epsilon_h^p \right\|_{L_2(\Omega)} \leq \Delta t \max_{t \in [t_i, t_{i+1}]} \|\dot{\gamma}(t) \mathbf{n}(t) - \gamma_{i+1} \mathbf{n}_{i+1}\|_{L_2(\Omega)} \quad (18)$$

where $\mathbf{n} = \text{dev} \sigma / \|\text{dev} \sigma\|$ denotes the outer normal to the yield surface.

Assuming further that the difference of plastic flow directions $\mathbf{n}(t)$ and \mathbf{n}_{i+1} in the current time interval is maximal for $t = t_i$ and choosing the plastic multiplier $\dot{\gamma}(t)$ as γ_{i+1} , the computable bound

$$\eta_{\Delta t} = \gamma_{i+1} \|\mathbf{n}_i - \mathbf{n}_{i+1}\| \Delta t \quad (19)$$

is obtained. The integrated staggered control of mesh size and time step was treated in [7]. In order to pay tribute to the accumulation error, the adapted time integration starts from the previous time step. The figures 6a,b show the spatial error distribution for $\eta_{\Delta t}$ at different loads. It is significant for this error indicator that the regions with beginning plasticity contribute high values of $\eta_{\Delta t}$ whereas already plastified sub-domains add only low values.

Examples with the convergence properties for graded meshes and adapted meshes due to the above error indicators are given in Figs. 11a – 11d.

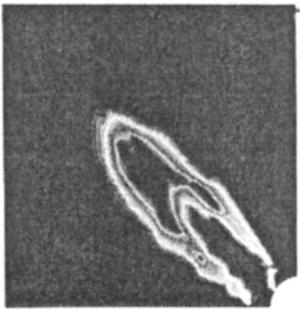


Figure 6a: $\eta_{\Delta t}$ at $\lambda = 4.5$



Figure 6b: $\eta_{\Delta t}$ at $\lambda = 4.75$

6. THE POSTERIOR EQUILIBRIUM METHOD (PEM)

6.1. Formulation of PEM for error estimation on patches

The posterior equilibrium method is based on local recovery analysis of improved tractions at the internal element boundaries with C^0 -continuity condition in normal directions. PEM can be explained as a stepwise hybrid displacement method or as a Trefftz method for Neumann problems of element patches, see also [11, 20, 15, 16, 22]. We formulate regularized variational problems on patches which equilibrate these new tractions with the known energy equivalent nodal forces from the previous finite element solution. Such it is possible to introduce an anisotropic error estimator for the discretization error and especially for the model error. The unknown locally equilibrated boundary tractions $\tilde{\mathbf{t}}_h$ on each patch k have to fulfill the weak equilibrium conditions (a hat denotes nodal quantities)

$$\int_{\partial\Omega_k} \tilde{\mathbf{t}}_h^T \mathbf{v}_h dO = (\widehat{\mathbf{p}}_{he}^T(\mathbf{u}_h) \widehat{\mathbf{v}}_h)_k \quad \forall \mathbf{v}_h \in \mathbf{V}_h \in \mathbf{V}(\Omega_e); \tilde{\mathbf{t}}_h \in \mathbf{L}_2(\Gamma_e) \quad (20)$$

related to the previous approximated solution \mathbf{u}_h . The resulting element nodal forces $\widehat{\mathbf{p}}_{he}$, the test functions (virtual displacements) \mathbf{v}_h and the unknown locally equilibrated boundary tractions $\tilde{\mathbf{t}}_h$ are given as

$$\widehat{\mathbf{p}}_{he}(\mathbf{u}_h) = \mathbf{K}_e \widehat{\mathbf{u}}_h; \quad \mathbf{v}_h = \mathbf{N}_v \widehat{\mathbf{v}}_h; \quad \tilde{\mathbf{t}}_h = \overline{\mathbf{N}}_t^T \widehat{\mathbf{t}}_h. \quad (21)$$

From eqns. (20 – 21) we get

$$\underbrace{\tilde{\mathbf{t}}_h^T \int_{\partial\Omega} \overline{\mathbf{N}}_t^T \mathbf{N}_v dO}_{\mathbf{T}_e} \widehat{\mathbf{v}}_h = \widehat{\mathbf{p}}_{he}^T(\mathbf{u}_h) \widehat{\mathbf{v}}_h \quad \forall \widehat{\mathbf{v}}_h \in \widehat{\mathbf{V}}_h \quad (22)$$

which results in $\mathbf{T}_e^T \widehat{\mathbf{t}}_h = \widehat{\mathbf{p}}_{he}(\mathbf{u}_h)$. Different from Ladevèze and Leguillon [15] and Brink and Stein [10], eqns. 20 and 21 fulfill the element equilibrium exactly, such that anisotropic error estimation is possible.

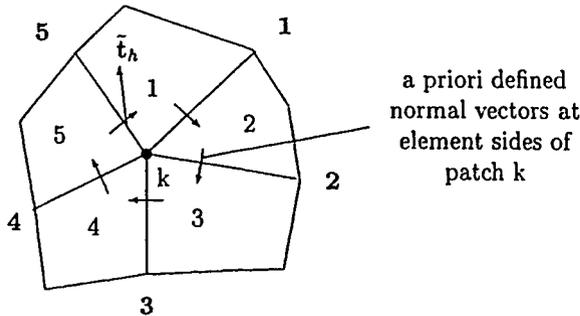


Figure 7: Patch k for node k ; element boundaries are described by side numbers $i=1-5$. A priori defined normal vectors at element sides i are plotted

Only C^0 -continuity conditions for \mathbf{t}_h in the normal direction of element surfaces are necessary, due to $\mathbf{t}_h \in \mathbf{L}_2(\Gamma_k)$ it is sufficient to describe the equilibrated tractions without C^0 -continuity conditions in tangential direction of element surfaces. Consequently, a patch-wise calculation is possible, see Ladevèze, Leguillon (1983) [15], Bank, Weiser (1985) [5] and Ainsworth, Oden (1992) [1].

To avoid coupling effects between neighboring patches it is recommendable to describe the new tractions with orthogonal shape functions $\overline{\mathbf{N}}_t$ with respect to the shape functions \mathbf{N}_v of the test space, see eqn. (22). \mathbf{N}_v and $\overline{\mathbf{N}}_t$ are different bases of the same approximation space. The orthogonality condition reads

$$\int_{\partial\Omega} \overline{\mathbf{N}}_t^i \mathbf{N}_{v,j} dO = \delta_j^i, \quad (23)$$

thus \mathbf{N}_v forms a covariant and $\overline{\mathbf{N}}_t$ the corresponding contravariant basis.

The equilibrium conditions, eqn. (22), for a patch k in Fig. 7 are given by

$$\begin{aligned} \text{Element } 1 : \quad & \widehat{t}_{h1} - \widehat{t}_{h5} = \widehat{p}_{h1}(\mathbf{u}_h) \\ 2 : \quad & \widehat{t}_{h2} - \widehat{t}_{h1} = \widehat{p}_{h2}(\mathbf{u}_h) \\ 3 : \quad & \widehat{t}_{h3} - \widehat{t}_{h2} = \widehat{p}_{h3}(\mathbf{u}_h) \\ 4 : \quad & \widehat{t}_{h4} - \widehat{t}_{h3} = \widehat{p}_{h4}(\mathbf{u}_h) \\ 5 : \quad & \widehat{t}_{h5} - \widehat{t}_{h4} = \widehat{p}_{h5}(\mathbf{u}_h) \end{aligned} \quad (24)$$

and result in

$$\overbrace{\begin{bmatrix} +1 & & & & -1 \\ -1 & +1 & & & \\ & -1 & +1 & & \\ & & -1 & +1 & \\ & & & -1 & +1 \end{bmatrix}}^{\mathbf{T}} \overbrace{\begin{bmatrix} \widehat{t}_{h1} \\ \widehat{t}_{h2} \\ \widehat{t}_{h3} \\ \widehat{t}_{h4} \\ \widehat{t}_{h5} \end{bmatrix}}^{\widehat{\mathbf{t}}_h} = \overbrace{\begin{bmatrix} \widehat{p}_{h1}(\mathbf{u}_h) \\ \widehat{p}_{h2}(\mathbf{u}_h) \\ \widehat{p}_{h3}(\mathbf{u}_h) \\ \widehat{p}_{h4}(\mathbf{u}_h) \\ \widehat{p}_{h5}(\mathbf{u}_h) \end{bmatrix}}^{\widehat{\mathbf{p}}_h(\mathbf{u}_h)} \quad (25)$$

where \mathbf{T} can be interpreted as a topology matrix. Eqn. (25) can be written as

$$\mathbf{T}^T \widehat{\mathbf{t}}_h = \bigcup_{e \in k} \mathbf{T}_e^T \widehat{\mathbf{t}}_h = \bigcup_{e \in k} \widehat{\mathbf{p}}_{he}(\mathbf{u}_h) = \widehat{\mathbf{p}}_h(\mathbf{u}_h). \quad (26)$$

We introduce covariant Lagrangean base functions and their contravariant base functions

$$\mathbf{N}_v(r, s) = \mathbf{N}_{\mathcal{L}}(r, s) = \mathbf{N}_{\mathcal{L}}(r) \otimes \mathbf{N}_{\mathcal{L}}(s), \quad (27)$$

$$\overline{\mathbf{N}}_t(r, s) = \frac{\overline{\mathbf{N}}_{\mathcal{L}}(r, s)}{\det \left(J(r, s) \right)_{\partial\Omega_e}} = \frac{\overline{\mathbf{N}}_{\mathcal{L}}(r) \otimes \overline{\mathbf{N}}_{\mathcal{L}}(s)}{\det \left(J(r, s) \right)_{\partial\Omega_e}}, \quad (28)$$

or the hierarchical Legendre base functions, respectively. The topology matrix \mathbf{T} , see eqn. (21),(26), is not regular and repeated as eqn. (i)

$$(i) \quad \mathbf{T}^T \widehat{\mathbf{t}}_h = \widehat{\mathbf{p}}_h(\mathbf{u}_h) \quad ; \quad \det(\mathbf{T}^T \mathbf{T}) = 0. \quad (29)$$

In 2D-problems there are one or more zero eigenvalues of $\mathbf{T}^T \mathbf{T}$, and in 3D-problems there are five or more zero eigenvalues.

The regularization of the local equation system for patch k is given by two additional conditions.

The *first additional condition* is posed on Neumann boundaries of the structural system where equilibrium with the FE-nodal forces is satisfied explicitly as

$$(ii) \quad \widehat{\mathbf{p}}_h(\mathbf{t}) = \widehat{\mathbf{t}}_h \quad \text{on } \Gamma_t. \quad (30)$$

The *second additional condition* is gained with FE-stresses $\sigma_h(\mathbf{u}_h)$ in Ω_e by postulating that the boundary tractions from the previous finite element solution \mathbf{u}_h are approximately equal to the new tractions in the weak sense.

$$(iii) \quad \int_{\partial\Omega_e} (\sigma_h(\mathbf{u}_h) \cdot \mathbf{n})^T \mathbf{v}_h dO \simeq \int_{\partial\Omega_e} \widetilde{\mathbf{t}}_h^T \mathbf{v}_h dO \quad \forall \mathbf{v}_h \in \mathbf{V}_h, \quad (31)$$

$$\underbrace{\int_{\partial\Omega_e} (\sigma_h(\mathbf{u}_h) \cdot \mathbf{n})^T \mathbf{N}_v dO}_{\widehat{\mathbf{p}}_h(\sigma_h)} \widehat{\mathbf{v}}_h \simeq \widetilde{\mathbf{t}}_h^T \int_{\partial\Omega_e} \underbrace{\frac{\overline{\mathbf{N}}_t^T}{\det(J(r,s))_{\partial\Omega_e}}}_{\overline{\mathbf{N}}_t^T} \underbrace{\mathbf{N}_v dO}_{\mathbf{N}_\mathcal{L}} \widehat{\mathbf{v}}_h \quad (32)$$

$$\widehat{\mathbf{p}}_h(\sigma_h)^T \widehat{\mathbf{v}}_h \simeq \widetilde{\mathbf{t}}_h^T \underbrace{\int_{\partial\Omega_e} \frac{1}{\det(J(r,s))_{\partial\Omega_e}} \overline{\mathbf{N}}_t^T \mathbf{N}_\mathcal{L} \det(J(r,s))_{\partial\Omega_e} dO}_{\mathbf{I}} \widehat{\mathbf{v}}_h \quad (33)$$

$$\Rightarrow \widehat{\mathbf{p}}_h(\sigma_h) = \widehat{\mathbf{t}}_h. \quad (34)$$

This results in a least-squares-approximation

$$(iii) \quad 1/2(\widehat{\mathbf{p}}_h(\sigma_h) - \widehat{\mathbf{t}}_h)^2 \rightarrow Min. \quad (35)$$

Summarizing the calculation of new tractions $\widetilde{\mathbf{t}}_h$:

$$\left. \begin{array}{l} \text{Nodal forces} \quad \mathbf{T}^T \widehat{\mathbf{t}}_h - \widehat{\mathbf{p}}_h(\mathbf{u}_h) = \mathbf{0} \\ \text{Boundary tractions} \quad \widehat{\mathbf{t}}_h - \widehat{\mathbf{p}}_h(\mathbf{t}) = \mathbf{0} \end{array} \right\} =: \mathbf{T}^T \widehat{\mathbf{t}}_h = \widehat{\mathbf{p}}_h \quad \begin{array}{l} \text{exactly fulfilled} \\ \text{conditions (i), (ii)} \end{array}$$

$$\text{Regularization} \quad 1/2(\widehat{\mathbf{p}}_h(\sigma_h) - \widehat{\mathbf{t}}_h)^2 \rightarrow Min \quad \begin{array}{l} \text{weakly fulfilled} \\ \text{condition (iii).} \end{array} \quad (36)$$

6.2. Higher order equilibration for improved effectivity index

Using higher order polynomials ($p > 3$) of the current approximation space and in case of solutions with very small errors the effectivity index tends to values much higher than 1. The reason is that the locally equilibrated tractions are not consistent to higher order deformation modes. The error is $\mathbf{e} = \mathbf{u} - \mathbf{u}_h \in \mathbf{V}$ and the discrete error of the hierarchically expanded test space reads $\mathbf{e}_{h+} = \mathbf{e}_h + \mathbf{e}_+$ with the interpolations $\mathbf{e}_h = \mathbf{N}_v \widehat{\mathbf{e}}_h \in \mathbf{V}_h$ and $\mathbf{e}_+ = \mathbf{N}_{v+} \widehat{\mathbf{e}}_+ \in \mathbf{V}_+$. If, e.g., the error is zero ($\mathbf{e} = \mathbf{e}_h + \mathbf{e}_+ = \mathbf{0}$) w.r.t. the locally higher

order approximation space than the following equation for the local Neumann problem on Ω_k holds

$$\begin{bmatrix} \mathbf{K}_h & \mathbf{L}_{h+} \\ \mathbf{L}_{h+}^T & \mathbf{K}_+ \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_h + \hat{\mathbf{e}}_h \\ \hat{\mathbf{e}}_+ \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{P}}_h \\ \hat{\mathbf{P}}_+ \end{bmatrix} = \begin{bmatrix} \int_{\partial\Omega_k} \tilde{\mathbf{t}}_h \mathbf{N}_h dx \\ \int_{\partial\Omega_k} \tilde{\mathbf{t}}_h \mathbf{N}_+ dx \end{bmatrix}, \quad (37)$$

with $\mathbf{K}_h := \int_{\Omega_k} \mathbf{B}_h^T \mathbf{C} \mathbf{B}_h dx$, $\mathbf{K}_+ := \int_{\Omega_k} \mathbf{B}_+^T \mathbf{C} \mathbf{B}_+ dx$, $\mathbf{L}_{h+} := \int_{\Omega_k} \mathbf{B}_h^T \mathbf{C} \mathbf{B}_+ dx$. This leads to the demand for $\mathbf{e} = \mathbf{e}_h + \mathbf{e}_+ = \mathbf{0}$, i.e.

$$\mathbf{L}_{h+}^T \mathbf{u}_h = \hat{\mathbf{P}}_+. \quad (38)$$

This condition can only be fulfilled by higher order equilibration with introduction of locally higher order equilibrated tractions $\tilde{\mathbf{t}}_{h+}$.

6.3. PEM for anisotropic error estimation in Prandtl-Reuß elastoplasticity

In case of Prandtl-Reuß elastoplasticity the error estimation has to be calculated in the tangential space. For linear elasticity the error estimation yields the local variational Neumann problem for the error $\mathbf{e} \in \mathbf{V}$ with higher test functions \mathbf{v}_+

$$a(\mathbf{e}, \mathbf{v}_{h+})_{\Omega_e} = L(\mathbf{v}_{h+})_{\partial\Omega_e} - a(\mathbf{u}_h, \mathbf{v}_{h+})_{\Omega_e} \quad \forall \mathbf{v}_{h+} \in \mathbf{V}_{h+}. \quad (39)$$

For elastoplasticity the following tangential space of nonlinear deformation and the linear form have to be introduced on element level

$$\bar{a}(\mathbf{e}, \mathbf{v}_{h+})_{\Omega_e} = \int_{\partial\Omega_e} \tilde{\mathbf{t}} \mathbf{v}_{h+} dO - \int_{\Omega_e} \mathbf{B}_{h+}^T \boldsymbol{\sigma}_h dV. \quad (40)$$

For anisotropic refinement with discrete test- and solution spaces ($\mathbf{V} \rightarrow \mathbf{V}_{h+}$) one has to take testwise $p_+ = 1$ and 2 for each local direction 1,2,3. For perfect plasticity the tangent is given by

$$\bar{a}(\mathbf{e}, \mathbf{v}_{h+})_{\Omega_e} = \int_{\Omega_e} \boldsymbol{\varepsilon}(\mathbf{e})(\mathbf{C} - 2\mu \mathbf{n} \otimes \mathbf{n}) \boldsymbol{\varepsilon}(\mathbf{v}_{h+}) dV \quad (41)$$

with the elasticity tensor \mathbf{C} and the relations of the current approximation

$$\begin{aligned} \text{flow direction} \quad \mathbf{n}_h &= \frac{dev \boldsymbol{\sigma}_h}{|dev \boldsymbol{\sigma}_h|} \\ \text{yield condition} \quad \Phi_h &= |dev \boldsymbol{\sigma}_h| - \sqrt{\frac{2}{3}} y_0 = 0 \\ \text{flow rule} \quad \dot{\boldsymbol{\varepsilon}}_h^p &= \dot{\lambda} grad_\sigma \Phi_h = \dot{\lambda} \mathbf{n}_h. \end{aligned}$$

The flow direction \mathbf{n}_h is given from FEM. Then the local anisotropic element error for Prandtl-Reuß elastoplasticity is

$$\eta_e = \sqrt{\bar{a}(\mathbf{e}, \mathbf{e})_{\Omega_e}} = \|\mathbf{e}\|_{E(\Omega_e)}. \quad (42)$$

6.4. Solution- and model-error estimation

After calculation of the equilibrated tractions the discretization error can be calculated by the difference of the equilibrated tractions $\tilde{\mathbf{t}}_h$ and the previous FE-tractions \mathbf{t}_h in the L_2 -norm or better by the energy norm of the difference of the new solution \mathbf{u}_{h+} of the

Neumann problem (for postprocessed tractions $\tilde{\mathbf{t}}_h$) minus the previous solution \mathbf{u}_h , see also Ladevèze, Leguillon [15] and Bank, Weiser [5].

Introducing the locally expanded test space $\mathbf{v}_{h+} \in \mathbf{V}_{h+} \subset \mathbf{V}$ and the locally expanded solution space $\mathbf{u}_{h+} \in \mathbf{V}_{h+} \subset \mathbf{V}$, the local variational problem is formulated as

$$a(\mathbf{u}_{h+}, \mathbf{v}_{h+})_{\Omega_e} = L(\mathbf{v}_{h+})_{\partial\Omega_e} \quad \forall \mathbf{v}_{h+} \in \mathbf{V}_{h+}; \mathbf{u}_{h+} \in \mathbf{V}_{h+} \quad (43)$$

with the locally enhanced test space

$$\mathbf{V}_{h+} := \{ \mathbf{v}_{h+} \in H_0^1(\Omega_e); \mathbf{v}_{h+} \text{ without rigid body modes} \} \quad (44)$$

and

$$L(\mathbf{v}_{h+})_{\partial\Omega_e} = \int_{\partial\Omega_e} \mathbf{v}_{h+}^T \tilde{\mathbf{t}}_h dO. \quad (45)$$

The local tractions $\tilde{\mathbf{t}}_h$ are gained from PEM. Mechanically, eqn. (43) describes the weak form of equilibrium of the tractions $\tilde{\mathbf{t}}_h$ and yields the displacements \mathbf{u}_{h+} without rigid body modes. The “equilibrium” discretization error estimator η_{De} in Ω_e is given by the difference of the current FE-approximation \mathbf{u}_h and the locally improved approximation \mathbf{u}_{h+} in the energy norm by

$$\eta_{De}^2 = a((\mathbf{u}_{h+} - \mathbf{u}_h), (\mathbf{u}_{h+} - \mathbf{u}_h))_{\Omega_e} \quad \text{and} \quad \eta_{DGI}^2 = \|\boldsymbol{\sigma}_{h+} - \boldsymbol{\sigma}_h\|_{E(\Omega_e)}. \quad (46)$$

The estimator, eqn. (47), yields an upper bound if the hierarchically expanded test spaces are sufficient. It has the same order of magnitude for convergence as the Babuška/Rheinboldt residual error estimator, but anisotropic error estimation and the interpolation constants are computed indirectly with higher accuracy. For model error estimation we introduce the local variational form a_M for an hierarchically expanded model, connected with the reduced model by a monomorphic transformation, and the corresponding discretized solution $\tilde{\mathbf{u}}_{h+}$ using again enhanced local test spaces

$$a_M(\tilde{\mathbf{u}}_{h+}, \mathbf{v}_{h+})_{\Omega_e} = L(\mathbf{v}_{h+})_{\partial\Omega_e} \quad , \quad \forall \mathbf{v}_{h+} \in \mathbf{V}_{h+} \quad \text{and} \quad \tilde{\mathbf{u}}_{h+} \in \mathbf{V}_{h+} \quad (47)$$

and yielding $\tilde{\mathbf{u}}_{h+}$ with the right-hand side

$$L(\mathbf{v}_{h+})_{\partial\Omega_e} = \int_{\partial\Omega_e} \mathbf{v}_{h+}^T \tilde{\mathbf{t}}_h dO. \quad (48)$$

Eqn. (47) describes the weak form of equilibrium of the new boundary tractions $\tilde{\mathbf{t}}_h$ (of an element patch k) for the expanded model within element Ω_e . Note that $\tilde{\mathbf{u}}_{h+}$ and \mathbf{u}_{h+} must have the same dimension.

Then, the “equilibrium” model error estimator η_{Me} in the energy norm within Ω_e reads

$$\eta_{Mu_e}^2 = a_M(\tilde{\mathbf{u}}_{h+} - \mathbf{u}_{h+}, \tilde{\mathbf{u}}_{h+} - \mathbf{u}_{h+})_{\Omega_e}; \quad \tilde{\mathbf{u}}_{h+} \in \mathbf{V}_{h+}; \quad \mathbf{u}_{h+} \in \mathbf{V}_{h+} \quad (49)$$

with

$$\eta_{M\sigma_e}^2 = \|\tilde{\boldsymbol{\sigma}}_{h+} - \boldsymbol{\sigma}_{h+}\|_{E_M(\Omega_e)}. \quad (50)$$

This model error is gained by the difference between the locally expanded approximation \mathbf{u}_{h+} (using the reduced $2\frac{1}{2}$ D model of a thin-walled structure) and the locally expanded approximation $\tilde{\mathbf{u}}_{h+}$ (using the full 3D- model) in the energy norm. This model error estimator, eqn. (49), is subject to locking phenomena for thin walled plates whereas the model error estimator, eqn. (49), yields consistent estimates.

To avoid substantial influences of the discretization error on the model error, the global discretization error has to be smaller than the global model error, $\eta_M^2 \leq \eta_D^2$.

7. FURTHER NUMERICAL RESULTS

The following numerical examples were performed for the benchmark system described in section 4. The results were computed for Hencky-type non-linear elastic materials and elastoplastic materials with Hill's flow theory and nonlinear isotropic hardening.

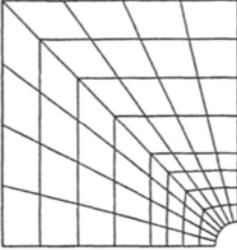


Figure 8a: Graded mesh No. 2 with 64 elements, see Fig. 1b

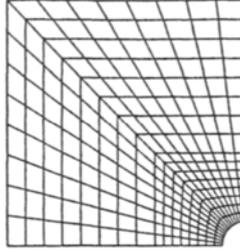


Figure 8b: Graded mesh No. 3 with 256 elements

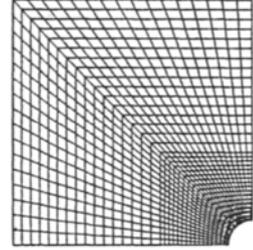


Figure 8c: Graded mesh No. 4 with 1024 elements

7.1. Hencky-type non-linear elasticity

Mesh number 2, Fig. 8a, was refined in 10 steps using the equilibrium criterion by Babuška and Miller, yielding 875 elements with 1816 degrees of freedom. The results of the adaptive computation are shown in the following Figs. 9a-d in comparison with the results of the sequence of graded meshes.

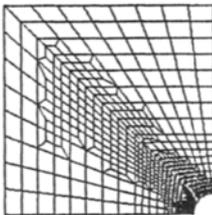


Figure 9a: Adapted mesh at $\lambda = 4.5$ for Hencky-plasticity

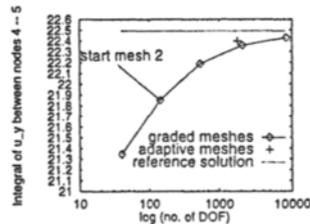


Figure 9b: $\int_4^5 |u_y| dx$ at $\lambda = 4.5$ for Hencky-plasticity

7.2. Prandtl-Reuß elastoplasticity with mixed hardening

In Figs. 10a-f, an adaptive computation starting from mesh 1, yielding finally 802 elements with 1674 degrees of freedom, is shown in comparison to the solutions of the sequence of graded meshes. Mesh No. 1 with 16 elements remains elastic throughout the whole loading process.

In the Figs. 11a-c below, the convergence rates for the three parts of spatial error are shown for a sequence of uniformly refined graded meshes from 256 (mesh no. 3) to 4096 elements (mesh no. 5) for a load of $\lambda = 4.5$. Adaptive computations were performed from

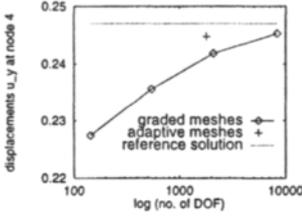


Figure 9c: Displacements u_y of node 4 at $\lambda = 4.5$ for Hencky-plasticity

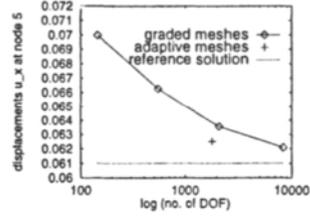


Figure 9d: Displacements u_x of node 5 at $\lambda = 4.5$ for Hencky-plasticity

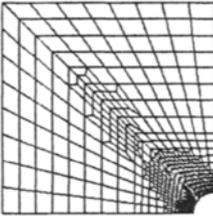


Figure 10a: Adapted mesh at $\lambda = 4.5$ for Prandtl-Reuß elastoplasticity

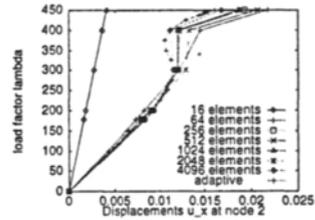


Figure 10b: Displacements u_x of node 2 over λ for Prandtl-Reuß elastoplasticity

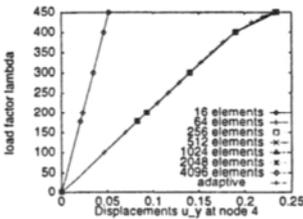


Figure 10c: Displacements u_y of node 4 over λ for Prandtl-Reuß elastoplasticity

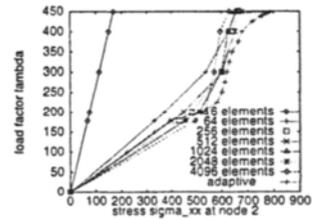


Figure 10d: Stresses σ_{yy} of node 2 over λ for Prandtl-Reuß elastoplasticity

start meshes number 2 and 4. Fig. 11a presents the error of equilibrium η_{eq} ; Fig. 11b depicts the evolution of the error η_{ϵ^p} w.r.t. the plastic strain rates, and Fig. 11c shows the evolution of the error η_{KT} of the Kuhn-Tucker-conditions. A higher rate of convergence is obtained for all the three adaptive computations.

Finally the convergence of the combined spatial error indicators was considered. Fig. 11d shows again that the convergence rate is higher for the adaptive computation starting from mesh number 4 than for graded meshes

Figs. 12a - 12h show results from elastoplastic plate with a hole, calculated with isotropic and anisotropic PEM. \circ : isotropic error estimation; \square : anisotropic error estimator.

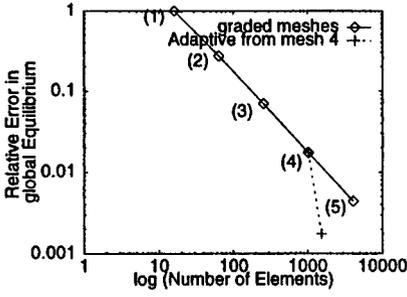


Figure 11a: Convergence due to η_{eq}

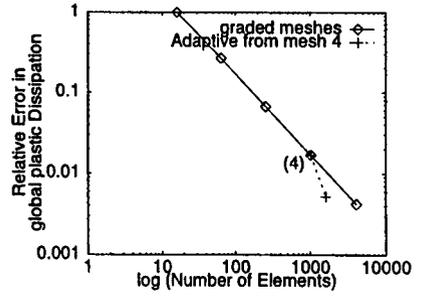


Figure 11b: Convergence due to η_{ϵ^p}

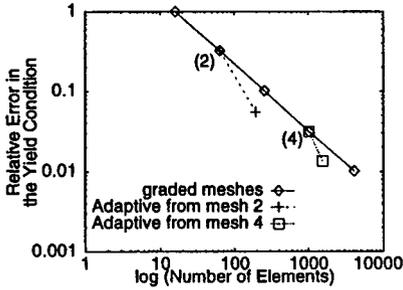


Figure 11c: Convergence due to η_{KT}

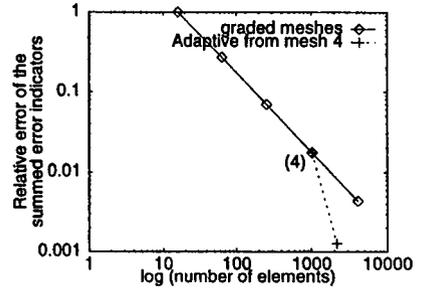


Figure 11d: Convergence due to the total spatial error indicator

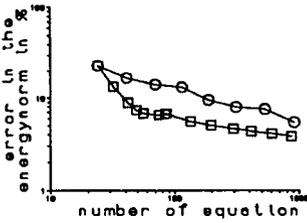


Figure 12a: Relative error in the energynorm

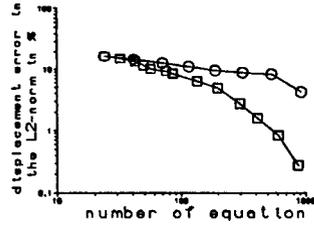


Figure 12b: Relative error in the L_2 -norm

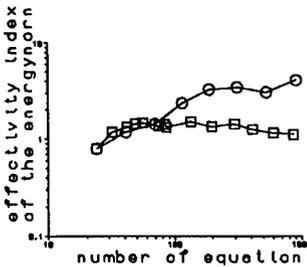


Figure 12c: Effectivity index

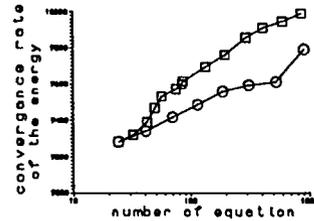


Figure 12d: Convergence of the energy

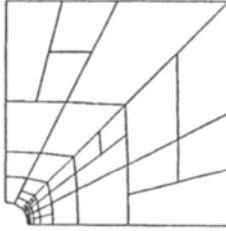


Figure 12e: 4th Mesh with anisotropic PEM

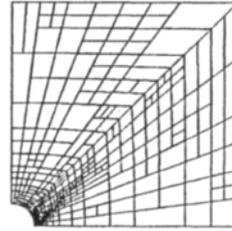


Figure 12f: 9th Mesh with anisotropic PEM

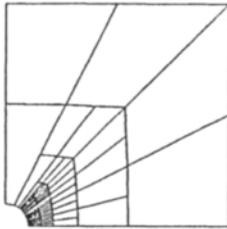


Figure 12g: 4th Mesh with isotropic PEM

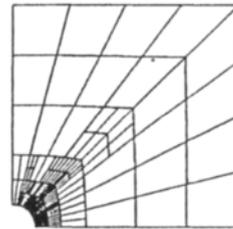


Figure 12h: 9th Mesh with isotropic PEM

8. CONCLUSIONS AND FUTURE WORK

It was shown that the primal FEM for elastoplasticity of Prandtl-Reuß-type with FEM in space and FDM in time yields essentially three spatial discretization errors, namely of equilibrium, of the Kuhn-Tucker conditions (yield condition and loading conditions) and of the plastic strain rates which can be interpreted as a control of the plastic dissipation condition. At the time there is not yet a common mathematical basis for these three errors due to the complicated trace-spaces in which the elastoplastic solutions are located.

Dual and dual hybrid FE-functionals, as applied in [9] for elasticity and in [19] for elastoplasticity, shall be investigated further on. Moreover local error analysis by using dual arguments is in progress.

The Posterior Equilibrium Method (PEM) yields physically consistent orthogonalized boundary tractions \hat{t}_h . With these tractions an error estimator was gained with solving local Neumann problems. For elastoplasticity the local Neumann problem was formulated in the tangential space. This method leads to anisotropic error estimators with upper bounds of the discretization error for elasticity and elastoplasticity and also an consistent model error estimator by locally changing from a reduced model to an enhanced model.

In the concept of operator split (FEM in space, FDM in time) the coupling of space-time discretization errors is only possible in a crude way using time step control with constant time increments for the whole spatial system. Our staggered scheme for space-time error control, [7], is promising, especially for un- and reloading processes. A true coupled space-time error analysis is only possible for FEM in space *and* time which becomes also attractive for parallel computing by e.g. splitting a time slab into p subproblems, p being the number of processors.

REFERENCES

- [1] M. Ainsworth and J. Oden. A procedure for a posteriori error estimation for h-p finite element methods. *Computer Methods in Applied Mechanics and Engineering*, 101:73–96, 1992.
- [2] M. Ainsworth and J. Oden. A posteriori error estimation in finite element analysis. Mathematics and computer science technical reports, University of Leicester, 1995.
- [3] I. Babuška and A. Miller. A-posteriori error estimates and adaptive techniques for the finite element method. Technical report, Institute for Physical Science and Technology, University of Maryland, 1981.
- [4] I. Babuška, T. Strouboulis, S. Gangaraj, and C. Upadhyay. A posteriori estimation and adaptive control of the pollution error in the h-version of the finite element method. *International Journal for Numerical Methods in Engineering*, 38:4207–4235, 1995.
- [5] R. E. Bank and A. Weiser. Some a posteriori error estimators for elliptic partial differential equations. *Mathematics of Computation*, 44:283–301, 1985.
- [6] F.-J. Barthold, M. Schmidt, and E. Stein. Error estimation and mesh adaptivity for elasto-plastic deformations. In *Proceedings of the Fifth International Conference on Computational Plasticity, Complas V, Barcelona*, pages 597–603, 1997.
- [7] F.-J. Barthold, M. Schmidt, and E. Stein. Error indicators and mesh refinements for finite-element-computations of elastoplastic deformations. *International Journal for Computational Mechanics*, to appear, 1997.
- [8] F.-J. Barthold and E. Stein. Error estimation and mesh adaptivity for elasto-plastic deformations. *ZAMM: Proc. ICIAM/GAMM*, 76 Sup. 1:159–162, 1996.
- [9] D. Braess, O. Klaas, R. Niekamp, E. Stein, and F. Wobschal. Error indicators for mixed finite elements in 2-dimensional linear elasticity. *Computer Methods in Applied Mechanics and Engineering*, 127:345–356, 1995.
- [10] U. Brink and E. Stein. On some mixed finite element methods for incompressible and nearly incompressible finite elasticity. *International Journal for Computational Mechanics*, 19:105–119, 1996.
- [11] H. Buefler and E. Stein. Zur Plattenberechnung mittels finiter Elemente. *Ing. Archiv*, 39:248–260, 1970.
- [12] J. Fish and S. Markolefas. Adaptive global-local refinement strategy based on the interior error estimates of the h-method. *International Journal for Numerical Methods in Engineering*, 37:827–838, 1994.

- [13] H. Hencky. Zur Theorie plastischer Deformationen und der hierdurch im Material hervorgerufenen Nebenspannungen. In *Proceedings of the 1st International Congress on Applied Mechanics, Delft, 1924*.
- [14] C. Johnson and P. Hansbo. Adaptive finite element methods in computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, 101:143–181, 1992.
- [15] P. Ladeveze and D. Leguillon. Error estimate procedure in the finite element method and applications. *SIAM J. Num. Anal.*, 20:485–509, 1983.
- [16] P. Ladevèze and E. Maunder. A general methodology for recovering equilibration finite element tractions and stress fields for plate and solid elements. in *First international workshop on Trefftz Method, 30. May – 1. June, 1996, Cracov, Poland, Archives of Mechanics*, pages 34–35, 1995.
- [17] R. Prandtl. Spannungsverteilung in plastischen Körpern. In *Proceedings of the 1st International Congress on Applied Mechanics, Delft, 1924*.
- [18] R. Rannacher and F.-T. Suttmeier. A posteriori error control in finite element methods via duality techniques: Application to perfect plasticity. Technical report, Preprints of SFB 350, Universität Heidelberg, 1997.
- [19] J. Schröder, O. Klaas, E. Stein, and C. Miehe. A physically nonlinear dual mixed finite element formulation. Technical report, Institut für Baumechanik und Numerische Mechanik, Universität Hannover, 1995.
- [20] E. Stein and R. Ahmad. An equilibrium method for stress calculation using finite element displacements models. *Computer Methods in Applied Mechanics and Engineering*, 10:175–198, 1977.
- [21] E. Stein and S. Ohnimus. Coupled model- and solution-adaptivity in the finite-element method. *Computer Methods in Applied Mechanics and Engineering*, in print:24, 1997.
- [22] E. Stein and S. Ohnimus. Equilibrium method for postprocessing and error estimation in the finite element method. *Computer Assisted Mechanics and Engineering Sciences*, 4:645–666, 1997.
- [23] O. C. Zienkiewicz and J. Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis. *International Journal for Numerical Methods in Engineering*, 24:337–357, 1987.
- [24] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and adaptive finite element refinement. *Computer Methods in Applied Mechanics and Engineering*, 101:207–224, 1992.

This Page Intentionally Left Blank

A Reliable A Posteriori Error Estimator for Adaptive Hierarchic Modelling

M. Ainsworth and M. Arnold ^{a *}

^aMathematics Department, Leicester University,
 Leicester LE1 7RH, United Kingdom

An *a posteriori* error estimator is obtained for estimating the error in non-uniform order hierarchic models of elliptic boundary value problems on thin domains, and is shown to provide an upper bound on the actual error measured in the energy norm. The estimator has many similarities with a known heuristic refinement indicator used for adaptive hierarchic modelling and in this sense provides some theoretical support for the existing technique. Numerical examples show that the performance of both estimators is comparable, both yielding good results, with the new technique performing slightly better.

1. Hierarchic Modelling

Let ω be a plane, polygonal domain with boundary $\partial\omega$, and let $\Omega = \omega \times (-d/2, d/2)$ be the domain of thickness $d > 0$. If the thickness parameter d is small compared with the dimensions of the mid-surface ω then Ω is often referred to as a *plate*.

Consider a steady state heat conduction problem with heat fluxes $f_+, f_- \in L_2(\omega)$ prescribed on on the upper and lower faces of the plate

$$\Omega_{\pm} = \{\mathbf{x} = (\mathbf{x}', x_3) : \mathbf{x}' \in \omega, x_3 = \pm d/2\}.$$

On the lateral boundary, $\Gamma = \partial\omega \times (-d/2, d/2)$, the temperature u of the plate is prescribed to be zero. The heat flux vector \mathbf{q} is assumed to be related to the temperature gradient according to

$$\mathbf{q}(\mathbf{x}) = - \begin{bmatrix} b(2x_3/d) & & \\ & b(2x_3/d) & \\ & & a(2x_3/d) \end{bmatrix} \nabla u(\mathbf{x}), \quad \mathbf{x} \in \Omega$$

where the thermal conductivities a and b are even functions independent of the thickness d and satisfy $a, b \in L_{\infty}(-1, 1)$ and

$$0 < \underline{a} \leq a(z), \quad 0 < \underline{b} \leq b(z).$$

In particular, these assumptions allow the material properties to be piecewise constant as would be the case for the laminated domain shown in Figure 1. In addition to the

^{*}The support of the Engineering and Physical Science Research Council through a research studentship is gratefully acknowledged.

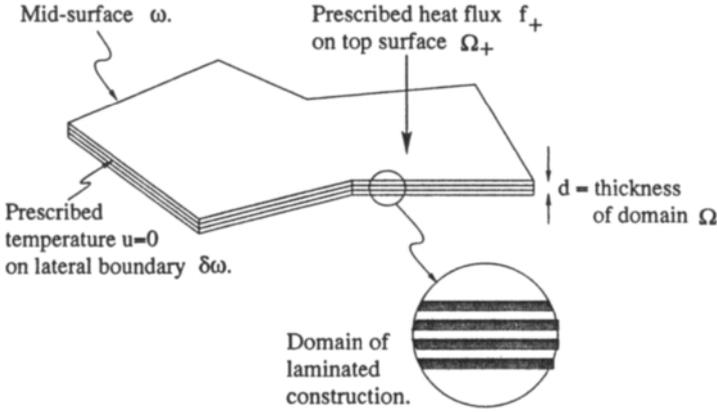


Figure 1. Plate-like domain of laminated construction.

constitutive equation, the boundary value problem may be formally stated as

$$\nabla \cdot \mathbf{q} = 0 \text{ in } \Omega; \quad \mathbf{n}_{\pm} \cdot \mathbf{q}(\mathbf{x}', \pm d/2) + f_{\pm} = 0 \text{ on } \Omega_{\pm}; \quad u = 0 \text{ on } \Gamma.$$

The variational form of the problem is

$$\text{find } u \in V : B(u, v) = L(v) \quad \forall v \in V$$

(1)

where V is the space

$$V = \left\{ v \in H^1(\Omega) : v = 0 \text{ on } \Gamma \right\}$$

with $B : V \times V \rightarrow \mathbb{R}$ and $L : V \rightarrow \mathbb{R}$ given by

$$B(u, v) = \int_{\Omega} \left\{ a \left(\frac{2x_3}{d} \right) \frac{\partial u}{\partial x_3} \frac{\partial v}{\partial x_3} + b \left(\frac{2x_3}{d} \right) \nabla' u \cdot \nabla' v \right\} dx' dx_3$$

and

$$L(v) = \int_{\omega} \{ f_+(\mathbf{x}')v(\mathbf{x}', d/2) + f_-(\mathbf{x}')v(\mathbf{x}', -d/2) \} dx'$$

where $\nabla' = (\partial/\partial x_1, \partial/\partial x_2, 0)$.

1.1. A Simple Plate Model

Generally, the true solution u will vary through the thickness. However, if the domain is thin, then the variation will be relatively small in comparison with the variation over the mid-surface. Therefore, it is natural to attempt to simplify the problem by seeking an approximate solution that is *constant through the thickness*. Mathematically, this amounts to seeking the solution of the problem: find $u_0 \in V_0$ such that

$$B(u_0, v) = L(v) \quad \forall v \in V_0$$

where $V_0 \subset V$ is the subspace of admissible functions that remain constant through the thickness,

$$V_0 = \{v \in H^1(\Omega) : v(\mathbf{x}', x_3) = \alpha_0(\mathbf{x}'), \quad \alpha_0 = 0 \text{ on } \partial\omega\}.$$

For smooth data, this problem is readily found to be equivalent to solving a partial differential equation posed over the mid-plane ω ,

$$-d\bar{b}\Delta' u_0 = f_+ + f_- \text{ in } \omega \quad (2)$$

with $u_0 = 0$ on $\partial\omega$, where \bar{b} is the value of the thermal conductivity coefficient b averaged through the thickness

$$\bar{b} = \frac{1}{d} \int_{-d/2}^{d/2} b(2x_3/d) dx_3 = \frac{1}{2} \int_{-1}^1 b(s) ds.$$

The equation (2) represents a dimensionally reduced, or *plate model*, for the original three dimension boundary value problem. It is natural to ask how accurately the solution of the reduced model approximates the full solution.

1.2. An A Posteriori/Priori Error Estimate for the Simple Plate Model

Suppose, for ease of exposition, that the prescribed fluxes are equal so that $f_+ = f_- = f$. If the error in the dimensionally reduced approximation is denoted by $e_0 = u - u_0 \in V$, then for any $v \in V$

$$B(e_0, v) = L(v) - B(u_0, v).$$

Let $v_0 \in V_0$ be defined by

$$v_0(\mathbf{x}') = \frac{2}{db} \int_{-d/2}^{d/2} b(2x_3/d) v(\mathbf{x}', x_3) dx_3,$$

then straightforward manipulations show that

$$B(w_0, v - v_0) = 0 \quad \forall w_0 \in V_0.$$

As a consequence,

$$B(e_0, v) = B(e_0, v - v_0) = L(v - v_0) - B(u_0, v - v_0) = L(v - v_0)$$

where use has been made of the standard orthogonality property of the Galerkin approximation. Now,

$$L(v - v_0) = \int_{\omega} f(\mathbf{x}') \{v(\mathbf{x}', d/2) + v(\mathbf{x}', -d/2) - v_0(\mathbf{x}')\} d\mathbf{x}'.$$

The following identity is obtained via the Peano Kernel Theorem,

$$v(\mathbf{x}', d/2) + v(\mathbf{x}', -d/2) - v_0(\mathbf{x}') = \int_{-d/2}^{d/2} K(2x_3/d) \frac{\partial v}{\partial x_3}(\mathbf{x}', x_3) dx_3$$

where

$$K(s) = \frac{1}{b} \int_0^s b(t) dt.$$

Therefore,

$$L(v - v_0) = \int_{\omega} f(\mathbf{x}') \int_{-d/2}^{d/2} K(2x_3/d) \frac{\partial v}{\partial x_3}(\mathbf{x}', x_3) dx_3 d\mathbf{x}'$$

and, with the aid of Cauchy-Schwarz inequalities,

$$|L(v - v_0)|^2 \leq \frac{d}{2} \left[\int_{-1}^1 \frac{1}{a(s)} K(s)^2 ds \right] \|f\|_{L_2(\omega)}^2 \|v\|^2$$

where $\|v\|^2 = B(v, v)$. Gathering results shows that

$$\|e_0\|^2 \leq \frac{d}{2} \left[\int_{-1}^1 \frac{1}{a(s)} K(s)^2 ds \right] \|f\|_{L_2(\omega)}^2.$$

In particular, if the plate is homogeneous so that $a = b = 1$, then $K(s) = s$ and the estimate simplifies to

$$\|e_0\|^2 \leq \frac{d}{3} \|f\|_{L_2(\omega)}^2.$$

These estimates show how the accuracy of the solution of the simplified model (2) depends on the data f , the material properties a and b , and the thickness of the plate. In fact, the error bound is *computable* and provides a numerical estimate for the accuracy of the zeroth order model.

2. Higher Order Plate Models

If the error in the simple plate model fails to meet the required tolerance, then it is necessary to turn to a higher order plate model. Let

$$\{\psi_j\}_{j=0}^{\infty} \subset H^1(-1, 1)$$

be a sequence of linearly independent functions that are dense, and let N be any non-negative integer. The subspace $V_N \subset V$ is defined by

$$V_N = \left\{ v : v = \sum_{j=0}^N \alpha_j(\mathbf{x}) \psi_j \left(\frac{2x_3}{d} \right), \alpha_j \in H_0^1(\omega) \right\}$$

and the model problem is approximated by solving

$$\text{find } u_N \in V_N : B(u_N, v) = L(v) \quad \forall v \in V_N. \quad (3)$$

As before, the problem is formally equivalent to solving a coupled system of PDEs on the mid-plane

$$-B\Delta' \alpha + \frac{4}{d^2} A\alpha = \frac{2}{d} f \text{ in } \omega$$

subject to $\alpha = 0$ on $\partial\omega$, where

$$[A]_{jk} = \int_{-1}^1 a(z) \psi_j'(z) \psi_k'(z) dz; \quad [B]_{jk} = \int_{-1}^1 b(z) \psi_j(z) \psi_k(z) dz$$

and

$$[f]_k = \psi_k(1)f_+(\mathbf{x}') + \psi_k(-1)f_-(\mathbf{x}'); \quad [\alpha]_k = \alpha_k(\mathbf{x}').$$

Thanks to the density property of the sequence $\{\psi_j\}_{j=0}^{\infty}$, the modelling error converges to zero in energy as the order N increases. The selection of the functions $\{\psi_j\}_{j=0}^N$ has been studied by Vogelius and Babuska [1]: define ψ_{2j} , $j = 0, 1, 2, \dots$ recursively by $\psi_{-2} = 0$ and

$$\int_{-1}^1 a(z)\psi'_{2j}(z)v'(z) dz + \int_{-1}^1 b(z)\psi_{2j-2}(z)v(z) dz = \begin{cases} v(1) + v(-1), & \text{if } j = 1 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

and define ψ_{2j+1} , $j = 0, 1, \dots$ by $\psi_{-1} = 0$ and

$$\int_{-1}^1 a(z)\psi'_{2j+1}(z)v'(z) dz + \int_{-1}^1 b(z)\psi_{2j-1}(z)v(z) dz = \begin{cases} v(1) - v(-1), & \text{if } j = 0 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

for all $v \in H^1(-1, 1)$.

In particular, if the coefficient functions a and b are constant then the optimal choice is to take ψ_j to be a polynomial of degree j . More generally, Vogelius and Babuska [2] show that this process generates a dense subspace.

The Gram-Schmidt procedure may be applied to orthogonalize the basis functions so that without loss of generality we may assume

$$\int_{-1}^1 b(z)\psi_j(z)\psi_k(z) dz = \lambda_j \delta_{jk} \quad (6)$$

where δ_{jk} is the Kronecker symbol.

2.1. An A Posteriori Error Estimate for Higher Order Models

It is possible to derive a posteriori error estimates for the error in the higher order plate models [3]. For instance, if the domain is homogeneous so that $a = b = 1$ and the data satisfies $f_+ = f_- = f$ then

$$\| \|u - u_{2N}\| \|^2 \leq \frac{d}{4N + 3} \left\| f - a(1) \frac{\partial u_{2N}}{\partial x_3} \Big|_{d/2} \right\|_{L_2(\omega)}^2. \quad (7)$$

The term on the right hand is the residual in the approximation to the boundary condition on the top (and bottom) surface. In particular, if $N = 0$, then the expression reduces to the bound obtained earlier for the error in the simple plate model.

Babuska and Schwab [3] showed this estimator to have a number of highly desirable properties. In addition to providing an upper bound, if the data is sufficiently regular then the upper bound will tend to the actual error as either the order $N \rightarrow \infty$ or as the thickness $d \rightarrow 0^+$.

2.2. Adaptive Plate Modelling

Typically, the true solution u will have markedly different behaviour through the thickness direction in different parts of the domain ω . For instance, if $f = 1$ then, away from the boundary $\partial\omega$, the variation through the thickness will be virtually constant and could

be well-resolved by a low order model. However, a very high order model would be needed to resolve the boundary layer in the true solution in the neighbourhood of $\partial\omega$. This suggests that there are advantages in using *non-uniform model orders N in different regions of the domain*.

Let $\{\omega_K : K \in \mathcal{P}\}$ be a partitioning of the mid-surface ω into the union of non-overlapping triangles or convex quadrilaterals such that neighbouring elements share a single common edge or vertex. Let $N(K)$ denote the order of the model to be used on the region ω_K , and define $\tilde{V} \subset V$ to be the subspace

$$\tilde{V} = \left\{ v \in V : v|_{\omega_K} \in H^1(\omega_K) \otimes \text{span}\{\psi_k : k = 0, \dots, N(K)\} \quad K \in \mathcal{P} \right\}.$$

The solution \tilde{u} of the non-uniform model is obtained by solving the problem: find $\tilde{u} \in \tilde{V}$ such that

$$B(\tilde{u}, v) = L(v) \quad \forall v \in \tilde{V}.$$

The selection of the model orders $\{N(K) : K \in \mathcal{P}\}$ will depend on the particular features of the specific problem under consideration. Adaptive, hierarchic modelling consists of starting from a low order model and then designing a sequence of non-uniform models for a given problem based on assessing the accuracy of the current approximation on each of the elements ω_K and locally increasing the order of the plate model in those elements where the error is too high. At the heart of the adaptive algorithm is an a posteriori estimator indicating the error on each element ω_K .

3. An A Posteriori Estimator for Adaptive Hierarchic Modelling

The purpose of this section is to describe a technique for estimating the error in a plate model of general, non-uniform order. To date, there have been no reliable error estimators for non-uniform order plate models. The basic idea behind the estimator given here stems from an idea in [4] for estimating the errors in fully discrete hierarchic models (where the error in the finite element approximation of the dimensionally reduced model is also taken into account). The estimator requires the solution of an auxiliary residual problem posed over a local patch consisting of a small number of elements, and is similar in spirit to an estimator proposed by Babuska and Rheinboldt [5] for finite element approximation.

3.1. Preliminaries

Let Ψ denote the vertices of the elements in the partitioning \mathcal{P} of the domain. Associated with each vertex \mathbf{x}_n is a first order (finite element type) Lagrange basis function θ_n defined by the conditions

$$\theta_n(\mathbf{x}_m) = \delta_{mn}, \quad \forall m, n \in \Psi.$$

The basis functions $\{\theta_n : n \in \Psi\}$ form a partition of unity on ω ,

$$\sum_{n \in \Psi} \theta_n(\mathbf{x}) = 1, \quad \mathbf{x} \in \omega$$

and satisfy

$$0 \leq \theta_n(\mathbf{x}) \leq 1, \quad \mathbf{x} \in \omega.$$

Moreover, the support ω_n of the basis function θ_n consists of the elements ω_K having a vertex at \mathbf{x}_n ,

$$\omega_n = \text{supp}(\theta_n) = \bigcup \{\omega_K : \mathbf{x}_n \in \omega_K\}.$$

If a family of partitions is used in the adaptive modelling procedure, then it will be assumed that the number of elements forming the support of a basis function is uniformly bounded over the whole family.

The basis functions are constructed using first order polynomials in the usual (finite element) way. Consequently, the gradients may be bounded as

$$|\nabla' \theta_n(\mathbf{x})| \leq \Delta_n^{-1}$$

where Δ_n depends on the dimensions of the elements forming the support of θ_n . For instance,

- if ω is an interval then Δ_n is equal to the size of the smallest element on which θ_n does not vanish identically;
- if $\omega \subset \mathbb{R}^2$ is partitioned into triangular or rectangular elements then $\sqrt{2}\Delta_n$ equals the length of the shortest edge that is connected to the vertex \mathbf{x}_n .

A variation of the arguments leading to the standard Poincaré inequality shows that there exists a positive constant C_S such that

$$\int_{-1}^1 b(s)v(s)^2 ds \leq C_S^2 \int_{-1}^1 a(s)v'(s)^2 ds$$

for all $v \in H^1(-1, 1)$ such that $\int_{-1}^1 b(s)v(s) ds = 0$. In particular, for a homogeneous material ($a = b = 1$) then $C_S = 1/\pi$. The following result will also prove useful:

Lemma 1 *For any $v \in V$ there exists $\bar{v} \in \bar{V}$ such that for all $n \in \Psi$*

$$|||\theta_n(v - \bar{v})||| \leq \sqrt{2} \left(1 + \left(\frac{C_S d}{2\Delta_n} \right)^2 \right)^{1/2} |||v|||_{\omega_n}.$$

Proof. Let $v \in V$ be given and let \bar{N} be the largest global modelling order,

$$\bar{N} = \max\{N(K) : K \in \mathcal{P}\}.$$

Without loss of generality, it may be assumed that the basis functions $\{\psi_j : j = 0, \dots, \bar{N}\}$ are selected so that $\psi_0 = 1$ and (by solving a generalized eigenvalue problem)

$$(b\psi_j, \psi_k) = (a\psi'_j, \psi'_k) = 0 \text{ for } j \neq k$$

where (\cdot, \cdot) denotes the inner product on $L_2(-1, 1)$. Let the expansion of the function v in terms of these basis functions be

$$v = \sum_{j=0}^{\bar{N}} \alpha_j \psi_j$$

where $\alpha_j \in H_0^1(\omega)$. A simple computation reveals that

$$\|v\|_{\omega_n}^2 = \sum_{j=0}^{\bar{N}} \left\{ \frac{d}{2} \|\nabla' \alpha_j\|_{L_2(\omega_n)}^2 (b\psi_j, \psi_j) + \frac{2}{d} \|\alpha_j\|_{L_2(\omega_n)}^2 (a\psi'_j, \psi'_j) \right\}.$$

Choose $\tilde{v} = \alpha_0 \psi_0 \in V_0 \subset \tilde{V}$. Then

$$\|\theta_n(v - \tilde{v})\|_{\omega_n}^2 = \sum_{j=1}^{\bar{N}} \left\{ \frac{d}{2} \|\nabla'(\theta_n \alpha_j)\|_{L_2(\omega)}^2 (b\psi_j, \psi_j) + \frac{2}{d} \|\theta_n \alpha_j\|_{L_2(\omega)}^2 (a\psi'_j, \psi'_j) \right\}.$$

The properties of the Lagrange basis functions imply that

$$\|\theta_n \alpha_j\|_{L_2(\omega)} \leq \|\alpha_j\|_{L_2(\omega_n)}$$

and

$$\begin{aligned} \|\nabla'(\theta_n \alpha_j)\|_{L_2(\omega)}^2 &\leq \left(\|\nabla' \alpha_j\|_{L_2(\omega_n)} + \Delta_n^{-1} \|\alpha_j\|_{L_2(\omega_n)} \right)^2 \\ &\leq \left(1 + \left(\frac{dC_S}{2\Delta_n} \right)^2 \right) \left(\|\nabla' \alpha_j\|_{L_2(\omega_n)}^2 + \left(\frac{2}{dC_S} \right)^2 \|\alpha_j\|_{L_2(\omega_n)}^2 \right). \end{aligned}$$

Therefore, for $j > 0$,

$$\begin{aligned} &\frac{d}{2} \|\nabla'(\theta_n \alpha_j)\|_{L_2(\omega)}^2 (b\psi_j, \psi_j) \\ &\leq \left(1 + \left(\frac{dC_S}{2\Delta_n} \right)^2 \right) \left(\frac{d}{2} (b\psi_j, \psi_j) \|\nabla' \alpha_j\|_{L_2(\omega_n)}^2 + \frac{2}{d} (a\psi'_j, \psi'_j) \|\alpha_j\|_{L_2(\omega_n)}^2 \right). \end{aligned}$$

The result follows at once by summing over both $j > 0$ and $n \in \Psi$, and recalling the expression for $\|v\|_{\omega}$.

3.2. Local Residual Problems

Let $n \in \Psi$ be fixed and define the space V^n to be

$$V^n = \{v \in V : \text{supp}(v) \subset \omega_n\}.$$

The largest local model order on the support ω_n of θ_n is denoted by

$$\bar{N}(n) = \max\{N(K) : \omega_K \subset \omega_n\}$$

with the associated local space \bar{V}^n is defined as

$$\bar{V}^n = \left\{ v \in V^n : v = \sum_{j=0}^{\bar{N}(n)} \beta_j \psi_j \right\}.$$

The local energy projection of the true error \tilde{e} onto the space \bar{V}^n is denoted by ϕ_n and is characterized by the condition: find $\phi_n \in \bar{V}^n$ such that

$$B(\phi_n, v) = B(\tilde{e}, v) \quad \forall v \in \bar{V}^n.$$

As a matter of fact, the function ϕ_n may be defined without explicit use of the true solution u since, for any $v \in \bar{V}^n \subset V$,

$$B(\tilde{e}, v) = B(u, v) - B(\tilde{u}, v) = L(v) - B(\tilde{u}, v),$$

and so ϕ_n may be obtained by solving the local residual problem

$$B(\phi_n, v) = L(v) - B(\tilde{u}, v) \quad \forall v \in \bar{V}^n. \tag{8}$$

3.3. Derivation of Error Estimator

The local projections ϕ_n are needed for the a posteriori error estimator that will now be derived. Let $v \in V$ be arbitrary and let \tilde{v} be chosen as in Lemma 1, then, thanks to the Galerkin orthogonality property,

$$B(\tilde{e}, v) = B(\tilde{e}, v - \tilde{v}) = B(\tilde{e}, w)$$

where $w = v - \tilde{v}$. Moreover, since the Lagrange basis functions form a partition of unity,

$$B(\tilde{e}, w) = B(\tilde{e}, w \sum_{n \in \Psi} \theta_n) = \sum_{n \in \Psi} B(\tilde{e}, w_n)$$

where $w_n = w\theta_n \in V^n$. Following [4], introduce the mapping $Q_n : V^n \mapsto \bar{V}^n$ given by the rule

$$Q_n v(\mathbf{x}', z) = \frac{2}{d} \sum_{j=0}^{\bar{N}(n)} \frac{1}{\lambda_j} \psi_j(2z/d) \int_{-d/2}^{d/2} b(2s/d) v(\mathbf{x}', s) \psi_j(2s/d) ds$$

where λ_j was defined in equation (6), and define $R_n : V^n \mapsto \bar{V}^n$ by

$$w_n = Q_n w_n + R_n w_n.$$

Now,

$$B(\tilde{e}, w_n) = B(\tilde{e}, R_n w_n) + B(\tilde{e}, Q_n w_n)$$

and then,

$$B(\tilde{e}, R_n w_n) = L(R_n w_n) - B(\tilde{u}, R_n w_n)$$

and

$$B(\tilde{e}, Q_n w_n) = L(Q_n w_n) - B(\tilde{u}, Q_n w_n) = B(\phi_n, Q_n w_n),$$

so that

$$B(\tilde{e}, w_n) = \underbrace{L(R_n w_n) - B(\tilde{u} + \phi_n, R_n w_n)}_{=:T_1} + \underbrace{B(\phi_n, w_n)}_{=:T_2}.$$

Integrating by parts in the thickness direction leads to

$$T_1 = \int_{\omega_n} \{R_n w_n(\mathbf{x}', d/2) r_+(\mathbf{x}') + R_n w_n(\mathbf{x}', -d/2) r_-(\mathbf{x}')\} d\mathbf{x}'$$

where

$$r_{\pm} = f_{\pm} \mp a(\pm 1) \left. \frac{\partial}{\partial z} (\tilde{u} + \phi_n) \right|_{\pm d/2}. \quad (9)$$

Applying Corollary 7 from [4] gives the following bound:

$$|T_1| \leq \left\{ \sqrt{\frac{d}{8} \kappa^+ (\bar{N}(n))} \|r_+ + r_-\|_{L_2(\omega_n)} + \sqrt{\frac{d}{8} \kappa^- (\bar{N}(n))} \|r_+ - r_-\|_{L_2(\omega_n)} \right\} M(w_n)$$

where the functions κ^\pm are defined in [4] (and are discussed later), and

$$M(w_n)^2 = \int_{\omega_n} \int_{-d/2}^{d/2} a(2z/d) \left| \frac{\partial w_n}{\partial z} \right|^2 dz dx'.$$

Observing ψ_0 is constant and using properties of the Lagrange basis functions shows

$$\left| \frac{\partial w_n}{\partial z} \right| = \left| \theta_n \frac{\partial w}{\partial z} \right| \leq \left| \frac{\partial w}{\partial z} \right| = \left| \frac{\partial(v - \bar{v})}{\partial z} \right| = \left| \frac{\partial v}{\partial z} \right|$$

and in turn reveals

$$|M(w_n)| \leq \|v\|_{\omega_n}.$$

Gathering these results gives

$$|T_1| \leq \left\{ \sqrt{\frac{d}{8} \kappa^+ (\bar{N}(n))} \|r_+ + r_-\|_{L_2(\omega_n)} + \sqrt{\frac{d}{8} \kappa^- (\bar{N}(n))} \|r_+ - r_-\|_{L_2(\omega_n)} \right\} \|v\|_{\omega_n}.$$

The second term is bounded more easily using Lemma 1 as follows

$$|T_2| \leq \|\phi_n\| \|w_n\| = \|\phi_n\| \|\theta_n(v - \bar{v})\| \leq \sqrt{2} \left(1 + \left(\frac{C_S d}{2\Delta_n} \right)^2 \right)^{1/2} \|\phi_n\| \|v\|_{\omega_n}.$$

Collecting these results

$$|B(\bar{e}, w_n)| \leq \eta_n \|v\|_{\omega_n}$$

where

$$\begin{aligned} \eta_n &= \sqrt{\frac{d}{8} \kappa^+ (\bar{N}(n))} \|r_+ + r_-\|_{L_2(\omega_n)} + \sqrt{\frac{d}{8} \kappa^- (\bar{N}(n))} \|r_+ - r_-\|_{L_2(\omega_n)} \\ &\quad + \sqrt{2} \left(1 + \left(\frac{C_S d}{2\Delta_n} \right)^2 \right)^{1/2} \|\phi_n\|. \end{aligned} \quad (10)$$

Now,

$$\sum_{n \in \Psi} \|v\|_{\omega_n}^2 = \sum_{n \in \Psi} \sum_{\omega_K \subset \omega_n} \|v\|_{\omega_K}^2 \leq \tau \sum_{K \in \mathcal{P}} \|v\|_K^2 = \tau \|v\|^2$$

where τ is the number of vertices of each element. Therefore,

$$|B(\bar{e}, v)| \leq \sum_{n \in \Psi} |B(\bar{e}, w_n)| \leq \left(\tau \sum_{n \in \Psi} \eta_n^2 \right)^{1/2} \|v\|$$

and hence we have proved:

Theorem 2 *Let \bar{e} denote the error in the non-uniform hierarchic model associated with the subspace $\bar{V} \subset V$. Then,*

$$\|\bar{e}\|^2 \leq \tau \sum_{n \in \Psi} \eta_n^2.$$

where η_n is defined in (10), τ is the number of vertices of the elements, r_\pm are defined in (9), ϕ_n denotes the solution of the local residual problem (8) and κ^\pm are given in Corollary 7 of [4].

A few remarks are in order:

1. The bound suggests using η_n^2 as an estimator for $\|\bar{e}\|^2$. However, since the patches overlap, resulting in each part of ω being counted τ times, we shall instead base the estimator on $\tau^{-1}\eta_n^2$.
2. For a homogeneous material ($a = b = 1$) the functions κ^\pm are given by

$$\kappa^+(N) = \begin{cases} 2/(2N + 3), & \text{if } N \text{ is even} \\ 2/(2N + 1), & \text{if } N \text{ is odd} \end{cases}$$

and

$$\kappa^-(N) = \begin{cases} 2/(2N + 1), & \text{if } N \text{ is even} \\ 2/(2N + 3), & \text{if } N \text{ is odd} \end{cases}$$

3. If $f_+ = f_- = f$ and a uniform model of order $2N$ is used then the solution of the local residual problems is $\phi_n = 0$. Consequently, the residuals satisfy $r_+ = r_- = f - a(1)\partial u_{2N}/\partial x_3$. In addition, if the material is homogeneous $a = b = 1$ then the estimator reduces to (7).
4. The factor involving the term $C_S d/2\Delta_n$ may blow up if the the elements are small compared with the thickness of the domain. Consequently, this suggests that one should not attempt to vary the model order over a scale significantly smaller than the thickness of the domain.
5. The estimator was based on solving residual problems formulated over patches composed of the elements. One obvious way to simplify the computation would be to solve residual problems posed on single elements. The local projection ϕ_K on the element ω_K would be characterized by: find $\phi_K \in \bar{V}^K$ such that

$$B(\phi_K, v) = L(v) - B(\bar{u}, v) \quad \forall v \in \bar{V}^K \quad (11)$$

where \bar{V}^K is defined in a similar way to \bar{V}^n with the element ω_K replacing the patch ω_n . The estimator would then be defined using the same expressions given above with the patches ω_n replaced by single elements ω_K . (Of course, the theory given earlier is no longer applicable.)

In fact, this estimator simplifies drastically if one notices that the solution of the local residual problem ϕ_K *vanishes* (since $\bar{V}^K \subset \bar{V}$ and the right hand side of (11) is identically zero). Therefore, the expression for the error indicator on element ω_K reduces to

$$\eta_K = \sqrt{\frac{d}{8}\kappa^+(N(K))} \|\bar{r}_+ + \bar{r}_-\|_{L_2(\omega_K)} + \sqrt{\frac{d}{8}\kappa^-(N(K))} \|\bar{r}_+ - \bar{r}_-\|_{L_2(\omega_K)} \quad (12)$$

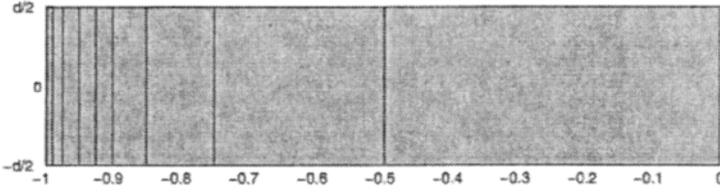


Figure 2. Domain and partition used to construct adaptive hierarchic model for numerical example.

where

$$\tilde{r}_{\pm} = f_{\pm} \mp a(\pm 1) \left. \frac{\partial \tilde{u}}{\partial x_3} \right|_{\pm d/2}.$$

This method for estimating the error in non-uniform hierarchic models was suggested in Babuska et al. [6].

4. Numerical Examples

The performance of the indicators described will be illustrated by solving the following problem:

$$-\Delta u = 0 \text{ in } \Omega = (-1, 1) \times (-d/2, d/2)$$

subject to $u = 0$ on the lateral boundary with prescribed heat fluxes $f_+ = f_- = 1$ on the upper and lower faces. The domain ω is partitioned symmetrically into 20 subdomains with breakpoints located at -1.00, -0.998, -0.990, -0.975, -0.950, -0.925, -0.900, -0.850, -0.750, -0.500, 0.00 as shown in Figure 2. The exact solution contains a boundary layer at $x = \pm 1$ whose accurate resolution requires the use of very high local model orders. Conversely, away from the boundary, the solution may be resolved with the use of relatively low model orders.

Table 1 shows the sequence of hierarchical models used to approximate the problem on a domain of thickness $d = 0.2$. The solution of the dimensionally reduced systems posed on the mid-surface is accomplished using a p -version finite element method designed so that the discretization error is negligible in comparison with the modelling error. The performance of three error estimators is shown in Table 1. In particular, the ratio of the estimated error to the actual error is shown for the following estimators:

1. η_K denotes the estimator based on (12) proposed in [6] obtained when the local residual problems are posed over single elements;
2. η_n denotes the error estimator based on (10) analysed in the previous section and shown to yield an upper bound;

Table 1

Details of sequence of non-uniform hierarchic models. The columns refer to each of the nine different hierarchic models used. Rows 1-10 show the model order used on each of the ten elements present in all the models. The final three rows show ratio of the estimated error to the actual error for three different error indicators.

		Adaptive Hierarchic Model								
		1	2	3	4	5	6	7	8	9
Element Number	1	4	6	8	10	12	12	14	14	18
	2	4	6	8	10	10	12	10	12	18
	3	4	6	6	6	6	6	6	6	16
	4	4	4	4	4	4	4	4	4	12
	5	4	4	4	4	4	4	4	4	6
	6	4	4	4	4	4	4	4	4	4
	7	2	2	2	2	2	2	2	2	2
	8	2	2	2	2	2	2	2	2	2
	9	2	2	2	2	2	2	2	2	2
	10	2	2	2	2	2	2	2	2	2
Effectivity Index	η_K	1.75	1.74	1.52	1.31	1.21	1.13	1.11	1.04	0.91
	η_n	1.70	1.68	1.51	1.35	1.30	1.25	1.25	1.20	1.06
	η_n^*	1.70	1.66	1.45	1.30	1.16	1.19	1.10	1.10	1.01

3. η_n^* denotes the error estimator based on (10) when the extra term $|||\phi_n|||$ is excluded.

The results confirm that the estimator η_n provides an upper bound on the actual error as the theory suggests.

REFERENCES

1. M. Vogelius and I. Babuska. On a dimensional reduction method. I. The optimal selection of basis functions. *Math. Comp.*, 37:31-46, 1981.
2. M. Vogelius and I. Babuska. On a dimensional reduction method. II. Some approximation theoretic results. *Math. Comp.*, 37:47-68, 1981.
3. I. Babuska and C. Schwab. A posteriori error estimation for hierarchic models of elliptic boundary value problems on thin domains. *SIAM J. Numer. Anal.*, 33:221-246, 1996.
4. M. Ainsworth. A posteriori error estimators for fully discrete hierarchical modelling on thin domains. *Numer. Math.*, To appear.
5. I. Babuska and W. C. Rheinboldt. Error estimates for adaptive finite element computations. *SIAM J. Numer. Anal.*, 18:736-754, 1978.
6. I. Babuska, I. Lee, and C. Schwab. On the a posteriori estimation of the modelling error for the heat conduction in a plate and its use for adaptive hierarchic modelling. *Appl. Numer. Math.*, 14:5-21, 1994.

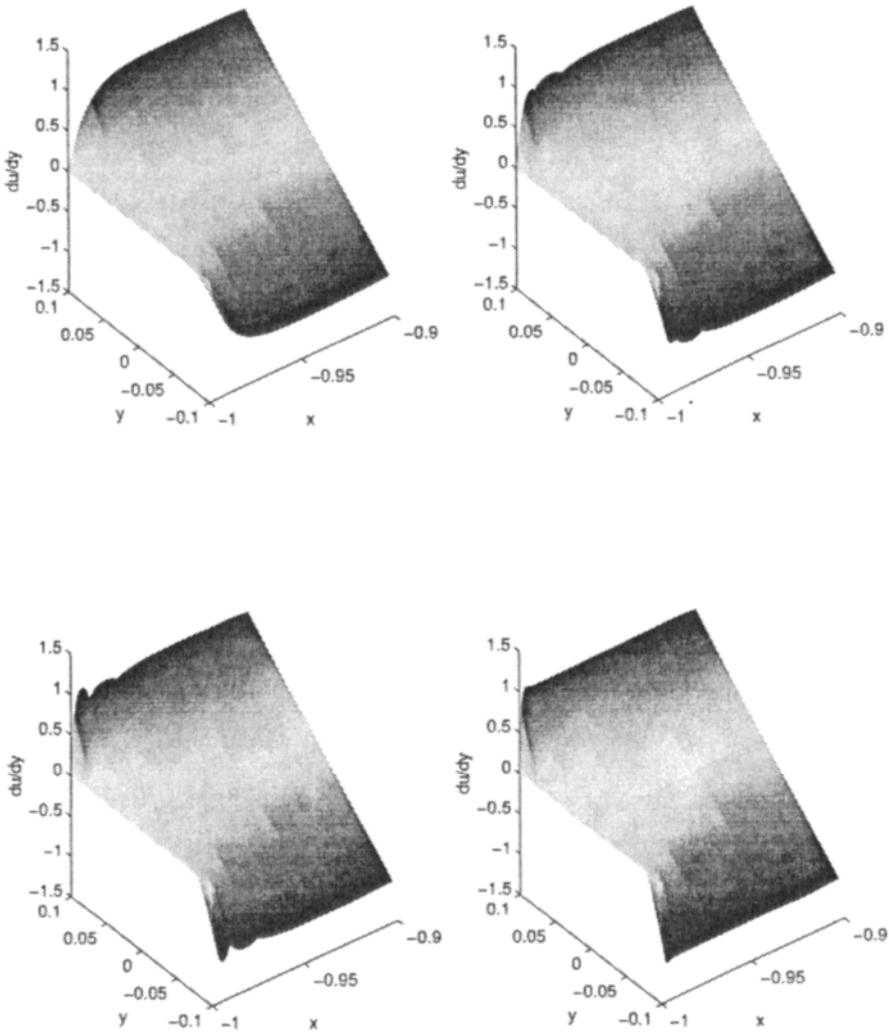


Figure 3. Plots showing the resolution of the derivative of the solution in the thickness direction near the edge of the plate of thickness $d = 0.2$ for Models 1, 3, 8 and 9. The details of the local model orders in each case are shown in Table 1.

A two-scale strategy and a posteriori error estimation for modeling heterogeneous structures

Nicolas Moës, J. Tinsley Oden and Kumar Vemaganti *

The Texas Institute for Computational and Applied Mathematics,
The University of Texas at Austin, Austin, TX 78712, U.S.A.

In this paper, we introduce a two-scale strategy to model scalar diffusion problems in heterogeneous structures and estimate a posteriori the quality of the model obtained. The modeling strategy involves the following steps: partition of the domain into subdomains, description of the temperature field as a “smooth” part plus a “correction” on each subdomain, construction of a prolongation operator on each subdomain by minimizing the local potential energy with respect to the correction part and, finally, definition of the “homogenized” problem through the minimization of the global potential energy expressed in term of the smooth part of the temperature field. By choosing properly the boundary conditions in the local minimization step, the strategy yields a continuous temperature field. To estimate the quality of this approximate solution, a second solution is locally built such that the heat flux field is in thermal equilibrium. Then, the energy in the difference of the two solutions built is linked to the exact energy error in these solutions using the Prager–Synge hypercircle theorem. Finally, the results are revisited in the case of an approximate computation of the prolongation operator and a 2-D numerical experiment is studied to analyze the proposed approach.

1. INTRODUCTION

In recent times, the use of the notion of hierarchical modeling has arisen as an approach to very complex problems involving the analysis of heterogeneous media. One approach, advanced in [1,2], is to construct an adaptive modeling strategy called the Homogenized Dirichlet Projection Method (HDPM) in which the various fine-scale constituents in a heterogeneous media are present in the characterization of the material constants, and a homogenized solution is calculated. Once an estimate is obtained, only the fine-scale information sufficient to produce solutions within a preset level of accuracy is used, with the result that models containing orders-of-magnitude fewer degrees of freedom can be used to obtain acceptable resolutions of fine-scale phenomena. In these approaches, the homogenization step is regarded as a process provided by a black-box-homogenization unit outside the HDPM algorithm.

*The authors gratefully acknowledges the support of this work by the U.S. Office of Naval Research under Grant N00014-95-1-0401 and the National Science Foundation under grant ECS-9422707. The assistance of Dr Carter Edwards in providing a C++ library useful in our calculations is also gratefully acknowledged.

In the present paper, we explore a slightly different strategy for hierarchical modeling in which the computation of an appropriate homogenized model is an intrinsic part of the hierarchical modeling process. Here a coarse scale model is produced which effectively projects fine-scale features onto a finite dimensional space. A prolongation operator of the coarse scale approximation into the fine scale solution is determined so as to minimize the potential energy based on the fine-scale features. A global coarse solution is likewise obtained through a minimization process. The boundary conditions for the local minimization are chosen so that the strategy yields a continuous temperature field. Then, to estimate the modeling error in this solution, we built locally a second solution for which the heat flux is in equilibrium. The energy in the difference of the two solutions may then be related to the modeling error.

Our method exploits a number of strategies proposed in other applications. For example, the relation mentioned above is provided by the classical Prager-Synge hypercircle arguments [3]; the second solution is built using Ladevèze equilibration technique [4]; the domain decomposition strategy is reminiscent of those proposed recently by Ladevèze and Dureisseix [5], and a similar approach to homogenization techniques for porous media is being developed by Arbogast [6] using mixed finite element techniques.

Following this introduction, we outline the details of the approach on a model class of scalar diffusion problems in n dimensions. We give a full development of techniques for constructing the coarse-scale homogenized models, and well-defined prolongations to the fine-scale model. This is followed by the local calculation of the equilibrated solution which is then used for a posteriori error estimation purposes. The practically important case of an approximate computation of the prolongation operator is studied and the meaning and validity of the results in this framework are analyzed. Finally, the technique is demonstrated through a representative numerical experiments and a summary is given of the major conclusions of this investigation.

2. THE MODEL PROBLEM

Vectors will be bold and scalars italic. For instance, the temperature will be denoted by u and the heat flux and the temperature gradient will be denoted by σ and ϵ , respectively. More complex operators will be uppercase (e.g. E for the conductivity tensor).

We consider a material body composed of a linearly-conductive material in thermal equilibrium under the action of given volumetric and surface heat sources, f_d and t_d respectively, see Fig. 1. The domain, $\Omega \subset \mathbb{R}^n$, $n = 1, 2, 3$, occupied by the material body is considered bounded and regular with a simply-connected domain with Lipschitz boundary Γ . The boundary Γ consists of a portion Γ_u where the temperature u_d is given and a portion Γ_t where surface heat sources t_d are prescribed, $\Gamma = \overline{\Gamma_u \cup \Gamma_t}$, $\Gamma_u \cap \Gamma_t = \emptyset$. The external sources are assumed to be regular and, in particular $f_d \in L^2(\Omega)$, $t_d \in L^2(\Gamma_t)$. The problem to be solved on Ω is to find a triple (u, ϵ, σ) such that the temperature constraints, the equilibrium equation and the constitutive law hold:

$$\left. \begin{aligned} \epsilon &= \nabla u && \text{on } \Omega, u = u_d \text{ on } \Gamma_u \\ \operatorname{div} \sigma &= -f_d && \text{on } \Omega, \sigma \cdot n = t_d \text{ on } \Gamma_t \\ \sigma &= E\epsilon && \text{on } \Omega \end{aligned} \right\} \quad (1)$$

\mathbf{n} is the outward normal to the boundary. The conductivity tensor E is a function of the position \mathbf{x} i.e. $E = E(\mathbf{x}) \in L^\infty(\Omega)^{n \times n}$ and E is symmetric and elliptic; i.e. for a.e. $\mathbf{x} \in \bar{\Omega}$, there are constants $\alpha_0, \alpha_1 > 0$, such that $\alpha_0 \mathbf{a}^t \mathbf{a} \leq \mathbf{a}^t E(\mathbf{x}) \mathbf{a} \leq \alpha_1 \mathbf{a}^t \mathbf{a}$ for any $\mathbf{a} \in \mathbb{R}^n$. More precisely, problem (1) corresponds exactly to a heat transfer problem if σ , f_d and t_d are the heat flux, the volumetric and surface prescribed heat sources, respectively (up to a sign).

Eliminating ϵ and σ in (1) we also have the classical heat-conduction formulation,

$$-\operatorname{div}(E\nabla u) = f_d \text{ on } \Omega, \quad u = u_d \text{ on } \Gamma_u, \quad E\nabla u \cdot \mathbf{n} = t_d \text{ on } \Gamma_t \quad (2)$$

A weak formulation of problem (2) is to find $u \in \mathcal{V}_d$ such that

$$B(u, v) = F(v) \quad \forall v \in \mathcal{V}_0 \quad (3)$$

where

$$\mathcal{V}_d = \{v \in \mathcal{V} : v = u_d \text{ on } \Gamma_u\}, \quad \mathcal{V}_0 = \{v \in \mathcal{V} : v = 0 \text{ on } \Gamma_u\}, \quad \mathcal{V} = H^1(\Omega)$$

$$B(u, v) = \int_{\Omega} E\epsilon(u) \cdot \epsilon(v) \, dx, \quad F(v) = \int_{\Omega} f_d v \, dx + \int_{\Gamma_t} t_d v \, ds$$

and we have used the notation $\epsilon(\bar{v}) = \nabla v$.

Owing to the assumed properties of E , $B(\cdot, \cdot)$ is symmetric positive and induces the energy norm on \mathcal{V} :

$$\|v\|_{E(\Omega)} = \{B(v, v)\}^{1/2} \quad (4)$$

Problem (3) possesses a unique solution u for given data and, as it is well known, this solution also minimizes the total potential “energy” J :

$$J(u) = \inf_{v \in \mathcal{V}_d} J(v), \quad J(v) = \frac{1}{2} B(v, v) - F(v) \quad (5)$$

3. TWO-SCALE MODELING STRATEGY

We are concerned with classes of problems in which the material body Ω possesses complex microstructures in which case $E = E(\mathbf{x})$ exhibits highly heterogeneous structure with possible nonperiodic oscillations at one or more scales smaller than the characteristic length of Ω itself. The present analysis focuses on cases in which the resolution of these fine scale features represents a computational problem of very large size, but within the capabilities of contemporary massively parallel computers; e.g., fine scale features may be of the order of $10^{-3}L$, L being a characteristic length scale of Ω . Our approach to the resolutions of the effects of the fine scale consists in the following steps.

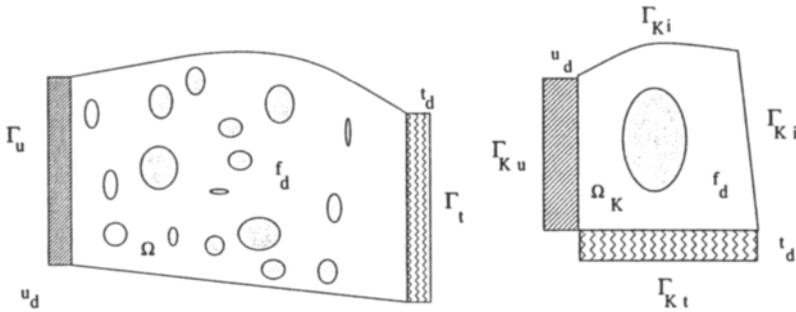


Figure 1. Notations for the model problem Figure 2. Notations for a subdomain

3.1. Partition of the domain

We introduce a partition \mathcal{P}_0 of the domain Ω into N_0 subdomains Ω_K :

$$\bar{\Omega} = \cup_{K=1}^{N_0} \bar{\Omega}_K, \quad \Omega_K \cap \Omega_L = \emptyset, \quad K \neq L \tag{6}$$

The boundary, Γ_K , of each subdomain is decomposed into the part of the boundary on which the temperature is prescribed, Γ_{Ku} , the part of the boundary on which surface heat sources are prescribed, Γ_{Kt} , and the remaining interior part, Γ_{Ki} , Fig. 2:

$$\bar{\Gamma}_K = \overline{\Gamma_{Ku} \cup \Gamma_{Kt} \cup \Gamma_{Ki}} \tag{7}$$

where

$$\Gamma_{Ku} = \Gamma_u \cap \Gamma_K, \quad \Gamma_{Kt} = \Gamma_t \cap \Gamma_K, \quad \Gamma_{Ki} = \Gamma_K \setminus (\Gamma_{Ku} \cup \Gamma_{Kt}) \tag{8}$$

We assume that Ω_K contains no holes or cracks and that its boundary does not cross any holes or cracks of Ω . These cases need special attention and will not be considered in this paper.

3.2. Decomposition of the spaces

On each subdomains Ω_K , we shall define the space $\mathcal{V}_K = H^1(\Omega_K)$ which is decomposed into the spaces \mathcal{V}_K^M and \mathcal{W}_K such that the following conditions hold:

- \mathcal{V}_K^M is a finite dimensional subspace of \mathcal{V}_K whose dimension is $dim \mathcal{V}_K^M = M \geq 1$. We assume we have a sequence of nested spaces everywhere dense in \mathcal{V}_K

$$\mathcal{V}_K^M \subset \mathcal{V}_K^{M+1} \subset \dots \subset \mathcal{V}_K; \quad \overline{\cup_{M>0} \mathcal{V}_K^M} = \mathcal{V}_K \tag{9}$$

- For any dimension $M \geq 1$, \mathcal{V}_K^M must at least contain all the constant temperature fields defined on Ω_K i.e.

$$\forall M \geq 1: \mathcal{R}_K \subset \mathcal{V}_K^M, \quad \mathcal{R}_K = \{v \in \mathcal{V}_K : \epsilon(v) = 0\} \tag{10}$$

- \mathcal{W}_K is a subspace of \mathcal{V}_K such that

$$\mathcal{V}_K = \mathcal{V}_K^M \oplus \mathcal{W}_K \quad (11)$$

We understand this direct sum condition in the following *algebraic* sense

$$\mathcal{V}_K^M \cap \mathcal{W}_K = \{0\} \quad (12)$$

$$\forall v^M \in \mathcal{V}_K^M, w \in \mathcal{W}_K : v^M + w \in \mathcal{V}_K \quad (13)$$

$$\forall v \in \mathcal{V}_K : \exists ! v^M \in \mathcal{V}_K^M, w \in \mathcal{W}_K : v = v^M + w \quad (14)$$

Note that for a given choice of \mathcal{V}_K^M , the choice of \mathcal{W}_K satisfying the previous conditions is not unique

Example An example of such a decomposition of \mathcal{V}_K is provided by the case in which \mathcal{V}_K^M is the space of all the polynomials of order $\leq M$ defined on Ω_K , and \mathcal{W}_K is the orthogonal complement in the $H^1(\Omega_K)$ norm, all the rigid body modes being excluded. Note that this example is particular since orthogonality is not required.

$$\mathcal{V}_K^M = \{v \in P_M(\Omega_K)\}$$

$$\mathcal{W}_K = \{w \in \mathcal{V}_K : \int_{\Omega_K} \nabla w \cdot \nabla v \, dx = 0 \quad \forall v \in \mathcal{V}_K^M, \int_{\Omega_K} w r \, dx = 0 \quad \forall r \in \mathcal{R}_K\}$$

We note that the finite dimensional space \mathcal{V}_K^M characterizes a “coarse scale” representation of the solution for which we will eventually define an *homogenized model*. On the other hand, \mathcal{W}_K describes a correction space to take into account the fine scale features of the solution in Ω_K .

3.3. Definition of the prolongation operator

Our next goal is to establish a connection between the coarse scale components of \mathcal{V}_K and the fine scale components of \mathcal{V}_K . This is accomplished if a prolongation operator P_K can be constructed by mapping \mathcal{V}_K^M into \mathcal{V}_K . In particular, we seek a prolongation of the form

$$P_K : \mathcal{V}_K^M \rightarrow \mathcal{V}_K : v^M \rightarrow v^M + C_K v^M \quad (15)$$

$$C_K : \mathcal{V}_K^M \rightarrow \mathcal{W}_K : v^M \rightarrow w(v^M) \quad (16)$$

where C_K is a correction operator.

3.4. Local construction of a prolongation

A natural way to construct the prolongation operator is to choose a perturbation w in \mathcal{W}_K which minimizes the total potential energy over Ω_K . Thus, we define $w = w^{\mathcal{D}}$ by

$$J_K(v^M + w^{\mathcal{D}}(v^M)) = \inf_{w \in \mathcal{W}_K^{\mathcal{D}}} J_K(v^M + w) \quad (17)$$

where $\mathcal{W}_K^{\mathcal{D}}$ is

$$\mathcal{W}_K^{\mathcal{D}} = \{w \in \mathcal{W}_K : w = 0 \text{ on } \Gamma_{Ku} \cup \Gamma_{Ki}\} \quad (18)$$

and

$$J_K(v) = \frac{1}{2} B_K(v, v) - F_K(v) \quad (19)$$

$$B_K(u, v) = \int_{\Omega_K} E \epsilon(u) \cdot \epsilon(v) \, dx, \quad F_K(v) = \int_{\Omega_K} f_d v \, dx + \int_{\Gamma_{Ki}} t_d v \, ds \quad (20)$$

The \mathcal{D} superscript used in the correction solution stresses the fact that we are considering Dirichlet type boundary condition on Γ_{Ki} in the local minimization problem. The corresponding correction and prolongation operators are denoted by $C_K^{\mathcal{D}}$ and $P_K^{\mathcal{D}}$, respectively. We shall see later that another useful choice of boundary condition is the Neumann type that will be denoted by the superscript \mathcal{N} . Concerning the imposed temperature, we assume that they may be taken into account exactly in the \mathcal{V}_K^M space, i.e. $\mathcal{V}_{K,d} \cap \mathcal{V}_K^M \neq \{0\}$ where

$$\mathcal{V}_{K,d} = \{v \in \mathcal{V}_K : v = u_d \text{ on } \Gamma_{Ku}\} \quad (21)$$

3.5. Definition of the homogenized problem

Knowing the prolongation operator, we can define the “homogenized” potential energy on each subdomain Ω_K in the following way:

$$J_K^{\text{hom},\mathcal{D}} : \mathcal{V}_K^M \rightarrow \mathbb{R} : v^M \rightarrow J_K^{\text{hom},\mathcal{D}}(v^M) = J_K(P_K^{\mathcal{D}} v^M) \quad (22)$$

Then we can sum these local potential energy to obtain the global “homogenized” potential energy on Ω

$$J^{\text{hom},\mathcal{D}} : \mathcal{V}^M \rightarrow \mathbb{R} : v^M \rightarrow J^{\text{hom},\mathcal{D}}(v^M) = \sum_{K=1}^{N_o} J_K^{\text{hom},\mathcal{D}}(v^M |_{\Omega_K}) \quad (23)$$

where

$$\mathcal{V}^M = \{v \in C^0(\Omega) : v |_{\Omega_K} \in \mathcal{V}_K^M, 1 \leq K \leq N_o\} \quad (24)$$

The homogenized problem then reads

$$J^{\text{hom},\mathcal{D}}(u^{M,\mathcal{D}}) = \inf_{v^M \in \mathcal{V}^M} J^{\text{hom},\mathcal{D}}(v^M) \quad (25)$$

where

$$\mathcal{V}_d^M = \{v \in \mathcal{V}^M : v = u_d \text{ on } \Gamma_u\} \quad (26)$$

We assume that this space is not empty.

3.6. Prolongated solution

Once the homogenized solution $u^{M,\mathcal{D}}$ is computed, we return to each subdomain and compute the prolonged solution $u^{\mathcal{D}}$.

$$u^{\mathcal{D}}|_{\Omega_K} = F_K^{\mathcal{D}} u^{M,\mathcal{D}}|_{\Omega_K} \quad 1 \leq K \leq N_o \quad (27)$$

The quality of the approximate solution $u^{\mathcal{D}}$ will be studied in section 4.

3.7. Local and homogenized variational principles

We now describe the variational principles associated with the minimum principles (25) and (17).

Local variational principle

The minimum principle (17) is equivalent to finding $w^{\mathcal{D}}(v^M) \in \mathcal{W}_K^{\mathcal{D}}$ such that

$$B_K(v^M + w^{\mathcal{D}}(v^M), w) = F_K(w) \quad \forall w \in \mathcal{W}_K^{\mathcal{D}} \quad (28)$$

We observe that the solution may be written as the sum of a particular solution $w_d^{\mathcal{D}}$ due to the linear functional $F_K(\cdot)$ and a term depending linearly on v^M denoted $w_0^{\mathcal{D}}(v^M)$:

$$w^{\mathcal{D}}(v^M) = w_0^{\mathcal{D}}(v^M) + w_d^{\mathcal{D}} \quad (29)$$

$w_0^{\mathcal{D}}(v^M)$ belongs to $\mathcal{W}_K^{\mathcal{D}}$ and satisfies

$$B_K(v^M + w_0^{\mathcal{D}}(v^M), w) = 0 \quad \forall w \in \mathcal{W}_K^{\mathcal{D}} \quad (30)$$

$w_d^{\mathcal{D}}$ belongs also to $\mathcal{W}_K^{\mathcal{D}}$ and satisfies

$$B_K(w_d^{\mathcal{D}}, w) = F_K(w) \quad \forall w \in \mathcal{W}_K^{\mathcal{D}} \quad (31)$$

Homogenized variational principle

We may now rewrite the expression (22) of $J_K^{\text{hom},\mathcal{D}}(\cdot)$ as

$$J_K^{\text{hom},\mathcal{D}}(v^M) = J_{K0}^{\text{hom},\mathcal{D}}(v^M) + J_K(w_d^{\mathcal{D}}) \quad (32)$$

where

$$J_{K0}^{\text{hom},\mathcal{D}}(v^M) = \frac{1}{2} B_K^{\text{hom},\mathcal{D}}(v^M, v^M) - F_K^{\text{hom},\mathcal{D}}(v^M) \quad (33)$$

$$B_K^{\text{hom},\mathcal{D}}(v^M, v^M) = B_K(v^M + w_0^{\mathcal{D}}(v^M), v^M + w_0^{\mathcal{D}}(v^M)) \quad (34)$$

$$F_K^{\text{hom},\mathcal{D}}(v^M) = F_K(v^M + w_0^{\mathcal{D}}(v^M)) - B_K(w_d^{\mathcal{D}}, v^M + w_0^{\mathcal{D}}(v^M)) \quad (35)$$

The variational principle associated with (25) is thus to find $u^{M,\mathcal{D}} \in \mathcal{V}_0^M$ such that

$$\sum_{K=1}^{N_o} B_K^{\text{hom},\mathcal{D}}(u^{M,\mathcal{D}}, v^M) = \sum_{K=1}^{N_o} F_K^{\text{hom},\mathcal{D}}(v^M) \quad \forall v^M \in \mathcal{V}_0^M \quad (36)$$

Compared to the initial variational principal (3), we see that the bilinear and linear functionals have been ‘‘homogenized’’ and that the homogenized problem is defined in a finite coarse scale space \mathcal{V}^M instead on the initial full space \mathcal{V} .

4. A POSTERIORI ERROR ESTIMATION

$u^{\mathcal{D}}$ is an approximate solution to the problem and differs from the reference solution u . We note that $u^{\mathcal{D}}$ belongs to $\mathcal{V}^{\mathcal{D}}$

$$\mathcal{V}^{\mathcal{D}} = \{v = v^M + w : v^M \in \mathcal{V}_d^M, w \in L^2(\Omega), w|_{\Omega_K} \in \mathcal{W}_K^{\mathcal{D}} \mid 1 \leq K \leq N_o\} \quad (37)$$

and satisfies Property 1, which is obvious since $\mathcal{V}^{\mathcal{D}} \subset \mathcal{V}_d$.

Property 1 $u^{\mathcal{D}}$ is an admissible temperature field, i.e. $u^{\mathcal{D}} \in \mathcal{V}_d$

In order to estimate the quality of the solution $u^{\mathcal{D}}$, we shall build locally from $u^{\mathcal{D}}$ another temperature field denoted $u^{\mathcal{N}}$ such that the corresponding heat flux will be in thermal equilibrium over the structure. This equilibrium property will allow us later to use the Prager–Synge hypercircle theorem and relate the difference in the solution $u^{\mathcal{D}}$ and $u^{\mathcal{N}}$ to the exact energy error in these fields. This strategy is inspired by the error-in-the-constitutive law approach [4].

The computation of $u^{\mathcal{N}}$ is carried out in two steps: the construction of equilibrated surface heat source and the resolution of Neumann type subdomain problems.

4.1. Construction of equilibrated surface heat source

Using the coarse scale Dirichlet solution $u^{M,\mathcal{D}}$, we build surface heat source t_K^M on the boundary $\Gamma_{K_u} \cup \Gamma_{K_i}$ of each subdomain such that they are continuous across Γ_{K_i} and they are in thermal equilibrium over each subdomain; i.e. we require t_K^M to be such that

$$F_K(r) + \hat{F}_K(r) = 0 \quad \forall r \in \mathcal{R}_K, \quad 1 \leq K \leq N_o \quad (38)$$

$$\sum_{K=1}^{N_o} \hat{F}_K(v) = 0 \quad \forall v \in \mathcal{V}_o \quad (39)$$

where

$$\hat{F}_K(v) = \int_{\Gamma_{K_u} \cup \Gamma_{K_i}} t_K^M v \, ds \quad (40)$$

Assuming that the subdomains Ω_K , $1 \leq K \leq N_o$, have been obtained using a finite element type decomposition of the structure and that the coarse scale spaces \mathcal{V}_K^M are finite element type local spaces, the construction of the equilibrated surface heat source may be carried out locally using the equilibration technique introduced in [4,7].

4.2. Subdomain Neumann type problems

The temperature field $u^{\mathcal{N}}$ is built in each subdomain separately: $u^{\mathcal{N}}|_{\Omega_K} = u_K^{\mathcal{N}}$ where $u_K^{\mathcal{N}} \in \mathcal{V}_K = H^1(\Omega_K)$ is such that

$$B_K(u_K^{\mathcal{N}}, v) = F_K(v) + \hat{F}_K(v) \quad \forall v \in \mathcal{V}_K \quad (41)$$

$$\int_{\Omega_K} (u_K^{\mathcal{N}} - u^{\mathcal{D}}|_{\Omega_K}) r \, dx = 0 \quad \forall r \in \mathcal{R}_K \quad (42)$$

The condition (42) allows us to define the solution uniquely.

In order to solve efficiently the problem (41-42), we take advantage of the decomposition of the space \mathcal{V}_K into the spaces \mathcal{V}_K^M and \mathcal{W}_K . We now look for u_K^N in the form

$$u_K^N = u_K^{M,N} + w^N, \quad u_K^{M,N} \in \mathcal{V}_K^M, \quad w^N \in \mathcal{W}_K \quad (43)$$

such that

$$B_K(u_K^{M,N} + w^N, v^M) = F_K(v^M) + \hat{F}_K(v^M) \quad \forall v^M \in \mathcal{V}_K^M \quad (44)$$

$$B_K(u_K^{M,N} + w^N, w) = F_K(w) + \hat{F}_K(w) \quad \forall w \in \mathcal{W}_K \quad (45)$$

The solution of (45) may be expressed as the sum of a particular solution w_d^N due to the linear functional in the right hand side and a term, denoted $w_0^N(u_K^{M,N})$, depending linearly on $u_K^{M,N}$:

$$w^N = w_0^N(u_K^{M,N}) + w_d^N \quad (46)$$

$w_0^N(u_K^{M,N})$ belongs to \mathcal{W}_K and satisfies

$$B_K(u_K^{M,N} + w_0^N(u_K^{M,N}), w) = 0 \quad \forall w \in \mathcal{W}_K \quad (47)$$

w_d^N belongs to \mathcal{W}_K and satisfies

$$B_K(w_d^N, w) = F_K(w) + \hat{F}_K(w) \quad \forall w \in \mathcal{W}_K \quad (48)$$

Note that the problem (49) differs from the problem (30) only through the boundary conditions. The problem (44) now reduces to find $u_K^{M,N} \in \mathcal{V}_K^M$ such that

$$B_K^{\text{hom},N}(u_K^{M,N}, v^M) = F_K^{\text{hom},N}(v^M) \quad \forall v^M \in \mathcal{V}_K^M \quad (49)$$

$$\int_{\Omega_K} (u_K^{M,N} - u^{M,D}|_{\Omega_K}) r \, dx = 0 \quad \forall r \in \mathcal{R}_K \quad (50)$$

where

$$B_K^{\text{hom},N}(u^M, v^M) = B_K(u^M + w_0^N(u^M), v^M + w_0^N(v^M)) \quad (51)$$

$$F_K^{\text{hom},N}(v^M) = F_K(v^M) + \hat{F}_K(v^M) - B_K(w_d^N, v^M) \quad (52)$$

The condition (42) has been replaced by the condition of (50) which allows us to deal only with the coarse components of the fields. Note however that conditions (42) and (50) are equivalent in the case where \mathcal{W}_K is defined so as to contain all w such that

$$\int_{\Omega_K} w r \, dx = 0 \quad \forall r \in \mathcal{R}_K \quad (53)$$

4.3. Property of the Neumann type solution

In each subdomain Ω_K , we may define the heat flux vector associated with the local solution u_K^N by $\sigma_K^N = E\epsilon(u_K^N)$ which is well defined in $(L^2(\Omega_K))^n$. We have the following property

Property 2 *There exists a heat flux field $\sigma^N \in (L^2(\Omega))^n$ such that*

$$\sigma^N|_{\Omega_K} = \sigma_K^N = E\epsilon(u_K^N), 1 \leq K \leq N_o \quad (54)$$

which is admissible i.e. $\sigma^N \in \mathcal{S}_d$ where

$$\mathcal{S}_d = \{ \tau \in (L^2(\Omega))^n, \operatorname{div} \tau \in L^2(\Omega) : \int_{\Omega} \tau \cdot \epsilon(v) \, dx = F(v), \forall v \in \mathcal{V}_0 \} \quad (55)$$

Proof: Owing to the definition of σ_K^N , we have

$$\int_{\Omega_K} \sigma_K^N \cdot \epsilon(v) \, dx = F_K(v) + \hat{F}_K(v), \forall v \in \mathcal{V}_K, 1 \leq K \leq N_o \quad (56)$$

Summing these relations over all the subdomains, we get

$$\sum_{K=1}^{N_o} \int_{\Omega_K} \sigma_K^N \cdot \epsilon(v) \, dx = \sum_{K=1}^{N_o} F_K(v) + \sum_{K=1}^{N_o} \hat{F}_K(v), \forall v \in \mathcal{V}^N \quad (57)$$

where

$$\mathcal{V}^N = \{ v \in L^2(\Omega) : v|_{\Omega_K} \in \mathcal{V}_K \} \quad (58)$$

Using (39) and the fact that $\mathcal{V}_0 \in \mathcal{V}^N$, (57) yields

$$\sum_{K=1}^{N_o} \int_{\Omega_K} \sigma_K^N \cdot \epsilon(v) \, dx = \sum_{K=1}^{N_o} F_K(v), \forall v \in \mathcal{V}_0 \quad (59)$$

We may take $v = \varphi \in \mathcal{D}(\Omega_K)$ (space of test functions) in (56) and using standard arguments, we show that $\operatorname{div} \sigma_K^N \in L^2(\Omega_K)$ and $\operatorname{div} \sigma_K^N + f_d = 0$ in $L^2(\Omega_K)$. From this and (59), it follows there exists a function σ^N such that $\operatorname{div} \sigma^N \in L^2(\Omega)$ and $\sigma^N|_{\Omega_K} = \sigma_K^N$, see [8].

4.4. Bounds on the error

The energy in the difference of the “Dirichlet”, u^D , and the “Neumann”, u^N , solutions may be linked to the exact energy error in these solutions using the Prager–Synge hyper-circle theorem [3]. Since the solution u^N belongs to $H^1(\Omega)$ only on each subdomain but not on the whole domain, we need to introduce the notation

$$\| \| v \| \|_{E(\Omega)} = \left(\sum_{K=1}^{N_o} \| v|_{\Omega_K} \|_{E(\Omega_K)}^2 \right)^{1/2} \quad (60)$$

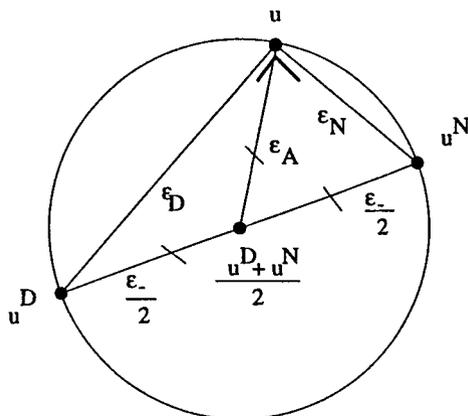


Figure 3. Geometrical interpretation of the Properties 3 and 4

Property 3 Defining the energy norm errors associated with the “Dirichlet” solution, the “Neumann” solution and their difference

$$e_D = \|u^D - u\|_{E(\Omega)}, \quad e_N = \|u^N - u\|_{E(\Omega)}, \quad e_- = \|u^D - u^N\|_{E(\Omega)} \quad (61)$$

we have $e_- = \sqrt{e_D^2 + e_N^2}$

The hypercircle theorem of Prager–Synge also gives us an interesting property concerning the average of the Dirichlet and Neumann solutions.

Property 4 Defining the energy norm associated with the average between the Dirichlet and the Neumann solution,

$$e_A = \left\| \frac{u^D + u^N}{2} - u \right\|_{E(\Omega)} \quad (62)$$

we have $e_A = \frac{1}{2}e_-$.

Proofs: The geometrical interpretation of property 3 and 4 is shown Fig. 3. Property 3 stems from

$$\sum_{K=1}^{N_0} B_K(u_K^N - u, u^D - u) = \int_{\Omega} (\sigma^N - \sigma) \cdot \epsilon(u^D - u) \, dx = 0 \quad (63)$$

since $\sigma^N, \sigma \in \mathcal{S}_d$ and $u^D, u \in \mathcal{V}_d$. Property 4 follows from

$$\begin{aligned} e_A^2 &= \left\| \frac{u^D - u}{2} + \frac{u^N - u}{2} \right\|_{E(\Omega)}^2 \\ &= \frac{1}{4}(e_D^2 + e_N^2 + 2 \sum_{K=1}^{N_0} B_K(u_K^N - u, u^D - u)) = \frac{1}{4}e_-^2 \end{aligned}$$

The last equality is obtained using Property 3 and relation (63).

5. FINITE ELEMENT APPROXIMATION

In the two-scale strategy presented in Section 3, the space $\mathcal{V}_K = H^1(\Omega_K)$ was decomposed into the spaces \mathcal{V}_K^M and \mathcal{W}_K . In practice we shall work with a finite element subspace, $\mathcal{V}_K^M \subset \mathcal{V}_K$. This subspace is decomposed into the spaces \mathcal{V}_K^M and \mathcal{W}_K^m so that the conditions described in the subsection 3.2 are fulfilled. Since we work with the subspace \mathcal{V}_K^m , the prolongation operator will now be discrete.

5.1. Discrete prolongation operator

We have to find $w^{m,D}(v^M) \in \mathcal{W}_K^{m,D}$ such that

$$B_K(v^M + w^{m,D}(v^M), w) = F_K(w) \quad \forall w \in \mathcal{W}_K^{m,D} \quad (64)$$

where

$$\mathcal{W}_K^{m,D} = \{w \in \mathcal{W}_K^m : w = 0 \text{ on } \Gamma_{Ku} \cup \Gamma_{Ki}\} \quad (65)$$

The prolongation operator is now denoted by $P_K^{m,D}$ to stress the fact that it depends on the choice of the subspace \mathcal{W}_K^m . The associate local and global homogenized potential are now

$$J_K^{M,m,D} : \mathcal{V}_K^M \rightarrow \mathbb{R} : v^M \rightarrow J_K^{M,m,D}(v^M) = J_K(P_K^{m,D}v^M) \quad (66)$$

$$J^{M,m,D} : \mathcal{V}^M \rightarrow \mathbb{R} : v^M \rightarrow J^{M,m,D}(v^M) = \sum_{K=1}^{N_o} J_K^{M,m,D}(v^M|_{\Omega_K}) \quad (67)$$

The homogenized problem is then

$$J^{M,m,D}(u^{M,m,D}) = \inf_{v^M \in \mathcal{V}^M} J^{M,m,D}(v^M) \quad (68)$$

and the prolonged solution reads

$$u^{m,D}|_{\Omega_K} = P_K^{m,D}u^{M,m,D}|_{\Omega_K} \quad 1 \leq K \leq N_o \quad (69)$$

5.2. Discrete reference solution

We see the solutions $u^{M,m,D}$ as an approximate solutions to a discrete reference solution, u^m , defined in what follows. We introduce the discrete global space \mathcal{V}^m defined on Ω as

$$\mathcal{V}^m = \{v \in C^0(\Omega) : v|_{\Omega_K} \in \mathcal{V}_K^m, 1 \leq K \leq N_o\} \quad (70)$$

Clearly $\mathcal{V}^m \subset \mathcal{V}$. As a particular case, in the numerical experiments, the meshes defining the space \mathcal{V}_K^m on each subdomain will be such that they match on the Γ_{Ki} boundaries. Thus, the space \mathcal{V}^m is defined on a global mesh obtained by assembling the local subdomains meshes. Finally, we define the discrete reference solution u^m as $u^m \in \mathcal{V}_d^m$ such that

$$B(u^m, v) = F(v) \quad \forall v \in \mathcal{V}_0^m \quad (71)$$

where

$$\mathcal{V}_d^m = \{v \in \mathcal{V}^m : v = u_d \text{ on } \Gamma_u\}, \quad \mathcal{V}_0^m = \{v \in \mathcal{V}^m : v = 0 \text{ on } \Gamma_u\} \quad (72)$$

5.3. A posteriori error estimation

u^m is our reference solution and $u^{m,\mathcal{D}}$ is our approximate solution. We wish to analyze which results of Section 4 still hold and in what sense. $u^{m,\mathcal{D}}$ belongs to the space $\mathcal{V}^{m,\mathcal{D}}$ defined by

$$\mathcal{V}^{m,\mathcal{D}} = \{v = v^M + w : v^M \in \mathcal{V}_d^M, w \in L^2(\Omega), w|_{\Omega_K} \in \mathcal{W}_K^{m,\mathcal{D}} \ 1 \leq K \leq N_o\} \quad (73)$$

Property 1 must now be understood as follows:

Property 5 $u^{m,\mathcal{D}}$ is an admissible discrete temperature field, i.e. $u^{m,\mathcal{D}} \in \mathcal{V}_d^m$

Proof: This property is obvious since $u^{m,\mathcal{D}} \in \mathcal{V}^{m,\mathcal{D}} \subset \mathcal{V}_d^m$.

Following the steps described in subsections 4.1 and 4.2, we are able to construct a Neumann solution $u_K^{m,\mathcal{N}}$ over each subdomain. The corresponding heat flux $\sigma_K^{m,\mathcal{N}} = E\epsilon(u_K^{m,\mathcal{N}})$ belongs to $(L^2(\Omega_K))^n$ and we have the property:

Property 6 There exists a heat flux field $\sigma^{m,\mathcal{N}} \in (L^2(\Omega))^n$ such that

$$\sigma^{m,\mathcal{N}}|_{\Omega_K} = \sigma_K^{m,\mathcal{N}} = E\epsilon(u_K^{m,\mathcal{N}}) \quad (74)$$

which is admissible in the discrete sense i.e. $\sigma^{m,\mathcal{N}} \in \mathcal{S}_d^m$ where

$$\mathcal{S}_d^m = \{\tau \in (L^2(\Omega))^n : \int_{\Omega} \tau \cdot \epsilon(v) \, dx = F(v), \forall v \in \mathcal{V}_0^m\} \quad (75)$$

Proof: Following the path of the proof of Property 2, we may obtain the relation

$$\sum_{K=1}^{N_o} \int_{\Omega_K} \sigma_K^{m,\mathcal{N}} \cdot \epsilon(v) \, dx = \sum_{K=1}^{N_o} F_K(v), \forall v \in \mathcal{V}_0^m \quad (76)$$

yielding the proof.

Properties 3 and 4 must be rewritten as Properties 7 and 8.

Property 7 Defining the energy norm errors associated with the “Dirichlet” solution, the “Neumann” solution and their difference

$$e_{\mathcal{D}}^m = \|u^{m,\mathcal{D}} - u^m\|_{E(\Omega)}, \quad e_{\mathcal{N}}^m = \| \|u^{m,\mathcal{N}} - u^m \| \|_{E(\Omega)}, \quad e_{-}^m = \| \|u^{m,\mathcal{D}} - u^{m,\mathcal{N}} \| \|_{E(\Omega)} \quad (77)$$

we have $e_{-}^m = \sqrt{(e_{\mathcal{D}}^m)^2 + (e_{\mathcal{N}}^m)^2}$

Property 8 Defining the energy norm associated with the average between the Dirichlet and the Neumann solution,

$$e_{\mathcal{A}}^m = \| \| \frac{u^{m,\mathcal{D}} + u^{m,\mathcal{N}}}{2} - u \| \|_{E(\Omega)} \quad (78)$$

we have $e_{\mathcal{A}}^m = \frac{1}{2}e_{-}^m$

Proofs: The proofs of Properties 7 and 8 are similar to the proofs of Properties 3 and 4. We note that the discrete exact stress $\sigma^m = E\epsilon(u^m)$ belongs to \mathcal{S}_d^m and thus we have

$$\sum_{K=1}^{N_o} B_K(u_K^{m,\mathcal{N}} - u^m, u^{m,\mathcal{D}} - u^m) = \int_{\Omega} (\sigma^{m,\mathcal{N}} - \sigma^m) \cdot \epsilon(u^{m,\mathcal{D}} - u^m) \, dx = 0 \quad (79)$$

since $\sigma^{m,\mathcal{N}}, \sigma^m \in \mathcal{S}_d^m$ and $u^{m,\mathcal{D}}, u^m \in \mathcal{V}_d^m$.

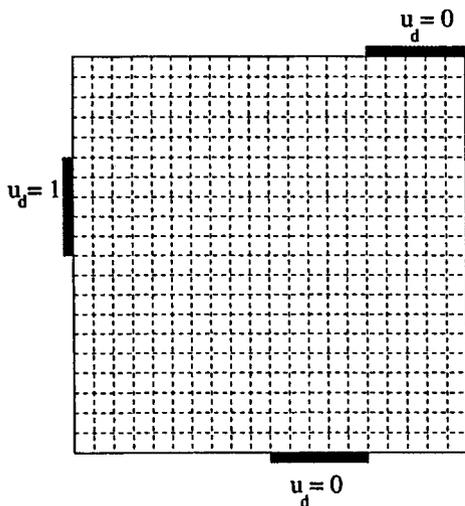


Figure 4. A square domain divided into 400 subdomains and prescribed temperature along three segments of the boundary

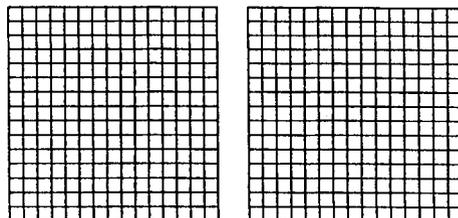


Figure 5. Mesh for a subdomain without and with inclusion

6. NUMERICAL EXPERIMENTS

6.1. Relative error, local contributions and effectivity indices

We shall need some notations to describe the numerical experiments. The relative error associated with the absolute errors e_D , e_N , e_- , e_A are defined using the energy of the exact solution.

$$\epsilon_D^m = e_D^m / \|u^m\|_{E(\Omega)}, \quad \epsilon_N^m = e_N^m / \|u^m\|_{E(\Omega)}, \quad \epsilon_-^m = e_-^m / \|u^m\|_{E(\Omega)}, \quad \epsilon_A^m = e_A^m / \|u^m\|_{E(\Omega)}$$

The contribution of a subdomain Ω_K to an absolute error will be denoted using a K index:

$$e_{D,K}^m = \|u^{m,D} - u^m\|_{E(\Omega_K)}, \quad e_{N,K}^m = \|u^{m,N} - u^m\|_{E(\Omega_K)} \tag{80}$$

$$e_{-,K}^m = \|u^{m,D} - u^{m,N}\|_{E(\Omega_K)}, \quad e_{A,K}^m = \left\| \frac{u^{m,D} + u^{m,N}}{2} - u \right\|_{E(\Omega_K)} \tag{81}$$

We also need to define the global effectivity indices:

$$\theta = \frac{e_-^m}{e_D^m}, \quad \theta^A = \frac{e_-^m}{2e_A^m} \tag{82}$$

and the local effectivity indices:

$$\theta_K = \frac{e_{-,K}^m}{2e_{D,K}^m}, \quad \theta_K^A = \frac{e_{-,K}^m}{2e_{A,K}^m}, \quad 1 \leq K \leq N. \tag{83}$$

Owing to Property 7, θ should be greater than 1 and owing to Property 8, θ^A should be equal to 1.

6.2. An example

We consider a square domain, $\Omega = [0, 20] \times [0, 20]$, Fig. 4, insulated on its boundary except along three segments where the temperature is prescribed. The domain is decomposed into 400 unit-square subdomains. Each subdomain is meshed using a 15×15 uniform mesh. Half of the subdomains are made of a material with uniform and isotropic conductivity, $E = 1$, and the other half contain a square inclusion of conductivity $E = 0.01$, Fig. 5. The distribution of the inclusions over the domain is obtained through a random process.

On each subdomain, Ω_K , \mathcal{V}_K^m is based on a 15×15 uniform mesh of four-node quadrilateral, Fig. 5. The space \mathcal{V}_K^M is defined using bi-linear functions:

$$\mathcal{V}_K^M = \{v \in \mathcal{V}_K^m : v = a_1(x - x_3)(y - y_3) + a_2(x - x_1)(y - y_3) + a_3(x - x_1)(y - y_1) + a_4(x - x_3)(y - y_1), a_i \in \mathbb{R}, i = 1 \cdots 4, (x, y) \in \Omega_K\} \quad (84)$$

where (x_i, y_i) , $i = 1 \cdots 4$, are the coordinates of the vertices of the subdomain. Note that we have indeed $\mathcal{V}_K^M \subset \mathcal{V}_K^m$. The correction space \mathcal{W}_K^m is defined by

$$\mathcal{W}_K^m = \{w \in \mathcal{V}_K^m : \int_{\Omega_K} \nabla w \cdot \nabla v \, dx = 0 \quad \forall v \in \mathcal{V}_K^M, \int_{\Omega_K} w r \, dx = 0 \quad \forall r \in \mathcal{R}_K\} \quad (85)$$

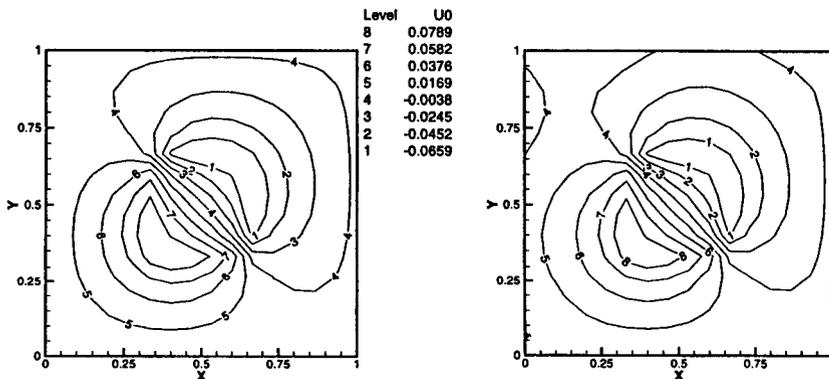


Figure 6. Correction field $w^{m,D}$, over a subdomain containing an inclusion, associated with a bilinear function v^M whose value is one in the bottom left corner. Left: Dirichlet boundary condition on each edge; right: Dirichlet boundary condition on the bottom and right edges and homogeneous Neumann boundary condition on the other two edges.

Figure 6 shows the correction function, $w^{m,D}(v^M)$, for a bilinear function whose value is one at the bottom left corner. Dirichlet conditions are applied on each edge for the left

Figure and only on the bottom and right edge for the right Figure, the two other edges being supplied with homogeneous Neumann-type boundary conditions.

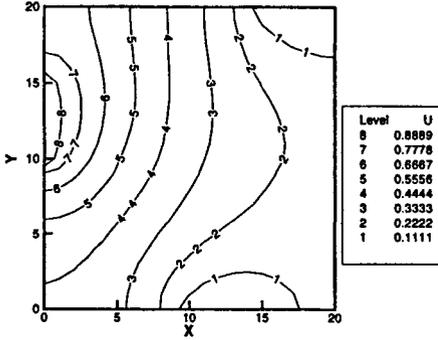


Figure 7. Homogenized solution with the Dirichlet approach: $u^{M,m,D}$

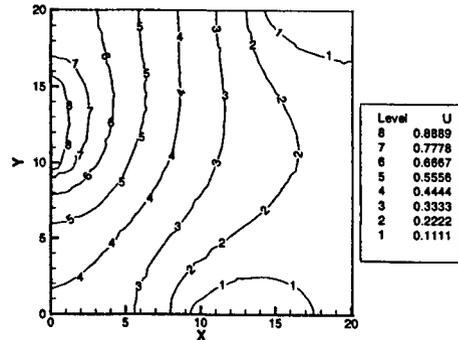


Figure 8. Prolongated solution with the Dirichlet approach: $u^{m,D}$

Figure 7 shows the homogenized solution, $u^{M,m,D}$ and Fig. 8 the prolonged solution $u^{m,D}$. We note that this latter solution involves local features that were not present in the homogenized solution. The exact relative energy error for this solution is $\epsilon_D^m = 17.6\%$. The contribution of each subdomain to the error, $e_{D,K}^m$, are given Fig. 9. The contributions are the highest where a change of boundary condition occur and are small inside the domain. This latter observation may come from the fact that the boundary of a subdomain never cross an inclusion.

The Neumann solution, $u^{m,N}$, is shown Fig. 10. It is discontinuous from subdomain to subdomain as expected. The exact relative error associated to this solution is $\epsilon_N^m = 24.7\%$. The relative energy in the difference of the Dirichlet and Neumann solutions gives the estimated error $\epsilon_-^m = 30.3\%$ so that the global effectivity index is $\theta = 30.3/17.6 = 1.7$, greater than one as expected. The local effectivity indices θ_K are shown Fig. 12. They are reasonably close to one on the subdomains contributing the most to the error. The average effectivity index is 1.46 and the standard deviation from this average is 0.78. Note that the equilibrated surface heat source were built in a crude manner since they are constant along the edges of the subdomains. Usually, linear representations are used for four-node quadrilateral elements [7].

The exact error associated with the average between the Dirichlet and Neumann solutions is $\epsilon_A^m = 15.2\%$ which is half the error in the difference of these two solutions (30.3%) as announced by Property 8. The local contributions $e_{A,K}^m$ to the error are given Fig. 11 and do not differ much from the contributions $e_{D,K}^m$ shown Fig. 9. Finally, Fig. 13 shows

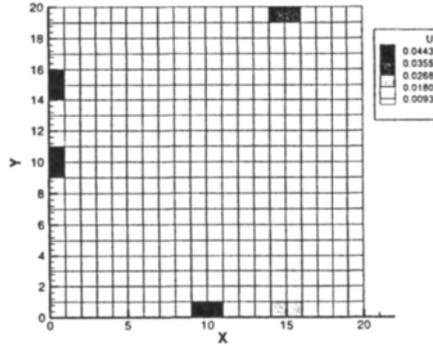


Figure 9. Subdomain contribution to the absolute error for the Dirichlet solution ($e_{\mathcal{D},K}^m$)

the local effectivity indices θ_K^A . The average effectivity index is 0.90 and the standard deviation from this average is 0.23.

The results for this example suggest that the average of the Dirichlet and Neumann solutions may be a better approximate solution than the Dirichlet solution itself since

- we can compute exactly the global energy error of the average without knowing the exact solution
- the local effectivity indices are closer to one for the average than they were in the case of the approximate Dirichlet solution
- the error in the average is smaller than for the Dirichlet solution (15.2% against 17.6%)

The first observation is supported by Property 8 and will always hold. On the other hand, the two other observations may not be true in general and pertain to the particular example considered.

7. CONCLUSIONS

We have presented a two-scale strategy to model scalar diffusion in heterogeneous structures. The structure is decomposed into subdomains over which the fine-scale components of the solution are expressed in terms of the coarse-scale component through a prolongation operator. The choice of Dirichlet types boundary conditions when solving for the prolongation operator yields a continuous approximate solution over the structure called

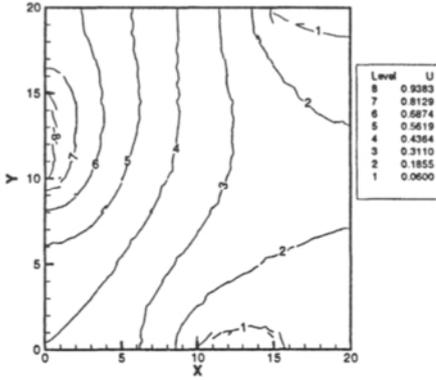


Figure 10. Neumann solution $u^{m,N}$

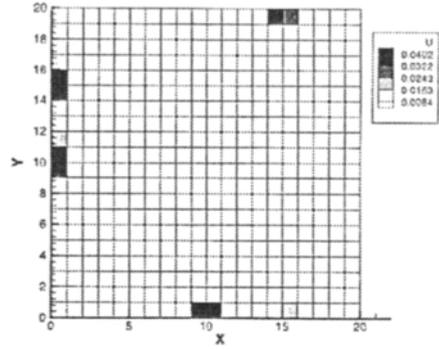


Figure 11. Subdomain contribution to the absolute error for the average solution ($e_{A,K}^m$)

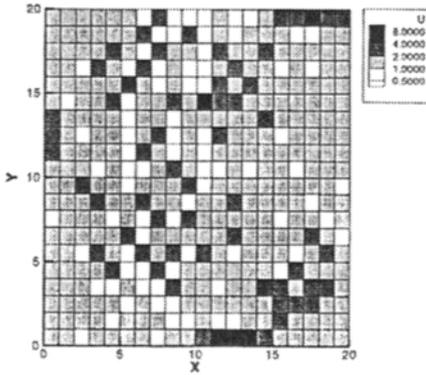


Figure 12. Local effectivity index on each subdomain, θ_K , for the dirichlet solution. The minimum and maximum values are 0.35 and 8.79, respectively

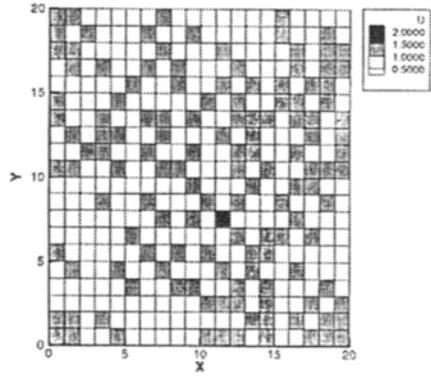


Figure 13. Local effectivity index on each subdomain for the average approximate solution, θ_K^A . The minimum and maximum values are 0.16 and 1.58, respectively

the Dirichlet solution. Using an equilibration technique and Neumann type subdomain problems we were able to build a second solution, called the Neumann solution, in thermal equilibrium over the entire structure. The energy in the difference of the two approximate solutions was found, using the Prager–Synge hypercircle theorem, to be an upper-bound to the error in the Dirichlet solution. The strategy was also analyzed in the important practical case of an approximate computation of the prolongation operator. A 2-D numerical experiment showed that the strategy yields a fast and fairly accurate first guess to the exact solution except in some zones generally near a point at which there is a change of boundary condition. We plan in a forthcoming paper to design a strategy to correct the solution where it is needed.

REFERENCES

1. T. I. Zohdi, J. T. Oden, and G. J. Rodin. Hierarchical modeling of heterogeneous bodies. *Comp. Meth. in Applied Mech. and Engrg.*, 138(1–4):273–298, 1996.
2. J. T. Oden and T. I. Zohdi. Analysis and adaptive modeling of highly heterogeneous elastic structures. *Comp. Meth. in Applied Mech. and Engrg.*, 148(3–4):367–391, 1997.
3. W. Prager and J. L. Synge. Approximation in elasticity based on the concept of functions space. *Quart. Appl. Math.*, 5:261–269, 1947.
4. P. Ladevèze and D. Leguillon. Error estimate procedure in the finite element method and application. *SIAM J. Num. Anal.*, 20(3):485–509, 1983.
5. P. Ladevèze and D. Dureisseix. *Une nouvelle stratégie de calcul parallèle et "micro/macro" en mécanique non linéaire*. Rapport interne 188, LMT, Ecole Normale Supérieure de Cachan, 1997.
6. T. Arbogast. Private communication.
7. P. Ladevèze, J. P. Pelle, and P. Rougeot. Error estimation and mesh optimization for classical finite elements. *Engrg. Computation*, 8:69–80, 1991.
8. F. Brezzi. *Mixed and hybrid finite element methods*. Springer series in computational mathematics. Springer-Verlag, 1991.

This Page Intentionally Left Blank

A modelling error estimator for dynamic structural model updating

P. Ladevèze

LMT - Cachan, ENS Cachan / CNRS / Univ. Paris 6, 61, avenue du President Wilson, 94235 Cachan Cedex, France

Abstract: The key-point in structural model updating is to first design an error that's between a mathematical model with its parameters and tests, and then to minimize this error. The proposed modelling error estimator can be considered as an extension to the works previously conducted on a posteriori error estimators in quantifying the quality of a finite element computation. It is based on the "mechanics concept" of constitutive relation error. Damping effects and non-linearities are taken into account. All available experimental results may be used as either free or forced vibrations. Applications deal with the improvement of finite element models, i.e. the mass, stiffness and damping matrices.

1. INTRODUCTION

Controlling and mastering both mechanical and numerical models has always been a major preoccupation in Mechanics, especially for engineers. So, what has changed? Quantitative tools have started to appear; they are capable of quantifying the quality of an approximate model with respect to a reference model.

In Mechanics and in particular for dynamic problems, there are not just two categories of model, but rather there are three :

- the continuum mechanics model,
- the numerical model, and
- experimental simulations.

Therefore, two situations do indeed exist. Most papers appearing in the present book are devoted to the situation whereby the continuum mechanics model (resp. the numerical model) is the reference (resp. the approximate model). The other situation studied in this paper relies on experimental simulations as the reference. The continuum mechanics model now becomes the approximate model.

For both of these situations, the first and most critical concern is to develop error measures which are able to quantify the quality of the approximate model, i.e. to build a distance between the approximate and reference models. Since both the data of the reference model and the approximate solution are assumed to be known, an a posteriori error estimator can be derived.

The second concern differs for the two studied situations. In the case where the numerical model is the approximate one, it is to determine the best computational parameters for a given level of accuracy, i.e. those which lead to a minimum numerical cost. For the other situation, the second issue is to correct the approximate model so that it provides a better fit for the experimental results.

References concerning the situation where the numerical model is the approximate one can be found in the present book. Concerning the situation studied herein where the reference has been defined through experimental simulations, many references are also available (see (Bui, Tanaka *et al.*, 1994)). Most of the reference authors have been part of "Automatics and Control", with for them the model being represented by a black box. In the field of Dynamic Structural Model Updating, mention should be made of: (Baruch, 1982), (Mottershead, 1988), (Ibrahim *et al.*, 1990), (Berger *et al.*, 1991), (Natke, 1991), (Farhat and Hemez, 1993), (Nobari, 1993), (Piranda *et al.*, 1993), (Berman, 1995), and (Kaouk and Zimmerman, 1995).

At Cachan, still with respect to Dynamic Structural Model Updating, we have been developing a rather different approach which demonstrates a strong mechanics content. Such an orientation could be considered as an extension to the works previously conducted on a posteriori estimators in quantifying the quality of a finite element computation, i.e. in studying the situation where the approximate model is the numerical one. More precisely, the updating method we have been developing over the past 15 years is based on the concept of a posteriori constitutive relation error. This concept was introduced in (Ladevèze, 1983) for improving models by means of experimental free vibrations tests. The numerical strategy has been developed in (Ladevèze, Reynier and Nedjar, 1994). The M.A.T. (Model Adjustment Technique) software was developed and implemented by the CNES (Bricout *et al.*, 1993) to constitute a post-processor of the MSC/NASTRAN. Applications deal with the improvement of finite element models, i.e. the mass, stiffness and damping matrices.

An additional step has recently been achieved (Ladevèze, 1994), (Ladevèze, Reynier and Maia, 1994), and (Chouaki *et al.*, 1997). We have taken into account damping effects and, more generally, non-linearities due to the material behavior. All available experimental results may be used as either free or forced-vibration responses. These test data, combined with a certain knowledge of the model, constitute the reference which allows building the "modified constitutive relation error" used to define the model's quality. The basic equations (e.g. equilibrium and compatibility equations), along with reliable experimental data (e.g. excitations and measurement locations), have been verified exactly.

The tuning procedure is iterative, with each iteration composed of two steps: (i) locating the most erroneous substructures, and (ii) correcting their structural parameters. This strategy is very well suited to the ill-posed property of tuning problems, thanks to the efficiency of the location step as well as the ability to introduce more information at each iteration.

The aim of this paper is to focus on the main features of the proposed modelling error estimator and its corresponding updating strategy. Applications representing industrial problems are also presented.

2. THE APPROXIMATE MODEL

2.1. Forced-vibration mathematical model - Classical formulation

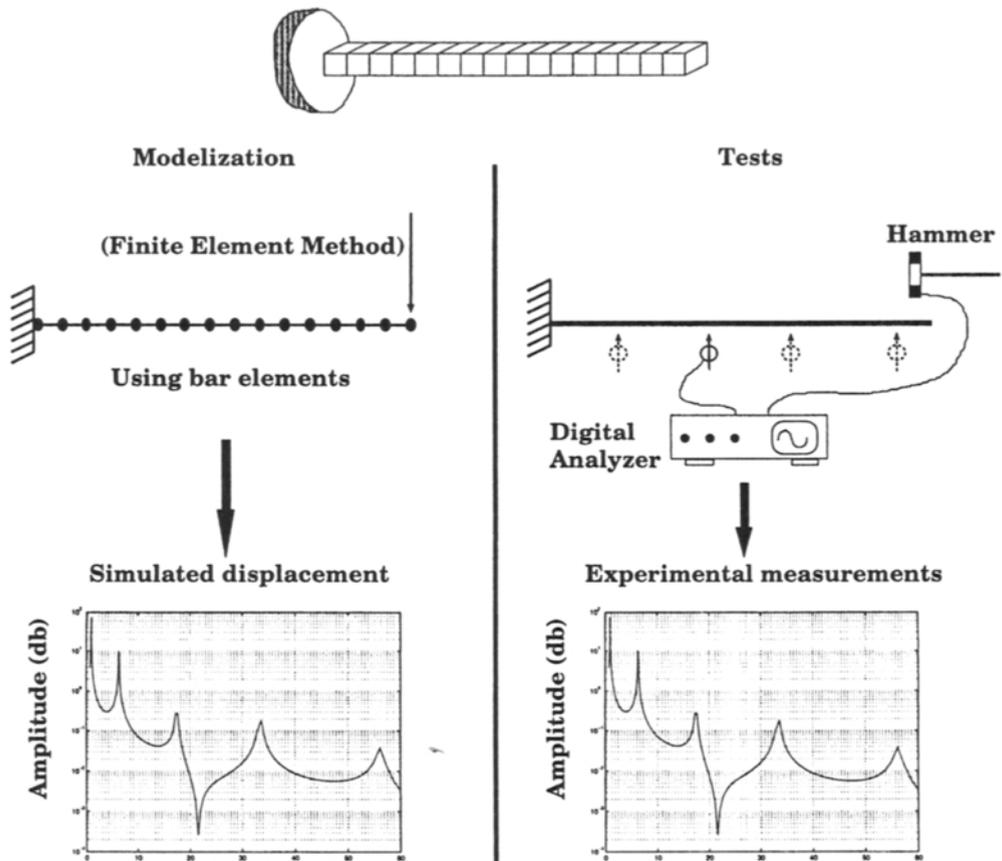


Figure 1: The studied structure: Excitations and geometry

Let us consider a structure occupying a domain Ω whose boundary is $\partial\Omega$. It is being submitted to a body force $f_{d,\omega}e^{i\omega t}$ (applied on Ω); a surface force $F_{d,\omega}e^{i\omega t}$ is given on a part $\partial_2\Omega$ of $\partial\Omega$. The displacement $U_{d,\omega}e^{i\omega t}$ is prescribed on the

complementary boundary part of $\partial_2\Omega$, namely $\partial_1\Omega$. ω is the given angular frequency; $\underline{U}_{d,\omega}$, $\underline{f}_{d,\omega}$ and $\underline{F}_{d,\omega}$ are the amplitudes.

The assumption of small displacements is then made; harmonic solutions are sought. Therefore, the problem becomes in the particular case where damping has not been taken into account :

Find the amplitude $\underline{U}_\omega(\underline{M})$ $\underline{M} \in \Omega$ such that :

$$\begin{aligned}
 & \bullet \underline{U}_\omega \in \mathcal{U} \quad \underline{U}_\omega = \underline{U}_{d,\omega} \text{ on } \partial_1\Omega \\
 & \bullet - \int_{\Omega} \text{Tr} \left[\mathbf{K} \boldsymbol{\varepsilon}(\underline{U}_\omega) \boldsymbol{\varepsilon}(\underline{U}^*) \right] d\Omega + \int_{\Omega} \underline{f}_{d,\omega} \circ \underline{U}^* d\Omega \\
 & \quad + \int_{\partial_2\Omega} \underline{F}_{d,\omega} \circ \underline{U}^* dS = - \int_{\Omega} \rho \omega^2 \underline{U}_\omega \circ \underline{U}^* d\Omega \\
 & \quad \forall \underline{U}^* \in \mathcal{U}_{ad,0} = \left\{ \underline{U}^* \mid \underline{U}^*|_{\partial_1\Omega} = 0, \underline{U}^* \in \mathcal{U} \right\}
 \end{aligned} \tag{1}$$

\mathcal{U} denotes the space of the displacement fields defined over Ω that possess finite energy; \mathbf{K} is Hooke's tensor, and ρ the mass density. The strain operator is written $\boldsymbol{\varepsilon}$.

A new writing for the previous classical formulation of the forced-vibration mathematical model can then be given by introducing the basic equations of Continuum Mechanics :

Find $s_\omega = \left\{ \underline{U}(\underline{M}), \underline{\Gamma}(\underline{M}), \boldsymbol{\sigma}(\underline{M}); \underline{M} \in \Omega \right\}_\omega$ which satisfies :

- kinematic constraints: $\underline{U} \in \mathcal{U}$

$$\underline{U}|_{\partial_1\Omega} = \underline{U}_{d,\omega} \tag{2}$$

- equilibrium equations : $(\underline{\Gamma}, \boldsymbol{\sigma}) \in \mathcal{S}$

$$\begin{aligned}
 & - \int_{\Omega} \text{Tr} \left[\boldsymbol{\sigma} \boldsymbol{\varepsilon}(\underline{U}^*) \right] d\Omega + \int_{\Omega} \underline{f}_{d,\omega} \circ \underline{U}^* d\Omega \\
 & \quad + \int_{\partial_1\Omega} \underline{F}_{d,\omega} \circ \underline{U}^* dS = \int_{\Omega} \underline{\Gamma} \circ \underline{U}^* d\Omega
 \end{aligned} \tag{3}$$

- constitutive relations :

$$\begin{aligned}\underline{\sigma} &= \mathbf{K} \underline{\varepsilon}(\underline{U}) \\ \underline{\Gamma} &= -\rho \omega^2 \underline{U}\end{aligned}\quad (4)$$

Remark

The constitutive relations are generally the least reliable equations. They are determined by the following structural parameters, defined over Ω :

$$\mathbf{K}(\underline{M}) \quad \rho(\underline{M}) \quad \underline{M} \in \Omega$$

Notations

Let \mathcal{S} be the space of the triplet $s = \left\{ \underline{U}(\underline{M}), \underline{\Gamma}(\underline{M}), \underline{\sigma}(\underline{M}); \underline{M} \in \Omega \right\}$ where $\underline{U} \in \mathcal{U}$ and $(\underline{\Gamma}, \underline{\sigma}) \in \mathcal{S}$. Solutions to the kinematic constraints (2) and the equilibrium equations (3) are called "admissible". \mathcal{S}_{ad}^ω denotes the corresponding subspace of \mathcal{S} .

The forced-vibration mathematical model can then be written :

Find $s_\omega \subset \mathcal{S}_{ad}^\omega$ which satisfies the constitutive relations :

$$\begin{aligned}\underline{\sigma}_\omega &= \mathbf{K} \underline{\varepsilon}(\underline{U}_\omega) \\ \underline{\Gamma}_\omega &= -\rho \omega^2 \underline{U}_\omega\end{aligned}\quad (5)$$

2.2. Error on the constitutive relation

Let $s = (\underline{U}, \underline{\Gamma}, \underline{\sigma}) \in \mathcal{S}_{ad}^\omega$. The global constitutive relation error is the sum of two terms which are associated with the two constitutive relations (5) :

$$\begin{aligned}\eta_{\omega}^2(s) &= \frac{\gamma}{2} \int_{\Omega} \frac{1}{\rho \omega^2} (\underline{\Gamma} + \rho \omega^2 \underline{U}) \circ (\underline{\Gamma} + \rho \omega^2 \underline{U}) \, d\Omega \\ &+ \frac{(1-\gamma)}{2} \int_{\Omega} \text{Tr} \left[(\underline{\sigma} - \mathbf{K} \underline{\varepsilon}(\underline{U})) \mathbf{K}^{-1} (\underline{\sigma} - \mathbf{K} \underline{\varepsilon}(\underline{U})) \right] d\Omega\end{aligned}\quad (6)$$

where γ is a weighting factor belonging to $[0,1]$. A typical value is $\gamma = \frac{1}{2}$. The norms involved in the global constitutive relation error are mechanical norms :

elastic energy norm and kinetic energy norm. Moreover, because $\eta_\omega(s) \geq 0$, it is easy to prove the following proposition :

Proposition 1: Conditions (i) and (i, i) are equivalent for $\gamma \in]0, 1[$

$$\begin{aligned} \text{(i)} \quad & \eta_\omega(s) = 0 \\ \text{(i, i)} \quad & \sigma = K \varepsilon(\underline{U}) \\ & \underline{\Gamma} = -\rho \omega^2 \underline{U} \quad \text{on } \Omega \end{aligned}$$

2.3. New formulation

Recognizing that $\eta_\omega(s)$ is always either positive or equal to zero, the following minimization formulation is reached :

Find $s_\omega = \{\underline{U}, \underline{\Gamma}, \sigma\}_\omega$ admissible ($s_\omega \in S_{ad}^\omega$) that minimizes the global error on the constitutive relation:

$$s \rightarrow \eta_\omega^2(s) \quad (7)$$

$$S_{ad}^\omega$$

Then, an approximate solution to such a new formulation exactly verifies the most reliable equations, which are the equilibrium equations and the kinematic constraints. The least reliable equations, i.e. the constitutive relations, are verified as well as possible.

Proposition 2: If a solution exists to the classical formulation, then the new formulation also has a solution and :

$$\eta_\omega^2(s_\omega) = 0$$

Proof

Let s_ω be a solution to the classical formulation. We then have :

- $s_\omega \in S_{ad}^\omega$
- $\sigma_\omega = K \varepsilon(\underline{U}_\omega)$
- $\underline{\Gamma}_\omega = -\rho \omega^2 \underline{U}_\omega$

It follows that $\eta_\omega(s_\omega) = 0$. Since $\eta_\omega \geq 0$, s_ω is also a solution to the minimization problem (7).

Remark 1: As opposed to the classical formulation, the new formulation yields a solution for ill-posed forced-vibration problems, yet its value of the global error on the constitutive relation can be different from 0. Such ill-posed problems are typical for identification purposes. A common situation is the one in which both displacements and forces are known over the entire boundary $\partial\Omega$ of the structure.

Remark 2: Extension to both damping and nonlinear behavior has recently been performed in (Ladevèze, 1994). Amplitudes then become complex numbers. Damping effects are taken into account by the following constitutive relations:

$$\begin{aligned}\sigma &= \mathbf{K} \varepsilon(\underline{\mathbf{U}}) + \mathbf{B}_\omega \varepsilon(\underline{\mathbf{U}}) \\ \underline{\Gamma} &= -f \omega^2 \underline{\mathbf{U}} + a_\omega \underline{\mathbf{U}}\end{aligned}\quad (8)$$

where \mathbf{B}_ω , a_ω are damping operators that depend both on the material and on ω .

The new formulation displays the same form as (7). The only modification concerns the expression of the global constitutive relation error. Our proposed error has been derived from Drucker's stability inequality (Drucker, 1964) which is verified by most stable materials (Ladevèze, 1985).

3. EXPERIMENTAL DATA

For $\omega \in [\omega_{\min}, \omega_{\max}]$, the excitation is defined by the data :

$$\mathcal{d}_\omega = \left(\underline{\mathbf{U}}_{d,\omega}, f_{d,\omega}, \underline{\mathbf{F}}_{d,\omega} \right) \quad (9)$$

All of the quantities are not necessarily measured data. Some of them are merely reliable informational elements. For the sake of simplicity, let us consider a structure submitted to a single excitation, i.e. to a force excitation at a given point $\underline{\mathbf{P}}$ and in a given direction. Let $F_{d,\omega}$ be its value. It is clear that both the force location and the force direction generally constitute reliable information. Such is not the case for its value $F_{d,\omega}$, with the measured value being $\tilde{F}_{d,\omega}$. More generally, it is possible to consider that the data \mathcal{d}_ω are not perfectly known. \mathcal{d}_ω depends on some quantity p which belongs to a certain space \mathcal{P} (example : $p = F_{d,\omega}$, $\mathcal{P} = \mathbf{R}$).

In addition to the excitation points, both forces and displacements are known for other points and other directions as well. All of the measured quantities are :

$$\mathbb{I} \cdot \tilde{\underline{\Gamma}}_\omega \quad \mathbb{I} \tilde{\underline{\mathbf{U}}}_\omega \quad (10)$$

where Π' , Π are projection operators. They serve to define the unreliable part of the experimental data. The reliable part is that portion of d_ω which does not depend on p ($p \in \mathcal{P}$).

4. QUALITY OF THE DYNAMIC STRUCTURAL MATHEMATICAL MODEL

4.1. Modified error on the constitutive relation at a given frequency ω

First, let us recall that $s_\omega = \{\underline{U}, \underline{\Gamma}, \underline{\sigma}\}_\omega$ has to be admissible, i.e. must verify the reliable equations and experimental data. The data that depend on $p \in \mathcal{P}$ can be introduced by employing some obvious notations :

$$s_\omega \in S_{ad}^\omega(p) \quad p \in \mathcal{P}$$

$$\left[e_\omega(s) \right]^2 = \left[\eta_\omega(s) \right]^2 + \frac{r}{1-r} \left\{ \gamma \left\| \Pi' \underline{\Gamma} - \Pi' \tilde{\Gamma}_\omega \right\|_K^2 + (1-\gamma) \left\| \Pi' \underline{U} - \Pi' \tilde{U}_\omega \right\|_E^2 \right\} \quad (11)$$

The norm $\| \cdot \|_K$ (resp. $\| \cdot \|_E$) is a kinematic energy norm (resp. elastic energy norm) obtained after condensation. The condensation can be static or otherwise, as chosen. r is a confidence coefficient belonging to the interval $[0,1]$; $r = 0.5$ is the current value. With the error measure being defined, the following problem is thus an extension to problem (7) :

Find s_ω admissible and $p \in \mathcal{P}$ which minimizes the modified error on the constitutive relation :

$$\begin{array}{l} s \rightarrow e_\omega(s) \\ S_{ad}^\omega(p) \\ p \in \mathcal{P} \end{array} \quad (12)$$

Remark: In determining the solution s_ω , the most important constraints are the admissibility conditions. They indicate the "possible shapes" of the solutions s_ω as defined on Ω ; these "shapes" are to be compared with the experimental values through the second term of the modified error on the constitutive relation. The first term of this error provides a distance between the model's constitutive relations and the various constitutive relations which can be associated with s_ω .

4.2. Quality of the model

Let us introduce the function z as depending on the frequency ω . $z(\omega)$ is the "weighting factor" at frequency ω . The function is scaled such that :

$$\int_{\omega_{\text{mini}}}^{\omega_{\text{maxi}}} z(\omega) d\omega = 1 \quad z \geq 0 \quad (13)$$

This weighting factor depends on the purpose of the model. The quality of the model is then defined by :

$$\varepsilon^2 = \int_{\omega_{\text{mini}}}^{\omega_{\text{maxi}}} z(\omega) \left[\frac{\eta_{\omega}(s_{\omega})}{D_{\omega}(s_{\omega})} \right]^2 d\omega \quad (14)$$

where $\frac{\eta_{\omega}(s_{\omega})}{D_{\omega}(s_{\omega})}$ is the relative error on the constitutive relation at frequency ω .

$[D_{\omega}(s_{\omega})]$ is a combination of kinetic and elastic energies with the dissipated energy for one cycle.

If the structure is divided into substructures $E \in \mathbf{E}$, the contribution ε_E of the structure E to the error can be easily defined, such that :

$$\varepsilon^2 = \sum_{E \in \mathbf{E}} \varepsilon_E^2 \quad (15)$$

A local error can also be introduced :

$$\varepsilon_{\text{loc}} = \max_{E \in \mathbf{E}} \varepsilon_E \quad (16)$$

Weighting factor examples :

- $z(\omega) = \frac{1}{\omega_{\text{maxi}} - \omega_{\text{mini}}} \quad \text{over } [\omega_{\text{mini}}, \omega_{\text{maxi}}]$
- $z(\omega) = \frac{1}{m} \sum_{i=1}^m \delta_{\omega_i}(\omega)$

(17)

δ_{ω_i} is the Dirac distribution associated with the value ω_i , where ω_i $i \in 1, \dots, m$ denote the free-vibration eigenfrequencies which are included between ω_{mini} and ω_{maxi} .

4.3. An initial example

Figure 1 details the studied problem wherein a single excitation is present. The frequency bandwidth [0-20 Hz] contains the first five eigenfrequencies. "Experimental data" are computed introducing a rigidity error of 100% for element 8. Figure 2 displays, for each element, its contribution ϵ_E to the error for different experimental situations. The first curve is related to the use of maximum measured quantities : the displacement associated with all d.o.f. are measured. It shows an excellent space localization of the defect. In contrast, in using 15% of the measured d.o.f., the defect's detection is far from attaining such a good level.

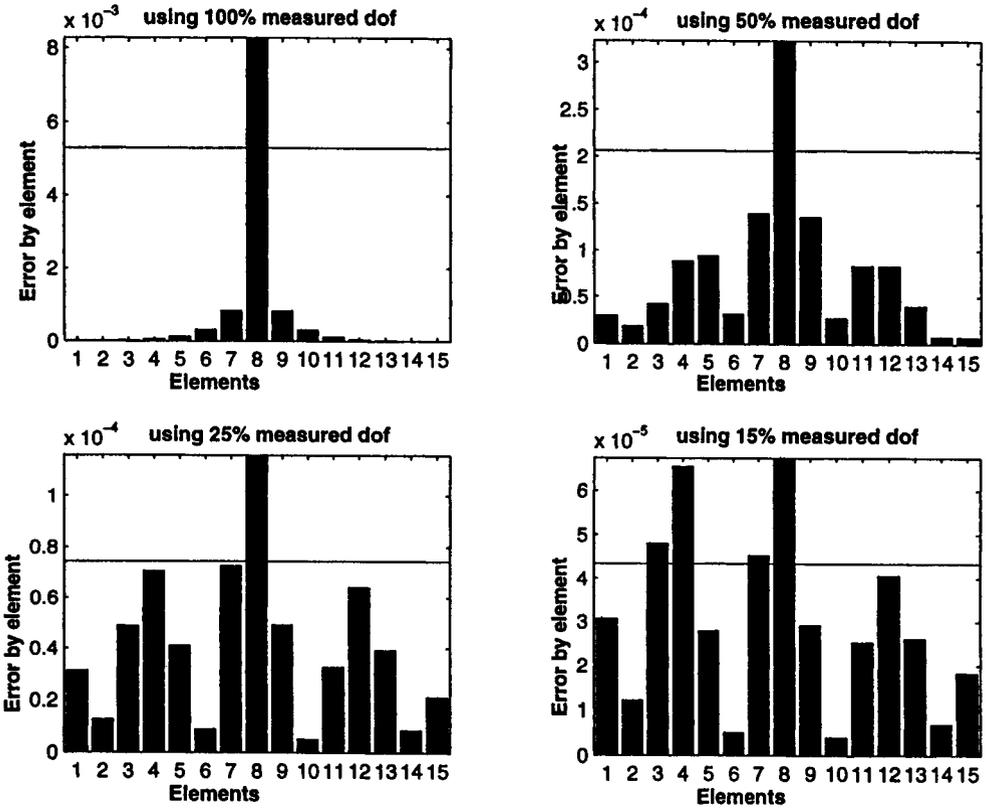


Figure 2 : Effect of the amount of experimental information

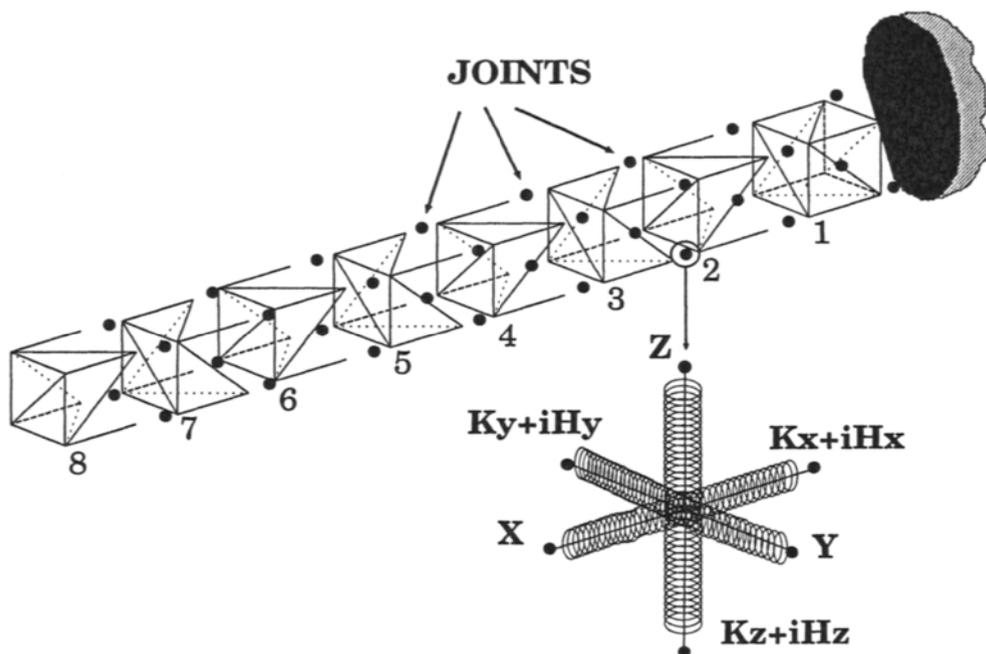


Figure 3 : Beam structure under testing

5. UPDATING METHOD FOR DYNAMIC STRUCTURAL MODELS

5.1. Principle

The mathematical model depends on structural parameters which are not necessarily well-defined : modulus, thickness, damping coefficient, etc. More specifically, these are :

- elastic coefficients: $\underline{K}(\underline{M})$
- mass density: $\rho(\underline{M})$
- damping operators: $\underline{B}_{\omega} \underline{a}_{\omega}$

These structural parameters are denoted by \underline{k} and the corresponding space \mathcal{k} . The updating problem is then to determine for \underline{k} the best value belonging to \mathcal{k} , i.e. minimizing over \mathcal{k} the global modified error on the constitutive relation. The following is then obtained :

Find $\underline{k} \in \mathcal{k}$ which minimizes :

$$\underline{k}' \rightarrow J(\underline{k}') = \int_{\omega_{\min i}}^{\omega_{\max i}} z(\omega) \left[e_{\omega}(s_{\omega}) \right]^2 d \quad (18)$$

As is the case with most inverse problems, this one is also ill-posed. This is why we are proposing herein an adaptive updating method which is iterative with two stages per iteration. Let us consider the n^{th} iteration :

- localization stage at iteration n

The updating procedure is to be stopped if the error ϵ is less than a certain value ϵ_0 , which serves to characterize the quality of the tests. If not, the most erroneous substructures are then detected by the criterion :

$$\epsilon_E \geq 0.8 \max_{E \in E} \epsilon_E \omega \quad (19)$$

Complementary information can also be introduced. This defines a small subset of \mathcal{E} : $\left[\mathcal{E} \right]_n$

- Correction stage at iteration n

The following "small" optimization problem has to be solved :

Find $\underline{k}_n \in \left[\mathcal{E} \right]_n$ which minimizes the modified error on the constitutive relation :

$$\underline{k}' \rightarrow J(\underline{k}')$$

$\left[\mathcal{E} \right]_n$

5.2. Updating examples

Industrial applications concern satellites, launchers, planes, etc. Here, we study the "Eight-Bay Truss" (see figure 3) whose real configuration is described in Figure 4. The "experimental results" are to be computed with such a configuration. Tests conducted and experimental data collected are shown in Figure 5.

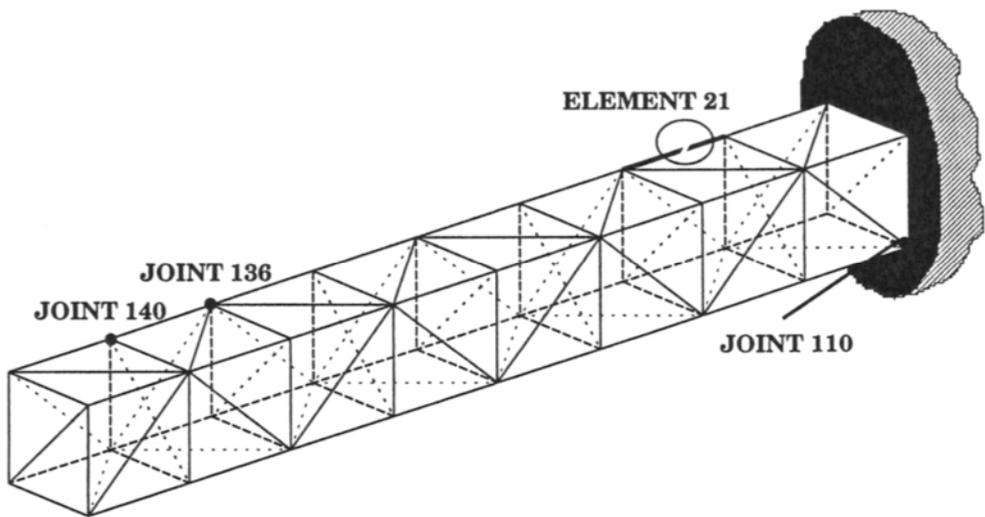


Figure 4 : The real Eight-Bay Truss

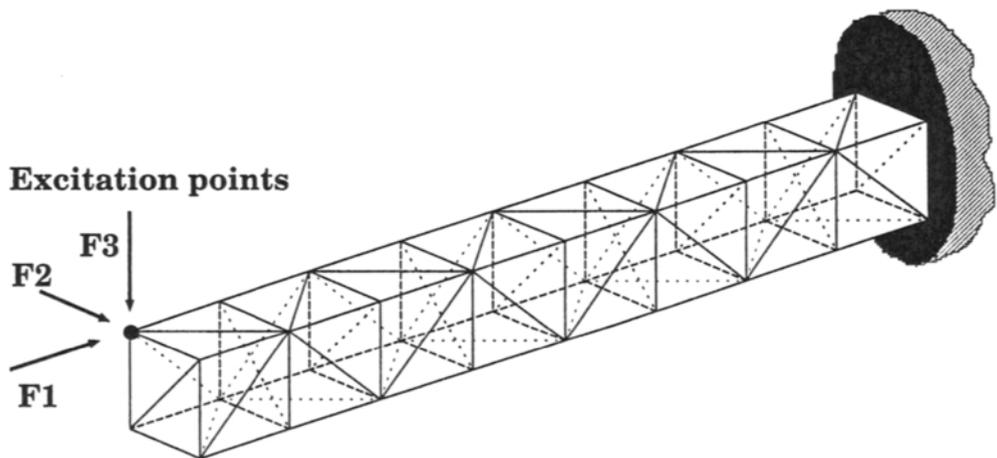


Figure 5 : Tests conducted and experimental data collected

Elements and initial values of the structural parameters	First step	Second step	Exact values
Bar 21 - Stiffness $p_k = 1$ - Mass : $p_m = 1$	$p_k := 0.0002$ $p_m = 1.0467$	corrected	$p_k := 0.0000$ $p_m = 1.0000$
Joint 110 - Stiffness : $p_k = 1$ - Mass : $p_m = 1$	Not localized	$p_k = 0.5025$ $p_m = 1.0007$	$p_k = 0.5000$ $p_m = 1.0000$
Error	1.109	0.147	-

Figure 6 : First and second updating steps

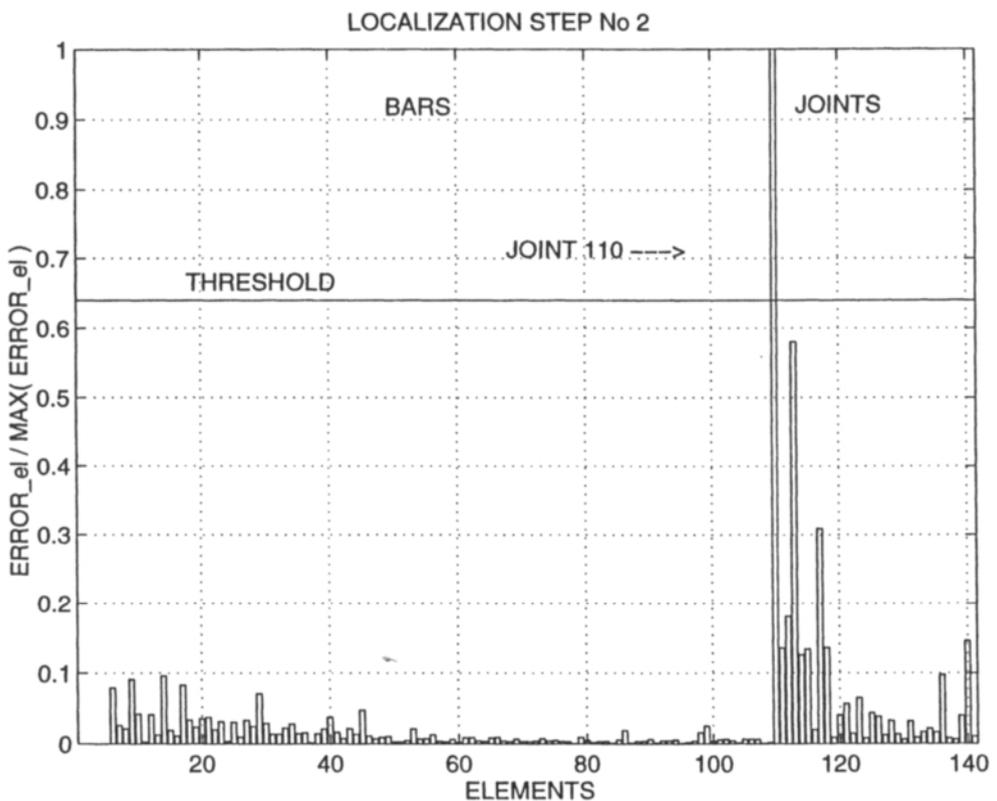


Figure 7 : Error contributions of the various elements - second updating step

Figure 6 presents the structural parametric corrections and the global error value for the first two updating steps. The chart of error contributions to the global error of the various elements is given in Figure 7 for the second updating

step; the joint 110 appears clearly as an incorrectly modelled zone. The updating procedure normally has to be stopped after the second step because the error value is small in comparison with the quality of the experimental results which have been contaminated by noise (level: $\pm 3\%$). During this step, the modelling errors related to joints 136 and 140 have not yet been detected. If one were to proceed further, it would be possible to detect and then correct these errors. Figure 8 shows, for the third updating step, the chart of error contribution to the global error of the various elements; modelling errors for joints 136 and 140 are clearly indicated.

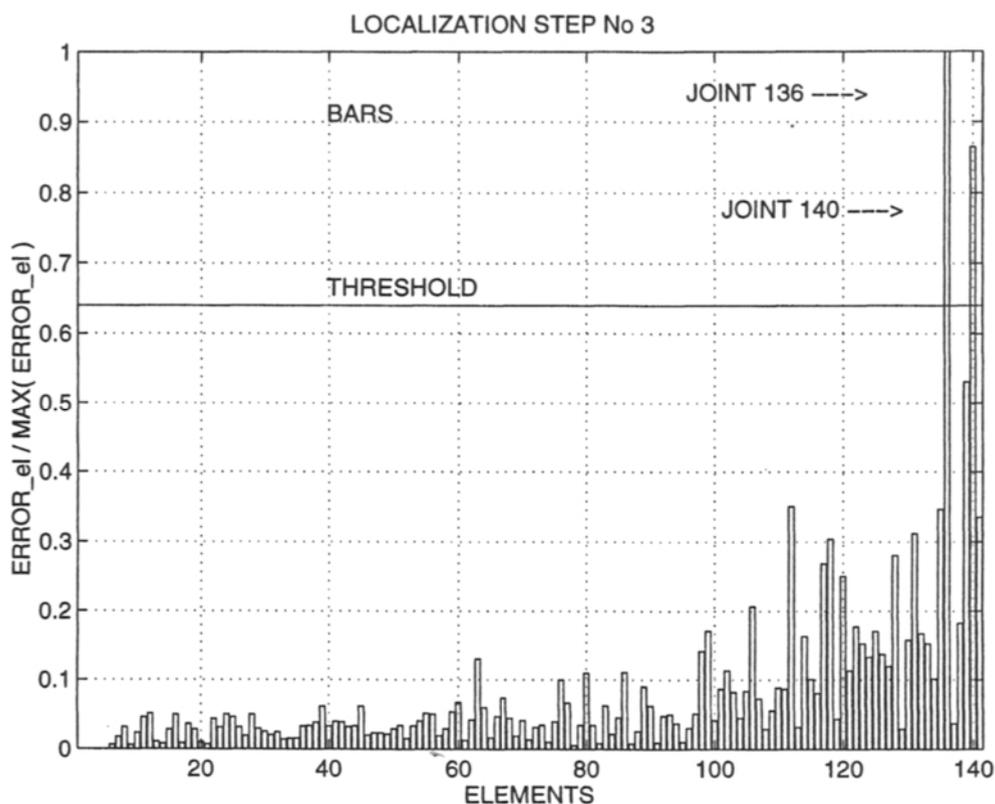


Figure 8 : Error contributions of the various elements - third eventual updating step

6. CONCLUSION

Remaining within the context of missmodelled complicated structures such as launchers, satellites or engines, and subsequent to the two issues raised in the introduction whose responses have been given herein, a third issue can now be raised: how to design a testing program that yields a satisfactory model after updating? This represents one of the key challenges for research over the next 15 years.

REFERENCES

1. H.D. Bui, M. Tanaka *et al.*, 1994, *Inverse Problems in Engineering Mechanics*, A.A. Balkema.
2. J. Bricout, F. Mercier, A. Moulin, P. Ladevèze, D. Nedjar and M. Reynier, 1993, A finite element updating method used with MSC/NASTRAN, *Proceedings of 20th Annual MSC European User's Conference*, Vienna.
3. D.C. Drucker, 1964, On the postulate of stability of material in the mechanics of continua, *J. de Mécanique*, 3, 2, 235-249.
4. P. Ladevèze, 1983, Recalage de modélisations des structures complexes, *Note Technique* n° 33.11.01.4, Aérospatiale, Les Mureaux.
5. P. Ladevèze, 1994, Erreur en relation de comportement en dynamique : théorie et applications au recalage de modèles de structures, *Rapport interne du LMT-Cachan*, n° 150.
6. H. Berger, R. Ohayon, L. Quetin, L. Barthe, P. Ladevèze and M. Reynier, 1991, Méthodes de recalage de modèles de structures en dynamique, *La Recherche Aérospatiale*, 5, 9-20.
7. C. Farhat and F.M. Hemez, 1993, Updating finite element dynamic models using an element-by-element sensitivity methodology, *AIAA Journal*, 31, n°9, 1702-1711.
8. J. Piranda, G. Lallement and S. Cogan, 1991, Parametric correction of finite element models by minimization of an output residual: Improvement of the sensitivity method, *Proceedings of the International Modal Analysis Conference IX*, Florence, Italy.
9. M. Baruch, 1982, Optimal Correction of Mass and Stiffness Matrices Using Measured Modes, *AIAA Journal*, 20, n°11, 1623-1626.
10. S.R. Ibrahim, C. Stavrinidis, E. Fissette and O. Brunner, 1990, A Direct Two-Response Approach for Updating Analytical Dynamic Models of Structures With Emphasis on Uniqueness, *Journal of Vibration and Acoustics*, 112, 107-111.
11. J.E. Mottershead, 1988, A Unified Theory of Recursive, Frequency Domain Filters With Application to System Identification in Structural Dynamics, *Journal of Vibration and Acoustics*, 110, 360-365.

12. P. Ladevèze, M. Reynier and N.M.M. Maia, 1994, Error on the constitutive relation in dynamics, *Inverse Problems in Engineering*, H.D. Bui, M. Tanaka *et al.* (Eds.) Balkema, Rotterdam, 251-256.
13. P. Ladevèze, D. Nedjar and M. Reynier, 1994, Updating of finite element models using vibration tests, *AIAA Journal*, 32, n°7, 1485-1495.
14. H.G. Natke, 1991, Error localisation within a spatially finite-dimensional mathematical model, *Computational Mechanics*, 8, 153-160.
15. F.M. Hemez and C. Farhat, 1995, Bypassing Numerical Difficulties Associated with Updating Simultaneously Mass and Stiffness Matrices, *AIAA Journal*, 33, n°3, 539-546.
16. J.E. Mottershead and M.I. Friswelle, 1993, Model updating in structural dynamics: A survey, *Journal of Sound and Vibration*, 167, n°2, 347-375.
17. M. Kaouk and D.C. Zimmerman, 1994, Structural Damage Assessment Using a Generalized Minimum Rank Perturbation, *AIAA Journal*, 32, n°4, 836-842.
18. A.S. Nobari, D.A. Robb and D.J. Ewins, 1993, Model updating and joint identification methods: Applications, restrictions and overlap, *Journal of Modal Analysis*, 8, n°2, 93-107.
19. P. Ladevèze, 1985, Sur une famille d'algorithmes en mécanique des Structures, *Comptes-Rendus Acad. Sc. Paris*, 300 II, 41-44.
20. A. Chouaki, P. Ladevèze and L. Proslie, 1997, An updating method for damping in structural dynamic models, *Proceedings IMAC XV*, Orlando, Florida, 558-564.
21. A. Berman, 1995, Multiple acceptable solutions in structural model improvement, *AIAA Journal*, 33, n°5, 924-927.

This Page Intentionally Left Blank

PART 3

LOCAL ERROR ESTIMATORS FOR LINEAR PROBLEMS

A-posteriori estimation of the error in the error estimate

I. Babuska^{a1}, T. Strouboulis^{b2}, S.K. Gangaraj^{b2}, K. Copps^{b2} and D.K. Datta^{b2}

^a Texas Institute for Computational and Applied Mathematics,
The University of Texas at Austin, Austin, TX 78712, U.S.A.

^b Department of Aerospace Engineering, Texas A&M University,
College Station, TX 77843-3141, U.S.A.

ABSTRACT

In this paper we address the problem of a-posteriori estimation of the error in the error estimate. We consider the case of estimates for the error in the derivatives, the strains, or the stresses, which are constructed in terms of locally-computed element error indicators of the element residual, or the least-squares recovery type. The estimates of the error in the error estimate have the same structure as the original error estimates, and are computed in terms of indicators of the error in the error indicators, which are determined by locally averaging (*recycling*) the original error indicators. The most accurate indicators of the error in the error indicators are obtained by employing a "harmonic" basis in the recycling of the indicators, namely, a basis which locally satisfies the partial differential equation and the boundary conditions.

¹ The work of this author was supported by the U.S. Office of Naval Research under Grants N00014-90-J-1030 and N00014-96-10019, and by the National Science Foundation under Grants DMS-91-20877 and DMS-95-01841.

² The work of these authors was supported by the U.S. Army Research Office under Grant DAAL03-G-028, by the National Science Foundation under Grant MSS-9025110, by the Texas Advanced Research Program under Grant TARP-71071, and by the U.S. Office of Naval Research under Grant N00014-96-1-0021 and Grant N00014-96-1-1015.

1. A-POSTERIORI ESTIMATION OF THE ERROR IN THE MISES STRESS

Let us consider the plane elasticity problem in a polygonal domain Ω (e.g., the domain shown in Fig. 1 which will be used as the model problem throughout the paper), with boundary $\Gamma \equiv \partial\Omega$, $\bar{\Gamma} = \bar{\Gamma}_D \cup \bar{\Gamma}_N$, $\Gamma_D \cap \Gamma_N = \emptyset$, where Γ_D is the Dirichlet part of the

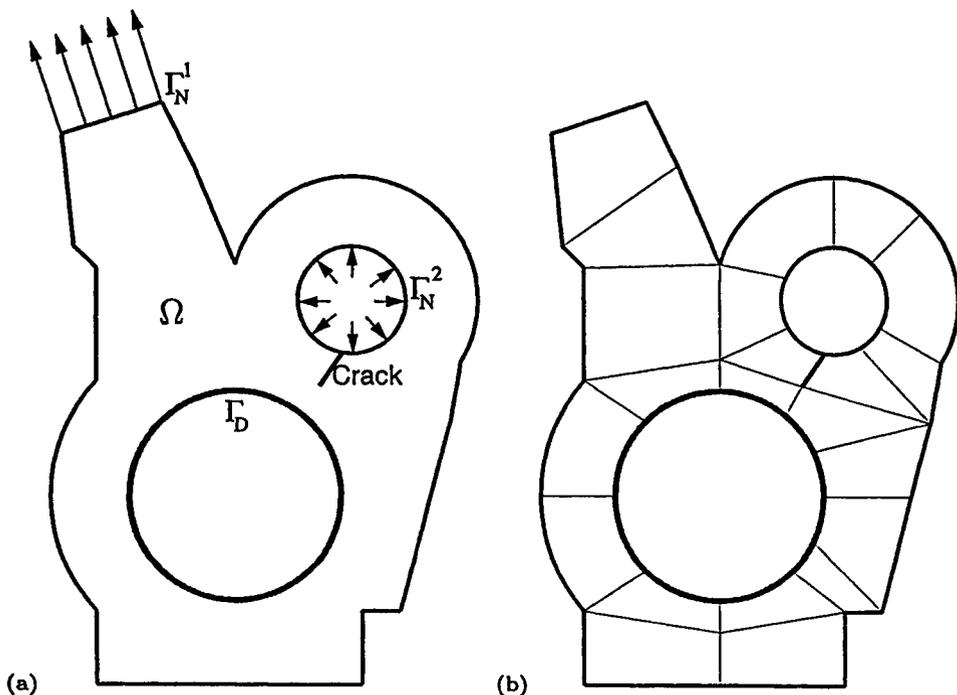


Fig. 1. An example of a polygonal domain. (a) The domain with the applied boundary tractions: $\mathbf{g}|_{\Gamma_N^1} = 5 \mathbf{n}_{\Gamma_N^1}$, $\mathbf{g}|_{\Gamma_N^2} = 3 \mathbf{n}_{\Gamma_N^2}$. (b) The partitioning of the domain into an initial mesh of curvilinear superelements.

boundary, where the displacements are prescribed (here we will assume that the boundary Γ_D is fixed), and Γ_N is the Neumann part of the boundary, where the tractions are prescribed. Let \mathbf{u}^{EX} , denote the exact solution which satisfies the variational problem:

Find $\mathbf{u}^{EX} \in H_{\Gamma_D}^1(\Omega) \stackrel{\text{def}}{=} \left\{ \mathbf{v} = (v_1, v_2) \mid \|\mathbf{v}\|_{\Omega} \stackrel{\text{def}}{=}} \sqrt{\mathcal{B}_{\Omega}(\mathbf{v}, \mathbf{v})} < \infty, \mathbf{v}|_{\Gamma_D} = \mathbf{0} \right\}$ such that

$$\mathcal{B}_{\Omega}(\mathbf{u}^{EX}, \mathbf{v}) \stackrel{\text{def}}{=} \int_{\Omega} \sum_{i,j=1}^2 \sigma_{ij}(\mathbf{u}^{EX}) \epsilon_{ij}(\mathbf{v}) = \mathcal{L}(\mathbf{v}) \stackrel{\text{def}}{=} \int_{\Omega} \mathbf{f} \cdot \mathbf{v} + \int_{\Gamma_N} \mathbf{g} \cdot \mathbf{v} \quad \forall \mathbf{v} \in H_{\Gamma_D}^1(\Omega) \quad (1)$$

Here $\mathfrak{B}_\Omega : \mathbf{H}_{\Gamma_D}^1(\Omega) \times \mathbf{H}_{\Gamma_D}^1(\Omega) \mapsto \mathbb{R}$ is the *energy inner product*; $\mathbf{H}_{\Gamma_D}^1(\Omega)$ is the *space of admissible displacements*; $\mathcal{L} : \mathbf{H}_{\Gamma_D}^1(\Omega) \mapsto \mathbb{R}$ is the *virtual work functional for the applied loads*, which are $\mathbf{f} \in \mathbf{L}^2(\Omega)$, the applied *body force* (in our model problem we will let $\mathbf{f} \equiv \mathbf{0}$), and $\mathbf{g} \in \mathbf{L}^2(\Gamma_N)$, the applied *boundary traction*; $\epsilon_{ij}(\mathbf{u}) \stackrel{\text{def}}{=} \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)$, $i, j = 1, 2$, are the components of infinitesimal strain corresponding to the displacement \mathbf{u} , σ_{ij} , $i, j = 1, 2$, are the components of stress, which are obtained from the strain components by employing Hooke's Law, $\sigma_{ij}(\mathbf{u}) = E_{ijkl} \epsilon_{kl}(\mathbf{u})$.

Let us assume that the domain Ω has been discretized by a mesh T_h which is obtained from the subdivision of an initial mesh of mapped squares (*superelements*, e.g., the mesh of mapped superelements shown in Fig. 1b), let $\mathbf{F}_\tau : \hat{\tau} \mapsto \tau$ denote the mapping which maps the master square (element) $\hat{\tau}$ into the element τ , and $\hat{\mathbf{S}}^p$ denote the space of bi- p polynomials over the master element. The finite element approximation \mathbf{u}^h is the solution of the discrete variational problem:

$$\text{Find } \mathbf{u}^h \in \mathbf{S}_{T_h, \Gamma_D}^p \stackrel{\text{def}}{=} \left\{ \mathbf{v} \in \mathbf{C}^0(\Omega) \mid \mathbf{v}|_\tau \circ \mathbf{F}_\tau \in \hat{\mathbf{S}}^p \quad \forall \tau \in T_h, \quad \mathbf{v}|_{\Gamma_D} = \mathbf{0} \right\} \text{ such that}$$

$$\mathfrak{B}_\Omega(\mathbf{u}^h, \mathbf{v}^h) = \mathcal{L}(\mathbf{v}^h) \quad \forall \mathbf{v}^h \in \mathbf{S}_{T_h, \Gamma_D}^p \quad (2)$$

Letting, $\mathbf{e}^h \stackrel{\text{def}}{=} \mathbf{u}^{EX} - \mathbf{u}^h$, denote the error, substituting $\mathbf{u}^{EX} = \mathbf{e}^h + \mathbf{u}^h$ into (1), and integrating by parts the term $\mathfrak{B}_\tau(\mathbf{u}^h, \mathbf{v})$ in each element τ , we see that the error \mathbf{e}^h is the solution of the *residual equation*:

Find $\mathbf{e}^h \in \mathbf{H}_{\Gamma_D}^1(\Omega)$ such that

$$\mathfrak{B}_\Omega(\mathbf{e}^h, \mathbf{v}) = \mathfrak{R}_{\Omega, \mathbf{u}^h}(\mathbf{v}) \stackrel{\text{def}}{=} \sum_{\tau \in T_h} \left(\int_\tau \mathbf{r}_{\mathbf{u}^h}^\tau \cdot \mathbf{v} + \sum_{\varepsilon \subseteq \partial\tau} \frac{1}{2} \int_\varepsilon \mathbf{J}_{\mathbf{u}^h}^\varepsilon \cdot \mathbf{v} \right) \quad \forall \mathbf{v} \in \mathbf{H}_{\Gamma_D}^1(\Omega) \quad (3)$$

Here $\mathfrak{R}_{\Omega, \mathbf{u}^h} : \mathbf{H}_{\Gamma_D}^1(\Omega) \mapsto \mathbb{R}$ is the virtual work functional for the residual loads, $\mathbf{r}_{\mathbf{u}^h}^\tau$, $\mathbf{J}_{\mathbf{u}^h}^\varepsilon$, $\varepsilon \subseteq \partial\tau$, $\tau \in T_h$, where

$$\mathbf{r}_{\mathbf{u}^h}^\tau \stackrel{\text{def}}{=} \mathbf{f} + \nabla \cdot (\mathbf{E} \boldsymbol{\epsilon}(\mathbf{u}^h)) \quad (4)$$

is the *interior residual* in the element τ (the residual distributed load in the interior of τ), and

$$\mathbf{J}_{\mathbf{u}^h}^\varepsilon(\mathbf{x}) \stackrel{\text{def}}{=} \begin{cases} \left(\boldsymbol{\sigma}(\mathbf{u}^h)|_{\tau^+} - \boldsymbol{\sigma}(\mathbf{u}^h)|_{\tau^-} \right)(\mathbf{x}) \mathbf{n}_\varepsilon(\mathbf{x}), & \mathbf{x} \in \varepsilon = \partial\tau \cap \partial\tau^* \\ 2(\mathbf{g}(\mathbf{x}) - \boldsymbol{\sigma}(\mathbf{u}^h)(\mathbf{x}) \mathbf{n}_{\Gamma_N}(\mathbf{x})), & \mathbf{x} \in \varepsilon \subseteq \Gamma_N \\ \mathbf{0}, & \mathbf{x} \in \varepsilon \subseteq \Gamma_D \end{cases} \quad (5)$$

is the *jump* in the traction on the edge ε (the residual line load on the edge ε). In the first case in (5) ε is an interior edge, τ and τ^* are the elements which share it, and \mathbf{n}_ε is the unit normal on ε which points into τ^* ; in the second case, ε is an edge on the Neumann boundary, and \mathbf{n}_{Γ_N} is the outward unit normal on Γ_N . The orthogonality of the error gives that

$$\mathfrak{R}_{\Omega, \mathbf{u}^h}(\mathbf{v}) = \mathfrak{B}_\Omega(\mathbf{e}^h, \mathbf{v}) = \mathcal{L}(\mathbf{v}) - \mathfrak{B}_\Omega(\mathbf{u}^h, \mathbf{v}) = 0 \quad \forall \mathbf{v} \in \mathbf{S}_{T_h, \Gamma_D}^p \quad (6)$$

namely, *the work done by the residual loads for any admissible displacement from the finite element space $\mathbf{S}_{T_h, \Gamma_D}^p$, vanishes identically.*

Following Ladeveze (see [1, 2, 3]), we will partition the residuals into element groups such that an analogue of the condition (6) holds in each element. For each interior edge $\varepsilon = \partial\tau \cap \partial\tau^*$, let $\mathbf{J}_{\mathbf{u}^h}^{\varepsilon, \tau}$ and $\mathbf{J}_{\mathbf{u}^h}^{\varepsilon, \tau^*}$, be such that

$$\mathbf{J}_{\mathbf{u}^h}^{\varepsilon, \tau} + \mathbf{J}_{\mathbf{u}^h}^{\varepsilon, \tau^*} = \mathbf{J}_{\mathbf{u}^h}^\varepsilon, \quad \varepsilon = \partial\tau \cap \partial\tau^* \quad (7)$$

and, for each edge on the boundary of the domain, $\varepsilon = \partial\tau \cap \Gamma$, let $\mathbf{J}_{\mathbf{u}^h}^{\varepsilon, \tau} = \mathbf{J}_{\mathbf{u}^h}^\varepsilon$. We will call the groups of loads $\mathbf{r}_{\mathbf{u}^h}^\tau$, $\mathbf{J}_{\mathbf{u}^h}^{\varepsilon, \tau}$, $\varepsilon \subseteq \partial\tau$, the *element residuals* of the element τ . Let $\partial\tau_D \stackrel{\text{def}}{=} \partial\tau \cap \Gamma_D$, and

$$\mathbf{S}_{\tau, \partial\tau_D}^p \stackrel{\text{def}}{=} \left\{ \mathbf{v} \in \mathbf{C}^0(\tau) \mid \mathbf{v} \circ \mathbf{F}_\tau \in \hat{\mathbf{S}}^p, \quad \mathbf{v}|_{\partial\tau_D} = \mathbf{0} \right\} \quad (8)$$

be the restriction of the finite element space $\mathbf{S}_{T_h, \Gamma_D}^p$, in the element τ . We will construct the split jumps $\mathbf{J}_{\mathbf{u}^h}^{\varepsilon, \tau}$ such that the virtual work of the element residuals vanishes for any displacement from $\mathbf{S}_{\tau, \partial\tau_D}^p$, namely

$$\mathfrak{R}_{\tau, \mathbf{u}^h}^{EQ,p}(\mathbf{v}) \stackrel{\text{def}}{=} \int_\tau \Gamma_{\mathbf{u}^h}^\tau \cdot \mathbf{v} + \sum_{\varepsilon \subseteq \partial\tau} \int_\varepsilon \mathbf{J}_{\mathbf{u}^h}^{\varepsilon, \tau} \cdot \mathbf{v} = 0 \quad \forall \mathbf{v} \in \mathbf{S}_{\tau, \partial\tau_D}^p \quad (9)$$

Here $\mathfrak{R}_{\tau, \mathbf{u}^h}^{EQ,p} : \mathbf{H}_{\partial\tau_D}^1(\tau) \stackrel{\text{def}}{=} \left\{ \mathbf{v} \in \mathbf{H}^1(\tau) \mid \mathbf{v}|_{\partial\tau_D} = \mathbf{0} \right\} \mapsto \mathbb{R}$, is the *equilibrated element residuum* of the element τ , namely the virtual work functional of the element residuals of the element τ . For each element which does not have an edge or a vertex on the Dirichlet boundary, namely $\partial\tau_D = \partial\tau \cap \Gamma_D = \emptyset$, the mapped affine functions belong to $\mathbf{S}_{\tau, \partial\tau_D}^p$, namely, we have $\hat{\mathbf{w}} \circ \mathbf{F}_\tau^{-1} \in \mathbf{S}_{\tau, \partial\tau_D}^p$ for any $\hat{\mathbf{w}} \in \mathfrak{P}^1(\hat{\tau})$, and letting $\mathbf{v} = \mathbf{i}_\ell \circ \mathbf{F}_\tau^{-1}$, and $\mathbf{v} = \hat{\mathbf{x}}_m \mathbf{i}_\ell \circ \mathbf{F}_\tau^{-1}$ into (9), we obtain the conditions of *equilibrium of forces and moments*

for the element residuals over the master element, namely

$$\left. \begin{aligned} \int_{\tau} \mathbf{r}_{\mathbf{u}^h}^{\tau} + \sum_{\varepsilon \subseteq \partial\tau} \int_{\varepsilon} \mathbf{J}_{\mathbf{u}^h}^{\varepsilon, \tau} &= \mathbf{0} \\ \int_{\tau} \mathbf{F}_{\tau} \times \mathbf{r}_{\mathbf{u}^h}^{\tau} + \sum_{\varepsilon \subseteq \partial\tau} \int_{\varepsilon} \mathbf{F}_{\tau} \times \mathbf{J}_{\mathbf{u}^h}^{\varepsilon, \tau} &= \mathbf{0} \end{aligned} \right\} \quad (10)$$

Note also that from (8) it follows that

$$\mathcal{R}_{\Omega, \mathbf{u}^h}(\mathbf{v}) = \sum_{\tau \in T_h} \mathcal{R}_{\tau, \mathbf{u}^h}^{EQ,p}(\mathbf{v}|_{\tau}) \quad \forall \mathbf{v} \in \mathbf{H}_{\Gamma_D}^1(\Omega) \quad (11)$$

and hence we have an *exact* splitting of the global residuum. In the numerical examples of this paper we employed the constructions of the split jumps $\mathbf{J}_{\mathbf{u}^h}^{\varepsilon, \tau}$ described in [4, 5]. (See also [6] for another method of splitting the jumps, which guarantees (10) but not (9).)

Let us now construct a splitting of the error corresponding to the above splitting of the global residuum. We will denote by $\mathbf{e}^{h, \tau}$, the component of the error due to the residual loads $\mathbf{r}_{\mathbf{u}^h}^{\tau}$, $\mathbf{J}_{\mathbf{u}^h}^{\varepsilon, \tau}$, $\varepsilon \subseteq \partial\tau$, of the element τ , namely, the solution of the variational problem:

Find $\mathbf{e}^{h, \tau} \in \mathbf{H}_{\Gamma_D}^1(\Omega)$ such that

$$\mathcal{B}_{\Omega}(\mathbf{e}^{h, \tau}, \mathbf{v}) = \mathcal{R}_{\tau, \mathbf{u}^h}^{EQ,p}(\mathbf{v}|_{\tau}) \quad \forall \mathbf{v} \in \mathbf{H}_{\Gamma_D}^1(\Omega) \quad (12)$$

In the case $\Gamma_D = \emptyset$, $\mathbf{e}^{h, \tau}$ is defined up to an arbitrary rigid body motion, and the equilibrium condition (10) is a necessary condition for the existence of $\mathbf{e}^{h, \tau}$. Recalling (11) and (3), by *superposition* we have

$$\mathbf{e}^h = \sum_{\tau \in T_h} \mathbf{e}^{h, \tau} \quad (13)$$

We will use the functions $\mathbf{e}^{h, \tau}$ to partition the error into two components with respect to any element of interest. Let $\bar{\tau}$ denote an element of interest, and let

$$\omega_{\bar{\tau}}^h \stackrel{\text{def}}{=} \left\{ \tau \in T_h \mid \partial\tau \cap \partial\bar{\tau} \neq \emptyset \right\} \quad (14)$$

be the patch of elements which are connected to the vertices of $\bar{\tau}$ (see, for example, Fig. 2, where the element $\bar{\tau}$ is shown shaded dark gray, and the patch $\omega_{\bar{\tau}}^h$ consists of $\bar{\tau}$ and the elements shown shaded light gray). We will employ the splitting

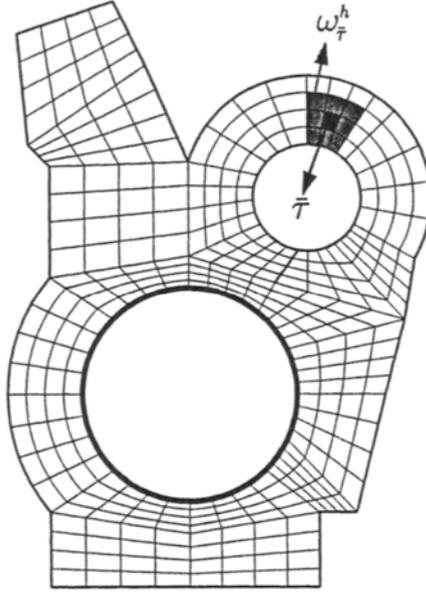


Fig. 2. *Local error and pollution error.* The patch $\omega_{\bar{\tau}}^h$ (light gray) which corresponds to the element $\bar{\tau}$; the patch consists of the elements connected to the vertices of $\bar{\tau}$, including $\bar{\tau}$.

$$\mathbf{e}^h = \mathbf{e}_{\bar{\tau}}^{h,loc} + \mathbf{e}_{\bar{\tau}}^{h,glob} \quad (15a)$$

where

$$\mathbf{e}_{\bar{\tau}}^{h,loc} \stackrel{\text{def}}{=} \sum_{\substack{\tau \in T_h \\ \tau \subseteq \omega_{\bar{\tau}}^h}} \mathbf{e}^{h,\tau}, \quad \mathbf{e}_{\bar{\tau}}^{h,glob} \stackrel{\text{def}}{=} \sum_{\substack{\tau \in T_h \\ \tau \subseteq \Omega - \omega_{\bar{\tau}}^h}} \mathbf{e}^{h,\tau} \quad (15b)$$

and we will call $\mathbf{e}_{\bar{\tau}}^{h,loc}|_{\bar{\tau}}$, the restriction of $\mathbf{e}_{\bar{\tau}}^{h,loc}$ in the element $\bar{\tau}$, the *local* or *near-field* error in $\bar{\tau}$, and $\mathbf{e}_{\bar{\tau}}^{h,glob}|_{\bar{\tau}}$, the restriction of $\mathbf{e}_{\bar{\tau}}^{h,glob}$ in the element $\bar{\tau}$, the *pollution* or *far-field* error in $\bar{\tau}$.

Remark 1. The component $\mathbf{e}_{\bar{\tau}}^{h,loc}$ is the solution of the *near-field* residual equation:

Find $\mathbf{e}_{\bar{\tau}}^{h,loc} \in \mathbf{H}_{\Gamma_D}^1(\Omega)$ such that

$$\mathfrak{B}_{\Omega}(\mathbf{e}_{\bar{\tau}}^{h,loc}, \mathbf{v}) = \mathfrak{R}^{\bar{\tau},loc}(\mathbf{v}) \stackrel{\text{def}}{=} \sum_{\substack{\tau \in T_h \\ \tau \subseteq \omega_{\bar{\tau}}^h}} \mathfrak{R}_{\tau, \mathbf{u}^h}^{EQ,p}(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{H}_{\Gamma_D}^1(\Omega) \quad (16)$$

Similarly, $\mathbf{e}_{\bar{\tau}}^{h,glob}$ is the solution of the *far-field* residual equation:

Find $\mathbf{e}_{\bar{\tau}}^{h,glob} \in \mathbf{H}_{\Gamma_D}^1(\Omega)$ such that

$$\mathfrak{B}_{\Omega}(\mathbf{e}_{\bar{\tau}}^{h,glob}, \mathbf{v}) = \mathfrak{R}^{\bar{\tau},glob}(\mathbf{v}) \stackrel{\text{def}}{=} \sum_{\substack{\tau \in T_h \\ \tau \subseteq \Omega - \omega_{\bar{\tau}}^h}} \mathfrak{R}_{\tau, \mathbf{u}^h}^{EQ,p}(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{H}_{\Gamma_D}^1(\Omega) \quad (17)$$

Remark 2. Let us underline that *the near-field residual equation (16) is never solved in practical computations*. Instead, the local error is estimated by an element error indicator which is determined, at a negligible cost, using local computations in $\omega_{\bar{\tau}}^h$.

Remark 3. In the numerical examples we will need $e_{\bar{\tau}}^{h,loc}$ in order to analyze the accuracy of the locally computed element error indicators, and we will employ the following *overkill* approximation:

Find $e_{\bar{\tau},ovk}^{h,loc} \in \mathbf{S}_{T_h^{ref},\Gamma_D}^{p+m}$ such that

$$\mathcal{B}_\Omega(e_{\bar{\tau},ovk}^{h,loc}, \mathbf{v}) = \mathcal{R}^{\bar{\tau},loc}(\mathbf{v}) = \sum_{\substack{\tau \in T_h \\ \tau \subseteq \omega_{\bar{\tau}}^h}} \mathcal{R}_{\tau,u^h}^{EQ,p}(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{S}_{T_h^{ref},\Gamma_D}^{p+m} \quad (18)$$

Here T_h^{ref} is a mesh obtained from T_h by refining several times the elements with a vertex at a corner point, and $m \geq 1$. The accuracy of $e_{\bar{\tau},ovk}^{h,loc}$ should be sufficiently high, such that it does not influence the conclusions of the analysis of the local error indicator.

Remark 4. Similarly, we will analyze estimates of the pollution error $e_{\bar{\tau}}^{h,glob}$, by comparing it to $e_{\bar{\tau},ovk}^{h,glob}$, the approximate solution of the far-field residual equation which satisfies:

Find $e_{\bar{\tau},ovk}^{h,glob} \in \mathbf{S}_{T_h^{ref},\Gamma_D}^{p+m}$ such that

$$\mathcal{B}_\Omega(e_{\bar{\tau},ovk}^{h,glob}, \mathbf{v}) = \mathcal{R}^{\bar{\tau},glob}(\mathbf{v}) = \sum_{\substack{\tau \in T_h \\ \tau \subseteq \Omega - \omega_{\bar{\tau}}^h}} \mathcal{R}_{\tau,u^h}^{EQ,p}(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{S}_{T_h^{ref},\Gamma_D}^{p+m} \quad (19)$$

The overkill approximation (19) can be used as a practical method for approximating the pollution error within the context of iterative solvers; see Oden and Feng [7].

Let us assume that we are interested in estimating the stresses in the element $\bar{\tau}$. We will employ the splitting

$$\boldsymbol{\sigma}(e^h)(\bar{\mathbf{x}}) = \underbrace{\left(\boldsymbol{\sigma}(e^h)(\bar{\mathbf{x}}) - \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(e^h) \right)}_{\text{local error}} + \underbrace{\frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(e^h)}_{\text{pollution error}} \quad (20)$$

We call the first term the local error because

$$\boldsymbol{\sigma}(e^h)(\bar{\mathbf{x}}) - \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(e^h) \approx \boldsymbol{\sigma}(e_{\bar{\tau}}^{h,loc}) \approx \hat{\mathbf{e}}_{\boldsymbol{\sigma}(u^h)}^{REC}(\bar{\mathbf{x}}) \quad (21)$$

and the second term the pollution error because

$$\frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(e^h) \approx \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(e_{\bar{\tau}}^{h,glob}) \quad (22)$$

Remark 5. Let us justify (22). Note that

$$\frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(\mathbf{e}^h) = \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(\mathbf{e}_{\bar{\tau}}^{h,loc}) + \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(\mathbf{e}_{\bar{\tau}}^{h,glob}) \quad (23)$$

First, let $\bar{\tau}$ be an element in the interior of the mesh, and assume that the meshes T_h are *quasi-uniform*, namely, we have

$$\gamma h \leq h_{\tau} \leq h \stackrel{\text{def}}{=} \max_{\tau \in T_h} h_{\tau} \quad \tau \in T_h \quad (24)$$

where $0 < \gamma \leq 1$ is a fixed parameter which characterizes the entire sequence of meshes. Further, assume that the meshes are sufficiently refined in the neighborhood of $\bar{\tau}$ (see [8] for the precise assumptions). In [8] we have proven that the local error is *locally periodic*, up to higher order terms, and we have

$$\left\| \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(\mathbf{e}_{\bar{\tau}}^{h,loc}) \right\|_M \approx \mathcal{C} h^{p+2} \quad (25)$$

where $\|\cdot\|_M$ is any matrix norm, e.g., the norm which chooses the maximum of the entries, and because

$$\left\| \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(\mathbf{e}_{\bar{\tau}}^{h,glob}) \right\|_M \approx \mathcal{C} h^{p+1} \quad (26)$$

the average local error may be neglected.

Remark 6. The majority of engineering computations are performed for the approximation of solutions of boundary value problems in complex domains (curvilinear polygonal domains with one or several corner points, or corner points smoothed by sharp fillets). Assume that the domain has n_{cp} corner points and let $\alpha \stackrel{\text{def}}{=} \min_{j=1,\dots,n_{cp}} \alpha_j^i$ be the smallest exponent in the asymptotic expansion

$$\mathbf{u}^{EX}(\bar{\mathbf{x}}) = \sum_{j=1}^{n_{cp}} \left(\sum_{i=1}^{\infty} \mathcal{C} (r_j(\bar{\mathbf{x}}))^{\alpha_i^j} \boldsymbol{\Phi}_i^j(\theta_j(\bar{\mathbf{x}})) \right) + \mathbf{u}_0^{EX}(\bar{\mathbf{x}}) \quad (27)$$

where $r_j(\bar{\mathbf{x}}), \theta_j(\bar{\mathbf{x}})$ are the polar coordinates of $\bar{\mathbf{x}}$ centered at the j th corner point, $\boldsymbol{\Phi}_i^j$, $i = 1, 2, \dots$, are the eigenfunctions for the infinite wedge corresponding to the j th corner point (see pp. 176–177 in [9]), and \mathbf{u}_0^{EX} is a smooth function which vanishes in the neighborhood of the corner points. Then, it can be proven (see [10]) that

$$\left\| \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(\mathbf{e}_{\bar{\tau}}^{h,glob}) \right\|_M \approx \mathcal{C} h^{2\alpha} \quad (28)$$

It follows that, in the cases that $2\alpha < p$, we have

$$\left\| \boldsymbol{\sigma}(\mathbf{e}_{\bar{\tau}}^{h,loc})(\bar{\mathbf{x}}) \right\|_M \approx \mathcal{C} h^p \ll \left\| \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(\mathbf{e}_{\bar{\tau}}^{h,glob}) \right\|_M \approx \mathcal{C} h^{2\alpha} \quad (29)$$

which means that *the local error at \bar{x} is negligible when compared to the element average of the pollution error*. Note that the condition $2\alpha < p$ is realized in the majority of engineering computations. Consider, for example, the case that the domain has a reentrant corner (e.g., the domain in Fig. 1); then, we have $\alpha < 1$, and for elements of quadratic degree or higher ($p \geq 2$), we have $2\alpha < 2 \leq p$. The above arguments also hold in the case that $\bar{\tau}$ is adjacent to a smooth boundary (see [11]). In the case of an element $\bar{\tau}$ with a vertex at j th corner point, we have

$$\left\| \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(\mathbf{e}_{\bar{\tau}}^{h,loc}) \right\|_M \approx \mathcal{C} h^{\alpha_1} \quad (30)$$

while

$$\left\| \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \boldsymbol{\sigma}(\mathbf{e}_{\bar{\tau}}^{h,glob}) \right\|_M \approx \mathcal{C} h^{2\alpha'} \quad (31)$$

where $\alpha' = \min_{\substack{k=1,\dots,n_{cp} \\ k \neq j}} \alpha_1^k$. It follows that, if $\alpha_1^j > 2\alpha'$, the average local error can be neglected.

Remark 7. Below, we will construct estimates for the average values of the derivatives of the pollution error in the element $\bar{\tau}$. From Remarks 6 and 7, it follows that these estimates can also be interpreted as estimates of the average error.

Let $\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}$ be the Green's function which corresponds to the average ij component of the stress in the element $\bar{\tau}$, namely, the solution of the variational problem:

Find $\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} \in \mathbf{H}_{\Gamma_D}^1(\Omega)$ such that

$$\mathfrak{B}_{\Omega}(\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}, \mathbf{v}) = \int_{\bar{\tau}} \sigma_{ij}(\mathbf{v}) = \int_{\partial\bar{\tau}} E_{ijkl} \epsilon_{kl}(\mathbf{v}) = \int_{\partial\bar{\tau}} E_{ijkl} (v_k n_{\partial\bar{\tau},\ell} + v_{\ell} n_{\partial\bar{\tau},k}) \quad \forall \mathbf{v} \in \mathbf{H}_{\Gamma_D}^1(\Omega) \quad (32)$$

where $\mathbf{n}_{\partial\bar{\tau}}$ is the exterior unit-normal on $\partial\bar{\tau}$. Letting $\mathbf{v} = \mathbf{e}_{\bar{\tau}}^{h,glob}$, into (32), employing the symmetry of the bilinear form $\mathfrak{B}_{\Omega}(\cdot, \cdot)$ and (19) we get

$$\frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \sigma_{ij}(\mathbf{e}_{\bar{\tau}}^{h,glob}) = \sum_{\substack{\tau \in T_h \\ \tau \subseteq \Omega - \omega_{\bar{\tau}}^h}} \mathfrak{R}_{\tau, \omega_{\bar{\tau}}^h}^{EQ,p}(\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}) = \sum_{\substack{\tau \in T_h \\ \tau \subseteq \Omega - \omega_{\bar{\tau}}^h}} \mathfrak{R}_{\tau, \omega_{\bar{\tau}}^h}^{EQ,p}(\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \mathcal{A}_{\omega_{\bar{\tau}}^h}^p \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}) \quad (33)$$

where $\mathcal{A}_{\omega_{\bar{\tau}}^h}^p \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}$ is the best approximation of $\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}$ from the restriction of the finite element space in $\omega_{\bar{\tau}}^h$. The second equality in (33), follows from the orthogonality condition $\mathfrak{R}_{\tau, \omega_{\bar{\tau}}^h}^{EQ,p}(\mathbf{v}) = 0$, for any $\mathbf{v} \in \mathbf{S}_{\bar{\tau}, \partial\bar{\tau}_D}^p$, which follows from the splitting of the jumps.

Let us also construct another representation of the average values of the derivatives of the pollution error. For each element $\tau \in T_h$, let $\hat{\mathbf{e}}_{\tau, \omega_{\bar{\tau}}^h}^{ERp}$, be the exact solution of the *element residual problem*:

Find $\hat{\mathbf{e}}_{\tau, \mathbf{u}^h}^{ERp} \in \mathbf{H}_{\partial\tau_D}^1(\tau)$ such that

$$\mathfrak{B}_\tau(\hat{\mathbf{e}}_{\tau, \mathbf{u}^h}^{ERp}, \mathbf{v}) = \mathfrak{R}_{\tau, \mathbf{u}^h}^{EQ,p}(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{H}_{\partial\tau_D}^1(\tau) \quad (34)$$

Remark 8. In the elements for which $\partial\tau_D = \partial\tau \cap \Gamma_D = \emptyset$, the function $\hat{\mathbf{e}}_{\tau, \mathbf{u}^h}^{ERp}$ is determined only up to a rigid body motion, and (10) is a necessary condition for its existence.

Employing (33) and (34), we get

$$\frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \sigma_{ij}(\mathbf{e}_{\bar{\tau}}^{h, glob}) = \sum_{\substack{\tau \in T_h \\ \tau \subseteq \Omega - \omega_{\bar{\tau}}^h}} \mathfrak{B}_\tau(\hat{\mathbf{e}}_{\tau, \mathbf{u}^h}^{ERp}, \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \mathcal{A}_{\omega_{\bar{\tau}}^h}^p \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}) \quad (35)$$

We will call the contribution

$$\mu_{\tau, \sigma_{ij}}^{EQ,p}(\bar{\tau}) \stackrel{\text{def}}{=} \mathfrak{R}_{\tau, \mathbf{u}^h}^{EQ,p}(\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \mathcal{A}_{\omega_{\bar{\tau}}^h}^p \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}) = \mathfrak{B}_\tau(\hat{\mathbf{e}}_{\tau, \mathbf{u}^h}^{ERp}, \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \mathcal{A}_{\omega_{\bar{\tau}}^h}^p \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}) \quad (36)$$

the *exact pollution* indicator in the element τ , corresponding to the ij component of the average stress in the element $\bar{\tau}$. We can then write (33) and (35) in the concise form

$$\frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \sigma_{ij}(\mathbf{e}_{\bar{\tau}}^{h, glob}) = \sum_{\substack{\tau \in T_h \\ \tau \subseteq \Omega - \omega_{\bar{\tau}}^h}} \mu_{\tau, \sigma_{ij}}^{EQ,p}(\bar{\tau}) \quad (37)$$

We will base the a-posteriori estimates for the derivatives, strains and stress of the pollution error on (35)–(37).

Let $\check{\mathbf{G}}_{\bar{\tau}, h}^{\sigma_{ij}}$ be the finite element approximation of $\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}$ from $\mathbf{S}_{T_h, \Gamma_D}^p$, namely, the solution of the discrete variational problem:

Find $\check{\mathbf{G}}_{\bar{\tau}, h}^{\sigma_{ij}} \in \mathbf{S}_{T_h, \Gamma_D}^p$ such that

$$\mathfrak{B}_\Omega(\check{\mathbf{G}}_{\bar{\tau}, h}^{\sigma_{ij}}, \mathbf{v}) = \int_{\partial\bar{\tau}} E_{ijkl}(v_k n_{\partial\bar{\tau}, \ell} + v_\ell n_{\partial\bar{\tau}, k}) \quad \forall \mathbf{v} \in \mathbf{S}_{T_h, \Gamma_D}^p \quad (38)$$

Then, we expect that

$$\boldsymbol{\sigma}(\hat{\mathbf{e}}_{\tau, \check{\mathbf{G}}_{\bar{\tau}, h}^{\sigma_{ij}}}^{ERp}) \approx \boldsymbol{\sigma}(\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \mathcal{A}_{\omega_{\bar{\tau}}^h}^p \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \boldsymbol{\chi}_{\bar{\tau}}^p) \quad (39)$$

where $\boldsymbol{\chi}_{\bar{\tau}}^p \in \mathbf{S}_{\tau, \partial\tau_D}^p$ is the local best approximation of the pollution error in $\check{\mathbf{G}}_{\bar{\tau}, h}^{\sigma_{ij}}$, and hence we can employ the approximation

$$\mathfrak{R}_{\tau, \mathbf{u}^h}^{EQ,p}(\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \mathcal{A}_{\omega_{\bar{\tau}}^h}^p \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}) = \mathfrak{R}_{\tau, \mathbf{u}^h}^{EQ,p}(\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \mathcal{A}_{\omega_{\bar{\tau}}^h}^p \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \boldsymbol{\chi}_{\bar{\tau}}^p) \approx \mathfrak{R}_{\tau, \mathbf{u}^h}^{EQ,p}(\hat{\mathbf{e}}_{\tau, \check{\mathbf{G}}_{\bar{\tau}, h}^{\sigma_{ij}}}^{ERp}) \quad (40)$$

Remark 9. Note that (39) and (40) do not hold in general (see [12] for a counterexample). However, they hold for the types of meshes and solutions encountered in engineering computations.

It follows that the exact pollution indicators, $\mu_{\tau,\sigma_{ij}}^{EQ,p}(\bar{\tau})$, can be approximated by the *computed* pollution indicators

$$\hat{\mu}_{\tau,\sigma_{ij}}^{EQ,p}(\bar{\tau}; ERp) \stackrel{\text{def}}{=} \mathfrak{G}_{\tau,\mathbf{u}^h}^{EQ,p}(\hat{\mathbf{e}}_{\tau,\mathbf{G}_{\sigma_{ij}}^{ERp}}^{ERp}) = \mathfrak{B}_{\tau}(\hat{\mathbf{e}}_{\tau,\mathbf{u}^h}^{ERp}, \hat{\mathbf{e}}_{\tau,\mathbf{G}_{\sigma_{ij}}^{ERp}}^{ERp}) \quad (41)$$

and we obtain the following estimate:

$$\frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \sigma_{ij}(\mathbf{e}_{\bar{\tau}}^{h,glob}) \approx \mathfrak{C}_{\sigma_{ij}^{avg(\bar{\tau})}}^{ERp} \stackrel{\text{def}}{=} \sum_{\substack{\tau \in T_h \\ \tau \subseteq \Omega - \omega_{\bar{\tau}}^{\sharp}}} \hat{\mu}_{\tau,\sigma_{ij}}^{EQ,p}(\bar{\tau}; ERp) \quad (42)$$

Remark 10. The function $\hat{\mathbf{e}}_{\tau,\mathbf{G}_{\sigma_{ij}}^{ERp}}^{ERp}$ cannot be obtained exactly, in general, and must be approximated in order to compute the pollution estimate; see (52)–(54) below.

Remark 11. Noting that

$$|\mu_{\tau,\sigma_{ij}}^{EQ,p}(\bar{\tau}) - \hat{\mu}_{\tau,\sigma_{ij}}^{EQ,p}(\bar{\tau}; ERp)| = \left| \mathfrak{B}_{\tau}(\hat{\mathbf{e}}_{\tau,\mathbf{u}^h}^{ERp}, (\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \mathcal{A}_{\omega_{\bar{\tau}}^{\sharp}}^p \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \boldsymbol{\chi}_{\bar{\tau}}^p) - \hat{\mathbf{e}}_{\tau,\mathbf{G}_{\sigma_{ij}}^{ERp}}^{ERp}) \right| \quad (43)$$

and using the orthogonality condition $\mathfrak{B}_{\tau}(\hat{\mathbf{e}}_{\tau,\mathbf{u}^h}^{ERp}, \boldsymbol{\chi}_{\bar{\tau}}^p) = \mathfrak{G}_{\tau,\mathbf{u}^h}^{EQ,p}(\boldsymbol{\chi}_{\bar{\tau}}^p) = 0$, and the Cauchy-Schwarz inequality we get

$$|\mu_{\tau,\sigma_{ij}}^{EQ,p}(\bar{\tau}) - \hat{\mu}_{\tau,\sigma_{ij}}^{EQ,p}(\bar{\tau}; ERp)| \leq ||| \hat{\mathbf{e}}_{\tau,\mathbf{u}^h}^{ERp} |||_{\bar{\tau}} ||| (\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \mathcal{A}_{\omega_{\bar{\tau}}^{\sharp}}^p \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}) - \hat{\mathbf{e}}_{\tau,\mathbf{G}_{\sigma_{ij}}^{ERp}}^{ERp} |||_{\bar{\tau}} \quad (44)$$

Hence, *the accuracy of the computed pollution indicators and the pollution estimate depends on the magnitude of the term $||| (\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \mathcal{A}_{\omega_{\bar{\tau}}^{\sharp}}^p \check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}) - \hat{\mathbf{e}}_{\tau,\mathbf{G}_{\sigma_{ij}}^{ERp}}^{ERp} |||_{\bar{\tau}}$, namely, the difference between the employed element error indicator and the error in the local best approximation of the extraction function $\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}}$.*

Remark 12. The exact pollution indicator can also be written in the form

$$\mu_{\tau,\sigma_{ij}}^{EQ,p}(\bar{\tau}) = \mathfrak{G}_{\tau,\mathbf{u}^h}^{EQ,p}(\check{\mathbf{G}}_{\bar{\tau}}^{\sigma_{ij}} - \check{\mathbf{G}}_{\bar{\tau},h}^{\sigma_{ij}}) \quad (45)$$

Above, we employed the representation (36) in terms of the error in the local best approximation, because *it underlines that it is not necessary to estimate the pollution error in $\check{\mathbf{G}}_{\bar{\tau},h}^{\sigma_{ij}}$.*

Let us also construct an estimate for the energy-norm of the error in the element $\bar{\tau}$.

We have

$$|||e^h|||_{\bar{\tau}} = \sqrt{|||e_{\bar{\tau}}^{h,loc}|||_{\bar{\tau}}^2 + |||e_{\bar{\tau}}^{h,glob}|||_{\bar{\tau}}^2 + 2\mathfrak{B}_{\bar{\tau}}(e_{\bar{\tau}}^{h,loc}, e_{\bar{\tau}}^{h,glob})} \quad (46)$$

Noting that

$$|\mathfrak{B}_{\bar{\tau}}(e_{\bar{\tau}}^{h,loc}, e_{\bar{\tau}}^{h,glob})| = |\mathfrak{B}_{\bar{\tau}}(e_{\bar{\tau}}^{h,loc}, e_{\bar{\tau}}^{h,glob} - \mathcal{A}_{\bar{\tau}}^p e_{\bar{\tau}}^{h,glob})| \quad (47)$$

it follows that for sufficiently refined meshes we have

$$|\mathfrak{B}_{\bar{\tau}}(e_{\bar{\tau}}^{h,loc}, e_{\bar{\tau}}^{h,glob})| \leq |||e_{\bar{\tau}}^{h,loc}|||_{\bar{\tau}} |||e_{\bar{\tau}}^{h,glob} - \mathcal{A}_{\bar{\tau}}^p e_{\bar{\tau}}^{h,glob}|||_{\bar{\tau}} \ll |||e_{\bar{\tau}}^{h,loc}|||_{\bar{\tau}} |||e_{\bar{\tau}}^{h,glob}|||_{\bar{\tau}} \quad (48)$$

See also [13] for numerical studies which illustrate (48). Hence, we can employ the approximation

$$|||e^h|||_{\bar{\tau}} \approx \sqrt{|||e_{\bar{\tau}}^{h,loc}|||_{\bar{\tau}}^2 + |||e_{\bar{\tau}}^{h,glob}|||_{\bar{\tau}}^2} \approx \mathfrak{E}_{\bar{\tau}}^{ERp} \stackrel{\text{def}}{=} \sqrt{(\mathfrak{E}_{\bar{\tau},loc}^{ERp})^2 + (\mathfrak{E}_{\bar{\tau},glob}^{ERp})^2} \quad (49)$$

where we have used the estimates

$$|||e_{\bar{\tau}}^{h,loc}|||_{\bar{\tau}} \approx \mathfrak{E}_{\bar{\tau},loc}^{ERp} \stackrel{\text{def}}{=} |||\hat{e}_{\bar{\tau},u^h}^{ERp}|||_{\bar{\tau}} \quad (50)$$

and

$$|||e_{\bar{\tau}}^{h,glob}|||_{\bar{\tau}} \approx \mathfrak{E}_{\bar{\tau},glob}^{ERp} \stackrel{\text{def}}{=} \sqrt{|\bar{\tau}| \sum_{i,j,k,l=1}^2 E_{ijkl}^{-1} \mathfrak{E}_{\sigma_{ij}^{avg(\bar{\tau})}}^{ERp} \mathfrak{E}_{\sigma_{kl}^{avg(\bar{\tau})}}^{ERp}} \quad (51)$$

The error indicators $\hat{e}_{\bar{\tau},u^h}^{ERp}$ will be approximated by

$$\hat{e}_{\bar{\tau},u^h}^{ERpBp+m} \stackrel{\text{def}}{=} \pi_{\mathbf{B}_{\bar{\tau},\partial\tau_D}^{p+m}}(\hat{e}_{\bar{\tau},u^h}^{ERp}) \quad (52)$$

where $\pi_{\mathbf{B}_{\bar{\tau},\partial\tau_D}^{p+m}} : H_{\partial\tau_D}^1(\tau) \mapsto \mathbf{B}_{\bar{\tau},\partial\tau_D}^{p+m} \stackrel{\text{def}}{=} \mathbf{S}_{\bar{\tau},\partial\tau_D}^{p+m} - \mathbf{S}_{\bar{\tau},\partial\tau_D}^p$ denotes the *energy-projection* into the *bubble space* $\mathbf{B}_{\bar{\tau},\partial\tau_D}^{p+m}$ (see Oden, Demkowicz, and coworkers [14]), and $\hat{e}_{\bar{\tau},u^h}^{ERpBp+m}$ will be determined by solving the following discrete element residual problem:

Find $\hat{e}_{\bar{\tau},u^h}^{ERpBp+m} \in \mathbf{B}_{\bar{\tau},\partial\tau_D}^{p+m}$ such that

$$\mathfrak{B}_{\bar{\tau}}(\hat{e}_{\bar{\tau},u^h}^{ERpBp+m}, \mathbf{v}) = \mathfrak{I}_{\bar{\tau},u^h}^{EQ,p}(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{B}_{\bar{\tau},\partial\tau_D}^{p+m} \quad (53)$$

The corresponding computed pollution indicators are

$$\hat{\mu}_{\bar{\tau},\sigma_{ij}}^{EQ,p}(\bar{\tau}; ERpBp+m) \stackrel{\text{def}}{=} \mathfrak{I}_{\bar{\tau},u^h}^{EQ,p}(\hat{e}_{\bar{\tau},\mathbf{G}_{\bar{\tau},h}^{\sigma_{ij}}}^{ERpBp+m}) = \mathfrak{B}_{\bar{\tau}}(\hat{e}_{\bar{\tau},u^h}^{ERpBp+m}, \hat{e}_{\bar{\tau},\mathbf{G}_{\bar{\tau},h}^{\sigma_{ij}}}^{ERpBp+m}) \quad (54)$$

and

$$\mathcal{E}_{\sigma_{ij}^{avg}(\bar{\tau})}^{ERpBp+m} \stackrel{\text{def}}{=} \sum_{\substack{\tau \in T_h \\ \tau \subseteq \Omega - \omega_h^{\frac{1}{2}}}} \hat{\mu}_{\tau, \sigma_{ij}}^{EQ,p}(\bar{\tau}; ERpBp+m) \quad (55)$$

The corresponding estimate for the energy norm of the local error is

$$\mathcal{E}_{\tau, loc}^{ERpBp+m} = \eta_{\tau, u^h}^{ERpBp+m} \stackrel{\text{def}}{=} \left\| \hat{\mathbf{e}}_{\tau, u^h}^{ERpBp+m} \right\|_{\tau} \quad (56)$$

Let us now consider the case that the finite element code does not have the capability of computing the above splittings of the residuals and the residual error indicators, but nevertheless, as it is now the case with many commercial codes, it can compute indicators of the error based on a local averaging of the displacement, the strain, or the stress. We call such an averaging a *recovery (REC)*, e.g., the *ZZ-SPR* recovery proposed by Zienkiewicz and Zhu [15, 16, 17], the *WA* or the *WAZ* recoveries proposed by Wiberg and coworkers [18], etc. In this case we can always compute element error indicators for the stress in the form

$$\hat{\mathbf{e}}_{\sigma(u^h)}^{REC} \stackrel{\text{def}}{=} \boldsymbol{\sigma}^{REC}(u^h) - \boldsymbol{\sigma}(u^h) \quad (57)$$

where *REC* \equiv *ZZ-SPR*, or *WA*, or *WAZ*, etc., we can approximate the pollution indicators by

$$\hat{\mu}_{\tau, \sigma_{ij}}(\bar{\tau}; REC) \stackrel{\text{def}}{=} \int_{\tau} (\hat{\mathbf{e}}_{\sigma(u^h)}^{REC})^T \mathbf{E}^{-1} \hat{\mathbf{e}}_{\sigma(\hat{\mathbf{G}}_{\tau, h}^{REC})}^{REC} \quad (58)$$

and we obtain the estimate of the pollution error in the derivatives of the error

$$\mathcal{E}_{\sigma_{ij}^{avg}(\bar{\tau})}^{REC} \stackrel{\text{def}}{=} \sum_{\substack{\tau \in T_h \\ \tau \subseteq \Omega - \omega_h^{\frac{1}{2}}}} \hat{\mu}_{\tau, \sigma_{ij}}(\bar{\tau}; REC) \quad (59)$$

and the corresponding estimate for the energy-norm of the local error

$$\mathcal{E}_{\tau, loc}^{REC} \stackrel{\text{def}}{=} \sqrt{\int_{\tau} (\hat{\mathbf{e}}_{\sigma(u^h)}^{REC})^T \mathbf{E}^{-1} \hat{\mathbf{e}}_{\sigma(u^h)}^{REC}} \quad (60)$$

Similarly, we can obtain the estimate of the energy norm of the pollution error $\mathcal{E}_{\tau, glob}^{REC}$ by employing the pollution indicators given in (58) into (42).

Remark 13. Let us note that *the element residual problems can also be understood as local recoveries for the stress*. For example, let

$$\boldsymbol{\sigma}^{ERpBp+m}(u^h) \stackrel{\text{def}}{=} \boldsymbol{\sigma}(u^h) + \boldsymbol{\sigma}(\hat{\mathbf{e}}_{\tau, u^h}^{ERpBp+m}) \quad (61)$$

and for the error indicator of the stress

$$\hat{e}_{\sigma(\mathbf{u}^h)}^{ERpBp+m} \stackrel{\text{def}}{=} \sigma(\hat{e}_{\bar{\tau}, \mathbf{u}^h}^{ERpBp+m}) \tag{62}$$

Employing $\hat{e}_{\sigma(\mathbf{u}^h)}^{ERpBp+m}$ into (57)–(60) we recover the estimates given in (52)–(56).

Remark 14. The representation of the error in the strains or stresses in terms of the energy-inner product of the error in the finite element solution with the error in the finite element approximation of the auxiliary function was first employed by Babuška and Miller [19, 20]. The same approach was recently employed in [21, 22] for the special case of bilinear elements ($p = 1$), and was based on *explicit* error indicators which are computed directly from the residuals. Let us also note that [19]–[22] do not employ a splitting of the error, and employ the same representation as (59), with the difference that the sum is taken over all the elements. In [23] we have shown that the error indicators for the Green’s function in the elements in $\omega_{\bar{\tau}}^h$ can be very inaccurate and they can degrade the accuracy of the estimates for the average error in the stress in the element $\bar{\tau}$, if they are included in the sum in (59).

Remark 15. The symbol \mathcal{E} is used throughout the paper to denote an estimate of the error for the quantity which appears as a subscript, while the superscript indicates the averaging employed in the computation of this estimate. For example, $\mathcal{E}_{\sigma(\mathbf{x})}^{ZZ-SPR}$ denotes an estimate for $\mathbf{e}_{\sigma(\mathbf{x})} \stackrel{\text{def}}{=} \sigma(\mathbf{e}^h)(\mathbf{x})$, which is computed using $\sigma^{ZZ-SPR}(\mathbf{u}^h)$.

Let us now illustrate the accuracy of the above estimates in the example problem depicted in Fig. 1 (see also [24] and [25] for details of the implementation of the above estimates).

Example 1. *The accuracy of the ERpBp + m and ZZ-SPR estimates for the local, the pollution, and the total errors.* We considered the domain, the loading, and the mesh of mapped biquadratic squares ($p = 2$) shown in Fig. 3. Fig. 3a shows the distribution of the von Mises stress

$$\sigma^{VM}(\mathbf{u}^h) \stackrel{\text{def}}{=} \frac{1}{\sqrt{3}} J_2(\sigma(\mathbf{u}^h)) \stackrel{\text{def}}{=} \frac{1}{\sqrt{3}} \sqrt{\sum_{i,j=1}^2 \sigma_{ij}^{DEV}(\mathbf{u}^h) \sigma_{ij}^{DEV}(\mathbf{u}^h)} \tag{63}$$

where $\sigma_{ij}^{DEV} \stackrel{\text{def}}{=} \sigma_{ij} - \frac{1}{3} \left(\sum_{k=1}^2 \sigma_{kk} \right) \delta_{ij}$, is the deviatoric stress, relative to the *reference* value

$$\sigma_{ref}(\mathbf{u}^h) \stackrel{\text{def}}{=} \max_{\mathbf{x} \in \Omega} (\sigma^{VM}(\mathbf{u}^h)(\bar{\mathbf{x}})) \tag{64}$$

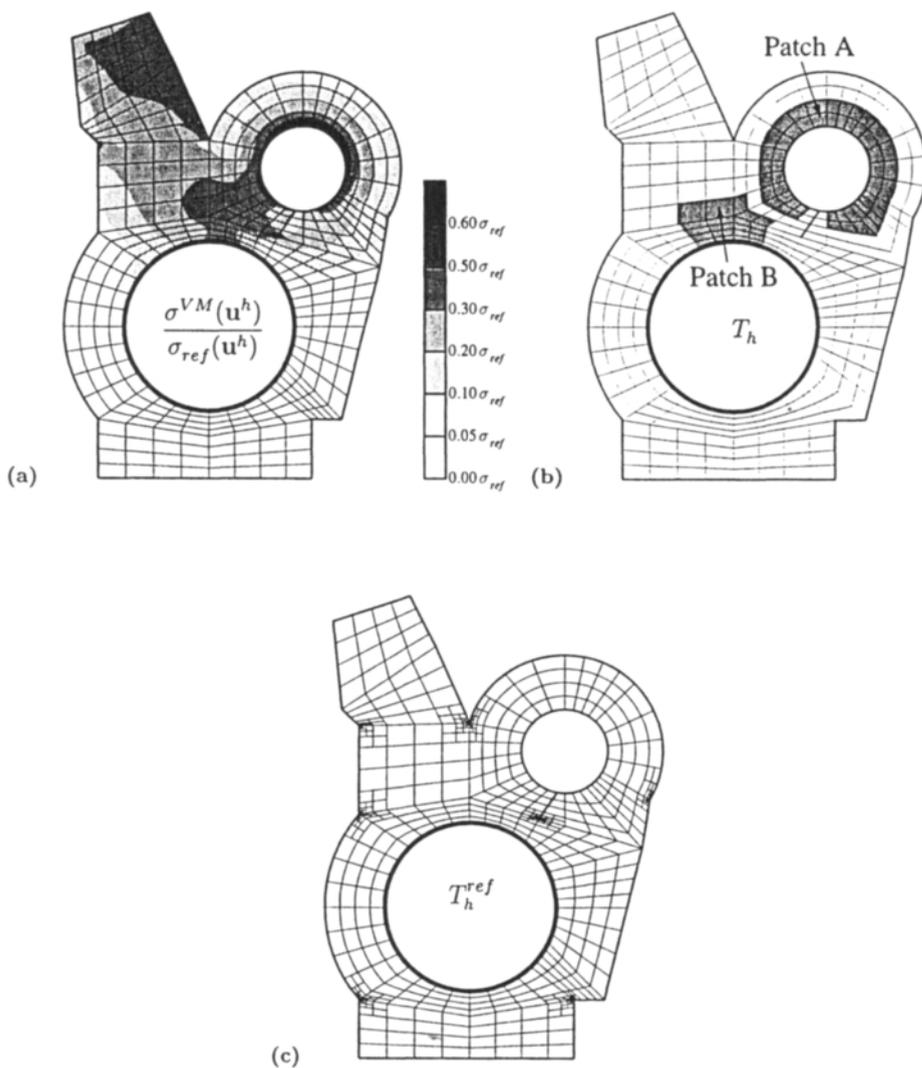


Fig. 3. An example which illustrates the a-posteriori estimation of the local error and the pollution error: (a) The mesh and the distribution of the von Mises stress, $\sigma^{VM}(\mathbf{u}^h)$, relative to the computed reference value $\sigma_{ref}(\mathbf{u}^h) = \max_{\bar{\mathbf{x}} \in \bar{\Omega}} (\sigma^{VM}(\mathbf{u}^h)(\bar{\mathbf{x}}))$. (b) The mesh T_h with the mesh patches A and B (shown shaded), in which the von Mises stress attains large finite values, and the estimates of its error will be computed and analyzed. (c) The mesh T_h^{ref} employed in the overkill approximation of the local and the total errors. This mesh was obtained from T_h by recursively subdividing four times each element with a vertex at a reentrant corner.

Note that here we are computing the maximum for a *fixed* mesh T_h and the computed maximum has finite value. This reference value does not have meaning as h tends to zero; nevertheless, here we are simply interested in scaling the von Mises stress, and we will analyze their error in areas away from the corner points where the exact values exist. From Fig. 3a we can see that the von Mises stress attains large finite values within the mesh-patches A and B shown in Fig. 3b. We will give the error in the von Mises stress and its estimates for the elements in these patches. In Fig. 4a we give the results for the exact relative error in the von Mises stress,

$$e_{\sigma^{VM}(\mathbf{u}^h),REL} \stackrel{\text{def}}{=} |\sigma^{VM}(\mathbf{u}^h + \mathbf{e}_{ovk}^h) - \sigma^{VM}(\mathbf{u}^h)| / \sigma^{VM}(\mathbf{u}^h) \quad (65)$$

where \mathbf{e}_{ovk}^h is the overkill approximation of the error computed by employing elements of degree $p + m = 4$, and the refined mesh T_h^{ref} shown in Fig. 3c (this mesh was obtained from the mesh T_h , shown in Fig. 1, by recursively subdividing four times every element with a vertex at a reentrant corner), in the elements in patch A , and in Fig. 4b we give the estimated relative error in the von Mises stress,

$$\mathcal{E}_{\sigma^{VM}(\mathbf{u}^h),REL}^{ER2B4} \stackrel{\text{def}}{=} \left| \frac{1}{\sqrt{3}} J_2(\sigma^{ER2B4}(\mathbf{u}^h) + \mathcal{E}_{\sigma^{av9}(\tau)}^{ER2B4}) - \sigma^{VM}(\mathbf{u}^h) \right| / \sigma^{VM}(\mathbf{u}^h) \quad (66)$$

By comparing Fig. 4a and Fig. 4b we note that $\mathcal{E}_{\sigma^{VM}(\mathbf{u}^h),REL}^{ER2B4}$ is a reliable estimate of the exact relative error in the von Mises stress of the finite element solution in all the elements in the mesh patch. In Fig. 4c we show the regions of the relative value of the von Mises stress of the exact local error

$$\sigma_{\tau,REL}^{VM}(\mathbf{e}_{\tau}^{h,loc}) \stackrel{\text{def}}{=} \sigma^{VM}(\mathbf{e}_{\tau}^{h,loc}) / \sigma^{VM}(\mathbf{u}^h) \quad (67)$$

which compare well with the corresponding regions of the von Mises stress of the estimated local error (Fig. 4d)

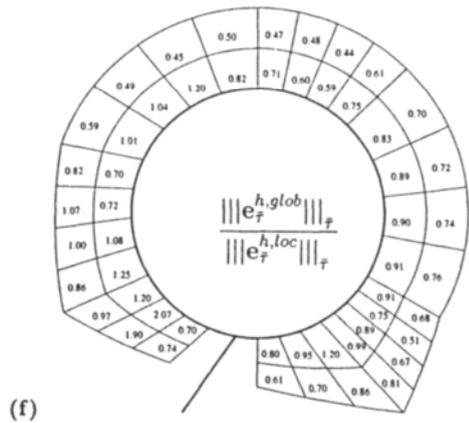
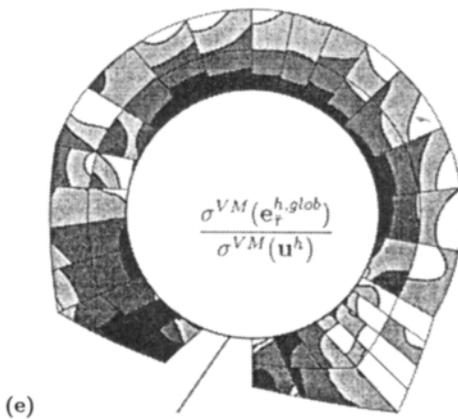
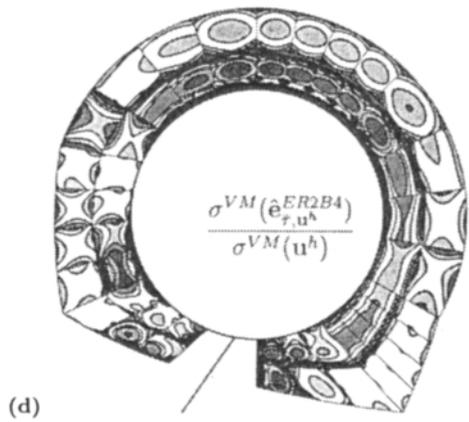
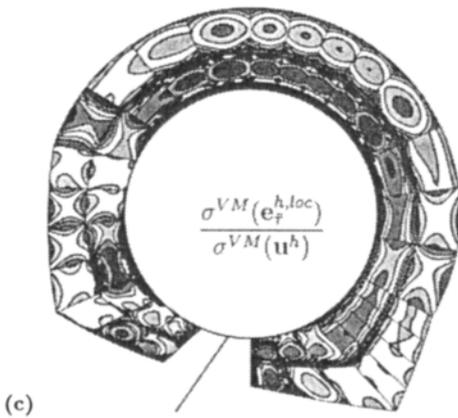
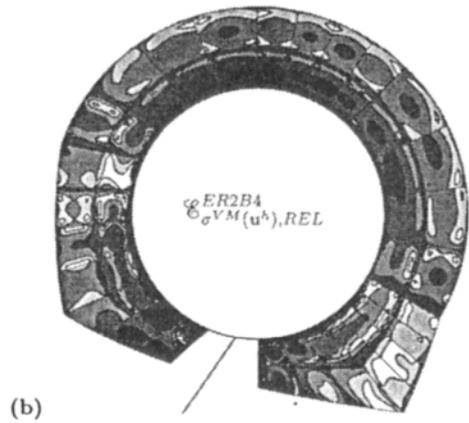
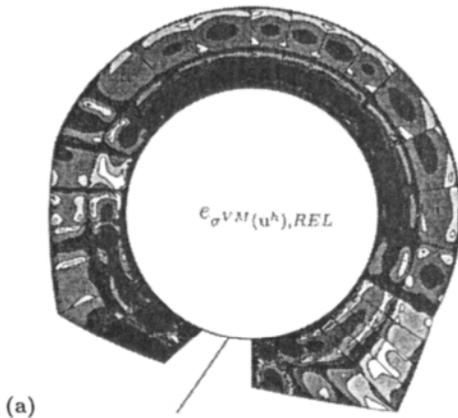
$$\sigma_{\tau,REL}^{VM}(\hat{\mathbf{e}}_{\tau,\mathbf{u}^h}^{ER2B4}) = \sigma^{VM}(\hat{\mathbf{e}}_{\tau,\mathbf{u}^h}^{ER2B4}) / \sigma^{VM}(\mathbf{u}^h) \quad (68)$$

In Fig. 4e we show the regions of the von Mises stress of the exact pollution error

$$\sigma_{\tau,REL}^{VM}(\mathbf{e}_{\tau}^{h,glob}) = \sigma^{VM}(\mathbf{e}_{\tau}^{h,glob}) / \sigma^{VM}(\mathbf{u}^h) \quad (69)$$

It can be seen that the pollution error in the stresses in an element is practically constant when compared with the oscillatory behavior of the local error over the element. In Fig. 4f we give the values of the pollution-ratios

$$pol_{\tau}^{\%} \stackrel{\text{def}}{=} \|\|\mathbf{e}_{\tau}^{h,glob}\|\|_{\tau} / \|\|\mathbf{e}_{\tau}^{h,loc}\|\|_{\tau} \quad (70)$$



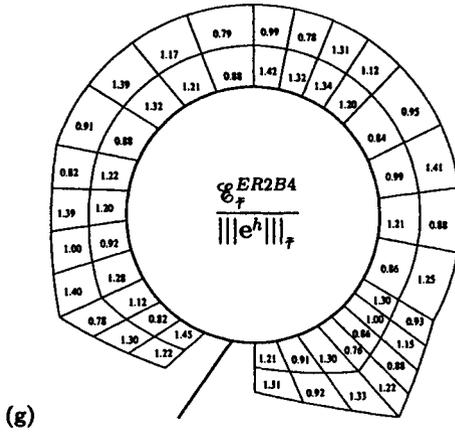


Fig. 4. An example which illustrates the a-posteriori estimation of the local error and the pollution error. The estimates based on the *ER2B4* indicators: Regions of 0–1%, 1–2.5%, 2.5–5%, and 5–10% (shown with increasing intensity of gray shading) for the patch *A* shown in Fig. 3b, for the: (a) Exact relative error in the von Mises stress $e_{\sigma^{VM}(\mathbf{u}^h),REL}$. (b) Estimated relative error in the von Mises stress $\mathcal{E}_{\sigma^{VM}(\mathbf{u}^h),REL}^{ER2B4}$. (c) Relative value of the von Mises stress of the local error $\sigma_{\tau,REL}^{VM}(e_{\tau}^{h,loc})$. (d) Relative value of the von Mises stress of the element error indicator function $\sigma_{\tau,REL}^{VM}(\hat{e}_{\tau,\mathbf{u}^h}^{ER2B4})$. (e) Relative value of the von Mises stress of the pollution error $\sigma_{\tau,REL}^{VM}(e_{\tau}^{h,glob})$. (f) Values of pollution ratios $pol_{\tau}^{\%}$ in the elements in the mesh patch. Note that the pollution error and the local error have comparable magnitudes in all the elements in patch *A*. (g) Values of the effectivity index κ_{τ}^{ER2B4} for the *ER2B4* estimate of the element energy norm of the error. It can be seen that the estimates for the local, the pollution, and the total errors, based on the *ER2B4* element error indicators, are rather accurate.

in the elements in patch *A*, from which we can see that the local error and the pollution error have comparable magnitudes in most of the elements. In Fig. 4g we give the element effectivity indices

$$\kappa_{\tau}^{ER2B4} \stackrel{\text{def}}{=} \mathcal{E}_{\tau}^{ER2B4} / |||e_{\sigma^k}^h|||_{\tau} \tag{71}$$

for the elements in patch *A*; we see that $0.8 \leq \kappa_{\tau}^{ER2B4} \leq 1.4$.

Let us also analyze the accuracy of the estimates based on the *ZZ-SPR* error indicators. In Fig. 5b we give the regions of the von Mises stress of the local error estimated by the *ZZ-SPR*

$$\sigma_{\tau,REL}^{VM,ZZ-SPR} \stackrel{\text{def}}{=} \frac{1}{\sqrt{3}} J_2(\hat{e}_{\sigma(\mathbf{u}^h)}^{ZZ-SPR}) / \sigma^{VM}(\mathbf{u}^h) \tag{72}$$

these regions compare well with the corresponding exact regions $\sigma_{\tau,REL}^{VM}(e_{\tau}^{h,loc})$ shown in Fig. 5a (Fig. 5a is identical to Fig. 4c). In Fig. 5c and 5e we give the element effectivity indices for the *ZZ-SPR* estimates of the element energy norms of the local and the

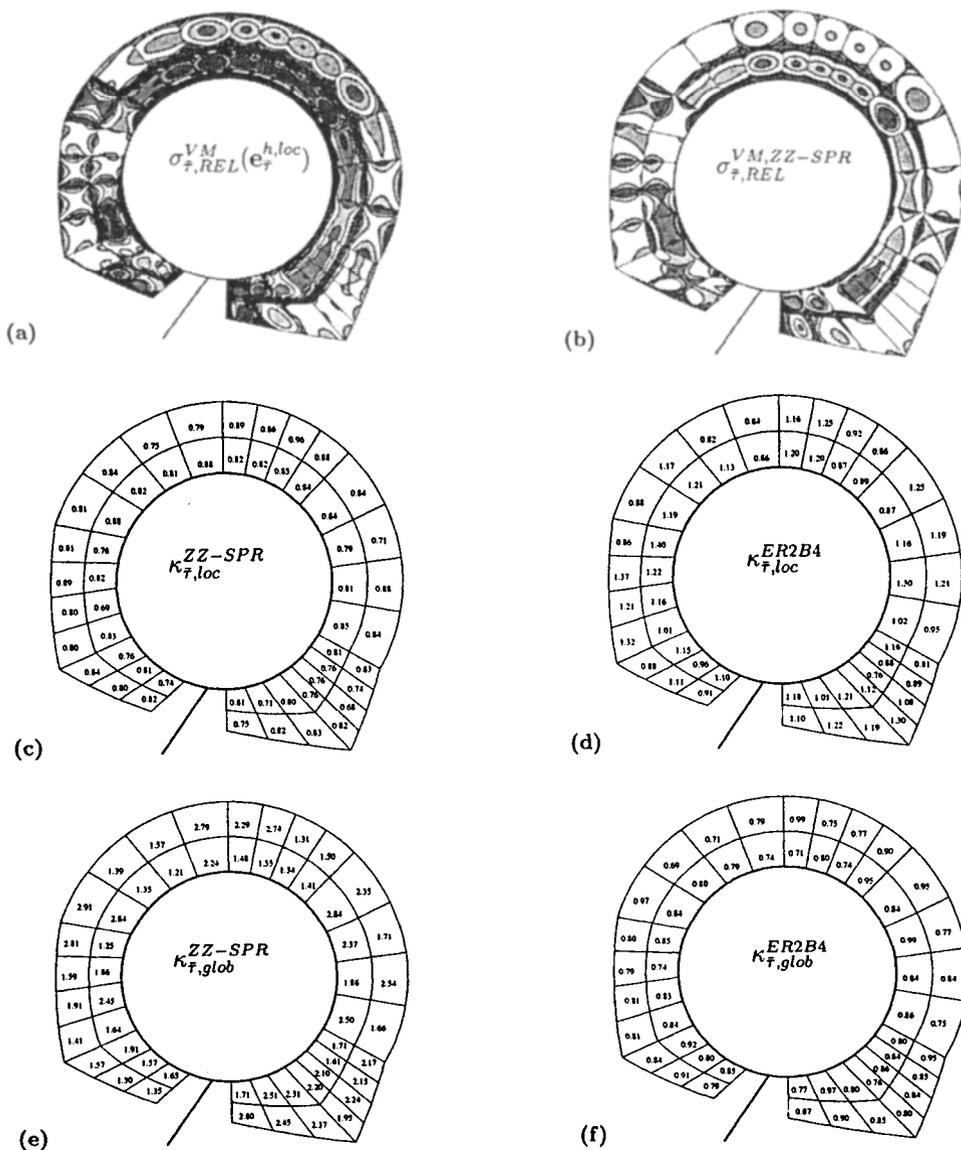


Fig. 5. An example which illustrates the *a-posteriori* estimation of the local error and the pollution error: The estimates based on the ZZ-SPR indicators, compared with the corresponding estimates based on the ER2B4 indicators for patch A. Regions of 0–1%, 1–2.5%, 2.5–5% and 5–10% (shown with increasing intensity of gray shading) for the relative value of the von Mises stress of: (a) $\sigma_{\tau,REL}^{VM}(e_{\tau}^{h,loc})$, (b) $\sigma_{\tau,REL}^{VM,ZZ-SPR}$. (c) Values of the element effectivity indices, $\kappa_{\tau,loc}^{ZZ-SPR}$. (d) Values of the element effectivity indices, $\kappa_{\tau,loc}^{ER2B4}$. (e) Values of the element effectivity indices, $\kappa_{\tau,glob}^{ZZ-SPR}$. (f) Values of the element effectivity indices, $\kappa_{\tau,glob}^{ER2B4}$. Note that the ER2B4 estimates of the pollution error are more effective than the corresponding ZZ-SPR estimates.

pollution error, namely

$$\kappa_{r,loc}^{ZZ-SPR} \stackrel{\text{def}}{=} \frac{c_{\tau,loc}^{ZZ-SPR}}{\|e_{\tau}^{h,loc}\|_{\tau}} \quad (73)$$

and

$$\kappa_{r,glob}^{ZZ-SPR} \stackrel{\text{def}}{=} \frac{c_{\tau,glob}^{ZZ-SPR}}{\|e_{\tau}^{h,glob}\|_{\tau}} \quad (74)$$

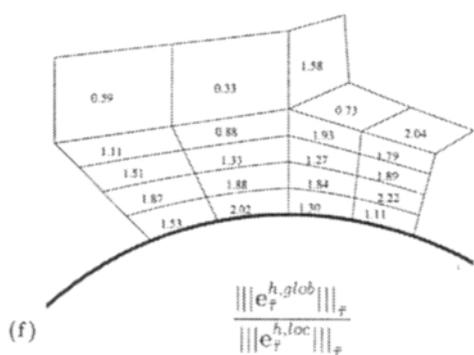
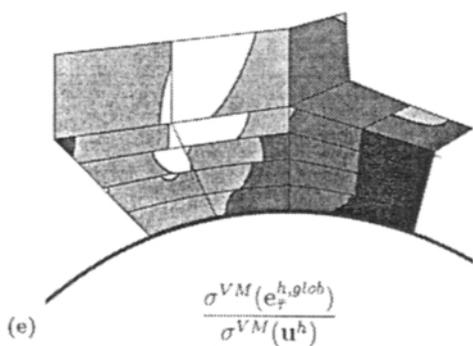
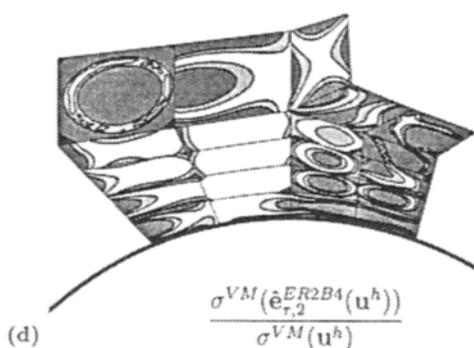
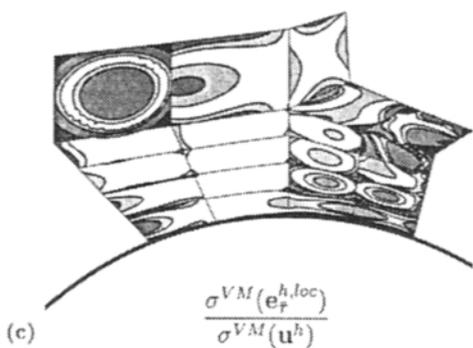
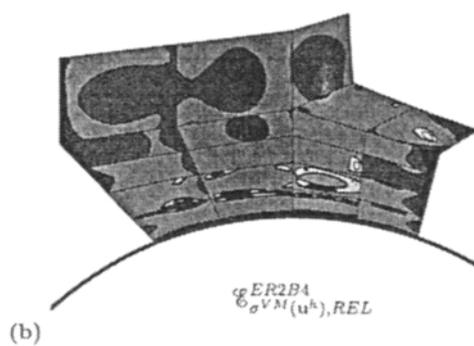
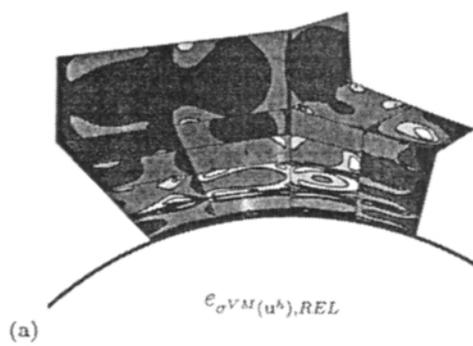
and in Fig. 5d and 5f the corresponding element effectivity indices, $\kappa_{r,loc}^{ER2B4}$ and $\kappa_{r,glob}^{ER2B4}$, for the estimates based on the *ER2B4* error indicators. It can be seen that the estimates for the pollution error based on the *ER2B4* error indicators are more accurate than the pollution estimates based on the *ZZ-SPR*. *This reflects the inferior accuracy of the ZZ-SPR pollution indicators in the elements with a vertex at a corner point.*

Let us also give the corresponding results for the elements in patch *B*. Fig. 6 gives the results for the estimates based on the *ER2B4* indicators in patch *B*, for the same quantities as the ones depicted in Fig. 4 for patch *A*. Fig. 7 compares the *ZZ-SPR* and the *ER2B4* estimates of the local and the pollution errors exactly as in Fig. 5 for patch *A*. Once more, we see that the pollution estimates based on the *ZZ-SPR* are less accurate than the corresponding estimates based on the *ER2B4*, due to the inferior quality of the *ZZ-SPR* indicators in the elements with a vertex at a corner point.

From the above example we clearly see that:

- (a) The estimates based on the *ER2B4* and the *ZZ-SPR* are sufficiently accurate for estimating the error in meshes of biquadratic elements.
- (b) The relative magnitude of the pollution error with respect to the local error in an element of interest cannot be predicted a-priori, and it is essential to have reliable estimates for both components of the error.
- (c) The accuracy of the pollution estimates depends on the accuracy of the error indicators over the entire mesh, and especially the elements with a vertex at a corner point.

Above we considered the case of bi-quadratic elements ($p = 2$) because they are used most often in practical computations. Similar results can also be obtained for linear ($p = 1$) and cubic ($p = 3$) elements. For example, in Fig. 8 we give the shades of the exact and the estimated relative error in the von Mises stress for bilinear ($p = 1$) and



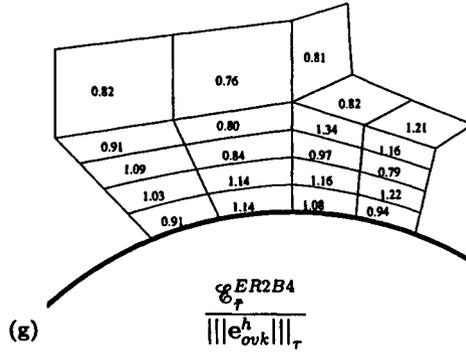


Fig. 6. An example which illustrates the a-posteriori estimation of the local error and the pollution error. The estimates based on the $ER2B4$ indicators: Regions of 0–1%, 1–2.5%, 2.5–5%, and 5–10% (shown with increasing intensity of gray shading) for the patch B shown in Fig. 3b, for the: (a) Exact relative error in the von Mises stress $e_{\sigma^{VM}(\mathbf{u}^h),REL}$. (b) Estimated relative error in the von Mises stress $\mathcal{E}_{\sigma^{VM}(\mathbf{u}^h),REL}^{ER2B4}$. (c) Relative value of the von Mises stress of the local error $\sigma^{VM}(e_T^{h,loc})/\sigma^{VM}(\mathbf{u}^h)$. (d) Relative value of the von Mises stress of the element error indicator function $\sigma^{VM}(\hat{e}_{T,2}^{ER2B4}(\mathbf{u}^h))/\sigma^{VM}(\mathbf{u}^h)$. (e) Relative value of the von Mises stress of the pollution error $\sigma^{VM}(e_T^{h,glob})/\sigma^{VM}(\mathbf{u}^h)$. (f) Values of pollution ratios $pol_T^{\%} = \|e_T^{h,glob}\|_T / \|e_T^{h,loc}\|_T$ in the elements in the mesh patch. (g) Values of the effectivity index $\kappa_T^{ER2B4} \stackrel{\text{def}}{=} \mathcal{E}_T^{ER2B4} / \|e_{ovk}^h\|_T$ for the $ER2B4$ estimate of the element energy norm of the error. Note that the conclusions are the same as the ones obtained in Fig. 4, namely, that the estimates based on the $ER2B4$ indicators are rather accurate.

bicubic ($p = 3$) elements, and the element effectivity indices for the estimate of the element energy-norm based on the $ERpB2p$ error indicators for the elements in patch A .

Let us now compute the error in the error estimates. We will denote the error in the error indicator by

$$e_{\sigma^{REC}(\mathbf{u}^h)} \stackrel{\text{def}}{=} \boldsymbol{\sigma}(e_T^{h,loc}) - \hat{e}_{\boldsymbol{\sigma}(\mathbf{u}^h)}^{REC} \quad (75)$$

and the relative error in the error indicator by

$$E_{REL}(\hat{e}_{\boldsymbol{\sigma}(\mathbf{u}^h)}^{REC}) \stackrel{\text{def}}{=} \|e_{\sigma^{REC}(\mathbf{u}^h)}\|_{T,E^{-1}} / \|e_T^{h,loc}\|_T \quad (76)$$

where $\|\cdot\|_{T,E^{-1}}$ is the *stress-energy* norm, namely

$$\|\boldsymbol{\sigma}\|_{T,E^{-1}} \stackrel{\text{def}}{=} \sqrt{\int_T \boldsymbol{\sigma}^T \mathbf{E}^{-1} \boldsymbol{\sigma}} \quad (77)$$

We will also denote the relative error in the estimate of the element average of the

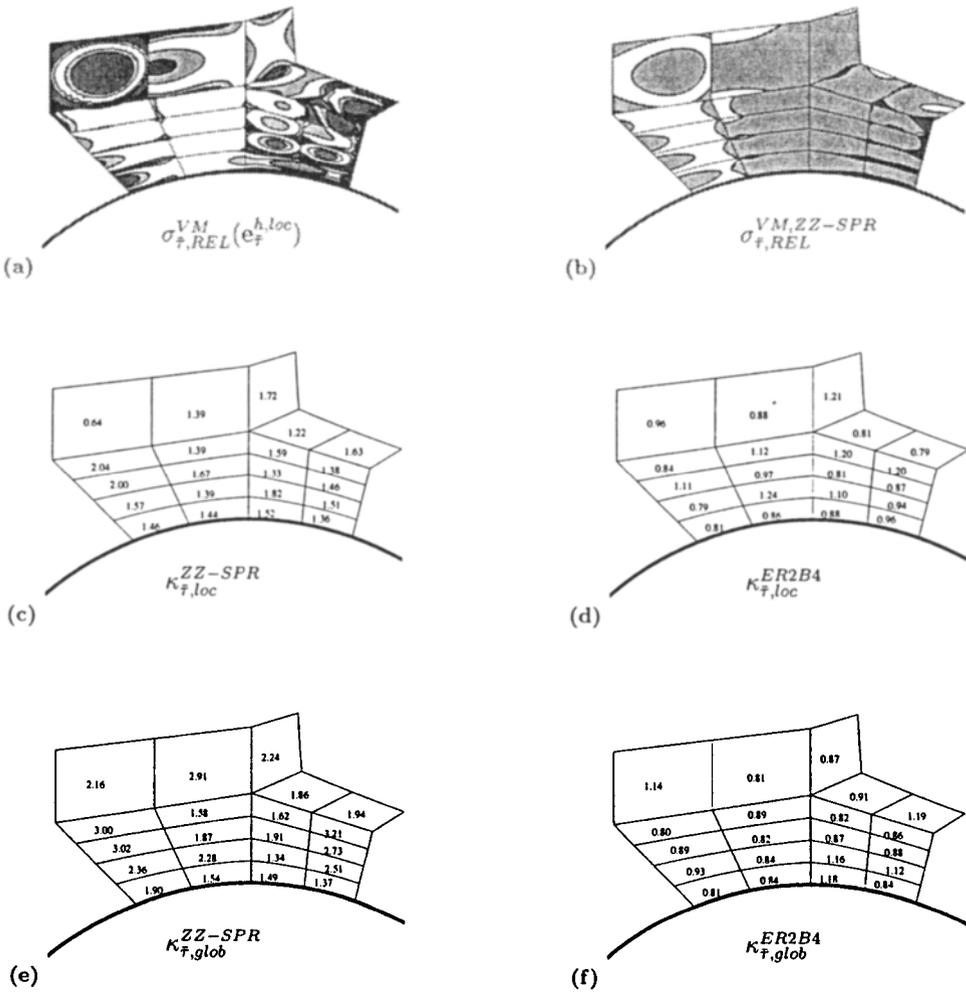
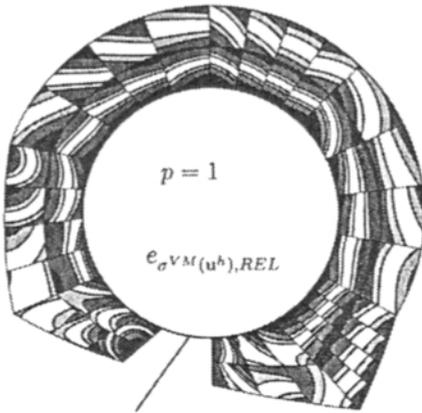
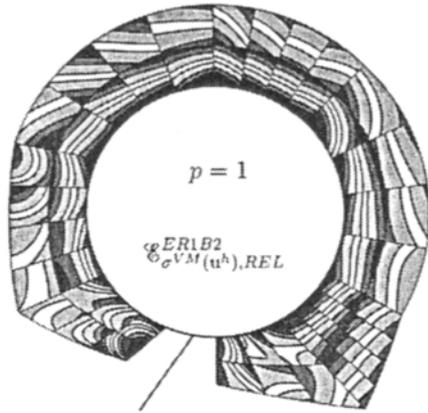


Fig. 7. An example which illustrates the a-posteriori estimation of the local error and the pollution error: The estimates based on the ZZ-SPR indicators, compared with the corresponding estimates based on the ER2B4 indicators for patch B. The regions of 0-1%, 1-2.5%, 2.5-5% and 5-10% (shown with increasing intensity of gray shading) for: (a) $\sigma_{\tau,REL}^{VM}(e_{\tau}^{h,loc})$, (b) $\sigma_{\tau,REL}^{VM,ZZ-SPR}$. (c) Values of the element effectivity indices, $\kappa_{\tau,loc}^{ZZ-SPR}$. (d) Values of the element effectivity indices, $\kappa_{\tau,loc}^{ER2B4}$. (e) Values of the element effectivity indices, $\kappa_{\tau,glob}^{ZZ-SPR}$. (f) Values of the element effectivity indices, $\kappa_{\tau,glob}^{ER2B4}$. Note that the ER2B4 estimates of the pollution error are more effective than the corresponding ZZ-SPR estimates.



(a)



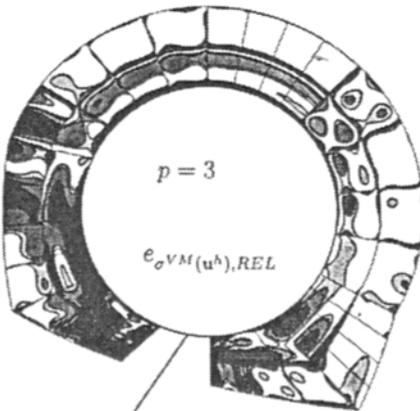
(b)



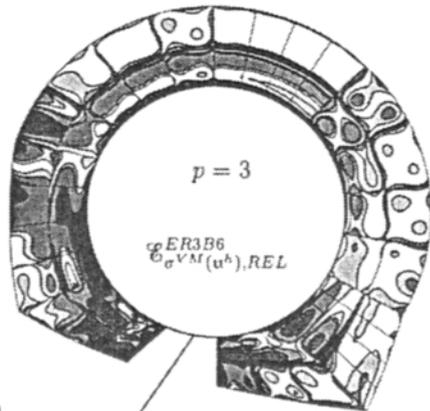
(c)



(d)



(e)



(f)

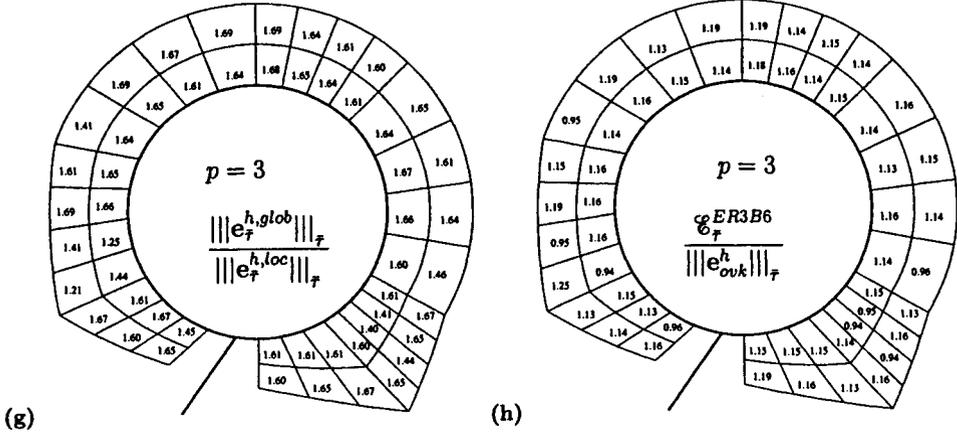


Fig. 8. An example which illustrates the a-posteriori estimation of the total error: The estimates based on the $ERpB2p$ indicators: The results given in (a)–(d) are for bilinear elements ($p = 1$). (a) Regions of the exact relative error in the von Mises stress $e_{\sigma_{VM}(u^h), REL}$ (0–5% blank, 5–10% light gray, 10–20% dark gray, 20–30% black). (b) Regions of the estimated relative error in the Mises stress $\mathcal{E}_{\sigma_{VM}(u^h), REL}^{ER2B4}$. (c) Values of the pollution ratios $pol_{\tau}^{\%} \stackrel{\text{def}}{=} \frac{|||e_{\tau}^{h, glob}|||_{\tau}}{|||e_{\tau}^{h, loc}|||_{\tau}}$. (d) The element effectivity indices $\kappa_{\tau}^{ER1B2} \stackrel{\text{def}}{=} \frac{\mathcal{E}_{\tau}^{ER1B2}}{|||e^h|||_{\tau}}$. The following results are for bicubic ($p = 3$) elements. (e) Regions of the exact relative error in the von Mises stress $e_{\sigma_{VM}(u^h), REL}$ (0–0.1% blank, 0.1–0.2% light gray, 0.2–0.5% dark gray, 0.5–1.0% black). (f) Regions of the estimated relative error in the Mises stress $\mathcal{E}_{\sigma_{VM}(u^h), REL}^{ER2B4}$. (g) Values of the pollution ratios $pol_{\tau}^{\%} \stackrel{\text{def}}{=} \frac{|||e_{\tau}^{h, glob}|||_{\tau}}{|||e_{\tau}^{h, loc}|||_{\tau}}$. (h) The element effectivity indices $\kappa_{\tau}^{ER3B6} \stackrel{\text{def}}{=} \frac{\mathcal{E}_{\tau}^{ER3B6}}{|||e_{\tau}^{h, ovk}|||_{\tau}}$.

derivative of the pollution error by

$$E_{REL}(\mathcal{E}_{\sigma_{ij}^{avg(\tau)}}^{REC}) \stackrel{\text{def}}{=} J_2 \left(\mathcal{E}_{\sigma_{ij}^{avg(\tau)}}^{REC} - \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \sigma_{ij}(e_{\tau}^{h, glob}) \right) / J_2 \left(\frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \sigma_{ij}(e_{\tau}^{h, glob}) \right) \quad (78)$$

and the relative error in the error estimate for the element energy norm of the error by

$$E_{REL}(\mathcal{E}_{\tau}^{REC}) \stackrel{\text{def}}{=} \left| \mathcal{E}_{\tau}^{REC} - |||e_{\tau}^h|||_{\tau} \right| / |||e^h|||_{\tau} \quad (79)$$

for $REC \equiv ER2B4$ and $REC \equiv ZZ-SPR$. In Fig. 9 and Fig. 10 we give the values of $E_{REL}(\hat{e}_{\sigma(u^h)}^{REC})$, $E_{REL}(\mathcal{E}_{\sigma_{ij}^{avg(\tau)}}^{REC})$ and $E_{REL}(\mathcal{E}_{\tau}^{REC})$ in the elements in Patch A and Patch B, respectively. We see that the maximum relative error in the $ER2B4$ error estimate of the element energy norm of the error in the elements in patches A and B is less than 30%, while it is more than 100% for the $ZZ-SPR$ estimate. From these results, we can conclude that *the error in the error estimate can be very significant and we must have the ability to estimate it.*

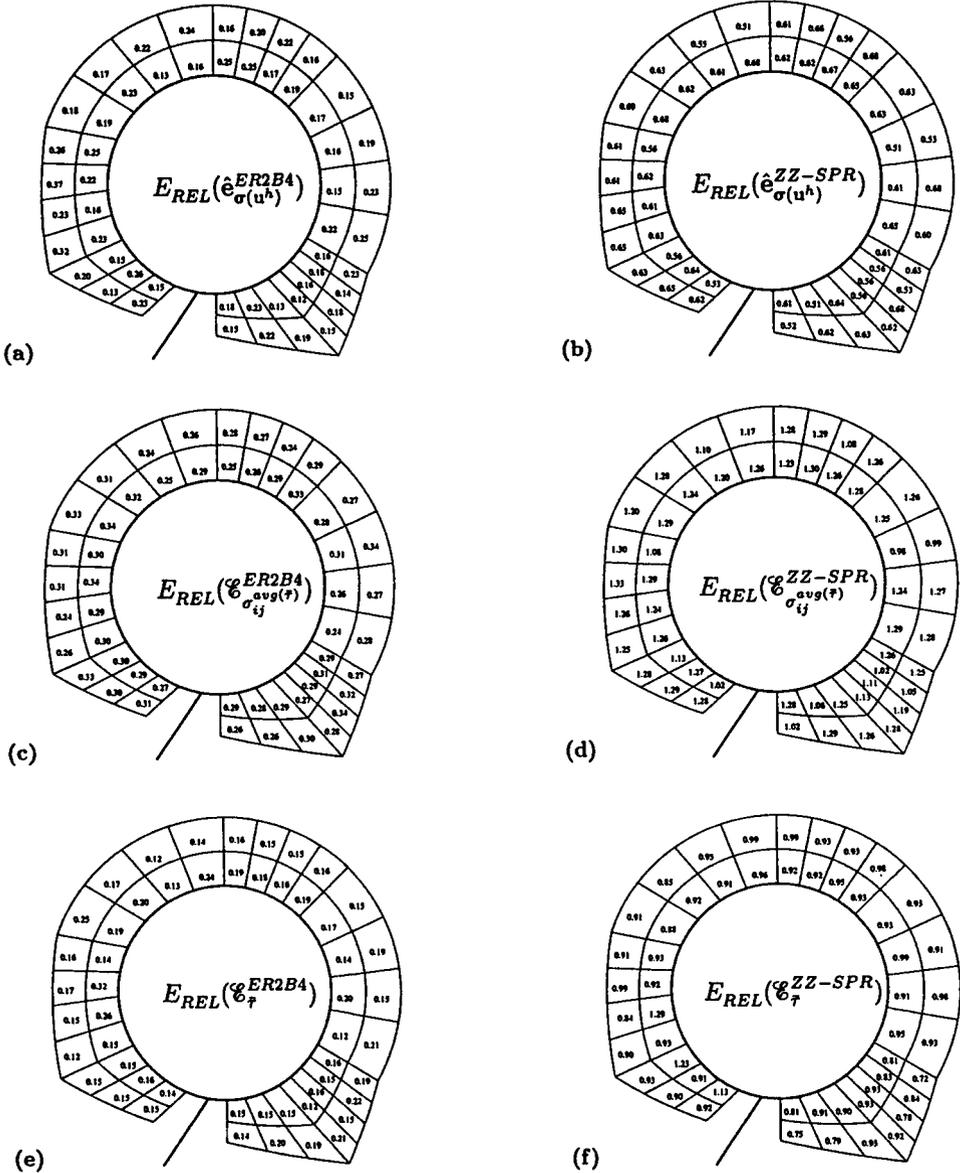


Fig. 9. An example which illustrates the a-posteriori estimation of the local error and the pollution error. The relative errors in the error estimates for the elements in patch A: The values of: (a) $E_{REL}(\hat{e}_{\sigma(u^h)}^{ER2B4})$, (b) $E_{REL}(\hat{e}_{\sigma(u^h)}^{ZZ-SPR})$, (c) $E_{REL}(\mathcal{E}_{\sigma^{avg}(\tau)}^{ER2B4})$, (d) $E_{REL}(\mathcal{E}_{\sigma^{avg}(\tau)}^{ZZ-SPR})$, (e) $E_{REL}(\mathcal{E}_{\tau}^{ER2B4})$, (f) $E_{REL}(\mathcal{E}_{\tau}^{ZZ-SPR})$. Note that the error in the ZZ-SPR estimates can be more than 100%, while the error in the ER2B4 estimates does not exceed 30%.

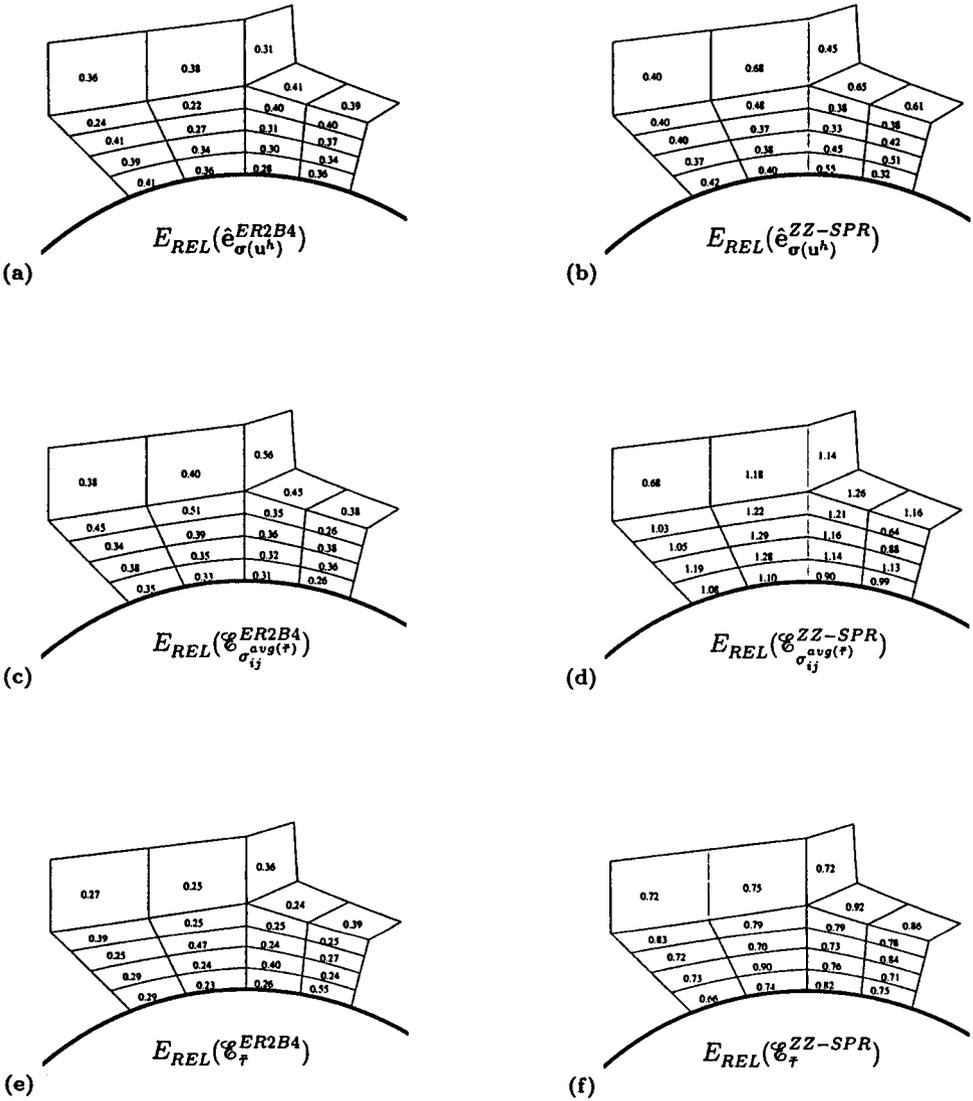


Fig. 10. An example which illustrates the a-posteriori estimation of the local error and the pollution error. The relative errors in the error estimates for the elements in patch B: The values of: (a) $E_{REL}(\hat{e}_{\sigma(u^h)}^{ER2B4})$, (b) $E_{REL}(\hat{e}_{\sigma(u^h)}^{ZZ-SPR})$, (c) $E_{REL}(c_{\sigma_{ij}^{avg}(\tau)}^{ER2B4})$, (d) $E_{REL}(c_{\sigma_{ij}^{avg}(\tau)}^{ZZ-SPR})$, (e) $E_{REL}(c_{\tau}^{ER2B4})$, (f) $E_{REL}(c_{\tau}^{ZZ-SPR})$.

2. A-POSTERIORI ESTIMATION OF THE ERROR IN THE ERROR ESTIMATE

In this section we construct *a-posteriori* estimates of the error in the error estimates by employing indicators of the local error in the recovered stress. The error in the error estimate $\mathcal{E}_{\sigma(\bar{x})}^{REC}$ can be split into two components, namely

$$\sigma(e^h)(\bar{x}) - \mathcal{E}_{\sigma(\bar{x})}^{REC} = \underbrace{\left(\sigma(e_{\bar{\tau}}^{h,loc})(\bar{x}) - \hat{e}_{\sigma(u^h)}^{REC}(\bar{x}) \right)}_{\text{error in the error indicator}} + \underbrace{\left(\sigma(e_{\bar{\tau}}^{h,glob})(\bar{x}) - \mathcal{E}_{\sigma^{avg}(\bar{\tau})}^{REC} \right)}_{\text{error in the pollution estimate}} \quad (80)$$

The error in the element error indicator will be estimated by

$$\sigma(e_{\bar{\tau}}^{h,loc}) - \hat{e}_{\sigma(u^h)}^{REC} \approx \hat{e}_{\sigma^{REC},REC^*}^{REC} \stackrel{\text{def}}{=} \sigma^{REC,REC^*}(u^h) - \sigma^{REC}(u^h) \quad (81)$$

where σ^{REC,REC^*} is a recovered stress which is constructed by locally averaging (recycling) $\sigma^{REC}(u^h)$ (see [26] for the asymptotic analysis of recycling). The error in the pollution estimate will be estimated by

$$\sigma(e_{\bar{\tau}}^{h,glob})(\bar{x}) - \mathcal{E}_{\sigma^{avg}(\bar{\tau})}^{REC} \approx \mathcal{E}_{\sigma^{avg}(\bar{\tau})}^{REC^*} \stackrel{\text{def}}{=} \sum_{\substack{\bar{\tau} \in T_h \\ \bar{\tau} \subseteq \Omega - \omega_{\bar{\tau}}^h}} \hat{e}_{\bar{\mu}_{\bar{\tau},\sigma}(\bar{\tau}; REC)}^{REC^*} \quad (82a)$$

where

$$\hat{e}_{\bar{\mu}_{\bar{\tau},\sigma_{ij}}(\bar{\tau}; REC)}^{REC^*} \stackrel{\text{def}}{=} \int_{\bar{\tau}} (\hat{e}_{\sigma(u^h)}^{REC})^T \mathbf{E}^{-1} \hat{e}_{\sigma^{REC}(\bar{\mathcal{G}}_{\bar{\tau},h}^{\sigma_{ij}})}^{REC^*} \quad (82b)$$

is the indicator of the error in the pollution indicator $\hat{\mu}_{\bar{\tau},\sigma_{ij}}(\bar{\tau}; REC)$. Let us now show how (82) was obtained. Noting that

$$\sigma(e_{\bar{\tau}}^{h,glob})(\bar{x}) \approx \frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \sigma(e_{\bar{\tau}}^{h,glob}) = \sum_{\substack{\bar{\tau} \in T_h \\ \bar{\tau} \subseteq \Omega - \omega_{\bar{\tau}}^h}} \mu_{\bar{\tau},\sigma}^{EQ,p}(\bar{\tau}) \quad (83)$$

the error in the pollution estimate is practically given by

$$\left(\frac{1}{|\bar{\tau}|} \int_{\bar{\tau}} \sigma(e_{\bar{\tau}}^{h,glob}) \right) - \mathcal{E}_{\sigma^{avg}(\bar{\tau})}^{REC} = \sum_{\substack{\bar{\tau} \in T_h \\ \bar{\tau} \subseteq \Omega - \omega_{\bar{\tau}}^h}} \left(\mu_{\bar{\tau},\sigma}^{EQ,p}(\bar{\tau}) - \hat{\mu}_{\bar{\tau},\sigma}(\bar{\tau}; REC) \right) \quad (84)$$

Further, the error in the pollution indicators is

$$\mu_{\bar{\tau},\sigma_{ij}}^{EQ,p}(\bar{\tau}) - \hat{\mu}_{\bar{\tau},\sigma_{ij}}(\bar{\tau}; REC) = \int_{\bar{\tau}} (\hat{e}_{\sigma(u^h)}^{ERp})^T \mathbf{E}^{-1} \left(\sigma(\bar{\mathcal{G}}_{\bar{\tau}}^{\sigma_{ij}} - \mathcal{A}_{\omega_{\bar{\tau}}^h}^p \bar{\mathcal{G}}_{\bar{\tau}}^{\sigma_{ij}}) - \hat{e}_{\sigma(\bar{\mathcal{G}}_{\bar{\tau},h}^{\sigma_{ij}})}^{REC} \right) \quad (85)$$

and employing the approximation

$$\boldsymbol{\sigma}(\check{\mathbf{G}}_{\tau}^{\sigma ij} - \mathcal{A}_{\omega_{\tau}^h}^p \check{\mathbf{G}}_{\tau}^{\sigma ij}) - \hat{\mathbf{e}}_{\boldsymbol{\sigma}(\check{\mathbf{G}}_{\tau,h}^{\sigma ij})}^{REC} \approx \boldsymbol{\sigma}(\check{\mathbf{G}}_{\tau}^{\sigma ij}) - \boldsymbol{\sigma}^{REC}(\check{\mathbf{G}}_{\tau,h}^{\sigma ij}) \approx \hat{\mathbf{e}}_{\boldsymbol{\sigma}^{REC*}(\check{\mathbf{G}}_{\tau,h}^{\sigma ij})}^{REC*} \quad (86)$$

we obtain the estimate (82). We obtained the last approximate equality in (86) by employing $\boldsymbol{\sigma}^{REC,REC*}(\check{\mathbf{G}}_{\tau,h}^{\sigma ij})$ to approximate $\boldsymbol{\sigma}(\check{\mathbf{G}}_{\tau}^{\sigma ij})$.

Let us now give some constructions of the *recycled* recovered stress, $\boldsymbol{\sigma}^{REC,REC*}$:

(i) *Polynomial recycling*:

Let

$$\boldsymbol{\sigma}^{REC,POL}(\mathbf{u}^h)|_{\tau} = \boldsymbol{\sigma}(\mathbf{u}_{\tau,\mathbf{u}^h}^{REC,POL}) \quad (87)$$

where $\mathbf{u}_{\tau,\mathbf{u}^h}^{REC,POL} \in \mathfrak{P}^{p+m}(\omega_{\tau}^h)$ is obtained by solving the following local minimization problem:

$$\|\boldsymbol{\sigma}(\mathbf{u}_{\tau,\mathbf{u}^h}^{REC,POL}) - \boldsymbol{\sigma}^{REC}(\mathbf{u}^h)\|_{\omega_{\tau}^h, \mathbf{D}, \{\zeta_k\}_{k=1}^{n.sp}} = \min_{\mathbf{u} \in \mathfrak{P}^{p+m}(\omega_{\tau}^h)} \|\boldsymbol{\sigma}(\mathbf{u}) - \boldsymbol{\sigma}^{REC}(\mathbf{u}^h)\|_{\omega_{\tau}^h, \mathbf{D}, \{\zeta_k\}_{k=1}^{n.sp}} \quad (88)$$

where

$$\|\boldsymbol{\sigma}\|_{\omega_{\tau}^h, \mathbf{D}, \{\zeta_k\}_{k=1}^{n.sp}}^2 \stackrel{\text{def}}{=} \sum_{k=1}^{n.sp} (\boldsymbol{\sigma}(\zeta_k))^T \mathbf{D} \boldsymbol{\sigma}(\zeta_k) \quad (89)$$

and $\{\zeta_k\}_{k=1}^{n.sp}$ is the set of mapped 2×2 Gauss-Legendre points in the elements in ω_{τ}^h . In the results given below we employed $\mathbf{D} = \mathbf{I}$ (the identity matrix); another possibility is $\mathbf{D} = \mathbf{E}^{-1}$, etc.

Remark 16. The polynomial recycling is rather general and can be easily implemented in any existing code which has the capability to compute a recovered stress.

(ii) *“Harmonic” recycling*:

Let $\{\mathbf{Q}_k^{\tau}\}_{k=1}^{\infty}$ a local “harmonic” basis in the element ω_{τ}^h , which satisfies

$$\nabla \cdot (\boldsymbol{\sigma}(\mathbf{Q}_k^{\tau})) = 0, \quad \boldsymbol{\sigma}(\mathbf{Q}_k^{\tau}) \mathbf{n}|_{\partial\tau_N} = \mathbf{g}, \quad \mathbf{Q}_k^{\tau}|_{\partial\tau_D} = 0 \quad (90)$$

where $\partial\tau_D \stackrel{\text{def}}{=} \partial\tau \cap \Gamma_D$, and $\partial\tau_N \stackrel{\text{def}}{=} \partial\tau \cap \Gamma_N$. In the cases that any of these sets is empty the corresponding boundary condition need not be enforced. We then have

$$\mathbf{u}^{EX}|_{\omega_{\tau}^h} = \sum_{k=1}^{\infty} \alpha_k \mathbf{Q}_k^{\tau} \quad (91)$$

Note that (90) does not specify uniquely the basis functions. For elements in the interior of the mesh, or elements adjacent to a smooth boundary, we will employ the Muskhelishvili complex potentials [27]

$$Q_{\ell,1}^r + i Q_{\ell,2}^r = \kappa \varphi(z) - z \overline{\varphi'(z)} - \psi(z) \quad (92)$$

where

$$\left. \begin{aligned} \varphi(z) &= A z \log z + B \log z + \sum_{k=1}^{p+m} a_k z^k + a_{p+m+k} z^{-k} \\ \psi(z) &= -\kappa A \log z + \sum_{k=1}^{p+m} b_k z^k + b_{p+m+k} z^{-k} \end{aligned} \right\} \quad (93)$$

where $z = x_1 + i x_2$, $i = \sqrt{-1}$, $\kappa \stackrel{\text{def}}{=} 3 - 4\nu$ for plane strain or $\kappa \stackrel{\text{def}}{=} \frac{3 - \nu}{1 + \nu}$ for plane stress, and A , B , a_k , b_k are complex and chosen such that Q_{ℓ}^r has finite value in ω_{τ}^h . In the computations we also constructed the basis, such that boundary conditions can be enforced on a circular arc (here we assume that the boundary of the domain can be represented by circular arcs).

For elements in the interior of the mesh, or adjacent to a smooth boundary, we will determine the recycled stress $\boldsymbol{\sigma}^{REC,HAR}(\mathbf{u}^h)|_{\tau} = \boldsymbol{\sigma}(\mathbf{u}_{\tau, \mathbf{u}^h}^{REC,HAR})$, where the recovered displacement

$$\mathbf{u}_{\tau, \mathbf{u}^h}^{REC,HAR} \in \mathfrak{Q}^{(N)}(\omega_{\tau}^h) \stackrel{\text{def}}{=} \left\{ \mathbf{Q} \mid \mathbf{Q} = \sum_{i=1}^N \alpha_i \mathbf{Q}_i^r \right\} \quad (94)$$

is obtained by solving the local minimization problem

$$\| \boldsymbol{\sigma}(\mathbf{u}_{\tau, \mathbf{u}^h}^{REC,HAR}) - \boldsymbol{\sigma}^{REC}(\mathbf{u}^h) \|_{\omega_{\tau}^h, \mathbf{D}, \{\boldsymbol{\zeta}_k\}_{k=1}^{n_{sp}}} = \min_{\mathbf{u} \in \mathfrak{Q}^{(N)}(\omega_{\tau}^h)} \| \boldsymbol{\sigma}(\mathbf{u}) - \boldsymbol{\sigma}^{REC}(\mathbf{u}^h) \|_{\omega_{\tau}^h, \mathbf{D}, \{\boldsymbol{\zeta}_k\}_{k=1}^{n_{sp}}} \quad (95)$$

Here $\{\boldsymbol{\zeta}_k\}_{k=1}^{n_{sp}}$ is the set of mapped 2×2 Gauss-Legendre points in the elements in ω_{τ}^h . We employed $N = 4(p + m)$ for interior elements, $N = 2(p + m)$ for elements adjacent to a straight boundary, $N = 4(p + m) + 2$ for elements adjacent to a curved boundary, and we let $m = p$.

For elements with a vertex at a corner point we will also employ the vertex basis functions

$$\mathbf{Q}_{\ell}^{x^0}(\mathbf{x}) = (r_{\mathbf{x}^0})^{\lambda_{\ell}} \boldsymbol{\Phi}_{\ell}(\theta_{\mathbf{x}^0}) \quad (96)$$

where \mathbf{x}^0 is the position vector of the corner point, and $r_{\mathbf{x}^0}$, $\theta_{\mathbf{x}^0}$ is a polar coordinate system centered at \mathbf{x}^0 . Here λ_{ℓ} and $\boldsymbol{\Phi}_{\ell}$ are, respectively, the ℓ th eigenvalue and the

eigenfunction for the infinite wedge, corresponding to the corner point. In this case we will employ a recovered displacement in the form

$$\mathbf{u}_{\tau, \mathbf{u}^h}^{REC, HAR} \in \overline{\mathcal{D}}^{(N)}(\omega_\tau^h) \stackrel{\text{def}}{=} \left\{ \mathbf{Q} \mid \mathbf{Q} = \sum_{i=1}^N \alpha_i \mathbf{Q}_i^T + \sum_{i=1}^4 \alpha_{N+i} \mathbf{Q}_i^{\mathbf{x}^0} \right\} \quad (97)$$

Analogously as in (87), will let $\hat{\mathbf{e}}_{\sigma_{REC}(\mathbf{u}^h)}^{HAR}$ denote the indicator of the error in the recovered stress, namely

$$\hat{\mathbf{e}}_{\sigma_{REC}(\mathbf{u}^h)}^{HAR} \stackrel{\text{def}}{=} \boldsymbol{\sigma}^{REC, HAR}(\mathbf{u}^h) - \boldsymbol{\sigma}^{REC}(\mathbf{u}^h) \quad (98)$$

and will compute the error in the pollution indicators as in (87)–(89)

$$\hat{\mathbf{e}}_{\hat{\mu}_{\tau, \mathbf{u}_{k, \ell}}^{REC}}^{HAR} \stackrel{\text{def}}{=} \int_{\tau} (\hat{\mathbf{e}}_{\sigma(\mathbf{u}^h)}^{EST})^T \mathbf{E}^{-1} \hat{\mathbf{e}}_{\sigma_{REC}(\mathbf{G}_{\tau, \lambda}^{\mathbf{u}_{k, \ell}})}^{HAR} \quad (99)$$

and the corresponding estimate of the error in $\mathcal{E}_{\mathbf{u}_{k, \ell}}^{REC}$ is

$$\mathcal{E}_{\mathcal{E}_{\mathbf{u}_{k, \ell}}^{REC}}^{HAR} \stackrel{\text{def}}{=} \sum_{\substack{\tau \in T_h \\ \tau \subseteq \Omega - \omega_\tau^h}} \hat{\mathbf{e}}_{\hat{\mu}_{\tau, \mathbf{u}_{k, \ell}}^{REC}}^{HAR}(\tau, REC) \quad (100)$$

Let us now give an example which compares the accuracy of the estimates of the error in the error estimate obtained by polynomial and harmonic recycling.

Example 2. *The accuracy of the estimates of the error in the error estimates.* Let us compare the exact relative error in the error-indicator

$$\sigma_{\tau, REL}^{VM, REC} \stackrel{\text{def}}{=} J_2(\mathbf{e}_{\sigma_{REC}(\mathbf{u}^h)}) / J_2(\boldsymbol{\sigma}(\mathbf{e}_{\tau}^{h, loc})) \quad (101)$$

where we recall that $\mathbf{e}_{\sigma_{REC}(\mathbf{u}^h)} \stackrel{\text{def}}{=} \boldsymbol{\sigma}(\mathbf{e}_{\tau}^{h, loc}) - \hat{\mathbf{e}}_{\sigma(\mathbf{u}^h)}^{REC}$ is the exact error in the error indicator, and the estimated relative error

$$\sigma_{\tau, REL}^{VM, REC, REC^*} \stackrel{\text{def}}{=} J_2(\hat{\mathbf{e}}_{\sigma_{REC}(\mathbf{u}^h)}^{REC^*}) / J_2(\boldsymbol{\sigma}(\mathbf{e}_{\tau}^{h, loc})) \quad (102)$$

and we measure the accuracy by computing the element effectivity indices

$$\kappa_{\tau}^{loc, REC, REC^*} \stackrel{\text{def}}{=} \|\hat{\mathbf{e}}_{\sigma_{REC}(\mathbf{u}^h)}^{REC^*}\|_{\tau, \mathbf{E}^{-1}} / \|\mathbf{e}_{\sigma_{REC}(\mathbf{u}^h)}\|_{\tau, \mathbf{E}^{-1}} \quad (103)$$

where $REC^* \equiv POL, HAR$. Let us first consider the case of polynomial recycling ($REC^* \equiv POL$). In Fig. 11 we give the shades of the exact relative error in the REC indicator $\sigma_{\tau, REL}^{VM, REC}$ (Fig. 11a, for $REC \equiv ER2B4$, Fig. 11b, for $REC \equiv ZZ-SPR$) and

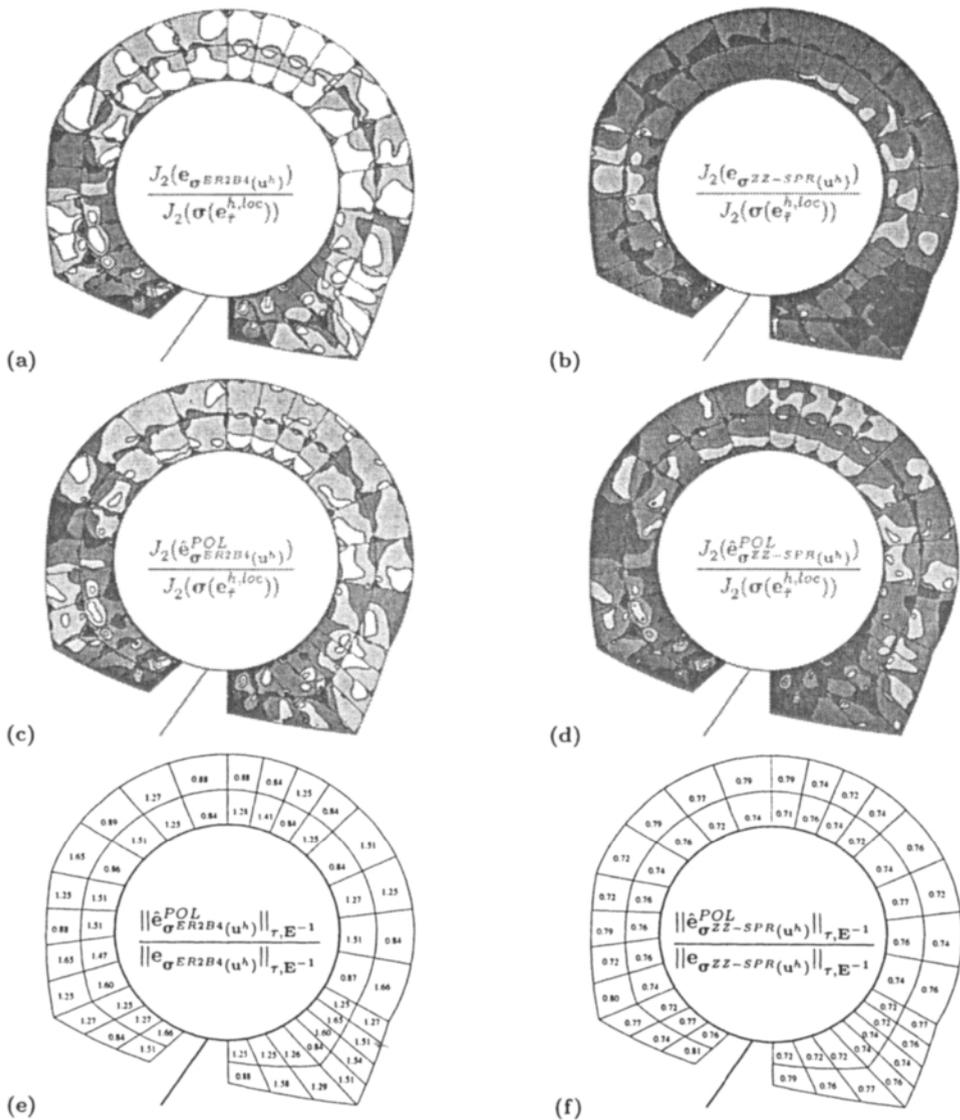


Fig. 11. An example which illustrates the accuracy of the estimate of the error in the error estimate: The accuracy of the indicators of the error in the error indicators in the elements of patch A. The regions of 0–5% (white), 5–10% (light gray), 10–20% (gray), 20–30% (dark gray), above 30% (black) for: The exact relative error in the error indicator, $J_2(\mathbf{e}_{\sigma_{REC}}(\mathbf{u}^h))/J_2(\sigma(\mathbf{e}_\tau^{h,loc}))$, for (a) $REC \equiv ER2B4$, and (b) $REC \equiv ZZ-SPR$. The estimated relative error in the error indicators, $J_2(\hat{\mathbf{e}}_{\sigma_{REC}}^{POL}(\mathbf{u}^h))/J_2(\sigma(\mathbf{e}_\tau^{h,loc}))$, for (c) $REC \equiv ER2B4$, and (d) $REC \equiv ZZ-SPR$, obtained using polynomial recycling. The values of the element effectivity indices $\|\hat{\mathbf{e}}_{\sigma_{REC}}^{POL}(\mathbf{u}^h)\|_{\tau, E^{-1}}/\|\mathbf{e}_{\sigma_{REC}}(\mathbf{u}^h)\|_{\tau, E^{-1}}$ for the estimate of the error in the error indicator, for: (e) $REC \equiv ER2B4$, (f) $REC \equiv ZZ-SPR$.

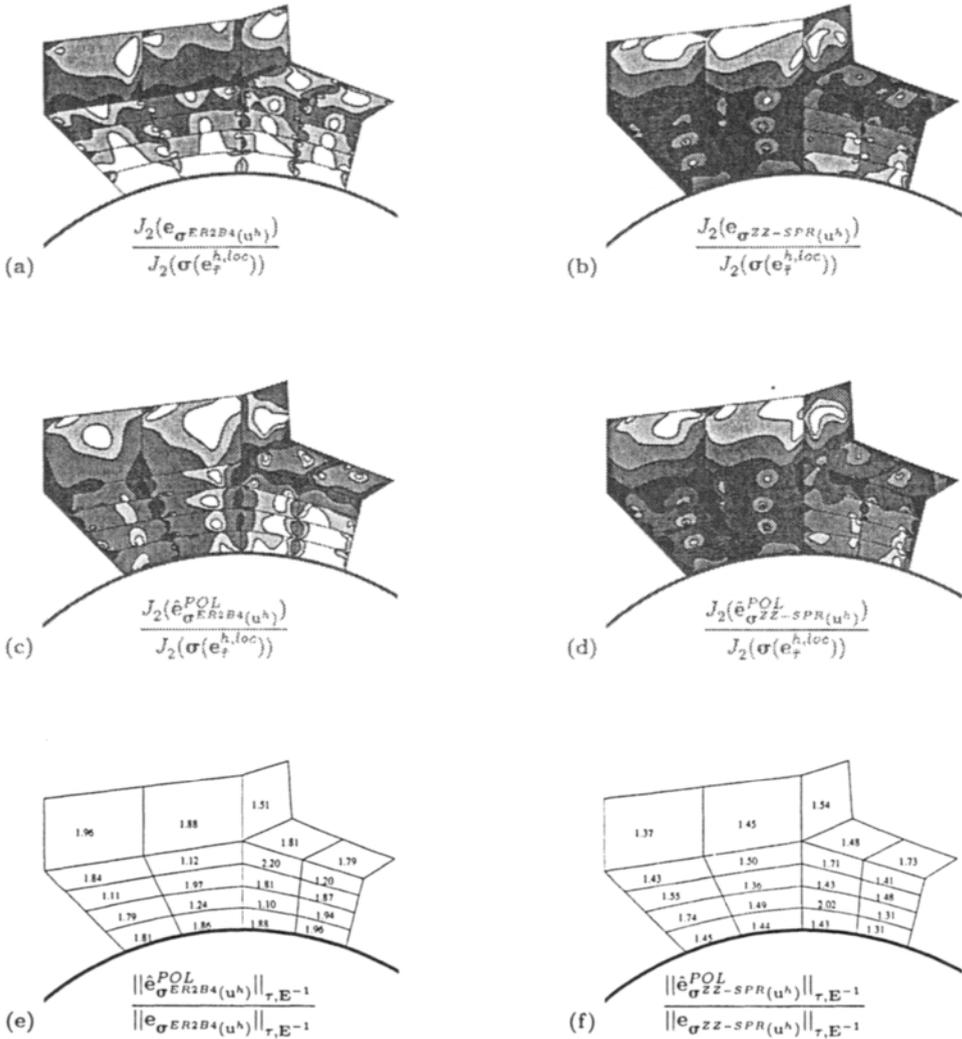


Fig. 12. An example which illustrates the accuracy of the estimate of the error in the error estimate: The accuracy of the indicators of the error in the error indicators in the elements of patch B . The regions of 0–5% (white), 5–10% (light gray), 10–20% (dark gray), 20–30% (gray), above 30% (black) for: The exact relative error in the error indicator, $J_2(\mathbf{e}_{\sigma_{REC}(u^h)})/J_2(\sigma(\mathbf{e}_T^{h,loc}))$, for (a) $REC \equiv ER2B4$, and (b) $REC \equiv ZZ-SPR$. The estimated relative error in the error indicators, $J_2(\hat{\mathbf{e}}_{\sigma_{REC}^{POL}(u^h)})/J_2(\sigma(\mathbf{e}_T^{h,loc}))$, for (c) $REC \equiv ER2B4$, and (d) $REC \equiv ZZ-SPR$, obtained using polynomial recycling. The values of the element effectivity indices $\|\hat{\mathbf{e}}_{\sigma_{REC}^{POL}(u^h)}\|_{\tau,E^{-1}}/\|\mathbf{e}_{\sigma_{REC}(u^h)}\|_{\tau,E^{-1}}$ for the estimate of the error in the error indicator, for: (e) $REC \equiv ER2B4$, (f) $REC \equiv ZZ-SPR$.

the corresponding estimated relative errors in the REC indicator $\sigma_{\tau,REL}^{VM,REC,POL}$ (Fig. 11c, for $REC \equiv ER2B4$, Fig. 11d, for $REC \equiv ZZ-SPR$), and the element effectivity indices $\kappa_{\tau}^{loc,REC,POL}$ (Fig. 11e, for $REC \equiv ER2B4$, Fig. 11f, for $REC \equiv ZZ-SPR$) for the elements in the patch A . In Fig. 12 we give the corresponding results for the elements in patch B . Note that the indicators $\hat{\epsilon}_{\sigma_{REC}(u^h)}^{POL}$ have effectivity indices in the range (1.0, 3.0). Let us now compute the estimates for the error in the stress in pollution estimate, i.e. $\mathcal{E}_{\sigma_{avg}(\tau)}^{POL}$ given in (85). The accuracy of $\mathcal{E}_{\sigma_{avg}(\tau)}^{POL}$ depends on the accuracy of the indicators of the error in the pollution indicators.

Let us now illustrate the accuracy of the ‘‘harmonic’’ recycling. In Fig. 13 we give the shades of the exact relative error in the REC indicator $\sigma_{\tau,REL}^{VM,REC}$ (Fig. 13a for $REC \equiv ER2B4$, Fig. 13d for $REC \equiv ZZ-SPR$), and the corresponding estimated relative errors in the REC indicator $\sigma_{\tau}^{VM,REC,HAR}$ (Fig. 13b for $REC \equiv ER2B4$, Fig. 13e for $REC \equiv ZZ-SPR$). In Fig. 13c and Fig. 13f we give the effectivity indices $\kappa_{\tau}^{loc,REC,HAR*}$ for $REC \equiv ER2B4$ and $REC \equiv ZZ-SPR$, respectively. Let us compare the estimated error in the error estimate $\mathcal{E}_{\sigma(x)}^{HAR}$ with the exact error in the error estimate given by

$$e_{\mathcal{E}_{\sigma(x)}^{REC}} \stackrel{\text{def}}{=} \sigma(e^h)(x) - \mathcal{E}_{\sigma(x)}^{REC} \quad (104)$$

for $REC \equiv ER2B4$ and $REC \equiv ZZ-SPR$. In Fig. 14 we give the shades of the exact relative error in $\mathcal{E}_{\sigma(x)}^{REC}$ (Fig. 14a for $REC \equiv ER2B4$, Fig. 14d for $REC \equiv ZZ-SPR$), and the corresponding estimated relative error regions (Fig. 14b for $REC \equiv ER2B4$, Fig. 14e for $REC \equiv ZZ-SPR$). In Fig. 14c and Fig. 14f we give the effectivity indices $\|\mathcal{E}_{\sigma(x)}^{HAR}\|_{\tau,E^{-1}} / \|e_{\mathcal{E}_{\sigma(x)}^{REC}}\|_{\tau,E^{-1}}$ for $REC \equiv ER2B4$ and $REC \equiv ZZ-SPR$, respectively. In Fig. 15 and Fig. 16 we give the results which are analogous to Fig. 13 and Fig. 14, respectively, for the patch B . We note that the effectivity indices in all the above cases is in the range (0.7, 1.3).

Let us now employ the ‘‘harmonic’’ recycling to predict the effectivity index of the global energy norm of the error, κ_{Ω}^{REC} . Recalling that

$$(\mathcal{E}_{\Omega}^{REC})^2 \stackrel{\text{def}}{=} \sum_{\tau \in T_h} \int_{\tau} \left(\sigma^{REC}(u^h) - \sigma(u^h) \right)^T \mathbf{E}^{-1} \left(\sigma^{REC}(u^h) - \sigma(u^h) \right) \quad (105)$$

letting $\sigma^{REC}(u^h) - \sigma(u^h) = \left(\sigma^{REC}(u^h) - \sigma(u) \right) + \left(\sigma(u) - \sigma(u^h) \right)$, and expanding, we

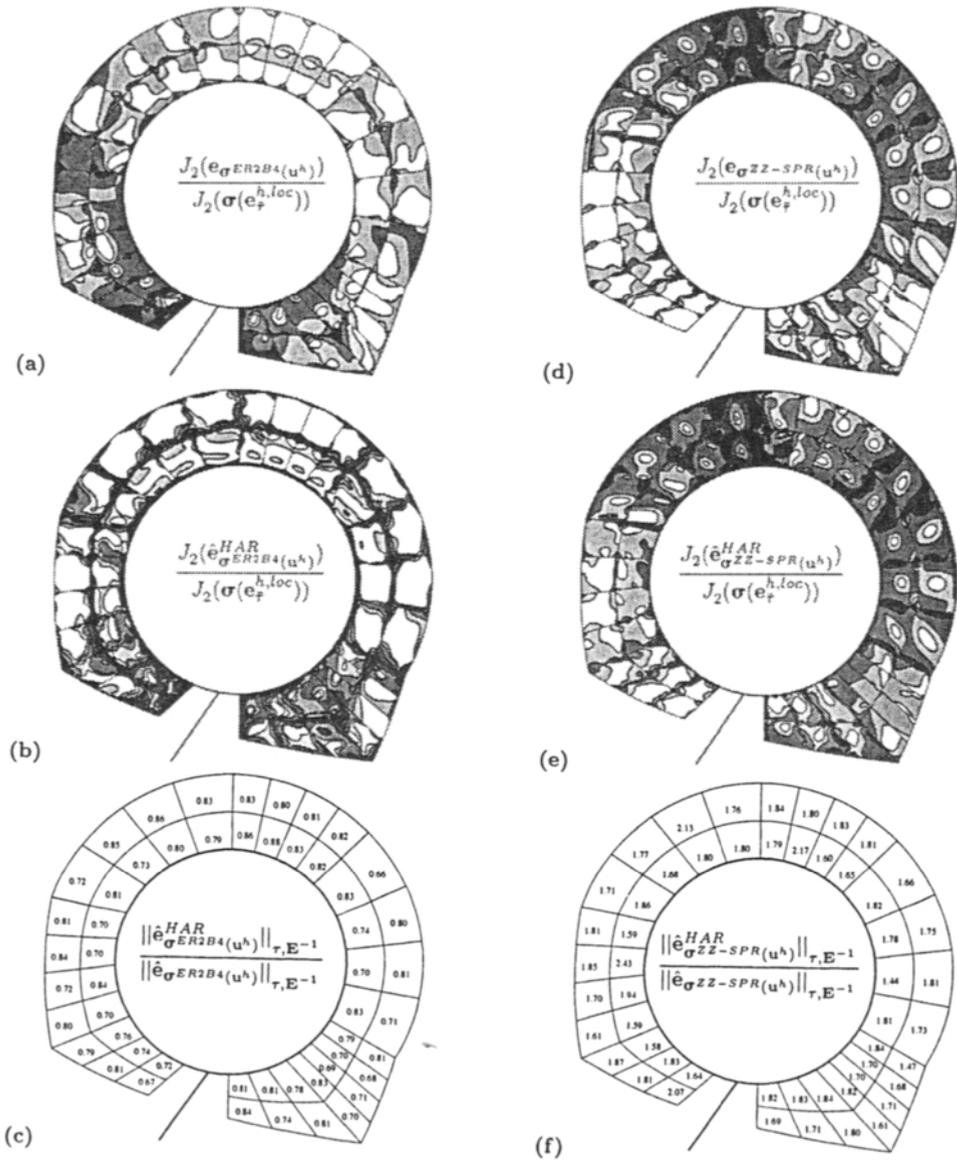


Fig. 13. An example which illustrates the estimate of error in the error indicator: Accuracy of the indicators of the error in the error indicators in the elements of patch A. The regions of 0-5% (white), 5-10% (light gray), 10-20% (gray), 20-30% (dark gray), above 30% (black) for: (a) $J_2(e_{\sigma ER2B4}(u^h))/J_2(\sigma(e_+^{h,loc}))$; (b) $J_2(\hat{e}_{\sigma ER2B4}^{HAR}(u^h))/J_2(\sigma(e_+^{h,loc}))$; (c) $\|\hat{e}_{\sigma ER2B4}^{HAR}(u^h)\|_{\tau,E-1}/\|\hat{e}_{\sigma ER2B4}(u^h)\|_{\tau,E-1}$; (d) $J_2(e_{\sigma ZZ-SPR}(u^h))/J_2(\sigma(e_+^{h,loc}))$; (e) $J_2(\hat{e}_{\sigma ZZ-SPR}^{HAR}(u^h))/J_2(\sigma(e_+^{h,loc}))$; (f) $\|\hat{e}_{\sigma ZZ-SPR}^{HAR}(u^h)\|_{\tau,E-1}/\|\hat{e}_{\sigma ZZ-SPR}(u^h)\|_{\tau,E-1}$.

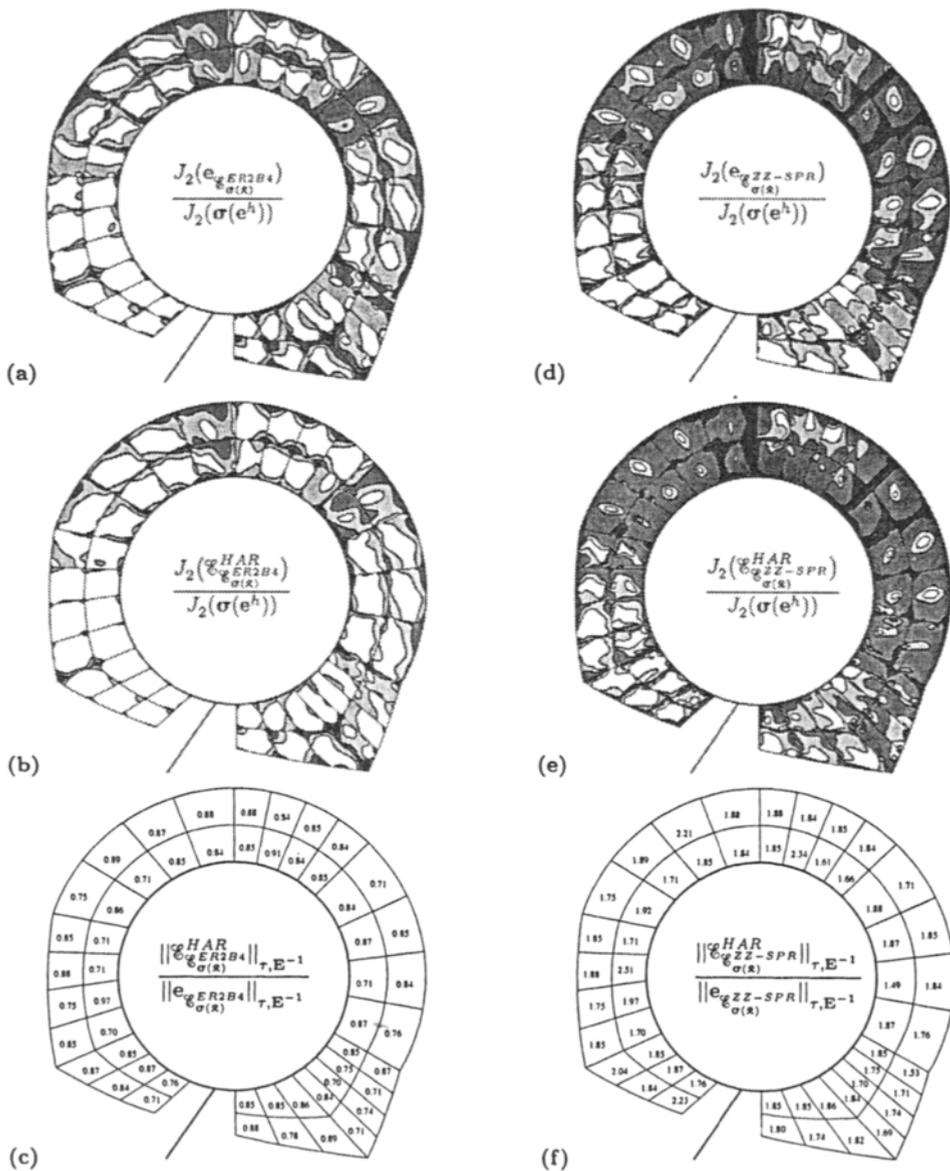


Fig. 14. An example which illustrates the estimate of error in the error indicator: Accuracy of the indicators of the error in the error indicators in the elements of patch *A*. The regions of 0–5% (white), 5–10% (light gray), 10–20% (gray), 20–30% (dark gray), above 30% (black) for: (a) $J_2(\mathbf{e}_{\mathcal{E}ER2B4}^{\sigma(R)})/J_2(\sigma(\mathbf{e}^h))$; (b) $J_2(\mathcal{E}_{\sigma(R)}^{HAR})/J_2(\sigma(\mathbf{e}^h))$; (c) $\|\mathcal{E}_{\sigma(R)}^{HAR}\|_{r,E-1}/\|\mathbf{e}_{\sigma(R)}^{ER2B4}\|_{r,E-1}$; (d) $J_2(\mathbf{e}_{\mathcal{E}ZZ-SPR}^{\sigma(R)})/J_2(\sigma(\mathbf{e}^h))$; (e) $J_2(\mathcal{E}_{\sigma(R)}^{HAR-SPR})/J_2(\sigma(\mathbf{e}^h))$; (f) $\|\mathcal{E}_{\sigma(R)}^{HAR-SPR}\|_{r,E-1}/\|\mathbf{e}_{\sigma(R)}^{ZZ-SPR}\|_{r,E-1}$.

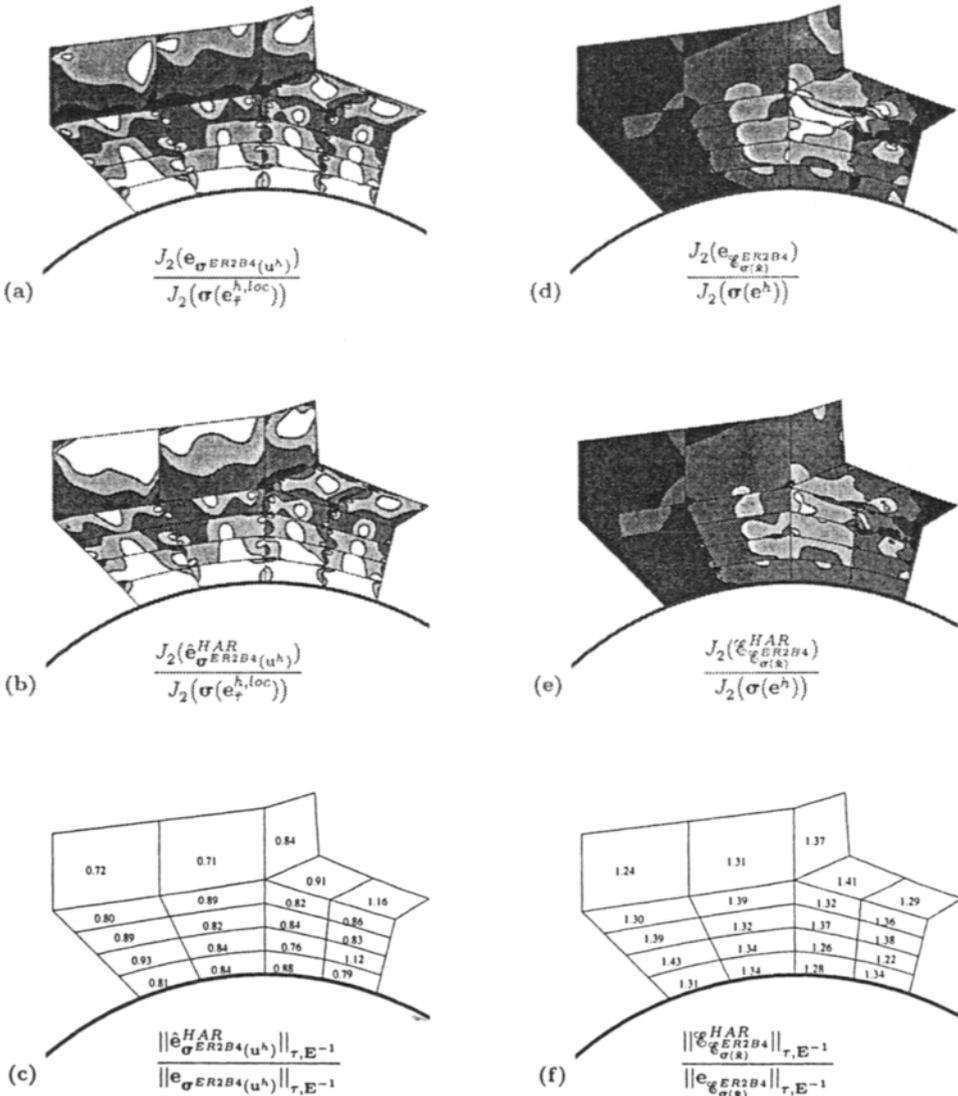


Fig. 15. An example which illustrates the estimate of error in the error indicator: Accuracy of the indicators of the error in the error indicators in the elements of patch B . The regions of 0-5% (white), 5-10% (light gray), 10-20% (gray), 20-30% (dark gray), above 30% (black) for: (a) $J_2(e_{\sigma_{ER2B4}(u^h)})/J_2(\sigma(e_r^{h,loc}))$; (b) $J_2(\hat{e}_{\sigma_{ER2B4}^{HAR}}(u^h))/J_2(\sigma(e_r^{h,loc}))$; (c) $\|\hat{e}_{\sigma_{ER2B4}^{HAR}}(u^h)\|_{\tau,E^{-1}}/\|e_{\sigma_{ER2B4}}(u^h)\|_{\tau,E^{-1}}$; (d) $J_2(e_{\mathcal{E}_{\sigma(R)}^{ER2B4}})/J_2(\sigma(e^h))$; (e) $J_2(\hat{\mathcal{E}}_{\sigma(R)}^{HAR})/J_2(\sigma(e^h))$; (f) $\|\hat{\mathcal{E}}_{\sigma(R)}^{HAR}\|_{\tau,E^{-1}}/\|\mathcal{E}_{\sigma(R)}^{ER2B4}\|_{\tau,E^{-1}}$.

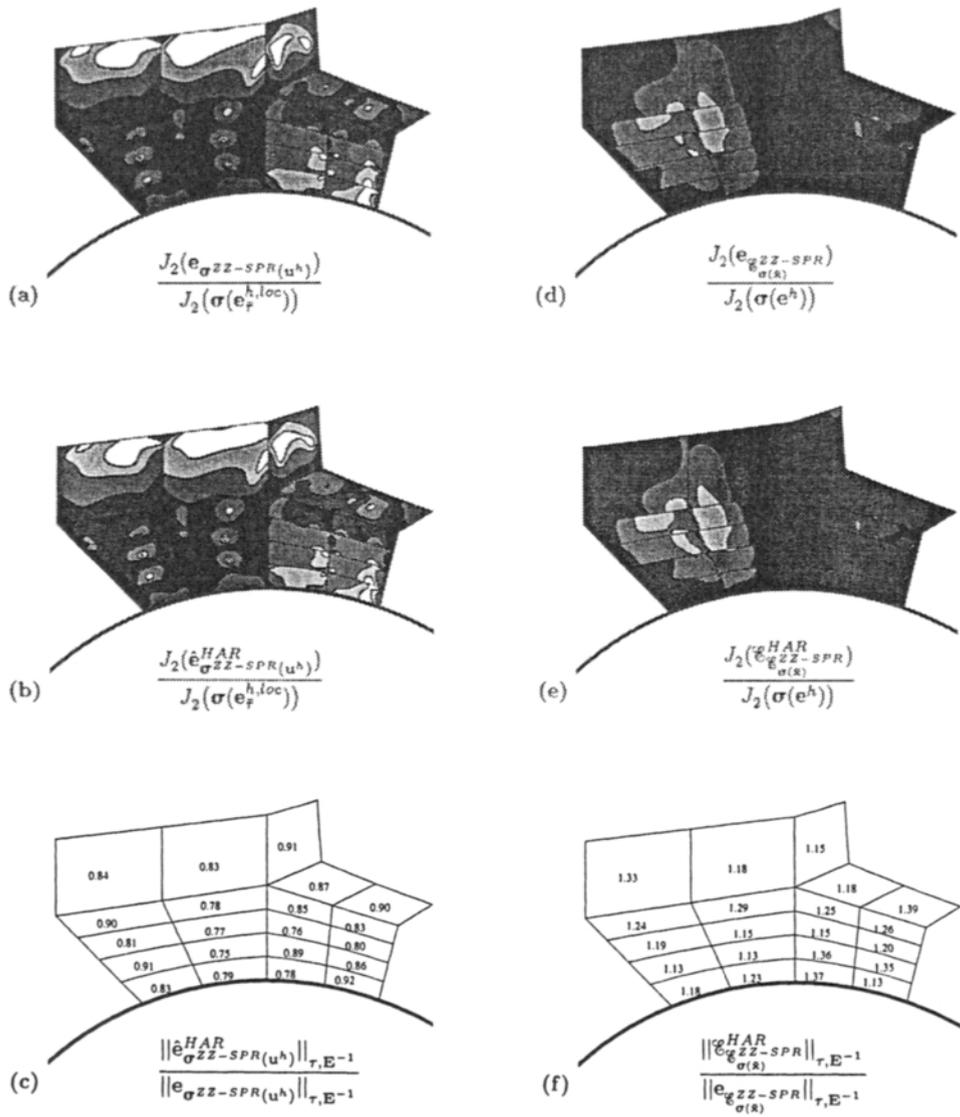


Fig. 16. An example which illustrates the estimate of error in the error indicator: Accuracy of the indicators of the error in the error indicators in the elements of patch *B*. The regions of 0–5% (white), 5–10% (light gray), 10–20% (gray), 20–30% (dark gray), above 30% (black) for: (a) $J_2(e_{\sigma^{ZZ-SPR}}(u^h))/J_2(\sigma(e^h_{\tau^{loc}}))$; (b) $J_2(\hat{e}_{\sigma^{ZZ-SPR}}^{HAR}(u^h))/J_2(\sigma(e^h_{\tau^{loc}}))$; (c) $\|\hat{e}_{\sigma^{ZZ-SPR}}^{HAR}(u^h)\|_{\tau,E^{-1}}/\|e_{\sigma^{ZZ-SPR}}(u^h)\|_{\tau,E^{-1}}$; (d) $J_2(e_{\hat{\sigma}^{ZZ-SPR}})/J_2(\sigma(e^h))$; (e) $J_2(\hat{e}_{\hat{\sigma}^{ZZ-SPR}}^{HAR})/J_2(\sigma(e^h))$; (f) $\|\hat{e}_{\hat{\sigma}^{ZZ-SPR}}^{HAR}\|_{\tau,E^{-1}}/\|e_{\hat{\sigma}^{ZZ-SPR}}\|_{\tau,E^{-1}}$.

get

$$(\mathcal{E}_{\Omega}^{REC})^2 - |||e^h|||_{\Omega}^2 = \left. \begin{aligned} & \sum_{\tau \in T_h} \int_{\tau} (\boldsymbol{\sigma}^{REC}(\mathbf{u}^h) - \boldsymbol{\sigma}(\mathbf{u}))^T \mathbf{E}^{-1} (\boldsymbol{\sigma}^{REC}(\mathbf{u}^h) - \boldsymbol{\sigma}(\mathbf{u})) \\ & + 2 \sum_{\tau \in T_h} \int_{\tau} (\boldsymbol{\sigma}^{REC}(\mathbf{u}^h) - \boldsymbol{\sigma}(\mathbf{u}))^T \mathbf{E}^{-1} (\boldsymbol{\sigma}(\mathbf{u}) - \boldsymbol{\sigma}(\mathbf{u}^h)) \end{aligned} \right\} \quad (106)$$

Employing $\boldsymbol{\sigma}^{REC, REC^*}(\mathbf{u}^h)$ instead of $\boldsymbol{\sigma}(\mathbf{u})$ in (106) we get the estimate

$$(\mathcal{E}_{\Omega}^{REC})^2 - |||e^h|||_{\Omega}^2 \approx \mathcal{E}_{(\mathcal{E}_{\Omega}^{REC^*})^2}^{REC^*} \quad (107)$$

where

$$\mathcal{E}_{(\mathcal{E}_{\Omega}^{REC^*})^2}^{REC^*} \stackrel{\text{def}}{=} \sum_{\tau \in T_h} \int_{\tau} (\hat{\boldsymbol{\sigma}}_{\boldsymbol{\sigma}^{REC^*}(\mathbf{u}^h)}^{REC^*})^T \mathbf{E}^{-1} \hat{\boldsymbol{\sigma}}_{\boldsymbol{\sigma}^{REC^*}(\mathbf{u}^h)}^{REC^*} - 2 \sum_{\tau \in T_h} \int_{\tau} (\hat{\boldsymbol{\sigma}}_{\boldsymbol{\sigma}^{REC^*}(\mathbf{u}^h)}^{REC^*})^T \mathbf{E}^{-1} \hat{\boldsymbol{\sigma}}_{\boldsymbol{\sigma}(\mathbf{u}^h)}^{REC^*} \quad (108)$$

We have

$$(\kappa_{\Omega}^{REC})^2 \stackrel{\text{def}}{=} \frac{(\mathcal{E}_{\Omega}^{REC})^2}{|||e^h|||_{\Omega}^2} \approx \mathcal{E}_{(\kappa_{\Omega}^{REC^*})^2}^{REC^*} \stackrel{\text{def}}{=} 1 + \frac{\mathcal{E}_{(\mathcal{E}_{\Omega}^{REC^*})^2}^{REC^*}}{(\mathcal{E}_{\Omega}^{REC})^2 - \mathcal{E}_{(\mathcal{E}_{\Omega}^{REC^*})^2}^{REC^*}} \quad (109)$$

We considered the mesh shown in Fig. 3b (we will call this mesh the *refined* mesh). We also employed a coarser mesh which is obtained from the initial mesh of the superelements by subdividing each element only once (the refined mesh was obtained from the initial mesh of superelements by employing two recursive subdivisions). In Table 1 we give the values of κ_{Ω}^{ER2B4} and κ_{Ω}^{ZZ-SPR} for the coarse mesh and the refined mesh and we also compare the estimated values of effectivity indices $\sqrt{\mathcal{E}_{(\kappa_{\Omega}^{REC^*})^2}^{HAR}}$. Note that we have $\sqrt{\mathcal{E}_{(\kappa_{\Omega}^{REC^*})^2}^{HAR}} \approx \kappa_{\Omega}^{REC}$ for both meshes.

Mesh	ER2B4		ZZ-SPR	
	κ_{Ω}^{ER2B4}	$\sqrt{\mathcal{E}_{(\kappa_{\Omega}^{ER2B4})^2}^{HAR}}$	κ_{Ω}^{ZZ-SPR}	$\sqrt{\mathcal{E}_{(\kappa_{\Omega}^{ZZ-SPR})^2}^{HAR}}$
Coarse mesh	1.877	2.034	2.213	2.469
Refined mesh	1.938	1.976	2.306	2.514

Table 1. An example which illustrates the accuracy of the estimates of the global effectivity index: The values of the global effectivity index κ_{Ω}^{REC} for the coarse mesh and the refined mesh (shown in Fig. 3b), and its estimate $\sqrt{\mathcal{E}_{(\kappa_{\Omega}^{REC^*})^2}^{HAR}}$. Note that $\sqrt{\mathcal{E}_{(\kappa_{\Omega}^{REC^*})^2}^{HAR}} \approx \kappa_{\Omega}^{REC}$ for both meshes.

3. CONCLUSIONS

The main conclusions of the paper are:

1. We have constructed reliable estimates of the error in the stress at any point (where the stress is finite) by splitting the error into two components with respect to the element which includes the point of interest. The local component (local error) is estimated by the error indicator, and the global component (pollution error) is estimated by a global extraction.
2. The estimates of the error in the stress in the engineering model problem are sufficiently accurate when the element residual error indicator or the *ZZ-SPR* error indicator are employed. Nevertheless, whenever the mesh is coarse, the effectivity index may be in the range (0.5, 2.2), and it is essential to estimate the error in the error estimate in order to guarantee the quality of the error estimate.
3. The a-posteriori estimation of the error in the error estimate is also based on splitting of the error in the error estimate, namely, the error in the error indicator and the error in the global extraction. The error in the error indicator and the error in the global extraction are estimated by locally averaging (recycling) the original error indicators.
4. The a-posteriori estimates of the error in the error estimates are most effective when the local character of the exact solution is taken into account in the recycling.

ACKNOWLEDGMENTS

The work of I. B. was supported by the Office of Naval Research under contracts N00014-90-J-1030 and N00014-96-10019, and by the National Science Foundation under Grant DMS-91-20877. The work of T.S. and S.K.G. was supported by the U.S. Army Research Office under Grant DAAL03-G-028, by the National Science Foundation under Grant MSS-9025110 and by the Office of Naval Research under Grant N00014-96-1-0021 and Grant N00014-96-1-1015.

REFERENCES

- [1] P. Ladeveze, 'Comparaison de modeles de milieux continus', *Thèse de Doctorat d'Etat es Science, Mathematiques*, L'Université P. et M. Curie, Paris, (1975).

- [2] P. Ladeveze and D. Leguillon, Error estimate procedure in the finite element method and applications, *SIAM J. Numer. Anal.*, **20**, 3, (1983) 485-509.
- [3] P. Rougeot, 'Sur le controle de la qualite des maillages elements finis', Ph.D. Thesis, Laboratoire de Mecanique et Technologie, E.N.S de Cachan-C.N.R.S-Universite Paris 6, Cachan, France (1989).
- [4] I. Babuška, T. Strouboulis, C.S. Upadhyay and S.K. Gangaraj, A model study of element residual estimators for linear elliptic problems: The quality of the estimators in the interior of meshes of triangles and quadrilaterals, *Computers and Structures*, **57**, (1995) 1009-1028.
- [5] I. Babuška, T. Strouboulis, C.S. Upadhyay and S.K. Gangaraj, A posteriori estimation and adaptive control of the pollution error in the h -version of the finite element method, *Int. J. Numer. Methods Engrg.*, **38**, (1995) 4207-4235.
- [6] M. Ainsworth and J.T. Oden, A-posteriori error estimation in finite element analysis, *Comput. Methods Appl. Mech. Engrg.: Computational Mechanics Advances*, **142**, (1997) 1-88.
- [7] J.T. Oden and Y. Feng, Local and pollution error estimation for finite element approximations of elliptic boundary value problems, *J. Comput. Appl. Math.*, **74**, (1996) 245-293.
- [8] I. Babuška, T. Strouboulis and C.S. Upadhyay, A model study of the quality of a-posteriori error estimators for linear elliptic problems. Error estimation in the interior of patchwise uniform grids of triangles, *Comput. Methods Appl. Mech. Engrg.*, **114**, (1994) 307-378.
- [9] B.A. Szabo and I. Babuška, *Finite Element Analysis*, John Wiley & Sons, Inc., New York, 1991.
- [10] L.B. Wahlbin, Local behavior in finite element methods, in: P.G. Ciarlet and J.L. Lions, eds., *Handbook of Numerical Analysis, Vol. II*, (North-Holland, Amsterdam, 1991) 357-522.
- [11] I. Babuška, T. Strouboulis and C.S. Upadhyay, A model study of the quality of a-posteriori error estimators for finite element solutions of linear elliptic problems, with particular reference to the behavior near the boundary, *Internat. J. Numer. Methods Engrg.*, **40**, (1997) 2521-2577.
- [12] I. Babuška and T. Strouboulis, *The Analysis and Design of Finite Element Methods*, in preparation.
- [13] I. Babuška, T. Strouboulis, A. Mathur and C.S. Upadhyay, Pollution-error in the h -version of the finite-element method and the local quality of a-posteriori error estimators, *Finite Elements in Analysis and Design*, **17**, (1994) 237-321.

- [14] J.T. Oden, L. Demkowicz, W. Rachowicz and T.A. Westermann, Toward a universal h - p adaptive finite element strategy, Part 2. A posteriori error estimation, *Comput. Methods Appl. Mech. Engrg.*, **77**, (1989) 113–180.
- [15] O.C. Zienkiewicz and J.Z. Zhu, The superconvergent patch recovery and a posteriori error estimators. Part 1. The recovery technique, *Int. J. Numer. Methods Engrg.*, **33**, (1992) 1331–1364.
- [16] O.C. Zienkiewicz and J.Z. Zhu, The superconvergent patch recovery and a posteriori error estimators. Part 2. Error estimates and adaptivity, *Int. J. Numer. Methods Engrg.*, **33**, (1992) 1365–1382.
- [17] O.C. Zienkiewicz and J.Z. Zhu, The superconvergent patch recovery (SPR) and adaptive finite element refinement, *Comput. Methods Appl. Mech. Engrg.*, **101**, (1992) 207–224.
- [18] N.E. Wiberg, F. Abdulwahab and S. Ziukas, 'Enhanced superconvergent patch recovery incorporation equilibrium and boundary conditions', *Internat. J. Numer. Methods Engrg.*, **37**, (1994) 3417–3440.
- [19] I. Babuška and A. Miller, The post-processing approach in the finite element method - Part 1: Calculation of displacements, stress and other higher derivatives of displacement, *Int. J. Numer. Methods Engrg.*, **20** (1984) 1085–1109.
- [20] I. Babuška and A. Miller, The post-processing approach in the finite element method - Part 3: A-posteriori error estimates and adaptive mesh selection, *Int. J. Numer. Methods Engrg.*, **20** (1984) 2311–2324.
- [21] R. Rannacher and F.-T. Suttmeier, A feedback approach to error control in finite element methods: Application to linear elasticity, Preprint 96-42, Institut für Angewandte Mathematik, Universität Heidelberg, Sept., (1996).
- [22] R. Becker and R. Rannacher, A feedback approach to error control in finite element methods: Basic analysis and examples, *East-West J. Numer. Math.*, **4**, (1996), 237–264.
- [23] I. Babuška, T. Strouboulis, S.K. Gangaraj, K. Copps and D.K. Datta, Practical aspects of a-posteriori estimation for reliable finite element analysis, *Computer and Structures* (in press).
- [24] I. Babuška, T. Strouboulis, S.K. Gangaraj and C.S. Upadhyay, Pollution error in the h -version of the finite element method and the local quality of the recovered derivatives, *Comput. Methods Appl. Mech. Engrg.*, **140**, (1997), 1–37.
- [25] I. Babuška, T. Strouboulis, D.K. Datta, K. Copps and S.K. Gangaraj, A-posteriori estimation and adaptive control of the error in the derivatives, strains and stresses at a point of interest, (in preparation).

- [26] I. Babuška, T. Strouboulis and S.K. Gangaraj, A-posteriori estimation of the error in the recovered derivatives of the finite element solution, *Comput. Methods Appl. Mech. Engrg.*, **150**, (1997) 369–396.
- [27] N.I. Muskhelishvili, *Some Basic Problems of the Mathematical Theory of Elasticity*, E.P. Noordhoff, Groningen, Netherlands, 1963.

This Page Intentionally Left Blank

Bounds for Linear–Functional Outputs of Coercive Partial Differential Equations : Local Indicators and Adaptive Refinement

J. Peraire ^a and A.T. Patera ^b

^aDepartment of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

^bDepartment of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

We present a framework for the efficient calculation of lower and upper bounds to outputs which are linear functionals of the solutions to symmetric or nonsymmetric second-order coercive partial differential equations. The method is based upon the construction of an augmented Lagrangian, in which the objective is a quadratic energy reformulation of the desired output, and the constraints are the finite element equilibrium conditions and interelement continuity requirements; the bounds are then derived by evoking the dual max-min problem for appropriately chosen candidate Lagrange multipliers. The bound computation comprises two components: several global calculations on a relatively coarse “working” finite-element triangulation \mathcal{T}_H consisting of K_H elements T_H ; and $2K_H$ independent T_H -local calculations on a relatively fine “truth” finite-element triangulation \mathcal{T}_h .

In this paper we focus on three new developments. First, we introduce a modified energy objective, and hence modified Lagrangian, that permits both more transparent interpretation and more ready generalization. Second, we demonstrate that the bound gap — the difference between the upper and lower bounds for the desired output — can be represented as the sum of positive contributions — local indicators — associated with the elements T_H of \mathcal{T}_H . Third, based on these local bound-gap error indicators, we develop adaptive strategies by which to reduce the bound gap — and hence improve our validated prediction for the output of interest — through optimal refinement of \mathcal{T}_H . The resulting method is applied to an illustrative problem in linear elasticity.

1. INTRODUCTION

The field of *a posteriori* error estimation and adaptive mesh refinement now has a long history in finite element analysis. The two goals of these related pursuits are, first, inexpensive confirmation of the accuracy of a particular finite element solution, and second, efficient improvement of the finite element solution by optimal adaptive mesh refinement. We now make these notions more precise.

We denote the exact solution to our coercive partial differential equation by u (say, the displacement), and the finite element approximation to u associated with a triangu-

lation \mathcal{T}_H by u_H . The finite element error is thus given by $e \equiv u - u_H$; we denote the (pseudo)metric in which we wish to measure the finite element error by $E(v)$. *A posteriori* procedures provide an estimate, $\mathcal{E}(\mathcal{T}_H)$, for the E -metric of the finite element error, $E(e)$; this estimate $\mathcal{E}(\mathcal{T}_H)$ is typically expressed as the sum of positive contributions \mathcal{E}_{T_H} associated with the K_H elements T_H of the triangulation \mathcal{T}_H . The elemental contributions \mathcal{E}_{T_H} are interpreted as local indicators for the purpose of subsequent mesh refinement strategies.

The general approaches to *a posteriori* error estimation may be categorized as “explicit” or “implicit” [3]. Explicit techniques are typically based on residual evaluation and *a priori* approximation and stability results [4,18,7]: the advantage is computational efficiency; the disadvantage is the presence of constants that can not be precisely evaluated. Implicit techniques [11,6,2] are based on the solution of K_H residual-forced T_H -local independent subproblems: the advantage is more precise quantification; the disadvantage is increased complexity and computational effort. Although both explicit and implicit methods typically provide error bounds in the sense that (roughly) there exists a constant $C_{\mathcal{E}}$ independent of \mathcal{T}_H and u such that $E(e) \leq C_{\mathcal{E}} \mathcal{E}_{T_H}$ [6,18], the constant $C_{\mathcal{E}}$ is typically not known. In what follows, we shall reserve the term “bound” for estimators for which $C_{\mathcal{E}}$ is known to high accuracy.

Most early work on *a posteriori* error estimation focused, first, on symmetric problems, and second, on the natural energy measure of the error, in which $E(v)$ is chosen to be the norm induced by the symmetric bilinear form associated with the weak formulation of the problem. In this case, very effective explicit and implicit techniques [4,6,18] can be developed that require essentially no regularity assumptions and contain only minimal unknown approximation (and perhaps coercivity) contributions to $C_{\mathcal{E}}$. Furthermore, implicit procedures can be developed [11,2] that provide rigorous *bounds* for the error; the unknown contributions to $C_{\mathcal{E}}$ are reduced to the (typically very small) inaccuracies incurred in the solution of the T_H -local subproblems. Many of these explicit and implicit methods can be readily generalized to nonsymmetric problems [18] and more general error metrics; in contrast, the earlier bound procedures [11,2] are developed only for coercive symmetric problems, and are fundamentally restricted to the natural energy norm.

There has recently been greatly increased interest in the extension of *a posteriori* estimation techniques to error metrics more directly relevant to engineering analysis. In particular, the quantity (or quantities) of interest in engineering studies is not the field variable u , or the error in the energy norm, but rather the output — the system performance metric — that reflects the specific goals and objectives of the design or optimization exercise. To be more precise, we denote this engineering output of interest by s (say, the force over part of the boundary); we further assume that s may be expressed as $s = l^O(u)$, where $l^O(v)$ is a (preferably bounded) affine functional. The finite element approximation to s , s_H , is then given by $s_H = l^O(u_H)$. It is clear that, in order to measure the error in the linear-functional output s , we should choose $E(v) \equiv |l^O(v)|$, since then $E(e) = |l^O(e)| = |l^O(u - u_H)| = |s - s_H|$.

Explicit error estimation techniques — based on the Aubin-Nitsche duality procedure [10] — and associated adaptive refinement strategies are now available for linear-functional outputs, $E(v) \equiv |l^O(v)|$ [7,8,16]. These procedures facilitate the assessment and improvement of finite element predictions for the desired engineering output, s . However,

in contrast to explicit indicators for the error in the energy norm, explicit indicators for the error in linear-functional outputs involve, first, more questionable assumptions, and second, more unknown contributions to $C_\mathcal{E}$. In particular, $C_\mathcal{E}$ now reflects not only undetermined approximation constants, but also regularity hypotheses on the adjoint (dual) variable ψ , and potentially inaccurate (\mathcal{T}_H) estimates for higher-derivative norms of ψ . Although these explicit techniques for output estimation can certainly yield good results [7,8,16], the goal of confirmation is less well-satisfied. Related reconstruction procedures share similar advantages and disadvantages [5].

We argue that engineering analysis and design is best served by *a posteriori* procedures that eliminate — or at least greatly reduce — the uncertainty in numerical predictions: rigorous and quantitative bounds for $|s - s_H|$ are required if simulation results can be reliably incorporated into the engineering design and optimization process. To this end, we have developed an *implicit* procedure for the prediction of *bounds* for linear-functional outputs of coercive partial differential equations [13,15,14]. The method is based upon the construction of an augmented Lagrangian, in which the objective is a quadratic energy reformulation of the desired output, and the constraints are the finite element equilibrium conditions and interelement continuity requirements; the bounds are then derived by evoking the dual max-min problem for appropriately chosen candidate Lagrange multipliers. As in all implicit procedures, the bound computation comprises two components: several global calculations on the relatively coarse “working” finite-element triangulation \mathcal{T}_H ; and $2K_H$ independent T_H -local calculations on a relatively fine “truth” finite-element triangulation \mathcal{T}_h .

More precisely, our procedure produces quantitative lower and upper bounds, $s_{LO}(\mathcal{T}_H)$ and $s_{UP}(\mathcal{T}_H)$, for $s_h = l^O(u_h)$, the finite element approximation to the output s on the “truth” triangulation \mathcal{T}_h . These bounds, in turn, engender a predictor for s_h ,

$$s_{pre}(\mathcal{T}_H) = \frac{1}{2}(s_{LO}(\mathcal{T}_H) + s_{UP}(\mathcal{T}_H)), \quad (1)$$

for which

$$|s_h - s_{pre}(\mathcal{T}_H)| \leq \Delta(\mathcal{T}_H). \quad (2)$$

Here

$$\Delta(\mathcal{T}_H) = \frac{1}{2}(s_{UP}(\mathcal{T}_H) - s_{LO}(\mathcal{T}_H)), \quad (3)$$

which we shall denote the (half) *bound gap*. It follows that $|s - s_{pre}|$ (respectively, $|s - s_H|$) is within $|s - s_h|$ of Δ (respectively, of $\Delta + |s_{pre} - s_H|$), and thus $C_\mathcal{E}$ is known to within the accuracy of the “truth” mesh. Note that whereas $|s_h - s_H|$ may in fact be smaller than $|s_h - s_{pre}|$, we prefer the predictor s_{pre} since the error *bound* on s_{pre} will be sharper.

Our framework borrows from earlier bound (and other implicit) approaches to *a posteriori* error estimation both in general methodology — the application of quadratic-linear duality theory [11,2] — and in specific technology — the interelement continuity component of the Lagrangian, and relatedly, the construction of equilibrated hybrid fluxes [11,6,2]. However, there are also important differences between the earlier approaches and our new proposal. In particular, earlier bound procedures are based upon a Lagrangian in which the objective function, not the constraints, recovers equilibrium: this

reliance on a variational principle limits these approaches to symmetric operators, and, even more fundamentally, to the natural-energy error norm. The Lagrangian in our approach is rather different: the objective is an energy reformulation of the desired output; and the constraints *include* the finite element equilibrium equations, as is common practice in optimization and control applications. These enhancements permit the treatment of much more general error measures, in particular linear-functional outputs, and the consideration of more general equations, in particular nonsymmetric operators [13,15]. Our method has also been extended to the Stokes equations [14,12] and certain nonlinear problems [14], however treatment of general noncoercive or nonlinear systems is not yet possible.

We refer the reader to our earlier publications [13,15,14,12] for a complete description of our approach. In this paper we focus on three important extensions. First, we develop a modified energy objective — and hence modified Lagrangian — that permits both more transparent interpretation (e.g., in terms of Aubin–Nitsche arguments) and more ready generalization (e.g., to inhomogeneous boundary conditions). Second, we demonstrate that the resulting bound gap, $\Delta(\mathcal{T}_H)$ of (3), can be represented as the sum of positive contributions, $\Delta_{\mathcal{T}_H}(\mathcal{T}_H)$ — local indicators — associated with the elements of \mathcal{T}_H . Third, based on these local bound-gap error indicators, we develop adaptive strategies by which to reduce the bound gap $\Delta(\mathcal{T}_H)$ — and hence the error bound on s_{pre} — through optimal refinement of \mathcal{T}_H . We emphasize that our focus is less on error estimation, and more on output prediction: we view the refinement of \mathcal{T}_H primarily as a means by which to improve our bounds. Relatedly, we anticipate that s_{LO} and s_{UP} are the quantities that will be of greatest interest in engineering design; s_{pre} and $\Delta(\mathcal{T}_H)$ are simply convenient artifices through which we can simultaneously sharpen both the lower and upper bounds.

The outline of the remainder of the paper is as follows. In Section 2 we describe the problem that will be the vehicle for our exposition: a general second-order scalar (nonsymmetric) coercive partial differential equation. We also introduce the finite element spaces and bilinear and linear forms that we will require in subsequent sections. In Section 3 we introduce our modified Lagrangian, and describe the associated bound procedure. In Section 4 we identify the local bound-gap indicators, and present our adaptive refinement strategy. Finally, in Section 5, we provide numerical results for a problem in linear elasticity.

2. PROBLEM STATEMENT AND DEFINITIONS

2.1. Governing Equation

We shall consider a scalar second-order linear coercive boundary-value problem characterized by the following variational statement : find $u \in X^D$, and $s \in \mathbb{R}$, such that

$$a(v, u) = l^N(v), \quad \forall v \in X, \tag{4}$$

and

$$s = l^O(u). \tag{5}$$

The field variable u is defined over a domain $\Omega \subset \mathbb{R}^2$ with boundary $\partial\Omega = \Gamma_D \cup \Gamma_N$; u must reside in $X^D = \{v + u_D \mid v \in X\}$, where $X = \{v \in H^1(\Omega) \mid v = 0 \text{ on } \Gamma_D\}$.

Here $H^1(\Omega)$ is the usual Sobolev space [1]; Γ_D is the portion of the domain boundary on which the essential (Dirichlet) boundary data, g_D , is imposed; u_D is any function in $H^1(\Omega)$ satisfying the essential boundary conditions, that is, $u_D|_{\Gamma_D} = g_D$; and Γ_N is the portion of $\partial\Omega$ on which natural (normal derivative) boundary conditions are imposed. Also, $a : H^1(\Omega) \times H^1(\Omega) \mapsto \mathbb{R}$ is a continuous bi-linear form (not necessarily symmetric), and $l^N : H^1(\Omega) \mapsto \mathbb{R}$ (respectively, $l^O : H^1(\Omega) \mapsto \mathbb{R}$) is a continuous linear (respectively, affine) functional. (Treatment of unbounded output functionals l^O is considered in [12].)

Finally, we introduce a bilinear symmetric coercive form $a^S : X \times X \mapsto \mathbb{R}$, defined as

$$a^S(v, w) = \frac{1}{2}(a(v, w) + a(w, v)), \quad \forall v, w \in X, \quad (6)$$

which induces the energy norm

$$\|v\|^2 = a^S(v, v), \quad \forall v \in X, \quad (7)$$

the “natural” metric in which to measure the finite element error.

2.2. Finite Element Spaces

Two triangulations of the computational domain Ω are considered: the working or design H -mesh, \mathcal{T}_H , consisting of K_H elements T_H ; and the “truth” h -mesh, \mathcal{T}_h , consisting of K_h elements T_h . We require that \mathcal{T}_h is a refinement of \mathcal{T}_H ; the geometric requirements on \mathcal{T}_H are discussed in the context of our refinement strategy.

To each of these meshes we associate regular piecewise-linear continuous finite element subspaces,

$$X_H = \{v \in X \mid v|_{T_H} \in \mathbf{P}_1(T_H), \forall T_H \in \mathcal{T}_H\}, \quad (8)$$

$$X_h = \{v \in X \mid v|_{T_h} \in \mathbf{P}_1(T_h), \forall T_h \in \mathcal{T}_h\}, \quad (9)$$

where $\mathbf{P}_1(T)$ denotes the space of linear polynomials over T . Similarly, we define $X_H^D = \{v + u_D \mid v \in X_H\}$, and $X_h^D = \{v + u_D \mid v \in X_h\}$. Note that, by construction, $X_H \subset X_h \subset X$, and $X_H^D \subset X_h^D \subset X^D$.

The algorithms to be presented require that our spaces and forms be expressed as sums of contributions over the H -elements T_H . Towards this end, we introduce the subdomain local spaces $Z_H(T_H)$ and $Z_h(T_H)$,

$$Z_H(T_H) = \mathbf{P}_1(T_H), \quad \forall T_H \in \mathcal{T}_H, \quad (10)$$

$$Z_h(T_H) = \{v|_{T_h} \in \mathbf{P}_1(T_h), \forall T_h \in \mathcal{R}_{T_H}\} \cap H^1(T_H), \quad \forall T_H \in \mathcal{T}_H, \quad (11)$$

where \mathcal{R}_{T_H} denotes the set of h -mesh elements contained in T_H . The global representation of $Z_H(T_H)$ and $Z_h(T_H)$ are the “broken” spaces \hat{X}_H and \hat{X}_h ,

$$\hat{X}_H = \{v \in L_2(\Omega) \mid v|_{T_H} \in Z_H(T_H), \forall T_H \in \mathcal{T}_H\}, \quad (12)$$

$$\hat{X}_h = \{v \in L_2(\Omega) \mid v|_{T_h} \in Z_h(T_H), \forall T_H \in \mathcal{T}_H\}, \quad (13)$$

where $L_2(\Omega)$ is the space of square-integrable functions.

Finally, the bilinear — $a(\cdot, \cdot)$, $a^S(\cdot, \cdot)$ — and linear — $l^N(\cdot)$, $l^O(\cdot)$ — forms are extended to accept discontinuous functions in the “broken” spaces by redefining these forms as a sum of H -element contributions. For instance, $a(\cdot, \cdot)$ is now written as

$$a(v, w) = \sum_{T_H \in \mathcal{T}_H} a_{T_H}(v|_{T_H}, w|_{T_H}), \quad \forall v, w \in \hat{X}_h; \quad (14)$$

similar expressions are obtained for a^S , l^N , and l^O .

2.3. Continuity Form

Let $\mathcal{E}(\mathcal{T}_H)$ (respectively, $\mathcal{E}(\mathcal{T}_h)$) denote the set of open edges in the triangulation \mathcal{T}_H (respectively, \mathcal{T}_h). We can then introduce a space of functions over the element edges $\gamma_H \in \mathcal{E}(\mathcal{T}_H)$,

$$Q_H = \{q|_{\gamma_H} \in P_1(\gamma_H), \forall \gamma_H \in \mathcal{E}(\mathcal{T}_H) \mid q|_{\Gamma_N} = 0\}; \quad (15)$$

analogously,

$$Q_h = \{q|_{\gamma_h} \in P_1(\gamma_h), \forall \gamma_h \in \mathcal{E}(\mathcal{T}_h) \cap \mathcal{E}(\mathcal{T}_H) \mid q|_{\Gamma_N} = 0\}. \quad (16)$$

It follows that $Q_H \subset Q_h \subset H^{-1/2}(\mathcal{E}(\mathcal{T}_H))$; the functions in these spaces can, of course, be discontinuous [9].

Next, we introduce the “jump” bilinear form $b : \hat{X}_h \times Q_h \mapsto \mathbb{R}$,

$$b(v, q) = \sum_{\gamma_H \in \mathcal{E}(\mathcal{T}_H)} \int_{\gamma_H} \mathcal{J}_{\gamma_H} v q|_{\gamma_H} ds, \quad (17)$$

where $\mathcal{J}_{\gamma_H} v$ is the *jump* in v across γ_H when γ_H is an interior edge, and the *trace* of v on γ_H when γ_H is on the boundary $\partial\Omega$. The form $b(\cdot, \cdot)$ can be used to enforce continuity on functions in \hat{X}_H and \hat{X}_h ; in particular,

$$X_H \equiv \{v \in \hat{X}_H \mid b(v, q) = 0, \forall q \in Q_H\}, \quad (18)$$

$$X_h \equiv \{v \in \hat{X}_h \mid b(v, q) = 0, \forall q \in Q_h\}. \quad (19)$$

Note $b(\cdot, \cdot)$ places no restriction on v on natural boundaries.

3. BOUND PROCEDURE

3.1. \mathcal{T}_H and \mathcal{T}_h Problems

The Galerkin finite element solution on the working mesh, $u_H \in X_H^D$, satisfies

$$a(w, u_H) = l^N(v), \quad \forall w \in X_H; \quad (20)$$

the associated output, s_H , is then computed as

$$s_H = l^O(u_H). \quad (21)$$

Analogously, the Galerkin finite element solution on the “truth” mesh, $u_h \in X_h^D$, satisfies

$$a(w, u_h) = l^N(v), \quad \forall w \in X_h; \quad (22)$$

the associated “truth” output, s_h , is then computed as

$$s_h = l^O(u_h). \quad (23)$$

In the smooth case, $|s - s_H| \sim \mathcal{O}(H^2)$, and $|s - s_h| \sim \mathcal{O}(h^2)$, where H and h are the diameters of \mathcal{T}_H and \mathcal{T}_h , respectively.

It will be assumed that the calculation of u_H , and hence s_H , can be performed at moderate cost. However, the h -mesh will typically be much finer than the H -mesh, in order to ensure that $|s - s_h|$ is suitably small; the computation of u_h and s_h in (22)–(23) will thus, most likely, be impractical. Our objective is, therefore, to devise a procedure that will yield sharp upper and lower bounds for s_h in an affordable manner.

3.2. Lagrangian Formulation

Noting that $u_h - u_H \in X_h$, equation (22) gives

$$a(u_h - u_H, u_h) - l^N(u_h - u_H) = 0, \quad (24)$$

which can be combined with (6) to obtain the following *energy equality*,

$$a^S(u_h - u_H, u_h - u_H) - l^N(u_h - u_H) + a(u_h - u_H, u_H) = 0. \quad (25)$$

Introducing the error $e_h \in X_h$ as the difference between the h -mesh and H -mesh approximations, $e_h = u_h - u_H$, and defining the residual

$$l^E(v) \equiv l^N(v) - a(v, u_H), \quad \forall v \in \hat{X}_h, \quad (26)$$

the energy equality can be rewritten as

$$a^S(e_h, e_h) - l^E(e_h) = 0. \quad (27)$$

Note that $a^S(e_h, e_h)$ is simply $\|e_h\|^2$.

We now introduce the set of functions $\mathcal{S} \subset \hat{X}_h$,

$$\mathcal{S} = \{v \in \hat{X}_h \mid a(w, u_H + v) = l^N(w), \forall w \in X_h; b(v, q) = 0, \forall q \in Q_h\}. \quad (28)$$

We observe that the second constraint enforces continuity and the homogeneous essential conditions, that is, $v \in X_h$, and the first constraint, from (22), then forces $v = e_h$. Therefore, the space \mathcal{S} consists of a single function, e_h . The following trivial minimization statement can then be written,

$$\pm s_h = \min_{v \in \mathcal{S}} \kappa(a^S(v, v) - l^E(v)) \pm l^O(u_H + v), \quad (29)$$

where $\kappa \in \mathbb{R}_+$ is a non-negative parameter that will be used later to optimize the computed bounds.

The constrained minimization problem (29) suggests the formation of the quadratic-linear Lagrangian $\mathcal{L} : \hat{X}_h \times X_h \times Q_h \mapsto \mathbb{R}$,

$$\mathcal{L}^\pm(v, \mu, q) = \kappa(a^S(v, v) - l^E(v)) \pm l^O(u_H + v) + a(\mu, u_H + v) - l^N(\mu) + b(v, q), \quad (30)$$

which allows us to express $\pm s_h$ as

$$\pm s_h = \inf_{v \in \hat{X}_h} \sup_{\mu \in X_h, q \in Q_h} \mathcal{L}^\pm(v, \mu, q). \quad (31)$$

We know that, at the saddlepoint $(e_h, \psi_h^\pm, p_h^\pm)$,

$$e_h = \arg \min_{v \in \hat{X}_h} \left[\sup_{\mu \in X_h, q \in Q_h} \mathcal{L}^\pm(v, \mu, q) \right], \quad (32)$$

$$(\psi_h^\pm, p_h^\pm) = \arg \max_{\mu \in X_h, q \in Q_h} \left[\inf_{v \in \hat{X}_h} \mathcal{L}^\pm(v, \mu, q) \right]. \quad (33)$$

We shall refer to μ as the adjoint (dual) variable, and q as the hybrid flux.

3.3. Duality

From quadratic-linear duality theory [17] and (31) it follows that

$$\pm s_h = \sup_{\mu \in X_h, q \in Q_h} \inf_{v \in \hat{X}_h} \mathcal{L}^\pm(v, \mu, q) \geq \inf_{v \in \hat{X}_h} \mathcal{L}^\pm(v, \bar{\mu}_h^\pm, \bar{q}_h^\pm), \quad \forall \bar{\mu}_h^\pm \in X_h, \forall \bar{q}_h^\pm \in Q_h, \quad (34)$$

with equality for $(\bar{\mu}_h^\pm, \bar{p}_h^\pm) = (\psi_h^\pm, p_h^\pm)$. Upper and lower bounds for the output s_h , s_{UP} and s_{LO} , respectively, are now readily constructed as

$$s_{UP} \equiv - \inf_{v \in \hat{X}_h} \mathcal{L}^-(v, \bar{\mu}_h^-, \bar{q}_h^-) \leq s_h \leq \inf_{v \in \hat{X}_h} \mathcal{L}^+(v, \bar{\mu}_h^+, \bar{q}_h^+) \equiv s_{LO}. \quad (35)$$

We note that (35) holds for any functions $\bar{\mu}_h^\pm \in X_h$ and $\bar{q}_h^\pm \in Q_h$. In particular, we will choose these functions to be members of $X_H \subset X_h$ and $Q_H \subset Q_h$.

3.4. Adjoint and Hybrid Flux Calculations

We consider here the problem of determining H -mesh approximations to the adjoint ψ_h^\pm , $\bar{\mu}_h^\pm$, and the hybrid flux p_h^\pm , \bar{q}_h^\pm , that will yield sharp bounds. Towards this end, we look for the saddlepoint of the Lagrangian (30) in the subspaces $\hat{X}_H \subset \hat{X}_h$, $X_H \subset X_h$, and $Q_H \subset Q_h$, in particular

$$(\psi_H^\pm, p_H^\pm) = \arg \max_{\mu \in X_H, q \in Q_H} [\min_{v \in \hat{X}_H} \mathcal{L}^\pm(v, \mu, q)], \quad (36)$$

and then set $\bar{\mu}_h^\pm = \psi_H^\pm$ and $\bar{q}_h^\pm = p_H^\pm$.

The stationarity conditions for (36) are obtained by requiring that the variations of \mathcal{L}^\pm with respect to v , μ , and q vanish: Find $e_H^\pm \in \hat{X}_H$, $\psi_H^\pm \in X_H$, and $p_H^\pm \in Q_H$ such that

$$b(w, p_H^\pm) = -\{\kappa (2a^S(w, e_H^\pm) - l^E(w)) \pm l^O(w) + a(\psi_H^\pm, w)\}, \quad \forall w \in \hat{X}_H, \quad (37)$$

$$a(w, u_H + e_H^\pm) = l^N(w), \quad \forall w \in X_H, \quad (38)$$

$$b(e_H^\pm, q) = 0, \quad \forall q \in Q_H. \quad (39)$$

First, equation (39) forces $e_H^\pm \in X_H$, which combined with (38) and (20) implies that $e_H^\pm = 0$. Next, equation (37) must be satisfied for all $w \in \hat{X}_H$, and thus, for all $w \in X_H \subset \hat{X}_H$. The adjoint, $\psi_H^\pm \in X_H$, can therefore be determined (precisely as in Aubin–Nitsche procedures [7]) as

$$a(\psi_H, w) = -l^O(w), \quad \forall w \in X_H, \quad (40)$$

with $\psi_H^\pm = \pm \psi_H$. Note that, in deriving (40), we exploit, first, $l^E(w) = 0$, $\forall w \in X_H$, which follows from (20), and second $b(w, p_H^\pm) = 0$, $\forall w \in X_H$, from (18).

Finally, equation (37) can now be evoked to determine the hybrid fluxes $p_H^\pm \in Q_H$,

$$b(w, p_H^\pm) = \kappa l^E(w) \mp l^O(w) - a(\psi_H^\pm, w), \quad \forall w \in \hat{X}_H. \quad (41)$$

Alternatively, we can find $p_{0H} \in Q_H$ and $p_{1H} \in Q_H$ such that

$$b(w, p_{0H}) = -l^O(w) - a(\psi_H, w), \quad \forall w \in \hat{X}_H, \quad (42)$$

$$b(w, p_{1H}) = l^E(w) (= l^N(w) - a(w, u_H)), \quad \forall w \in \hat{X}_H, \quad (43)$$

and then set $p_H^\pm = \pm p_{0H} + \kappa p_{1H}$.

Note that p_{0H} equilibrates the H -mesh adjoint (40), whereas p_{1H} equilibrates the original H -mesh solution (20).

Equations (42) and (43) represent a solvable but indeterminate system. The computation of an acceptable compatible solution is known as the *equilibration problem*. We follow here the approach proposed in [11,6,2], which involves the solution of an indeterminate system at each vertex of the \mathcal{T}_H grid, the size of which is given by the number of edges that meet at the vertex. Details about this procedure can be found elsewhere [15,12].

3.5. Local Problems

We consider here the solution of the minimization problems in (35). The minimizers of $\mathcal{L}^\pm(v, \psi_H^\pm, p_H^\pm)$, $\hat{e}_h^\pm \in \hat{X}_h$, will satisfy the following stationarity condition,

$$2\kappa a^S(w, \hat{e}_h^\pm) = \kappa l^E(w) \mp l^O(w) - a(\psi_H^\pm, w) - b(w, p_H^\pm), \quad \forall w \in \hat{X}_h, \quad (44)$$

which we recognize as $2K_H$ independent T_H -local symmetric Neumann (or Robin) sub-domain problems. Note that, although a^S is coercive over $X_h \times X_h$, it is typically only semi-definite over $\hat{X}_h \times \hat{X}_h$; however, thanks to (41), if the singular modes of \hat{X}_h are shared by \hat{X}_H , (44) will be solvable, and our bounds finite. The equations for \hat{e}_h^+ and \hat{e}_h^- can be rearranged into the following problems for \hat{e}_{0h} and \hat{e}_{1h} ,

$$2a^S(w, \hat{e}_{0h}) = -l^O(w) - a(\psi_H, w) - b(w, p_{0H}), \quad \forall w \in \hat{X}_h, \quad (45)$$

$$2a^S(w, \hat{e}_{1h}) = l^N(w) - a(w, u_H) - b(w, p_{1h}), \quad \forall w \in \hat{X}_h, \quad (46)$$

from which we construct $\hat{e}_h^\pm = \pm \frac{1}{\kappa} \hat{e}_{0h} + \hat{e}_{1h}$. Note that \hat{e}_{0h} and \hat{e}_{1h} are related to the error in the adjoint ψ_H , and the error in the solution u_H , respectively.

3.6. Bounds Calculation

To evaluate the bounds, it remains only to insert the computed values for \hat{e}_h^\pm , ψ_H^\pm , and p_H^\pm into our Lagrangian. In particular, evoking (44) with $w = \hat{e}_h^\pm$, and (20) with $w = \psi_H^\pm$, we obtain from (35)

$$s_{UP}(\mathcal{T}_H, \kappa) \equiv -\mathcal{L}^-(\hat{e}_h^-, \psi_H^-, p_H^-) = l^O(u_H) + \kappa a^S(\hat{e}_h^-, \hat{e}_h^-), \quad (47)$$

$$s_{LO}(\mathcal{T}_H, \kappa) \equiv \mathcal{L}^+(\hat{e}_h^+, \psi_H^+, p_H^+) = l^O(u_H) - \kappa a^S(\hat{e}_h^+, \hat{e}_h^+), \quad (48)$$

where the arguments of s_{UP} and s_{LO} indicate that the quality of the bounds depends on the working mesh \mathcal{T}_H and the parameter κ . Note that (47), (48) indicate that our bounds for s_h are also bounds for s_H . Note also that, for the ‘‘compliance’’ case — $a(\cdot, \cdot)$ symmetric, $l^O = l^N$, homogeneous essential conditions, and $\kappa = 1 - \psi^+ = -u_H$ and $p_H^+ = 0$; it follows that $\hat{e}_h^+ = 0$, and thus $s_{LO} = l^O(u_H)$.

We can also express our bounds in terms of \hat{e}_{0h} , \hat{e}_{1h} ,

$$s_{UP}(\mathcal{T}_H, \kappa) = l^O(u_H) - 2a^S(\hat{e}_{0h}, \hat{e}_{1h}) + \frac{1}{\kappa} a^S(\hat{e}_{0h}, \hat{e}_{0h}) + \kappa a^S(\hat{e}_{1h}, \hat{e}_{1h}), \quad (49)$$

$$s_{LO}(\mathcal{T}_H, \kappa) = l^O(u_H) - 2a^S(\hat{e}_{0h}, \hat{e}_{1h}) - \frac{1}{\kappa} a^S(\hat{e}_{0h}, \hat{e}_{0h}) - \kappa a^S(\hat{e}_{1h}, \hat{e}_{1h}), \quad (50)$$

which we recall is valid for any positive value of the parameter κ . The choice of κ will affect the (half) bound gap,

$$\Delta(\mathcal{T}_H, \kappa) \equiv \frac{1}{2}(s_{UP} - s_{LO}) = \frac{1}{\kappa} a^S(\hat{e}_{0h}, \hat{e}_{0h}) + \kappa a^S(\hat{e}_{1h}, \hat{e}_{1h}), \quad (51)$$

but will not affect the output predictor,

$$s_{pre}(\mathcal{T}_H) \equiv \frac{1}{2}(s_{UP} + s_{LO}) = l^O(u_H) - 2a^S(\hat{e}_{0h}, \hat{e}_{1h}). \quad (52)$$

(Recall from (2) that $|s_h - s_{pre}(\mathcal{T}_H)| \leq \Delta(\mathcal{T}_H)$.) Since \hat{e}_{0h} and \hat{e}_{1h} do not depend on the choice of κ , we can readily find that κ which minimizes the bound gap, and hence renders the lower and upper bounds as sharp as possible. This procedure gives

$$\kappa^* = \sqrt{\frac{a^S(\hat{e}_{0h}, \hat{e}_{0h})}{a^S(\hat{e}_{1h}, \hat{e}_{1h})}}, \quad (53)$$

and the optimized bounds

$$s_{UP}(\mathcal{T}_H, \kappa^*) \equiv s_{UP}(\mathcal{T}_H) = l^O(u_H) - 2a^S(\hat{e}_{0h}, \hat{e}_{1h}) + 2 \|\hat{e}_{0h}\| \|\hat{e}_{1h}\|, \quad (54)$$

$$s_{LO}(\mathcal{T}_H, \kappa^*) \equiv s_{LO}(\mathcal{T}_H) = l^O(u_H) - 2a^S(\hat{e}_{0h}, \hat{e}_{1h}) - 2 \|\hat{e}_{0h}\| \|\hat{e}_{1h}\|, \quad (55)$$

and associated bound gap

$$\Delta(\mathcal{T}_H, \kappa^*) \equiv \Delta(\mathcal{T}_H) = 2 \|\hat{e}_{0h}\| \|\hat{e}_{1h}\|. \quad (56)$$

(Note that, in the compliance case, $\kappa^* = 1$.) We thus see that our procedure is a *bound formulation* of the classical Aubin–Nitsche result; an implicit counterpart to existing explicit schemes for linear functional outputs [7,8]. In the smooth case we expect that $\|\hat{e}_{0h}\|$ and $\|\hat{e}_{1h}\|$ will each be $\mathcal{O}(H)$, and thus $\Delta(\mathcal{T}_H) \sim \mathcal{O}(H^2)$ will vanish at the same rate as $|s - s_H|$ as $H \rightarrow 0$.

For the special situation in which $\kappa = 1$ and $l^O(u_H + v) \equiv l^E(v)$, we see from (27) that our upper bound corresponds to an upper bound for the error in the energy norm, and thus in this case our method is a (rather trivial) extension of the method of [2] to nonsymmetric operators. (The lower bound is zero, consistent with our compliance result.) Note that, for both the symmetric and nonsymmetric problems, $\psi_H^- = 0$, however only for the symmetric problem is $\psi_h^- = 0$. It follows that only for the symmetric problem is e_h the minimizer of $\mathcal{L}^\pm(v, \psi_H^-, \cdot)$ ($= 2J(v)$ of [2] in this particular case) over all $v \in X_h$. This highlights the distinction between our *energy functional* [14], $a^S(v, v) - l^E(v) \pm l^O(u_H + v)$, and a variational principle: the former is only required to *evaluate* to $l^O(u_h)$ for $v = e_h$, while the latter must attain its *minimum* $l^O(u_h)$ at $v = e_h$. We are able to consider more general objective functions since the task of enforcing $v = e_h$ has been moved to a constraint.

Finally, we briefly summarize computational complexity. On \mathcal{T}_H , we require two global calculations, one for u_H and one for ψ_H , and two (local) hybrid flux equilibrations. On \mathcal{T}_h we need only invert $2K_H$ independent T_H -local *symmetric* Neumann operators. It follows from the superlinearity of most general solution algorithms that the effort to compute our bounds will be much less than the effort to directly compute u_h and s_h . Note that for each additional output of interest, we require only *one* additional global calculation on \mathcal{T}_H , and *one* additional equilibration, and only $1K_H$ additional T_H -local Neumann problems.

4. ADAPTIVITY

4.1. Local Indicators

The procedure presented to calculate bounds can be implemented on any pair of meshes, \mathcal{T}_H and \mathcal{T}_h , which satisfy the requirement that \mathcal{T}_h is a refinement of \mathcal{T}_H . Since the cost of computing the bounds is essentially a function of the number of elements K_H in \mathcal{T}_H , it is desirable to construct *optimized* triangulations that maximize the bound accuracy (minimize the bound gap) for a given number of degrees-of-freedom. In this section we present an algorithm for generating such optimized grids in an adaptive iterative manner.

We first note that the bound gap (56) can be expressed as a sum of elemental contributions,

$$\Delta(\mathcal{T}_H) = \sum_{T_H \in \mathcal{T}_H} \Delta_{T_H}(\mathcal{T}_H), \quad (57)$$

where

$$\Delta_{T_H}(\mathcal{T}_H) = \frac{1}{\kappa^*} a_{T_H}^S(\hat{e}_{0h}, \hat{e}_{0h}) + \kappa^* a_{T_H}^S(\hat{e}_{1h}, \hat{e}_{1h}) \quad (58)$$

$$= \sqrt{\frac{a^S(\hat{e}_{1h}, \hat{e}_{1h})}{a^S(\hat{e}_{0h}, \hat{e}_{0h})}} a_{T_H}^S(\hat{e}_{0h}, \hat{e}_{0h}) + \sqrt{\frac{a^S(\hat{e}_{0h}, \hat{e}_{0h})}{a^S(\hat{e}_{1h}, \hat{e}_{1h})}} a_{T_H}^S(\hat{e}_{1h}, \hat{e}_{1h}) \quad (59)$$

Note that $\Delta_{T_H}(\mathcal{T}_H)$ is non-negative and can thus be directly interpreted as the *contribution* to the bound gap from element T_H .

We can now describe our iterative adaptive strategy. Starting from an initial grid \mathcal{T}_H^0 with bound gap $\Delta(\mathcal{T}_H^0)$, we generate a sequence of triangulations $\{\mathcal{T}_H^k, k = 1, 2, \dots\}$, with corresponding bound gaps $\{\Delta(\mathcal{T}_H^k), k = 1, 2, \dots\}$, such that each triangulation \mathcal{T}_H^k is a refinement of the preceding triangulation \mathcal{T}_H^{k-1} . The approach does not guarantee that $\Delta(\mathcal{T}_H^k) \leq \Delta(\mathcal{T}_H^{k-1})$ for any particular k , but it does ensure that, for a sufficiently large k , $\Delta(\mathcal{T}_H^k) \leq \Delta^{arg}$, where $\Delta^{arg} > 0$ is a specified positive gap target.

In order to identify the elements in \mathcal{T}_H^{k-1} that need to be refined, we first calculate the largest elemental contribution Δ_{max}^{k-1} to the bound gap $\Delta(\mathcal{T}_H^{k-1})$,

$$\Delta_{max}^{k-1} = \max_{T_H \in \mathcal{T}_H^{k-1}} \Delta_{T_H}(\mathcal{T}_H^{k-1}), \quad (60)$$

and then select for refinement all those elements $T_H \in \mathcal{T}_H^{k-1}$ for which

$$\Delta_{T_H}(\mathcal{T}_H^{k-1}) \geq \alpha \Delta_{max}^{k-1}. \quad (61)$$

Here, $0 < \alpha < 1$ is a parameter which controls the fraction of elements to be refined at each iteration. At present, this parameter is specified *a priori*, and is independent of k . More sophisticated strategies for choosing α can, of course, be devised; for instance, we can determine α^k such that, based on *a priori* scalings, a given target gap $\Delta^{arg,k}$ should be satisfied by the refined grid \mathcal{T}_H^k .

The refinement process is terminated when a selected accuracy goal is achieved. In the applications that follow, the accuracy measure is the relative (half) bound gap θ^k ,

$$\theta^k = \frac{\Delta(\mathcal{T}_H^k)}{s_{pre}(\mathcal{T}_H^k)}. \quad (62)$$

The iteration is thus halted when $\theta^k \leq \theta^{obj}$ where θ^{obj} , is the prescribed accuracy.

4.2. Grid Refinement

The triangulation \mathcal{T}_H^k is generated from \mathcal{T}_H^{k-1} by refinement of the selected triangles (T_H such that $\Delta_{T_H}(\mathcal{T}_H^{k-1}) > \alpha \Delta_{max}^{k-1}$) triangles into *four* similar triangles. This refinement strategy leads to the generation of *hanging nodes* whenever only one of the two triangles sharing an edge is refined. Since an element T_H with a hanging node — a node on ∂T_H which is *not* a vertex of T_H — requires special computational treatment, we permit only one hanging node per element. We thus, first, refine all those elements satisfying equation (61), and then, in addition, recursively refine those elements that have more than one hanging node. At the conclusion of the process, only two types of elements (having either zero or one hanging node) are present. This approach has the benefit that fairly smooth grids are generated, and that only one type of *special* element needs to be developed.

We note that in unrefined elements the T_H -local operators required in (45), (46) are unchanged from iteration $k - 1$ to iteration k . This can be used (in conjunction with direct solution methods) to greatly reduce the cost of subsequent refinements.

4.3. Computational Treatment of Hanging Nodes

Consider a triangle with nodes IJK , with a hanging node M located at the midpoint of edge IJ . It follows from (8) that functions in the finite element space X_H^k associated with the triangulation \mathcal{T}_H^k must be continuous. This implies that any function v in X_H^k must vary linearly over the edge IJ , and thus the value of v at M , v_M , will be determined by the values of v at nodes I and J , v_I and v_J , respectively: $v_M = (v_I + v_J)/2$.

A convenient way to treat hanging nodes when solving equations (20) and (40) is to split the triangle IJK into two standard triangles, IMK and MJK , and to proceed in forming the equation system as if M is a standard node. The equation corresponding to the hanging node M can then be conveniently removed from the system by simple row and column operations: one half of row (respectively, column) M is added to rows (respectively, columns) I and J . Performing this operation for every hanging node produces a system that involves only the true unknowns at the standard nodes. Once the system has been solved, the value of the solution at the hanging nodes is easily recovered.

The next step is the solution of the equilibration problems (42) and (43). Here, we also follow the approach of splitting the triangle IJK into the two triangles, IMK and MJK . The equilibration problem, however, can not be carried out in the usual fashion, since the solution at points I , M , and J will generally not be in equilibrium. In fact, it can be shown that the effect of constraining v_M to be the average of v_I and v_J is equivalent to the application of a “force” $\lambda \in \mathbb{R}$ at nodes I and J , and a force $-\lambda$ at node M . The magnitude of λ — a Lagrange multiplier — can be determined from the load imbalance at node M after all elements sharing M have been considered. With the addition of the elemental forces to element IMK (respectively, MJK) — λ to node I (respectively, J) and $-\lambda$ to node M — the flux equilibration procedure (42) and (43) can then be carried out in its standard form.

Finally, the two elements IMK and MJK are combined into the original element IJK before the local Neumann problems (45) and (46) are solved. The element hybrid fluxes over the edge MK will cancel, and the equilibrating forces λ introduced to compute the hybrid fluxes can be removed; the equations (45) and (46) remain solvable, since the sum

and moment of the λ forces applied is zero. The only special treatment in this case is that the hybrid flux over the edge IJ is defined in terms of two piecewise linear functions.

5. RESULTS

Results are presented for the two-dimensional generalized planar stress linear elasticity equations. Although here the solution field u is the displacement vector, the theory presented for the scalar problem can be trivially extended to handle this case [15]. The particular problem considered is that of one-fourth of a symmetrically loaded plate with two square holes, as shown in Figure 1. The domain boundaries are denoted by Γ^j , $j = 1, \dots, 6$. A horizontal unit distributed force is applied along Γ^3 , and homogeneous essential boundary conditions are enforced on the normal displacement on Γ^1 , Γ^2 , and Γ^5 ; stress-free condition apply on all other boundaries and components. The Young's modulus and Poisson ratio are taken to be 1 and 0.3, respectively.

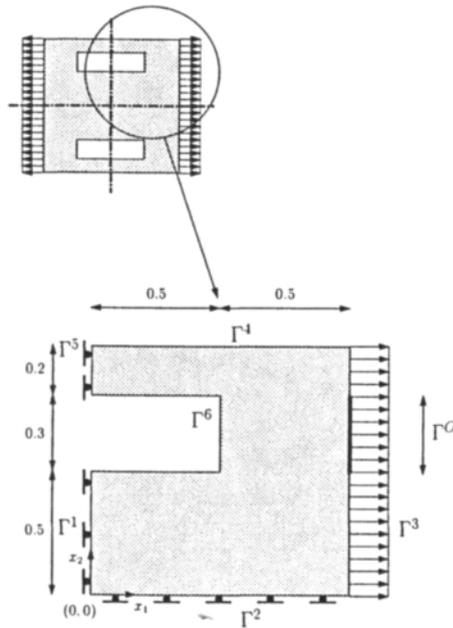


Figure 1. Geometry for linear elasticity example: one-fourth of a symmetrically loaded plate with two square holes. A normal traction is applied on Γ^3 ; the outputs of interest are the normal force on Γ^5 , and the average normal displacement over the segment $\Gamma^O \subset \Gamma^3$.

Two outputs are considered: the normal force on the boundary segment Γ^5 , denoted by s^1 , and the average horizontal displacement over the segment Γ^O , denoted by s^2 . we express the force output as a bounded linear functional [15],

$$l_1^O(v) = \int_{\Omega} \frac{\partial \chi_i}{\partial x_j} \sigma_{ij}(u) dA, \quad (63)$$

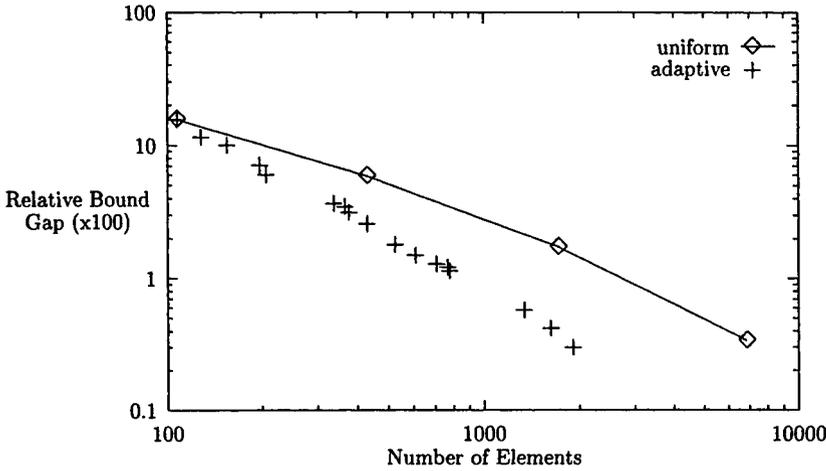


Figure 2. Bound gap $\Delta(\mathcal{T}_H)$ as a function of the number of elements K_H , for the force output s^1 , for meshes generated using uniform and adaptive refinement.

where summation over repeated indices is assumed ($i, j = 1, 2$), $\sigma_{ij}(u)$ is the stress tensor, and χ_i is a vector function defined as

$$\chi_1 = \begin{cases} 1 - 2x_1 & : (x_1, x_2) \in \Omega' \\ 0 & : (x_1, x_2) \in \Omega/\Omega' \end{cases}, \quad \chi_2 = 0. \tag{64}$$

Here $\Omega' = \{(x_1, x_2) \mid 0 \leq x_1 \leq 0.5, 0.8 \leq x_2 \leq 1.0\}$. Since there are no volume forces, or tractions applied at either Γ^4 or Γ^6 , (63) is equivalent, for sufficiently smooth functions, to the integral along Γ^5 of the horizontal component of the traction vector. The average normal displacement linear functional is simply written as

$$l_2^O(v) = \frac{1}{|\Gamma^O|} \int_{\Gamma^O} v_1 ds, \tag{65}$$

which is bounded for all Γ^O of non-zero measure.

The same initial triangulation \mathcal{T}_H^0 and “truth” mesh \mathcal{T}_h have been considered for both outputs. The initial triangulation consists of 108 triangles. The “truth” mesh has been generated from this initial triangulation by four refinement interactions in which *all* triangles are subdivided; thus, the number of elements in the “truth” mesh is $27,648 = 108 \times 4^4$. In the adaptive process, elements were selected for refinement based on equation (61) with $\alpha = 0.3$ for all the iterations. The accuracy goal, θ^{obj} , for the relative (half) bound gap, (62), is 0.003 (or 0.3%).

For the force output s^1 , 16 refinement iterations are required starting from a relative bound gap of 15.64%; only 12 refinement iterations are required for the displacement

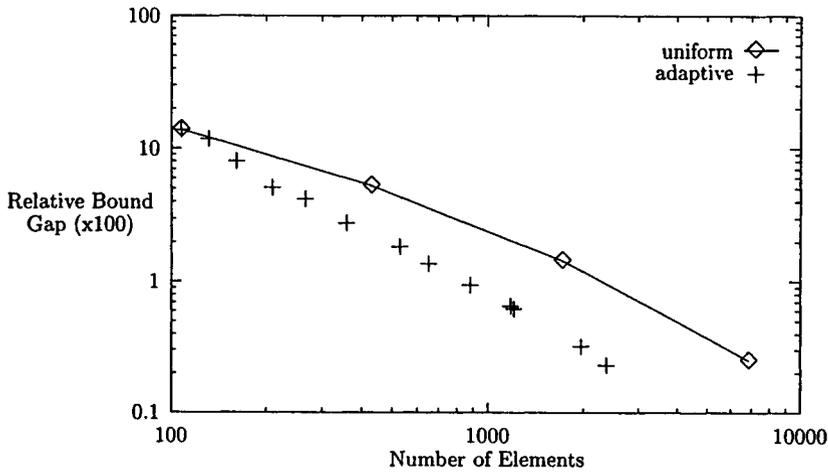


Figure 3. Bound gap $\Delta(\mathcal{T}_H)$ as a function of the number of elements K_H , for the displacement output s^2 , for meshes generated using uniform and adaptive refinement.

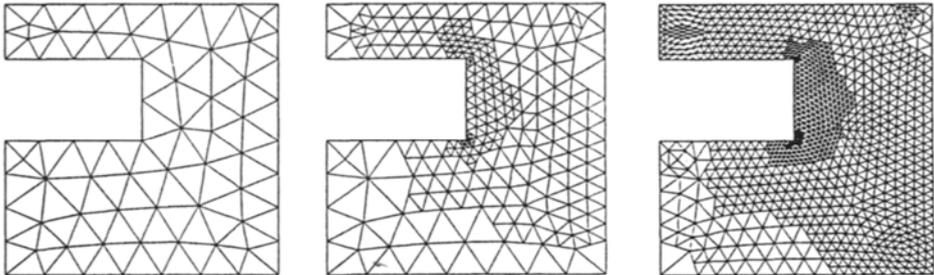


Figure 4. Adaptively generated meshes for normal force output s^1 . Figure shows \mathcal{T}_H^0 , \mathcal{T}_H^8 , and \mathcal{T}_H^{16} , with $K_H^0 = 108$, $K_H^8 = 432$, and $K_H^{16} = 1917$ elements; predicted outputs are $s_{pre}^{1,0} = -0.3198$, $s_{pre}^{1,8} = -0.3215$, and $s_{pre}^{1,16} = -0.3220$; and relative bound gaps are $\theta^{1,0} = 15.64\%$, $\theta^{1,8} = 2.59\%$, and $\theta^{1,16} = 0.30\%$.

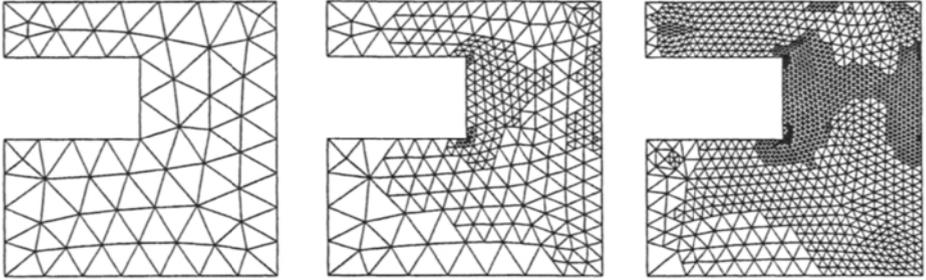


Figure 5. Adaptively generated meshes for the displacement output s^2 . Figure shows \mathcal{T}_H^0 , \mathcal{T}_H^6 , and \mathcal{T}_H^{12} , with $K_H^0 = 108$, $K_H^6 = 531$, and $K_H^{12} = 2382$ elements; predicted outputs are $s_{pre}^{2,0} = 0.4056$, $s_{pre}^{2,6} = 0.4204$, and $s_{pre}^{2,12} = 0.4226$; and relative bound gaps are $\theta^{2,0} = 13.82\%$, $\theta^{2,6} = 1.83\%$, and $\theta^{2,12} = 0.23\%$.

output s^2 , for which the initial relative bound gap is 13.82%. Figures 2 and 3 show the relative bound gap as a function of the number of elements, at each adaptive iteration, for the force and displacement outputs, respectively. We also show the bound gap computed for a sequence of grids obtained by uniform refinement, the finest of which still has four times fewer elements than the “truth” mesh \mathcal{T}_h . For both outputs we observe that, the uniformly refined meshes require approximately four times as many elements as the adaptive meshes for comparable levels of accuracy.

We note that, say for the force output, $|(s_{pre}^{1,16} - s_h)/s_{pre}^{1,16}| = 0.0001$, whereas $\theta^{1,16} = 0.003$, and thus the effectivity of the estimator in the classical sense is not as good as we may hope, even though the convergence rate of $\Delta(\mathcal{T}_H)$ is optimal. We emphasize, however, that our bounds provide *certainty* that $|(s_{pre}^{1,16} - s_h)/s_{pre}^{1,16}| \leq 0.003$ — and more importantly, that s_h lies between $s_{LO}^{1,16} = -0.3230$ and $s_{UP}^{1,16} = -0.3211$ — without calculating s_h on the uniformly highly refined (and hence expensive) “truth” mesh \mathcal{T}_h . Future work will address techniques by which to reduce the ratio $\Delta(\mathcal{T}_H)/|s_h - s_{pre}|$ as $H \rightarrow 0$; this, in turn, will permit to achieve θ^{obj} on coarser meshes \mathcal{T}_H .

Finally, in Figures 4 and 5, we show the initial, an intermediate, and the final adaptive triangulations obtained for the force and displacement outputs, respectively. Refinement around the singular re-entrant corners is apparent in both cases but, in addition, in the displacement output case, further refinement is observed near the points where l_2^Q is “discontinuous”.

In conclusion, we have shown that the goals of *validated outputs* and *optimal mesh refinement* can be simultaneously satisfied within our bounds framework. Clearly much work is required to translate these “end results” into computationally effective algorithms; our focus in the current paper is on the development of the appropriate foundation.

6. ACKNOWLEDGEMENTS

Support for this work has been provided by DARPA and ONR under Grant N00014-91-J-1889, the AFOSR under Grant F49620-97-1-0052, and NASA under Grants NAG 1-1613, NAG 4-105, NAG 1-1587, and NCC 3-483.

REFERENCES

1. R.A. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.
2. M. Ainsworth and J.T. Oden, *A unified approach to a posteriori error estimation using element residual methods*, Numer. Math., **65** (1993), pp. 23-50.
3. M. Ainsworth and J.T. Oden, *A posteriori error estimation in finite element analysis*, TICAM Report 96-19, 1996.
4. I. Babuška and W.C. Rheinboldt, *Analysis of optimal finite element meshes in \mathbb{R}^1* , Math. Comp., **33** (1979), pp. 435-463.
5. I. Babuška, T. Strouboulis, C.S. Upadhyay, and S.K. Gangaraj, *A posteriori estimation and adaptive control of the pollution error in the h -version of the finite element method*, Int. J. Numer. Methods Engrg., **38** (1995), pp. 4207-4235.
6. R.E. Bank and A. Weiser, *Some a posteriori error estimators for elliptic partial differential equations*, Math. Comp., **44**:170 (1985), pp. 283-301.
7. R. Becker and R. Rannacher, *Weighted a posteriori error control in finite element methods*, IWR Preprint 96-1 (SFB 359), Heidelberg, 1996.
8. R. Becker and R. Rannacher, *A feedback approach to error control in finite element methods: basic analysis and examples*, IWR Preprint 96-52 (SFB 359), Heidelberg, 1996.
9. F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, New York, 1991.
10. K. Eriksson and C. Johnson, *An adaptive finite element method for linear elliptic problems*, Math. Comp., **50** (1988), pp. 361-383.
11. P. Ladeveze and D. Leguillon, *Error estimation procedures in the finite element method and applications*, SIAM J. Numer. Anal., **20** (1983), pp. 485-509.
12. M. Paraschivoiu, *A Posteriori Finite Element Bounds for Linear-Functional Outputs of Coercive Partial Differential Equations and of the Stokes Problem*, Ph.D. Thesis, Department of Mechanical Engineering, M.I.T., October, 1997.
13. M. Paraschivoiu and A.T. Patera, *A hierarchical duality approach to bounds for the outputs of partial differential equations*, Comp. Meth. Appl. Mech. Engrg., to appear.
14. M. Paraschivoiu, J. Peraire, Y. Maday, and A.T. Patera, *Fast bounds for Outputs of Partial Differential Equations*, in Proceedings Workshop on Optimal Design and Control, Washington, D.C., ed. J. A. Burns, E. M. Cliff, and B. Grossman, to appear.
15. M. Paraschivoiu, J. Peraire, and A.T. Patera, *A posteriori finite element bounds for linear-functional outputs of elliptic partial differential equations*, Comp. Meth. Appl. Mech. Engrg., to appear.
16. R. Rannacher and F.-T. Suttmeier, *A feedback approach to error control in finite element methods: application to linear elasticity*, Preprint 96-42 (SFB 359), Heidelberg, 1996.

17. G. Strang, *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, Wellesley, Massachusetts, 1986.
18. R. Verfürth, *A posteriori error estimation and adaptive mesh-refinement techniques*, *J. Comp. Appl. Math.*, **50** (1994), pp. 67-83.

On adaptivity and error criteria for meshfree methods

Ted Belytschko^a, Wing Kam Liu^b and Mike Singer^c

^a Walter Murphy Professor of Computational Mechanics, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, U.S.A.

^b Professor of Mechanical Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, U.S.A.

^c Graduate Student, Theoretical and Applied Mechanics, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, U.S.A.

1. INTRODUCTION

Much of theory of error indicators and error estimators which has been developed in the context of finite elements is applicable to all Galerkin methods. Consequently, the methodologies apply directly to meshfree methods based on Galerkin methods. Among the tools which can be used directly are: relationships between residuals and local errors, concepts such as pollution error, and the fundamental techniques for error extraction.

Meshfree methods offer several unique opportunities in the implementation of adaptivity:

1. it is easy to add nodes to a model, since a corresponding element structure need not be developed;
2. error indicators of the residual type can be used more easily since the displacements can easily be constructed to be C_1 , so that one can compute useful residuals at points;
3. meshfree approximations are closely related to wavelets, so some useful techniques and knowledge from that field can be exploited.

The first point is easily appreciated, but there are some provisos on this statement. While it is simple to add nodes in a meshfree model, for effectiveness, the procedure must usually be more complicated because simply inserting nodes between existing nodes degrades the quality of the resulting approximation. For example, in a one dimensional problem, if two nodes are simply added in a high gradient region, the error is decreased very little due to the poor behavior of the approximation. Instead, the nodes should usually be graded somewhat around the area where nodal insertion is required for extra resolution.

There is a rich theory in error indicators of the residual type and their application in finite elements has been hampered by the presence of residuals of two types:

1. residuals interior to the elements, which can be evaluated directly from the partial differential equation;
2. residuals on element interfaces, which consist of the violation of the required jump conditions.

Combining these two types of residuals often involves the introduction of extra scaling factors. Moreover, it becomes difficult to define the residual, and consequently the error, easily in a pointwise manner, which is desirable in certain implementations of adaptivity as the new model is

being constructed.

For meshfree solutions, these error indicators are much more attractive because their higher continuity is markedly different than finite element solutions, particularly with low order elements. Low order finite element solutions exhibit jumps in the derivatives, such as strains and stresses in elastostatics, in areas of insufficient resolution. On the other hand, in meshfree solutions, in areas of insufficient resolution the derivatives exhibit wiggles or oscillations. Thus jumps in finite element solutions turn to wiggles in meshfree solutions.

Another interesting aspect of meshfree methods is their relationship to wavelet theory, which provides a large horizon of new methods for error estimation. In Section 4, we will demonstrate this relationship to wavelets and how it can be applied to error indicators.

Meshfree methods originated with Smoothed Particle Hydrodynamics (SPH) developed by Lucy (1977), Gingold and Monaghan (1977), and Monaghan (1988). With moving least square interpolants, Belytschko et al.(1994) and Lu et al.(1994) developed the Element Free Galerkin (EFG) method and successfully simulated the static and dynamic crack problems. Other work includes the hp-Cloud method by Duarte and Oden (1995), Partition of Unity FEM by Babuska and Melenk (1995), Free-Mesh method by Yagawa (1995) and Finite Point method by Oñate et al.(1995); these are all examples of meshfree methods that do not need meshes or grids in their formulation. Therefore, these methods can all be grouped in the category of mesh (or grid) free particle methods.

Reproducing Kernel Particle Method (RKPM), developed by Liu et al.(1993, 1995c, 1995d), is similar to SPH in the sense that it begins the formulation from the kernel estimate of a function. The multiple scale Reproducing Kernel Particle combines multiresolution analysis and wavelet theory with a meshless method. However, RKPM modifies the kernel function by introducing a correction function in order to enhance its accuracy near or on the boundaries of the problem domain. This modified kernel function employing the correction function was first suggested by Liu et al. (1993, 1995d) in both continuous and discrete forms. Due to this correction function, RKPM's kernel function satisfies the consistency conditions (Liu et al.(1995d)).

2. KERNEL REPRESENTATION

2.1. Meshfree Approximation Functions

We consider the approximation of a single function $u(x)$ in a domain Ω . The domain Ω is assumed to be described by the usual methods of computational geometry. Within the domain, a set of nodes x_i , $i=1$ to n_N , is constructed and the parameter associated with the approximation at node i is denoted by u_i .

A rationale for the first meshfree methods was provided by invoking the notion of a kernel approximation for $u(x)$ on a domain Ω is generated by

$$u^R(x) = \int_{\Omega} u(y)\phi(x-y, h)d\Omega_y \quad (2.1)$$

where $u^R(x)$ is the approximation, $\phi(x-y, h)$ is a kernel or weight function, and h is a measure of the size of the support; in the SPH literature, it is often called a smoothing function. For purposes of developing approximations, discrete analogs of (2.1) are needed. The discrete form of (2.1) is obtained by numerical quadrature of the right-hand side.

2.2. Reproducing Conditions

The derivation of a meshfree discretization begins with the approximation of a function $u(x)$ in a domain Ω by

$$u^R(x) = \int_{\Omega} u(\bar{x}) \bar{\phi}_a(x - \bar{x}) d\bar{x} \quad (2.2a)$$

From this convolution, the given function $u(x)$ is reproduced up to any degree desired by choosing a suitable window function $\bar{\phi}_a(x)$. The Taylor series expansion of u about x is given as

$$u(\bar{x}) = u(x) - (x - \bar{x})u'(x) + \frac{1}{2!}(x - \bar{x})^2 u''(x) + \dots + \frac{(-1)^n}{n!}(x - \bar{x})^n u^{(n)}(x) \\ + \frac{(-1)^{n+1}}{(n+1)!}(x - \bar{x})^{n+1} u^{(n+1)}(x + \xi(\bar{x} - x)) \quad \xi \in (0, 1) \quad (2.2b)$$

Substituting equation (2.2b) into equation (2.2a) yields

$$u^R(x) = u(x) \int_{\Omega} \bar{\phi}_a(x - \bar{x}) d\bar{x} - u'(x) \int_{\Omega} (x - \bar{x}) \bar{\phi}_a(x - \bar{x}) d\bar{x} + \frac{u''(x)}{2!} \int_{\Omega} (x - \bar{x})^2 \bar{\phi}_a(x - \bar{x}) d\bar{x} \\ + \dots + \frac{(-1)^n}{n!} u^{(n)}(x) \int_{\Omega} (x - \bar{x})^n \bar{\phi}_a(x - \bar{x}) d\bar{x} \\ + \frac{(-1)^{n+1}}{(n+1)!} u^{(n+1)}(x + \xi(\bar{x} - x)) \int_{\Omega} (x - \bar{x})^{n+1} \bar{\phi}_a(x - \bar{x}) d\bar{x} \quad (2.3)$$

The integrals in equation (2.3) can be used to define the moments of the window function as

$$\bar{m}_k(a, x) = \int_{\Omega} (x - \bar{x})^k \bar{\phi}_a(x - \bar{x}) d\bar{x} \quad k = 0, \dots, n \quad (2.4)$$

Rewriting equation (2.3) using equation (2.4) gives

$$u^R(x) = u(x) \bar{m}_0(a, x) - u'(x) \bar{m}_1(a, x) + \frac{u''(x)}{2!} \bar{m}_2(a, x) + \dots \\ + \frac{(-1)^n}{n!} u^{(n)}(x) \bar{m}_n(a, x) + \frac{(-1)^{n+1}}{(n+1)!} u^{(n+1)}(x + \xi(\bar{x} - x)) \bar{m}_{n+1}(a, x) \quad (2.5)$$

In order to reproduce the function up to degree n , the following *reproducing conditions* need to be satisfied [Liu et al. (1996a), (1996b)]:

$$\bar{m}_k(a, x) = \delta_{k0} \quad k = 0, 1, 2, \dots, n \quad (2.6)$$

The window function $\bar{\phi}_a(x - \bar{x})$ can be expressed as

$$\bar{\phi}_a(x - \bar{x}) = \sum_{k=0}^n b_k(a, x) (x - \bar{x})^k \phi_a(x - \bar{x}) \\ = \mathbf{P}^T(x - \bar{x}) \mathbf{b}(a, x) \phi_a(x - \bar{x})$$

$$= C_a(x; x - \bar{x})\phi_a(x - \bar{x}) \quad (2.7)$$

with

$$P(x - \bar{x}) = \begin{bmatrix} 1 \\ (x - \bar{x}) \\ \vdots \\ (x - \bar{x})^n \end{bmatrix} \quad \text{and} \quad \mathbf{b}(a, x) = \begin{bmatrix} b_0(a, x) \\ b_1(a, x) \\ \vdots \\ b_n(a, x) \end{bmatrix} \quad (2.8)$$

The function $\bar{\phi}_a(x)$ is an arbitrarily chosen window function which does not necessarily satisfy the reproducing conditions. The scalar product of polynomial base vector $P^T(x - \bar{x})$ and $\mathbf{b}(a, x)$ is called the correction function $C_a(x; x - \bar{x})$ for the window function. The coefficient $\mathbf{b}(a, x)$ can be obtained by using the reproducing condition. Substituting equation (2.7) into the definition of the moments of window function $\bar{\phi}_a(x)$ yields

$$\begin{aligned} \bar{m}_k(a, x) &= \int_{\Omega} (x - \bar{x})^k \left[\sum_{k=0}^n b_k(a, x) (x - \bar{x})^k \phi_a(x - \bar{x}) \right] d\bar{x} \\ &= b_0(a, x)m_k(a, x) + b_1(a, x)m_{k+1}(a, x) + \dots + b_n(a, x)m_{k+n}(a, x) \end{aligned} \quad (2.9)$$

where $m_k(a, x)$ is the k-th moment of the original window function $\phi_a(x - \bar{x})$. Equation (2.9) can also be written in matrix form as

$$\begin{bmatrix} \bar{m}_0(a, x) \\ \bar{m}_1(a, x) \\ \vdots \\ \bar{m}_n(a, x) \end{bmatrix} = \begin{bmatrix} m_0(a, x) & m_1(a, x) & \cdots & m_n(a, x) \\ m_1(a, x) & m_2(a, x) & & m_{n+1}(a, x) \\ \vdots & \vdots & \ddots & \vdots \\ m_n(a, x) & m_{n+1}(a, x) & \cdots & m_{2n}(a, x) \end{bmatrix} \begin{bmatrix} b_0(a, x) \\ b_1(a, x) \\ \vdots \\ b_n(a, x) \end{bmatrix}$$

or

$$\bar{\mathbf{m}}(a, x) = \mathbf{M}(a, x)\mathbf{b}(a, x) \quad (2.10)$$

where $\mathbf{M}(a, x)$ is called the moment matrix and can be written as

$$\mathbf{M}(a, x) = \int_{\Omega} P(x - \bar{x})\phi_a(x - \bar{x})P^T(x - \bar{x})d\bar{x} \quad (2.11)$$

Equation (2.6) which gives the reproducing conditions is rewritten in vector form as

$$\begin{bmatrix} \bar{m}_0(a, x) \\ \bar{m}_1(a, x) \\ \vdots \\ \bar{m}_n(a, x) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{P}(0) \quad (2.12)$$

Combining equation (2.10) with equation (2.12) immediately leads to

$$\mathbf{M}(a, x)\mathbf{b}(a, x) = \mathbf{P}(0) \quad (2.13)$$

The coefficients $\mathbf{b}_k(a, x)$ can then be obtained by solving equation (2.13) to yield

$$\mathbf{b}(a, x) = \mathbf{M}^{-1}(a, x)\mathbf{P}(0) \quad (2.14)$$

2.3. Discrete Form of the Reproducing Conditions

In a computation, discretization is inevitable so the continuous form has to be discretized into a set of points or particles. Recall equation (2.2a):

$$u^R(x) = u(x) * \bar{\phi}_a(x) = \int_{\Omega} u(\bar{x})\bar{\phi}_a(x - \bar{x})d\bar{x} \quad (2.15)$$

On an arbitrary particle distribution, a discrete convolution can be defined by

$$u^h(x) = \sum_{I=1}^{NP} u(x_I)\bar{\phi}_a(x - x_I)\Delta V \quad (2.16)$$

where ΔV is the nodal volume distributed to each node or particle. The number of particles in the discrete system is denoted by NP . The window function, $\bar{\phi}_a(x - x_I)$, normalized with respect to the dilation parameter, is

$$\bar{\phi}_a(x - x_I) = \frac{1}{a}\bar{\phi}\left(\frac{x - x_I}{a}\right) \quad (2.17)$$

Then, the discrete moment of the window function is

$$\bar{m}_k(a, x) = \sum_{I=1}^{NP} (x - x_I)^k \bar{\phi}_a(x - x_I)\Delta x_I, \quad k = 0, 1, \dots, n \quad (2.18)$$

The window function $\bar{\phi}_a(x - x_I)$ can be expressed as the product of the correction function $C_a(x; x - x_I)$ and the originally chosen window function $\phi_a(x - x_I)$ as

$$\bar{\phi}_a(x - x_I) = C_a(x; x - x_I)\phi_a(x - x_I) \quad (2.19)$$

Here, we define the shape function as

$$N_I(x) = C_a(x; x - x_I)\phi_a(x - x_I)\Delta x_I \quad (2.20)$$

Thus, the interpolation equation can be written as

$$u^R(x) = \sum_{I=1}^{NP} N_I(x)u(x_I) \equiv \sum_{I=1}^{NP} N_I(x)u_I \quad (2.21)$$

3. MULTIREOLUTION KPM

From the analysis of Fourier transformations, it is noted that the RKPM kernel function can be regarded as a low-pass filter in the reconstruction procedure (Liu and Chen (1995) and Liu et al.(1996b)). The multiresolution RKPM is defined by a family of those kernel functions. The wavelet which corresponds to each level defined by those kernel functions is

$$\psi_{m+1}(x) = \phi_m(x) - \phi_{m+1}(x) \tag{3.1}$$

Therefore, a multiple scale decomposition of any response can be written as

$$\begin{aligned} u^h(x) &= u_o(x) && \text{finest scale} \\ &= u_1(x) + w_1(x) && \text{two-level decomposition} \\ &= u_2(x) + w_1(x) + w_2(x) && \text{three-level decomposition} \\ &\vdots && \vdots \\ &= u_m(x) + \sum_{i=1}^m w_i(x) && (m+1)\text{-level decomposition} \end{aligned} \tag{3.2}$$

where

$$u_m(x) = \int_{\Omega} C(2^m a, x, \bar{x}) \phi_m(x - \bar{x}) u(\bar{x}) d\bar{V} \tag{3.3}$$

$$w_i(x) = \int_{\Omega} \psi_i(x - \bar{x}) u(\bar{x}) d\bar{V} \tag{3.4}$$

with

$$\psi_i(x - \bar{x}) = C(2^{i-1} a, x, \bar{x}) \phi_{i-1}(x - \bar{x}) - C(2^i a, x, \bar{x}) \phi_i(x - \bar{x}) \tag{3.5}$$

where C is the correction function and a is the dilation parameter given in a previous section. For applications, the integrals in equations (3.3)-(3.4) are converted to their discrete forms. The discrete counterparts of the above equations (3.3)-(3.5) are written as

$$u_m(x) = \sum_{j=1}^{NP} N_j(x, x_j; 2^m a) u_j \tag{3.6}$$

$$w_i(x) = \sum_{j=1}^{NP} [N_j(x, x_j; 2^{i-1} a) - N_j(x, x_j; 2^i a)] u_j \tag{3.7}$$

in which N_j is the shape function of RKPM. The two dimensional two-level decomposition of a function $u^h(x)$ in discrete form is given by

$$\begin{aligned}
u^h(x) &= \sum_{j=1}^{NP} N_j(x; a) \mu_j \\
&= \sum_{j=1}^{NP} N_j(x; 2a) \mu_j + \sum_{j=1}^{NP} [N_j(x; a) - N_j(x; 2a)] \mu_j \\
&= \sum_{j=1}^{NP} N_j^l(x) \mu_j + \sum_{j=1}^{NP} N_j^h(x) \mu_j \\
&= u_l(x) + w_l(x)
\end{aligned} \tag{3.8}$$

The shape function is decomposed into a low scale (or scaling function) component, $N_j^l(x)$, and a high scale (or wavelet) component, $N_j^h(x)$.

According to the multi-decomposition procedures of RKPM shape functions, a two-level decomposition of the response of a mechanical system can be conducted. If the solution of a mechanical system is given, the low scale and high scale components are obtained by the decomposition of equation (3.8). A review and other literature about multiresolution analysis and multiple scale RKPM is given in Liu et al. (1995a, 1996a, 1996b, 1995b, 1996c, 1996d, 1997a).

4. ERROR ESTIMATE BASED ON RESIDUAL

The projected solutions of RKM in the a -scale and the $2a$ -scale are given as

$$u^{R_a}(x) = u(x) + \frac{(-1)^{n+1}}{(n+1)!} u^{(n+1)}(x) \left[i^{n+1} a^{n+1} \hat{\phi}^{(n+1)}(0) \right] + h.o.t.^a \tag{4.1}$$

$$u^{R_{2a}}(x) = u(x) + \frac{(-1)^{n+1}}{(n+1)!} u^{(n+1)}(x) \left[i^{n+1} (2a)^{n+1} \hat{\phi}^{(n+1)}(0) \right] + h.o.t.^{2a} \tag{4.2}$$

The RKM wavelet solution in $2a$ -scale is in the form of

$$w^{R_{2a}}(x) = (2^{n+1} - 1) \frac{(-1)^{n+1}}{(n+1)!} u^{(n+1)}(x) \left[i^{n+1} a^{n+1} \hat{\phi}^{(n+1)}(0) \right] + (h.o.t.^{2a} - h.o.t.^a) \tag{4.3}$$

Neglecting the h.o.t.'s, the wavelet solution is proportional to the error term

$$w^{R_{2a}}(x) \approx (2^{n+1} - 1) error(x) \tag{4.4}$$

Eq. (4.4) reveals a possibility that the wavelet solution can be an index for the error estimate of the computed solution.

5. ADAPTIVITY BASED ON LOCALIZED SHEAR DEFORMATION

It has been shown, in Liu et al.(1996) and references therein, that the multiple scale RKPM performs well for high gradient linear problems and its associated adaptivity. Some large deformation problems, such as shear band problems, also involve high gradient properties. In the following, the measure of shear deformation for multiple scale analysis is suggested. Additionally, the validity of the measure is demonstrated by detecting the location of high gradients.

A formula is proposed here for the high scale component of the second invariant I_2 of the Cauchy-Green deformation tensor. The second invariant of the left or the right deformation tensor is given as

$$I_2 = \frac{1}{2}[(trG)^2 - tr(G)^2] = \frac{1}{2}[(trC)^2 - tr(C)^2] \quad (5.1)$$

where the left G and right C Cauchy-Green deformation tensors are defined as

$$G = F \cdot F^T \quad \text{and} \quad C = F^T \cdot F \quad (5.2)$$

For plane strain problems, the deformation gradient F is defined as

$$F = \begin{bmatrix} \frac{\partial x_i}{\partial X_j} \end{bmatrix} = \begin{bmatrix} 1 + \frac{\partial u_1}{\partial X_1} & \frac{\partial u_1}{\partial X_2} \\ \frac{\partial u_2}{\partial X_1} & 1 + \frac{\partial u_2}{\partial X_2} \end{bmatrix} \quad (5.3)$$

The high scale component of I_2 can be obtained by inserting $u_i = v_i + w_i$, into equation (5.3) to obtain

$$F = \begin{bmatrix} 1 + \frac{\partial v_1}{\partial X_1} + \frac{\partial w_1}{\partial X_1} & \frac{\partial v_1}{\partial X_2} + \frac{\partial w_1}{\partial X_2} \\ \frac{\partial v_2}{\partial X_1} + \frac{\partial w_2}{\partial X_1} & 1 + \frac{\partial v_2}{\partial X_2} + \frac{\partial w_2}{\partial X_2} \end{bmatrix} \quad (5.4)$$

It is noted that no extra degree of freedom is needed in this multiple scale analysis and the components of v_i and w_i are defined according to Eq. (3.8) using the corresponding scaling and wavelet window functions. Next, if equation (5.4) is substituted into equations (5.2) and (5.3), the second invariant of the Cauchy-Green deformation strain tensors may be written as a sum of its low and high scale terms

$$I_2 = I_2^{low} + I_2^{high} \quad (5.5)$$

The high scale component is defined by grouping the terms involving only the high scale term of displacement, w_i , and is written as

$$I_2^{high} = \left[\left(1 + \frac{\partial w_1}{\partial X_1} \right) \left(1 + \frac{\partial w_2}{\partial X_2} \right) - \frac{\partial w_1}{\partial X_2} \frac{\partial w_2}{\partial X_1} \right]^2 - 1 \quad (5.6)$$

The terms collected, with the omission of the high scale components obtained above, are defined as the low scale components of the second invariant of the Cauchy-Green deformation tensor.

In order to apply the multiple scale method to large deformation problems, the explicit RKPM formulation, based on the reference configuration as presented by Jun and Liu (1996), is employed. Here, an adaptive procedure based on the high scale component of I_2 is presented. The example for this model is a plane strain compression problem. The initial shape of the material is given in Figure 1 (a). The upper boundary is moving down with a constant velocity $V_0 = 5 \times 10^{-1}$ (in/sec) while the bottom surface is fixed. For convenience, a constant time step of 2.5×10^{-7} (sec) is used. The size of the material is 1.0 (in) x 2.0 (in) and the total number of particles is 961. The upper and lower boundaries are perfectly bonded, which generates severe shear deformation in the material. The deformed shape at $t = 3.5 \times 10^{-6}$ (sec) is illustrated in Figure 1 (b) and the contours of the high scale components of I_2 are plotted in Figure 1 (c).

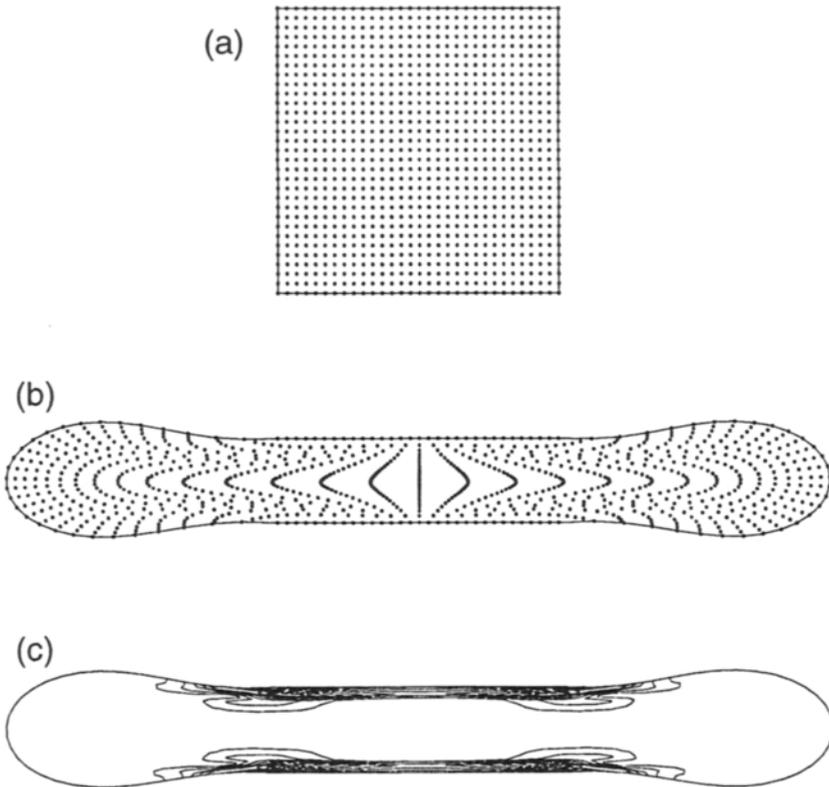


Figure 1: Compression problem for adaptive procedures; (a) initial shape, (b) deformed shape, (c) contours of high scale component of I_2 on deformed configuration.

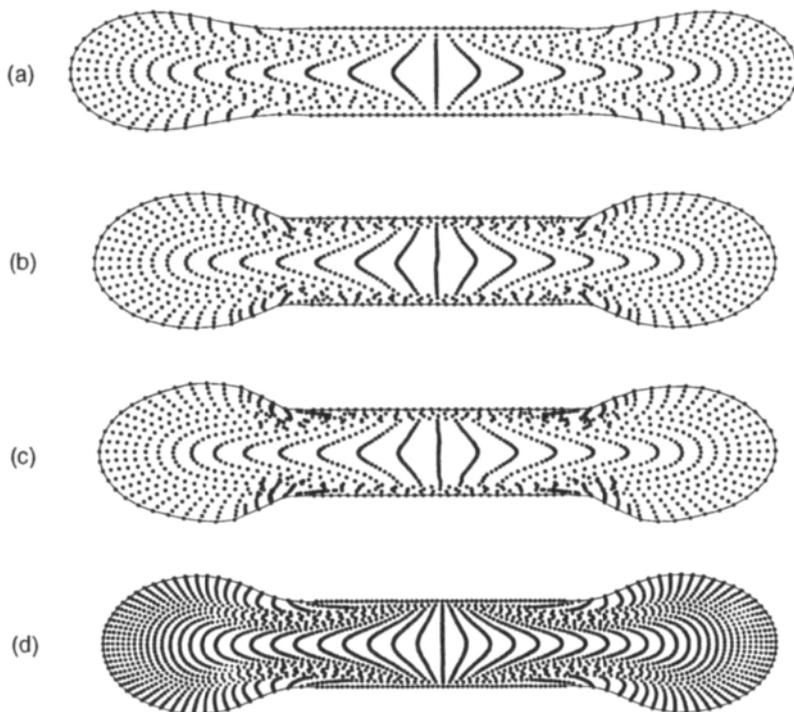


Figure 2: Comparison of deformation; (a) coarse distribution (961 particles), (b) 1st refinement (1113 particles), (c) 2nd refinement (1213 particles), (d) fine distribution (2601 particles).

It is seen that the high gradient values of I_2 are concentrated around the upper and lower boundaries. New particles are added at the center of the cells which share a node indicating high scale location. In this method, the algorithm for the addition of nodes is similar to that of the finite element method. We have recently changed the algorithm to include the insertion of an additional node between any two high gradient nodes. This new algorithm is valid for arbitrarily distributed particles. The high scale is selected if the absolute value of I_2 is larger than 30% of the highest peak of I_2 . From the contours of I_2 in Figure 1 (c), it can be seen that the high gradient values are localized at the corners of the material. The deformations at $t=3.25 \times 10^{-6}$ (sec) of each particle refinement are presented in Figure 2. The initial coarse discretization has 961 nodes. The first and second refinements have 1113 nodes and 1213 nodes, respectively. For comparison, the results for the finer particle distributions of 2601 and 5041 nodes are illustrated in Figure 2. It is shown that the first refinement provides a deformed shape which is very close to the deformed shapes of the finer particle distributions.

7. CONCLUSION

The framework for meshfree methods based on kernel approximations has been presented. The modification of the kernel required to satisfy reproducing conditions is also described. A wavelet interpretation of the kernel approximation enables the approximation to be subdivided into high and low scale (short wavelength and long wavelength) components. It is suggested that the highest scale component can be used as an error indicator. As an example of adaptivity driven by this indicator has been given.

8. ACKNOWLEDGMENTS

The support of this research by grants from the Office of Naval Research and the National Science Foundation is gratefully acknowledged. Support was also provided by the Army HPC Research Center under the auspices of the Department of the Army, Army Research Laboratory. The content does not necessarily reflect the position or the policy of the government and no official endorsement should be inferred.

8. REFERENCES

- Babuska, I. and Melenk, J. M., [1995]; The Partition of Unity Finite Element Method, *Univ. of Maryland, Technical Note BN-1185*.
- Belytschko, T., Lu, Y. Y. and Gu, L., [1994]; Element Free Galerkin Methods, *International Journal for Numerical Methods in Engineering*, **37**, 229-256.
- Duarte, C. A. and Oden, J. T. [1995]; Hp Clouds - A Meshless Method to Solve Boundary -Value Problems, *TICAM Report*, **95-05**.
- Gingold, R. A. and Monaghan, J. J., [1977]; Smoothed Particle Hydrodynamics: Theory and Application to Non-Spherical Stars, *Mon. Not. Roy. Astron. Soc.*, **181**, 375-389.
- Jun, S., Liu, W. K. and Belytschko, T., [1996]; Explicit Reproducing Kernel Particle Methods for Large Deformation Problems, *International Journal of Numerical Methods for Engineering*, in press.
- Liu, W. K., Adee, J. and Jun, S., [1993]; Reproducing Kernel Particle Methods for Elastic and Plastic Problems, *Advanced Computational Methods for Material Modeling*, eds. Benson, D. J., and Asaro, R. A., AMD 180 and PVP 268, ASME, 175-190.
- Liu, W. K., Chang, C. T., Chen, Y. and Uras, R. A., [1995a]; *Multiresolution Reproducing Kernel Particle Methods in Acoustic Problems, Acoustics, Vibration and Rotating Machines (ASME DE-Vol. 84-2, Part-B)*, 881-900.
- Liu, W. K. and Chen, Y., [1995]; Wavelet and Multiple Scale Reproducing Kernel Methods, *International Journal of Numerical Methods in Fluids*, **21**, 901-931.

- Liu, W. K., Chen, Y., Chang, C. T. and Belytschko, T., [1996a]; Advances in Multiple Scale Kernel Particle Methods, *Computational Mechanics*, **18-2**, 73-111.
- Liu, W. K., Chen, Y., Jun, S., Chen, J. S., Belytschko, T., Pan, C., Uras, R. A. and Chang, C. T., [1996b]; Overview and Applications of the Reproducing Kernel Particle Methods, *Archives of Computational Methods in Engineering*, **3**, **1**, 3-80.
- Liu, W. K., Chen, Y. and Uras, R. A., [1995b]; Enrichment of the Finite Element Method with Reproducing Kernel Particle Method, *Current Topics in Computational Mechanics*, eds. Cory, J. F. Jr. and Gordon, J. L., ASME PVP, **305**, 253-258.
- Liu, W. K., Chen, Y., Uras, R. A. and Chang, C. T., [1996c]; Generalized Multiple Scale Reproducing Kernel Particle Method, *Computer Methods in Applied Mechanics and Engineering*, **139**, 91-158.
- Liu, W. K., Jun, S., Li, S., Adee, J., and Belytschko, T., [1995c]; Reproducing Kernel Particle Methods for Structural Dynamics, *International Journal of Numerical Methods for Engineering*, **38**, 1655-1679.
- Liu, W. K., Jun, S., Sihling, D. T., Chen, Y. and Hao, W., [1996d]; *Multiresolution Reproducing Kernel Particle Method for Computational Fluid Dynamics*, presented in the 3rd US-JAPAN Symposium on Finite Element Methods in Large-Scale Computational Fluid Dynamics, also accepted for publication in *International Journal for Numerical Methods in Fluids*.
- Liu, W. K., Jun, S., Zhang, Y., [1995d]; Reproducing Kernel Particle Methods, *International Journal of Numerical Methods in Fluids*, **20**, 1081-1106.
- Liu, W. K. and Oberste-Brandenburg, C., [1993]; Reproducing Kernel and Wavelet Particle Methods, *Aerospace Structures: Nonlinear Dynamics and System Response*, eds. Cusumano, J. P., Pierre, C., and Wu, S. T., AD 33, ASME, 39-56.
- Lu, Y. Y., Belytschko, T. and Gu, L., [1994]; A New Implementation of the Element Free Galerkin Methods, *Computer Methods in Applied Mechanics and Engineering*, **113**, 397-414.
- Lucy, L., [1977]; A Numerical Approach to Testing The Fission Hypothesis, *A. J.*, **82**, 1013-1024.
- Monaghan, J. J., [1988]; An Introduction to SPH, *Comp. Phys. Comm.*, **48**, 89-96.
- Oñate, E., Idelson, S., Fischer, T. and Zienkiewicz, O.C., [1995]; A Finite Point Method for Analysis of Fluid Flow Problems, presented in *9th Int. Conference on Finite Elements in Fluids*, Venezia, Italy, 15-21 October.
- Yagawa, G., [1995]; Some Remarks on Free Mesh Method : A Kind of Meshless Finite Element Method. presented in *International Conference on Computational Engineering Science*, Hawaii, USA, August.

PART 4

ERROR ESTIMATORS FOR NON LINEAR TIME-DEPENDENT PROBLEMS AND ADAPTIVE COMPUTATIONAL METHODS

This Page Intentionally Left Blank

A posteriori constitutive relation error estimators for nonlinear finite element analysis and adaptive control

Pierre Ladevèze^a and Nicolas Moës^{b*}

^aLaboratoire de Mécanique et Technologie,
E.N.S. Cachan / C.N.R.S. / Université Paris 6,
61, Avenue du Président Wilson, 94235 Cachan Cedex, France.

^bThe Texas Institute for Computational and Applied Mathematics,
The University of Texas at Austin,
Austin, TX 78712, U.S.A.

Today quite a lot of a posteriori error estimators are available to control finite element calculations. We describe here error estimators based on constitutive relation residuals which have been developed for 20 years, in particular at Cachan. This approach has a strong physical meaning and is quite general. Different errors on constitutive relation can be easily introduced to measure the quality of finite element computations of plastic and viscoplastic structures whose behavior is described by internal variables. These measures take into account, over the studied time interval, all the classical sources of error involved in the computation: the space discretization (the mesh), the time discretization and the iterative technique used to solve the nonlinear discrete problem. To quantify more specifically each source of error, we introduce quantities called indicators. Numerical experiments show that the indicators are linked to the error in a limit sense. The errors and the indicators may then be used to simultaneously adapt the mesh, the time discretization and the stopping criteria to meet a prescribed accuracy. The adaptive control is illustrated on two plane stress problems using the Prandtl-Reuss plastic model and its viscoplastic version.

1. INTRODUCTION

Nowadays, numerical simulations are used to solve more and more complex problems. The control of such computations has become a critical issue for their performance. Concerning a posteriori error estimate for the finite element approximation of linear elliptic problems, a vast literature is available. Three main approaches must be distinguished. Chronologically speaking, the first one is based on the concept of error in the constitutive law [1]. This concept is a priori independent of the type of numerical approximation used. Its application in the case of finite element approximations may be found in [2,3,1].

*The second author gratefully acknowledges the support of this work by the U.S. Office of Naval Research under Grant N00014-95-1-0401 and the National Science Foundation under grant ECS-9422707.

The second approach, introduced by Babuska and Rheinboldt [4,5], then developed by Zienkiewicz, Gago, Kelly [6,7] and more recently by Oden et al. [8,9], uses the equilibrium residuals through local problems to estimate the error. The last one, developed by Zienkiewicz and Zhu [10–12], consists of comparing the finite element solution to a smoother one obtained by special averaging techniques. A validation of these a posteriori error estimators may be found in [13,14].

Here, we are interested in plastic and viscoplastic problems under the small strains, small displacements and isothermal assumptions. These problems are time-dependent nonlinear problems. The literature available is much narrower compare to the linear case. Also, it is important to distinguish between the nonlinear time-dependent and time-independent problems. Concerning the later type, let us mention Babuska [15] for the design of estimates for nonlinear elasticity problems of rods, Johnson [16] in the field of Hencky-type plasticity, and Verfurth for the proof of bounds on the error [17]. For nonlinear time-dependent problems, viscoplasticity problems were treated in [18], strain localization problems in [19–21] and large strains problems in [22–25]. In most cases, techniques devised for linear problems or time-independent nonlinear problems are used at each time step so that the error due to the time discretization may hardly be taken into account.

We focus here our attention here on solid mechanics but the design of a posteriori error estimator for nonlinear fluid related problems is also an active area of research [26–28].

In nonlinear time-dependent problems, three error sources must be distinguished: the space discretization, the time discretization and the iterative technique used to solve the nonlinear discrete problem. Several error estimators based on the constitutive law residuals have been introduced in Cachan to take these three sources into account, mainly the Drucker error and the dissipation error.

The Drucker error is based on a sufficient condition insuring the stability of the material (Drucker's inequality [29]). It was introduced in [30] and first applied in [31] for plane stress problems and three-node triangular elements. A procedure to adapt the mesh is also described in this paper. Later, the error has been reused and enhanced in order to conduct a simultaneous adaptive control of the space and time discretization for three and six-node triangular elements in plane and axisymmetric problems [32]. The Drucker error only assumes that the functional law linking the present stress state to the strain history satisfies Drucker's stability condition.

Nowadays, characterization of material behaviors using internal variable formulation is increasingly employed at the theoretical, experimental and computational levels. In this case, a more natural error has been elaborated in [33] and [34]. It was named dissipation error since the error is estimated through the residuals of the laws describing the evolution of the state of the material, thus the dissipation process. It's usefulness has been evaluated in the framework of the classical incremental finite element method in [35].

As for the Drucker error, the dissipation error does take into account all the error sources involved in the computation. For adaptive purposes, it is important to know the contribution to the total error of each error source. We define three errors: the time error, the space error and the iteration error. For a given time (space resp.) discretization, the time (space resp.) error is the limit of the error (dissipation or Drucker) as the mesh size (time-step size resp.) tends to zero. For a given stopping criteria of the iterative strategy,

the iteration error is the limit of the error (dissipation or Drucker) as the mesh size and time-step size tends to zero. Space, time and iteration error are theoretical quantities that are estimated through error indicators: space, time and iteration error indicators, respectively. In [32], space and time error indicator were introduced for the Drucker error and allowed the adaptation of the mesh and the time discretization. Concerning the dissipation error, space and time error indicators were introduced in [35] and an iteration indicator in [36]. A simultaneous time, space and iteration adaptive scheme is also presented in [36]. Basically, the error indicators are introduced by modifying the reference problem: to obtain the time (space resp.) error indicator, the reference problem is taken as the space-discrete (time- resp.) problem. For the iteration error indicator, the reference is the time-space discrete problem.

The present paper is basically a review of the use of the error in the constitutive law in the control of finite element approximations of plastic and viscoplastic problems. However, concerning the Drucker error, we take the opportunity to introduce an iteration error indicator that was missing and we define a new space error indicator in the philosophy described at the end of the previous paragraph. So that now, time, space and iteration error indicators, based on the same philosophy, are available for the dissipation and Drucker approach.

The plan of the paper is as follows. In section 2, the reference problem and the formulation of the material behavior are recalled. The next section is devoted to the error in the constitutive law for nonlinear problem. Then, in section 4 the implementation of the errors in the finite element framework is explained. Error indicators are reviewed in section 5 and two new ones are proposed. They are tested on examples. Finally, in the last section, simultaneous time, space and iteration control of the computation are shown for two examples.

2. THE MODEL PROBLEM

Concerning the notations, vectors will be underlined ($\underline{u}, \underline{v}, \dots$) and second-order tensors shaded. For instance, the stress and strain tensors will be denoted by σ and ϵ , respectively. This notation will also be used to denote the additional internal variables. More complex operators will be in bold (e.g. \mathbf{K} for the Hooke's tensor).

The solid medium under study occupies a domain Ω , bounded by $\partial\Omega$, which is independent of t (small strains and displacements assumption). The environment of the medium is schematized for all $t \in [0, T]$, with an imposed displacement \underline{u}_d on a part $\partial_1\Omega$ of the boundary, a traction \underline{F}_d on $\partial_2\Omega$ (complementary to $\partial_1\Omega$), and a body load \underline{f}_d on the domain Ω . The partition of $\partial\Omega$ in $\partial_1\Omega$ and $\partial_2\Omega$ is assumed constant in time.

\mathcal{U} denotes the space of the displacement field \underline{u} defined on Ω , and \mathcal{S} the space of the stress field, also defined on Ω . The extensions of these two spaces to the entire time range $[0, T]$ will be denoted $\mathcal{U}^{[0, T]}$ and $\mathcal{S}^{[0, T]}$.

The solution must satisfy the kinematic constraints, the equilibrium equations and the constitutive equations describing the behavior of the material.

2.1. Kinematic constraints and equilibrium equations

The displacement solution, \underline{u} , must belong to the space $\mathcal{U}_{ad}^{[0,T]}$:

$$\underline{u} \in \mathcal{U}_{ad}^{[0,T]} = \{ \underline{v} \in \mathcal{U}^{[0,T]} \text{ such that } \underline{v} = \underline{u}_d \text{ on } \partial_1 \Omega \times [0, T] \} \quad (1)$$

The stress solution, σ , must belong to space $\mathcal{S}_{ad}^{[0,T]}$:

$$\sigma \in \mathcal{S}_{ad}^{[0,T]} = \{ \tau \in \mathcal{S}^{[0,T]} \text{ satisfying (3) , } \forall t \in [0, T] \} \quad (2)$$

$$\int_{\Omega} \tau : \epsilon(\underline{v}) \, d\Omega - \int_{\Omega} \underline{f}_d \cdot \underline{v} \, d\Omega - \int_{\partial_2 \Omega} \underline{F}_d \cdot \underline{v} \, dS = 0, \quad \forall \underline{v} \in \mathcal{U}_0 \quad (3)$$

where $:$ denotes the dot product of second order tensors; \cdot the scalar product and

$$\mathcal{U}_0 = \{ \underline{v} \in \mathcal{U} \text{ such that } \underline{v} = 0 \text{ on } \partial_1 \Omega \} .$$

2.2. Constitutive relation

The functional approach expresses the stress state at a given time t as a function of the past history of the strain rate:

$$\sigma(t) = \mathcal{A}(\dot{\epsilon}(t'), t' \leq t) \quad (4)$$

In the internal variable approach, the state of the material is characterized at each point by the total strain ϵ , the inelastic strain ϵ^P and a set of internal variables denoted by X . The associated variables are the quantity Y for X and the stress σ for ϵ , ϵ^P and ϵ^e . The expression of the dissipation is given by

$$\sigma : \dot{\epsilon}^P - Y \circ \dot{X} \quad (5)$$

The second term specifies the contribution of (X, Y) to the dissipation. If X denotes a column of \mathbb{R}^n , then Y is also a column of \mathbb{R}^n and we have

$$Y \circ \dot{X} = Y^t \dot{X}$$

where t stands for the usual transposition. More precisely, two spaces \mathbf{e} and \mathbf{f} are placed in duality by the following bilinear form:

$$\begin{array}{ccc} \left[\begin{array}{c} \dot{\epsilon}^P \\ -\dot{X} \end{array} \right], \left[\begin{array}{c} \sigma \\ Y \end{array} \right] & \rightarrow & \sigma : \dot{\epsilon}^P - Y \circ \dot{X} \\ \mathbf{e} \times \mathbf{f} & \rightarrow & \mathbb{R} \end{array}$$

As we work with the small strains assumption, the strain is given by the symmetric part of the gradient of the displacement. In an orthonormal basis, we have

$$[\epsilon(\underline{v})]_{ij} = \frac{1}{2}(v_{i,j} + v_{j,i})$$

and we also have the additivity relation of the elastic strain rate $\dot{\epsilon}^e$ and the inelastic strain rate $\dot{\epsilon}^P$: $\dot{\epsilon} = \dot{\epsilon}^e + \dot{\epsilon}^P$. The material behavior is described in term of state laws, evolution laws and initial conditions.

The state laws

According to the first principle of thermodynamics, a free energy ψ , depending only on the state variables ϵ , ϵ^p and \mathbf{X} can be introduced. The following classical modeling assumptions are made

- ψ depends only on the elastic strain ϵ^e and the internal variables \mathbf{X}
- $\psi(\epsilon^e, \mathbf{X}) = \psi_e(\epsilon^e) + \psi_p(\mathbf{X})$
- linear elasticity: $\psi_e(\epsilon^e) = \frac{1}{2} \mathbf{K} \epsilon^e : \epsilon^e$, where \mathbf{K} is Hooke's tensor.

The derivation of ψ yields the state equations $\boldsymbol{\sigma} = \mathbf{K} \epsilon^e$, $\mathbf{Y} = \mathbf{G}(\mathbf{X})$ where $\mathbf{G}(\mathbf{X})$ is the derivative of ψ_p with respect to \mathbf{X} .

The evolution laws

The second principle of thermodynamics, written as $\boldsymbol{\sigma} : \dot{\epsilon}^p - \mathbf{Y} \circ \dot{\mathbf{X}} \geq 0$, imposes a constraint on the evolution laws relating $(\dot{\epsilon}^p, -\dot{\mathbf{X}})$ to $(\boldsymbol{\sigma}, \mathbf{Y})$. This law can be written

$$\begin{bmatrix} \dot{\epsilon}^p \\ -\dot{\mathbf{X}} \end{bmatrix} = \mathbf{B} \left(\begin{bmatrix} \boldsymbol{\sigma} \\ \mathbf{Y} \end{bmatrix} \right) \quad (6)$$

\mathbf{B} is an operator relevant to the material. It must be positive to respect the second principle of thermodynamics. A typical way to define the operator \mathbf{B} is to give a scalar function $\varphi^*(\boldsymbol{\sigma}, \mathbf{Y})$, generally convex, called the potential of dissipation, and to write

$$\begin{pmatrix} \dot{\epsilon}^p \\ -\dot{\mathbf{X}} \end{pmatrix} \in \begin{pmatrix} \partial_{\boldsymbol{\sigma}} \varphi^*(\boldsymbol{\sigma}, \mathbf{Y}) \\ \partial_{\mathbf{Y}} \varphi^*(\boldsymbol{\sigma}, \mathbf{Y}) \end{pmatrix} \quad (7)$$

where $(\partial_{\boldsymbol{\sigma}} \varphi^*, \partial_{\mathbf{Y}} \varphi^*)$ denotes the subdifferential of φ^* at $(\boldsymbol{\sigma}, \mathbf{Y})$. This defines a *standard material*. When the potential is differentiable, the subdifferential becomes a classical gradient and the belonging an equality. The interest of a standard model lies in the following classical property. The second principle of thermodynamics is fulfilled if the potential satisfies

$$\varphi^* \text{ convex,} \quad \varphi^*(0, 0) = 0, \quad \varphi^*(\cdot, \cdot) \geq 0 \quad (8)$$

Concerning the initial conditions, we assume that the material is initially virgin:

$$\epsilon^p = 0, \quad \mathbf{X} = 0, \quad \text{for } t = 0 \quad (9)$$

Finally, note that by eliminating the internal variables in the state laws and evolution laws, one can always, at least formally, associate a stress-strain functional law to an internal variable formulation.

Examples

As an example of material behavior described by internal variables, consider the Prandtl-Reuss plastic model. In addition to the plastic strain, the model involves another scalar

internal variable p that can be interpreted as the cumulative plastic strain. The free energy is of the form

$$\psi(\epsilon^e, p) = \frac{1}{2} \mathbf{K} \epsilon^e : \epsilon^e + g(p)$$

where g is a function characterizing the hardening law. By derivation, we obtain the state laws $\sigma = \mathbf{K} \epsilon^e$, $R = g'(p)$, where g' denotes the derivative of g with respect to p and R is the associate variable to p . Classical hardening laws are $R = k_y p$ (linear hardening), $R = k_y p^{1/m}$ (power hardening) and $R = R_M (1 - \exp(-\gamma p))$ (exponential hardening) where k_y , m , R_M and γ are constant material parameters. Prandtl-Reuss materials are standard with a dissipation potential given by

$$\varphi^*(\sigma, R) = \begin{cases} 0 & \text{if } z \leq 0 \\ +\infty & \text{if } z > 0 \end{cases} \quad \text{where } z(\sigma, R) = \|\sigma^D\| - (R + R_0) \quad (10)$$

σ^D is the deviatoric part of the stress tensor, R_0 the initial yield stress, and $\|\sigma^D\| = (\sigma^D : \sigma^D)^{1/2}$.

A viscoplastic version of the Prandtl-Reuss plastic model may be obtained by the regularization of the plastic potential; for instance, with a power law:

$$\varphi^*(\sigma, R) = \frac{k_v}{n_v + 1} \langle z \rangle_+^{n_v + 1} \quad (11)$$

k_v , n_v are positive constant materials parameters and $\langle z \rangle_+$ denotes the positive part of z : $\langle z \rangle_+ = \frac{z + |z|}{2}$.

3. ERROR IN THE CONSTITUTIVE LAW

The notion of error in the constitutive law has been introduced in [37]. It relies on splitting the equations of the problem into two groups.

When the constitutive law is of the functional type, the first group combines both the equilibrium equations and the kinematic constraints, and the second group contains the constitutive law. The quality of an approximate solution satisfying the first group (i.e. an admissible solution) is quantified by the non-fulfillment of the second group of equations (constitutive law). If Drucker's stability inequality, [29], holds for the material, a natural way to measure the error can be obtained [32,31].

When the constitutive law is described in terms of internal variables, the notion of admissibility must be revised. Indeed, the state equations – associated with the free energy – and the evolution laws – associated with the dissipative phenomena – must be distinguished in the formulation. In [34], the state equations are included in the definition of admissibility, the error being measured on the evolution laws alone.

3.1. The Drucker error

The Drucker error is based on a sufficient condition ensuring the stability of the material. Let (σ, ϵ) and (σ', ϵ') be two stress-strain couples satisfying the behavior described by (4) on $I = [0, T]$, with $(\sigma, \epsilon) = (\sigma', \epsilon') = 0$ at $t = 0$. The material is said to be stable in the

Drucker sense if it satisfies (12). Moreover, if (13) is satisfied, the material is said to be strictly stable in the Drucker sense

$$\forall t \in I, \int_0^t (\boldsymbol{\sigma} - \boldsymbol{\sigma}') : (\dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}') dt \geq 0 \quad (12)$$

$$\forall t \in I, \int_0^t (\boldsymbol{\sigma} - \boldsymbol{\sigma}') : (\dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}') dt = 0 \Leftrightarrow \forall t \in I, \begin{cases} \boldsymbol{\epsilon} = \boldsymbol{\epsilon}' \\ \boldsymbol{\sigma} = \boldsymbol{\sigma}' \end{cases} \quad (13)$$

We shall need the convenient notation $\gamma_t(\cdot, \cdot)$:

$$\gamma_t(\boldsymbol{\sigma}, \boldsymbol{\epsilon}) = \int_0^t (\boldsymbol{\sigma} - \boldsymbol{\sigma}')(\dot{\boldsymbol{\epsilon}}' - \dot{\boldsymbol{\epsilon}}) dt \quad (14)$$

where $(\boldsymbol{\sigma}, \boldsymbol{\epsilon})$ and $(\boldsymbol{\sigma}', \boldsymbol{\epsilon}')$ are two stress-strain couples satisfying the behavior on $I = [0, T]$ with $(\boldsymbol{\sigma}, \boldsymbol{\epsilon}) = (\boldsymbol{\sigma}', \boldsymbol{\epsilon}') = 0$ at $t = 0$.

Let $(\hat{\boldsymbol{u}}, \hat{\boldsymbol{\sigma}})$ be a displacement-stress couple defined on $\Omega \times [0, T]$ such that $\hat{\boldsymbol{u}}$ is kinematically admissible and $\hat{\boldsymbol{\sigma}}$ is statically admissible, i.e.

$$\hat{\boldsymbol{u}} \in \mathcal{U}_{ad}^{[0, T]}, \quad \hat{\boldsymbol{\sigma}} \in \mathcal{S}_{ad}^{[0, T]} \quad (15)$$

This couple satisfies all the equations of the problem except, in general, the constitutive relation. It is an approximate solution to the problem (1)-(3)-(4). To the strain field, $\hat{\boldsymbol{\epsilon}} = \boldsymbol{\epsilon}(\hat{\boldsymbol{u}})$, one can associate the stress field $\hat{\boldsymbol{\sigma}}'$ using the constitutive relation (4). Similarly, one can associate the strain field $\hat{\boldsymbol{\epsilon}}'$ to the stress field $\hat{\boldsymbol{\sigma}}$ using the reverse relation. The couple $(\hat{\boldsymbol{u}}, \hat{\boldsymbol{\sigma}})$ is the exact solution of the problem if and only if

$$\gamma_t(\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\epsilon}}) = 0 \quad \text{on } \Omega \times [0, T] \quad (16)$$

To estimate the quality of the couple $(\hat{\boldsymbol{u}}, \hat{\boldsymbol{\sigma}})$ as the approximate solution to the problem, one may define the absolute error as

$$e^2 = \sup_{t \in [0, T]} \int_{\Omega} \gamma_t(\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\epsilon}}) d\Omega \quad (17)$$

and the relative error by $\epsilon = e/D$ where

$$D^2 = 2 \sup_{t \in [0, T]} \int_{\Omega} \gamma_t(\hat{\boldsymbol{\sigma}}, 0) + \gamma_t(0, \hat{\boldsymbol{\epsilon}}) d\Omega \quad (18)$$

The contribution of the time interval $[0, t]$ to the error is defined by ϵ_t

$$\epsilon_t^2 = \sup_{t' \in [0, t]} \int_{\Omega} \gamma_{t'}(\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\epsilon}}) d\Omega / D^2 \quad (19)$$

3.2. The dissipation error

The problem is divided precisely into two groups:

- The first group – related to the free energy – defines the admissibility of a solution. It combines the equilibrium equations, the kinematic constraints and the state equations. The initial conditions are also included in this group.

- The second group – related to the dissipation – only includes the evolution laws. The quality of an admissible solution will be estimated through the quality of satisfaction of these laws.

The practical definition of the error depends on the form of the evolution laws. We shall here consider standard materials that possess a normal standard formulation. This class is important and gathers many practical stable material. For the definition of the dissipation error for other types of evolution laws, see [34,35]. A standard material possess a normal standard formulation if by definition there exist a change of variables $\tilde{X} = R(X)$, $\tilde{Y} = S(Y)$ such that

- the state equations become linear: $\sigma = K(\epsilon - \epsilon^p)$, $\tilde{Y} = \Lambda \tilde{X}$ where Λ is a constant symmetric definite positive operator;
- the standard character of the evolution laws is kept with the potential φ^* satisfying (8)

$$\begin{pmatrix} \dot{\epsilon}^p \\ -\dot{\tilde{X}} \end{pmatrix} \in \begin{pmatrix} \partial_{\sigma} \varphi^*(\sigma, \tilde{Y}) \\ \partial_{\tilde{Y}} \varphi^*(\sigma, \tilde{Y}) \end{pmatrix} \quad \tilde{X} = 0, \epsilon^p = 0, \quad \text{for } t = 0 \quad (20)$$

- the new formulation still yield the same stress-strain functional law.

The condition for a standard material to possess a *normal* standard formulation are not too restrictive in practice, see [33]. For instance, the Prandtl-Reuss model previously described fulfills them. For the sake of simplicity in notations, the tilde will be dropped, so that the state laws an evolution laws are now written

$$\sigma = K(\epsilon - \epsilon^p), \quad Y = \Lambda X \quad (21)$$

$$\begin{pmatrix} \dot{\epsilon}^p \\ -\dot{X} \end{pmatrix} \in \begin{pmatrix} \partial_{\sigma} \varphi^*(\sigma, Y) \\ \partial_Y \varphi^*(\sigma, Y) \end{pmatrix} \quad X = 0, \epsilon^p = 0, \quad \text{for } t = 0 \quad (22)$$

Finally, note that a standard material possessing a normal standard formulation is strictly stable in the Drucker sense. In other words, the associated stress-strain relation satisfies (13). See [38] for the proof.

With the help of the Legendre-Fenchel transform, one can associate a dual potential $\varphi(\dot{\epsilon}^p, -\dot{X})$ to the primal one $\varphi^*(\sigma, Y)$.

$$\varphi(\dot{\epsilon}^p, -\dot{X}) = \sup_{(\sigma, Y) \in f} \left(\sigma : \dot{\epsilon}^p - Y \circ \dot{X} - \varphi^*(\sigma, Y) \right) \quad (23)$$

The Legendre-Fenchel transform possesses two interesting classical properties:

$$\eta(\dot{\epsilon}^p, \dot{X}, \sigma, Y) \geq 0 \quad \forall (\sigma, Y, \dot{\epsilon}^p, -\dot{X}) \quad (24)$$

$$\eta(\dot{\epsilon}^p, \dot{X}, \sigma, Y) = 0 \Leftrightarrow \begin{pmatrix} \dot{\epsilon}^p \\ -\dot{X} \end{pmatrix} \in \begin{pmatrix} \partial_{\sigma} \varphi^*(\sigma, Y) \\ \partial_Y \varphi^*(\sigma, Y) \end{pmatrix} \quad (25)$$

where the condensed following notation has been used:

$$\eta(\dot{\epsilon}^P, \dot{X}, \sigma, Y) = \varphi(\dot{\epsilon}^P, -\dot{X}) + \varphi^*(\sigma, Y) - \sigma : \dot{\epsilon}^P + Y \circ \dot{X} \quad (26)$$

The inequality (24) is usually called the Legendre-Fenchel inequality, and the equality relation (25) simply means that η is zero if and only if the evolution laws are satisfied.

For example, we give below the expression of the dual potentials for the Prandtl-Reuss model. We shall use the convex sets

$$\begin{aligned} C_1^* &= \{(\sigma, R) \in \mathbf{f} \mid z \leq 0\}, & C_1 &= \{(\dot{\epsilon}^P, -\dot{p}) \in \mathbf{e} \mid \text{Tr}[\dot{\epsilon}^P] = 0\}, \\ C_2^* &= \{(\sigma, R) \in \mathbf{f} \mid R \geq 0\}, & C_2 &= \{(\dot{\epsilon}^P, -\dot{p}) \in \mathbf{e} \mid \|\dot{\epsilon}^P\| - \dot{p} \leq 0\} \end{aligned}$$

where Tr is the trace operator. In the plastic case, the dual potential are given by

$$\varphi^*(\sigma, R) = \Psi_{C_1^* \cap C_2^*}, \quad \varphi(\dot{\epsilon}^P, -\dot{p}) = R_0 \|\dot{\epsilon}^P\| + \Psi_{C_1 \cap C_2} \quad (27)$$

where Ψ_A is the indicator function associated with the convex domain A . The introduction of the condition $R \geq 0$ is explained in [35]. In the visco-plastic case, we have

$$\begin{aligned} \varphi^*(\sigma, R) &= \frac{k_v}{n_v + 1} (z)_+^{n_v + 1} + \Psi_{C_2^*} \\ \varphi(\dot{\epsilon}^P, -\dot{p}) &= R_0 \|\dot{\epsilon}^P\| + \frac{k_v}{n'_v + 1} \left(\frac{\|\dot{\epsilon}^P\|}{k_v} \right)^{n'_v + 1} + \Psi_{C_1 \cap C_2} \end{aligned}$$

where n'_v is the inverse of n_v .

Dissipation error for normal standard formulation

Let us consider an admissible solution $(\hat{u}, \hat{\epsilon}^e, \hat{\epsilon}^P, \hat{X}, \hat{\sigma}, \hat{Y})$, i.e.. a solution satisfying

$$\hat{u} \in U_{ad}^{[0,T]}, \quad \hat{\sigma} \in S_{ad}^{[0,T]} \quad (28)$$

$$\hat{\sigma} = K \hat{\epsilon}^e = K(\epsilon(\hat{u}) - \hat{\epsilon}^P), \quad \hat{Y} = \Lambda \hat{X}, \quad \text{on } \Omega \times [0, T] \quad (29)$$

$$\hat{\epsilon}^P = \hat{X} = 0 \text{ at } t = 0 \text{ on } \Omega \quad (30)$$

This solution is the exact solution of the problem if and only if

$$\eta(\hat{\epsilon}^P, \hat{X}, \hat{\sigma}, \hat{Y}) = 0 \text{ on } \Omega \times [0, T]. \quad (31)$$

Thus, the absolute error, e , may be defined as

$$e = \int_0^T \int_{\Omega} \eta(\hat{\epsilon}^P, \hat{X}, \hat{\sigma}, \hat{Y}) \, d\Omega dt \quad (32)$$

and the relative error by $\epsilon = e/D$ where

$$\begin{aligned} D &= 2 \sup_{t \in [0, T]} d_t, \\ d_t &= \int_0^t \int_{\Omega} \sup \{ \varphi^*(\hat{\sigma}, \hat{Y}) + \varphi(\hat{\epsilon}^P, \hat{X}), \frac{R_0}{\|\hat{\sigma}^D\|} |\psi_e(\hat{\epsilon}^e)| \} \, d\Omega dt + \int_{\Omega} |\hat{\sigma}|_K^2 + |\hat{Y}|_{\Lambda}^2 \, d\Omega \quad (33) \end{aligned}$$

We have used the notation $|\sigma|_K^2 = 1/2K^{-1}\sigma : \sigma$ and $|Y|_\Lambda^2 = 1/2\Lambda^{-1}Y \circ Y$. The denominator is chosen so that the relative error is reasonable in elasticity in comparison to classical error norm for such problems, see [35]. The contribution of the time interval $[0, t]$ to the error is defined by ϵ_t

$$\epsilon_t = \int_0^t \int_\Omega \eta(\widehat{\epsilon}^p, \widehat{X}, \widehat{\sigma}, \widehat{Y}) \, d\Omega dt / D \quad (34)$$

The norm chosen allows writing a direct link between the error and the gap occurring between the exact and admissible solutions:

$$e = \int_0^T \int_\Omega \eta(\widehat{\epsilon}^p, \widehat{X}, \overline{\sigma}, \overline{Y}) + \eta(\overline{\epsilon}^p, \overline{X}, \widehat{\sigma}, \widehat{Y}) \, d\Omega dt + \int_\Omega |\widehat{\sigma} - \overline{\sigma}|_K^2 + |\widehat{Y} - \overline{Y}|_\Lambda^2 \, d\Omega \Big|_T$$

where $(\overline{\epsilon}^p, \overline{X}, \overline{\sigma}, \overline{Y})$ denotes the exact solution. This property has been established in [34] and is an extension of the famous Prager-Synge theorem [39].

4. IMPLEMENTATION OF THE ERRORS FOR THE FEM

We consider the classical incremental finite element method. The Drucker and dissipation error cannot be directly measured on the finite element solution because it is not generally admissible. For instance the finite element stresses are only in equilibrium in a weak, average, sense. We thus need to build an admissible solution from the finite element one. But, let us first detail the discrete problem.

4.1. The discrete problem

The problem to be solved on $[0, T]$ is divided into a succession of resolutions over $[t_n, t_{n+1}]$ ($n = 0, \dots, N$; $t_0 = 0$; $t_{N+1} = T$). Assuming the solution is known up until the time instant t_n , one must then build the solution over $[t_n, t_{n+1}]$. Both time and space discretizations need to be performed.

Time discretization

A common choice is to assume a linear evolution of the solution over $[t_n, t_{n+1}]$. Thus, the problem reduces to the search of the solution at t_{n+1} . The equilibrium equations and kinematic constraints are written at t_{n+1} and the constitutive law is discretized to express algebraically the field at t_{n+1} in term of the field at t_n . The discretization of the functional law (4) over $[t_n, t_{n+1}]$ may formally be written

$$\sigma_{n+1} - \sigma_n = \mathcal{A}_n(\epsilon_{n+1} - \epsilon_n) \quad (35)$$

The notation \mathcal{A}_n reminds that the stresses at t_{n+1} are no longer expressed as a functional of the strain rate history on $[t_n, t_{n+1}]$, but as an algebraic, nonlinear, function of the strain increase $(\epsilon_{n+1} - \epsilon_n)$. In the case of an internal variable formulation of the behavior, the state laws are written at t_{n+1} :

$$\sigma_{n+1} = K(\epsilon_{n+1} - \epsilon_{n+1}^p), \quad Y_{n+1} = \Lambda X_{n+1} \quad (36)$$

and we consider the generalized mid-point rule to discretize the evolution laws:

$$\begin{pmatrix} \dot{\epsilon}_{n+1}^p \\ -\dot{X}_{n+1} \end{pmatrix} \in \begin{pmatrix} \partial_{\sigma} \varphi^*(\sigma_{\theta}, Y_{\theta}) \\ \partial_Y \varphi^*(\sigma_{\theta}, Y_{\theta}) \end{pmatrix}, \quad X_0 = 0, \epsilon_0^p = 0 \quad (37)$$

where we have used the notation

$$\dot{x}_{n+1} = \frac{x_{n+1} - x_n}{t_{n+1} - t_n}, \quad x_{\theta} = (1 - \theta)x_n + \theta x_{n+1}, \quad 0 \leq \theta \leq 1 \quad (38)$$

The relations (35) and (36)–(37) may be viewed as “discrete constitutive law” compared to their continuous counter-part (4) and (21)–(22), respectively.

Space discretization

Concerning the space discretization, the displacement space, \mathcal{U} , is replaced a finite element subset: $\mathcal{U}_h \subset \mathcal{U}$. The equilibrium equations (3) is imposed in the weaker sense:

$$\int_{\Omega_h} \tau : \epsilon(\underline{v}) \, d\Omega_h - \int_{\Omega} \underline{f}_d \cdot \underline{v} \, d\Omega - \int_{\partial_2 \Omega} \underline{F}_d \cdot \underline{v} \, dS = 0, \quad \forall \underline{v} \in \mathcal{U}_{h,0} \quad (39)$$

where

$$\mathcal{U}_{h,0} = \{\underline{v} \in \mathcal{U}_h \text{ such that } \underline{v} = 0 \text{ on } \partial_1 \Omega\} \quad (40)$$

The notation Ω_h , instead of Ω , in the first term of (39) expresses the fact that due to the nonlinear stress-strain relation, this term cannot be in general integrated exactly and therefore is expressed in term of a numerical integration.

iterative technique

The system to be solved at each time step is an algebraic nonlinear system. This system is solved using an iterative technique, typically Newton’s method. As the iterations proceed, two types of solutions cohabit: a solution satisfying the discrete kinematic and static equations, and a solution satisfying the discrete kinematic and constitutive relation. When the difference between the two types of solutions is under a given tolerance δ_{tol} , for instance in the L^2 norm of the difference in stresses (choice for the numerical experiments to be shown later), the iterative procedure stops and the finite element code generally gives the solution satisfying the discrete constitutive relation.

4.2. Construction of an admissible solution

From the finite element solution, one must construct an admissible solution prior to evaluate the error. For the Drucker error, the admissibility is defined by (15) and the error is measured by (17)–(18). For the dissipation error, the admissibility is defined by (28)–(29)–(30) and the error is measured by (32)–(33). The construction techniques are detailed in [31,32] for the Drucker error and in [35] for the dissipation error.

4.3. Examples: two plane stress problems

Here are two examples of error estimation on problems for which the discretization is coarse. These discretizations will be enhanced in the last section of this paper using the tools described in the next section. The parameter θ in the mid-point rule integration, (37), is taken as $\theta = 1$.

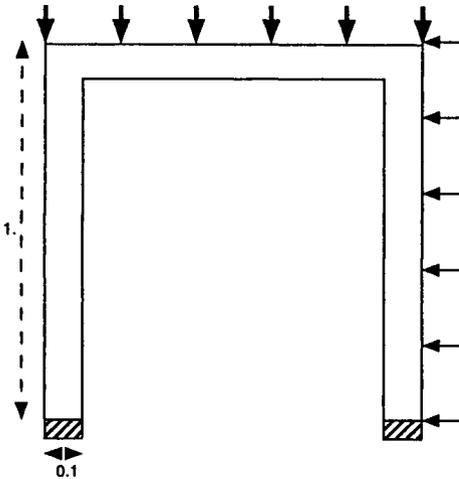


Figure 1. The geometry and applied loading on the frame.

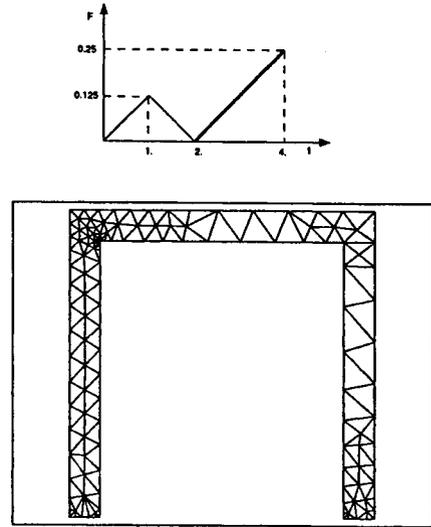


Figure 2. Evolution of the loading and initial mesh optimized in elasticity, 194 six-node triangles.

A frame problem

A frame, Fig. 1, is submitted to a growing then decreasing load on its right side (thin line in Fig. 2). Then, an increasing pressure is applied to its upper part (thick line in Fig. 2). The Prandtl-Reuss plastic model with linear hardening is considered. The dimensionless material parameters are $R_0 = 1$, $k_y = 8.16$, $E = 244.95$ and $\nu = 0.3$. E and ν denote Young's modulus and Poisson's ratio, respectively. The mesh has been optimized in elasticity for an error of 10%, Fig. 2. It is composed of 194 six-node triangular elements. Three time steps are used and the tolerance for the iterative technique is set to $\delta_{tol} = 10^{-2}$. The errors obtained are 28.15% for the dissipation error and 26.90% for the Drucker error.

Extension of a strip weakened by an angular notch

Fig. 3 shows the geometry and the applied loading which evolves monotonically, Fig. 5. We consider the same plastic material as for the previous example and its viscoplastic version with $n_v = 2$ and $k_v = 2.25$. The associate one dimensional behaviors are compared Fig. 6 at a given strain rate speed, $0.05s^{-1}$. The mesh used, Fig. 4, is composed of 88 three-node triangular elements and was optimized in elasticity for 10% error. A single time-step is used and the tolerance is $\delta_{tol} = 10^{-3}$. The dissipation errors obtained are 24% and 17.79% in plasticity and visco-plasticity, respectively. Concerning the Drucker

error, we obtain 27.55% and 26.26%, respectively.

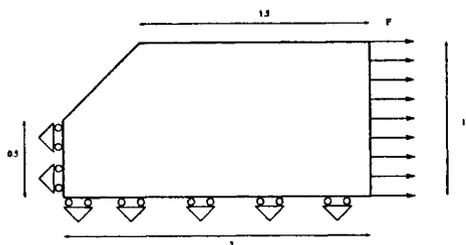


Figure 3. The geometry and applied loading for the angular notch.

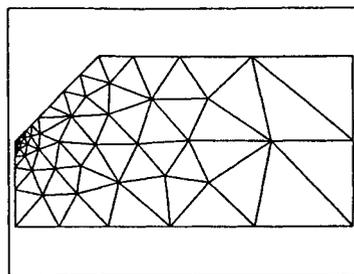


Figure 4. Initial mesh optimized in elasticity for 10% error.

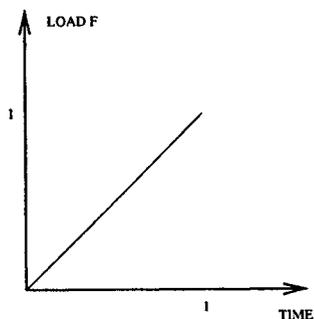


Figure 5. Evolution of the loading.

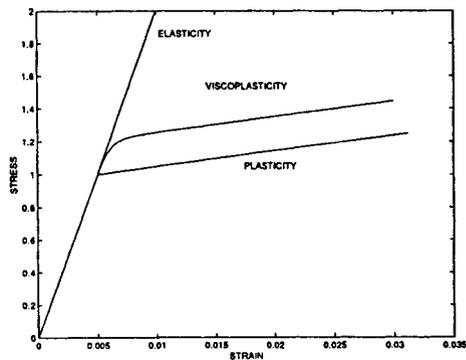


Figure 6. Stress-strain curves for three different models. The viscoplasticity curve corresponds to a strain speed of $0.05s^{-1}$.

5. ERROR INDICATORS

Consider an initial coarse computation. To know how the mesh and the time discretization need to be refined to meet a prescribed accuracy, the single total error information is not enough. One needs to know the contribution to the error of the mesh, the time discretization and the iterative technique.

These contributions are estimated through quantities called error indicators: time, space and iteration error indicator. Time and space error indicator were developed in [38,

32] for the Drucker error and in [35] for the dissipation error. An iteration error indicator was also proposed in [36].

To introduce these indicators, four types of problem are distinguished: the continuous time-space problem (the model problem), the space discrete problem (continuous in time), the time discrete problem (continuous in space) and the time-space discrete problem.

5.1. Time error indicator

When the reference is the continuous time-space problem, the error associated to the finite element solution is due to the space discretization, the time discretization and the iterative technique. We shall call this error the total error. Let us now suppose that the mesh size tends to zero. The total error then tends to a limit that we shall call the time error. The time error is estimated through a quantity called the time error indicator. To compute the time error indicator, the reference is taken as the space discrete problem, so that the error associated to the finite element solution is now only due to the time discretization and the iterative technique. The equations describing the space discrete problem are divided into two groups. The first group gathers the kinematic constraints, the *weak* form of the equilibrium equations (plus the state equations and the initial conditions for the dissipation approach) and the second group gathers the rest of the equations: functional constitutive relation for the Drucker approach and evolution laws for the dissipation approach. The time error indicator is measured on an admissible solution, i.e., a solution satisfying the first group, by the non-fulfillment of the second group of equations. In other words, the time error indicator is obtained using the same strategy as for the total error except that the reference problem is different.

5.2. Space error indicator

The space error is defined as the limit of the total error as the size of the time steps tends to zero. It is estimated by a quantity called the space error indicator. In [32], the space error indicator for the Drucker approach was basically obtained by the difference between the total error and the time error indicator. Another way, is to take the reference problem as the time discrete problem and to proceed as for the time error indicator. The difficulty is that in this problem, the constitutive law has now a discrete form. It turns out that with the generalized mid-point rule the discrete evolution laws (37) are still written in a standard way so that that the error may still be measured using the η quantity (26), see [35]. Concerning the discrete constitutive relation (35), a question arise: under what conditions does the Drucker stability condition still apply and in what sense? A sufficient condition is given below.

Let (σ_k, ϵ_k) and (σ'_k, ϵ'_k) be two stress-strain sequences defined at every instant t_k , $k \in \{0 \dots N + 1\}$, satisfying the discrete constitutive relation (35) with $\sigma_0 = \epsilon_0 = \sigma'_0 = \epsilon'_0 = 0$. We shall say that a discrete constitutive relation, is strictly stable in the Drucker sense if there exist a parameter α , $0 \leq \alpha \leq 1$, such that the two following conditions are satisfied:

$$\forall n \in \{0 \dots N\}, \sum_{k=0}^n (\sigma_\alpha - \sigma'_\alpha) : (\dot{\epsilon}_{k+1} - \dot{\epsilon}'_{k+1}) \Delta t_k \geq 0 \quad (41)$$

$$\begin{aligned} \forall n \in \{0 \cdots N\}, \sum_{k=0}^n (\boldsymbol{\sigma}_\alpha - \boldsymbol{\sigma}'_\alpha) : (\dot{\boldsymbol{\epsilon}}_{k+1} - \dot{\boldsymbol{\epsilon}}'_{k+1}) \Delta t_k = 0 \\ \Leftrightarrow (\boldsymbol{\sigma}_k, \boldsymbol{\epsilon}_k) = (\boldsymbol{\sigma}'_k, \boldsymbol{\epsilon}'_k), \forall k \in \{0 \cdots N + 1\} \end{aligned} \quad (42)$$

The notations (38) have been reused, θ being replaced by α , and $\Delta t_k = t_{k+1} - t_k$.

A sufficient condition to have strict stability is to define the operator \mathcal{A}_n as the one obtained by eliminating the internal and state variables in the discrete internal variable formulation (36)–(37), with $1/2 \leq \theta \leq 1$. The parameter α is then simply θ . The key for the proof is the following, easy to establish, relation:

$$\begin{aligned} (\boldsymbol{\sigma}_\theta - \boldsymbol{\sigma}'_\theta) : (\dot{\boldsymbol{\epsilon}}_{k+1} - \dot{\boldsymbol{\epsilon}}'_{k+1}) = \eta(\dot{\boldsymbol{\epsilon}}_{k+1}^p, \dot{\boldsymbol{X}}_{k+1}, \boldsymbol{\sigma}'_\theta, \boldsymbol{Y}'_\theta) + \eta(\dot{\boldsymbol{\epsilon}}_{k+1}^{p'}, \dot{\boldsymbol{X}}'_{k+1}, \boldsymbol{\sigma}_\theta, \boldsymbol{Y}_\theta) + \\ \mathbf{K}^{-1}(\boldsymbol{\sigma}_\theta - \boldsymbol{\sigma}'_\theta) : (\dot{\boldsymbol{\sigma}}_{k+1} - \dot{\boldsymbol{\sigma}}'_{k+1}) + \Lambda^{-1}(\boldsymbol{Y}_\theta - \boldsymbol{Y}'_\theta) \circ (\dot{\boldsymbol{Y}}_{k+1} - \dot{\boldsymbol{Y}}'_{k+1}) \end{aligned} \quad (43)$$

where the set of prime and non-prime variables satisfy the system (36)–(37). Taking into account the zero initial boundary conditions and the algebraic relation

$$x_\theta = \frac{1}{2}(x_{k+1} + x_k) + \frac{1}{2}(2\theta - 1)(x_{k+1} - x_k) \quad (44)$$

we obtain by summing the relation (43) over the time steps

$$\begin{aligned} \sum_{k=0}^n (\boldsymbol{\sigma}_\theta - \boldsymbol{\sigma}'_\theta) : (\dot{\boldsymbol{\epsilon}}_{k+1} - \dot{\boldsymbol{\epsilon}}'_{k+1}) \Delta t_k = \\ \sum_{k=0}^n \left(\eta(\dot{\boldsymbol{\epsilon}}_{k+1}^p, \dot{\boldsymbol{X}}_{k+1}, \boldsymbol{\sigma}'_\theta, \boldsymbol{Y}'_\theta) + \eta(\dot{\boldsymbol{\epsilon}}_{k+1}^{p'}, \dot{\boldsymbol{X}}'_{k+1}, \boldsymbol{\sigma}_\theta, \boldsymbol{Y}_\theta) \right) \Delta t_k + \\ |\boldsymbol{\sigma}_{n+1} - \boldsymbol{\sigma}'_{n+1}|_{\mathbf{K}}^2 + |\boldsymbol{Y}_{n+1} - \boldsymbol{Y}'_{n+1}|_{\Lambda}^2 + \\ (2\theta - 1) \sum_{k=0}^n \left(|\dot{\boldsymbol{\sigma}}_{k+1} - \dot{\boldsymbol{\sigma}}'_{k+1}|_{\mathbf{K}}^2 + |\dot{\boldsymbol{Y}}_{k+1} - \dot{\boldsymbol{Y}}'_{k+1}|_{\Lambda}^2 \right) \Delta t_k^2 \end{aligned}$$

The result stem from this last relation which is positive for $\theta \geq 1/2$ and is zero if and only if the prime and non-prime variables coincide.

Let $(\hat{\boldsymbol{u}}_k, \hat{\boldsymbol{\sigma}}_k)$ be a displacement–stress couple defined on Ω for all the time instant t_k , $k \in \{0 \cdots N + 1\}$ ($\{t_k\}$ in short). We assume that this couple is admissible for the time-discrete problem, i.e.

$$\hat{\boldsymbol{u}}_k \in \mathcal{U}_{ad}^{\{t_k\}} = \{\boldsymbol{v} \in \mathcal{U}^{\{t_k\}} \text{ such that } \boldsymbol{v} = \boldsymbol{u}_d \text{ on } \partial_1 \Omega \times \{t_k\}\} \quad (45)$$

$$\hat{\boldsymbol{\sigma}}_k \in \mathcal{S}_{ad}^{\{t_k\}} = \{\boldsymbol{\tau} \in \mathcal{S}^{\{t_k\}} \text{ satisfying (3)} \forall t \in \{t_k\}\} \quad (46)$$

This couple satisfies all the equations of the time discrete problem except, in general, the discrete constitutive relation (35). It is an approximate solution to the time discrete problem. To the strain field, $\hat{\boldsymbol{\epsilon}}_k = \boldsymbol{\epsilon}(\hat{\boldsymbol{u}}_k)$, one can associate the stress field $\hat{\boldsymbol{\sigma}}'_k$ using the relation (35). Similarly, one can associate the strain field $\hat{\boldsymbol{\epsilon}}'_k$ to the stress field $\hat{\boldsymbol{\sigma}}_k$ using the reverse relation. The absolute space error indicator, I_{spce} , may be defined by

$$I_{\text{spce}}^2 = \sup_{n \in \{0 \cdots N\}} \int_{\Omega} \sum_{k=0}^n (\hat{\boldsymbol{\sigma}}_\theta - \hat{\boldsymbol{\sigma}}'_\theta) : (\dot{\boldsymbol{\epsilon}}'_{k+1} - \dot{\boldsymbol{\epsilon}}_{k+1}) \Delta t_k \, d\Omega \quad (47)$$

and the relative space error indicator by $i_{spce} = I_{spce}/D_{spce}$ where

$$D_{spce}^2 = 2 \sup_{n \in \{0 \dots N\}} \int_{\Omega} \sum_{k=0}^n (\hat{\sigma}_{\theta} : \dot{\epsilon}'_{k+1} + \hat{\sigma}'_{\theta} : \dot{\epsilon}_{k+1}) \Delta t_k \, d\Omega \tag{48}$$

5.3. Iteration error indicator

The iteration error is defined as the limit of the total error as both the size of the elements and the size of the time steps tend to zero. It is estimated by the iteration error indicator. The reference is taken as the fully discrete space-time problem, so that the error associated to the finite element solution is only due to the iterative technique. This lead to the iteration error indicator introduced in [36] for the dissipation approach. Concerning the Drucker approach, expressions similar to (47)–(48) are used.

		Dissipation					
		mesh1	mesh2	mesh3	mesh4	mesh5	mesh6
ϵ		40.94	28.99	25.74	23.96	20.80	18.92
i_{time}		19.29	18.33	17.68	17.65	17.65	17.65
i_{spce}		34.08	15.42	10.90	8.40	4.05	1.61
i_{tot}		53.37	33.75	28.58	26.05	21.70	19.26
$i_{ite} * 10^2$		0.38	0.16	0.07	0.17	0.33	0.53
		Drucker					
		mesh1	mesh2	mesh3	mesh4	mesh5	mesh6
ϵ		44.75	27.12	23.57	20.89	16.83	15.14
i_{time}		15.23	14.70	14.45	14.46	14.45	14.44
i_{spce}		44.67	23.56	19.27	15.69	9.06	4.88
i_{tot}		47.19	27.77	24.09	21.34	17.06	15.24
$i_{ite} * 10^2$		0.80	0.46	0.15	0.31	0.51	0.76

Table 1
Influence of a growing quality of the space discretization on the error and the indicators. The computations were carried out with a stopping criteria of $\delta_{tot} = 10^{-3}$.

5.4. Numerical experiments

We present here numerical experiments to analyze the behavior of the error and the indicators as either the mesh, the time discretization of the stopping criteria is modified. We consider the frame problem. All the results are given in relative % values. Table 1 shows the influence of the quality of the mesh and Table 2 the influence of the quality of the time discretization. i_{time} , i_{spce} and i_{ite} denote respectively the time, space and iteration error indicator Mesh1, mesh2, ... are increasingly refined meshes. To see how the time and space indicators are related to the total error, ϵ , we have computed the quantity i_{tot} which is defined by $i_{tot} = i_{time} + i_{spce}$ for the dissipation approach and by $i_{tot} = \sqrt{i_{time}^2 + i_{spce}^2}$ for the Drucker approach. We see that

nb steps	Dissipation					Drucker				
	3	6	12	24	48	2	4	8	16	32
ϵ	25.74	16.36	13.85	13.76	14.00	23.57	20.91	19.79	19.72	19.80
i_{time}	17.69	4.87	1.00	0.34	0.21	14.43	8.50	3.97	2.02	1.04
i_{spce}	10.91	11.67	12.48	13.39	13.87	19.27	19.72	19.96	20.15	20.21
i_{tot}	28.60	16.54	13.48	13.73	14.08	24.07	21.47	20.35	20.25	20.24
$i_{ite} * 10^2$	2.89	6.48	7.83	10.25	14.39	5.78	10.25	14.75	16.67	18.15

Table 2

Influence of a growing quality of the time discretization on the error and the indicators. The computations were carried out with a stopping criteria of $\delta_{tol} = 10^{-2}$.

δ_{tol}	Dissipation					
	10^0	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
ϵ	26.60	16.42	16.36	16.37	16.37	16.37
i_{time}	21.62	5.32	4.87	4.82	4.82	4.82
i_{spce}	22.74	11.75	11.67	11.67	11.67	11.67
i_{ite}/δ_{tol}	17.78	5.59	6.48	4.48	5.08	5.17
ϵ	Drucker					
	10^0	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
ϵ	27.78	20.94	20.91	20.91	20.91	20.91
i_{time}	20.57	8.57	8.50	8.50	8.50	8.50
i_{spce}	28.41	19.74	19.72	19.72	19.72	19.72
i_{ite}/δ_{tol}	21.25	9.30	10.25	6.57	8.55	10.71
Nb ite	28	72	106	130	154	180

Table 3

Influence of the stopping criteria on the error and indicators for a coarse time-space discretization.

Dissipation						
δ_{tol}	10^0	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
ϵ	12.40	6.05	5.37	5.36	5.36	5.36
i_{time}	10.74	2.39	0.64	0.55	0.55	0.55
i_{spce}	12.59	5.13	4.47	4.47	4.47	4.47
$i_{\text{ite}}/\delta_{\text{tol}}$	11.09	17.01	8.23	6.20	7.54	8.04
Drucker						
ϵ	15.26	9.42	9.36	9.35	9.34	9.34
i_{time}	12.68	3.54	3.07	3.07	3.07	3.07
i_{spce}	16.68	9.55	9.38	9.37	9.37	9.37
$i_{\text{ite}}/\delta_{\text{tol}}$	14.24	25.69	16.96	12.97	14.51	15.03
Nb ite	40	112	194	262	322	386

Table 4

Influence of the stopping criteria on the error and indicators for a fine time-space discretization.

- first, the error decreases but then tends to stabilize;
- the time indicator is almost insensitive to the quality of the mesh and approximate very well the time error;
- the space indicator barely depend on the number of time steps and approximate very well the space error;
- the space and time indicators decrease monotonically with respect to the quality of the mesh and the number time steps, respectively;
- the total error, ϵ , is close or bounded from above by the quantity i_{tot} ;
- for the stopping criteria chosen, the iteration error indicator, i_{ite} , is very small compared to the space and time indicators.

Table 3 and 4 give the influence of the stopping criteria for a coarse and fine space-time discretization, respectively. N_{ite} is the total number of iteration in the computation. We note that

- the error decreases but then quickly stabilizes;
- under a critical stopping criteria, δ_{crit} , the error and the space and time indicators are insensitive to the stopping criteria;
- the critical stopping criteria is smaller for a finer time-space discretization;
- the iteration indicator is more or less linear with respect to the stopping criteria;
- N_{ite} grows roughly logarithmically with the decrease of the stopping criteria.

6. ADAPTIVE CONTROL

Using the Drucker error, an adaptive strategy to control the space and time discretization for three- and six-node triangular elements in plane and axisymmetric problems was presented in [32]. Concerning the dissipation error, a simultaneous time, space and iteration adaptive scheme was presented in [36]. We summarize here the basic features of the strategy. Basically, the idea is to meet the prescribed accuracy while minimizing the numerical effort. The prescribed accuracy is linked to the time and space indicators through the i_{tot} quantity. The numerical effort is expressed as the number of time steps times the number of elements raised to a power (2 in the examples).

The time and space indicators need to be related to the mesh size and the time step size. We assume that the space error indicator has a local convergence $O(h^p)$ (from numerical experiments, $p = 1$ and $p = 2$ for three- and six-node triangular elements, respectively), h being the local size of the element. Concerning the time error indicator, we assume a local convergence $O(\Delta t^q)$ ($q = 1$ and $q = 2$ for the Drucker and dissipation time error indicator, respectively), Δt being the local size of the time step.

Minimizing the numerical in order to reach the prescribed accuracy yield a new mesh size map and a new time step size map, which are used to remesh in space and "time" the problem. When the p and q parameters are uniform, the minimization of the numerical effort implies that the prescribed accuracy, ϵ_0 , should be shared in the ratios

$$i_{\text{spce},0} = \frac{4/p}{4/p + 1/q} \epsilon_0, \quad i_{\text{time},0} = \frac{1/q}{4/p + 1/q} \epsilon_0 \quad (49)$$

for the objective space and time error indicators respectively, in the dissipation error case and in the ratios

$$i_{\text{spce},0} = \left(\frac{4/p}{4/p + 1/q}\right)^{1/2} \epsilon_0, \quad i_{\text{time},0} = \left(\frac{1/q}{4/p + 1/q}\right)^{1/2} \epsilon_0 \quad (50)$$

for the Drucker case.

The stopping criteria for the iterative procedure is chosen such that the iteration error indicator is about on tenth of the minimum of the space and time error indicators. This allows that enough - and not too many - iterations are made to avoid the influence of the non-convergence of the iterative technique.

Finally, note that when the prescribed accuracy is too far from the actual error obtained with the initial coarse computation, some intermediate steps are introduced in the adaptive strategy.

The frame example

Tables 5 and 6 summarize the adaptive procedure using the dissipation and Drucker error, respectively. The prescribed accuracy is 5% and six-node triangular elements are used. We note that the numerical effort needed to reach the 5% error is more important in the Drucker error sense than in the dissipation error sense. The optimized meshes at the last step are shown Fig. 7 and 8 and the evolutions of the contribution to the error with respect to the time for the successive stages, Fig.7. These contributions were defined by equations (19) and (34). Note that the time stepping, indicated by the squares, stars, ... is

not uniform and takes into account the unloading. The Von Mises stresses obtained at the final time are shown Fig.10. They are compared to the Von Mises stresses obtained with the Hencky–Mises model, Fig.11. Since the loading does not evolve monotonically, the difference is important. Indeed, the Hencky–Mises model assumes a simple monotonous evolution of the stresses which is obviously not the case here.

	ϵ	i_{spce}	i_{time}	i_{ite}	nb elts	nb steps	δ_{tol}
Initially	28.15	14.33	17.90	0.03	194	3	1.10^{-2}
Asked	10	8.	2.	0.20			
Obtained	12.30	10.03	2.18	0.38	253	7	$7.41 \cdot 10^{-2}$
Asked	5	4.	1.	0.10			
Obtained	5.37	4.48	0.74	0.17	661	11	$1.91 \cdot 10^{-2}$

Table 5

Summary of the optimization procedure using the dissipation error for the frame problem.

	ϵ	i_{spce}	i_{time}	i_{ite}	nb elts	nb steps	δ_{tol}
Initially	26.90	23.26	14.56	0.06	194	3	1.10^{-2}
Asked	15.	12.2	8.7	0.9			
Obtained	17.02	15.34	8.45	0.77	262	5	$16.2 \cdot 10^{-2}$
Asked	10.	8.2	5.8	0.6			
Obtained	10.54	10.29	4.44	1.8	502	7	$12.5 \cdot 10^{-2}$
Asked	5.	4.1	2.9	0.3			
Obtained	5.75	5.69	2.63	0.40	1286	12	$2.07 \cdot 10^{-2}$

Table 6

Summary of the optimization procedure using Drucker error for the frame problem.

Extension of a strip weakened by an angular notch

Finally, we consider the strip weakened by an angular notch with three different models: the elastic model, the Prandtl–Reuss plastic model and its viscoplastic version. Table 7 summarizes the adaptive procedure in the plastic and viscoplastic cases, using the Drucker error and three-node triangular elements. The prescribed accuracy is again 5%. For an elastic problem, the Drucker error coincide with the classical energy norm error used in elasticity, see [32]. The Von Mises stresses obtained at the final time are quite different for the three models, Fig.12, 14 and 16. The optimized meshes are also quite different, Fig. 13, 15 and 17.

Plastic computation							
	ϵ	i_{space}	i_{time}	i_{ite}	nb elts	nb steps	δ_{tol}
Initially	27.55	21.49	20.53	0.44	88	1	1.10^{-3}
Asked	10.	8.9	4.4	0.4			
Obtained	9.68	10.13	4.86	3.9	411	5	9.10^{-3}
Asked	5.	4.4	2.2	0.2			
Obtained	4.81	4.74	2.12	0.26	1744	7	4.610^{-3}
Viscoplastic computation							
	ϵ	i_{space}	i_{time}	i_{ite}	nb elts	nb steps	δ_{tol}
Initially	26.26	16.95	22.42	0.04	88	1	1.10^{-3}
Asked	10.	8.9	4.4	0.4			
Obtained	9.47	8.40	5.78	1.9	296	5	10.10^{-3}
Asked	5.	4.4	2.2	0.2			
Obtained	4.73	4.44	1.94	0.58	982	8	1.10^{-3}

Table 7

Summary of the optimization procedure using Drucker error for the angular notch problem.

7. CONCLUSIONS

The error in the constitutive law concept allows to define error with strong mechanical foundations. When the constitutive law is characterized by a functional stress-strain formulation, the error is based on a sufficient condition insuring the stability of the material, Drucker's stability condition. When the constitutive law is described in terms of internal variables, the error is measured on the equations associated the dissipation process in the material, i.e. the evolution laws. An extension of the famous Prager-Synge theorem may be obtained in this later case.

The error itself is not enough to carry an efficient adaptive procedure. The contributions of each error source to the error are also needed. These contributions are estimated through quantities called error indicators: time, space and iteration error indicators. They are defined in the same way as the error except that the reference problem is different: space discrete, time discrete and space-time discrete problems, respectively. The behavior of the indicators was shown to be robust.

When used along with the error, the error indicators allows to adapt simultaneously the quality of the time and space discretization, and the choice of the stopping criteria for the iterative technique. This was illustrated on two examples. The first examples was the occasion to recall the danger in using an Hencky-Mises formulation when the loading is not evolving monotonically. We also noticed that the numerical effort needed to reach a given accuracy was more important in the Drucker approach than in the dissipation approach. This point, which is not surprising since the foundations of the two errors are different, should be investigated in future works.

REFERENCES

1. P. Ladevèze. *Les bases de la méthode de l'erreur en relation de comportement pour le contrôle des calculs éléments finis: Les travaux des années 1975*. Rapport interne 163, LMT, Ecole Normale Supérieure de Cachan, 1995.
2. P. Ladevèze and D. Leguillon. Error estimate procedure in the finite element method and application. *SIAM J. Num. Anal.*, 20(3):485–509, 1983.
3. P. Ladevèze, J. P. Pelle, and P. Rougeot. Error estimation and mesh optimization for classical finite elements. *Engrg. Computation*, 8:69–80, 1991.
4. I. Babuska and W. C. Rheinboldt. A posteriori estimates for the finite element method. *Int. J. for Num. Meth. in Engrg.*, 12:1597–1615, 1978.
5. I. Babuska and W. C. Rheinboldt. Adaptive approaches and reliability estimators in finite element analysis. *Comp. Meth. in Applied Mech. and Engrg.*, 17/18:519–540, 1979.
6. D. W. Kelly, J. Gago, O. C. Zienkiewicz, and I. Babuska. A posteriori error analysis and adaptive processes in finite element method. part 1: Error analysis. *Int. J. for Num. Meth. in Engrg.*, 19:1593–1619, 1983.
7. J. Gago, D. W. Kelly, O. C. Zienkiewicz, and I. Babuska. A posteriori error analysis and adaptive processes in finite element method. part 2: Adaptive mesh refinement. *Int. J. for Num. Meth. in Engrg.*, 19:1921–1656, 1983.
8. J. T. Oden, L. Demkowicz, W. Rachowicz, and T. A. Westermann. Toward a universal h-p adaptive finite element strategy, part 2: A posteriori error estimation. *Comp. Meth. in Applied Mech. and Engrg.*, 77:113–180, 1989.
9. M. Ainsworth and J. T. Oden. A posteriori error estimation in finite element analysis. *Comp. Meth. in Applied Mech. and Engrg.*, 142(1-2):1–88, 1997.
10. O. C. Zienkiewicz and J. Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis. *Int. J. for Num. Meth. in Engrg.*, 24:337–357, 1987.
11. O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and a posteriori error estimates, part 1: The recovery technique. *Int. J. for Num. Meth. in Engrg.*, 33:1331–1364, 1992.
12. O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and a posteriori error estimates, part 2: Error estimates and adaptativity. *Int. J. for Num. Meth. in Engrg.*, 33:1365–1382, 1992.
13. I. Babuska, T. Strouboulis, C. S. Upadhyay, S. K. Gangaraj, and K. Copps. Validation of a posteriori error estimators by numerical approach. *Int. J. for Num. Meth. in Engrg.*, 37:1073–1123, 1994.
14. I. Babuska, T. Strouboulis, and C. S. Upadhyay. A model study of the quality of a posteriori error estimators for linear elliptic problems: Error estimation in the interior of patchwise uniform grids of triangles. *Comp. Meth. in Applied Mech. and Engrg.*, 114:307–378, 1994.
15. I. Babuska and W. C. Rheinboldt. Computational error estimates and adaptive processes for some nonlinear structural problems. *Comp. Meth. in Applied Mech. and Engrg.*, 34:895–937, 1982.
16. C. Johnson and P. Hansbo. Adaptive finite element methods in computational mechanics. *Comp. Meth. in Applied Mech. and Engrg.*, 101:143–181, 1992.

17. R. Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques*. Wiley-Teubner, ISBN 0-471-96795-5, 1996.
18. J. M. Bass and J. T. Oden. Adaptive finite element methods for a class of evolution problems in viscoplasticity. *Int. J. Engrg Sci.*, 25(6):623-653, 1987.
19. D. Aubry and B. Tie. Relationships between error estimation and adaptive computations in strain localization. In *Proceedings of the 9th Conference on Engineering Mechanics, College Station*, 1992.
20. D. Peric, J. Yu, and D. R. J. Owen. On error estimates and adaptativity in elastoplastic solids: Applications to the numerical simulation of strain localisation in classical and Cosserat continua. *Int. J. for Num. Meth. in Engrg.*, 37(8):1351-1379, 1994.
21. G. Pijaudier-Cabot, L. Bode, and A. Huerta. Arbitrary lagrangian-eulerian finite element analysis of strain localization in transient problems. *Int. J. for Num. Meth. in Engrg.*, 38(24):4171-4191, 1995.
22. P. Bussy. *Optimisation et fiabilité des calculs par éléments finis en non linéarité géométrique*. Thèse, Université de Pierre et Marie Curie, 1984.
23. P. Bussy, G. Coffignal, and P. Ladevèze. Reliability and optimization of non linear finite element analyses. In *Proc. of NUMETA '85, Swansea*, 1985.
24. O.C. Zienkiewicz and G.C. Huang. Note on localization phenomena and adaptive finite-element analysis in forming processes. *Communications in Applied Numerical Methods*, 6(2):71-76, 1990.
25. L. Fourment and J.L. Chenot. Error estimators for viscoplastic materials: application to forming processes. *Int. J. for Num. Meth. in Engrg.*, 12(5):469-490, 1995.
26. J. Baranger, K. Najib, and D. Sandri. Numerical analysis of a three-fields model for a quasi-newtonian flow. *Comp. Meth. in Applied Mech. and Engrg.*, 109(3-4):281-292, 1993.
27. J. T. Oden, W. Wu, and M. Ainsworth. An a posteriori error estimate for the finite element approximations of the navier-stokes equations. *Comp. Meth. in Applied Mech. and Engrg.*, 111:185-202, 1994.
28. R. Becker and R. Rannacher. Weighted a posteriori error control in FE methods. In *Proc. ENUMATH-95, Paris*, 1995.
29. D. C. Drucker. On the postulate of stability of material in the mechanics of continua. *Journal de Mécanique*, 3(2):235-249, 1964.
30. P. Ladevèze. Sur une famille d'algorithmes en mécanique des structures. *Compte-Rendus de l'Académie des Sciences, Série II*, 300:41-44, 1985.
31. P. Ladevèze, G. Coffignal, and J. P. Pelle. Accuracy of elastoplastic and dynamic analysis. In I. Babuska, J. Gago, Oliveira, and O. C. Zienkiewicz, editors, *Accuracy estimates and adaptive refinements in Finite Element computations*, pages 181-203. J. Wiley, 1986.
32. L. Gallimard, P. Ladevèze, and J. P. Pelle. Error estimation and adaptativity in elastoplasticity. *Int. J. for Num. Meth. in Engrg.*, 39:189-217, 1996.
33. P. Ladevèze. La méthode à grand incrément de temps pour l'analyse de structures à comportement non linéaires décrit par variables internes. *Compte-Rendus de l'Académie des Sciences, Série II*, 309(11):1095-1099, 1989.
34. P. Ladevèze. *Mécanique non linéaire des structures. Nouvelle approche et méthodes de calcul non incrémentales*. Hermes, 1996.

35. P. Ladevèze and N. Moës. A new a posteriori error estimation for nonlinear time-dependent finite element analysis. *Comp. Meth. in Applied Mech. and Engrg.* to appear.
36. P. Ladevèze and N. Moës. Adaptive control for time-dependent nonlinear finite element analysis. *Computers and Structures*. submitted.
37. P. Ladevèze. Nouvelles procédures d'estimations d'erreur relative à la méthode des éléments finis. In *proc. Journées Eléments finis, Rennes, 1977*.
38. G. Coffignal. *Optimisation et fiabilité des calculs éléments finis en élastoplasticité*. Thèse d'état, Université de Pierre et Marie Curie, 1987.
39. W. Prager and J. L. Synge. Approximation in elasticity based on the concept of functions space. *Quart. Appl. Math.*, 5:261-269, 1947.



Figure 7. Final mesh optimized with the dissipation error, 661 six-node triangles.

Figure 8. Final mesh optimized with Drucker error, 1286 six-node triangles.

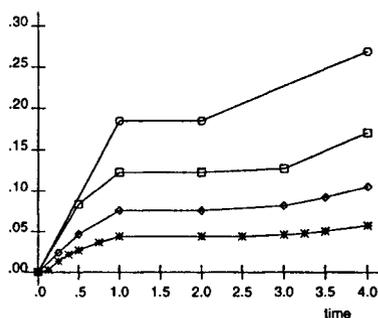
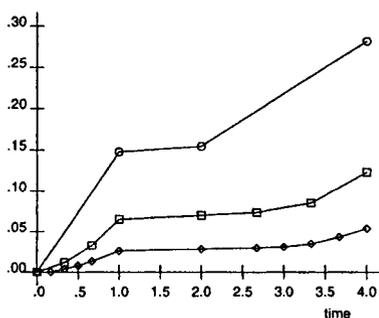


Figure 9. Evolution of the contribution to the error with respect to the time for the successive adaptive steps, dissipation error on the left and Drucker error on the right.

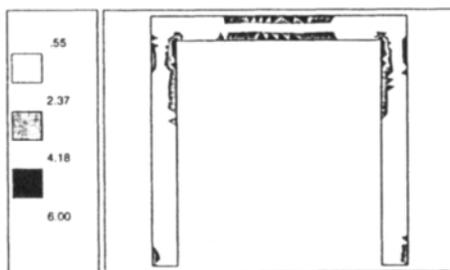
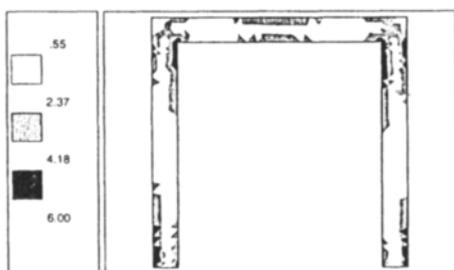


Figure 10. Von Mises stresses obtained with the Prandtl-Reuss plastic model.

Figure 11. Von Mises stresses obtained with the Hencky-Mises model.

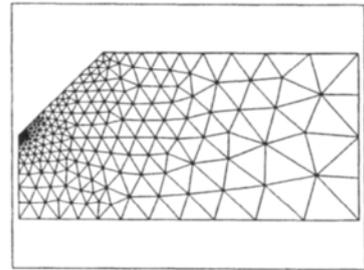
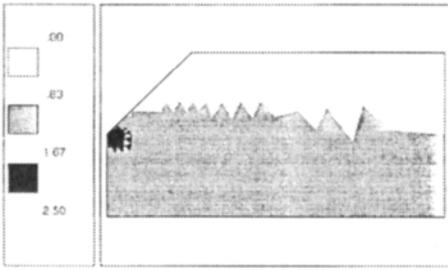


Figure 12. Von Mises stresses in elasticity.

Figure 13. Optimized mesh: 456 three-nodes triangles.

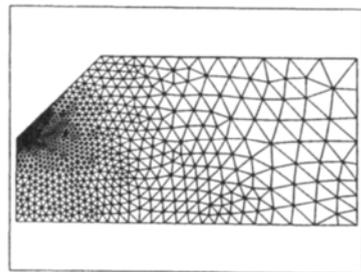
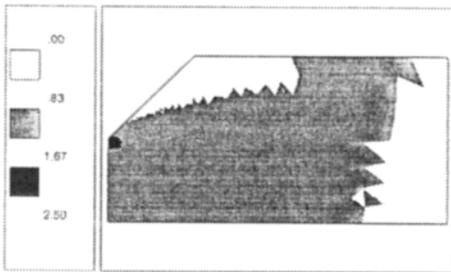


Figure 14. Von Mises stresses in plasticity.

Figure 15. Optimized mesh: 1744 three-nodes triangles.

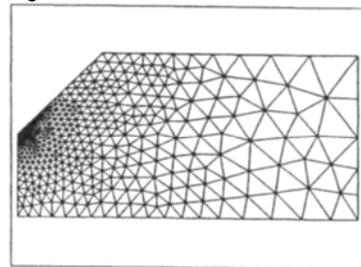
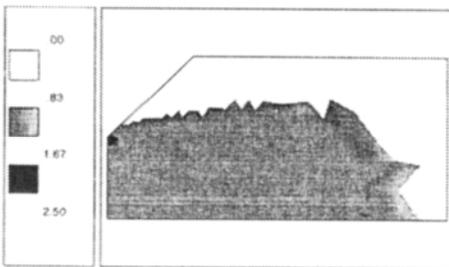


Figure 16. Von Mises stresses in viscoplasticity.

Figure 17. Optimized mesh: 982 three-nodes triangles.

A Review of A Posteriori Error Estimation Techniques for Elasticity Problems

R. Verfürth

Fakultät für Mathematik, Ruhr-Universität Bochum, D-44780 Bochum, Germany
E-mail address: rv@silly.num1.ruhr-uni-bochum.de

1. INTRODUCTION

Adaptive mesh-refinement techniques based on a posteriori error estimators have become an indispensable tool in large scale scientific computing. In what follows we will present some of the most popular error estimators and will discuss their common features, their main differences, and the underlying basic principles.

The essential point will be the equivalence of the norm of the error, which should be estimated, and a corresponding dual norm of the residual, which only involves the given data of the problem and the computed numerical solution. This equivalence is a consequence of the stability of the infinite dimensional variational problem. Thus, in a posteriori error analysis the situation is quite different from a priori error analysis where the stability of the discrete problem is the main ingredient.

The error estimators considered here can be viewed as a device for rendering computable the above-mentioned dual norm of the residual. They differ in the way they try to achieve this goal. We will show that all yield upper and lower bounds on the error and that they are all equivalent in the sense that the ratio of any two of them remains bounded independently of the mesh-size.

The upper bounds on the error are always global ones. This is due to the fact that upper bounds involve the inverse of the differential operator which is a global operator. Some of the lower bounds on the errors are local ones. This is possible since lower bounds involve the original differential operator which is a local operator.

Upper bounds, of course, are mandatory for reliability, i.e. to ensure that the error is below a given tolerance. Lower bounds on the other hand, are indispensable for efficiency, i.e. to ensure that the required tolerance is achieved with a nearly minimal amount of work. Moreover, locality of the lower bounds guarantees that the spatial distribution of the error is properly reproduced by the error estimator. Unfortunately, this point is often not sufficiently taken into account.

In order to simplify the exposition and to keep the technicalities at a minimum we present the error estimators within the framework of displacement methods for linearized elasticity. In the last quarter of this contribution we show how the results and techniques can be extended to displacement methods for non-linear elasticity.

In what follows we only consider displacement methods based on a primal variational principle. We completely omit mixed methods based on a dual or hybrid primal/dual variational principle. Since these methods play an important role, e.g. in the context of locking phenomena, this is a serious omission. It is due to two reasons. First, a profound consideration of dual methods would unduely overload the presentation. Second and more important, the theory of a posteriori error estimators for mixed methods is not as well developed as for primal methods. First approaches can be found, e.g. in [6, 7, 17, 18] and the literature cited therein. But all these are concerned with membrane problems and cannot directly be applied to linearized elasticity.

2. THE EQUATIONS OF LINEARIZED ELASTICITY

We consider the equations of linearized elasticity

$$-\operatorname{div} \sigma(u) = f \text{ in } \Omega; \quad u = u_D \text{ on } \Gamma_D; \quad n \cdot \sigma(u) = t_N \text{ on } \Gamma_N \tag{1}$$

in a connected bounded domain $\Omega \subset \mathbb{R}^n, n = 2$ or 3 , with a polyhedral boundary $\Gamma = \Gamma_D \cup \Gamma_N, \Gamma_D \cap \Gamma_N = \emptyset$. As usual, $u : \Omega \rightarrow \mathbb{R}^n$ denotes the sought displacement, f is the external load, t_N are the prescribed boundary tractions, and n denotes the unit outward normal. The stress tensor $\sigma(u)$ is related to the strain tensor $\varepsilon(u) := (\frac{1}{2}(\partial_i u_j + \partial_j u_i))_{1 \leq i, j \leq n}$ by the material law $\sigma(u)_{ij} = \sum_{1 \leq k, l \leq n} C_{ijkl} \varepsilon(u)_{kl}$ with a symmetric positive definite $n^2 \times n^2$ matrix C . In order to guarantee unique solvability of problem (1) we assume that the Dirichlet part Γ_D of Γ has a positive $(n - 1)$ -dimensional Lebesgue-measure.

For any open subset ω of Ω with Lipschitz boundary γ , we denote by $H^k(\omega), k \in \mathbb{N}, L^2(\omega) = H^0(\omega)$, and $L^2(\gamma)$ the usual Sobolev and Lebesgue-spaces equipped with the standard norms $\|\varphi\|_{k;\omega} := \|\varphi\|_{H^k(\omega)}, \|\varphi\|_\omega := \|\varphi\|_{L^2(\omega)}$, and $\|\varphi\|_\gamma := \|\varphi\|_{L^2(\gamma)}$. If $\omega = \Omega$, we will omit the index ω . We use the same notations for the norms of vector-valued and matrix-valued functions.

Set $V_D := \{v \in H^1(\Omega)^n : v = 0 \text{ on } \Gamma_D\}$. Then the standard weak displacement formulation of problem (1) is to find $u \in u_D + V_D$ such that

$$\int_\Omega \sigma(u) \cdot \varepsilon(v) = \int_\Omega f \cdot v + \int_{\Gamma_N} t_N \cdot v \quad \forall v \in V_D. \tag{2}$$

Since Γ_D has a positive $(n - 1)$ -dimensional Lebesgue-measure, problem (2) admits a unique solution.

3. FINITE ELEMENT DISCRETIZATION

Let $\mathcal{T}_h, h > 0$, be a family of partitions of Ω into triangles or quadrilaterals, if $n = 2$, or into tetrahedra or hexahedra, if $n = 3$. Each \mathcal{T}_h must be consistent with Γ_D and Γ_N , i.e. Γ_D and Γ_N are the union of edges, if $n = 2$, resp. of faces, if $n = 3$, of elements in \mathcal{T}_h . Moreover, \mathcal{T}_h must satisfy the following two conditions:

- (1) *admissibility*: Any two elements of \mathcal{T}_h are either disjoint, or share a vertex, or share an edge or, if $n = 3$, share a face.
- (2) *shape regularity*: $c_{\mathcal{T}} := \sup\{h_K/\rho_K : K \in \mathcal{T}_h, h > 0\} < \infty$.

Here, h_K and ρ_K denote the diameter of K and the diameter of the largest ball inscribed into K , respectively. If $n = 2$, shape regularity means that the smallest angle of all elements is bounded away from zero.

Let $V_h \subset V_D$ be any conforming finite element space associated with \mathcal{T}_h . We assume that it contains all continuous, piecewise linear or n -linear functions. The finite element discretization (in displacement form) of problem (1) then consists in finding $u_h \in u_D + V_h$ such that

$$\int_{\Omega} \sigma(u_h) \cdot \varepsilon(v_h) = \int_{\Omega} f \cdot v + \int_{\Gamma_N} t_N \cdot v_h \quad \forall v_h \in V_h. \quad (3)$$

Since Γ_D has a positive $(n - 1)$ -dimensional Lebesgue-measure problem (3) admits a unique solution.

4. AUXILIARY RESULTS

For $K \in \mathcal{T}_h$ we denote by $\mathcal{N}(K)$ and $\mathcal{E}(K)$ the set of its vertices and the set of its edges, if $n = 2$, or of its faces, if $n = 3$, respectively. Let $\mathcal{N}_h := \bigcup_{K \in \mathcal{T}_h} \mathcal{N}(K)$, $\mathcal{E}_h := \bigcup_{K \in \mathcal{T}_h} \mathcal{E}(K)$ be the sets of all vertices resp. of all edges or faces in \mathcal{T}_h . \mathcal{N}_h and \mathcal{E}_h naturally split into the sets $\mathcal{N}_{h,\Omega}, \mathcal{N}_{h,D}, \mathcal{N}_{h,N}$ resp. $\mathcal{E}_{h,\Omega}, \mathcal{E}_{h,D}, \mathcal{E}_{h,N}$ of all vertices resp. edges or faces in Ω , on Γ_D , and Γ_N , respectively. With each $E \in \mathcal{E}_{h,\Omega}$ we associate a unit vector n_E orthogonal to E and denote by $J_E(\varphi)$ the jump of a given piecewise continuous function φ across E in direction n_E . Of course, $J_E(\varphi)$ depends on the orientation of n_E , but quantities like $J_E(\varphi n_E)$ do not.

For any $K \in \mathcal{T}_h, E \in \mathcal{E}_h$, and $x \in \mathcal{N}_h$ we denote by

- ω_K the union of all elements sharing an edge, if $n = 2$, resp. a face, if $n = 3$, with K ,
- ω_E the union of all elements having E as an edge, if $n = 2$, resp. as a face, if $n = 3$,
- ω_x the union of all elements having x as a vertex.

Given a vertex $x \in \mathcal{N}_h$ we denote by λ_x the corresponding nodal shape function, i.e. the continuous, piecewise linear or n -linear function which takes the value 1 at x and which vanishes at all other vertices. Note that λ_x vanishes identically outside ω_x . With each $K \in \mathcal{T}_h$ and each $E \in \mathcal{E}_h$ we associate a cut-off function ψ_K and ψ_E , resp. by

$$\psi_K := \alpha_K \prod_{x \in \mathcal{N}(K)} \lambda_x, \quad \psi_E := \alpha_E \prod_{x \in \mathcal{N}(E)} \lambda_x. \quad (4)$$

Here, $\mathcal{N}(E)$ denotes the set of all vertices of E ; the constants α_K and α_E are determined by the requirements $\max_{x \in K} \psi_K(x) = 1$ and $\max_{x \in E} \psi_E(x) = 1$. Note, that ψ_K and ψ_E vanish identically outside K and ω_E , respectively.

Finally, we define a quasi-interpolation operator I_h by

$$I_h \varphi := \sum_{x \in \mathcal{N}_{h,\Omega} \cup \mathcal{N}_{h,N}} \lambda_x \pi_x \varphi$$

where $\pi_x \varphi$ is the mean value of φ on ω_x . Note, that $I_h \varphi$ vanishes on Γ_D . For vector-valued functions, I_h is defined by applying it to the components of the function. In particular, we have $I_h v \in V_h$ for all $v \in L^2(\Omega)^n$. For any $v \in V_D$ we can prove the following interpolation error estimate (cf. [15] or adapt Exercise 3.2.3 in [9])

$$\left\{ \sum_{K \in \mathcal{T}_h} h_K^{-2} \|v - I_h v\|_K^2 + \sum_{E \in \mathcal{E}_h} h_E^{-1} \|v - I_h v\|_E^2 \right\}^{1/2} \leq c_I \|v\|_{1;\Omega}. \quad (5)$$

The constant c_I only depends on the shape parameter c_T .

5. EQUIVALENCE OF ERROR AND RESIDUAL

Denote by u and u_h the unique solutions of problems (2) and (3), resp. and by $e := u - u_h$ the error of the finite element discretization. From equations (2) and (3) we immediately get the error representation

$$\int_{\Omega} \sigma(e) \cdot \varepsilon(v) = \int_{\Omega} f \cdot v + \int_{\Gamma_N} t_N \cdot v - \int_{\Omega} \sigma(u_h) \cdot \varepsilon(v) \quad \forall v \in V_D. \quad (6)$$

The right-hand side of this equation implicitly defines the residual $R(u_h)$ of u_h as an element of the dual space of V_D :

$$\langle R(u_h), v \rangle := \int_{\Omega} f \cdot v + \int_{\Gamma_N} t_N \cdot v - \int_{\Omega} \sigma(u_h) \cdot \varepsilon(v) \quad \forall v \in V_D. \quad (7)$$

The corresponding norm of $R(u_h)$ is given by $\|R(u_h)\|_{-1} := \sup_{\|v\|_1=1} \langle R(u_h), v \rangle$. From the definition of $R(u_h)$, the error representation, and the Cauchy-Schwarz inequality we get for all $v \in V_D$:

$$\langle R(u_h), v \rangle = \int_{\Omega} \sigma(e) \cdot \varepsilon(v) \leq \|\sigma(e)\| \|\varepsilon(v)\| \leq c^* \|e\|_1 \|v\|_1.$$

The constant c^* only depends on the largest eigenvalue of the matrix C in the material law. This estimate and the definition of $\|R(u_h)\|_{-1}$ imply that $\|R(u_h)\|_{-1} \leq c^* \|e\|_1$. From the positive definiteness of C and from Korn's inequality we conclude on the other hand that

$$\int_{\Omega} \sigma(e) \cdot \varepsilon(e) \geq c_1 \|\varepsilon(e)\|_1^2 \geq c_* \|e\|_1^2.$$

The constant c_1 only depends on the smallest eigenvalue of C ; the constant c_* in addition depends on Ω and on Γ_D . This estimate and the definition of $\|R(u_h)\|_{-1}$ yield the estimate $c_* \|e\|_1 \leq \|R(u_h)\|_{-1}$. Combining both estimates of $\|R(u_h)\|_{-1}$ we finally arrive at

$$c_* \|e\|_1 \leq \|R(u_h)\|_{-1} \leq c^* \|e\|_1. \quad (8)$$

I.e., the energy norm of the error is bounded from below and from above by a dual norm of the residual of the finite element solution. Since, by definition, $R(u_h)$ only includes the given data f and t_N and the computed finite element solution u_h , its norm may be used as an a posteriori error estimator. Unfortunately, the norm $\|R(u_h)\|_{-1}$ is not directly amenable to numerical calculations. Its computation would require the solution of an infinite dimensional variational problem. Most a posteriori error estimators try to approximate $\|R(u_h)\|_{-1}$ by quantities which are much easier to compute.

6. A RESIDUAL ERROR ESTIMATOR

Since $V_h \subset V_D$ equations (2) and (3) imply Galerkin orthogonality, i.e.

$$\langle R(u_h), v_h \rangle = 0 \quad \forall v_h \in V_h.$$

Since $I_h v \in V_h$ for all $v \in V_D$, this in particular yields

$$\langle R(u_h), v \rangle = \langle R(u_h), v - I_h v \rangle \quad \forall v \in V_D. \quad (9)$$

Next, we derive an L^2 -representation of $R(u_h)$ which originally was defined as an element of a dual space being "larger" than L^2 . To this end we start from equation (7) and perform integration by parts elementwise. This yields for all $v \in V_D$

$$\langle R(u_h), v \rangle = \sum_{K \in \mathcal{T}_h} \int_K R_K(u_h) \cdot v + \sum_{E \in \mathcal{E}_h} \int_E R_E(u_h) \cdot v \quad (10)$$

where

$$\begin{aligned} R_K(u_h) &:= \operatorname{div} \sigma(u_h) + f, \quad K \in \mathcal{T}_h, \\ R_E(u_h) &:= \begin{cases} -J_E(n_E \cdot \sigma(u_h)) & , E \in \mathcal{E}_{h,\Omega}, \\ t_N - n \cdot \sigma(u_h) & , E \in \mathcal{E}_{h,N}, \\ 0 & , E \in \mathcal{E}_{h,D}. \end{cases} \end{aligned} \quad (11)$$

This is the L^2 -representation of $R(u_h)$. Note that

$R_K(u_h)$, $K \in \mathcal{T}_h$, is the elementwise residual of u_h w.r.t. the differential equation (1a), $R_E(u_h)$, $E \in \mathcal{E}_{h,N}$, is the residual of u_h w.r.t. the traction boundary condition (1c), $R_E(u_h)$, $E \in \mathcal{E}_{h,\Omega}$, is the interelement-jump of that trace operator that associates the weak formulation (2) with the strong formulation (1).

Now, we consider an arbitrary $v \in V_D$ with $\|v\|_1 = 1$. Equations (9) and (10), the error estimates (5), and the Cauchy-Schwarz inequality then yield the estimate

$$\begin{aligned} \langle R(u_h), v \rangle &= \langle R(u_h), v - I_h v \rangle \\ &\leq c \left\{ \sum_{K \in \mathcal{T}_h} h_K^2 \|R_K(u_h)\|_K^2 + \sum_{E \in \mathcal{E}_h} h_E \|R_E(u_h)\|_E^2 \right\}^{1/2} \|v\|_1. \end{aligned}$$

Here, the constant c only depends on the shape parameter $c_{\mathcal{T}}$. For abbreviation, we set

$$\begin{aligned} \eta_{R,K} &:= \left\{ h_K^2 \|R_K(u_h)\|_K^2 + \sum_{E \in \mathcal{E}(K)} \beta_E h_E \|R_E(u_h)\|_E^2 \right\}^{1/2}, \\ \eta_R &:= \left\{ \sum_{K \in \mathcal{T}_h} \eta_{R,K}^2 \right\}^{1/2}, \end{aligned} \quad (12)$$

where $\beta_E = \frac{1}{2}$, if $E \in \mathcal{E}_{h,\Omega}$, and $\beta_E = 1$, otherwise. Then the above estimate and the definition of $\|R(u_h)\|_{-1}$ imply that $\|R(u_h)\|_{-1} \leq c\eta_R$. Together with the first inequality in (8) this yields the following *upper bound on the error*

$$\|e\|_1 \leq c_*^{-1} c\eta_R. \quad (13)$$

In order to establish the converse estimate, we denote by f_h and $t_{N,h}$ some finite element approximations of f and t_N . For example, these may be piecewise constant. We then define residuals $\tilde{R}_K(u_h)$ and $\tilde{R}_E(u_h)$ as $R_K(u_h)$ and $R_E(u_h)$ with f and t_N replaced by f_h and $t_{N,h}$, respectively. Consider an arbitrary element $K \in \mathcal{T}_h$ and an arbitrary edge or, if $n = 3$, face E of K . Inserting the functions $w_K := \psi_K \tilde{R}_K(u_h)$ and $w_E := \psi_E \tilde{R}_E(u_h)$ as test-functions in equation (10) one arrives at the following *lower bound on the error*

$$\eta_{R,K} \leq c \left\{ \|e\|_{1;\omega_K} + \sum_{K' \subset \omega_K} h_{K'} \|f - f_h\|_{K'} + \sum_{E \in \mathcal{E}(K) \cap \mathcal{E}_{h,N}} h_E^{1/2} \|t_N - t_{N,h}\|_E \right\}. \quad (14)$$

The constant c depends on the constant c^* in inequality (8), the shape parameter $c_{\mathcal{T}}$, and the polynomial degree of the functions $f_h, t_{N,h}$, and of those in V_D . The second and third terms on the right-hand side of estimate (14) only depend on the given data and are usually higher-order perturbations.

In conclusion, η_R yields, up to higher order perturbation terms, global upper and local lower bounds on the error of the finite element solution.

7. ERROR ESTIMATES BASED ON THE SOLUTION OF AUXILIARY LOCAL PROBLEMS

The residual error estimator of the previous section tries to approximate the dual norm $\|R(u_h)\|_{-1}$ of the residual $R(u_h)$ by a suitable L^2 -representation of the residual. The estimators which we will describe now try to lift the residual into a subspace of $H^1(\Omega)^n$ by solving suitable local discrete problems similar to the original problem (3) and by evaluating the energy norm of the solutions of these local problems.

We start with an estimator which originally was introduced by Babuška and Rheinboldt [2, 3]. Consider an arbitrary vertex $x_0 \in \mathcal{N}_{h,\Omega} \cap \mathcal{N}_{h,N}$ and keep it fixed in what follows. With x_0 we associate a finite element space V_{x_0} on ω_{x_0} such that its elements vanish on $\partial\omega_{x_0} \setminus \Gamma_N$ and in all vertices contained in ω_{x_0} and are continuous, piecewise polynomials of a sufficiently high degree (this condition will be made more precise later on). Then the following auxiliary problem admits a unique solution: Find $u_{x_0} \in V_{x_0}$ such that

$$\int_{\omega_{x_0}} \sigma(u_{x_0}) \cdot \varepsilon(v) = \int_{\omega_{x_0}} f \cdot v + \int_{\Gamma_N \cap \partial\omega_{x_0}} t_N \cdot v - \int_{\omega_{x_0}} \sigma(u_h) \cdot \varepsilon(v) \quad \forall v \in V_{x_0}. \quad (15)$$

Set

$$\eta_{D,x_0} := \|\varepsilon(u_{x_0})\|_{\omega_{x_0}}, \quad \eta_D := \left\{ \sum_{x \in \mathcal{N}_{h,\Omega} \cup \mathcal{N}_{h,N}} \eta_{D,x}^2 \right\}^{1/2}. \quad (16)$$

This will be the error estimator. Problem (15) can be interpreted as a discrete elasticity problem on ω_{x_0} with load f , traction t_N on $\partial\omega_{x_0} \cap \Gamma_N$, and displacement u_h on $\partial\omega_{x_0} \setminus \Gamma_N$.

In order to see that η_D yields upper and lower bounds on the error $e = u - u_h$, we first observe that the right-hand side of equation (15) is equal to $\int_{\omega_{x_0}} \sigma(e) \cdot \varepsilon(v)$. Inserting u_{x_0} as a test-function v in (15) therefore immediately yields

$$\underline{c} \|\varepsilon(u_{x_0})\|_{\omega_{x_0}}^2 \leq \int_{\omega_{x_0}} \sigma(u_{x_0}) \cdot \varepsilon(u_{x_0}) = \int_{\omega_{x_0}} \sigma(e) \cdot \varepsilon(u_{x_0}) \leq c^* \|e\|_{1;\omega_{x_0}} \|\varepsilon(u_{x_0})\|_{\omega_{x_0}}$$

and hence the *lower error bound* $\eta_{D,x_0} \leq \frac{c^*}{\underline{c}} \|e\|_{1;\omega_{x_0}}$. The constant \underline{c} only depends on the smallest eigenvalue of the matrix C in the material law. Note, that this estimate holds without any restriction on the polynomial degree of the functions in V_{x_0} .

In order to show that η_D also yields upper bounds on the error, we must specify the condition that the polynomial degree of the functions in V_{x_0} is sufficiently high. More precisely, we assume that V_{x_0} contains, for each $K \subset \omega_{x_0}$ and each $E \in \mathcal{E}_h$ having x_0 as a vertex, the functions w_K and w_E of the previous section.

In order to better understand this condition, assume, e.g., that we approximate problem (1) by bilinear finite elements on a quadrilateral mesh in \mathbb{R}^2 . Then, V_{x_0} must have at least two degrees of freedom for each $K \subset \mathcal{T}_h$ contained in ω_{x_0} and each $E \in \mathcal{E}_h$ having x_0 as a vertex. Since in the generic case ω_{x_0} consists of four quadrilaterals, problem (15) is a linear system of equations with 16 unknowns.

One can then prove that, up to higher order perturbation terms, η_D yields *global upper and local lower bounds on the error*. Moreover, η_D is equivalent to η_R , i.e., up to higher order perturbation terms and multiplicative constants, which do not depend on h , each quantity gives upper and lower bounds on the other one.

Next, we consider an estimator which originally was introduced by Bank and Weiser [4]. Consider an arbitrary element $K \in \mathcal{T}_h$ and keep it fixed in what follows. We associate with K a finite element space V_K on K such that its elements vanish on $\partial K \cap \Gamma_D$ and at the vertices of K and have a sufficiently high polynomial degree (this condition, again, will be made more precise later on). Then the following auxiliary problem admits a unique solution: Find $u_K \in V_K$ such that

$$\int_K \sigma(u_K) \cdot \varepsilon(v) = \int_K R_K(u_h) \cdot v + \int_{\partial K \setminus \Gamma_D} R_E(u_h) \cdot v \quad \forall v \in V_K. \quad (17)$$

Set

$$\eta_{N,K} := \|\varepsilon(u_K)\|_K, \quad \eta_N := \left\{ \sum_{K \in \mathcal{T}_h} \eta_{N,K}^2 \right\}^{1/2}. \quad (18)$$

This will be the error estimator. Problem (17) may be interpreted as a discrete elasticity problem on K with load $R_K(u_h)$, traction $R_E(u_h)$ on $\partial K \setminus \Gamma_D$, and zero displacement on $\partial K \cap \Gamma_D$.

Inserting, u_K as a test function v in equation (17) and using the scaling properties of u_K which are due to its vanishing nodal values, we get the estimate $\eta_{N,K} \leq c \eta_{R,K}$ with a constant c which depends on $c_{\mathcal{T}}$ and the lowest eigenvalue of the matrix C in the material law. Combining this estimate with inequality (14) shows that, up to higher order perturbation terms, $\eta_{N,K}$ yields a lower bound on the energy norm of the error

on ω_K . This holds without any condition on the polynomial degree of the elements of V_K .

In order to establish upper bounds on the error, we must once more specify the condition that the polynomial degree of the functions in V_K is sufficiently high. More precisely, we assume that the functions w_K and w_E , $E \in \mathcal{E}(K) \setminus \mathcal{E}_{h,D}$, are contained in V_K . If, e.g., we approximate problem (1) by bilinear elements on a quadrilateral mesh in \mathbb{R}^2 , this means that V_K must have at least two interior degrees of freedom plus two degrees of freedom for each edge of K which is not part of the Dirichlet boundary Γ_D . Thus, in the generic case, problem (17) is a linear system of equations with 10 unknowns. Since the functions w_E and w_K are contained in V_K we may proceed as for η_D and conclude that η_N yields global upper and local lower bounds on the error. Moreover, η_N is, as η_D , equivalent to η_R .

8. HIERARCHICAL ERROR ESTIMATORS

Consider a finite element space W_h which satisfies $V_h \subset W_h \subset V_D$ and which either consists of higher order elements or corresponds to a refinement of \mathcal{T}_h . If, e.g., V_h consists of bilinear elements on a quadrilateral mesh in \mathbb{R}^2 , W_h may either consist of biquadratic elements on the same mesh or of bilinear elements on a refined mesh which is obtained by cutting each quadrilateral into four new ones by connecting the midpoints of edges. Denote by $w_h \in W_h$ the unique solution of problem (3) with V_h replaced by W_h . Assume that the following *saturation assumption* holds: There is a constant β with $0 \leq \beta < 1$ such that

$$\|u - w_h\|_1 \leq \beta \|u - u_h\|_1. \quad (19)$$

From the saturation assumption (19) and the triangle inequality we immediately conclude that $\|u_h - w_h\|_1$ yields upper and lower bounds on the energy norm of the error $u - u_h$ with amplification factors $(1 - \beta)^{-1}$ and $(1 + \beta)$, respectively. Note, that, in contrast to the previous sections, both bounds are global ones.

Thus, the energy norm of $u_h - w_h$ could be used as an error estimator. This device, however, is not efficient since the computation of w_h is at least as costly as the one of u_h . In order to obtain a more efficient error estimation, we use a hierarchical splitting $W_h = V_h + Z_h$ and assume that the spaces V_h and Z_h are nearly orthogonal, i.e. there is a constant γ with $0 \leq \gamma < 1$ such that the following *strengthened Cauchy-Schwarz inequality* holds

$$\left| \int_{\Omega} \varepsilon(v) \cdot \varepsilon(z) \right| \leq \gamma \|\varepsilon(v)\| \|\varepsilon(z)\| \quad \forall v \in V_h, z \in Z_h. \quad (20)$$

A strengthened Cauchy-Schwarz inequality holds true if, e.g., V_h and W_h consist of bilinear and biquadratic elements, respectively.

Denote by $z_h \in Z_h$ the unique solution of

$$\int_{\Omega} \sigma(z_h) \cdot \varepsilon(\zeta_h) = \int_{\Omega} f \cdot \zeta_h + \int_{\Gamma_N} t_N \cdot \zeta_h - \int_{\Omega} \sigma(u_h) \cdot \varepsilon(\zeta_h) \quad \forall \zeta_h \in Z_h.$$

From the definitions of $u_h, w_h,$ and z_h and from the strengthened Cauchy-Schwarz inequality we infer that $\|z_h\|_1$ yields upper and lower bounds on $\|u_h - w_h\|_1$ and, consequently, on the energy norm of the error. Thus, it can be used as an error estimator. At first sight, the computation of z_h seems to be cheaper than the one of w_h since the dimension of Z_h is smaller than that of W_h . It, however, still requires the solution of a global system and is therefore as expensive as the calculation of u_h and w_h . But, in most applications the functions in Z_h vanish at the vertices of \mathcal{T}_h . This in particular implies that the stiffness matrix corresponding to Z_h is spectrally equivalent to a suitably scaled lumped mass matrix. Therefore, z_h can be replaced by a quantity z_h^* which can be computed by solving a diagonal linear system of equations.

More precisely, on Z_h there is a bilinear form b such that the corresponding stiffness matrix is diagonal and such that

$$\lambda \|z_h\|_1^2 \leq b(z_h, z_h) \leq \Lambda \|z_h\|_1^2 \quad \forall z_h \in Z_h \quad (21)$$

holds with constants $0 < \lambda \leq \Lambda$ that are independent of h . Let $z_h^* \in Z_h$ be the unique solution of

$$b(z_h^*, \zeta_h) = \int_{\Omega} f \cdot \zeta_h + \int_{\Gamma_N} t_N \cdot \zeta_h - \int_{\Omega} \sigma(u_h) \cdot \varepsilon(\zeta_h) \quad \forall \zeta_h \in Z_h. \quad (22)$$

The definitions of z_h and of z_h^* and inequality (21) then imply that $\|z_h^*\|_1$ yields upper and lower bounds on the energy norm of the error. Moreover, problem (22) can be solved element by element.

9. AVERAGING TECHNIQUES

In [21, 22] Zienkiewicz and Zhu proposed an error estimator which is based on an averaging or postprocessing of the computed stress tensor $\sigma(u_h)$. In order to explain the underlying idea we resort to the simplest case and assume that V_h consists of linear elements on a triangular mesh. The stress tensor then is piecewise constant. Suppose that we dispose of an approximation $S(u_h)$ of $\sigma(u)$ such that

$$\|\sigma(u) - S(u_h)\| \leq \beta \|\sigma(u) - \sigma(u_h)\|$$

holds with a constant $0 \leq \beta < 1$. Then by the triangle inequality $\|\sigma(u_h) - S(u_h)\|$ yields upper and lower bounds on the energy norm of the error and may be used as an error estimator.

The idea for computing $S(u_h)$ is to project $\sigma(u_h)$, which is piecewise constant, onto the space of continuous, piecewise linear stress tensors. In order to obtain an easy-to-compute projection it is performed with respect to a lumped mass matrix. More precisely, denote by $S_h^{1,-1}$ and $S_h^{1,0}$ the spaces of all discontinuous resp. continuous, piecewise linear stress tensors corresponding to \mathcal{T}_h . Since $\sigma(u_h)$ is piecewise constant it is contained in $S_h^{1,-1}$. On $S_h^{1,-1}$ we define a mesh-dependent scalar-product by

$$(\sigma_h, \tau_h)_h := \sum_{K \in \mathcal{T}_h} \frac{|K|}{3} \left\{ \sum_{x \in \mathcal{N}(K)} \sigma_h|_K(x) \cdot \tau_h|_K(x) \right\}.$$

Here, $|K|$ is the area of K and $\varphi|_K(x)$ is the value of φ restricted to K at the vertex x . Now, $S(u_h)$ is defined as the $(\cdot, \cdot)_h$ -projection of $\sigma(u_h)$ onto $S_h^{1,0}$:

$$(S(u_h), \tau_h)_h = (\sigma(u_h), \tau_h)_h \quad \forall \tau_h \in S_h^{1,0}. \tag{23}$$

Since the quadrature rule $\int_K \varphi \sim \frac{|K|}{3} \sum_{x \in \mathcal{N}(K)} \varphi(x)$ is exact for all linear functions, we deduce from (23) by inserting the corresponding shape function that, at a given vertex, $S(u_h)$ is the average of the $\sigma(u_h)$ on the neighbouring elements:

$$S(u_h)(x) = \sum_{K \subset \omega_x} \frac{|K|}{|\omega_x|} \sigma(u_h)|_K \quad \forall x \in \mathcal{N}_h. \tag{24}$$

The ZZ-estimator of Zienkiewicz-Zhu is now given by

$$\eta_Z := \left\{ \sum_{K \in \mathcal{T}_h} \eta_{Z,K}^2 \right\}^{1/2}, \quad \eta_{Z,K} := \|S(u_h) - \sigma(u_h)\|_K.$$

It can be shown (cf. [13] and Section 1.5 in [14]) that η_Z yields upper and lower bounds on the energy norm of the error.

When using bilinear elements on quadrilateral meshes one may retain the above expressions for $S(u_h)$ and η_Z with $\sigma(u_h)|_K$ replaced by $\sigma(u_h)(x_K)$, where x_K is the barycentre of K .

10. THE EQUILIBRATED RESIDUAL METHOD

In [11] Ladevèze and Leguillon proposed a technique for a posteriori error estimation which is based on a dual variational principle. In what follows we will shortly sketch the underlying idea. For a more detailed presentation we refer to [11] and [1].

We define a norm and a quadratic functional on V_D by

$$\|v\|^2 := \int_{\Omega} \sigma(v) \cdot \varepsilon(v), \quad J(v) := \frac{1}{2} \int_{\Omega} \sigma(v) \cdot \varepsilon(v) - \langle R(u_h), v \rangle.$$

From the definition of $\|\cdot\|, J$, and $R(u_h)$ we conclude that $-\frac{1}{2} \|e\|^2 = J(e) \leq J(v)$ for all $v \in V_D$. Hence, the energy norm of the error can be computed by solving the variational problem $J(v) \rightarrow \min$ in V_D .

Unfortunately, this is an infinite dimensional minimization problem. In order to obtain a more tractable problem we want to replace V_D by the broken space $V_T := \{v \in L^2(\Omega)^n : v|_K \in H^1(K)^n \quad \forall K \in \mathcal{T}_h, v = 0 \text{ on } \Gamma_D\}$. Obviously, a function $v \in V_T$ belongs to V_D if and only if $J_E(v) = 0$ for all $E \in \mathcal{E}_{h,\Omega}$. It can be proven that the space of Lagrange multipliers corresponding to this constraint is given by $M := \{A \in L^2(\Omega)^{n \times n} : \operatorname{div} A \in L^2(\Omega)^n, n \cdot A = 0 \text{ on } \Gamma_N\}$.

Next, we must extend the residual $R(u_h)$, which is a linear functional on V_D , to a linear functional on the larger space V_T . To this end we associate with each $E \in \mathcal{E}_h$

a smooth vector field g_E . The choice of g_E is arbitrary subject to the constraint that $g_E = t_{N|E}$ for all $E \in \mathcal{E}_{h,N}$. The particular choice of the fluxes g_E for the interelement boundaries will later on determine the error estimation method; for $E \subset \Gamma_D$ the value of g_E is completely irrelevant. Once we have chosen the g_E 's we can associate with each element $K \in \mathcal{T}_h$ a vector field g_K defined on ∂K such that $\sum_{K \in \mathcal{T}_h} \int_{\partial K} g_K \cdot v = \sum_{E \in \mathcal{E}_h} \int_E g_E \cdot J_E(v)$ for all $v \in V_T$. Here, we use the convention that $J_E(v) := v$ if $E \subset \Gamma$. We then define the extension $\tilde{R}(u_h)$ of $R(u_h)$ to V_T by setting for all $v \in V_T$

$$\langle \tilde{R}(u_h), v \rangle := \sum_{K \in \mathcal{T}_h} \left\{ \int_K f \cdot v - \int_K \sigma(u_h) \cdot \varepsilon(v) + \int_{\partial K} g_K \cdot v \right\} - \sum_{E \in \mathcal{E}_{h,\Omega}} \int_E g_E \cdot J_E(v).$$

Moreover, there is a $\mu_* \in M$ such that the last term in $\langle \tilde{R}(u_h), v \rangle$ can be identified with $\mu_*(v)$. Now, we define a Lagrangian functional on $V_T \times M$ by

$$\mathcal{L}(v, \mu) := \frac{1}{2} \sum_{K \in \mathcal{T}_h} \int_K \sigma(v) \cdot \varepsilon(v) - \langle \tilde{R}_h(u_h), v \rangle - \mu(v).$$

Since M is the space of Langrange multipliers for the constraint $J_E(v) = 0$ we conclude that

$$-\frac{1}{2} \|e\|^2 = \inf_{v \in V_D} J(v) = \inf_{w \in V_T} \sup_{\mu \in M} \mathcal{L}(w, \mu) = \sup_{\mu \in M} \inf_{w \in V_T} \mathcal{L}(w, \mu).$$

Hence, we get for all $\mu \in M$ that $-\frac{1}{2} \|e\|^2 \geq \inf_{w \in V_T} \mathcal{L}(w, \mu)$. The particular choice $\mu = \mu_*$ therefore yields

$$\|e\|^2 \leq -2 \sum_{K \in \mathcal{T}_h} \inf_{w_K \in V_K} J_K(w_K), \quad (25)$$

where

$$J_K(w) := \frac{1}{2} \int_K \sigma(w) \cdot \varepsilon(w) - \int_K f \cdot w + \int_K \sigma(u_h) \cdot \varepsilon(w) - \int_{\partial K} g_K \cdot w.$$

Each choice of $w_K \in V_K, K \in \mathcal{T}_h$, will yield a lower bound for the energy norm of the error.

Thanks to (25) the computation of the energy norm of the error is reduced to a family of minimization problems on the elements in \mathcal{T}_h . However, for each $K \in \mathcal{T}_h$, the corresponding variational problem still is infinite dimensional. In order to overcome this difficulty we first rewrite J_K . Using integration by parts we see that

$$J_K(w) = \frac{1}{2} \int_K \sigma(w) \cdot \varepsilon(w) - \int_K R_K(u_h) \cdot w - \int_{\partial K} \hat{R}_K(u_h) \cdot w$$

where $\hat{R}_K(u_h) := g_K - n_K \cdot \sigma(u_h)$. Given $K \in \mathcal{T}_h$ we set

$$W_K := \{\tau \in L^2(K)^{n \times n} : \operatorname{div} \tau \in L^2(K)^n, \operatorname{div} \tau = R_K(u_h), n_K \cdot \tau = \hat{R}_K(u_h)\}.$$

The complementary energy principle then tells us that

$$\inf_{w \in V_K} J_K(w) = \sup_{\tau \in W_K} -\frac{1}{2} \int_K \tau \cdot \tau.$$

Together with (25) this implies that

$$\|e\|^2 \leq -2 \sum_{K \in \mathcal{T}_h} \sup_{\tau_K \in W_K} -\frac{1}{2} \int_K \tau_K \cdot \tau_K.$$

Hence, any choice of $\tau_K \in W_K$, $K \in \mathcal{T}_h$, gives an upper bound for the energy norm of the error. This is the announced dual variational principle.

The concrete realization of the equilibrated residual method now depends on the choice of the τ_K 's and on the definition of the g_K 's. Ladevèze and Leguillon [11], e.g. chose g_K to be the average of $n \cdot \sigma(u_h)$ from the neighbouring elements plus a suitable piecewise linear function on ∂K . The functions τ_K are often chosen as the solution of a maximization problem on a finite dimensional subspace of W_K corresponding to higher order finite elements. With a proper choice of g_K and τ_K one may thus recover the Bank-Weiser estimator η_N of Section 7.

11. COMPLEMENTS

The quality of an error estimator is often measured by its *efficiency index*, i.e. the ratio of the estimated versus the true error. It is called *efficient* if its efficiency index together with its inverse remain bounded for all mesh-sizes. It is called *asymptotically exact* if its efficiency index tends to 1 when the mesh-size converges to zero. The estimators considered in the previous sections are efficient regardless of the mesh-topology. Asymptotic exactness, however, is in general based on super-convergence and only holds for particular mesh-topologies. The Bank-Weiser estimator η_N of Section 7, e.g., is asymptotically exact on a Courant triangulation but not on a criss-cross triangulation. On a uniform quadrilateral mesh η_N is also asymptotically exact. This property, however, is lost when the mesh is slightly distorted in a non-uniform way.

The residual error estimator of Section 6 consists of two contributions: the element residuals and the jump-terms across interelement boundaries. On a uniform square grid in \mathbb{R}^2 it can be shown that either the element residuals or the jump-terms are dominant depending on whether the polynomial degree of V_h is even or odd (cf. [19, 20]). This result strongly depends on the special structure of the mesh. For simplicial finite elements a similar result is not known. Recently, however, it was shown that for linear elements on a triangular grid the element residuals may be disregarded (cf. [8]).

In the previous sections we have only considered energy norm error estimates. Sometimes, however, one is more interested in estimating the error with respect to another norm, e.g. the L^2 -norm or the maximum-norm, or with respect to a given functional, e.g. the shear stress in a particular point or region. This task can often be achieved by combining the previously used techniques with an appropriate duality argument (cf. [5,

12, 14, 16]). For example, one can thus show that $h_K \eta_{R,K}$ yields upper and lower bounds on $\|u - u_h\|$. The same holds for the Babuška-Rheinboldt and Bank-Weiser estimators of Section 7 provided one replaces in definitions (16) and (18) the energy norm by the corresponding L^2 -norm. In a similar way one obtains maximum-norm error estimates by a duality argument and sharp a priori estimates for the Green's function.

12. THE EQUATIONS OF NON-LINEAR ELASTICITY AND THEIR DISCRETIZATION

Formally, the elliptic system under consideration is still given by equation (1). But, now, the matrix C in the material law, the external force f , and the boundary traction t_N may depend on the displacement u and the strain tensor $\varepsilon(u)$.

To simplify the presentation, we will assume in what follows that the boundary displacement u_D vanishes. Then the standard variational formulation of problem (1) is to find $u \in X$ such that

$$\langle F(u), v \rangle := \int_{\Omega} \sigma(u) \cdot \varepsilon(v) - \int_{\Omega} f \cdot v - \int_{\Gamma_N} t_N \cdot v = 0 \quad \forall v \in Y \quad (26)$$

where $X = W_D^{1,p}(\Omega)^n := \{v \in W^{1,p}(\Omega)^n : v = 0 \text{ on } \Gamma_D\}$ and $Y = W_D^{1,r}(\Omega)^n := \{v \in W^{1,r}(\Omega)^n : v = 0 \text{ on } \Gamma_D\}$. The Lebesgue exponents $1 < p, r < \infty$ depend on the non-linearities. In the simplest case we will have $p = r = 2$, i.e. $X = Y = V_D$.

Any finite element discretization of problem (26) can now be written as

$$\langle F_h(u_h), v_h \rangle = 0 \quad \forall v_h \in Y_h. \quad (27)$$

Here, $F_h : X_h \rightarrow Y_h^*$ is a continuous non-linear mapping between a finite element space X_h approximating X and the dual of a finite element space Y_h approximating Y . In the simplest case we will have $X_h = Y_h$ and $\langle F_h(u_h), v_h \rangle = (F(u_h), v_h)$. In general the definitions of F_h , X_h , and Y_h will have to take into account variational crimes such as numerical quadrature, selected reduced integration etc.

13. EQUIVALENCE OF ERROR AND RESIDUAL FOR THE NON-LINEAR PROBLEM

Let $u \in X$ and $u_h \in X_h$ be a solution of problem (26) and an approximate solution of problem (27), respectively. Note that in general problems (26) and (27) will have several solutions and that we take into account additional errors which are committed when solving the discrete algebraic problem (27) only approximately by some iterative method, e.g. Newton's method.

We want to derive computable upper and lower bounds for $\|u - u_h\|_X$. For linearized elasticity we have seen in Section 5 that the first and essential step in achieving this goal consists in establishing an equivalence between $\|u - u_h\|_X$ and a suitable dual norm of

the residual of u_h . This is also the case for the present non-linear problem. We therefore assume that u satisfies the following two conditions:

- (1) u is a *regular solution* of problem (26). I.e. the derivative $DF(u)$ of F at u is a bounded linear map of X onto Y^* with a bounded inverse $DF(u)^{-1}$.
- (2) DF is *locally Lipschitz continuous* at u . I.e. there are numbers $R > 0$ and $\beta > 0$ such that $\|DF(u) - DF(v)\|_{\mathcal{L}(X, Y^*)} \leq \beta \|u - v\|_X$ holds for all $v \in X$ with $\|u - v\|_X \leq R$.

Condition (1) is the non-linear analogue of estimate (8) which means that the weak formulation of problem (1) defines a bounded linear operator of V_D into its dual having a bounded inverse. Condition (1) excludes limit, turning, or bifurcation points. The treatment of those singular solutions requires some additional techniques which we cannot present here (cf. Section 2.2 of [14]). Condition (2) is not needed for linear problems since then the mapping $u \rightarrow DF(u)$ is constant. It requires extra smoothness on the data C, f , and t_N such as, e.g., twice continuous differentiability. Note that for a linear operator $A : X \rightarrow Y^*$ its norm is given by $\|A\|_{\mathcal{L}(X, Y^*)} = \sup_{\|v\|_X=1} \sup_{\|\varphi\|_{Y^*}=1} \langle Av, \varphi \rangle$.

Recalling that $F(u) = 0$ and using the identity

$$\langle DF(u)(v - u), \varphi \rangle = \langle F(v), \varphi \rangle + \int_0^1 \langle [DF(u) - DF(u + t(v - u))](v - u), \varphi \rangle dt$$

which holds for all $v \in X$ and $\varphi \in Y$ and conditions (1) and (2), one can prove (cf. Proposition 2.1 in [14]) that the estimate

$$\frac{1}{2} \|DF(u)\|_{\mathcal{L}(X, Y^*)}^{-1} \|F(v)\|_{Y^*} \leq \|u - v\|_X \leq 2 \|DF(u)^{-1}\|_{\mathcal{L}(Y^*, X)} \|F(v)\|_{Y^*}$$

holds for all $v \in X$ with $\|u - v\|_X \leq R_0 := \min\{R, \beta^{-1} \|DF(u)^{-1}\|_{\mathcal{L}(Y^*, X)}^{-1}, 2\beta^{-1} \|DF(u)\|_{\mathcal{L}(X, Y^*)}\}$.

Hence, if $\|u - u_h\|_X \leq R_0$, the error $\|u - u_h\|_X$ is bounded from above and from below by the residual $\|F(u_h)\|_{Y^*} = \sup_{\|\varphi\|_{Y^*}=1} \langle F(u_h), \varphi \rangle$. For linear problems this is nothing else than the dual norm $\|R(u_h)\|_{-1}$ of the residual. The amplification factors in the above equivalence between error and residual are the operator norms $\|DF(u)\|_{\mathcal{L}(X, Y^*)}$ and $\|DF(u)^{-1}\|_{\mathcal{L}(Y^*, X)}$ of the derivative of F at u and of its inverse. Note that $\|DF(u)^{-1}\|_{\mathcal{L}(Y^*, X)} = \{\inf_{\|v\|_X=1} \sup_{\|\varphi\|_{Y^*}=1} \langle DF(u)v, \varphi \rangle\}^{-1}$. The product $\|DF(u)\|_{\mathcal{L}(X, Y^*)} \|DF(u)^{-1}\|_{\mathcal{L}(Y^*, X)}$ is the condition number of the linearization of F at u . It measures the sensitivity of the non-linear problem (26) with respect to small perturbations and is independent of the particular finite element scheme. Thus, *independently of the finite element discretization, there will be a large gap between the upper and lower bounds when the non-linear mechanical system is ill-conditioned*, i.e. if it is very sensitive to small perturbations.

Note that the above equivalence between error and residual only holds if the finite element solution u_h is sufficiently close to the true displacement u , i.e. if $\|u - u_h\|_X \leq R_0$. This is not surprising since we are dealing with a non-linear problem which may have a large variety of solutions. For linear problems which are uniquely solvable this extra condition is of course not needed.

14. A RESIDUAL ERROR ESTIMATOR FOR THE NON-LINEAR PROBLEM

In Section 6 we have seen that the dual norm of the residual can be bounded from below and from above by a weighted sum of element residuals and of jump-terms across interelement boundaries. The main tools for establishing this were an L^2 -representation of the residual, local error estimates for a quasi-interpolation operator, and suitable local cut-off functions. This technique can be embedded in a general framework which then allows the treatment of non-linear problems. We do not have the place to present this framework here. Instead we refer to Sections 3.1 and 3.3 of [14] and only present the final result.

Denote by $q := \frac{r}{r-1}$ the dual Lebesgue exponent of r . In analogy to (12) we then set

$$\eta_{R,K} := \left\{ h_K^q \int_K |R_K(u_h)|^q + \sum_{E \in \mathcal{E}(K)} \beta_E h_E \int_E |R_E(u_h)|^q \right\}^{1/q}, \quad \eta_R := \left\{ \sum_{K \in \mathcal{T}_h} \eta_{R,K}^q \right\}^{1/q}.$$

Note that, in evaluating $R_K(u_h)$ and $R_E(u_h)$ according to (11), one has to use the corresponding non-linear material law. Then we obtain the following a posteriori error estimates:

$$\|u - u_h\|_X \leq c_1 \left\{ \eta_R + \sup_{\substack{\varphi_h \in Y_h \\ \|\varphi_h\|_Y=1}} \langle F(u_h) - F_h(u_h), \varphi_h \rangle + \sup_{\substack{\varphi_h \in Y_h \\ \|\varphi_h\|_Y=1}} \langle F_h(u_h), \varphi_h \rangle \right\} \quad (28)$$

$$\eta_{R,K} \leq c_2 \|u - u_h\|_{W^{1,p}(\omega_K)}. \quad (29)$$

The constants c_1 and c_2 depend on the shape parameter $c_{\mathcal{T}}$, $\|DF(u)\|_{\mathcal{L}(X,Y^*)}$, and $\|DF(u)^{-1}\|_{\mathcal{L}(Y^*,X)}$. The second term on the right-hand side of estimate (28) is the consistency error of the discretization. It can be bounded a priori and it vanishes when using a conforming finite element discretization without numerical integration. The third term on the right-hand side of estimate (28) is the residual of the algebraic system (27). It takes into account errors which stem from an approximate solution of (27) with an iterative method. Both perturbation terms were not present in the linear case since there we assumed that we use a conforming method with exact integration and that we exactly solve the discrete problem.

Under some additional assumptions about elliptic regularity (cf. [16] and Remark 3.15 in [14]) one can prove estimates similar to (28) and (29) also for the L^p -norm of the error. One then has to replace $\eta_{R,K}$ and η_R by $h_K \eta_{R,K}$ and $\{\sum_{K \in \mathcal{T}_h} h_K^q \eta_{R,K}^q\}^{1/q}$, respectively.

15. ERROR ESTIMATORS BASED ON THE SOLUTION OF AUXILIARY LOCAL PROBLEMS FOR THE NON-LINEAR PROBLEM

The Babüska-Rheinboldt error estimator η_D of Section 7 can be generalized to non-linear problems (cf. Sections 2.3 and 3.3 in [14]). Using the notations of Section 7 it is given by

$$\eta_{D,x_0} := \|\epsilon(u_{x_0})\|_{L^p(\omega_{x_0})}, \quad \eta_D := \left\{ \sum_{x \in \mathcal{N}_{h,\Omega} \cup \mathcal{N}_{h,N}} \eta_{D,x}^p \right\}^{1/p}$$

where $u_{x_0} \in V_{x_0}$ is the unique solution of

$$\int_{\omega_{x_0}} \hat{\sigma}(u_{x_0}) \cdot \varepsilon(v) = \int_{\omega_{x_0}} f \cdot v + \int_{\Gamma_N \cap \partial \omega_{x_0}} t_N \cdot v - \int_{\omega_{x_0}} \sigma(u_h) \cdot \varepsilon(v) \quad \forall v \in V_{x_0}. \quad (30)$$

The differences with the linear problem treated in Section 7 are:

- (1) The L^2 -norm in (16) is replaced by an L^p -norm.
- (2) As in (15) we use on the right-hand side of (30) the original – now non-linear – material law. But on the left hand side of (30) it is replaced by the simpler linear material law $\hat{\sigma}(v)_{ij} = \sum_{1 \leq k, l \leq n} C(u_h(x_0), \pi_0 \sigma(u_h))_{ijkl} \varepsilon(v)_{kl}$. Here, $\pi_0 \sigma(u_h)$ is some averaging of the possibly discontinuous stress tensor $\sigma(u_h)$ over neighbouring elements of x_0 . If $\sigma(u_h)$ is continuous we may take $\pi_0 \sigma(u_h) = \sigma(u_h)(x_0)$. Thus, for the auxiliary local problem (30) the non-linear material law is linearized and its coefficients are frozen to appropriate constant values.

It can be proven (cf. Proposition 3.14 in [14]) that η_D yields upper and lower bounds for $\|u - u_h\|_X$ similar to estimates (28) and (29). Similarly, one can prove that $\|u_{x_0}\|_{L^p(\omega_{x_0})}$ yields upper and lower bounds on the L^p -norm of the error (cf. [16] and Remark 3.15 in [14]).

16. COMPLEMENTS ON THE NON-LINEAR PROBLEM

The constants relating the estimated error with the true error, i.e. the constants c_1 and c_2 in estimates (28) and (29), consist of two multiplicative ingredients:

- (1) A factor c_I which only depends on the shape parameter c_T and which stems from local inverse estimates (for lower bounds on the error) or local interpolation error estimates (for upper bounds on the error).
- (2) A factor c_P which is problem dependent and which essentially is $\|DF(u)\|_{\mathcal{L}(X, Y^*)}$ (for lower bounds on the error) or $\|DF(u)^{-1}\|_{\mathcal{L}(Y^*, X)}$ (for upper bounds on the error).

The factor c_I is relatively harmless. In principle, it can be estimated explicitly (cf. [15]) and is of moderate size. The factor c_P , on the other hand, is much more dangerous. It is related to the condition of the variational problem, i.e. its sensitivity with respect to small perturbations. Since it depends on the unknown solution u of the variational problem it cannot be computed explicitly. Often only crude a priori estimates for c_P are available which for complex problems are nearly useless due to their unrealistic size. Hence, one must try to get more realistic estimates of this constant. This is an important point of current research (cf. [5, 10]). Many approaches try to replace $\|DF(u)\|_{\mathcal{L}(X, Y^*)}$ and $\|DF(u)^{-1}\|_{\mathcal{L}(Y^*, X)}$ by the extreme eigenvalues of $DF_h(u_h)$ and to estimate these parallel to the computation of the numerical solution u_h by solving an appropriate dual problem.

REFERENCES

- [1] M. Ainsworth and J.T. Oden: A posteriori error estimators for 2nd order elliptic systems II. An optimal order process for calculating self-equilibrated fluxes. *Comput. Math. Appl.* 26, 75 – 87 (1993).
- [2] I. Babuška and W. C. Rheinboldt: Error estimates for adaptive finite element computations. *SIAM J. Numer. Anal.* 15, 736 – 754 (1978).
- [3] ———: A posteriori error estimates for the finite element method. *Int. J. Numer. Meth. Engrg.* 12, 1597 – 1615 (1978).
- [4] R. E. Bank and A. Weiser: Some a posteriori error estimators for elliptic partial differential equations. *Math. Comput.* 44, 283 – 301 (1985).
- [5] R. Becker and R. Rannacher: Weighted a posteriori error control in FE methods. Preprint 96 -1 (SFB 359), Universität Heidelberg, 1996.
- [6] D. Braess and R. Verfürth: A posteriori error estimators for the Raviart-Thomas element. *SIAM J. Numer. Anal.* 33, 2431 – 2444 (1996).
- [7] C. Carstensen: A posteriori error estimate for the mixed finite element method. *Math. Comput.* 66, 465 – 476 (1997).
- [8] C. Carstensen and R. Verfürth: Edge residuals dominate a posteriori error estimates for low order finite element methods. Bericht Nr. 216, Ruhr-Universität Bochum, 1997.
- [9] P. G. Ciarlet: *The Finite Element Method for Elliptic Problems*. North Holland, Amsterdam, 1978.
- [10] C. Johnson and P. Hansbo: Adaptive finite element methods in computational mechanics. *Comp. Math. Appl. Mech. Engrg.* 101, 143 – 181 (1992).
- [11] P. Ladevèze and D. Leguillon: Error estimate procedure in the finite element method and applications. *SIAM J. Numer. Anal.* 20, 485 – 509 (1983).
- [12] R.H. Nochetto: Pointwise a posteriori error estimates for elliptic problems on highly graded meshes. *Math. Comput.* 64, 1 – 22 (1995).
- [13] R. Rodriguez: Some remarks on Zienkiewicz-Zhu estimator. *Int. J. Numer. Meth. in PDE* 10, 625 – 635, (1994).
- [14] R. Verfürth: *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*. Wiley-Teubner, Stuttgart, 1996.
- [15] ———: Sharp error estimates for some quasi-interpolation operators. Bericht, Ruhr-Universität Bochum, 1997.
- [16] ———: A posteriori error estimates for non-linear problems. L^r -estimates for finite element discretizations of elliptic equations. Bericht Nr. 199, Ruhr-Universität Bochum, 1996 (RAIRO M^2AN to appear).
- [17] B. Wohlmuth: *Adaptive Multilevel-Finite-Elemente Methoden zur Lösung elliptischer Randwertprobleme*. Dissertation, Universität München, 1995.
- [18] B. Wohlmuth and R. H. W. Hoppe: A comparison of a posteriori error estimators for mixed finite elements. Report Nr. 350, Universität Augsburg, 1996.
- [19] D. Yu: Asymptotically exact a posteriori error estimators for elements of bi-odd degree. *Chinese J. Num. Math. and Appl.* 13, 64 – 78 (1991).
- [20] ———: Asymptotically exact a posteriori error estimators for elements of bi-even degree. *Chinese J. Num. Math. and Appl.* 13, 82 – 90 (1991).

- [21] O. C. Zienkiewicz and J. Z. Zhu: Adaptive techniques in the finite element method. *Commun. Appl. Numer. Methods* 4, 197 – 204 (1988).
- [22] ———: A simple error estimator and adaptive procedure for practical engineering analysis. *Int. J. Numer. Meth. Engrg.*, 24, 337 – 357 (1987).

A Posteriori Error Control and Mesh Adaptation for FE models in Elasticity and Elasto-Plasticity

R. Rannacher and F.T. Suttmeier * *

*Institut für Angewandte Mathematik, Universität Heidelberg, INF 294,
D-69120 Heidelberg, Germany

A new technique for *a posteriori* error control and adaptive mesh design is presented for finite element models in elasticity and elasto-plasticity. It is based on weighted *a posteriori* error estimates for general error functionals which are derived by duality arguments. This approach was originally proposed in [6] as a natural extension of the ideas developed by C. Johnson, et al., [11] and [13], for adaptive finite element methods. The conventional strategies for mesh refinement in finite element methods are mostly based on *a posteriori* error estimates for the global energy norm in terms of local residuals of the computed solution. These estimates reflect the approximation properties of the trial functions by local interpolation constants while the stability property of the continuous model enters through a global coercivity constant. However, meshes generated on the basis of such *global* error estimates may not be appropriate in computing local quantities like point values or contour integrals and in the case of very heterogeneous or nonlinear material behavior. More accurate and efficient error estimation can be achieved by using suitable weighting factors which can be obtained numerically in the course of a feed-back process from the solutions of discretised dual problems. This general approach is discussed here for finite element models in linear elasticity and perfect plasticity.

1. Introduction

The fundamental problem in the flow theory of linear-elastic perfect-plastic material in classical notation reads

$$\begin{aligned} \operatorname{div} \sigma &= -f, \quad \varepsilon(\dot{u}) = A : \dot{\sigma} + \lambda && \text{in } \Omega, \\ \lambda : (\tau - \sigma) &\leq 0 \quad \forall \tau \text{ with } \mathcal{F}(\tau) \leq 0, \quad \lambda : \dot{\sigma} = 0 && \text{in } \Omega, \\ \dot{u} &= 0 \quad \text{on } \Gamma_D, \quad \sigma \cdot n = g && \text{on } \Gamma_N. \end{aligned} \tag{1}$$

where σ and u are the stress tensor and displacement vector, respectively, and a stress-free initial state $\sigma(0) = 0$ is assumed. This idealized model describes the deformation of an elasto-plastic body occupying a bounded domain $\Omega \subset \mathbb{R}^d$ ($d = 2$ or 3) under the action of a body force f and a surface traction g along Γ_N . Along the remaining part of

*This work has been supported by the German Research Association (DFG) under contract no. RA 306/8-2, Institut für Angewandte Mathematik, Universität Heidelberg, INF 294, D-69120 Heidelberg, Germany

the boundary, $\Gamma_D = \partial\Omega \setminus \Gamma_N$, the body is fixed. We assume a quasi-static process, i.e., acceleration effects are neglected. Further, the displacement u is supposed to be small, so that the strain tensor can be written as $\varepsilon(u) = \frac{1}{2}(\nabla u + \nabla u^T)$. The material tensor A is assumed to be symmetric and positive definite. The plastic growth is denoted by λ , and \mathcal{F} is the von Mises yield function, $\mathcal{F}(\tau) = |\tau^D| - \sigma_0$, with $\sigma_0 > 0$ and τ^D being the deviatoric part of τ . For the choice $\sigma_0 = \infty$, (1) reduces to the standard problem in linear elasticity.

This problem is to be solved by the finite element Galerkin method on adaptively optimized meshes. By variational arguments, we derive weighted *a posteriori* error estimates for controlling arbitrary functionals of the error. This new approach leads to effective strategies for designing economical meshes and to practically useful bounds for the error.

The implementation of the finite element code for the test computations in this paper is based on the object-oriented FE package DEAL (*Differential Equations Analysis Library*)[5] which enabled our group to do studies on adaptivity for various kinds of problems (see, e.g., [4],[15],[19] and [7]). Special thanks go to Guido Kanschat who started together with the second author in 1991 in Bonn the development of this software project.

2. Variational formulations of the plasticity problem

For applying a finite element method, one has to rewrite problem (1) in a variational setting. Typically the behaviour of plastic materials is an evolutionary process over a time interval $I := [0, T]$, where the load functions may be given in the form $f = f(t, x)$, $g = g(t, x)$, $t \in I$, $x \in \Omega$. We introduce the function spaces

$$\begin{aligned} L^2(\Omega)^d &:= L^2(\Omega, \mathbb{R}^d), & L^2(\Omega)_{sym}^{d \times d} &:= L^2(\Omega, \mathbb{R}_{sym}^{d \times d}), \\ V &:= \{v \in H^1(\Omega, \mathbb{R}^d), v = 0 \text{ on } \Gamma_D\}, \\ H^{div} &:= \{\tau \in L^2(\Omega)_{sym}^{d \times d}, \operatorname{div} \tau \in L^2(\Omega)^d\}, \\ H_{f,g}^{div} &:= \{\tau \in H^{div}, -\operatorname{div} \tau = f \text{ in } \Omega, \tau \cdot n = g \text{ on } \Gamma_N\}, \end{aligned}$$

and define for any such function space Σ the *admissible set* as $\Pi(\Sigma) = \{\tau \in \Sigma, \mathcal{F}(\tau) \leq 0\}$. Further, (\cdot, \cdot) and $\|\cdot\|$ denote the L^2 -inner product and norm over Ω and $(\cdot, \cdot)_{\Gamma_N}$ is the L^2 -inner product over the curve segment Γ_N . Then, following [14], we can state the *dual-mixed* formulation of (1): Find a pair $\{v, \sigma\} : I \rightarrow L^2(\Omega)^d \times \Pi(H_{*,g}^{div})$ with $\sigma(0) = 0$, satisfying

$$\begin{aligned} (A\dot{\sigma}, \tau - \sigma) + (v, \operatorname{div} \tau - \operatorname{div} \sigma) &\geq 0 \quad \forall \tau \in \Pi(H_{*,0}^{div}), \\ -(\operatorname{div} \sigma, \varphi) &= (f, \varphi) \quad \forall \varphi \in L^2(\Omega)^d, \end{aligned} \tag{2}$$

where $v = \dot{u}$ is the displacement velocity and the $*$ in $H_{*,0}^{div}$ indicates that the corresponding value is not fixed. Integrating by parts in (2) leads to the *primal-mixed* variational formulation: Find a pair $\{v, \sigma\} : I \rightarrow V \times \Pi(L^2(\Omega)_{sym}^{d \times d})$ with $\sigma(0) = 0$, satisfying

$$\begin{aligned} (A\dot{\sigma}, \tau - \sigma) - (\varepsilon(v), \tau - \sigma) &\geq 0 \quad \forall \tau \in \Pi(L^2(\Omega)_{sym}^{d \times d}), \\ (\sigma, \varepsilon(\varphi)) &= (f, \varphi) + (g, \varphi)_{\Gamma_N} \quad \forall \varphi \in V. \end{aligned} \tag{3}$$

Neglecting the rate dependence in (3), we obtain a material behaviour of *Hencky-type* (c.f. [10]). This means that a pair $\{u, \sigma\}$ of displacement and stress is sought satisfying

$$\begin{aligned} (A\sigma, \tau - \sigma) - (\varepsilon(u), \tau - \sigma) &\geq 0 \quad \forall \tau \in \Pi(L^2(\Omega)_{sym}^{d \times d}), \\ (\sigma, \varepsilon(\varphi)) &= (f, \varphi) + (g, \varphi)_{\Gamma_N} \quad \forall \varphi \in V. \end{aligned} \quad (4)$$

Examining (4), we see that σ is just the projection of $A^{-1}\varepsilon(u)$ onto $\Pi(L^2(\Omega)_{sym}^{d \times d})$. In this note, we consider the case of the linear elastic isotropic material law

$$\sigma = 2\mu\varepsilon^D(u) + \kappa \operatorname{div} u I,$$

with material dependent constants $\mu > 0$ and $\kappa > 0$. Hence, observing the von Mises flow rule, (4) becomes

$$a(u; \varphi) := (\Pi(2\mu\varepsilon^D(u)), \varepsilon(\varphi)) + (\kappa \operatorname{div} u, \operatorname{div} \varphi) = (f, \varphi) + (g, \varphi)_{\Gamma_N} \quad \forall \varphi \in V, \quad (5)$$

where

$$\Pi(2\mu\varepsilon^D(u)) := \begin{cases} 2\mu\varepsilon^D(u) & , \text{ if } |2\mu\varepsilon^D(u)| \leq \sigma_0, \\ \frac{\sigma_0}{|\varepsilon^D(u)|} \varepsilon^D(u) & , \text{ if } |2\mu\varepsilon^D(u)| > \sigma_0. \end{cases}$$

Usually, in the plastic zone there holds $\kappa \gg |\Pi(\varepsilon^D(u))|$, i.e., nearly incompressible material behaviour occurs. In this case, the relation between $\varepsilon(u)$ and σ becomes stiff causing a poor approximation behaviour for discretisations based on (5). One may account for this difficulty by introducing an auxiliary variable $p := \kappa \operatorname{div} u$ which plays the role of a *pressure*. Then, introducing the *pressure space* $Q := L^2(\Omega)$, a pair $\{u, p\} \in V \times Q$ is determined by the saddle-point problem, see [21],

$$\begin{aligned} (\Pi(2\mu\varepsilon^D(u)), \varepsilon(\varphi)) + (\operatorname{div} \varphi, p) &= (f, \varphi) + (g, \varphi)_{\Gamma_N} \quad \forall \varphi \in V, \\ (\operatorname{div} u, q) - \frac{1}{\kappa}(p, q) &= 0 \quad \forall q \in Q, \end{aligned} \quad (6)$$

where, in the case $\Gamma_D = \partial\Omega$, Q is supplemented by the condition $(q, 1) = 0$.

3. The linear elastic case

First, we consider the linear-elastic case. The primal formulation obtained from (5) reads

$$a(u, \varphi) = (f, \varphi) + (g, \varphi)_{\Gamma_N} \quad \forall \varphi \in V, \quad (7)$$

with the linear energy form $a(\cdot, \cdot) := (C\varepsilon(\cdot), \varepsilon(\cdot))$, where $C := A^{-1}$. Starting from (7) the choice of finite element subspaces $V_h \subset V$ is straightforward. Here we confine ourselves to the lowest-order approximation by linear or (iso-parametric) d-linear shape functions. The underlying meshes are assumed to satisfy the usual regularity conditions (shape regularity, cf. [8]) and to properly match the decomposition $\partial\Omega = \Gamma_D \cup \Gamma_N$. Further, in order to ease the refinement and coarsening process, hanging nodes are allowed, but only one per element edge. The width of the finite element mesh \mathbb{T}_h is characterised in terms of a

piecewise constant mesh size function $h = h(x)$, $0 < h \leq h_0$, where $h_K := h|_K = \text{diam}(K)$ and $h_{max} = \max_{K \in \mathcal{T}_h} \{h_K\}$. The discrete solutions $u_h \in V_h$ are determined by the equation

$$a(u_h, \varphi_h) = (f, \varphi_h) + (g, \varphi_h)_{\Gamma_N} \quad \forall \varphi_h \in V_h. \tag{8}$$

If problem (7) is sufficiently regular, we have the following well known *a priori* convergence result (see, e.g., [8]),

$$\|u - u_h\| + h_{max} \|\sigma - \sigma_h\| = \mathcal{O}(h_{max}^2), \tag{9}$$

where the discrete stresses are obtained locally in the form $\sigma_h|_K = C\varepsilon(u_h)|_K$.

3.1. A posteriori error estimate

Now, the *a posteriori* error analysis starts with selecting an error functional $J(\cdot)$ which is to be controlled. We assume that the corresponding *dual problem*

$$a(\varphi, z) = J(\varphi) \quad \forall \varphi \in V. \tag{10}$$

possesses a solution $z \in V$. This is automatically satisfied if $J(\cdot)$ is well-defined on V . Taking $\varphi = e := u - u_h$ in (10) and using the usual Galerkin orthogonality relation $a(e, \varphi) = 0$, $\varphi \in V_h$, we obtain the error identity

$$J(e) = a(e, z) = a(e, z - z_h), \tag{11}$$

with an arbitrary function $z_h \in V_h$. Then, by element-wise integration by parts, it follows that

$$J(e) = \sum_{K \in \mathcal{T}_h} \left\{ (f + \text{div } C\varepsilon(u_h), z - z_h)_K - \frac{1}{2} (n \cdot [C\varepsilon(u_h)], z - z_h)_{\partial K} \right\}, \tag{12}$$

where $[C\varepsilon(u_h)]$ denotes the jump of $C\varepsilon(u_h)$ across the interelement boundary. On edges Γ belonging to the boundary part Γ_N , we understand our notation in the sense that $n \cdot [C\varepsilon(u_h)]|_{\Gamma} = (n \cdot C\varepsilon(u_h) - g)|_{\Gamma}$, in order to properly include the non-homogeneous surface traction. At this point, we have assumed that the domain Ω is polygonal (or polyhedral) in order to ease the approximation of the boundary. In the presence of curved parts of the boundary $\partial\Omega$ the error formula (12) contains additional terms representing the error caused by the polygonal approximation of the boundary (see, e.g., [7] or [9]). In the tests presented below these terms are suppressed. From the error representation (12) we conclude the following theorem (see [17]).

Theorem 3.1. *Under the foregoing assumptions, the finite element scheme (8) admits the a posteriori error estimate*

$$|J(u - u_h)| \leq \eta(u_h) := \sum_{K \in \mathcal{T}_h} \omega_K \varrho_K, \tag{13}$$

with the local residuals ϱ_K and weighting factors ω_K defined by

$$\begin{aligned} \varrho_K &= \left(h_K^2 \|f + \text{div } C\varepsilon(u_h)\|_K^2 + \frac{1}{2} h_K \|n \cdot [C\varepsilon(u_h)]\|_{\partial K}^2 \right)^{1/2}, \\ \omega_K &= \left(h_K^{-2} \|z - z_h\|_K^2 + \frac{1}{2} h_K^{-1} \|z - z_h\|_{\partial K}^2 \right)^{1/2}. \end{aligned}$$

For z_h being an suitable local approximation of z , the weights ω_K in (13) can be estimated by

$$\omega_K \leq C_{i,K} h_K \|\nabla^2 z\|_K, \quad (14)$$

with certain *interpolation constants* $C_{i,K}$. Following the approach proposed in [6], in the *a posteriori* error estimate (13) the exact *dual solution* z is replaced by an approximation \tilde{z}_h which is computed on the same mesh on which u_h has been obtained. Accordingly, the weights ω_K are approximated by

$$\omega_K \approx \tilde{\omega}_K := C_{i,K} h_K^2 |\nabla_h^2 \tilde{z}_h(x_K)|, \quad (15)$$

where x_K is the midpoint of K and $\nabla_h^2 \tilde{z}_h$ are certain second-order difference quotients of \tilde{z}_h approximating $\nabla^2 z$. The local interpolation constants $C_{i,K}$ essentially depend on the geometry of the elements K and vary in the range $C_{i,K} \approx 0.1 - 1$. In the computational tests presented below, we have used the conservative value $C_{i,K} = 1$. Further, in the case that the functional $J(\cdot)$ is "irregular", i.e., if it is not properly defined on the solution space V , one has to work with a regularised functional $J_\epsilon(\cdot)$ defined on V , such that $|J_\epsilon(u) - J(u)| \approx TOL_L$ on each refinement level L . Otherwise, local over-refinement may occur. An example is the contour integral over normal stresses, discussed below. We remark that the explicit *a posteriori* error estimator of Theorem 3.1 may be combined with the locally implicit approach of Bank and Weiser [3] in order to better balance the two residual terms (see [2]).

The weighted error estimate (13) will be compared against the traditional *energy* error estimate (see, e.g., [1] or [22])

$$\|u - u_h\|_E \leq \eta_E(u_h) := C_s C_i \left(\sum_{K \in \mathcal{T}_h} \varrho_K^2 \right)^{1/2}, \quad (16)$$

with the local residuals ϱ_K as defined above.

3.2. Numerical test for the linear elastic case

The approach described above is applied for a model problem in linear elasticity employing the two dimensional plane strain model (cf. [17]). A square elastic disc with a crack is subjected to a constant boundary traction acting on half of the upper boundary (see Figure 1). Along the right-hand side and the lower boundary the disc is fixed and the remaining part of the boundary (including the crack) is left free. This problem is interesting as its solution develops a singularity at the tip of the crack where a stress concentration occurs with an asymptotic behaviour of the form $\sigma \approx r^{-1/2}$ (expressed in terms of polar coordinates (r, θ)). The material parameters are chosen as commonly used for aluminium, i.e., $\mu = 80193.8 N/mm^2$, $\kappa = 164206 N/mm^2$ (see, e.g., [16]). For our tests, we consider the computation of the total normal stress over the clamped part of the boundary which is measured by the ("irregular") functional

$$J(\varphi) = \int_{\Gamma_D} n \cdot C \varepsilon(\varphi) \cdot n \, ds. \quad (17)$$

Three different mesh refinement processes are compared, where the solution on a very fine adapted mesh with about 200000 cells is taken as *reference* solution u_{ref} for determining

the relative error $RelErr$ on coarser meshes, while $Ratio$ is the overestimation factor of the error estimator,

$$RelErr := \frac{|J(u_h - u_{ref})|}{|J(u_{ref})|}, \quad Ratio := \frac{\eta(u_h)}{|J(u_h - u_{ref})|}.$$

The parameters L and N indicate the refinement level and the corresponding number of cells, respectively. In the computation the error functional $J(\cdot)$ has been regularised as described above.

1) On a sequence of uniformly refined meshes, the *weighted* estimator (13) with the approximation (15) is evaluated. The results are shown in Table 1.

Table 1

The *weighted* error estimator on uniformly refined meshes (from [17])

L	N	$J(u_h)$	Rel.Err	Ratio
1	256	0.017080	0.0283	1.80
2	1024	0.019528	0.0181	1.76
3	4096	0.021137	0.0113	1.70
4	16384	0.022161	0.0070	1.66
5	65536	0.022802	0.0043	1.62
	∞	0.023850		

2) A sequence of refined meshes is generated on the basis of the energy-error estimator (16). The so-called *fixed fraction strategy* is used, i.e., in each refinement cycle the elements are ordered according to the size of the local error indicators and then a fixed fraction, here 30%, of all elements is refined resulting in about a doubling of the number N of cells. The results are shown in Table 2 and in Figure 2. The overestimation factor "Ratio" is stated, though the global energy-error estimator $\eta_E(u_h)$ cannot be expected to give a reliable bound for the local quantity $J(\cdot)$.

3) A sequence of refined meshes is generated on the basis of the *weighted* error estimator (13) with the approximation (15). This time the so-called *fixed reduction strategy* is used, i.e., in each refinement cycle the varying tolerance is reduced by a fixed factor to $TOL_L = 0.5\eta(u_h)$. Since, as mentioned above, the functional $J(\cdot)$ is irregular, it is regularised in computing the dual solution according to

$$J_\epsilon(\varphi) := \frac{1}{|\Gamma_\epsilon|} \int_{\Gamma_\epsilon} n \cdot C\epsilon(\varphi) \cdot n \, dx, \quad \Gamma_\epsilon = \{x \in \mathbb{R}^2, \text{dist}(x, \Gamma_D) < \frac{1}{2}\epsilon\}, \quad (18)$$

with $\epsilon = TOL_L$. The results obtained are shown in Table 2 and in Figure 2.

In this test case the *weighted* estimator leads to slightly more economical meshes than the energy estimator but also provides a reliable (as well as efficient) bound for the error functional considered. We remark that the square-geometry used in this test is particularly favorable for the energy error estimator η_E as the necessity of properly resolving the strong crack-singularity dominates the computation of the localized quantity $J(u)$. The superiority of our selective estimator over the energy-error estimator would be more pronounced if the distance between the corner and the support of $J(\cdot)$ were larger (see the examples in [7]).

Table 2

Adaptive refinement based on the *energy-error* estimator (left) and the *weighted* error estimator (right) (from [17])

L	N	$J(u_h)$	Rel.Err	Ratio	L	N	$J(u_h)$	Rel.Err	Ratio
4	2659	0.020528	0.0139	0.79	4	2113	0.022157	0.0070	1.96
5	6193	0.021538	0.0096	0.80	5	4435	0.022795	0.0044	1.92
6	13423	0.022319	0.0064	0.84	6	8830	0.023198	0.0027	1.86
7	31336	0.022811	0.0043	0.86	7	15886	0.023428	0.0017	1.79
8	65332	0.023153	0.0029	0.90	8	29947	0.023593	0.0010	1.79

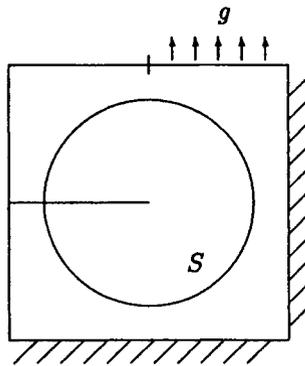


Figure 1. Geometry sketch of the test problem "square disc with slit"

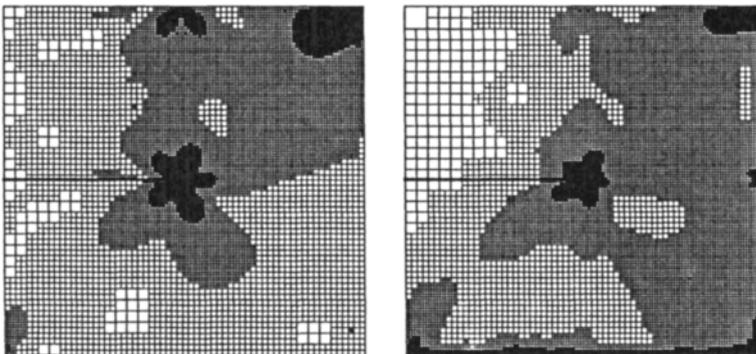


Figure 2. Finest meshes obtained by the *energy-error* estimator (left) and the *weighted* error estimator (right) (from [17])

4. The elasto-plastic case

Next, we consider the elasto-plastic case. As starting point for discretising problems of Prandtl–Reuss–type, we choose the primal–mixed formulation (3). Splitting the continuous loading process $f(t), g(t)$ into a sequence of incremental load steps,

$$f^n = f^{n-1} + k_n f_0, \quad g^n = g^{n-1} + k_n g_0,$$

with pseudo-time step lengths $k_n := t_n - t_{n-1}$, we obtain a sequence of Hencky–type problems,

$$\begin{aligned} (A\sigma^n, \tau - \sigma^n) - (A\sigma^{n-1} + k_n \varepsilon(v^n), \tau - \sigma^n) &\geq 0 \quad \forall \tau \in \Pi L^2(\Omega)_{sym}^{d \times d}, \\ (\sigma^n, \varepsilon(\varphi)) &= (f^n, \varphi) + (g^n, \varphi)_{\Gamma_N} \quad \forall \varphi \in V, \end{aligned} \quad (19)$$

with the initial values $\sigma^0 = 0$ and $u^0 = 0$. According to the above discussion, each such *load step* is equivalent to a nonlinear problem of the form

$$(C(\varepsilon(v^n)), \varepsilon(\varphi)) = (f^n, \varphi) + (g^n, \varphi)_{\Gamma_N} \quad \forall \varphi \in V, \quad (20)$$

where

$$C(\varepsilon(v^n)) := \Pi(\sigma^{n-1} + k_n C\varepsilon(v^n)).$$

Solving this equation gives us an update for the displacement velocity $v^n \in V$ and then the new stress and displacement as

$$\sigma^n := \Pi(\sigma^{n-1} + k_n C\varepsilon(v^n)) \in \Pi L^2(\Omega)_{sym}^{d \times d}, \quad u^n := u^{n-1} + k_n v^n \in V.$$

This process is equivalent to applying the backward Euler time–stepping scheme to the nonstationary problem (3). We assume that the incremental steps $k_n f_0$, and $k_n g_0$ are chosen small enough that the resulting time–discretization error can be neglected compared to the error resulting from the finite element approximation. Our numerical tests show that this is realistic, at least for the type of problems considered. The stationary Hencky model (4) may be viewed as the approximation of the Prandtl–Reuss model by one time step of length $k = 1$ starting from the initial state $\sigma^0 = 0, u^0 = 0$.

The nonlinear problems (20) are approximated by a damped Newton iteration. Starting from the result at the preceding load level, $v^{n,0} := v^{n-1}$, the step $v^{n,i-1} \rightarrow v^{n,i}$ reads

$$(C'(\varepsilon(v^{n,i-1}))\varepsilon(\delta v^{n,i}), \varepsilon(\varphi)) = (f^n, \varphi) + (g^n, \varphi)_{\Gamma_N} - (C(\varepsilon(v^{n,i-1})), \varepsilon(\varphi)) \quad \forall \varphi \in V, \quad (21)$$

with the Jacobian $C'(\cdot)$ of the function $C(\cdot)$, followed by the update

$$v^{n,i} = v^{n,i-1} + \lambda_i \delta v^{n,i}.$$

The damping parameter $0 < \lambda_i \leq 1$ is determined in the form $\lambda_i = 2^{-r}$ such that the residual norm is decreased. Hence, on the continuous level the nonlinear plasticity problem (3) is reduced to a sequence of linear problems of the form

$$(C'(\varepsilon(v))\varepsilon(w), \varepsilon(\varphi)) = l(\varphi) \quad \forall \varphi \in V, \quad (22)$$

with certain functionals $l(\cdot)$ defined on V . These are now discretized by the finite element method,

$$(C'(\varepsilon(v_h))\varepsilon(w_h), \varepsilon(\varphi)) = l(\varphi) \quad \forall \varphi \in V_h, \quad (23)$$

where V_h is a finite element space as defined before. The resulting linear subproblems are solved by the CR-method with multigrid acceleration. The use of multigrid is rather natural as in the course of the mesh refinement process a sequence of nested meshes is automatically generated. Since the theme of this note is the aspect of a *posteriori* error estimation and mesh design, we do not go into the details of the algebraic solution techniques and instead refer to [19] and [20].

Again our error estimation uses the general approach described above applied to the stationary nonlinear problems (20). In each load step a stationary Hencky-type problem is solved using an adaptive finite element discretization. The finite element meshes are optimized separately within each load step in accordance to the particular target functional $J(\cdot)$ leading to a dynamic development of refinement and coarsening over the whole loading process. We remark that, in contrast to really dynamic problems involving acceleration terms, the control of the time-stepping error in the quasi-stationary Prandtl-Reuss model is less critical. Since in each incremental load step the displacement velocity $v^n = \dot{u}^n$ is updated, we do not expect much accumulation of these local errors over the loading path. Hence it seems justified to estimate the global spatial error in this process by simply treating it as a sequence of stationary problems. This approach, of course, would be disastrous in the case of a parabolic problem like the heat equation.

The loading process is started from the initial state $\sigma^0 \equiv 0$. In each load step $t_{n-1} \rightarrow t_n$, the nonlinear problem (20) has to be solved resulting in an equation for the discrete deformation velocity $v_h^n \in V_h$,

$$(\Pi(\sigma_h^{n-1} + k_n C \varepsilon(v_h^n)), \varepsilon(\varphi)) = (f^n, \varphi) + (g^n, \varphi)_{\Gamma_N} \quad \forall \varphi \in V_h. \quad (24)$$

From this, the stress update is obtained,

$$\sigma_h^n := \Pi(\sigma_h^{n-1} + k_n C \varepsilon(v_h^n)), \quad (25)$$

where the projection Π is as defined above. Further, starting from the initial displacement $u_h^0 \equiv 0$, corresponding approximate displacements are obtained,

$$u_h^n := u_h^{n-1} + k_n v_h^n.$$

Hence, (24) may be rewritten as

$$(\Pi(\sigma_h^{n-1} + C \varepsilon(u_h^n - u_h^{n-1})), \varepsilon(\varphi)) = (f^n, \varphi) + (g^n, \varphi)_{\Gamma_N} \quad \forall \varphi \in V_h. \quad (26)$$

From now on, we suppress the superscript n and consider Hencky-type problems of the form

$$(\Pi(\hat{\sigma} + C \varepsilon(u_h - \hat{u})), \varepsilon(\varphi)) = (f, \varphi) + (g, \varphi)_{\Gamma_N} \quad \forall \varphi \in V_h, \quad (27)$$

with given initial stress $\hat{\sigma}$ and deflection \hat{u} . From (27), we obtain deflections $u_h \in V_h$ and associated stresses

$$\sigma_h := \Pi(\hat{\sigma} + C \varepsilon(u_h) - C \hat{u}),$$

approximating the solutions $u \in V$ and $\sigma \in \Pi L^2(\Omega)_{sym}^{d \times d}$ of the corresponding continuous problem

$$(\Pi(\hat{\sigma} + C\varepsilon(u - \hat{u})), \varepsilon(\varphi)) = (f, \varphi) + (g, \varphi)_{\Gamma_N} \quad \forall \varphi \in V. \quad (28)$$

Clearly, for $\hat{\sigma} = 0$ and $\hat{u} = 0$, this reduces to the usual Hencky model (5).

4.1. A posteriori error estimate for the Hencky model

Combining (27) and (28), we obtain the *nonlinear* Galerkin orthogonality relation

$$(C(\varepsilon(u)) - C(\varepsilon(u_h)), \varepsilon(\varphi)) = \int_0^1 (C'(\varepsilon(su + (1-s)u_h))\varepsilon(u - u_h), \varepsilon(\varphi)) ds = 0, \quad (29)$$

for $\varphi \in V_h$. Suppose that the quantity $J(u)$ has to be computed. For representing the error $J(e)$, we use the solution z of the linear dual problem

$$L^*(u, u_h; z, \varphi) = J(\varphi) \quad \forall \varphi \in V, \quad (30)$$

with the adjoint of the bilinear form

$$L(u, u_h; \varphi, \psi) := \int_0^1 (C'(\varepsilon(su + (1-s)u_h))\varepsilon(\varphi), \varepsilon(\psi)) ds.$$

We assume that this *dual solution* is well defined. This, of course, requires the underlying problem to satisfy certain structural conditions. By the orthogonality relation (29), there holds

$$J(e) = L(u, u_h; e, z - z_h),$$

with a suitable $z_h \in V_h$. Then, analogously as in the linear case, we obtain the error representation

$$J(e) = \sum_{K \in \mathcal{T}_h} \left\{ (f - \operatorname{div} C(\varepsilon(u_h)), z - z_h)_K - \frac{1}{2} (n \cdot [C(\varepsilon(u_h))], z - z_h)_{\partial K} \right\}.$$

From this, we conclude the following theorem for the elasto-plastic case (see [18]).

Theorem 4.1. *Under the foregoing assumptions, the primal finite element scheme (26) admits the a posteriori error estimate*

$$|J(u - u_h)| \leq \eta(u_h) := \sum_{K \in \mathcal{T}_h} \omega_K \varrho_K, \quad (31)$$

with the local residuals ϱ_K and weight factors ω_K defined by

$$\begin{aligned} \varrho_K &:= \left(h_K^2 \|f - \operatorname{div} C(\varepsilon(u_h))\|_K^2 + \frac{1}{2} h_K \|n \cdot [C(\varepsilon(u_h))]\|_{\partial K}^2 \right)^{1/2}, \\ \omega_K &:= \left(h_K^{-2} \|z - z_h\|_K^2 + \frac{1}{2} h_K^{-1} \|z - z_h\|_{\partial K}^2 \right)^{1/2}. \end{aligned}$$

In order to evaluate the *a posteriori* bound (31), we replace the unknown solution u in the bilinear form $L(u, u_h; \cdot, \cdot)$ by the currently computed approximation u_h , and solve the corresponding perturbed dual problem by the same method as used in computing u_h , yielding an approximation $\tilde{z}_h \in V_h$ to the exact dual solution z ,

$$L^*(u_h, u_h; \tilde{z}_h, \varphi) = J(\varphi) \quad \forall \varphi \in V_h. \quad (32)$$

The weights ω_K may then again be approximated as in the linear case described above. We emphasize that the computation of the weights requires only to solve *linear* problems and normally only amounts to a small fraction of the total cost within a Newton iteration for the nonlinear problem.

We will compare our weighted error estimator (31) against two more traditional approaches.

1) *The ZZ-approach*: The error indicator proposed by Zienkiewicz and Zhu [23] for finite element models in structural mechanics is based on the idea of higher-order stress recovery by local averaging. The element-wise error $\|\sigma - \sigma_h\|_K$ is thought to be well represented by the auxiliary quantity $\eta_K := \|\mathcal{M}_h \sigma_h - \sigma_h\|_K$, where $\mathcal{M}_h \sigma_h$ is a local (super-convergent) approximation of σ . The corresponding (heuristic) global error estimator reads

$$\|\sigma - \sigma_h\| \approx \eta_{ZZ} := \left(\sum_{K \in \mathbb{T}_h} \|\mathcal{M}_h \sigma_h - \sigma_h\|_K^2 \right)^{1/2}. \quad (33)$$

For our purpose we assume the discrete stresses to be constant over each cell. One possible construction of $\mathcal{M}_h \sigma_h$ is the patch-wise L^2 -projection $P_K \sigma_h$ onto the space of (bi-)linear shape functions. Here the nodal value at a point of the triangulation determining $\mathcal{M}_h \sigma_h$ is obtained by averaging the cell-wise constant values of σ_h of those cells having this point in common. For cells containing hanging nodes this process is appropriately modified.

2) *An energy error estimator*: Johnson and Hansbo [13] proposed an error estimator for the primal-mixed formulation of the Hencky model which is based on monotonicity properties of the energy form and, under some additional heuristic assumptions, bounds the error in the global energy norm. Let Ω_h^e and Ω_h^p denote the union of elements where the discrete solution behaves elastic and plastic, respectively. Then, the estimator reads

$$\|\sigma - \sigma_h\| \approx \eta_E := C_i \left(\sum_{K \in \mathbb{T}_h} \eta_K^2 \right)^{1/2}, \quad (34)$$

with the local error indicators

$$\eta_K^2 := \begin{cases} h_K^4 \max_K |R(u_h)|^2, & \text{if } K \in \Omega_h^e, \\ h_K \max_K |C\varepsilon(u_h)| \int_K |R(u_h)| dx, & \text{if } K \in \Omega_h^p, \end{cases}$$

where on each element $K \in \mathbb{T}_h$ the local residual is defined by

$$R(u_h) := |\operatorname{div} C\varepsilon(u_h)| + \frac{1}{2} h_K^{-1} |[n \cdot C\varepsilon(u_h)]|.$$

Here, C_i is some interpolation constant usually set to one. This estimator is rather heuristic, as it relies on the assumption that the plastification zone is already correctly captured on the current mesh. Furthermore, it is of only sub-optimal order in the plastic

zone which results in mesh over-refinement in Ω_h^p , though the stresses are suspected to be rather smooth there. The ZZ-estimator (33) does not suffer from this deficiency as it essentially relies on the smoothness of σ . Hence, we are led to modify the estimator (34) by replacing the obviously too crude bound $\max_K |C\varepsilon(u_h)|$ in the plastic zone by $\max_K |C\varepsilon(u_h) - \mathcal{M}_h C\varepsilon(u_h)|$. This gives us the local (still heuristic) error indicators

$$\eta_K^2 := \begin{cases} h_K^4 \max_K |R(u_h)|^2, & \text{if } K \in \Omega_h^e, \\ h_K \max_K |C\varepsilon(u_h) - \mathcal{M}_h C\varepsilon(u_h)| \int_K |R(u_h)| dx, & \text{if } K \in \Omega_h^p. \end{cases}$$

4.2. Numerical test for the Hencky model

As the first example, we take the model case "disc with crack" from above (c.f. [18]). The material values are as before $\kappa = 164206$, $\mu = 80193.8$, and further $\sigma_0 = 450$. The boundary traction is assumed in the form $g = tg_0$, with $g_0 = 100$ and $t > 0$. In the case of the Hencky model, we choose $t_{lim} = 2.234$, since already for $t = 2.3$ plastic collaps occurs. Again, we compute the mean normal stress over the clamped part of the boundary,

$$J_\sigma(u) = \int_{\Gamma_D} n \cdot C(\varepsilon(u)) \cdot n ds. \tag{35}$$

The refinement strategy is as described above and the quantities *Rel.Err* and *Ratio* have the same meaning as before.

Our weighted error estimator turns out to be rather sharp even on relatively coarse meshes, see Table 3. This indicates that the strategy of evaluating the weights ω_K computationally works also for the present nonlinear problem. Further, this approach yields more economical meshes than the other error estimators (Figure 3). This is in agreement with our observations in the linear elastic case.

Table 3

Results for $J_\sigma(u_h)$ with adaptivity based on the weighted *a posteriori* error estimator (from [18])

N	$J_\sigma(u_h)$	<i>Rel.Err</i>	<i>Ratio</i>
1000	2.2224e+02	1.5757e-02	1.6650e+00
4000	2.2405e+02	7.7387e-03	1.7360e+00
8000	2.2475e+02	4.6647e-03	1.6205e+00
12000	2.2509e+02	3.1559e-03	1.5455e+00
16000	2.2532e+02	2.1360e-03	1.7726e+00

As a second test, we apply our adaptive approach to a standard benchmark problem described as follows (c.f. [20]). A geometrically 2-dimensional square disc with a hole is subjected to a constant boundary traction acting upon two opposite sides.

We use again the plain-strain approximation and assume perfectly plastic material behavior. In virtue of symmetry the consideration can be restricted to a quarter of the domain as shown in Figure 4. The height and width of the quarter corresponding to lines $\overline{45}$ and $\overline{15}$ are 100, and the radius of the hole is 10. The material parameters are the same as for the example above. The boundary traction is written in the form $g = tg_0$, with $g_0 = 100$ and load factor $t \in [0, 6]$.

Among the quantities to be computed are the component σ_{22} of the stress tensor and the horizontal deflection u_1 at point 2, and the horizontal displacement u_1 at point 5. Again, the result on a very fine adapted mesh with about 200000 cells is taken as reference solution u_{ref} . Some of the results of these benchmark computations are summarized in Figures 5, 8 and 9. They show that, comparing the relative errors on adaptive grids produced by the "weighted", the "energy" and the "ZZ"-approach, the first one yields more economical meshes. We remark that in this case the additional cost for the adaptive process is equivalent to two steps of the Newton iteration and therefore amounts to less than 20 % of the total cost.

4.3. Decoupled pressure discretisation

We employ a stabilised finite element discretisation corresponding to (6) in the form (c.f. [12])

$$\begin{aligned} (\Pi(2\mu\varepsilon^D(u_h)), \varepsilon(\varphi)) + (\operatorname{div} \varphi, p_h) &= (f, \varphi) \quad \forall \varphi \in V_h, \\ (\operatorname{div} u_h, q) - \frac{1}{\kappa}(p_h, q) - \sum_{K \in \mathcal{T}_h} \delta_K h_K^2 (\nabla p_h, \nabla q)_K &= 0 \quad \forall q \in Q_h, \end{aligned} \quad (36)$$

where V_h and Q_h are constructed by the standard conforming bilinear shape functions. For comparing the accuracy of the standard primal formulation to the mixed displacement/pressure formulation, we use the first model problem "disc with a crack" and consider the computation of the integral $J_u(u)$

$$J_u(u) = \int_S u \cdot n \, ds = \int_{\Omega_S} \operatorname{div} u \, dx, \quad (37)$$

where S is a suitable circular path around the tip of the crack, Ω_S is the domain with boundary $\partial\Omega_S = S$, and n the outer normal unit vector along S . The results are presented in Table 4. It turns out that by the displacement/pressure formulation significantly higher accuracy can be achieved.

Table 4

Results for $J_u(u_h)$ on adaptive grids for the primal and the displacement/pressure discretisation (from [21])

N	$J_u(u_h)$	
	primal	pressure
4000	1.6875e-04	1.696680e-04
8000	1.6926e-04	1.699004e-04
16000	1.6963e-04	1.699354e-04
32000	1.6986e-04	1.700872e-04
∞	1.7020e-04	

4.4. Numerical test for the Prandtl-Reuss model

Finally, we present some results obtained with our adaptive method for the time dependent Prandtl-Reuss model. The test case is again the benchmark problem "disc with

a whole" (see Figure 4) and, this time, we compute the stress component σ_{11} at point 7 by the scheme (24). The resulting sequence of meshes for $k = 0.0025$ with a constant number of $N \approx 3200$ mesh cells is shown in Figure 7. Further, we test the dependence of the solution on variations of the loading step k and the number N of mesh cells. Figure 6 (left) shows the results for a sequence of reduced loading steps $k = 0.08 - 0.01$ with a fixed number of $N = 8000$ mesh cells. Figure 6 (right) shows the results obtained with a fixed load-step $k = 0.01$ for a sequence of adapted meshes with increasing numbers of mesh cells $N = 1000 - 8000$.

Eventually, in Table 5 we show the dependence of the critical load level t_f for which plastification begins and the limit load level t_b at which final break down occurs.

Table 5

Critical load level t_f and limit load level t_b computed with fixed $k = 0.01$ and varying N (left) and for varying k and fixed $N = 8000$ (right) (from [20])

N	t_f	t_b	k	t_f	t_b
1000	3.095	4.655	0.08	1.880	4.680
2000	2.975	4.655	0.04	1.860	4.660
4000	2.485	4.655	0.02	1.950	4.655
8000	2.115	4.655	0.01	2.115	4.655

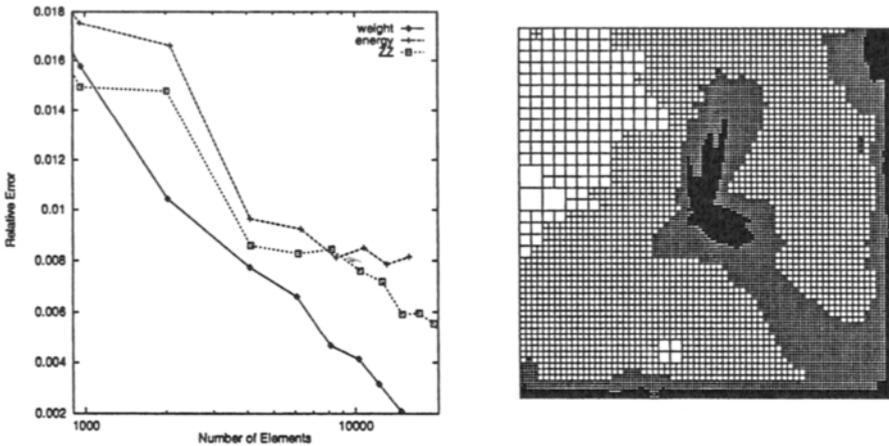


Figure 3. Relative error for $J_\sigma(u_h)$ on grids based on the different estimators and structure of "optimal" grid for $J_\sigma(u_h)$ with $N \approx 8100$ (from [18])

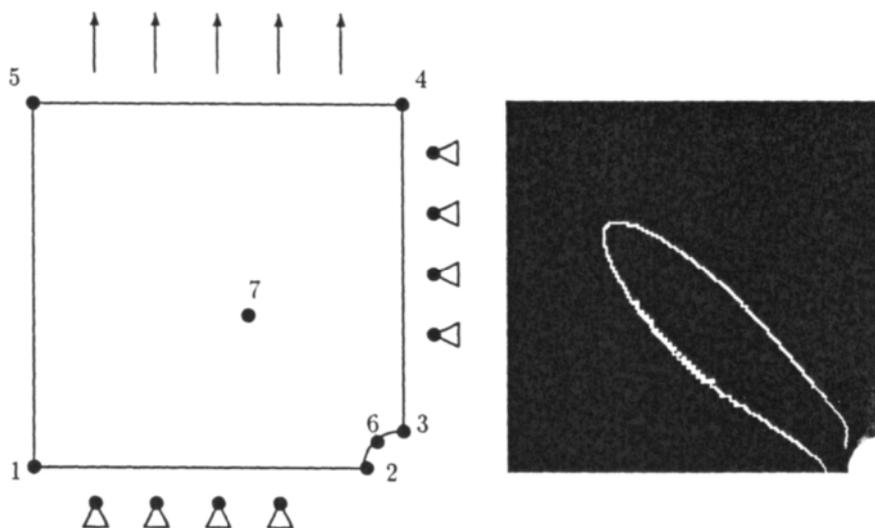


Figure 4. Geometry of the benchmark problem and plot of $|\sigma^D|$ (plastic region black, transition zone white) computed on a mesh with $N \approx 10000$ cells

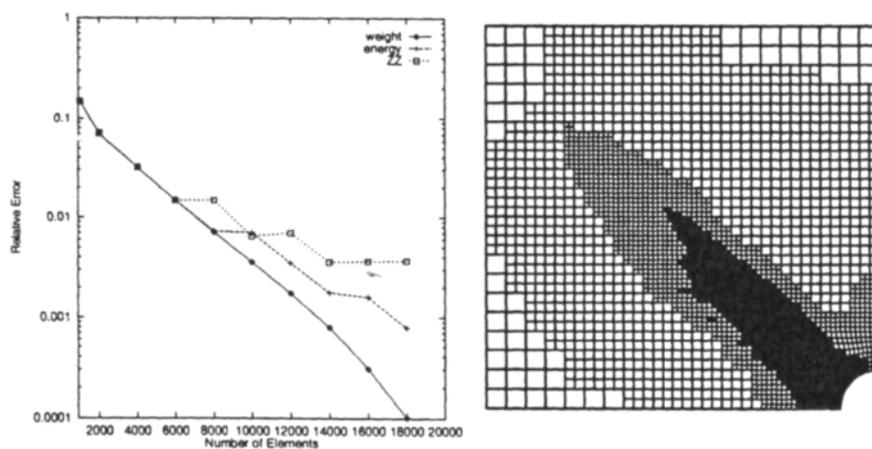


Figure 5. Relative error for computation of σ_{22} at point 2 using different estimators and "optimal" grid with about 10000 cells (from [20])

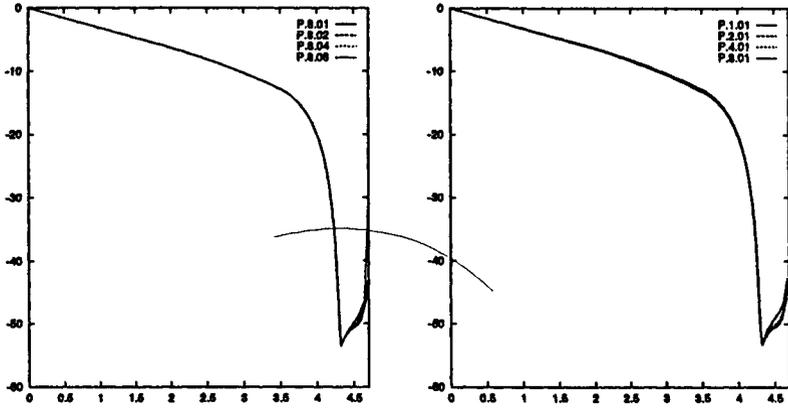


Figure 6. Computation of σ_{11} at point 7 over the loading pass $1.5 \leq t \leq 5$: Dependence on the loading step k (left) and on the number of mesh cells N (right) (from [20])

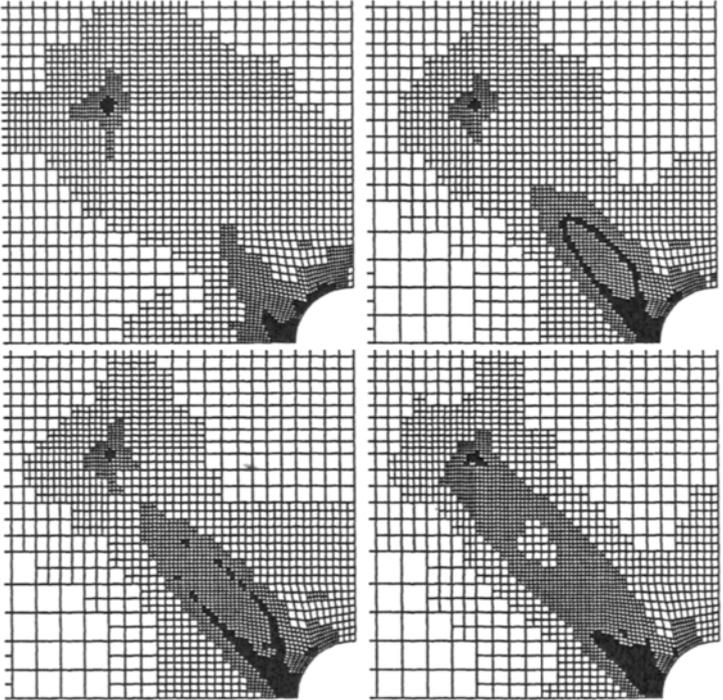


Figure 7. Computation of σ_{11} at point 7 over the loading pass $3.6 \leq t \leq 4.2$: Sequence of adapted grids (zoom) with about 3200 mesh cells (from [19])

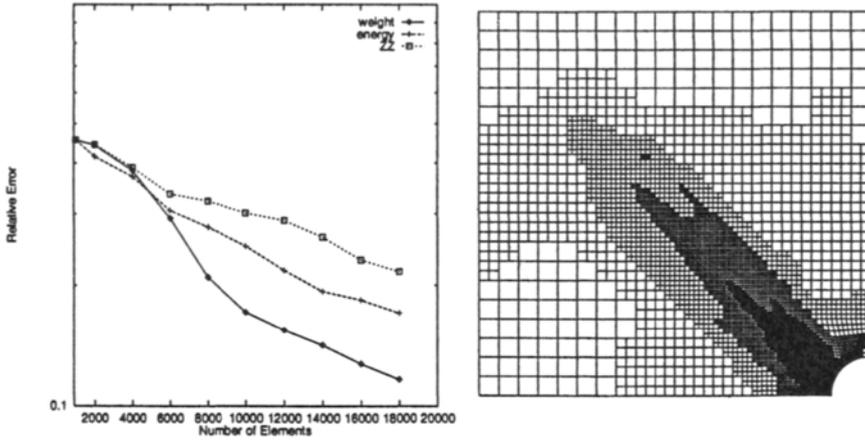


Figure 8. Relative error for computation of u_1 at point 2 using different estimators and "optimal" grid with about 10000 cells (from [20])

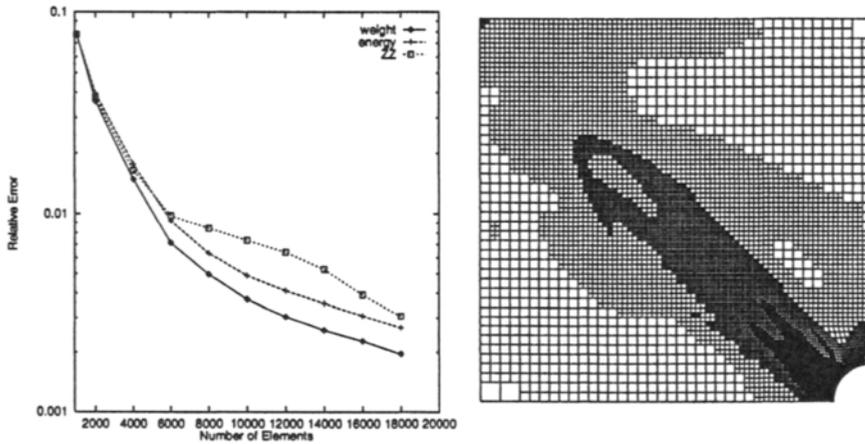


Figure 9. Relative error for computation of u_1 at point 5 using different estimators and "optimal" grid with about 10000 cells (from [20])

REFERENCES

1. I. Babuška and W.C. Rheinboldt: *Error estimates for adaptive finite element computations*, SIAM J. Num. Anal. 15, 736-754 (1978).
2. E. Backes: *Gewichtete a posteriori Fehleranalyse bei der adaptiven Finite-Elemente-Methode: Ein Vergleich zwischen Residuen- und Bank-Weiser-Schätzer*, Diplomarbeit, Institut für Angewandte Mathematik, Universität Heidelberg, 1997.

3. R.E. Bank and A. Weiser: *Some a posteriori error estimators for elliptic partial differential equations*, Math. Comput. 44, 283-301 (1985).
4. R. Becker: *An Adaptive Finite Element Method for the Incompressible Navier-Stokes Equations on Time-Dependent Domains*, Dissertation, Institut für Angewandte Mathematik, Universität Heidelberg, 1995.
5. R. Becker, G. Kanschat and F.T. Suttmeier: *DEAL - differential equations analysis library*, available via <http://gaia.iwr.uni-heidelberg.de/DEAL.html>, 1995
6. R. Becker and R. Rannacher: *Weighted a posteriori error control in FE methods*, Preprint 96-1 (SFB 359), Universität Heidelberg, Proc. ENUMATH'95, Paris, 18-22 Sept., 1995.
7. R. Becker and R. Rannacher: *A feed-back approach to error control in finite element methods: Basic analysis and examples*, EAST-WEST J. Numer. Math 4, 237-264 (1996).
8. S.C. Brenner and R.L. Scott: *The Mathematical Theory of Finite Element Methods*, Springer, Berlin-Heidelberg-New York, 1994.
9. W. Dörfler and M. Rumpf: *An adaptive strategy for elliptic problems including a posteriori controlled boundary approximation*, Preprint Nr. 1/1996-16.01.1996, Mathematische Fakultät, Universität Freiburg (1996).
10. G. Duvaut and J. L. Lions: *Inequalities in Mechanics and Physics*, Springer, Berlin-Heidelberg-New York, 1976.
11. K. Eriksson and C. Johnson: *An adaptive finite element method for linear elliptic problems*, Math. Comp. 50, 361-383 (1988).
12. L.P. Franca and R. Stenberg: *Error analysis of some Galerkin least squares methods for the elasticity equations*, SIAM J. Numer. Anal. 28 (1991).
13. C. Johnson and P. Hansbo: *Adaptive finite element methods in computational mechanics*, in Computer Methods in Applied Mechanics and Engineering 101, North-Holland, Amsterdam, 1992.
14. C. Johnson: *Existence theorems for plasticity problems*, J.Math. Pures Appl. 55, 1976
15. G. Kanschat: *Parallel and Adaptive Galerkin Methods for Radiative Transfer Problems*, Dissertation, Institut für Angewandte Mathematik, Universität Heidelberg, 1996.
16. T. Lehmann and U. Blix: *On a coupled thermo-mechanical process in the necking problem*, Int. J. Plasticity 1, 175-188 (1985).
17. R. Rannacher and F.T. Suttmeier: *A feed-back approach to error control in finite element methods: Application to linear elasticity*, Comp. Mech. 19/5, 1997
18. R. Rannacher and F.T. Suttmeier: *A posteriori error control in finite element methods via duality techniques: application to perfect plasticity*, Preprint, Institut für Angewandte Mathematik, Universität Heidelberg, 1997.
19. F.-T. Suttmeier: *Adaptive Finite Element Approximation of Problems in Elasto-Plasticity Theory*, Thesis, Institut für Angewandte Mathematik, Universität Heidelberg, 1996.
20. F.-T. Suttmeier: *An Adaptive Finite Element Scheme for a Benchmark Problem in Perfect Plasticity*, Preprint, Institute for Applied Mathematics, University of Heidelberg.
21. F.-T. Suttmeier: *An Adaptive Displacement/Pressure Finite Element Scheme for Treating Incompressibility Effects in Elasto-Plastic Materials*, Preprint, Institute for Applied Mathematics, University of Heidelberg.
22. R. Verfürth: *A Review of A Posteriori Estimation and Adaptive Mesh-Refinement Techniques*, Advances in Numerical Mathematics, Wiley/Teubner, New York-Stuttgart, 1996.
23. O.C. Zienkiewicz and J.Z. Zhu: *A simple error estimator and adaptive procedure for practical engineering analysis*, Int. J. Numer. Methods Engrg. 24 (1987).

An adaptive finite element approach in associated and non-associated plasticity considering localization phenomena

W. Wunderlich, H. Cramer, G. Steinl,

Technische Universität München, Germany

Abstract

In an adaptive finite element approach elasto-plastic problems with associated and non-associated flow rules are investigated. The first part of the paper deals with the underlying numerical formulation for a classical continuum model and a hierarchical h -adaptive mesh refinement strategy. Essential ingredients of the adaptive process are a suitable error indicator and transfer operations for the mapping of history-dependent state variables between different meshes. These are imbedded in a nonlinear incremental finite element procedure. For non-associated plasticity a standard continuum approach may lead to an ill-posed problem. Therefore, in the second part a generalization in the framework of a Cosserat theory is considered. The underlying equations possess a similar structure, and the adaptive finite element formulation can be extended in a straightforward manner. Numerical examples demonstrate the general applicability of the approach to elastic-plastic problems including associated as well as non-associated plasticity. They show the superior behaviour of the Cosserat formulation in the case of localization phenomena also for non-associated plasticity.

1. INTRODUCTION

The quality of finite element solutions of elastic plastic problems is very much influenced on how the structural system is discretized, especially when the load carrying capacity has to be determined. Adaptive mesh refinement strategies provide a good basis to improve the accuracy of the solutions. Essential ingredients of adaptive finite element calculations are the estimation of the discretization error and the design of the refined meshes. In the nonlinear analysis of path-dependent problems the mesh refinement has to be controlled by suitable error measures and integrated into the incremental solution procedure.

In the first part of the paper we present an adaptive finite element formulation for a classical continuum model including associated and non-associated plasticity. A standard displacement formulation with conforming shape functions for the displacements is used as the basis of the underlying finite element approach. The mesh refinement is controlled by an error indicator calculated after each load step in the course of the nonlinear incremental solution. This indicator represents a modification of an error estimator derived in [1] which takes into account the stress residuals within the elements and on their boundaries as well as the incremental strains.

Based on the error distributions successive refined meshes are generated by simply subdividing those elements for which the error exceeds a permissible tolerance. This leads to a hierarchical finite element approach in which the shape functions of the coarser meshes are retained in the basis of the shape functions of the finer meshes. In path-dependent problems this facilitates the mapping of state variables like stresses and internal variables between different meshes.

The performance of the approach is demonstrated by the results of numerical calculations for different problems. The results illustrate the necessity of adaptive calculations, especially when limit states are investigated.

In non-associated plasticity localization phenomena are encountered which may lead to an ill-posed problem when classical continuum theory is applied. Therefore, in the second part of the paper a generalization of the approach in the framework of a Cosserat continuum model is presented. In this formulation an internal length scale is introduced which leads to strain concentrations in shear bands of finite width. This is illustrated by the results of numerical calculations which also demonstrate the superior behaviour of the extended formulation in contrast to the standard continuum model. The problem common to all formulations including an internal length scale is that they need an extremely fine discretization. Therefore, an adaptive approach is absolutely necessary for the analysis of problems where localization phenomena have to be taken into account.

2. FORMULATION OF THE CLASSICAL ELASTIC-PLASTIC FINITE ELEMENT MODEL

The governing equations of the standard elastic plastic continuum model are outlined first. The equilibrium is defined by the local equations

$$\sigma_{ji,j} + \bar{f}_i = 0 \quad \text{in } V, \quad (1)$$

$$\sigma_{ji} n_j - \bar{t}_i = 0 \quad \text{on } S_s. \quad (2)$$

A variational or weak form of these equations is given by

$$-\int_V \delta u_i (\sigma_{ji,j} + \bar{f}_i) dV + \int_{S_s} \delta u_i (\sigma_{ji} n_j - \bar{t}_i) dS = 0, \quad (3)$$

which is equivalent to the common principle of virtual work:

$$\int_V \delta \varepsilon_{ij} \sigma_{ij} dV - \int_V \delta u_i \bar{f}_i dV - \int_{S_s} \delta u_i \bar{t}_i dS = 0. \quad (4)$$

The constitutive law applied is based on an elastic-plastic model with isotropic hardening. The total strain increment is decomposed into an elastic and a plastic part

$$d\varepsilon_{ij} = d\varepsilon_{ij}^e + d\varepsilon_{ij}^p, \quad (5)$$

where $d\varepsilon_{ij}$, $d\varepsilon_{ij}^e$ and $d\varepsilon_{ij}^p$ denote the total, elastic and plastic strain components, respectively. For infinitesimal deformations the kinematic relations are given by

$$d\varepsilon_{ij} = \frac{1}{2} (du_{i,j} + du_{j,i}). \quad (6)$$

The elastic constitutive relation is defined by a hypoelastic stress–strain law

$$d\varepsilon_{ij}^e = C_{ijkl} d\sigma_{kl} . \quad (7)$$

For the description of the plastic response the yield surface is expressed in terms of the stress invariants I_1 , J_2 and J_3 , representing the first invariant of the stress tensor and the second and third invariant of the stress deviator, respectively :

$$f(\sigma, \kappa) = f(I_1, J_2, J_3, \kappa) = 0 , \quad (8)$$

$$\text{where } I_1 = \sigma_{ij} \delta_{ij} , \quad s_{ij} = \sigma_{ij} - \frac{1}{3} I_1 \delta_{ij} ,$$

$$J_2 = \frac{1}{2} s_{ij} s_{ij} , \quad J_3 = \frac{1}{3} s_{ij} s_{jk} s_{ki} .$$

Isotropic hardening is included by an internal variable κ .

A special type of yield function that is used in our numerical examples is given by

$$f = m(\theta) \bar{\sigma} + \kappa (\eta I_m - k) \leq 0 , \quad (9)$$

$$\text{where } m(\theta) = \left(\frac{(1 - \sin 3\theta) + \gamma^4 (1 + \sin 3\theta)}{2 \gamma^4} \right)^{0.25} ,$$

$$\bar{\sigma} = \sqrt{J_2} , \quad \sin(3\theta) = -\frac{\sqrt{27} J_3}{2 \sqrt{J_2^3}} .$$

This allows the representation of different yield criteria like von Mises, Drucker–Prager as well as a Coulomb–type criterion with a smooth yield surface in the π –plane sections. A graphical representation in the principal stress space is given in Figure 1.

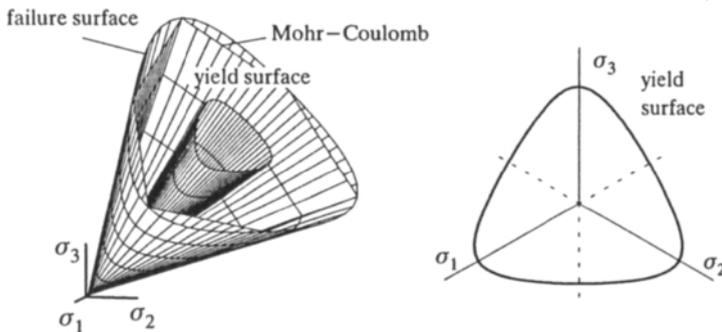


Figure 1: Graphical representation of the yield and failure surfaces in principal stress space and in the π –plane

Plastic deformations are defined by the flow rule in the form

$$d\varepsilon_{ij}^p = \frac{\partial g}{\partial \sigma_{ij}} d\lambda , \quad d\lambda \geq 0 , \quad (10)$$

where g denotes the plastic potential. Classical associated plasticity is included with $g \equiv f$, whereas the more general non–associated case is obtained by a definition of g which is independent of f . In the numerical calculations a special form of plastic potential is used that allows the description of both associated and non–associated plastic flow:

$$g(\sigma) = f(\sigma) - \bar{\eta} I_1 . \quad (11)$$

The consistency condition

$$df = \frac{\partial f}{\partial \sigma_{ij}} d\sigma_{ij} + \frac{\partial f}{\partial \kappa} d\kappa = 0 \quad (12)$$

and a law for the evolution of the parameter κ representing strain-hardening or work-hardening completes the equations of the elastic plastic model:

$$\frac{\partial f}{\partial \kappa} d\kappa = \frac{\partial f}{\partial \kappa} \frac{\partial \kappa}{\partial \varepsilon_{ij}^p} \frac{\partial g}{\partial \sigma_{ij}} d\lambda = -h d\lambda \quad (13)$$

The finite element formulation follows the common path based on the principle of virtual work. Generally, eight-node or nine-node quadrilaterals are used for plane strain problems with standard isoparametric displacement functions within each element:

$$\mathbf{u} = \mathbf{N} \mathbf{v} \quad (14)$$

\mathbf{N} represents the shape functions and \mathbf{v} are the generalized displacements at the element nodes. The strain field may be computed by the discrete strain matrix \mathbf{B} by

$$\boldsymbol{\varepsilon} = \mathbf{B} \mathbf{v} \quad (15)$$

The discrete version of the equilibrium equations results in

$$\int_V \mathbf{B}^T \boldsymbol{\sigma} dV = \int_V \mathbf{N}^T \bar{\mathbf{f}} dV + \int_{S_i} \mathbf{N}^T \bar{\mathbf{t}} dS \quad (16)$$

The stress integration algorithm follows a conventional linearization procedure of the nonlinear incremental equations where the elastic-plastic constitutive relations are evaluated locally at the Gauss points.

3. ERROR INDICATORS AND ADAPTIVE MESH REFINEMENT

A-posteriori error estimators for finite element solutions of linear elliptic problems are usually based on bilinear forms associated with the energy norm. As an extension to the case of associated plasticity, an appropriate bilinear form for the linearized elastic-plastic problem can be given that represents an incremental form of the complementary free energy function as:

$$\Psi_t = \int_V d\varepsilon_{ij} d\sigma_{ij} dV = \int_V d\varepsilon_{ij}^e d\sigma_{ij} dV + \int_V d\varepsilon_{ij}^p d\sigma_{ij} dV \quad (17)$$

Inserting Equations (7), (10), (12) and (13) under the condition that Drucker's stability postulates apply leads to a quadratic form which is positive for a hardening material:

$$\Psi_t = \int_V d\sigma_{ij} C_{ijkl} d\sigma_{kl} dV + \int_V h d\lambda^2 dV \quad (18)$$

A corresponding global error of the finite element solution may be obtained in terms of finite increments of stress and strain by

$$\|e\| = \int_V (\Delta\sigma - \Delta\sigma^h) (\Delta\varepsilon - \Delta\varepsilon^h) dV, \quad (19)$$

where $\Delta\sigma$, $\Delta\varepsilon$ and $\Delta\sigma^h$, $\Delta\varepsilon^h$ are the stress and strain increments of the exact and the finite element solution, respectively. In accordance with the work of Johnson and Hansbo [1] an estimate is used that represents an error measure in which the maximum norms of the residuals of the equilibrium equation within the elements and of the stress jumps on the element boundaries are weighted by the L_1 -norm of the total strain increments:

$$\|e\| = \|\Delta\varepsilon\|_{L_1} (C_1 h \|r\|_{L_\infty} + C_2 \frac{h}{h_k} \|J\|_{L_\infty}), \quad (20)$$

$$\text{where } r_i = \sigma_{ji,j} + \bar{f}_i \quad \text{in } V_e \quad (21)$$

$$\text{and } J_i = \frac{1}{2}(\sigma_{ji}^+ - \sigma_{ji}^-)n_j \quad \text{on } S_e \quad (22a)$$

$$\text{or } J_i = (\sigma_{ji} n_j - \bar{t}_i) \quad \text{on } S_s. \quad (22b)$$

Due to the assumptions stated above the mathematical foundation of the estimator is restricted to associated plasticity. But it still can serve as a reasonable heuristical error indicator also for problems with non-associated plasticity. Due to the combination of the norms of the residuals with the norm of the strain increments it represents a weighted residual that measures errors in the equilibrium equations also in the non-associated case. Moreover, the appearance of the norm of the strain increments in this error measure may also serve for the detection of zones of strain localization.

Within each load step the mesh refinement is controlled by the restriction that the error measure in Equation (20) should be less than a permissible fixed percentage η , say 1–5%, of a similar weighted norm of the stresses themselves:

$$\|e\| \leq \eta \|\Sigma\|, \quad (23)$$

$$\text{where } \|\Sigma\| = \|\Delta\varepsilon\|_{L_1} \|\sigma\|_{L_\infty}. \quad (24)$$

This is satisfied if the restriction for the contribution $\|e\|_e$ of each element holds:

$$\|e\|_e \leq \eta \frac{\|\Sigma\|}{N_{el}}, \quad (25)$$

where N_{el} is the total number of elements. Successive adaptive meshes are generated by a hierarchical subdivision of those elements for which the error exceeds the permissible value. To ensure conformity at the interface between refined and non-refined zones (see Figure 2) the displacements of inconforming nodes are coupled to the neighbouring ones through additional kinematic conditions (hanging nodes).

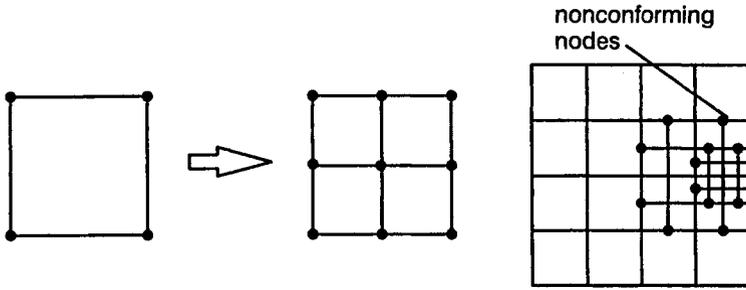


Figure 2: Refinement of 4-node quadrilaterals

An incremental adaptive solution procedure for path-dependent problems requires the projection of various state variables between meshes of different discretization levels. These state variables are displacements, strains, stresses and internal history variables. The hierarchical refinement implies that the shape functions for the displacements of coarser meshes are embedded in the function space of the elements in a refined mesh. Therefore, for the displacement field transfer standard interpolation may be used to determine the discrete displacements in subdivided elements:

$$\hat{\mathbf{u}}^{new}(\xi, \eta) = \sum_{i=1}^{nodes} N_i(\xi, \eta) \hat{\mathbf{U}}_i^{old} . \tag{26}$$

A transfer operator for stresses, strains and internal variables which are defined by their discrete values at the (nip) Gauss points (see Figure 3) is obtained also in a straightforward manner by a simple interpolation:

$$\hat{\mathbf{z}}^{new}(\xi, \eta) = \sum_{i=1}^{nip} M_i(\xi, \eta) \hat{\mathbf{Z}}_i^{old} . \tag{27}$$

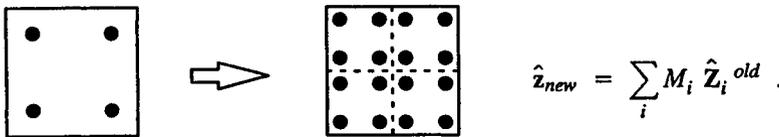


Figure 3: Transfer of data at integration points

Due to the hierarchical refinement this is equivalent to a variational formulation proposed in [2].

4. EXTENSION OF THE FINITE ELEMENT MODEL FOR THE ELASTOPLASTIC COSSERAT CONTINUUM

The classical continuum model may lead to an ill-posed problem due to the unstable material behaviour in the case of non-associated plasticity. Therefore, an extension of the adaptive finite element formulation for a Cosserat continuum is presented in this section which should lead to a regularized problem.

The regularization effect is obtained by the introduction of additional kinematic and static variables which are the micro-rotations ω_i , the micro-curvatures χ_{ij} , and corresponding stress couples μ_{ij} .

An extended set of equilibrium equations is given in terms of the now non-symmetric stress tensor σ_{ij} by Equation (1) and the stress couple tensor in addition to (1) by

$$\mu_{ji,j} + e_{ijk} \sigma_{jk} = 0 \quad \text{in } V, \quad (28)$$

where e_{ijk} denotes the permutation tensor. The kinematic relations are defined as

$$\varepsilon_{ij} = u_{j,i} - e_{ijk} \omega_k \quad \text{in } V, \quad (29)$$

$$\chi_{ij} = \omega_{j,i} \quad \text{in } V. \quad (30)$$

An equivalent variational form of these equations is given by the corresponding principle of virtual work with

$$\int_V (\delta \varepsilon_{ij} \sigma_{ij} + \delta \chi_{ij} \mu_{ij}) dV - \int_V \delta u_i \bar{f}_i dV - \int_{S_t} \delta u_i \bar{t}_i dS = 0. \quad (31)$$

The stress and strain tensors are splitted up into a symmetric part $\sigma_{(ij)}$, $\varepsilon_{(ij)}$ and a skew-symmetric part $\sigma_{[ij]}$, $\varepsilon_{[ij]}$.

In accordance with the classical formulation the strain components as well as the micro-curvature components are decomposed additively into their elastic and plastic parts as in Equation (5) and in

$$d\chi_{ij} = d\chi_{ij}^e + d\chi_{ij}^p. \quad (32)$$

The elastic constitutive relation between the symmetric parts of the stress and the elastic strain increments is identical with the classical formulation

$$d\varepsilon_{(ij)}^e = C_{ijkl} d\sigma_{(kl)}. \quad (33)$$

The components of their skew-symmetric parts are related by the so called Cosserat shear modulus G_c as

$$d\varepsilon_{[ij]}^e = \frac{1}{G_c} d\sigma_{[ij]}. \quad (34)$$

Because the stress couples and the curvature tensor are related by a type of bending modulus an internal length scale l_c is introduced into the formulation. The elastic relation may be given in terms of this internal length and the shear modulus G by

$$d\chi_{ij}^e = \frac{1}{G l_c^2} d\mu_{ij}. \quad (35)$$

The representation of the yield function is chosen as a generalization of the classical formulation with some additional terms arising from the introduction of the skew-symmetric stress and the stress couple tensors:

$$f(\sigma_{ij}, \mu_{ij}, \kappa) = f(\sigma_{(ij)}, \sigma_{[ij]}, \mu_{ij}, \kappa) = f(I_1, J_2, J_3, \sigma_{[ij]}, \mu_{ij}, \kappa) \leq 0, \quad (36)$$

$$\text{where } I_1 = \sigma_{(ij)} \delta_{ij}, \quad s_{ij} = \sigma_{(ij)} - \frac{1}{3} I_1 \delta_{ij},$$

$$J_2 = \frac{1}{2} s_{ij} s_{ij}, \quad J_3 = \frac{1}{3} s_{ij} s_{jk} s_{ki}.$$

A possible choice that is used in the numerical calculations is given by

$$f = \left\{ (m(\theta) \bar{\sigma})^2 + j_2 \sigma_{[ij]} \sigma_{[ij]} + j_3 \frac{1}{l_c^2} \mu_{ij} \mu_{ij} \right\}^{\frac{1}{2}} + \kappa (\eta I_1 - k) \leq 0, \quad (37)$$

$$\text{where } \bar{\sigma} = \sqrt{J_2}, \quad \sin(3\theta) = -\frac{\sqrt{27} J_3}{2 \sqrt{J_2^3}}.$$

With $j_2 = j_3 = 0$ the classical formulation of the yield function is included as a special case.

The plastic deformations are given by the flow rule where the plastic potential g is defined as in Equation (11) with

$$d\varepsilon_{ij}^p = \frac{\partial g}{\partial \sigma_{ij}} d\lambda,$$

$$d\chi_{ij}^p = \frac{\partial g}{\partial \mu_{ij}} d\lambda, \quad d\lambda \geq 0. \quad (38)$$

The structure of the fundamental equations of the classical continuum model is retained in the Cosserat theory. Therefore, the adaptive finite element formulation is just straightforward. Standard shape functions are used for the interpolation of the displacements and micro-rotations as in equation (14). Restricted to plane strain conditions the generalized displacement vector takes the form

$$\mathbf{u} = \{ u_1, u_2, \omega_3 \}. \quad (39)$$

The nodal displacement vector \mathbf{v} includes nodal displacements and rotations. The generalized strain field $\tilde{\boldsymbol{\varepsilon}}$ is redefined as

$$\tilde{\boldsymbol{\varepsilon}} = \{ \boldsymbol{\varepsilon}, \boldsymbol{\chi} \} = \{ \varepsilon_{ij}, \chi_{ij} \} \quad (40)$$

which may be computed by a modified discrete strain matrix \mathbf{B} with

$$\tilde{\boldsymbol{\varepsilon}} = \mathbf{B} \mathbf{v}. \quad (41)$$

When the generalized stress field $\tilde{\boldsymbol{\sigma}}$ is defined corresponding to the strain field $\tilde{\boldsymbol{\varepsilon}}$ with

$$\tilde{\boldsymbol{\sigma}} = \{ \boldsymbol{\sigma}, \boldsymbol{\mu} \} = \{ \sigma_{ij}, \mu_{ij} \} \quad (42)$$

the discrete equilibrium equations are given by

$$\int_V \mathbf{B}^T \bar{\boldsymbol{\sigma}} dV = \int_V \mathbf{N}^T \bar{\mathbf{f}} dV + \int_{S_i} \mathbf{N}^T \bar{\mathbf{t}} dS . \quad (43)$$

According to equation (19) the global error measure of the finite element solution has thus to be redefined in terms of increments of the generalized stress and strain field by

$$\begin{aligned} \|e\| &= \int_V (\Delta \bar{\boldsymbol{\sigma}} - \Delta \bar{\boldsymbol{\sigma}}^h) (\Delta \bar{\boldsymbol{\varepsilon}} - \Delta \bar{\boldsymbol{\varepsilon}}^h) dV \\ &= \int_V \left((\Delta \boldsymbol{\sigma} - \Delta \boldsymbol{\sigma}^h) (\Delta \boldsymbol{\varepsilon} - \Delta \boldsymbol{\varepsilon}^h) + (\Delta \boldsymbol{\mu} - \Delta \boldsymbol{\mu}^h) (\Delta \boldsymbol{\chi} - \Delta \boldsymbol{\chi}^h) \right) dV . \end{aligned} \quad (44)$$

The generalization of the error indicator from the classical continuum, Equ. (20), to the Cosserat model leads to:

$$\begin{aligned} \|e\| &= \|\Delta \boldsymbol{\varepsilon}\|_{L_1} (C_1 h \|\mathbf{r}\|_{L_\infty} + C_2 \frac{h}{h_k} \|\mathbf{J}\|_{L_\infty}) \\ &+ \|\Delta \boldsymbol{\chi}\|_{L_1} (C_1 h \|\mathbf{m}\|_{L_\infty} + C_2 \frac{h}{h_k} \|\mathbf{M}\|_{L_\infty}) , \end{aligned} \quad (45)$$

$$\text{where } r_i = \sigma_{ji,j} + \bar{f}_i \quad \text{in } V_e , \quad (46)$$

$$J_i = \frac{1}{2} (\sigma_{ji}^+ - \sigma_{ji}^-) n_j \quad \text{on } S_e \quad (47a)$$

$$\text{or } J_i = (\sigma_{ji} n_j - \bar{t}_i) \quad \text{on } S_s , \quad (47b)$$

$$m_i = \mu_{jij} + e_{ijk} \sigma_{jk} \quad \text{in } V_e , \quad (48)$$

$$\text{and } M_i = \frac{1}{2} (\mu_{ji}^+ - \mu_{ji}^-) n_j \quad \text{on } S_e \quad (49a)$$

$$\text{or } M_i = \mu_{ji} n_j \quad \text{on } S_s . \quad (49b)$$

This error indicator consists of two additive parts in which the first is identical with that for the classical model. In the second part maximum norms of the residual moments of the equilibrium equations within the elements and of the couple stress jumps on the element boundaries are weighted by the L_1 -norm of the total micro-curvature increments. Based on this error indicator the mesh refinement including the transfer of the state variables can now be performed in the same manner as described in the previous section.

5. NUMERICAL EXAMPLES

The approach described in this paper was applied to various examples in which different aspects were investigated. As a *benchmark problem* for several research groups a metal sheet with a hole under uniaxial extension was analyzed. Several adaptive methods were compared for elastic and elastic-ideally plastic material with a von Mises yield condition and associated flow rule.

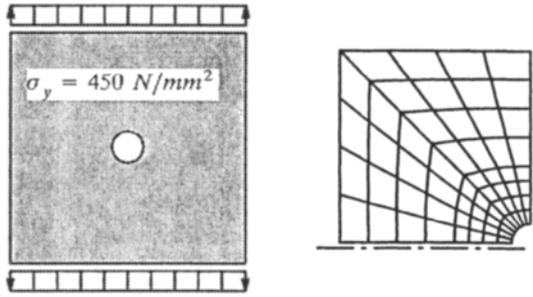


Figure 4: Metal sheet with a hole, system and finite element idealization

For the results of our research group the reader is referred to the literature [3]. They confirmed that the error indicator as well as the transfer process described are very well suited for the adaptive nonlinear finite element method.

In a further example a *strip footing* on cohesive soil with a small angle of internal friction and an associated flow rule was considered in [3,4]. The yield condition was formulated by a modified Coulomb criterion with a smooth yield surface. In this example results of different discretizations were compared with those obtained by the adaptive process.

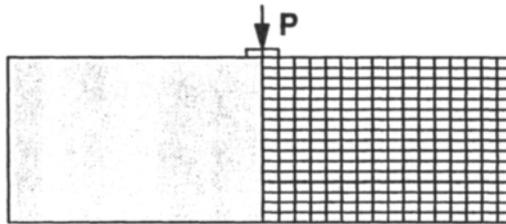


Figure 5: Strip footing on a cohesive soil, system and initial mesh a)

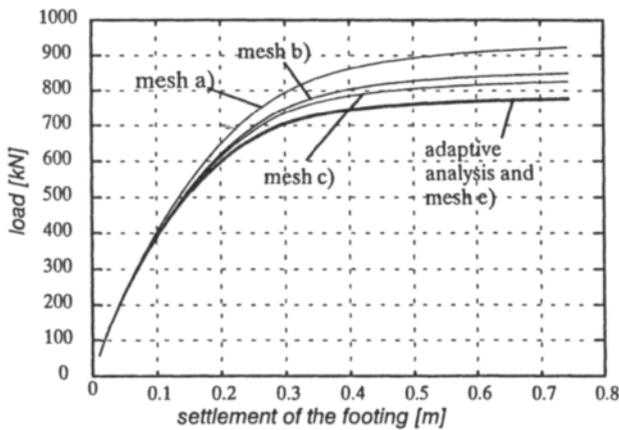


Figure 6: Load-displacement curves for different discretizations

The corresponding load–displacement curves for different meshes are given in Figure 6. The solution with the finest mesh is almost identical with the adaptive solution. This is not surprising as this discretization represents the final adaptive mesh whereas the solutions with the three other meshes show quite considerable deviations, however. This demonstrates the improvement of finite element computations by adaptive mesh refinements.

As an example for a non–associated flow rule a *strip footing near a slope* under vertical loading was investigated in [3,4]. It was observed that the error indicator is also capable to detect zones with localized shear deformations.

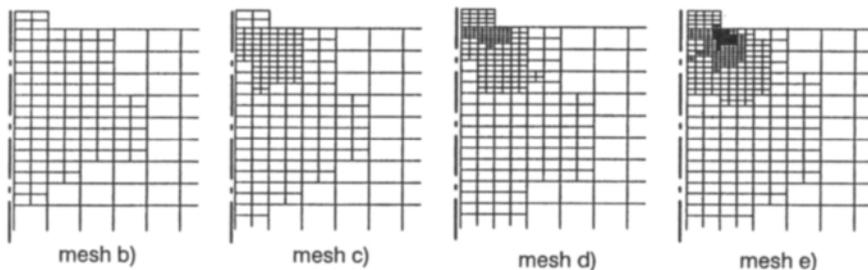


Figure 7: Sequence of refined meshes for $\eta = 0.025$

A frictional material with nearly no cohesion and a friction angle of $\phi = 25^\circ$ was assumed with a yield condition according to Equation (9) with a non–associated flow rule (10), (11).

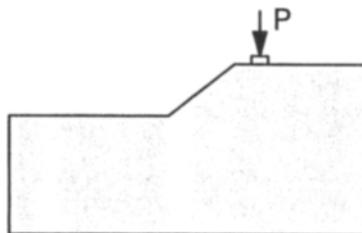


Figure 8: Strip footing near a slope under vertical loading

The load–settlement curve of the footing obtained in the adaptive finite element calculation is given in Figure 9. The comparison with results of non–adaptive finite element calculations illustrates that a localization will not be noticed at all without adaptive mesh refinements. This may lead to a complete misjudgement of the carrying behaviour of a structure. But also numerical difficulties were encountered for further adaptive refinement leading to a non–converged solution. This is due to an unstable material behaviour caused by the non–associated flow rule accompanied by a sharp decrease of the load carrying capacity.

In this case the governing differential equations of the classical continuum lose its ellipticity leading to an ill–posed problem. Therefore, in numerical simulations an extreme mesh sensitivity may be observed which is the crucial point in adaptive finite element calculations.

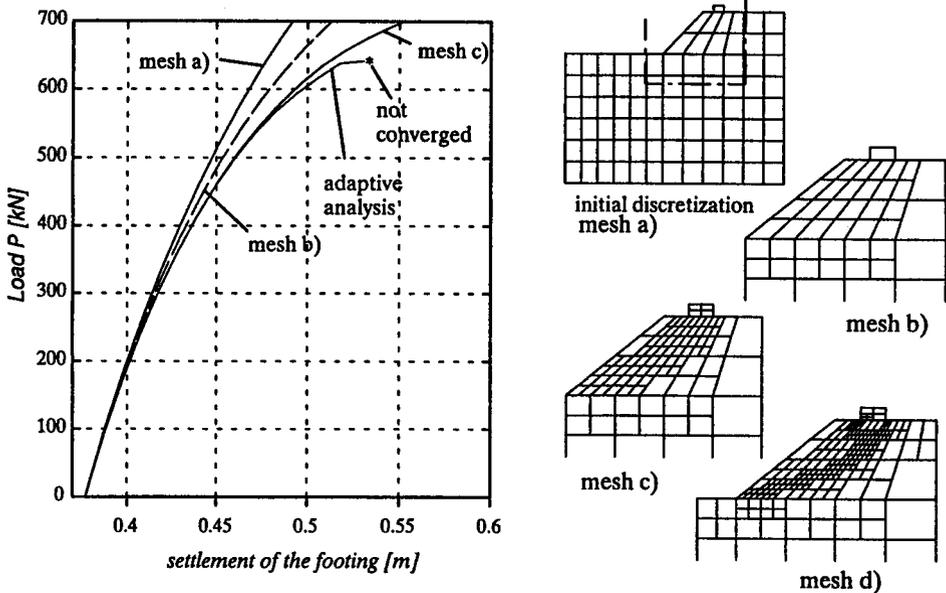


Figure 9: Load–settlement–curves for different discretizations

Thus, the formulation of the localization regime has to be extended to generalized continuum models like the polar continuum theory with additional rotational degrees of freedom employed in this paper. For illustration of this approach two examples with non-associated flow rule are investigated and described in more detail: the numerical simulation of a *plane strain compression test* and the *strip footing near a slope* similar to the example mentioned above but differing in the material description.

5.1. Localization in a plane strain compression test

The different behaviour of the classical finite element model and the Cosserat model as well as the influence of different internal length scales shall be illustrated in this example. A plane strain compression test for a generalized Drucker–Prager material with non-associated flow rule is considered. The geometric and material characteristics are shown in Figure 10. To evaluate the influence of the internal length scale in the Cosserat model different analyses are performed with the three different parameters $l_c = 0.1$ mm, 1.0 mm, and 10.0 mm.

In addition, the system is analysed using the classical continuum model. Different uniform discretizations are considered to illustrate the mesh dependence (see Figure 11).

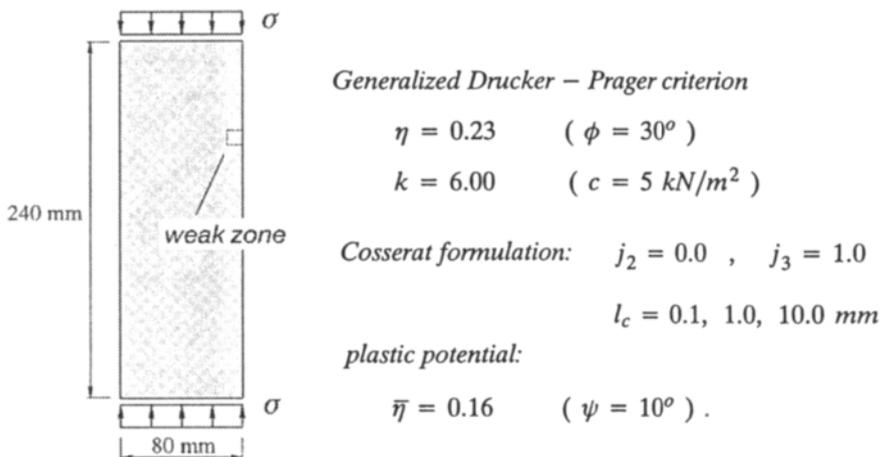


Figure 10: Geometry and material characteristics of the plane strain compression test

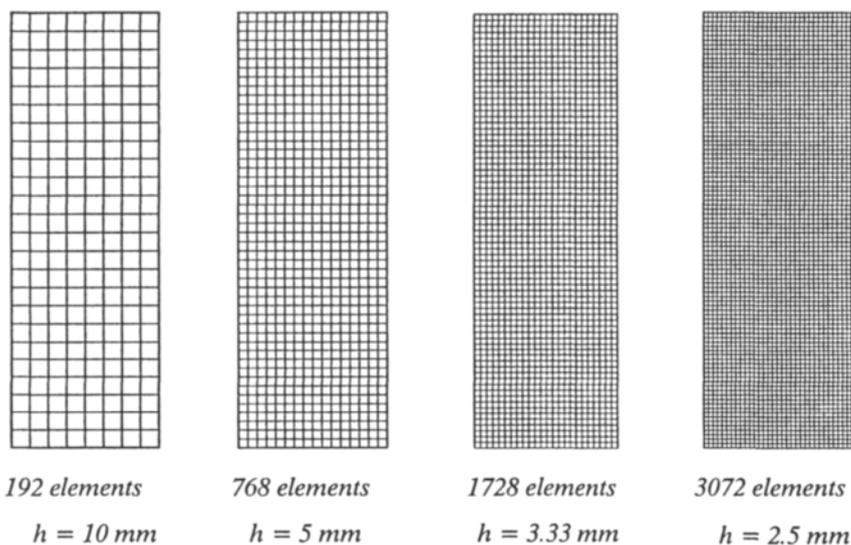


Figure 11: Finite element discretizations for the plane strain compression test

The load–displacement curves of the finite element calculations with the classical continuum model are shown in Figure 12. Localization initiated by a small weak zone with lower strength is achieved for all discretizations as illustrated in Figure 12. However, the results show the strong mesh dependence leading to different load–displacement characteristics. The width of the observed shearbands depends on the mesh size and no convergence may be achieved with any further mesh refinement.

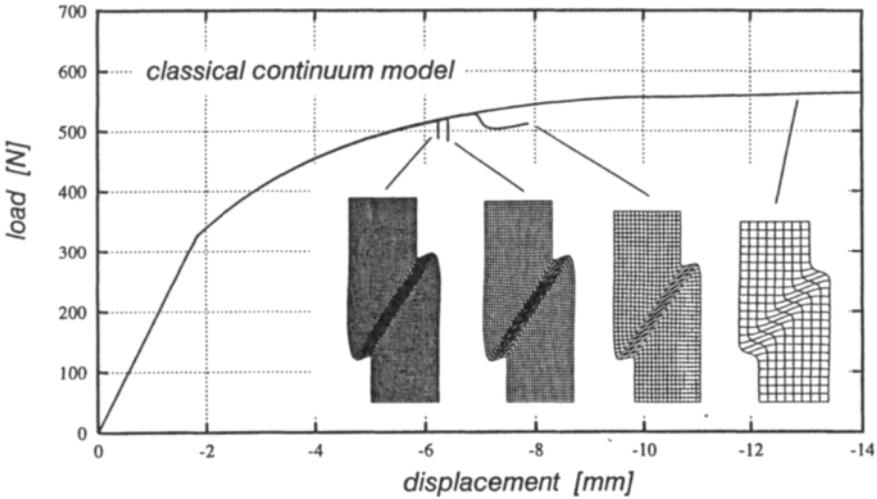


Figure 12: Load–displacement characteristics and deformations of the classical continuum model for different discretizations

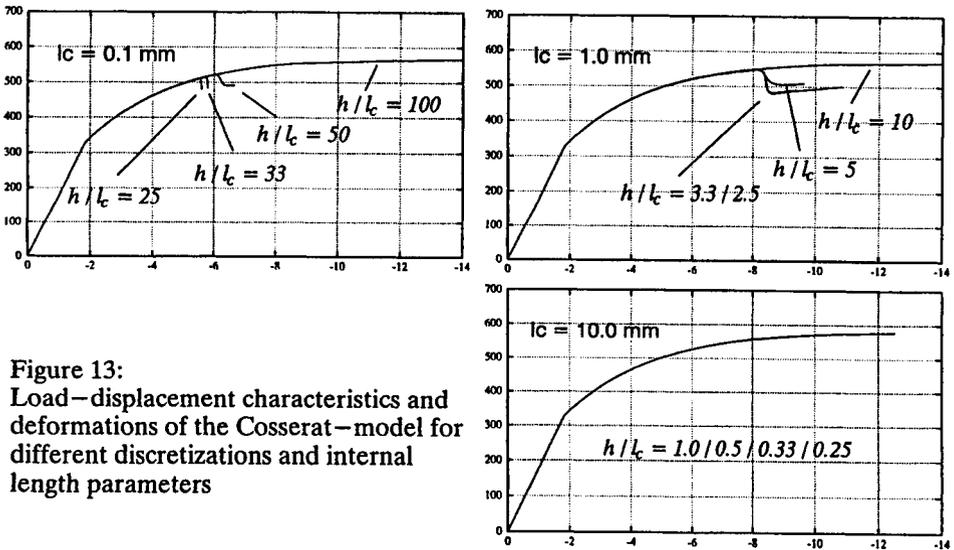


Figure 13: Load–displacement characteristics and deformations of the Cosserat–model for different discretizations and internal length parameters

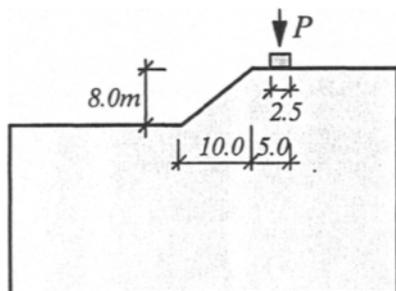
Results obtained by the Cosserat–continuum model for different values of the internal length parameter are presented in Figure 13. They illustrate the strong influence of the size of the internal length parameter. The load–displacement curves of the different finite element discretizations for the small length parameter $l_c = 0.1 \text{ mm}$ show no convergence of the finite element solution. For the mean value of $l_c = 1.0 \text{ mm}$ however, a converged finite element solution with a continuously decreasing load–displacement curve is achieved when the ratio of the mesh size to the internal length parameter has the order of one in the

localization zone. On the other hand, when l_c approaches the system size as illustrated by the results for $l_c = 10.0\text{ mm}$, a much stiffer behaviour of the system with no decrease in the load displacement curve is observed.

Therefore, it is very essential to specify the internal length parameter in an appropriate order of magnitude. In addition, the finite element size in zones of strain localization should also be of the same order.

5.2. Strip footing near a slope

As a first test of the Cosserat model in a real-life problem an adaptive finite element analysis of a strip footing near a slope under vertical loading is investigated.



$$E = 30000\text{ kN/m}^2, \quad \nu = 0.20$$

$$\phi = 25^\circ, \quad \psi = 5^\circ, \quad c = 5.0\text{ kN/m}^2$$

$$\eta = 0.144, \quad \bar{\eta} = 0.11, \quad k = 4.59$$

$$\alpha = 1.0$$

$$m(\theta) = 1.0, \quad j_2 = 0.0, \quad j_3 = 1.0$$

Figure 14: System and material characteristics of the strip footing near a slope

An elastic ideal plastic behaviour of the soil was assumed with a generalized Drucker–Prager criterion and a non-associated flow rule. The system and material characteristics are shown in Figure 14. The internal length parameter was specified to $l_c = 4.0\text{ cm}$.

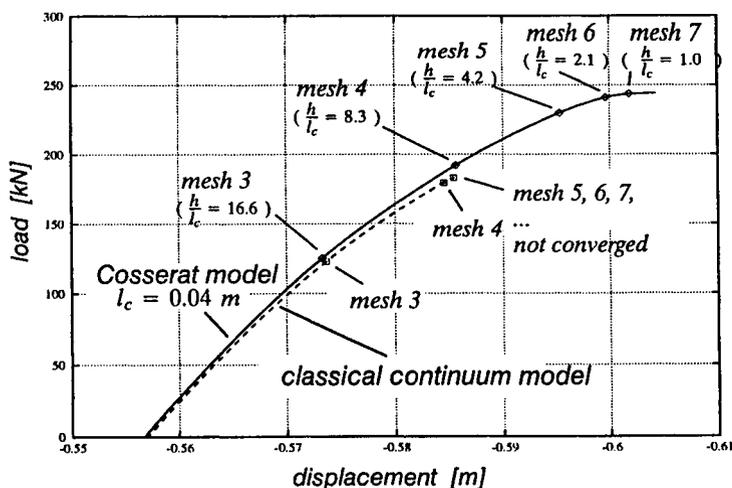


Figure 15: Load–displacement curves, Cosserat model and classical model

The load–displacement curve of the load–controlled calculation is represented in Figure 15. A convergent adaptive finite element analysis with six hierarchical refinement steps can be performed until a horizontal tangent is reached when the Cosserat model is applied. A sequence of the adaptively refined meshes is shown in Figure 16 which illustrates the evolution of shearbands within the system. For comparison the load displacement curve of an adaptive calculation with the classical continuum model is represented in Figure 15 by the dashed line. In contrast to the Cosserat model it was not possible to obtain a converged solution in this case. The analysis could be performed up to a distinct load level where an uncontrollable sequence of mesh refinements was initiated.

These first results illustrate the superior behaviour of the Cosserat–model for non–associated plasticity. Of course, further experience with problems of similar complexity is necessary in which especially the influence of the size of the internal length scale is investigated.

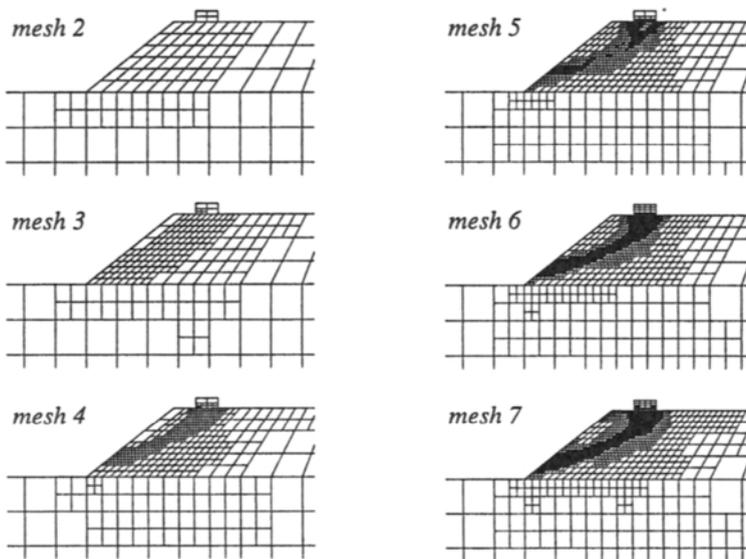


Figure 16: Sequence of adaptively refined meshes of the Cosserat model

References

- [1] Johnson C., Hansbo P., "Adaptive finite element methods in computational mechanics", *Comp. Meth. Appl. Mech. Eng.*, 101, 143–181, 1992.
- [2] Ortiz, M., Quigley, J. J., "Adaptive mesh refinement in strain localization problems", *Comp. Meth. Appl. Mech. Eng.*, 90, 781–804, 1991.
- [3] H. Cramer, M. Rudolph, G. Steinl, W. Wunderlich, "A hierarchical adaptive finite element strategy for elastic–plastic problems". In: "Advances in Finite Element Technology", B.H.V. Topping (ed.), Edinburgh, 1996, 151–159.
- [4] W. Wunderlich, H. Cramer, M. Rudolph, "Adaptive finite element approach to nonlinear geomechanical problems". In: "3rd Asian–Pacific Conference on Computational Mechanics", Vol.3, Seoul, 1996, 2151–2156.

An alpha-adaptive approach for stabilized finite element solution of advective-diffusive problems with sharp gradients

E. Oñate¹, J. García¹ and S. Idelsohn²

¹International Center for Numerical Methods in Engineering,
Edificio C1, Campus Norte UPC, 08034 Barcelona, Spain

²Universidad Nacional del Litoral, Santa Fe, Argentina
Visiting Scientist at 1

1. INTRODUCTION

The standard Galerkin finite element method is known to fail for solution of advective-diffusive problems for moderate and high values of the advective terms [1,2]. Over the years a number of techniques have been proposed to obtain the so called stable (or oscillation-free) solutions. Original remedies were based on the heuristic addition of the right amount of balancing diffusion to the original problem [1-4].

A more rigorous approach is based in adding to the Galerkin finite element formulation the sum over all elements of the integrals over the element interiors of the residual of the original differential equation times an "ad-hoc" perturbation of the weighting functions and the so called stabilization parameter. By choosing adequately the perturbation function the standard SUPG [5], GLS [6], Taylor-Galerkin [7], Characteristic approximation [1,8] and Subgrid Scale [9] methods can be recovered as shown in [10]. As for the stabilization parameter, this can be interpreted either as a "characteristic length" of the discrete problem, as a proportion of a typical element dimension, or as the "intrinsic time" taken for a particle to travel half the characteristic length at the advective speed.

The precise computation of any of the equivalent forms of the crucial stabilization parameter can only be attempted for simple one dimensional (1D) problems such the sourceless 1D advection-diffusion case [1,2]. Attempts to generalize the computation of this parameter were due to Idelsohn [1] using a pseudo-variational principle. Hughes [9] and later Brezzi and co-workers [12-14] have proposed a numerical expression for the stabilization parameter involving an approximation of the element Green's function using bubble shape functions. None of these procedures has however succeeded so far to present evidence of its usefulness for practical multidimensional problems.

In [15,16] Oñate proposed a different approach for computing the stabilization parameter. The method is based in introducing "a priori" the stabilizing terms within the differential equations governing the balance of fluxes over a finite domain. This

kind of *finite increment calculus* (FIC) procedure allows to obtain any stable discretized scheme using finite difference, finite element or finite volume methods in a straight forward manner. For instance it can be shown that the Galerkin finite element form of the new stabilized governing equations is identical to that obtained with the well known SUPG and Characteristic-Galerkin methods, among others [15,16].

The interest of the FIC approach is that it leads naturally to an iterative scheme for evaluating the stabilization parameter in terms of the residuals of the numerical solution. The efficiency of the new approach for computing the streamline stabilization parameter in a variety of 1D and 2D advective-diffusive problems was reported in [15-17].

In this paper the FIC method is used as the basis for a new "alpha-adaptive" procedure (where alpha denotes the stabilization parameters) for obtaining stable solution in advective-diffusive problems where arbitrary sharp transverse gradients are present. The new stabilization technique can be viewed as an alternative class of adaptive methods where the numerical solution is enhanced by searching "adaptively" the optimal value of the streamline and transverse (crosswing) stabilization parameters while keeping the mesh and the finite element approximation unchanged. Indeed the basic alpha-adaptive process can be enhanced by combining it with standard h , p or hp adaptive schemes.

In the first part of the paper the basis of the FIC stabilized method for advective-diffusive problems are explained. Next the algorithm for computing the streamline and transverse stabilization parameters via the new "alpha-adaptive" procedure is described. Finally, the efficiency and accuracy of the new approach are shown in two examples of application.

2. STABILIZED GOVERNING EQUATIONS FOR ADVECTIVE-DIFFUSIVE TRANSPORT

2.1 One dimensional advective-diffusive problem

Let us consider for simplicity the standard advective-diffusive transport problem to be solved in a one-dimensional domain of length l (Figure 1a). Figure 1b shows a typical segment AB of length $\overline{AB} = h$ where balance (equilibrium) of fluxes must be satisfied. The values of the diffusive flow rate q and the advective transport rate $u\phi$ at a point A with coordinate $x_A = x_B - h$ can be approximated in terms of values at point B using third order Taylor's expansion. A linear variation of the source term Q over the segment is also assumed. Under these assumptions and using Fourier's law the governing balance equation can be obtained as [15,16].

$$\boxed{r - \frac{h}{2} \frac{dr}{dx} = 0} \quad , \quad 0 < x < l \quad (1a)$$

with

$$r = -\nu \frac{d(u\phi)}{dx} + \frac{d}{dx} \left(k \frac{d\phi}{dx} \right) + Q \quad (1b)$$

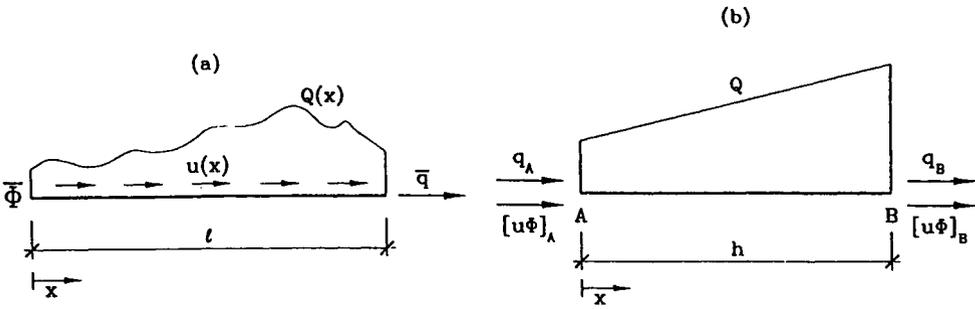


Figure 1. (a) One-dimensional advection-diffusion problem. (b) Finite balance domain AB

In eq.(1b) ν and k are the advective and diffusive material parameters, respectively.

Note that for $h \rightarrow 0$ (i.e. when the length of the balancing domain is infinitesimal) then the standard form of the governing equation for 1D advective-diffusive transport ($r = 0$) is recovered.

The essential (Dirichlet) boundary condition is the standard one given by

$$\phi - \bar{\phi} = 0 \quad \text{on } x = 0 \tag{2}$$

where $\bar{\phi}$ is the prescribed unknown field at the Dirichlet boundary.

For consistency the stabilized form of the Neumann boundary condition is needed. This can be obtained by invoking again the balance law in a segment AB next to a boundary point. For convenience the length of this segment is taken as half of the characteristic length h for the interior domain points [15,16].

Assuming now second order expansion for the advective and diffusive fluxes and taking the source Q to be constant over AB , the balance equation is obtained as [15,16]

$$\boxed{-\nu u \phi + k \frac{d\phi}{dx} + \bar{q} - \frac{h}{2} r = 0} \quad \text{on } x = l \tag{3}$$

where r is given by eq.(16). Obviously for $h \rightarrow 0$ the standard form of the Neumann boundary condition is recovered.

Equation (1a) can now be solved together with eqs.(2) and (3). These equations are the starting point to derive stabilized numerical schemes using any discretization procedure.

The extension of this stabilization concept to the transient case can be found in [15,16].

2.2 Two dimensional advective-diffusive problem

The concepts of previous section will be extended now to the solution of advection-diffusion problems in a two-dimensional domain Ω with boundary Γ . Let us consider a finite rectangular domain of dimensions h_x and h_y in directions x and y , respectively. Both the advective and diffusive fluxes are assumed to vary linearly along the four sides

of the balance domain (Figure 2). The flux balance equation will be obtained using the following Taylor expansions: diffusive term, third order expansion; advective term, third order expansion; source term, second order expansion.

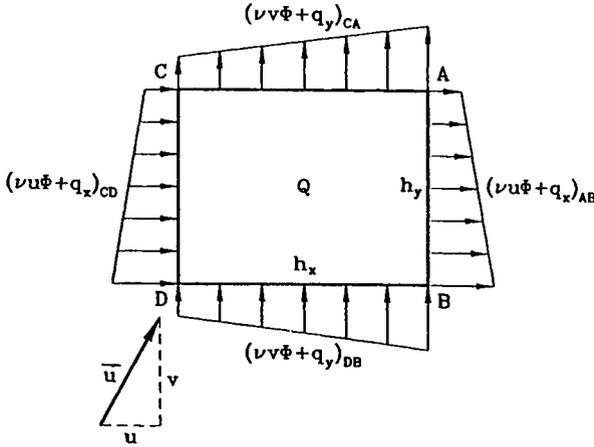


Figure 2. Balance domain for 2D advection-diffusion problem. Advective and diffusive fluxes are assumed to vary linearly along the sides

The balance of fluxes across the four sides of the rectangular domain of Figure 2 gives after some algebra [16]

$$\boxed{r - \frac{1}{2} \mathbf{h}^T \nabla r = 0} \quad \text{in } \Omega \tag{4}$$

where

$$r = -\nu \nabla^T \mathbf{f} + \nabla^T (\mathbf{D} \nabla \phi) + Q \tag{5}$$

and

$$\mathbf{h} = [h_x, h_y]^T \tag{6}$$

In eq.(5)

$$\mathbf{f} = [u\phi, v\phi]^T, \quad \nabla = \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right]^T, \quad \mathbf{D} = \begin{bmatrix} k_x & 0 \\ 0 & k_y \end{bmatrix} \tag{7}$$

The boundary conditions are written as

$$\phi - \bar{\phi} = 0 \quad \text{on } \Gamma_\phi \tag{8}$$

where Γ_ϕ is the Dirichlet boundary, where the variable is prescribed, and

$$\boxed{-\nu \mathbf{n}^T \mathbf{u} \phi + \mathbf{n}^T \mathbf{D} \nabla \phi + \bar{q}_n - \frac{1}{2} \mathbf{h}^T \mathbf{n} r = 0} \quad \text{on } \Gamma_q \tag{9}$$

where \bar{q}_n is the prescribed total flux across the Neumann boundary Γ_q with $\Gamma = \Gamma_\phi \cup \Gamma_q$ and $\mathbf{n} = [n_x, n_y]^T$ is the normal vector. Eq.(9) has been obtained by balance of fluxes in a finite boundary domain [15,16].

The standard differential equations are simply obtained by neglecting the stabilizing terms in eqs.(4) and (9) (i.e making $\mathbf{h} = 0$). The extension to three-dimensional problems is straightforward and identical stabilized expressions are obtained.

REMARK 1

It is interesting to note that the finite element Galerkin form of the new stabilized governing equations leads to a set of discretized equations identical to those obtained with the standard SUPG formulation [15]. Alternatively, the stabilized transient form leads to the well known Characteristic-Galerkin procedure [15]. This indicates that the new governing equations can be considered as the *intrinsic stabilized equations* of the problem.

2.3 The concept of intrinsic time

It is usual to accept that \mathbf{h} and \mathbf{u} are parallel, so that $\mathbf{h} = \frac{h}{|\mathbf{u}|}\mathbf{u}$. The distance $h = (h_x^2 + h_y^2)^{1/2}$ is then called the characteristic length of the 2D advective-diffusive problem. The *intrinsic time parameter* is now defined as [6]

$$\tau = \frac{h}{2|\mathbf{u}|} \quad (10)$$

Note that this coincides with the time taken for a particle to travel the distance $h/2$ at the speed $|\mathbf{u}|$.

The assumption of the characteristic length vector \mathbf{h} being parallel to the velocity vector \mathbf{u} is a *simplification* which eliminates any transverse diffusion effect. This assumption is the basis of the standard SUPG approach. However it is well known that when arbitrary sharp transverse layers are present, additional transverse (or crosswind) diffusion is required to capture these discontinuities. Different "ad hoc" expressions for the transverse diffusion terms, typically of non linear nature, have been proposed [18–20]. Indeed the introduction of this additional stabilizing effect can be simply reproduced in the FIC approach here proposed by abandoning the assumption of \mathbf{h} being parallel to \mathbf{u} and keeping the two characteristic lengths h_x and h_y as "free" stabilization parameters. The computation of these two parameters is described in the following section.

3. COMPUTATION OF THE STABILIZATION PARAMETERS

Let us consider the finite element solution of an advective-diffusive problem. The standard interpolation within an element e with n nodes can be written as

$$\phi \simeq \hat{\phi} = \sum_{i=1}^n N_i \phi_i \quad (11)$$

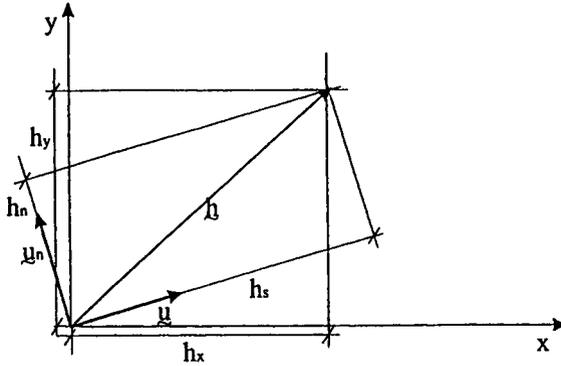


Figure 3. Characteristic length in global and velocity axes.

where N_i are the element shape functions and ϕ_i are nodal values of the approximate function $\hat{\phi}$. Substitution of eq.(11) into eq.(4) gives

$$\hat{r} - \frac{1}{2} \mathbf{h}^T \nabla \hat{r} = r_{\Omega} \quad \text{in } \Omega \quad (12)$$

where $\hat{r} = r(\hat{\phi})$.

Let us now define the average residual of a particular numerical solution over an element as

$$r^{(e)} = \frac{1}{\Omega^{(e)}} \int_{\Omega^{(e)}} r_{\Omega} d\Omega \quad (13)$$

Substituting eq.(13) into (12) gives

$$r^{(e)} = \hat{r}^{(e)} - \left(\frac{1}{2} \mathbf{h}^T \nabla \hat{r} \right)^{(e)} \quad (14)$$

where

$$a^{(e)} \equiv \frac{1}{\Omega^{(e)}} \int_{\Omega^{(e)}} a d\Omega \quad (15)$$

For simplicity the characteristic length vector will be assumed to be constant over each element, i.e. $\mathbf{h} = \mathbf{h}^{(e)}$. With this assumption eq.(14) can be simplified to

$$r^{(e)} = \hat{r}^{(e)} - \frac{1}{2} \left[\mathbf{h}^{(e)} \right]^T (\nabla \hat{r})^{(e)} \quad (16)$$

Let us express the characteristic length vector in terms of the components along the velocity vector \mathbf{u} and the normal velocity direction \mathbf{u}_n (Figure 3) as

$$\mathbf{h} = \frac{1}{|\mathbf{u}|} [h_s \mathbf{u} + h_n \mathbf{u}_n] \quad (17)$$

where $\mathbf{u}_n = [-v, u]^T$ and h_s and h_n are streamline and transverse (crosswind) characteristic lengths, respectively.

Substituting (17) into (16) gives

$$r^{(e)} = \hat{r}^{(e)} - \frac{1}{2|\mathbf{u}|} \left[h_s \mathbf{u}^T + h_n \mathbf{u}_n^T \right]^{(e)} (\nabla \hat{r})^{(e)} \quad (18)$$

The characteristic lengths h_s and h_n can be expressed now as a proportion of a typical element dimension $l^{(e)}$

$$h_s^{(e)} = \alpha_s^{(e)} l^{(e)} \quad , \quad h_n = \alpha_n^{(e)} l^{(e)} \quad (19)$$

where $\alpha_s^{(e)}$ and $\alpha_n^{(e)}$ are the streamline and transverse stabilization parameters, respectively. In the examples shown next $l^{(e)}$ has been taken equal to the length of the longest side of each triangular element.

Clearly for $\alpha_n^{(e)} = 0$ just the streamline diffusive effect, typical of the SUPG approach, is reproduced.

Let us consider now that an enhanced numerical solution has been found for a given finite element mesh. This can be simply achieved by projecting into the original mesh an improved solution obtained via global/local smoothing or superconvergent recovery of derivatives [21,22]. If $r_1^{(e)}$ and $r_2^{(e)}$ respectively denote the element residuals of the original and the enhanced numerical solutions for a given mesh it is obvious that

$$r_1^{(e)} - r_2^{(e)} \geq 0 \quad (20)$$

Eq. (11) assumes that r_1 is positive. Clearly for the negative case the inequality should be appropriately reversed.

Combining eqs.(18),(19) and (20) gives

$$[\alpha_s \mathbf{u}^T + \alpha_n \mathbf{u}_n^T]^{(e)} (\nabla \hat{r}_2^{(e)} - \nabla \hat{r}_1^{(e)}) \geq \frac{2|\mathbf{u}|}{l^{(e)}} (\hat{r}_2^{(e)} - \hat{r}_1^{(e)}) \quad (21)$$

3.1 Computation of α for elements at the boundaries

The stabilized balance equation at a boundary can be written after discretization as

$$-\nu \mathbf{n}^T \mathbf{u} \hat{\phi} + \mathbf{n}^T \mathbf{D} \nabla \hat{\phi} + q - \frac{1}{2} \mathbf{h}^T \mathbf{n} \hat{r} = r_\Gamma \quad (22)$$

where q represents the prescribed normal flux at a Neumann boundary, or alternatively the unknown normal flux at the Dirichlet boundary where ϕ is prescribed.

Following the arguments used previously the equation defining the stabilization parameters at a boundary element can be obtained as

$$[\alpha_s \mathbf{u}^T + \alpha_n \mathbf{u}_n^T]^{(e)} \mathbf{n} (\hat{r}_2^{(e)} - \hat{r}_1^{(e)}) \geq -\nu \mathbf{n}^T \mathbf{u} (\hat{\phi}_2^{(e)} - \hat{\phi}_1^{(e)}) + \mathbf{n}^T \mathbf{D} (\nabla \hat{\phi}_2^{(e)} - \nabla \hat{\phi}_1^{(e)}) + q_2^{(e)} - q_1^{(e)} \quad (23)$$

where $(\cdot)^{(e)}$ denotes average values over a boundary domain and indexes 1 and 2 refer to the original and enhanced solutions, respectively. The enhanced nodal values $\hat{\phi}_2^{(e)}$ can be obtained by superconvergent nodal recovery of primary variables [22].

Note that the equality sign in eqs.(21) and (23) provides the value of the stabilization parameters ensuring no growth of the numerical error. In reference [15] it is proved that this yields the standard critical value of α in the simplest sourceless one dimensional problem solved with linear elements.

Eqs.(21) and (23) are the basis for the alpha-adaptive scheme to be described in next section.

4. ALPHA-ADAPTIVE STABILIZATION SCHEME

The following scheme can be devised to obtain an stable numerical solution in an adaptive manner.

- (1) Solve the stabilized problem defined by eqs.(4), (8) and (9) using the FEM with an initial guess of the stabilization parameters, i.e.

$$\alpha_s^{(e)} = \alpha_s^{(e)} \quad , \quad \alpha_n^{(e)} = \alpha_n^{(e)} \quad (24)$$

- (2) Recover an enhanced derivatives field. Evaluate $\hat{r}^{(1)}$, $\hat{r}^{(2)}$, $\nabla \hat{r}_1^{(e)}$ and $\nabla \hat{r}_2^{(e)}$.
- (3) Compute an enhanced valued of the streamline stabilization parameter $\alpha_s^{(e)}$ by

$${}^1\alpha_s^{(e)} = \frac{2|\mathbf{u}|}{l^{(e)}\mathbf{u}^T(\nabla \hat{r}_2^{(e)} - \nabla \hat{r}_1^{(e)})} [\hat{r}_2^{(e)} - \hat{r}_1^{(e)} - \alpha_n^{(e)}\mathbf{u}_n^T(\nabla \hat{r}_2^{(e)} - \nabla \hat{r}_1^{(e)})] \quad (25)$$

If the element lays in one of the boundaries the expression for $\alpha_s^{(e)}$ as deduced from eq.(23) should be used.

- (4) Repeat steps (1)–(3) until convergence is found for the value of $\alpha_s^{(e)}$ while keeping $\alpha_n^{(e)}$ constant.
- (5) Repeat steps (1)–(4) for computing $\alpha_n^{(e)}$ while keeping $\alpha_s^{(e)}$ constant and equal to the previously converged value. In the first iteration $\alpha_n = \alpha_n^{(e)} + \varepsilon$ where ε is a small value should be used. The updated value of $\alpha_n^{(e)}$ is computed as

$${}^i\alpha_n^{(e)} = \frac{2|\mathbf{u}|}{l^{(e)}\mathbf{u}_n^T(\nabla \hat{r}_2^{(e)} - \nabla \hat{r}_1^{(e)})} [\hat{r}_2^{(e)} - \hat{r}_1^{(e)} - \alpha_s^{(e)}\mathbf{u}^T(\nabla \hat{r}_2^{(e)} - \nabla \hat{r}_1^{(e)})] \quad (26)$$

Again for a boundary element the expression for $\alpha_n^{(e)}$ deduced from eq.(23) should be used.

- (6) Once $\alpha_n^{(e)}$ has been found steps (1)–(5) can be repeated to obtain yet more improved values of both $\alpha_s^{(e)}$ and $\alpha_n^{(e)}$.

Note that for $\alpha_n^{(e)} = 0$ above adaptive scheme provides the value of the critical streamline stabilization parameter $\alpha_s^{(e)}$ corresponding to the well known SUPG procedure. It can be shown that for the simplest 1D sourceless advective-diffusive case solved with linear elements the well known critical value $\alpha_s^{(e)} = 1 - \frac{1}{\gamma^{(e)}}$, where $\gamma^{(e)} = \frac{u l^{(e)}}{2k}$ is the element Peclet number is obtained. Indeed accounting for the cross-

wind stabilization parameter α_n has proved to be essential for obtaining stable solution in presence of arbitrary transverse sharp layers.

In the examples shown next the enhanced derivative field has been obtained by the simplest nodal averaging procedure. It has also been found useful to smooth the distribution of the $\alpha_s^{(e)}$ and $\alpha_n^{(e)}$ values and this has been done again using nodal averaging. Note also that the number of iterations in the above adaptive process is substantially reduced if the initial guess for $\alpha_s^{(e)}$ and $\alpha_n^{(e)}$ are not far from the final converged values. This can be ensured by using as initial value for $\alpha_s^{(e)}$ the standard expression derived from the straight forward extension of the simple 1D case, whereas the initial guess ${}^o\alpha_n^{(e)} = 0$ provides a good approximation in zones far from sharp layers non orthogonal to the velocity vector.

5. EXAMPLES

5.1 Example 1. Two dimensional advective-diffusive problem with no source, diagonal velocity and uniform Dirichlet boundary conditions

The first 2D example chosen is the solution of the standard advection-diffusion equation in a square domain of unit size with

$$k_x = k_y = 1 \quad , \quad \mathbf{u} = [1, 1]^T \quad , \quad \nu = 1 \times 10^{10} \quad , \quad Q = 0$$

The following Dirichlet boundary conditions are assumed

$\phi = 0$ along the boundary lines $x = 0$ and $y = 0$

$\phi = 100$ along the boundary line $x = 1$

$q_n = 0$ along the boundary line $y = 1$

The expected solution in this case is a uniform distribution of $\phi = 0$ over the whole domain except in the vicinity of the boundary $y = 1$ where a boundary layer is formed.

The domain has been discretized with a uniform mesh of 800 three node triangles as shown in Figure 4. The initial values ${}^o\alpha_s^{(e)} = {}^o\alpha_n^{(e)} = 0$ have been taken in all elements.

Figure 5 shows the initial distributions of ϕ for $\alpha_s^{(e)} = \alpha_n^{(e)} = 0$ (standard Galerkin solution). Note the strong oscillations obtained as expected.

The final converged solution for ϕ after 7 iterations is displayed in Figure 6. Note that the boundary layer originated in the vicinity of the boundary at $y = 1$ is well reproduced with minimum oscillations. These oscillations grows considerably higher if the value of the transverse stabilization parameter $\alpha_n^{(e)}$ is kept equal to zero during the adaptive process, thus yielding the standard SUPG solution, as shown in Figure 7.

Figure 8 shows finally the smoothed distribution of the stabilization vector $\boldsymbol{\alpha} = \alpha_s \mathbf{u} + \alpha_n \mathbf{u}_n$. Note that in the central part of the domain the $\boldsymbol{\alpha}$ vectors are aligned with the velocity direction (i.e. $\alpha_n = 0$), whereas in the vicinity of the boundaries the effect of the transverse stabilization parameter α_n leads to a noticeable change of the direction of $\boldsymbol{\alpha}$.

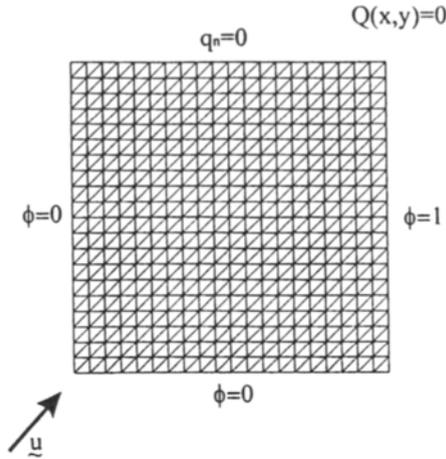


Figure 4. Example 1. Sourceless advective-diffusive problem with diagonal velocity. Finite element mesh of 800 linear triangles.

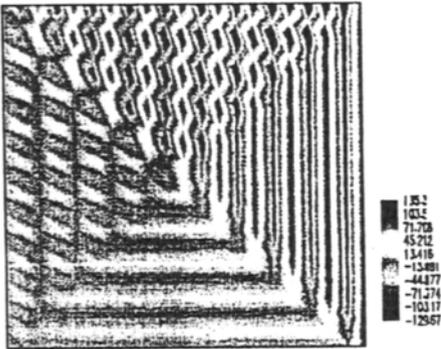


Figure 5. Example 1. Initial oscillatory distribution of ϕ for $\alpha_n^{(e)} = \alpha_n^{(e)} = 0$.

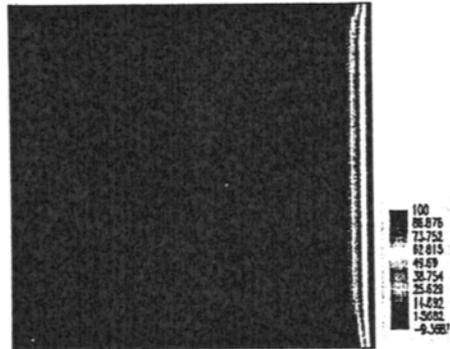


Figure 6. Example 1. Final distribution of ϕ after 7 iterations.

5.2 Example 2. Two dimensional advective-diffusive problem with no source and non uniform Dirichlet boundary conditions

The advection-diffusion equations are now solved with

$$\Omega =] - \frac{1}{2}, \frac{1}{2}[\times] - \frac{1}{2}, -\frac{1}{2}[\quad , \quad \mathbf{u} = [\cos \theta, -\sin \theta]^T$$

$$k_x = k_y = 10^{-6}, \quad Q(x, y) = 0 \quad , \quad \bar{\phi}(x, y) = \begin{cases} 1 & \text{if } (x, y) \in \Gamma_{\phi_1} \\ 0 & \text{if } (x, y) \in \Gamma_{\phi_2} \end{cases}$$

with $\Gamma_{\phi_1} = \{-1/2\} \times [1/4, 1/2] \cup] - 1/2, 1/2[\times \{1/2\}$, $\Gamma_{\phi_2} = \Gamma_{\phi} - \Gamma_{\phi_1}$ and $\Gamma_q = 0$.

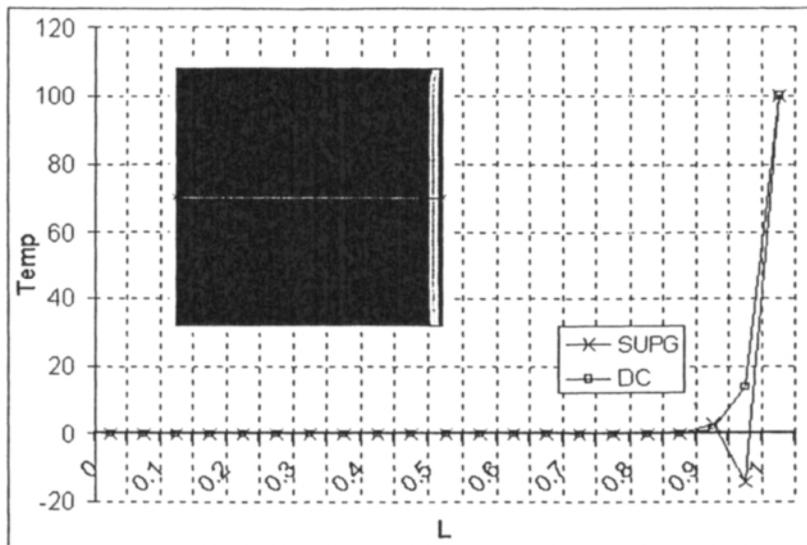


Figure 7. Distribution of ϕ along a central line obtained with the present discontinuity capturing method (DC) and the SUPG formulation ($\alpha_n = 0$).

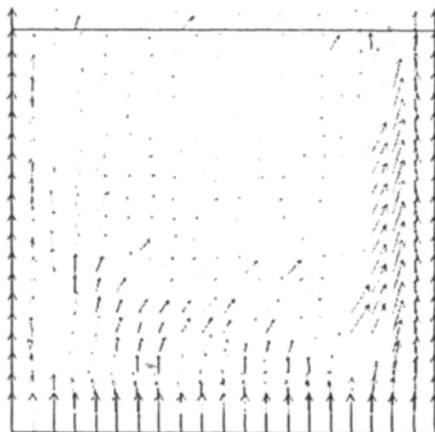


Figure 8. Example 1. Final distribution of the stabilization vector $\alpha = \alpha_s \mathbf{u} + \alpha_n \mathbf{u}_n$.

A unstructured mesh of 902 linear triangles has been chosen (Figure 9.). The problem has been chosen for an angle of \mathbf{u} given by $\tan \theta = 2$. Once again the initial values ${}^o\alpha_s^{(e)} = {}^o\alpha_n^{(e)} = 0$ have been taken.

Figure 10 shows the oscillatory distribution of ϕ obtained for the first solution, as expected. The final distribution of ϕ after 7 iterations is displayed in Figures 11 and

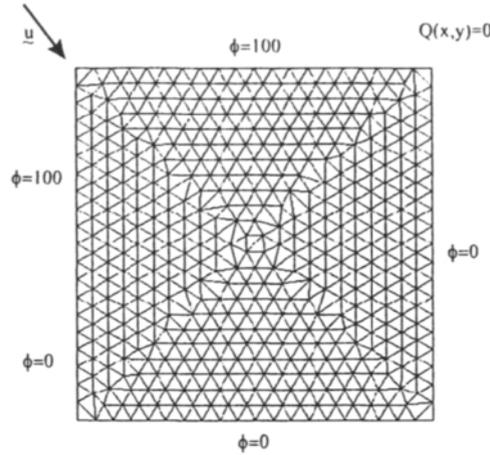


Figure 9. Example 2. Two dimensional advective-diffusive problem with zero source and non uniform boundary conditions. Geometry and unstructured finite element mesh of 902 linear triangles.

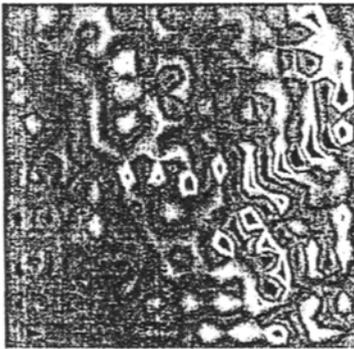


Figure 10. Example 2. Initial oscillatory solution for ϕ obtained for $\alpha_n^{(e)} = \alpha_n^{(e)} = 0$.

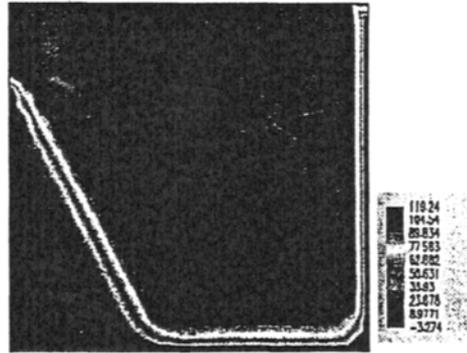


Figure 11. Final solution of ϕ after 7 iterations.

12. Note that both the boundary layers at the edges and the internal sharp layer are captured with minor oscillations. These oscillations are more pronounced near the right hand side edge (Figure 13 and 15) when $\alpha_n^{(e)} = 0$ is taken through out the adaptive process (SUPG solution).

Figure 14 shows finally the distribution of the stabilization vector $\alpha = \alpha_s u + \alpha_n u_n$. Again note that the direction of α in the smooth part of the solution is aligned with that of the velocity vector, whereas the effect of the transverse stabilization term is very pronounced near the sharp gradient boundary regions. This leads to a change in the direction of α in these zones.

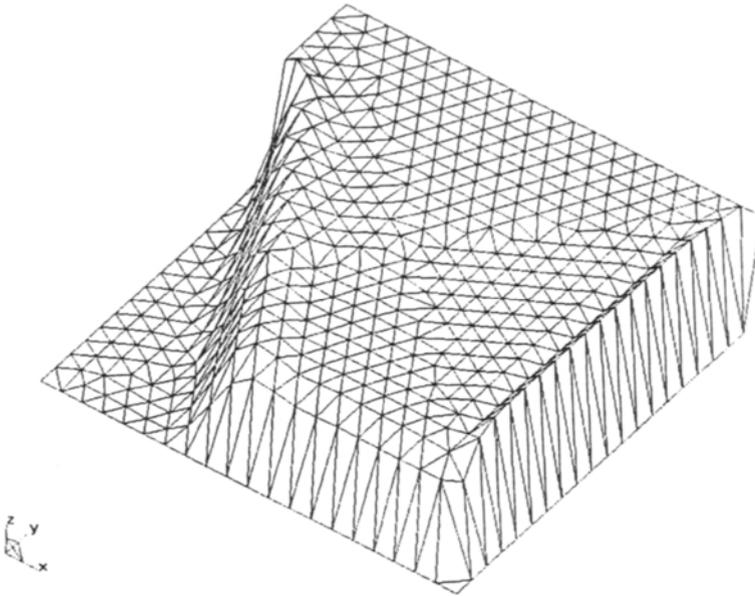


Figure 12. Example 2. Final distribution of ϕ after 7 iterations.

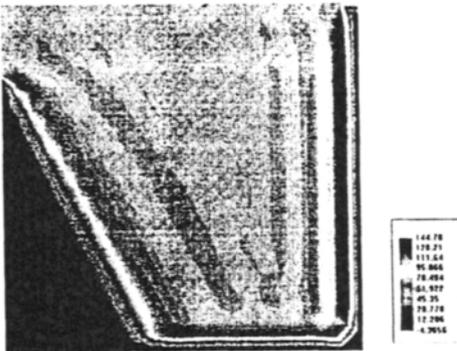


Figure 13. Example 2. Final distribution of ϕ obtained with $\alpha_n = 0$ (SUPG method).

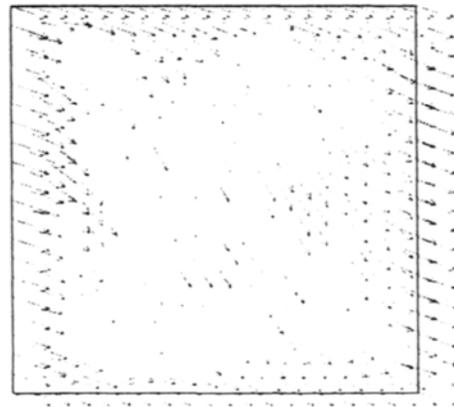


Figure 14. Example 2. Final distribution of the stabilization vector $\alpha = \alpha_s \mathbf{u} + \alpha_n \mathbf{u}_n$.

CONCLUSIONS

The new stabilized form of the governing differential equations derived via a “finite increment calculus” approach seems to be the natural root for obtaining stable finite element methods for advective-diffusive problems. The stabilized governing

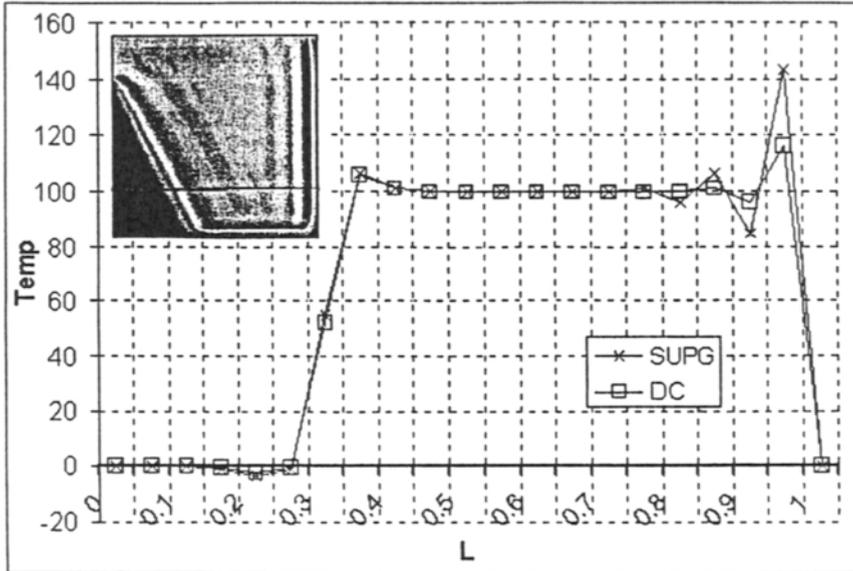


Figure 15. Distribution of ϕ along a central line obtained with the present discontinuity capturing method (DC) and the SUPG formulation ($\alpha_n = 0$).

equations are also the basis for computing line the streamline and crosswind stabilization parameters necessary to capture arbitrary sharp transverse layers. The new stabilization approach can be interpreted as a class of adaptive methods where the numerical solution is enhanced by progressively improving the value of the stabilization parameter, while keeping the mesh and the finite element approximation unchanged. The efficiency of this alpha-adaptive procedure has been shown for two problems with sharp gradients where accounting for the crosswind stabilization parameter has proved to be essential to obtain accurate solutions.

REFERENCES

1. C. Hirsch, "Numerical computations of internal and external flow", Vol. 2, John Wiley, 1990.
2. O.C. Zienkiewicz and R.L. Taylor, "The finite element method", Mc.Graw Hill, Vol. I., 1989, Vol. II., 1991.
3. D.W. Kelly, S. Nakazawa, O.C. Zienkiewicz and J.C. Heinrich, "A note on upwind and anisotropic balancing dissipation in finite element approximation to convective diffusion problems", *Int. J. Num. Meth. Engng.*, 15, pp. 1705-1711, 1980.
4. J.C. Heinrich, P.S. Hayakorn and O.C. Zienkiewicz, "An upwind finite element scheme for two dimensional convective transport equations", *Int. J. Num. Meth. Engng.*, 11, 131-143, 1977.
5. A. Brooks and T.J.R. Hughes, "Streamline upwind/Petrov-Galerkin formulation for convection dominated flows with particular emphasis on the incompressible Navier-Stokes

- equations", *Comput. Meth. Appl. Mech. Engng*, **32**, 199–259, 1982.
6. T.J.R. Hughes, L.P. Franca and G.M. Hulbert, "A new finite element formulation for computational fluid dynamics: VIII. The Galerkin/least-squares method for advective-diffusive equations", *Comput. Meth. Appl. Mech. Engng.*, **73**, pp. 173–189, 1989.
 7. J. Donea, "A Taylor-Galerkin method for convective transport problems", *Int. J. Num. Meth. Engng.*, **20**, pp. 101–119, 1984.
 8. O.C. Zienkiewicz and R. Codina, "A general algorithm for compr. and incompr. flow. Part I: The split characteristic based scheme", *Int. J. Num. Meth. in Fluids*, **20**, 869–85, 1995.
 9. T.J.R. Hughes, "Multiscale phenomena: Greens functions, subgrid scale models, bubbles and the origins of stabilized meth.", *Comp. Meth. Appl. Mech. Engng*, Vol. **127**, 387–401, 1995.
 10. R. Codina, "Comparison of some finite element methods for solving the diffusion-convection-reaction equation", Publication CIMNE, No. 101, Barcelona, 1996.
 11. S. Idelsohn, "Upwind techniques via variational principles", *Int. J. Num. Meth. Engng.*, **28**, N. 4, 669–784, 1989.
 12. F. Brezzi, M.O. Bristeau, L.P. Franca, M. Mallet and G. Rogé, "A relationship between stabilized finite element methods and the Galerkin method with bubble functions", *Comput. Meth. Appl. Mech. Engn.*, Vol. **96**, pp. 117–129, 1992.
 13. F. Brezzi, D. Marini and A. Russo, "Pseudo residual-free bubbles and stabilized methods", *Computational Methods in Applied Sciences '96*, J. Periaux et. al. (Eds.), J. Wiley, 1996.
 14. F. Brezzi, L.P. Franca, T.J.R. Hughes and A. Russo, " $b = \int g$ ", *Comput. Meth. Appl. Mech. Engn.*, Vol. **145**, pp. 329–339, 1997ss.
 15. E. Oñate, "Derivation of stabilized equations for advective-diffusive transport and fluid flow problems", *Comp. Meth. Appl. Mech. Engng*, **151**, 1–2, 233–67, 1998.
 16. E. Oñate, "On the stabilization of numerical solutions for advective-diffusive transport and fluid flow problems", Publication CIMNE no. 81, Barcelona 1996.
 17. E. Oñate, J. García and S. Idelsohn, "Computation of the stabilization parameter for the finite element solution of advective-diffusive problems", Publication CIMNE no. 100, Barcelona 1997. To be published in *Int. J. Num. Meth. Fluids*.
 18. J.J.R. Hughes, M. Mallet and A. Mizulami, "A new finite element formulation for computational fluid dynamics: II Beyond SUPG", *Comput. Meths. Appl. Mech. Engng.*, **54**, 341–355, 1986.
 19. T.E. Tezduyar and Y.J. Park, "Discontinuity capturing finite element formulation for non linear convection-diffusion-reaction equations", *Comput Meths. Appl. Mech. Engng.*, **59**, 307–325, 1986.
 20. R. Codina, "A discontinuity-capturing crosswind-dissipation for the finite element solution of the convection-diffusion equations", *Comp. Meth. Appl. Mech. Engng.*, **110**, 325–342, 1993.
 21. O.C. Zienkiewicz and J.Z. Zhu, "The superconvergent patch recovery (SPR) and adaptive finite element refinement", *Comp. Meth. Appl. Mech. Engng.*, **101**, 207–224, 1992.
 22. N.-E. Wiberg, F. Abdulwahab and X.D. Li, "Error estimation and adaptive procedures based on superconvergent patch recovery", *Archives Comp. Meth. Engng.*, **4**, 3, 203–242, 1997.

This Page Intentionally Left Blank

Adaptive strategy for transient/coupled problems Applications to thermoelasticity and elastodynamics

D. Aubry, D. Lucas, Tie B.

LMSS-Mat, CNRS, Ecole Centrale Paris, 92295 Chatenay Malabry Cedex, France

The interest in adaptive strategies for the finite element method is growing at a fast pace. Researchers are now fully convinced that this should be part of any numerical modelling and try to incorporate the methodology to their problems. Even the industry begins to require such an approach to increase the quality of the numerical results and to get more effective analyses to increase the ratio between computer costs and target error level.

The adaptive strategy has first been applied to linear elliptic problems, e.g. linear infinitesimal elasticity (Ainsworth et al. [1], Ewing [6], Verfürth [16], Tie [11]). It is now enlarged to nonlinear and transient problems (Tie et al. [12] [14], Eriksson et al. [5], Safjan et al. [10], Tie et al. [13], [15], Wiberg [17]). It is the aim of this paper to illustrate the application of the technique of residuals to adaptive refinement of coupled problems and to elastic wave propagation.

There are many other approaches to a posteriori error estimations. Well-known alternate techniques are presented in this workshop and they will not be recalled here. Some of them appear difficult to be extended easily to coupled and transient problems. It is believed that the residual approach is versatile enough to be easily applied to more complex problems. The two basic ideas used here rely on a space-time Galerkin formulation which provides a variational formulation of the error with respect to the residual, and the adjoint state which gives an upper bound of the error by the norm of the residuals weighted by some power of h and Δt .

1. COUPLED THERMOELASTICITY

The coupled equations of thermoelasticity are considered first. Let $u(t)$, $\tau(t)$ be respectively the displacement and temperature fields which satisfy the following coupled equations associated respectively with the conservation of the linear momentum and of the energy for given body forces f and thermal flux φ on a domain Ω :

$$\text{Div } C:\varepsilon(u(t)) - \alpha \text{ grad } \tau(t) + f(t) = 0 \quad (1)$$

$$c \tau'(t) = \varphi(t) + K \Delta \tau(t) - \alpha \operatorname{div} u'(t) \quad (2)$$

where C stands for the elasticity tensor, α the dilatation coefficient, c the heat capacity, K the conductivity. A symbol with a prime like τ' denotes the time derivative. Boundary and initial conditions should clearly be added to these equations, and it will be assumed here that homogeneous Dirichlet boundary conditions prevail on the boundary of the domain and the following initial conditions occur:

$$u(0) = u_0, \tau(0) = \tau_0 \quad (3)$$

Classically these equations are decoupled either in the stationary regime or when the deformation term in the conservation of energy may be neglected. This will not be assumed here and full coupling may occur between the displacement and temperature dependant variables. The time interval is $I =]0, \Delta t[$ which may be considered as a time step as well. The effect of successive time steps will not be analyzed here.

1.1 Space-time variational residual error

A space-time weak form of the above coupled equations (Eriksson et al. [5]) is easily built by introducing first a displacement trial function space $u(t) \in V_u$ and a temperature trial function space $\tau(t) \in V_\tau$ and then a displacement test function space $\bar{u} \in \bar{V}_u$ and a temperature trial function space $\bar{\tau} \in \bar{V}_\tau$:

$$\int_I (C:\varepsilon(u(t)), \varepsilon(\bar{u})) - \int_I (\alpha \tau(t), \operatorname{div} \bar{u}) = \int_I (f, \bar{u}), \forall \bar{u} \in \bar{V}_u \quad (4)$$

$$\int_I (c \tau'(t), \bar{\tau}) + \int_I (\alpha \operatorname{div} u'(t), \bar{\tau}) + \int_{\Delta t} (K \operatorname{grad} \tau(t), \operatorname{grad} \bar{\tau}) = \int_I (\varphi, \bar{\tau}), \forall \bar{\tau} \in \bar{V}_\tau \quad (5)$$

$$(\tau(0), \bar{\tau}) = (\tau_0, \bar{\tau}), (u(0), \bar{u}) = (u_0, \bar{u}) \quad (6)$$

In the above equation the following notation is used for two fields a and b which may be either scalars, vectors or tensors: $(a, b) = \int_{\Omega} a \cdot b dV$, where $a \cdot b$ stands for the standard appropriate dot product depending on the nature of the implied variables.

It is now possible to proceed to the space-time discretisation of the weak form by introducing two subspaces V_{uh} and $V_{\tau h}$ of the trial spaces V_u and V_τ . Basically they are assumed to be polynomial of degree one with respect to time on I and generated by a classical finite element space triangulation with respect to space. As for the test spaces \bar{V}_{uh} and $\bar{V}_{\tau h}$ they will be assumed to be generated by functions constant with respect to time on I and by finite element classical basis functions with respect to space.

$$u_h(t) = (t/\Delta t) u_{h1} + (1 - (t/\Delta t)) u_{h0} \quad (7)$$

Clearly more general polynomial dependency may be assumed. The approximations u_h and τ_h to the displacement and temperature are solutions of the equation underneath:

$$\int_I (C:\varepsilon(u_h), \varepsilon(\bar{u}_h)) dt - \int_I (\alpha \tau_h, \operatorname{div} \bar{u}_h) dt = \int_I (f, \bar{u}_h) dt, \forall \bar{u}_h \in \bar{V}_{uh} \quad (8)$$

$$\int_I (c \tau_h', \bar{\tau}_h) dt + \int_I (\alpha \operatorname{div} u_h', \bar{\tau}_h) dt + \int_I (K \operatorname{grad} \tau_h, \operatorname{grad} \bar{\tau}_h) dt = \int_I (\varphi, \bar{\tau}_h) dt, \forall \bar{\tau}_h \in \bar{V}_{\tau h} \quad (9)$$

The preceding equations are thus equivalent to:

$$(C:\varepsilon(\langle u_h \rangle), \varepsilon(\bar{u}_h)) - (\alpha \langle \tau_h \rangle, \text{div } \bar{u}_h) = (\langle f \rangle, \bar{u}_h), \quad \forall \bar{u}_h \in \bar{V}_{uh} \quad (10)$$

$$(c(\tau_h - \tau_{h0}), \bar{\tau}_h) + (\alpha \text{div}(u_{h1} - u_{h0}), \bar{\tau}_h) + (K \text{grad } \langle \tau_h \rangle, \text{grad } \bar{\tau}_h) = (\langle \varphi \rangle, \bar{\tau}_h), \quad \forall \bar{\tau}_h \in \bar{V}_{th} \quad (11)$$

where the following notation is used: $\langle \zeta \rangle = \Delta t (\zeta_{h1} + \zeta_{h0})/2$. The displacement and temperature errors are now introduced: $e_u = u - u_h$, $e_\tau = \tau - \tau_h$. They satisfy the two following equations :

$$\int_1 \{ (C:\varepsilon(e_u), \varepsilon(\bar{u})) - (\alpha e_\tau, \text{div } \bar{u}) \} dt = \int_1 \{ (f, \bar{u}) - (C:\varepsilon(u_h), \varepsilon(\bar{u})) + (\alpha \tau_h, \text{div } \bar{u}) \} dt \quad (12)$$

$$\int_1 \{ (c e_\tau', \bar{\tau}) + (\alpha \text{div } e_u', \bar{\tau}) + (K \text{grad } e_\tau, \text{grad } \bar{\tau}) \} dt = \int_1 \{ (\varphi, \bar{\tau}) - (c \tau_h', \bar{\tau}) - (\alpha \text{div } u_h', \bar{\tau}) - (K \text{grad } \tau_h, \text{grad } \bar{\tau}) \} dt \quad (13)$$

This is a residual equation (Eriksson et al. [5]) because the right-hand side may be transformed to residual mechanical equilibrium and thermal and fluxes across each element. By choosing $\bar{u} = e_u'$ and $\bar{\tau} = e_\tau$ in the preceding equation we get the following equation:

$$\|e_u(\Delta t)\|^2 + |e_\tau(\Delta t)|^2 + \int_1 \|e_\tau\|^2 dt = \int_1 [(R_u, e_u) + (R_\tau, e_\tau')] dt \quad (14)$$

with: $\|e_u\|^2 = (C:\varepsilon(e_u), \varepsilon(e_u))$; $|e_\tau|^2 = (c e_\tau, e_\tau)$; $\|e_\tau\|^2 = (K \text{grad } e_\tau, \text{grad } e_\tau)$.

By using a Gronwall type of inequality it is possible to arrive at a bound on both errors with respect to the residuals. Then when the test functions $(\bar{u}, \bar{\tau})$ belong to $\bar{V}_{uh} \times \bar{V}_{th}$ then the two right-hand sides vanish. Thus to evaluate the compound error (e_u, e_τ) it is required to test the residual on a larger space than $\bar{V}_{uh} \times \bar{V}_{th}$, for instance $\bar{V}_{uh/2} \times \bar{V}_{th/2}$. The cost implied by such a strategy should of course be small compared to the overall computation. Consequently only local thermoelastic problems are solved on patches of potentially refined elements to estimate the error in displacement and temperature. The concept of adjoint state in the next paragraph will allow to arrive at a better representation of the error with respect to the residual.

1.2 Error representation with adjoint state

Eriksson et al. [5] has shown recently that the concept of adjoint state is particularly helpful to get an error representation. This approach is followed here and extended to thermo-elasticity. Let us define the adjoint-state (u^*, τ^*) which is a thermo-elastic solution with specially designed final conditions:

$$\text{Div } C:\varepsilon(u^*) - \alpha \text{grad } \tau^* = 0, \quad -c \tau^* = K \Delta \tau^* + \alpha \text{div } u^* \quad (15)$$

$$u^*(\Delta t) = e_u(\Delta t); \quad \tau^*(\Delta t) = e_\tau(\Delta t)$$

Then it is easy to show that:

$$\|e_u(\Delta t)\|^2 + |e_\tau(\Delta t)|^2 = (\varepsilon(u^*), C:\varepsilon(e_u))(0) + (\tau^*, c e_\tau)(0) + \int_1 \{ (\varepsilon(u^*), C:\varepsilon(e_u)) - (\text{div } u^*, \alpha e_\tau) + (\tau^*, c e_\tau') + (\tau^*, \alpha \text{div } e_u') + (\text{grad } \tau^*, K \text{grad } e_\tau) \} dt$$

But here appears on the right-hand side the residual equation with the test functions equal to the adjoint state. The trick now is to take benefit of the time-space Galerkin formulation

which implies that the residual vanishes on the finite element space so that the interpolate of the adjoint state may be introduced in the right-hand side above. Finally this term may be bounded above by the local residuals and a norm of the adjoint state. But if regularity results are assumed the norm of the adjoint state is itself bounded by the data ($e_n(\Delta t)$, $e_r(\Delta t)$). This then implies an upper-bound on the error by the thermo-elastic residual. Clearly the analysis presented above was restricted to a given time step Δt but can be considered to several time steps. Then the projection phase from one mesh to the other as soon as mesh refinement is taken into account in the initial error.

1.3 Numerical results

As a numerical example, the solidifying test proposed by Celentano et al. [4] is considered. It consists in casting aluminium into an instrumented steel cylindrical mould. In our analysis the problem is transformed into a plane strain problem (Figure 1), and no interface treatment is made between aluminium and steel. The analysis begins when the mould is completely filled with aluminium in the liquid state. The initial temperature are assumed to be 700°C for the casting and 200°C for the mould. This initial temperature is kept at the exterior mould wall. We refer to [4] and Tie et al. [14] for the numerical value of the parameters.

The initial mesh is uniform and composed of 4-nodes quadrilateral elements (Figure 1). During the numerical computation, a series of nested FE meshes and a corresponding series of embedded FE spaces are built using h-version hierarchical bases. All newly added degrees of freedom are relative corrections with respect to FE solutions defined on coarser meshes (Tie [11],[12]). For each time step, several adaptive meshes can be generated. The global system of equations have a hierarchical block structure and is resolved by the hierarchical multimesh iterative solvers, whose convergence behaviour keeps relatively lightly deteriorated, despite the highly local mesh refinement. In Figure 2, the adaptive refined meshes at different time steps are presented. On the one hand, the zone around the casting-mould interface is refined at the beginning of the cast process, large variation of the stress field is observed at this zone. On the other hand, the moving solidification front is captured and followed up by the local mesh refinement. It is obvious that the local unrefinement of mesh behind this moving front would be interesting for decreasing the numerical cost of the computation.

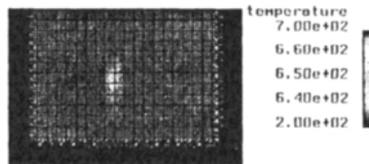


Figure 1. Solidification test: Problem description and initial FE mesh

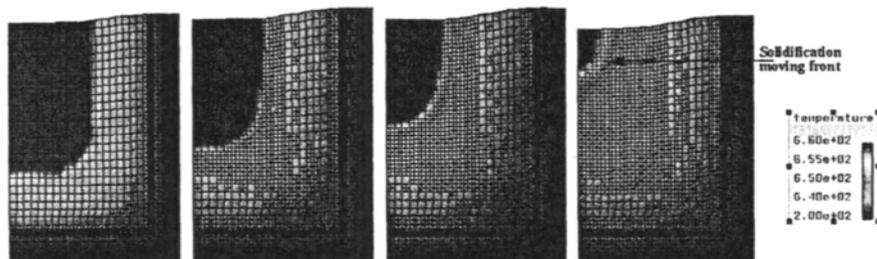


Figure 2. Adaptive mesh refinement and moving solidification front development

As far as phase changes effects are concerned, the problem is nonlinear. The residual approach presented above is simply applied here without more rigorous analysis. Furthermore local residuals are only measured on the spatial FE mesh at each time step as only spatial mesh refinement is considered here. It is however obvious that a whole time-space control and refinement would be possible and simple if it had been based on the residual analysis presented above.

2. TRANSIENT ELASTODYNAMICS

It is intended in the second part of this paper to show now that the two key ingredients that is the error residual and the adjoint state may also be extended to transient elastodynamics. The classical Navier equations are recalled :

$$\rho \mathbf{u}'' = \text{Div } \boldsymbol{\sigma}(\mathbf{u}) + \mathbf{f} \quad (16)$$

with Dirichlet boundary conditions and initial conditions : $\mathbf{u}(0)=\mathbf{u}_0$, $\mathbf{u}'(0)=\mathbf{v}_0$. First of all the system will be recast in a first-order system by introducing the velocity as a supplementary unknown and by adding a compatibility condition suggested by Johnson which will warranty conservation of energy:

$$\rho \mathbf{v}' = \text{Div } \boldsymbol{\sigma}(\mathbf{u}) + \mathbf{f}, \quad \text{Div } \boldsymbol{\sigma}(\mathbf{u}') = \text{Div } \boldsymbol{\sigma}(\mathbf{v}) \quad (17)$$

Then a space-time variational formulation (Hughes et al. [7], Hulbert [8], Johnson [9]) is obtained on the time interval $I=]0,\Delta t[$:

$$\int_1 (\rho \mathbf{v}', \bar{\mathbf{v}}) = \int_1 (\mathbf{f}, \bar{\mathbf{v}}) - \int_1 (\boldsymbol{\sigma}(\mathbf{u}), \boldsymbol{\varepsilon}(\bar{\mathbf{v}})) \quad \forall \bar{\mathbf{v}} \in \mathbf{V}_v \quad (18)$$

$$\int_1 (\boldsymbol{\sigma}(\mathbf{u}'), \boldsymbol{\varepsilon}(\bar{\mathbf{u}})) = \int_1 (\boldsymbol{\sigma}(\mathbf{v}), \boldsymbol{\varepsilon}(\bar{\mathbf{u}})) , \quad \forall \bar{\mathbf{u}} \in \mathbf{V}_u \quad (19)$$

It is now possible to proceed to the space-time discretisation of the weak form by introducing two subspaces \mathbf{V}_{vh} and \mathbf{V}_{uh} of the trial spaces \mathbf{V}_v and \mathbf{V}_u . Basically they are assumed to be polynomial of degree one with respect to time on I and generated by a classical finite element space triangulation with respect to space. As for the test spaces $\bar{\mathbf{V}}_{vh}$ and $\bar{\mathbf{V}}_{uh}$ they will be assumed to be generated by functions constant with respect to time on I and finite element classical basis functions with respect to space.

$$\mathbf{u}_h(t) = (t/\Delta t) \mathbf{u}_{h1} + (1 - (t/\Delta t)) \mathbf{u}_{h0} \quad (20)$$

Clearly more general polynomial dependency may be assumed. The space-time discretization leads to the following variational equation :

$$\int_I (\rho v_h', \bar{v}_h) = \int_I (f, \bar{v}_h) - \int_I (\sigma(u_h), \varepsilon(\bar{v}_h)) , \forall \bar{v}_h \in \bar{V}_{v_h} \quad (21)$$

$$\int_I (\sigma(u_h'), \varepsilon(\bar{u}_h)) = \int_I (\sigma(v_h), \varepsilon(\bar{u}_h)) , \forall \bar{u}_h \in \bar{V}_{u_h} \quad (22)$$

This is equivalent to the following standard discrete scheme with the unknowns (v_{h1}, u_{h1}) :

$$(\rho (v_{h1} - v_{h0}), \bar{v}_h) = \Delta t (\langle f \rangle, \bar{v}_h) - \Delta t (\sigma(\langle u_h \rangle), \varepsilon(\bar{v}_h)) \quad \forall \bar{v}_h \in \bar{V}_{v_h} \quad (23)$$

$$(\sigma(u_{h1} - u_{h0}), \varepsilon(\bar{u}_h)) = \Delta t (\sigma(\langle v_h \rangle), \varepsilon(\bar{u}_h)) , \forall \bar{u}_h \in \bar{V}_{u_h} \quad (24)$$

2.1 Residual error equation

Let us again define the error by: $e_u = u_h - u$, $e_v = v_h - v$. Then it satisfies the following variational residual equation which cancels on the space $\bar{V}_{v_h} \times \bar{V}_{u_h}$ itself :

$$\int_I (\rho e_v', \bar{v}) + \int_I (\sigma(e_v), \varepsilon(\bar{v})) = \int_I (f, \bar{v}) - \int_I (\sigma(u_h), \varepsilon(\bar{v})) - \int_I (\rho v_h', \bar{v}) , \forall \bar{v} \in \bar{V}_v \quad (25)$$

$$\int_I (\sigma(e_u' - e_u), \varepsilon(\bar{u})) = - \int_I (\sigma(u_h' - v_h), \varepsilon(\bar{u})) , \forall \bar{u} \in \bar{V}_u \quad (26)$$

Note here how the concept of residual is here generalised especially with respect to the second equation to the compatibility between the displacement derivative and the velocity, where in the context of finite differences it would be considered as a truncation error. Again it is possible by testing this residual equation on a finer mesh or through local Dirichlet elastodynamics problems to implement an adaptive strategy both with respect to space and time discretization. To arrive at a nice representation of the error the concept of the adjoint state will be again used in the next paragraph.

2.2 Residual error representation by adjoint state

In the case of the elastodynamic equation the adjoint state (v^*, u^*) adapted to error representation is defined by the following backwards wave equation (Eriksson et al. [6]):

$$\rho v^{*'} = \text{Div } \sigma(u^*), \quad \text{Div } \sigma(u^{*'}) = \text{Div } \sigma(v^*) \quad (27)$$

$$u^*(\Delta t) = e_u(\Delta t) , \quad v^*(\Delta t) = e_v(\Delta t) \quad (28)$$

Now if we multiply the first equation by e_v and the second equation by e_u , perform adequate integration by parts and add the two equations, we get :

$$\begin{aligned} |e_v(\Delta t)|^2 + |e_u(\Delta t)|^2 &= (v^*, \rho e_v)(0) + (\varepsilon(u^*), \sigma(e_v))(0) \\ &+ \int_I (v^*, \rho e_v') + \int_I (\varepsilon(v^*), \sigma(e_u)) + \int_I (\varepsilon(u^*), \sigma(e_u' - e_u)) \end{aligned} \quad (29)$$

which is again the variational residual equation applied on the adjoint state plus the initial error. Then taking into account the Galerkin projection result, the projection of the adjoint state may be introduced without perturbing the above equation. The estimation of the norm of the error then boils down to the estimation of the distance between (v^*, u^*) and its

interpolate and a regularity results which give an upperbound of the adjoint state with respect to its final value $(e_v(\Delta t), e_u(\Delta t))$.

2.3 Numerical examples

Two numerical examples will now be given where the refinement procedure is controlled by the residual defined above. As indicated in the first section, the local residuals are only measured and controlled on the spatial meshes at each time step. The first example is a square membrane in plane stress. The left boundary is clamped and the right one is submitted to a uniform Heaviside loading. The lateral boundary is free. The results are shown at different time steps. The generated P-wave is clearly seen together with refinement in the vicinity of the front of the wave, before it reaches the left end. Then due to the clamped boundary condition at the left, waves at the corner are triggered, diffract from them and interfere with the pure reflected P-wave from the left boundary. Clearly such a complex behaviour could be obtained only with a much refined mesh and a coarser one would not capture wave interferences.

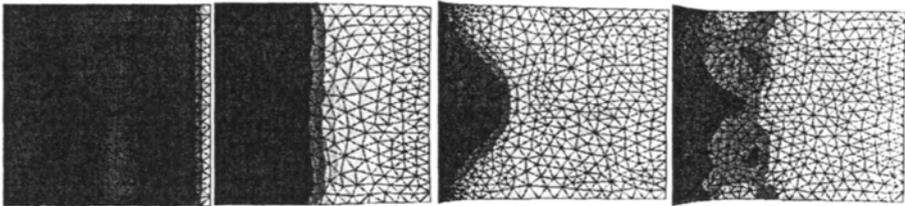


Figure 3. Displacements norms at different time steps

In the second example, there is a further geometrical singularity with a reentrant corner (L shaped domain). On a vertical segment at the upper left end again a Heaviside traction is applied. The initial mesh is made of 68 triangles. Deformed and adapted meshes are shown again at different time steps (Figure 4).

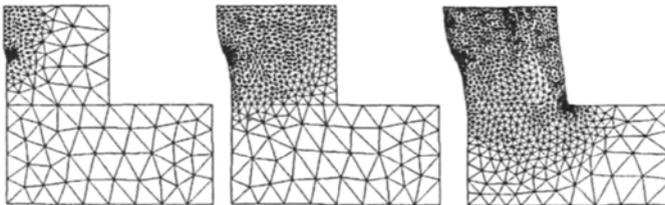


Figure 4. Adapted and deformed meshes at different time steps

3. CONCLUSION

The residual technique has been extended to a coupled system such as thermoelasticity and transient elastodynamics. Using a space-time Galerkin approach it embodies both the classical residual and the truncation error usually expressed as a compatibility condition

between the velocity and the time derivative of the displacement. The adjoint state allows to represent elegantly the error at the end of a time step with respect to the residuals and thus to design an adaptive strategy based on them (Aubry et al. [3]). It is believed to be an efficient technique versatile enough to be extended to coupled and transient problems where the lack of intuition and insight into the singularities of the exact solution makes adaptive approaches mandatory. In both cases only much refined meshes would be able to capture sharp temperature or wave fronts. Obviously the whole potential of the approach will be of greatest interest in the three-dimensional case applied to complex structures.

REFERENCES

1. Ainsworth M., Oden J.T. (1993) A unified approach to a posteriori estimation using residual methods, *Numer. Math.*, 65, 23-50.
2. Aubry D., Lucas D., Tié B. (1997) A unified approach to adaptive computation, Giens, to be published in *European Journal for Finite Element Methods*
3. Aubry D., Tie B., (1997) The adjoint state as a general tool to get a posteriori error representation in coupled and transient problems, to be published.
4. Celentano D., Oller S. and Oñate E. (1996) A coupled thermomechanical model for the solidification of cast metals, *Int. J. Solids Structures*, Vol. 33, No 5, pp647-673.
5. Eriksson K., Estep D., Hansbo P., Johnson C. (1996) *Computational differential equations*, Cambridge University Press.
6. Ewing R. E. (1990) A posteriori error estimation, *Comp. Meth. Appl. Mech. Eng.*, 82, 59-72.
7. Hughes T.J.R., Hulbert G.M. (1988) Space-time FEM for elastodynamics, formulations and error estimates, *Comp. Meth. Appl. Mech. Engin.*, 66, pp 339-363.
8. Hulbert G.M., Hughes T.J.R. (1990) Space-time finite element methods for second-order hyperbolic equations, *Comp. Meth. Appl. Mech. Eng.*, 84, 327-348.
9. Johnson C. (1993) Discontinuous Galerkin finite element methods for second order hyperbolic problems, *Comp. Meth. Appl. Mech. Eng.*, 107, 117-129.
10. Safjan A., Oden J.T. (1993) High-order Taylor-Galerkin and adaptive h-p methods for second-order hyperbolic systems: applications to elastodynamics, *Comp. Meth. Appl. Mech. Eng.*, 103, 187-230.
11. Tie B. (1993) *Éléments finis adaptatifs et hiérarchiques en élastoplasticité, localisation des déformations*, PhD thesis, Ecole Centrale Paris.
12. Tie B., Aubry D. (1996) Hierarchical and adaptive FE strategy for non-linear and coupled structural computation, pp 516-522, *Proc. of 2nd ECCOMAS Conference on Numer. Meth. Engin.*
13. Tie B., Aubry D. (1996) Integrated approach to a posteriori error estimate, adaptive mesh refinement and hierarchical multi-mesh solver, in preparation.
14. Aubry D., Tié B. (1997) Adaptive mesh refinement for phase-change effects capture in thermoelasticity, Bochum, IUTAM symposium, to be published
15. Tie B., Aubry D. (1997) Coupled effects on the local mesh refinement and shear band formation in fluid-saturated porous media, in preparation.
16. Verfürth R. (1996) *A review of a posteriori error estimation and adaptive mesh-refinement techniques*, Wiley
17. Wiberg N-E, Zeng L., Li X. (1992) Error estimation and adaptivity in elastodynamics, *Comp. Meth. Appl. Mech. Eng.*, 101, 369-395.

Error estimation and adaptive finite element analysis of softening solids

Antonio Huerta, Pedro Díez, Antonio Rodríguez-Ferran ^a and Gilles Pijaudier-Cabot ^b

^aDepartamento de Matemática Aplicada III, E.T.S. de Ingenieros de Caminos, UPC, Campus Nord, Módulo C-2, E-08034 Barcelona (Spain)

^bLMT – E.N.S. de Cachan, C.N.R.S., Université Paris 6
61, avenue du Président Wilson, F-94235 Cachan Cedex (France) and Institut
Universitaire de France

1. INTRODUCTION

Adaptive finite element computations are nowadays regarded as a natural approach for the analysis of localization in strain softening solids, because the localization area —where a very fine mesh is needed— is usually not known *a priori*. These adaptive strategies have three main ingredients. The first one is an algorithm for increasing/decreasing the richness of the interpolation in a particular area of the computational domain. For instance, a good mesh generator if *h*-refinement [11,21] is used. In this work, the mesh generator developed by Sarrate [19] is used. The second one is an error estimator or error indicator, which is employed to locate where there is a need for refinement/de-refinement. Finally, the third ingredient is a remeshing criterion, which must be used to translate the output of the error analysis into the input of the mesh generator (for instance, the distribution of desired mesh sizes). The criterion proposed by Li and Bettess [9] is used here.

Only the error analysis is discussed here, because it is probably the most crucial component of the adaptive strategy for softening problems, [12]. The main difficulty is that it must tackle the material nonlinearity associated to path-dependent constitutive models. Two basic approaches can be used: error indicators or error estimators. In the literature, adaptivity in softening problems has been associated to error indicators, [13,16,20], which are based on heuristic considerations. In contrast to that, a tool for assessing the error measured in the energy norm is proposed in this work. The obtained approximation to the error is asymptotically exact, that is, tends to the actual error if the element size tends to zero [2,3]. In that sense, this tool is an error estimator.

The nonlinear error estimator is presented in Section 2. Its application to adaptive computations in softening problems is shown by means of numerical examples in Section 3, where two regularized constitutive models are employed: Mazars nonlocal damage and Perzyna viscoplasticity. Finally, some concluding remarks are made in Section 4.

2. ERROR ESTIMATION

As previously said, an error estimator is a key feature in any adaptive procedure. This section is devoted to briefly describe a residual-type a-posteriori error estimator (introduced in [5]) which can be used in general nonlinear cases.

Using a mesh of characteristic size h , the Finite Element Method (FEM) provides a discrete equilibrium equation where the unknown is the nodal displacement vector \mathbf{u}_h :

$$\mathbf{f}_h^{\text{int}}(\mathbf{u}_h) = \mathbf{f}_h^{\text{ext}}, \quad (1)$$

where $\mathbf{f}_h^{\text{int}}(\mathbf{u}_h)$ is the vector of nodal internal forces associated with \mathbf{u}_h and $\mathbf{f}_h^{\text{ext}}$ is the discretized external force term.

Once Eq. (1) is solved, the solution \mathbf{u}_h is affected by an error that has to be estimated. Since the actual displacements are unknown, the actual error cannot be computed. However, using a much finer mesh of characteristic size \tilde{h} ($\tilde{h} \ll h$), the FEM gives a new solution $\mathbf{u}_{\tilde{h}}$ which is much more accurate than \mathbf{u}_h because the regularized model ensures that the Finite Element Analysis converges as the element size goes to zero [7]. This solution can be taken as a reference solution and, consequently, the actual error can be fairly replaced by the reference error $\mathbf{e}_{\tilde{h}} := \mathbf{u}_{\tilde{h}} - \mathbf{u}_h$.

Nevertheless, the determination of $\mathbf{u}_{\tilde{h}}$ (or $\mathbf{e}_{\tilde{h}}$) requires to solve an equation analogous to Eq. (1) but in the finer mesh:

$$\mathbf{f}_{\tilde{h}}^{\text{int}}(\mathbf{u}_{\tilde{h}}) = \mathbf{f}_{\tilde{h}}^{\text{int}}(\mathbf{u}_h + \mathbf{e}_{\tilde{h}}) = \mathbf{f}_{\tilde{h}}^{\text{ext}}. \quad (2)$$

This problem is much more expensive than the original one and it is unaffordable from a computational point of view.

In the remainder of this section a method for approximating $\mathbf{e}_{\tilde{h}}$ by low cost local computations is presented. That is, instead of solving Eq. (2), $\mathbf{e}_{\tilde{h}}$ is approximated by solving a set of local problems. This method is splitted in two phases. First, a simple residual problem is solved inside each element and an interior estimate is obtained. Second, a new family of simple problems is considered and the interior estimate is complemented adding a new contribution. The first phase is called interior estimation and the second one is called patch estimation.

2.1. An error estimator based on local computations

Interior estimation

The natural partition of the domain in order to solve local problems is the set of elements of the "coarse" computational mesh (denoted by Ω_k , $k = 1, \dots$).

A finer reference mesh is constructed by the assembly of submeshes discretizing each element. These elementary submeshes are built from a discretization of the reference element mapped into the elements of the actual mesh (see Figure 1).

Then, the elementary submeshes can be used to solve the error equation (2) on each element Ω_k of the original mesh. However, the solution of such problems requires proper boundary conditions for the error. Most of residual type error estimators [1,4,8], solve Eq. (2) prescribing the error flux around each element Ω_k . This is imposed using a flux splitting procedure which is generally involved and expensive from a computational point of view. Here, the elementary problems are solved in a straightforward manner imposing

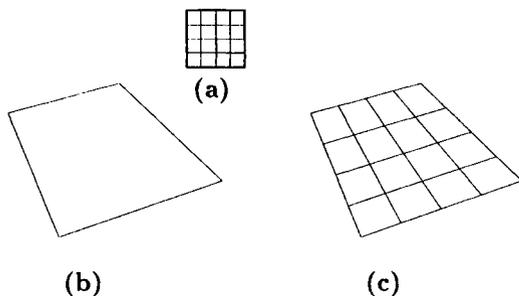


Figure 1. (a), reference submesh mapped into (b), an element, to get (c), an elementary submesh

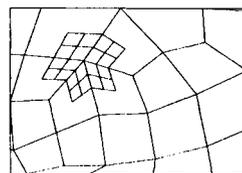


Figure 2. Patch submesh centered in a node of the computational mesh

the displacement error to vanish along the boundary of each element Ω_k (see [5]). That is, the error is prescribed to zero in all the boundary nodes of the elementary submesh. An approximation to the error within each element Ω_k is obtained solving each of these discrete local problems, this local approximation is called interior estimate and denoted by ϵ_k . The vector ϵ_k is an approximation to the restriction of the reference error $e_{\bar{h}}$ inside the element Ω_k .

Once the elementary problems are solved, the local interior estimates can be assembled to build up a global estimate ϵ having values in the whole domain Ω ,

$$\epsilon = \sum_k \epsilon_k. \quad (3)$$

The previous choice of the homogeneous boundary condition implies that $\|\epsilon\| \ll \|e_{\bar{h}}\|$. The reference error $e_{\bar{h}}$ is, most probably, nonzero along the element edges, thus ϵ may be a poor approximation to $e_{\bar{h}}$. In other words, the information contained in the flux jumps is ignored.

Patch estimation and complete estimate

Once the interior estimate is computed, it is necessary to add the contribution of the flux jumps, that is, to improve the error estimation by adding non-zero values in the element boundaries. This can be done following the same idea of the interior estimation, precluding the direct computation of flux jumps and avoiding any flux splitting procedure.

In this second phase, a different set of subdomains, called patches, is considered, each one overlapping a few number of elements and covering a part of the elementary boundaries. Using the elementary submeshes of Figure 1, the most natural choice for patch subdomains is to associate them with the nodes of the mesh: each patch is associated with a node and includes a fourth of every element sharing that node (see Figure 2 for an illustration).

The idea is to use this new partition to define new local problems for the error and to solve them. Local boundary conditions are imposed in the same fashion as in the previous phase (interior estimate). A new approximation to the error is obtained. This new approximation takes non-zero values in the boundary of the elements, where the interior estimate ϵ vanishes. In order to solve these problems each patch must be discretized by a patch submesh. The discretization of Eq. (2) using this patch submesh leads to a system of equations analogous to the interior estimation case. Since patches cover the edges of the elements, the restriction of Eq. (2) to each patch accounts for the flux jumps.

For the patch estimate, local and global estimates can be computed following the same strategy used for the interior estimate. In order to properly add the contribution of the patch estimates to the previously computed interior estimates, the patch estimate must be forced to verify an additional restriction: it must be orthogonal to the global interior estimate ϵ . This orthogonality condition can be easily implemented using the Lagrange multiplier technique.

Thus, the interior estimate ϵ is completed and a new approximation to $e_{\bar{h}}$ is found.

2.2. Error equations for linear and nonlinear problems

A method for approximating the reference error by low-cost local computations has been described. The particularization of the error equation (2) to linear and nonlinear problems is discussed next. These are the equations that must be solved locally in order to estimate the error.

If the problem is linear, $\mathbf{f}^{\text{int}}(\mathbf{u})$ is a linear function of \mathbf{u} and, consequently,

$$\mathbf{f}_h^{\text{int}}(\mathbf{u}_h) = \mathbf{K}_h \mathbf{u}_h \quad \text{and} \quad \mathbf{f}_{\bar{h}}^{\text{int}}(\mathbf{u}_{\bar{h}}) = \mathbf{K}_{\bar{h}} \mathbf{u}_{\bar{h}}, \quad (4)$$

where \mathbf{K}_h and $\mathbf{K}_{\bar{h}}$ stand for the stiffness matrices associated with the coarse mesh and the finer mesh respectively. These equations can be easily manipulated and a linear equation for the reference error is found:

$$\mathbf{K}_{\bar{h}} \mathbf{e}_{\bar{h}} = \mathbf{f}_{\bar{h}}^{\text{ext}} - \mathbf{f}_{\bar{h}}^{\text{int}}(\mathbf{u}_h) =: -\mathbf{r}_{\bar{h}}(\mathbf{u}_h), \quad (5)$$

where $\mathbf{r}_{\bar{h}}(\mathbf{u}_h)$ is the residual.

Figure 3 shows a graphic illustration of the meaning of the reference error and its relation with the residual. The reference error, $\mathbf{e}_{\bar{h}} = \mathbf{u}_{\bar{h}} - \mathbf{u}_h$, and the residual, $\mathbf{r}_{\bar{h}}(\mathbf{u}_h) := \mathbf{f}_{\bar{h}}^{\text{int}}(\mathbf{u}_h) - \mathbf{f}_{\bar{h}}^{\text{ext}}$, are related in terms of the stiffness matrix in the reference mesh, $\mathbf{K}_{\bar{h}}$, as indicated in Eq. (5). Although this illustration may appear trivial in the linear case, a variation of this figure provides a good understanding for the nonlinear case.

On the contrary, if the problem is nonlinear, Eq. (5) does not hold and the only available equation for the error is Eq. (2). This is a general nonlinear equation and must be solved using any standard nonlinear solver, see Figure 4. However, it is important to notice that the approximate solution \mathbf{u}_h can be used as a good initial guess for $\mathbf{u}_{\bar{h}}$ and, consequently, the solution of the nonlinear equations providing $\mathbf{u}_{\bar{h}}$ is straightforward (typically only two iterations are needed to reach the prescribed accuracy).

Moreover, here, the unknown $\mathbf{e}_{\bar{h}}$ is small compared with $\mathbf{u}_{\bar{h}}$ and, consequently, this nonlinear problem is much easier than the original one. If this assumption is true and the tangent stiffness matrix is available, the internal force vector can be approximated by a linear Taylor expansion, that is,

$$\mathbf{f}_{\bar{h}}^{\text{int}}(\mathbf{u}_h + \mathbf{e}_{\bar{h}}) \approx \mathbf{f}_{\bar{h}}^{\text{int}}(\mathbf{u}_h) + \mathbf{K}_{T,\bar{h}}(\mathbf{u}_h) \mathbf{e}_{\bar{h}}, \quad (6)$$

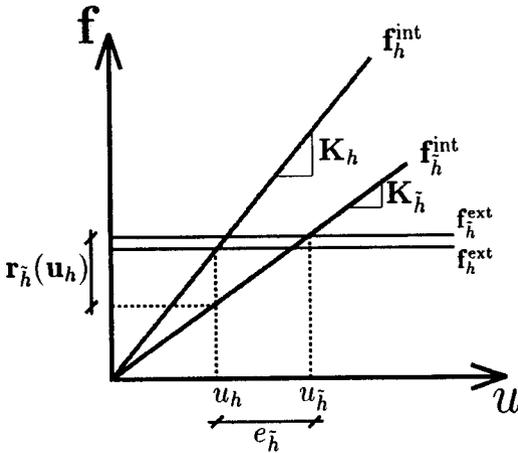


Figure 3. Graphic interpretation of reference error in linear problems

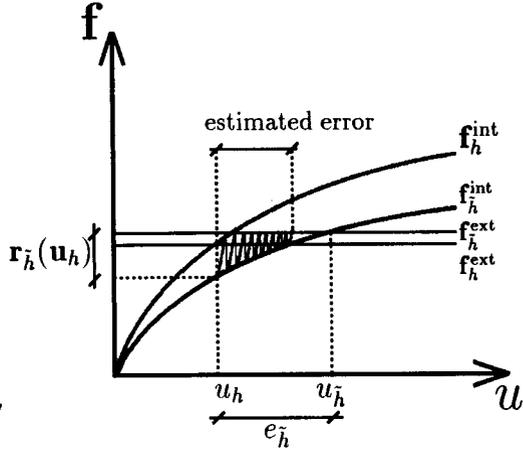


Figure 4. Graphic interpretation of reference error in nonlinear problems and fully nonlinear error estimation

where $\mathbf{K}_{T,\tilde{h}}(\mathbf{u}_h)$ is the tangent matrix describing the linear behavior of the system (discretized with the finer mesh) around \mathbf{u}_h .

Replacing Eq. (6) in Eq. (2), a new error equation is found:

$$\mathbf{K}_{T,\tilde{h}}(\mathbf{u}_h)\mathbf{e}_{\tilde{h}} = -\mathbf{r}_{\tilde{h}}(\mathbf{u}_h), \tag{7}$$

where $\mathbf{r}_{\tilde{h}}(\mathbf{u}_h)$ is the residual, following the definition given in Equation (5). It is worth noting that Eq. (7) is linear and has exactly the same structure of Eq. (5).

Following the idea of the graphic illustration of Figure 3 and Figure 4, Figure 5 shows how the nonlinear case can be treated using a tangent approximation.

Once the error equation is set as a linear equation (for linear problems or nonlinear problems with available tangent stiffness matrix) or a fully nonlinear equation (if tangent matrix is not available), the reference error is approximated solving the local problems described in the previous section. Thus, the estimation of the error is splitted in two steps. First, elementary problems are solved with null error boundary conditions, and an interior estimate is computed. Second, adding the contribution of the patches, a complete estimate is obtained.

3. NUMERICAL EXAMPLES

3.1. Example 1: Mazars nonlocal damage model

The Mazars damage model is used for describing the behavior of brittle materials, such as concrete and other geomaterials [10,18]. A non-decreasing scalar damage factor D , which ranges between 0 (intact material) and 1 (completely damaged material), is

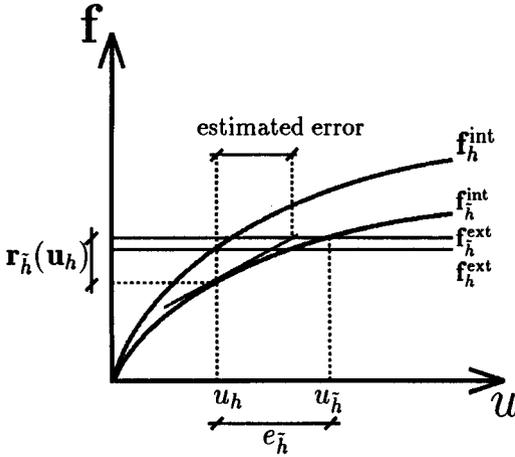


Figure 5. Graphic interpretation of reference error in nonlinear problems and error estimation using tangent approximation

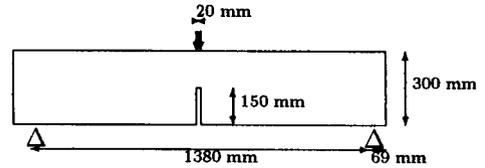


Figure 6. Three-point bending test on notched beam

employed to represent the loss of stiffness caused by damage. The evolution of D is controlled by the principal positive strains. The one-dimensional stress-strain law shows a clear softening behavior both in tension and compression. Because of softening, some kind of regularization is needed. The nonlocal regularization [15] can be applied to the Mazars model [10]. The basic idea is to use an average value of positive strains around each point as the internal variable that controls D . Because of this nonlocal regularization, tangent stiffness matrices are not readily available for the Mazars nonlocal damage model [17], and the linear tangent error equation (7) cannot be employed.

These ideas are illustrated with a simple test: the three-point bending of a notched beam, see Figure 6. The same material parameters of reference [18] are used. Figure 7 shows the final damage distribution and deformed shape. As expected, both damage and strains concentrate in the midspan. Figure 8 shows the load-displacement curve, with a clear softening structural response.

Two different adaptive computations were carried out with an acceptability criterion (maximum acceptable error) of 0.5%. In the first one, the beam lies on two point supports. Figure 9 shows the sequence of meshes obtained during the adaptive process. The error is considerably reduced (from 2.50% to 0.91%) in the first two steps. However the error is not further reduced in the subsequent steps, despite the increase in the number of elements, and the goal of 0.5% error cannot be attained. This is mainly due to the singularities associated to the point supports. As the mesh is refined near the supports, more errors are detected, and this process does not converge.

A completely different result is obtained if singularities are suppressed. In the second

case, the beam lies on two distributed (20mm wide) supports. Figure 10 shows the results of the adaptive computation. In only two steps, the error is reduced from 1.93% to an acceptable 0.48%. As expected, small elements are needed in the midspan and near the supports.

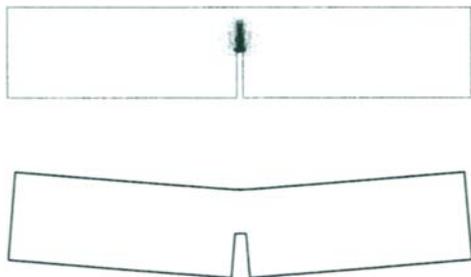


Figure 7. Damage distribution and deformed shape ($\times 200$)

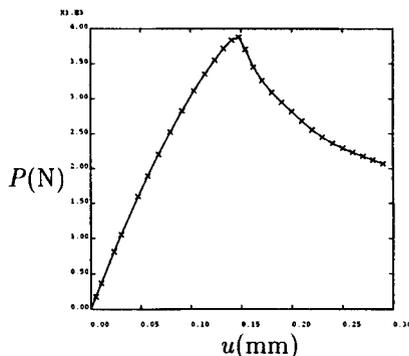


Figure 8. Load-displacement curve

3.2. Example 2: Perzyna viscoplasticity

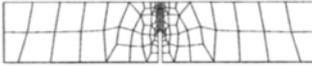
Due to the rate-dependent effects, viscoplastic constitutive models lead to regularized problems. Here, a simple viscoplastic model by Perzyna [14] is employed. For this model, tangent stiffness matrices are readily available, so the tangent version of the error equation, Eq. (7), may be employed. This is illustrated here by means of a numerical example, which also shows that an adaptive strategy based on error estimation is essential to capture complex failure mechanisms.

A plane strain rectangular ($50 \times 100 \text{ mm}^2$) specimen with two circular openings (2 mm radius, 11 mm vertical distance between centers) is subjected to uniaxial compression, see Figure 11. Two cases are considered: large and small horizontal separation between openings (examples 2a and 2b respectively), because this parameter has a great influence on the collapse mode. Due to the central symmetry, only one half of the specimen is used in computations. The acceptability criterion is set to 1.5%. The material parameters for this test can be found in [6].

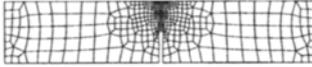
Example 2a (distant openings)

If the horizontal distance between the circular openings is large enough, one shear band is developed aligned with the two openings. The remeshing process (see Figure 13) leads to a mesh with a large number of elements concentrated along the two edges of the band. Starting with the initial mesh, which is roughly uniform, the error estimator detects large errors along this band, so more elements are added during the remeshing process.

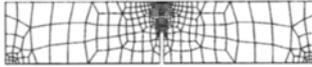
Figure 12 shows the general behavior of the solution, with a clear softening behavior in the force-displacement curve. The equivalent inelastic strain is concentrated along the shear band, both in the original and the final meshes of the remeshing process. That is, the captured collapse mechanism is the same in both meshes.



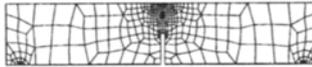
Mesh 0; 122 elements; error of 2.50%



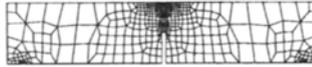
Mesh 1; 416 elements; error of 1.81%



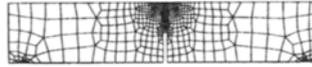
Mesh 2; 324 elements; error of 0.91%



Mesh 3; 364 elements; error of 0.95%

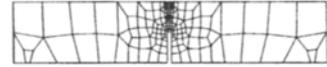


Mesh 4; 528 elements; error of 0.94%

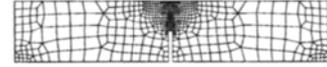


Mesh 5; 582 elements; error of 0.91%

Figure 9. Sequence of meshes with point supports



Mesh 0; 162 elements; error of 1.93%



Mesh 1; 576 elements; error of 0.52%



Mesh 2; 480 elements; error of 0.48%

Figure 10. Sequence of meshes with distributed supports

Example 2b (*close openings*)

On the contrary, if the circular openings are closer, the behavior of the solution is much more complex and the original mesh is not able to reproduce such a mechanism.

Figure 14 shows the sequence of meshes in this case. It is worth noting that, according to the concentration of elements, two bands are developed in the final mesh. In fact, the resulting bands are not aligned with the imperfections (as in example 2a), but have the opposite inclination. Indeed Figure 15 shows how the computed equivalent inelastic strain and the deformation evolve along the remeshing process. Only after two remeshing steps the mesh captures two bands. In the previous meshes the discretization is not accurate enough and only one band is completely developed. Since large deformations are considered, once the first band evolves enough, the kinematic mechanism associated with this band locks: then a second band appears as a new deformation mode with less energy.

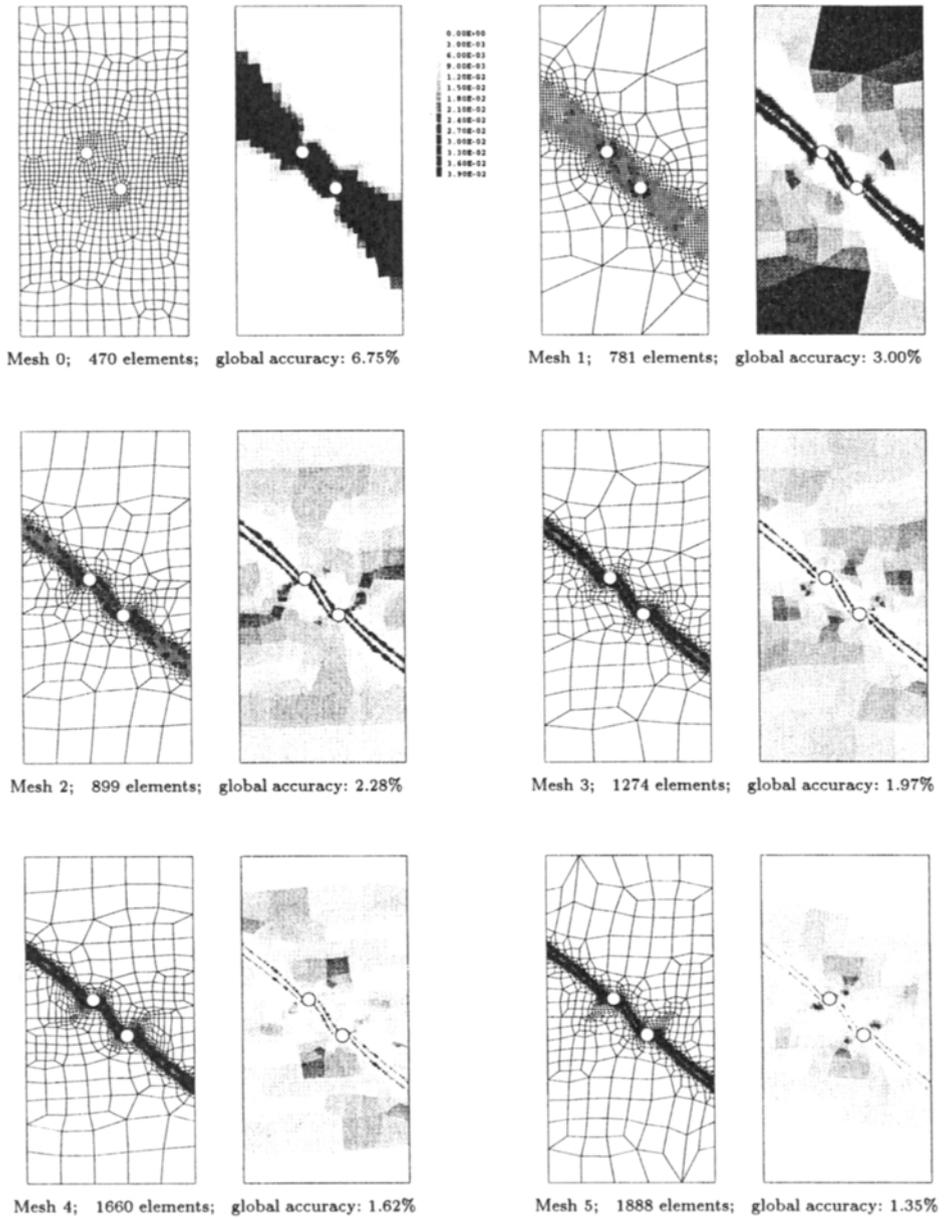


Figure 13. Example 2a. Remeshing process for a prescribed accuracy of 1.5%: sequence of meshes and estimated error distributions

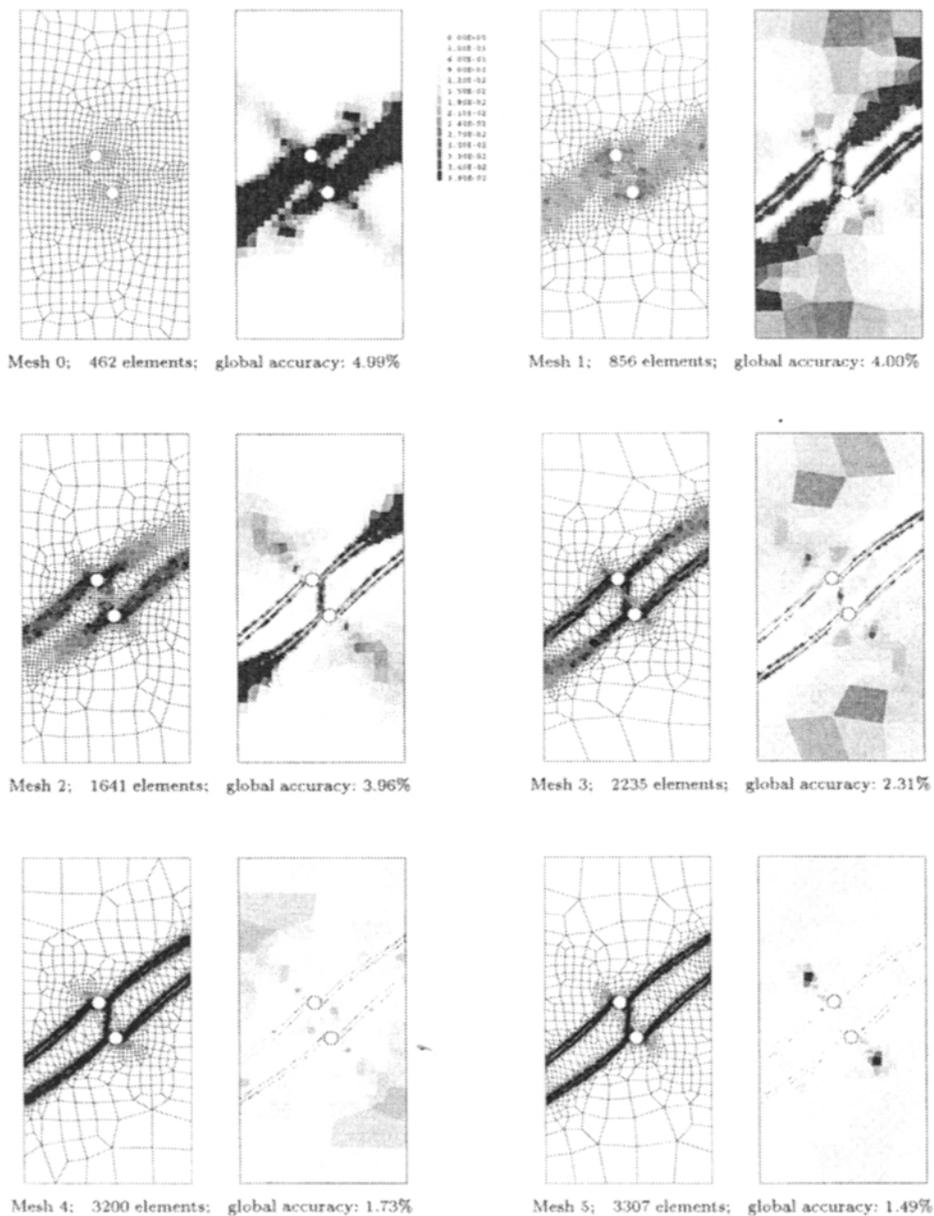


Figure 14. Example 2b. Remeshing process for a prescribed accuracy of 1.5%: sequence of meshes and estimated error distributions

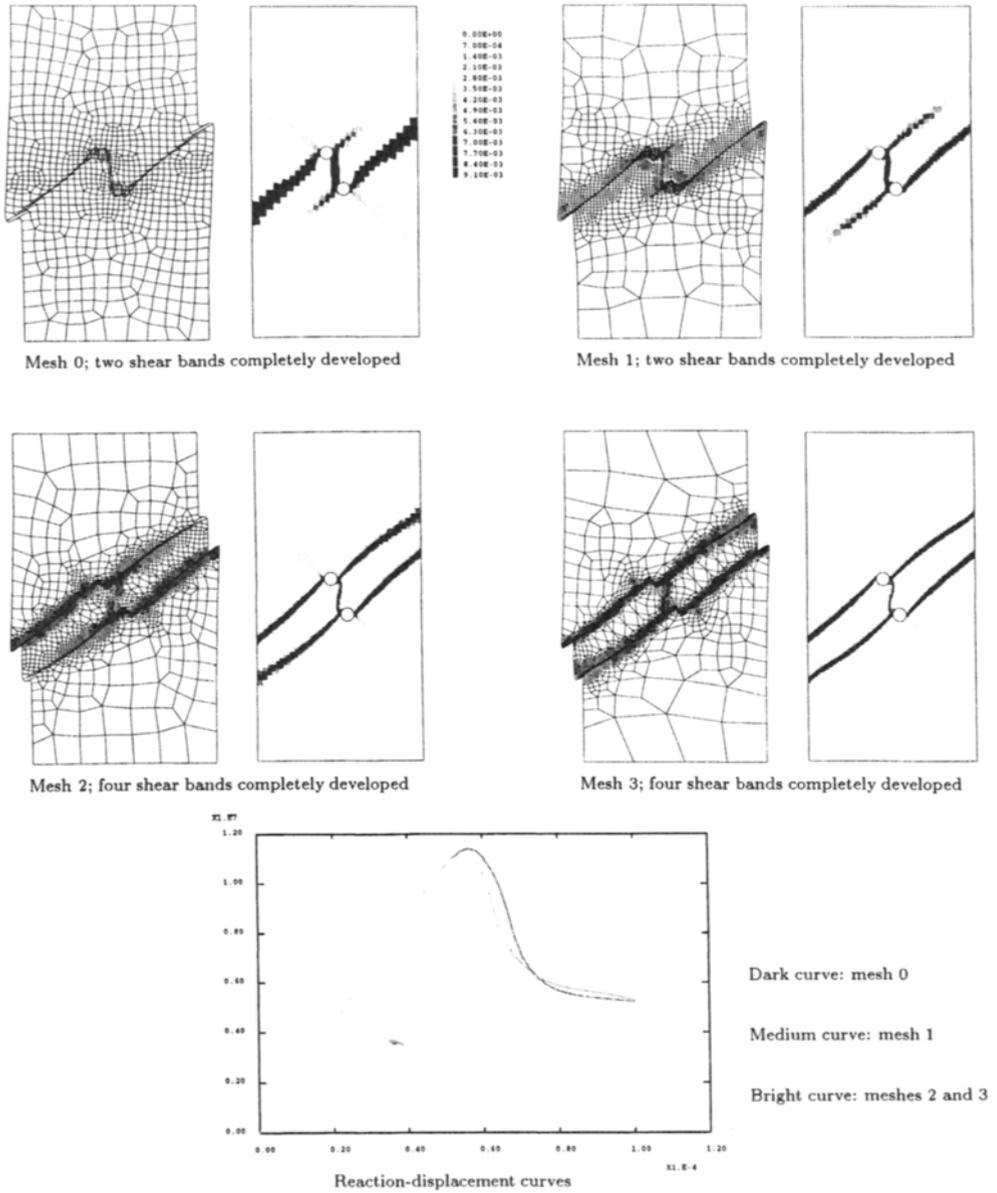


Figure 15. Numerical bifurcation in the first meshes: mesh deformation amplified 40 times and equivalent inelastic strain contours

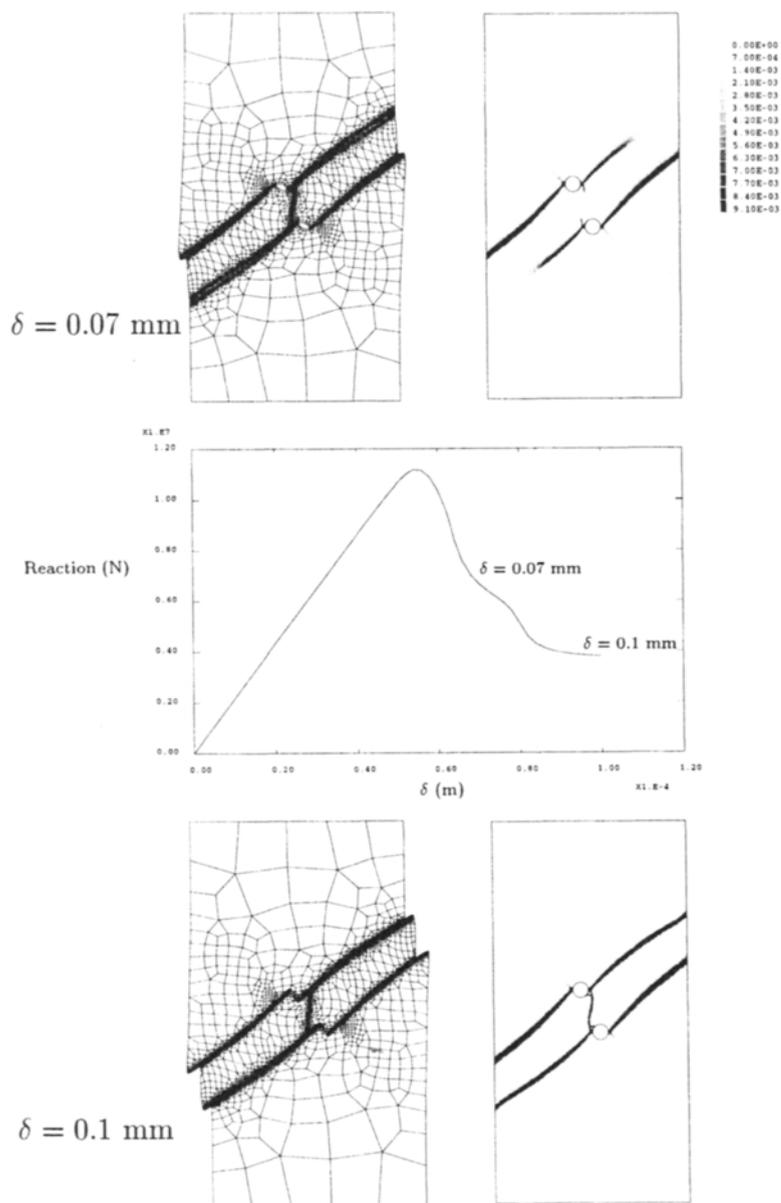


Figure 16. Two consecutive failure mechanisms for the final mesh: mesh deformation amplified 40 times and equivalent inelastic strain contours at two moments of the loading history

Figure 15 shows also how the force-displacement curves for meshes 0 and 1 are qualitatively different from meshes 2 to 5. In fact the shape of the force-displacement curve for meshes 2,3,4 and 5 are practically identical and have two inflections in the descending branch. In Figure 16 it is shown how the inflections correspond to the formation of a new failure mechanism. The solution given by the last mesh is obviously more accurate than the original one because the energy of deformation (area under the force-displacement curve) is lower. In fact, since the error is controlled in the energy norm, one can be sure that the actual curve, associated with the exact solution, is not too far from the obtained curve (the error in energy norm is less than 1.5% and, consequently, the difference of the area under the curves is less than 1.5%). The first meshes are not able to reproduce the behavior of the actual solution because the elements along the later band (which develops in a further stage of the loading process) are too large and, consequently, the discretization is too stiff. Then, the size of the elements in this zone does not allow the inception of softening. On the contrary, once the remeshing process introduces small enough elements along the second band, the second mechanism can be captured.

Thus, this example demonstrates that adaptivity based on error estimation is an essential tool for the determination of a priori unpredictable final solutions. Without this adaptive strategy, the initial mesh (mesh 0 in figure 14) and the resulting solution could be regarded as correct, and the second mechanism would not be detected.

4. CONCLUDING REMARKS

Adaptivity based on objective error estimation has been applied for softening mechanical problems. This computational strategy allows to deal with the two scales of the problem, providing a good mesh in order to capture both the geometry (macroscale) and the behavior in the localization zone (microscale). The generality of the proposed approach is shown by using two very different regularized constitutive models: Mazars nonlocal damage and Perzyna viscoplasticity. The numerical examples illustrate the effectiveness of the adaptive strategy, which is capable of detecting non-trivial features of the solution, such as singularities or unexpected complex collapse modes.

REFERENCES

1. Ainsworth, M. and J.T. Oden, (1993), "A unified approach to a posteriori error estimation using element residual methods", *Numerische Mathematik*, **65**, 23–50.
2. Babuška, I., R. Durán and R. Rodríguez (1992), "Analysis of the efficiency of an a posteriori error estimator for linear triangular finite elements", *SIAM (Society for Industrial and Applied Mathematics) Journal of Numerical Analysis*, **29**, 947–964.
3. Babuška, I. and R. Rodríguez, (1993) "The problem of the selection of an a posteriori error indicator based on smoothening techniques", *International Journal for Numerical Methods in Engineering*, **36**, 539–567.
4. Bank, R.E. and A. Weiser (1985), "Some a posteriori error estimators for elliptic partial differential equations", *Mathematics of Computation*, **44**, 283–301.
5. Díez, P. (1996), *Un nuevo estimador de error para el método de los elementos finitos*, Doctoral Thesis, Universitat Politècnica de Catalunya, Barcelona.

6. Díez, P., M. Arroyo and A. Huerta, "Adaptivity based on error estimation for viscoplastic softening materials", submitted to *Mechanics of Cohesive-Frictional Materials*.
7. Huerta, A., and G. Pijaudier-Cabot (1994), "Discretization influence on the regularization by two localization limiters," *Journal of Engineering Mechanics*, ASCE, **120**, 1198-1218.
8. Ladevèze, P., J.P. Pelle and Ph. Rougeot (1991), "Error estimation and mesh optimization for classical finite elements", *Engineering Computations*, **8**, 69-80.
9. Li, L.Y. and P. Bettess (1995), "Notes on mesh optimal criteria in adaptive finite element computations", *Communications in numerical methods in engineering*, **11**, 911-915.
10. Mazars, J., G. Pijaudier-Cabot and C. Saouridis (1991), "Size effect and continuous damage in cementitious materials", *International Journal of Fracture*, **51**, 159-173.
11. Oden, J.T., L. Demkowicz, W. Rachowicz and T.A. Westermann (1989) "Toward a universal h - p adaptive finite element strategy, part 2. A posteriori error estimation", *Computer Methods in Applied Mechanics and Engineering*, **77**, 113-180.
12. Ortiz, M., and J.J. Quigley IV (1991), "Adaptive mesh refinement in strain localization problems," *Computer Methods in Applied Mechanics and Engineering*, **90**, 781-804.
13. Pastor, M., J. Peraire and O.C. Zienkiewicz (1991), "Adaptive remeshing for shear band localisation problems," *Archive of Applied Mechanics*, **61**, 30-39.
14. Perzyna, P. (1966), "Fundamental problems in viscoplasticity", *Recent Advances in Applied Mechanics*, Academic Press, New York, **9**, 243-377.
15. Pijaudier-Cabot, G., and Z.P. Bazant (1987), "Nonlocal Damage Theory," *Journal of Engineering Mechanics*, ASCE, **113**, 1512-1533.
16. Pijaudier-Cabot, G., L. Bodé and A. Huerta (1995), "Arbitrary Lagrangian-Eulerian finite element analysis of strain localization in transient problems" *International Journal for Numerical Methods in Engineering*, **38**, 4171-4191.
17. Pijaudier-Cabot, G., and A. Huerta (1991), "Finite element analysis of bifurcation in nonlocal strain softening solids", *Computer Methods in Applied Mechanics and Engineering*, **90**, 905-919.
18. Saouridis, C. and J. Mazars (1988), "A multiscale approach to distributed damage and its usefulness for capturing structural size effect", CNRS-NSF Workshop *Strain localization and size effect due to cracking and damage*, E.N.S. Cachan, France.
19. Sarrate, J. (1996), *Modelización numérica de la interacción fluido-sólido rígido: desarrollo de algoritmos, generación de mallas y adaptividad*, Doctoral Thesis, Universitat Politècnica de Catalunya, Barcelona.
20. Yu, J., D. Peric and D.R.J. Owen (1992), "Adaptive finite element analysis of a strain localisation problem for the elastoplastic Cosserat continuum," *Proceedings of the 3rd International Conference on Computational Plasticity*, COMPLAS III, Eds. D.R.J. Owen et al., Pineridge Press, 551-566.
21. Zhu, J.Z., E. Hinton and O.C. Zienkiewicz (1993), "Mesh enrichment against mesh regeneration using quadrilateral elements", *Communications in numerical methods in engineering*, **9**, 547-554.

This Page Intentionally Left Blank

Aspects of adaptive strategies for large deformation problems at finite inelastic strains

D. Perić, M. Dutko and D.R.J. Owen

Department of Civil Engineering, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, United Kingdom

The main objective of this work is to discuss aspects of adaptive strategies for finite element simulation of large deformation problems in the presence of finite elastic-(visco)plastic strains. The need for a rigorous treatment of both theoretical and algorithmic issues is emphasised and the practically important aspects of automatic data generation and adaptive mesh refinement procedures are discussed. A set of numerical examples is provided to illustrate the effectiveness and robustness of the developed approach.

1 INTRODUCTION

The formal structure of adaptive finite element methods for linear elliptic problems is now well understood thus forming a solid foundation upon which effective and reliable techniques of error estimation and adaptive refinement may be established. Although certain issues still remain unresolved, it may be said that nowadays, adaptive strategies for linear problems can be routinely performed within finite element computations. On the contrary, although some advances have been recorded for certain classes of nonlinear problems (see e.g. Babuška *et al.* (1986) for some early contributions) only a limited amount of published work exists on *a posteriori* error estimates and adaptive approaches for history dependent nonlinear problems in solid mechanics. Notable exceptions are contributions by Ladevèze *et al.* (1986), Jin *et al.* (1989), Belytschko *et al.* (1989), Ortiz and Quigley (1991), Johnson and Hansbo (1992), Lee and Bathe (1994), Perić *et al.* (1994,1996), Gallimard *et al.* (1996) and Barthold *et al.* (1997).

On the practical side, since for a large number of industrially relevant solid mechanics problems, the optimal mesh configuration changes continually throughout the deformation process, the introduction of adaptive mesh refinement processes is crucial for the solution of large scale industrial problems.

In this work some practical aspects of adaptive strategies are discussed that are relevant for finite element simulation of large deformation problems in the presence of finite elastic-(visco)plastic strains. The need for a rigorous treatment of both theoretical

and algorithmic issues is emphasised and the practically important aspects of automatic data generation and adaptive mesh refinement procedures are discussed.

For problems involving non-linear behaviour, and plastic deformation in particular, issues related to the transfer of variables from the old mesh to the new mesh are crucially important for preserving the accuracy, robustness and convergence properties of the finite element solution. The problems in this area are especially acute when deformations are large, since the adaptive procedure must account for frequent substantial changes in geometry during the process.

Error indicators, used in this work, are of the post-processing type which employ the smoothing of the appropriate variables over the local mesh of finite elements. The standard stress error in the energy norm is reformulated in the generalised energy norm that is suitable for history dependent evolution processes. Also in view of the dissipative nature of inelastic problems of evolution, an error indicator based on the plastic dissipation functional and the rate of plastic work is considered.

The benefits of the above computational development will be illustrated by the solution of a range of industrially relevant problems.

2 COMPUTATIONAL MODELLING OF INELASTIC MATERIAL BEHAVIOUR

Some important aspects of computational treatment of inelastic solids at finite strains are briefly reviewed in this section. These relate both to the mathematical formulation and essential ingredients of the numerical integration scheme.

Due to the nature of large deformation problems at large inelastic strains, which are for many practical applications mostly pressure insensitive in the plastic region, the problem of the proper treatment of *incompressibility* of plastic flow becomes important. Aspects of technology of elements capable of the treatment of incompressibility are discussed in Section 2.2.

2.1 Finite strain plasticity

2.1.1 Multiplicative decomposition

The main hypothesis underlying the approach employed for finite strain elasto-plasticity is the multiplicative split of the deformation gradient into elastic and plastic parts, i.e.,

$$\mathbf{F} := \mathbf{F}^e \mathbf{F}^p \quad (1)$$

This assumption, firstly introduced by Lee (1969), admits the existence of a local unstressed *intermediate configuration*. Due to its suitability for the computational treatment of finite strain elasto-plasticity, the hypothesis of multiplicative decomposition is currently widely employed in the computational mechanics literature (Eterovic & Bathe, 1990; Simo, 1992; Perić *et al.*, 1992).

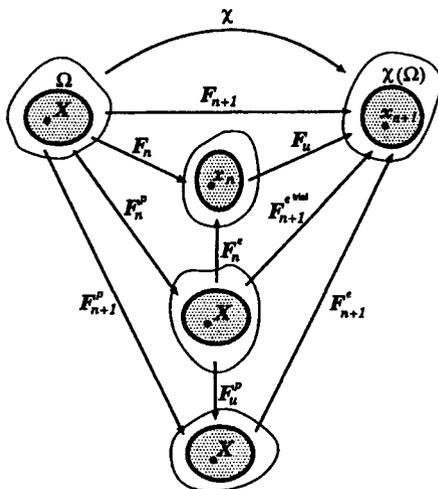


Figure 1 : Multiplicative decomposition of deformation gradient.

2.1.2 Stress update for finite strains

In the context of finite element analysis of path dependent problems, the load path is followed incrementally and a numerical approximation to the material constitutive law is needed to update stresses as well as the internal variables of the problem within each increment. Then, given the values of the variables $\{\sigma_n, \mathbf{q}_n\}$ at the beginning of a generic increment $[t_n, t_{n+1}]$ an algorithm for integration of the evolution equations is required to obtain the updated values $\{\sigma_{n+1}, \mathbf{q}_{n+1}\}$ at the end of the increment.

In the present work, the *backward Euler scheme* is employed for time integration and the *Newton-Raphson algorithm* is used in the solution of the resulting set of nonlinear equations. For convenience, the operations, on the kinematic level, of the algorithm for integration of the constitutive equations at finite strains are summarized in Box 1 while Box 2 describes stress update for small strains. We note that the small strain algorithm represented in Box 2 describes a simple procedure for von Mises elasto-plasticity.

2.2 Finite element technology

It is a well known fact that the performance of low order kinematically based finite elements is extremely poor near the incompressible limit. Large deformation problems under dominant plastic deformations and the assumption of isochoric plastic flow are included in this class of analysis. In such situations, spurious *locking* frequently occurs as a consequence of the inability of low order interpolation polynomials to adequately represent general volume preserving displacement fields. However, due to their simplicity, low order elements are often preferred in large scale computations and several formulations have been proposed to allow their use near the incompressible limit. Within the context of

Box 1. Algorithm for integration of constitutive equations

- (i) For given displacement \mathbf{u} and increment of displacement $\Delta\mathbf{u}$, evaluate total and incremental deformation gradient

$$\mathbf{F}_{n+1} := \mathbf{1} + \text{Grad}\chi_{n+1}[\mathbf{u}_{n+1}], \quad \mathbf{F}_u := \mathbf{1} + \text{Grad}\chi_{n+1}[\Delta\mathbf{u}_{n+1}]$$

- (ii) Evaluate elastic trial deformation gradient and elastic trial Fingers tensor

$$\mathbf{b}_{n+1}^{e \text{ trial}} := \mathbf{F}_u(\mathbf{b}_n^e)(\mathbf{F}_u)^T$$

- (iii) Compute eigenvalues (principal stretches $\lambda^{e \text{ trial}}$) and eigenvectors (rotation tensor $\mathbf{R}_{n+1}^{e \text{ trial}}$) of elastic trial Fingers tensor $\mathbf{b}_{n+1}^{e \text{ trial}}$

- (iv) Evaluate elastic trial left strain tensor and its logarithmic strain measure

$$\begin{aligned} \mathbf{V}_{n+1}^{e \text{ trial}} &:= (\mathbf{R}_{n+1}^{e \text{ trial}})^T (\lambda^{e \text{ trial}}) \mathbf{R}_{n+1}^{e \text{ trial}} \\ \mathbf{e}_{n+1}^{e \text{ trial}} &:= (\mathbf{R}_{n+1}^{e \text{ trial}})^T \ln[\lambda^{e \text{ trial}}] \mathbf{R}_{n+1}^{e \text{ trial}} \end{aligned}$$

- (v) Perform stress updating procedure for small strain

- (vi) Update Cauchy stress and internal variable

$$\begin{aligned} \boldsymbol{\sigma}_{n+1} &:= \det[\mathbf{F}_{n+1}]^{-1} \boldsymbol{\tau}_{n+1} \\ \mathbf{V}_{n+1}^e &:= \exp[\mathbf{e}_{n+1}^e], \quad \mathbf{b}_{n+1}^e := (\mathbf{V}_{n+1}^e)^2 \end{aligned}$$

geometrically linear theory, the class of assumed enhanced strain methods described by Simo and Rifai (1990), which incorporates popular procedures such as the classical incompatible modes formulation (Taylor *et al.*, 1976) and *B-bar* methods (Hughes, 1980), is well established and is employed with success in a number of existing commercial finite element codes. In the geometrically non-linear regime, however, the enforcement of incompressibility is substantially more demanding and the development of robust and efficient low order finite elements is by no means trivial. Different approaches have been introduced in the computational literature; the class of mixed variational methods developed by Simo *et al.* (1985), the mixed *u/p* formulation proposed by Sussman and Bathe (1987), the non-linear *B-bar* methodology adopted by Moran *et al.* (1990) and the family of enhanced elements of Simo and Armero (1992) are particularly important. However, due to the occurrence of pathological hourglassing patterns, a serious limitation on the applicability of enhanced elements has been identified by de Souza Neto *et al.* (1995) for the elasto-plastic finite strain case.

2.2.1 A low order element for large deformations of near-incompressible materials

The implementation of a simple 4-node quadrilateral and a simple 8-node hexahedron for finite strain analysis of nearly incompressible solids has been presented by de Souza Neto *et al.* (1996). Briefly, the elements are based on the concept of multiplicative devia-

Box 2. Stress updating procedure – Small Strains

(i) Elastic predictor

- Evaluate trial elastic stress

$$\boldsymbol{\tau}_{n+1}^{\text{trial}} := \mathbf{h} : \mathbf{e}_{n+1}^{\text{trial}}$$

- Check plastic consistency condition

$$\text{IF } \Phi^{\text{trial}} := J_2(\boldsymbol{\tau}_{n+1}^{\text{trial}}) - K(R_n) - \tau_y \leq 0 \text{ THEN}$$

$$\text{Set } (\cdot)_{n+1} = (\cdot)_{n+1}^{\text{trial}} \text{ and RETURN}$$

$$\text{ELSE go to (ii)}$$

 (ii) Plastic corrector (solve the system for $\boldsymbol{\tau}_{n+1}$, and $\Delta\gamma$)

$$\left\{ \begin{array}{l} J_2(\boldsymbol{\tau}_{n+1}) - K(R_n + \Delta\gamma) - \tau_y \\ \boldsymbol{\tau}_{n+1} - \mathbf{h} : (\mathbf{e}_{n+1}^{\text{trial}} - \Delta\gamma \partial_r \phi_{n+1}) \end{array} \right\} = \left\{ \begin{array}{l} 0 \\ 0 \end{array} \right\}$$

with

$$\partial_r \phi_{n+1} = \frac{3}{2} \frac{\text{dev} \boldsymbol{\tau}_{n+1}}{J_2(\boldsymbol{\tau}_{n+1})}$$

 (iii) Update elastic part of the strain \mathbf{e}^e and plastic consistency parameter R

$$\mathbf{e}_{n+1}^e := \mathbf{h}^{-1} : \boldsymbol{\tau}_{n+1}, \quad R_{n+1} := R_n + \Delta\gamma$$

(iv) RETURN

toric/volumetric split in conjunction with the replacement of the compatible deformation gradient field with an assumed modified counterpart. The resulting formulation can be used regardless of the material model adopted. In addition, the strain driven format of the algorithms for integration of inelastic constitutive equations of the purely kinematic formulation is maintained.

The key idea underlying this formulation is the use of an assumed *modified* deformation gradient to compute the stresses. Firstly, the volumetric/deviatoric split is applied to the deformation gradient \mathbf{F} at the gauss point of interest as well as to the deformation gradient \mathbf{F}_0 that results from the conventional displacement interpolation at the *centroid* of the element. The *modified deformation gradient*, $\bar{\mathbf{F}}$, is then defined as the composition of the deviatoric component of \mathbf{F} with the volumetric component of \mathbf{F}_0 , i.e.,

$$\bar{\mathbf{F}} := \mathbf{F}_d(\mathbf{F}_0)_v = \left(\frac{\det[\mathbf{F}_0]}{\det[\mathbf{F}_1]} \right)^{1/3} \mathbf{F} \quad (2)$$

Having defined the modified deformation gradient, the proposed elements are obtained by replacing \mathbf{F} with $\bar{\mathbf{F}}$ to compute the Cauchy stress at each gauss point, i.e. $\boldsymbol{\sigma} = \hat{\boldsymbol{\sigma}}(\bar{\mathbf{F}})$.

3 ADAPTIVE SOLUTION UPDATE

The history dependent nature of the process necessitates transfer of all relevant problem variables from the old mesh to the new one, as successive remeshing is applied during

the process simulation. As the mesh is adapted, with respect to an appropriate error estimator, the solution procedure, in general, cannot be re-computed from the initial state, but has to be continued from the previously computed state. Hence, some suitable means for transferring the state variables between meshes, or *transfer operators*, needs to be defined. A class of transfer operators for large strain elasto-plastic problems occurring in large deformation problems is defined in Section 3.

3.1 Error indicators

The extension of the error estimation based on the *plastic dissipation functional* and the rate of *plastic work* described by Perić *et al.* (Perić *et al.*, 1994) for large strain elasto-plasticity has been found to be an appropriate choice to the adaptive solution of large deformation problems at finite elasto-plastic strains.

3.2 Mesh regeneration

An unstructured meshing approach is used for the mesh generation and subsequent mesh adaptation. The algorithm employed is based on the *Delaunay triangulation* technique which is particularly suited to local mesh regeneration. An extension of the Delaunay scheme to quadrilateral elements is also available and creates the possibility of employment of the low order elements described in Section 2.2.

3.3 Transfer operations for evolving meshes

After creating a new mesh, the transfer of displacement and history-dependent variables from the old mesh to a new one is required. Several important aspects of the transfer operation have to be addressed ((Perić *et al.*, 1996), (Lee & Bathe, 1994)):

- (i) consistency with the constitutive equations,
- (ii) requirement of equilibrium (which is fundamental for implicit FE simulation),
- (iii) compatibility of the history-dependent internal variables transfer with the displacement field on the new mesh,
- (iv) compatibility with evolving boundary conditions,
- (v) minimisation of the numerical diffusion of transferred state fields.

To describe the transfer operation, let us define a state array ${}^h\mathbf{A}_n = ({}^h\mathbf{u}_n, {}^h\mathbf{e}_n, {}^h\boldsymbol{\sigma}_n, {}^h\mathbf{q}_n)$ where ${}^h\mathbf{u}_n, {}^h\mathbf{e}_n, {}^h\boldsymbol{\sigma}_n, {}^h\mathbf{q}_n$ denote values of the displacement, strain tensor, stress tensor and a vector of internal variables at time t_n for the mesh h . Assume, furthermore, that the estimated error of the solution ${}^h\mathbf{A}_n$ respects the prescribed criteria, while these are violated by the solution ${}^h\mathbf{A}_{n+1}$. In this case a new mesh $h+1$ is generated and a new solution ${}^{h+1}\mathbf{A}_{n+1}$ needs to be computed. As the backward Euler scheme is adopted the internal variables ${}^{h+1}\mathbf{q}_n$ for a new mesh $h+1$ at time t_n need to be evaluated. In this way the state ${}^{h+1}\tilde{\mathbf{A}}_n = (\bullet, \bullet, \bullet, {}^{h+1}\mathbf{q}_n)$ is constructed, where a symbol $\tilde{}$ is used to denote a reduced

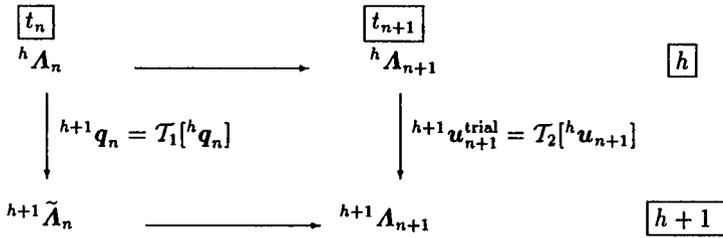


Figure 2 : Transfer operator diagram

state array. It should be noted that this state characterises the history of the material and, in the case of a fully implicit scheme, provides sufficient information for computation of a new solution ${}^{h+1}\Lambda_{n+1}$.

Conceptually, Figure 2 summarises a typical transfer operation that includes both, the mapping of the internal variables and mapping of the displacement field. The implementation of the given general transfer operation is performed for the case of evolving finite element meshes composed of constant strain triangles in the following fashion:

• Mapping of the internal variables - Transfer operator T_1

Algorithm of the transfer operator T_1 comprises the following steps:

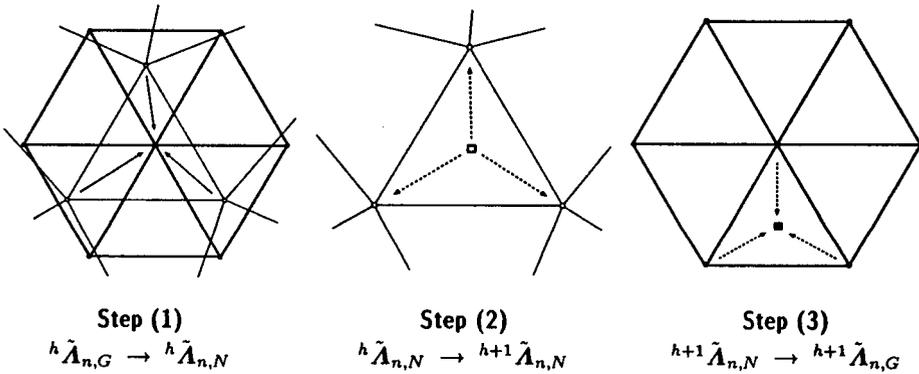


Figure 3 : A procedure illustrating the implementation of the transfer operator ${}^{h+1}q_{n,G} = T_1[{}^h q_{n,G}]$ for finite element meshes composed of three noded triangles.

- (1) The Gauss point components of the old mesh ${}^h q_{n,G}$ are projected to nodes ${}^h q_{n,N}$ using the finite element shape functions. The nodal point averages are then performed resulting in ${}^h q_{n,N}^*$.
- (2) In the second step the nodal components of the state variables ${}^h q_{n,N}^*$ will be transferred from the old mesh h to a new mesh $h + 1$ resulting in ${}^{h+1} q_{n,N}^*$. This step of the transfer operation is the most complex one and can be subdivided as follows:

- *Construction of the background triangular mesh*

For each node A of the new mesh $h + 1$ with known coordinates ${}^{h+1}\mathbf{x}_{n,A}$ the so-called background element is found in the old mesh h , i.e. element ${}^h\Omega^{(e)}$ for which ${}^{h+1}\mathbf{x}_{n,A} \in {}^h\Omega^{(e)}$.

- *Evaluation of the local coordinates*

Knowing the coordinates of all nodal points for both old and new meshes, the local coordinates $({}^hr_A, {}^hs_A)$ within the background element corresponding to the position of nodal point A in the new mesh can be found by solving the following equation

$${}^{h+1}\mathbf{x}_{n,A} = \sum_{b=1}^3 {}^hN_b({}^hr_A, {}^hs_A) {}^h\mathbf{x}_{n,b} \quad (3)$$

where hN_b represent shape functions of the element ${}^h\Omega^{(e)}$. Since three noded elements are used for the background mesh, local coordinates for each node of the new mesh $h + 1$ can be obtained by resolving the linear system (3).

- *Transfer of the nodal values*

By using the shape functions ${}^hN_b({}^hr_A, {}^hs_A)$ the state variables ${}^h\tilde{\mathbf{A}}_{n,B} = {}^h\mathbf{q}_{n,B}^*$ are mapped from the nodes B of the old mesh h to the nodes A of the new mesh $h + 1$. This mapping may be expressed as

$${}^{h+1}\tilde{\mathbf{A}}_{n,A} = \sum_{b=1}^3 {}^hN_b({}^hr_A, {}^hs_A) {}^h\tilde{\mathbf{A}}_{n,b} \quad (4)$$

(3) The state variables at the Gauss points of the new mesh ${}^{h+1}\mathbf{q}_{n,G}^*$ can be easily obtained by employing the shape functions of the element ${}^{h+1}\Omega^{(e)}$, i.e.

$${}^{h+1}\tilde{\mathbf{A}}_{n,A} = \sum_{a=1}^3 {}^{h+1}N_a(r_G, s_G) {}^{h+1}\tilde{\mathbf{A}}_{n,a} \quad (5)$$

where (r_G, s_G) are the Gauss point coordinates.

- *Mapping of the displacements - Transfer operator \mathcal{T}_2*

Since the displacement field over new mesh $h + 1$ is fully prescribed by the nodal values ${}^{h+1}\mathbf{u}_{n,a}$ and element shape functions of the new mesh ${}^{h+1}N_a({}^{h+1}r_a, {}^{h+1}s_a)$ defined for each element ${}^{h+1}\Omega^{(e)}$ of the new mesh, the task of transferring displacements (i.e. transfer operator \mathcal{T}_2) is performed by repeating the step *Mapping of the nodal values*, in the procedure describing the transfer operator \mathcal{T}_1 .

4 NUMERICAL VERIFICATION

To illustrate the adaptive strategy, its application to examples of some industrial large deformation problems exhibiting finite elasto-plastic strains are presented in this section.

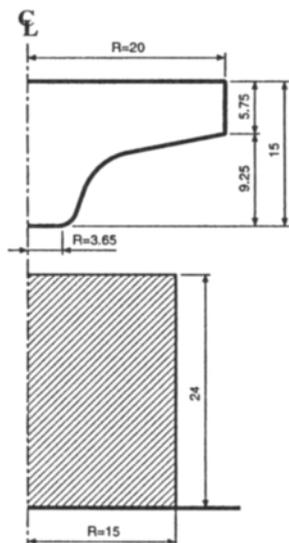


Figure 4 : Axisymmetric piercing.
Geometry.

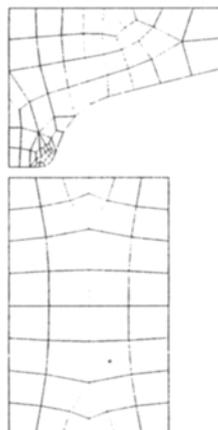


Figure 5 : Axisymmetric piercing.
Initial mesh.

EXAMPLE 4.1 *Axisymmetric piercing.* The finite element simulation of the axisymmetric piercing of a cylindrical workpiece is presented. The geometry of the problem is shown in Figure 4 and the initial mesh in Figure 5. The workpiece is assumed to be made of an elastic-plastic material with Young's modulus $E = 210\text{GPa}$, Poisson ratio $\nu = 0.3$, yield stress $\sigma = 100\text{MPa}$ and linear hardening with hardening modulus $H = 900\text{MPa}$, while the punch is assumed to be rigid. Frictional contact between workpiece and tool is defined by a Coulomb law with coefficient of friction $\mu = 0.1$.

In analysis an error indicator based on the rate of plastic work is used. The initial mesh consists of 101 quadrilateral elements and the final mesh contains 426 elements. Convergence of the finite element solution is established on the basis of the standard Euclidean norm of the out-of-balance forces with a tolerance of 10^{-3} . No difficulties related to the convergence have been observed during the simulation despite frequent remeshings.

Distribution of effective plastic strain on deformed meshes at various stages of the process is shown in Figure 6. The deformed meshes show no hourglassing patterns, which is in agreement with analyses of a similar class of problems carried out in (Souza Neto *et al.*, 1996).

EXAMPLE 4.2 *Plane strain spike forming.* The finite element simulation of plane strain spike forming is presented. The geometry of the problem is shown in Figure 7 and the initial mesh in Figure 8. The workpiece is assumed to be made of an elastic-plastic material with Young's modulus $E = 125\text{GPa}$, Poisson ratio $\nu = 0.3$ and yield stress $\sigma = 40\text{MPa}$ with an exponential hardening law. In the present FE analysis the punch is assumed to be rigid. A Coulomb law with a coefficient of friction $\mu = 0.2$ defines the

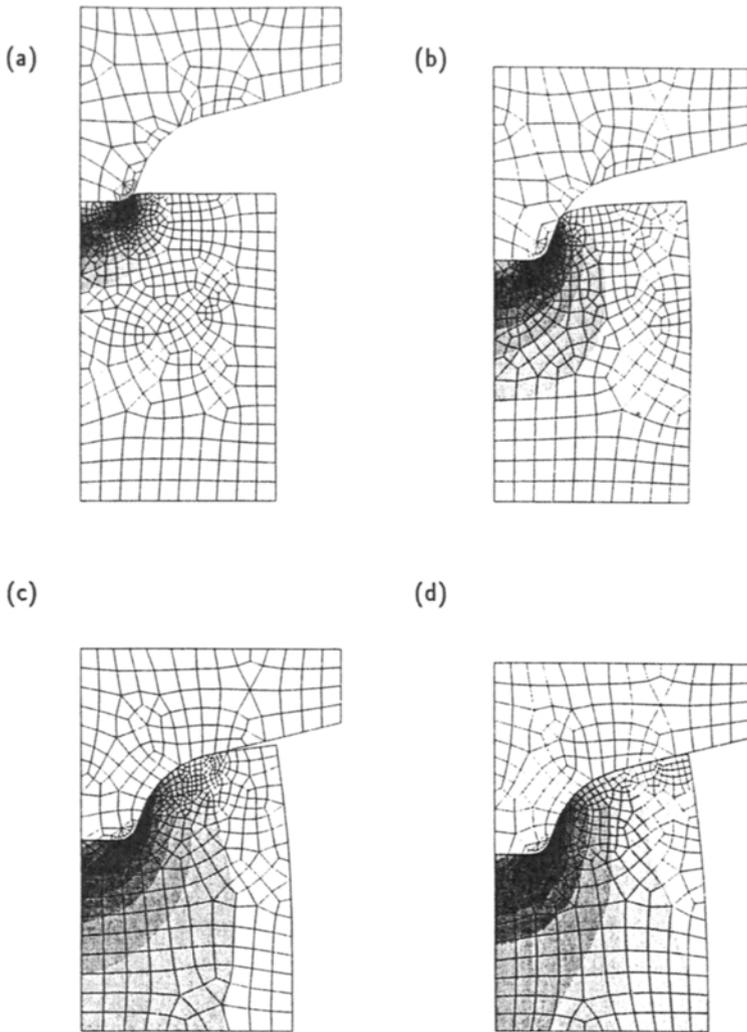


Figure 6 : Axisymmetric piercing. Evolution of effective plastic strain. (a) $U = 0.62$, (b) $U = 5.66$, (c) $U = 9.83$, (d) $U = 11.00$.

frictional contact between workpiece and tool. A convergence tolerance of 10^{-3} of the finite element solution is established on the basis of the standard Euclidean norm of the

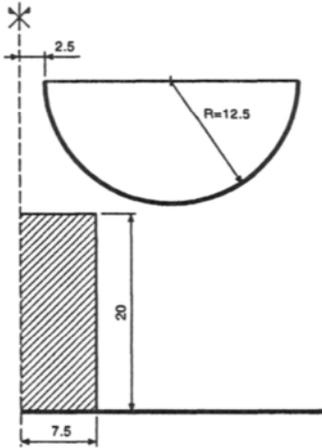


Figure 7 : Plane strain spike forming. Geometry.

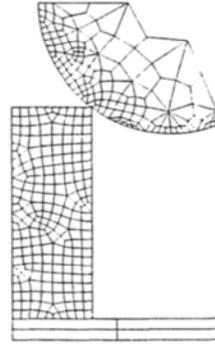


Figure 8 : Plane strain spike forming. Initial mesh.

out-of-balance forces.

In total 12 mesh adaptations were performed, which involve complete new mesh definitions related to the error indicator based on the rate of plastic work. The number of elements varies in subsequent meshes from initially 302 quadrilateral elements to finally 353 elements. Deformed meshes obtained after adaptive remeshing are shown in Figure 9. Similarly to the previous analysis, meshes at various stages show no hourglassing patterns despite very large deformation.

EXAMPLE 4.3 *Impact of an elasto-plastic projectile on a ceramic block.* Geometry and loading for this problem are depicted in Figure 10(a). The projectile is assumed to be made of an elastic-plastic copper material with Young's modulus $E = 117[\text{GPa}]$, Poisson ratio $\nu = 0.35$ and yield stress $\sigma = 400[\text{MPa}]$ with linear isotropic hardening modulus $H = 100\text{MPa}$. The target is made of an elastic-plastic brittle material with Young's modulus $E = 211.69[\text{GPa}]$, Poisson ratio $\nu = 0.286$, yield stress $\sigma = 958[\text{MPa}]$ and fracture energy $D_f = 0.1[\text{MPa}]$.

A Coulomb law with coefficient of friction $\mu = 0.9$ defines frictional contact between the projectile and target and the initial velocity of the projectile is $v_o = 1800[\text{m s}^{-1}]$. In analysis an error indicator based on L^2 stress norm is used. In this case, analysis is performed by a transient dynamic explicit time integration approach in view of the high loading rates.

Continuous change in geometry of the projectile near the contact area necessitates frequent remeshings during the process. In addition, mesh adaptation is required to provide for accurate description of multiple fracturing of the block. It should be emphasised that the present discrete element model allows for crack propagation both, along element boundaries and through the finite elements. The deformation of the projectile, the adapted finite element meshes and fracturing process of the block are depicted in Figure 10(b)-(c).

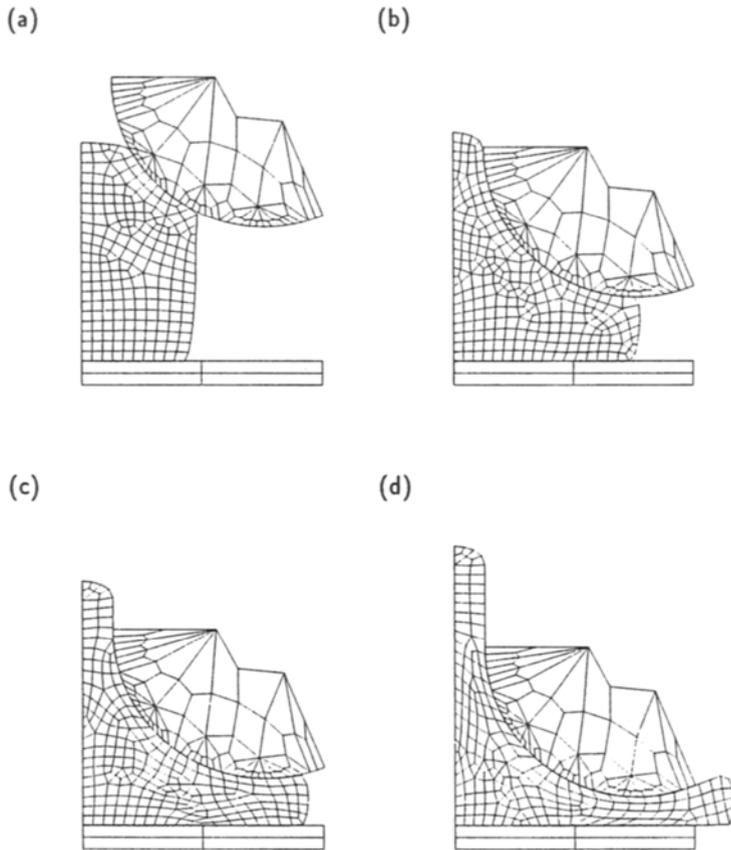


Figure 9 : Plane strain spike forming. Evolution of deformed mesh. (a) $U = 6.3$, (b) $U = 12.1$, (c) $U = 13.5$, (d) $U = 15.0$.

5 CONCLUSIONS

Some recent advances in the adaptive finite element analysis of large deformation problems of elasto-(visco)plastic solids at finite strains have been described in this work. It has been indicated that, for this class of problems, a successful and practically applicable adaptive strategy depends on both the theoretical understanding of inelastic material behaviour under finite strains and the associated numerical implementation.

It is emphasised that, although adaptive strategies are, at present, routinely performed for linear elliptic problems, their extension to nonlinear elliptic problems – in

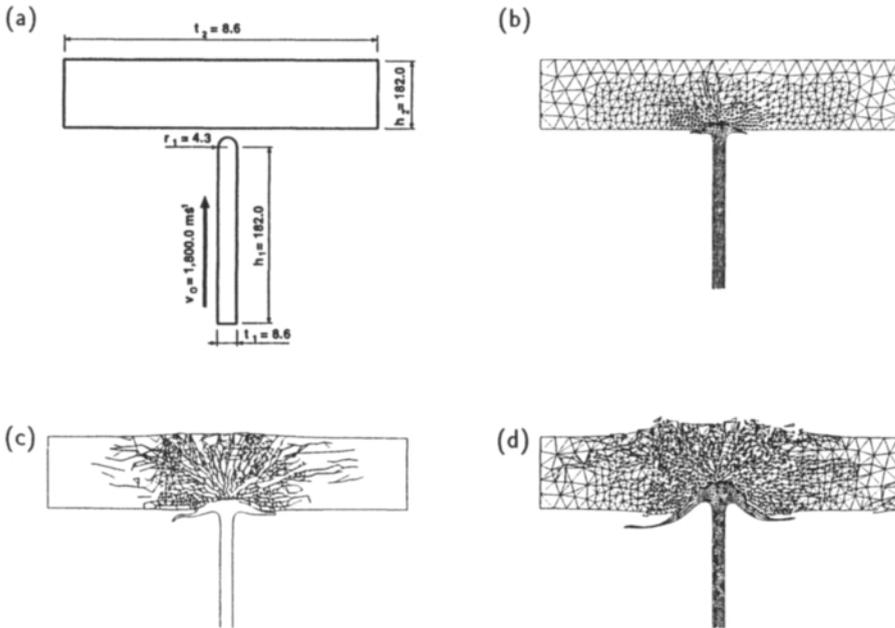


Figure 10 : Impact of an elasto-plastic projectile on a ceramic block: (a) Geometry and loading conditions. Evolution of deformed mesh at various time instants (b) $t=0.013$ [ms], (c) $t=0.026$ [ms], (d) $t=0.050$ [ms].

particular to large deformation problems of inelastic solids at finite strains – is by no means trivial. Apart from the issues briefly mentioned in this paper, several important aspects of adaptive strategies related to nonlinear industrial applications which need further attention include the introduction of various types of error estimators and their comparative analysis, and the consideration of alternative data transfer strategies.

References

- BABUŠKA, I., ZIENKIEWICZ, O.C., GAGO, J., & DE OLIVEIRA, E.R. (eds). 1986. *Accuracy Estimates and Adaptive Refinements in Finite Element Computations*. New York: John Wiley.
- BARTHOLD, F.-J. SCHMIDT, M., & STEIN, E. 1997. Error Estimation Mesh Adaptivity for Elasto-Plastic Deformations. *Pages 597–602 of: OWEN, D.R.J., OÑATE, E., & HINTON, E. (eds), Computational Plasticity V: Fundamentals and Applications*. CIMNE.
- BELYTSCHKO, T., WONG, B.L., & PLAKACZ, E.J. 1989. Fission-Fusion Adaptivity in Finite Elements for Nonlinear Dynamics of Shells. *Comput. Struct.*, **33**, 1307–1323.

- ETEROVIC, A.L., & BATHE, K.-J. 1990. A Hyperelastic Based Large Strain Elasto-Plastic Constitutive Formulation with Combined Isotropic-Kinematic Hardening Using the Logarithmic Stress and Strain Measures. *Int. J. Num. Meth. Engng.*, **30**, 1099–1114.
- GALLIMARD, L., LADEVÈZE, P., & PELLE, J.-P. 1996. Error Estimation and Adaptivity in Elastoplasticity. *Int. J. Num. Meth. Engng.*, **39**, 189–217.
- HUGHES, T.J.R. 1980. Generalization of Selective Integration Procedures to Anisotropic and Nonlinear Media. *Int. J. Num. Meth. Engng.*, **15**, 1413–1418.
- JIN, H., BERNSPANG, L., LARSSON, R., & WIBERG, N.E. 1989. Mesh Generation and Mesh Refinement Procedures for Computational Plasticity. *Pages 347–357 of: OWEN, D.R.J., OÑATE, E., & HINTON, E. (eds), Proceedings of the Second International Conference on Computational Plasticity: Models, Software and Applications – Barcelona, September 1989.* Swansea: Pineridge Press.
- JOHNSON, C., & HANSBO, P. 1992. Adaptive Finite Element Methods in Computational Mechanics. *Comp. Meth. Appl. Mech. Engng.*, **101**, 143–181.
- LADEVÈZE, P., COFIGNAL, G., & PELLE, J.P. 1986. Accuracy of Elastoplastic and Dynamic Analysis. *Pages 181–203 of: BABUŠKA, I., ZIENKIEWICZ, O.C., GAGO, J., & DE OLIVEIRA, E.R. (eds), Accuracy Estimates and Adaptive Refinements in Finite Element Computations.* New York: John Wiley.
- LEE, E.H. 1969. Elastic-Plastic Deformation at Finite Strains. *J. Appl. Mech.*, **36**, 1–6.
- LEE, N.-S., & BATHE, K.-J. 1994. Error Indicators and Adaptive Remeshing in Large Deformation Finite Element Analysis. *Fin. Elem. Anal. Design*, **16**, 99–139.
- MORAN, B., ORTIZ, M., & SHIH, F. 1990. Formulation of Implicit Finite Element Methods for Multiplicative Finite Deformation Plasticity. *Int. J. Num. Meth. Engng.*, **29**, 483–514.
- ORTIZ, M., & QUIGLEY IV, J.J. 1991. Adaptive Mesh Refinement in Strain Localization Problems. *Comp. Meth. Appl. Mech. Engng.*, **90**, 781–804.
- PERIĆ, D., OWEN, D.R.J., & HONNOR, M.E. 1992. A Model for Finite Strain Elasto-Plasticity Based on Logarithmic Strains: Computational Issues. *Comp. Meth. Appl. Mech. Engng.*, **94**, 35–61.
- PERIĆ, D., YU, J., & OWEN, D.R.J. 1994. On Error Estimates and Adaptivity in Elastoplastic Solids: Application to the Numerical Simulation of Strain Localization in Classical and Cosserat Continua. *Int. J. Num. Meth. Engng.*, **37**, 1351–1379.
- PERIĆ, D., HOCHARD, CH., DUTKO, M., & OWEN, D.R.J. 1996. Transfer Operators for Evolving Meshes in Small Strain Elasto-plasticity. *Comp. Meth. Appl. Mech. Engng.*, **137**, 331–344.
- SIMO, J.C. 1992. Algorithms for Static and Dynamic Multiplicative Plasticity that Preserve the Classical Return Mapping Schemes of the Infinitesimal Theory. *Comp. Meth. Appl. Mech. Engng.*, **99**, 61–112.
- SIMO, J.C., & ARMERO, F. 1992. Geometrically Non-Linear Enhanced Strain Mixed Methods and the Method of Incompatible Modes. *Int. J. Num. Meth. Engng.*, **33**, 1413–1449.

- SIMO, J.C., & RIFAI, S. 1990. A Class of Mixed Assumed Strain Methods and the Method of Incompatible Modes. *Int. J. Num. Meth. Engng.*, **29**, 1595-1638.
- SIMO, J.C., TAYLOR, R.L., & PISTER, K.S. 1985. Variational and Projection Methods for the Volume Constraint in Finite Deformation Elasto-Plasticity. *Comp. Meth. Appl. Mech. Engng.*, **51**, 177-208.
- SOUZA NETO, E.A. DE, PERIĆ, D., HUANG, G.C., & OWEN, D.R.J. 1995. Remarks on the Stability of Enhanced Strain Elements in Finite Elasticity and Elastoplasticity. *Comm. Num. Meth. Engng.*, **11**, 951-961.
- SOUZA NETO, E.A. DE, PERIĆ, D., DUTKO, M., & OWEN, D.R.J. 1996. Design of Simple Low Order Finite Elements for Large Strain Analysis of Nearly Incompressible Solids. *Int. J. Solids Struct.*, **33**, 3277-3296.
- SUSSMAN, T., & BATHE, K.-J. 1987. A Finite Element Formulation for Nonlinear Incompressible Elastic and Inelastic Analysis. *Comp. Struct.*, **26**, 357-409.
- TAYLOR, R.L., BERESFORD, P.J., & WILSON, E.L. 1976. A Non-Conforming Element for Stress Analysis. *Int. J. Num. Meth. Engng.*, **10**, 1211-1219.

This Page Intentionally Left Blank

Adaptive solutions in industrial forming process simulation

T. Coupez, L. Fourment and J.L. Chenot

Ecole des Mines de Paris, CEMEF, u.a. CNRS 1374, B.P. 207,
06904 Sophia Antipolis Cedex, France

Abstract

Adaptive solution in forming process simulation are presented, covering the theoretical aspects of the a posteriori error estimation as well as the practical aspects of the dynamic mesh optimization algorithm. A simple error estimate for time-dependent nonlinear problems is described with a special attention to the 2D forging application. Example of 3D automatic remeshing with a local size mesh enforcement is presented for a polymer cutting application. A new level of optimization allowed by anisotropic mesh techniques is introduced in the context of 3D extrusion, with primarily results of error estimation with anisotropic meshes. A 3D colliding jets example based on a fixed mesh approach is analyzed in term of flow front position error, both a priori and a posteriori, in the L^2 norm.

1. INTRODUCTION

The numerical simulation of forming processes has been early concerned with the need of automatic remeshing, leading to control effectively the mesh size locally. In fact, most of the forming applications presented below need to deal with meshing or remeshing technique just in order to carry out a calculation. It is well admitted that forging simulation cannot be done without remeshing. It is unavoidable when using a Lagrangian method. In this case the mesh is deformed along with the material and the mesh distortion clearly limits the level of computable deformation. Although enhanced formulations makes it possible to delay the mesh distortion, the mesh topology becomes poorly adapted to the shape of the studied domain. In order to overtake these limitations, automatic remeshing procedure have been developed both in 2D and in 3D. This paper is based on the presentation of different results of simulation.

2. FINITE ELEMENT SOLVER AND ADAPTIVE ANALYSIS

Most of the adaptive schemes are based on the convergence properties and the associated a priori error estimate of the finite element method. It assumes that a mesh independent solution exists and that the numerical solutions approximate it. More precisely, it is expected that the error decreases along with the mesh size parameter. In this context, the behavior of finite element method is well known for linear elasticity problems and it has been extended to incompressible materials both in fluid and solid applications through the mixed finite element theory [5].

Both in metal and polymer forming, the viscoplastic behavior of the material can be modeled by a flow formulation, the viscosity being a function of the second invariant of the strain rate tensor $\varepsilon(v)$:

$$\eta(|\varepsilon(v)|) = K(\sqrt{2}|\varepsilon(v)|)^{m-1}. \quad (1)$$

On a time depending domain Ω , inertia effects and volume forces being negligible, the velocity v and pressure p fields are solution of the following problem :

$$\left\{ \begin{array}{l} \nabla \cdot (2\eta(|\varepsilon(v)|)\varepsilon(v)) - \nabla p = 0 \\ \nabla \cdot v = 0 \quad \text{in } \Omega \\ \text{on } \partial\Omega \cap \partial O \quad \left\{ \begin{array}{l} (v - v^O) \cdot n > 0 \quad \text{and } \tau = 0 \\ \text{or } (v - v^O) \cdot n = 0, \quad \text{and } \tau = \alpha(|v - v^O|)(v - v^O), \end{array} \right. \end{array} \right. \quad (2)$$

The inequality characterizes the unilateral contact boundary condition which plays a crucial role in the applications presented here. v^O denotes the velocity of the contacting rigid bodies (the forming tools) described by the domain O , n being the outside normal of its boundary ∂O . The relationship which gives the tangent shear stress τ at the interface matter/tool as a function of the relative tangent velocity follows a power friction law. The viscoplasticity coefficient m usually ranges between 0 and 1. It is often around 0.1 in hot metal forming applications..

The solvers behind the applications presented in this paper are designed to be used with triangle elements in 2D and tetrahedral elements in 3D, making it possible to deal with general meshing techniques detailed after. In 2D the mixed finite element will be based on the Crouseix-Raviart element ($P2 + /P1$) and on the $P2/P0$ element [16]. In 3D, we use a first order element entering in the mini-element family [1]. The bubble is constructed with a pyramidal function which is condensed at the element level, without taking into account the nonlinear part of the behavior law [9]. This element enters in the equivalent class of stable/stabilized mixed formulation [20, 15, 27]. An important advantage of this formulation is to enable a fully parallel iterative solution [14].

Despite of the unilateral contact condition, this flow formulation can be analyzed in term of a generalized Stokes problem. From the numerical analysis standpoint, the convergence of the finite element method for the viscoplastic problem (quasi-Newtonian flow) has been analyzed in term of the a priori estimate in [3], using the $W^{1,m+1}$ norm, m being the viscoplastic power index. In term of the a posteriori estimate, the residual technique

introduced for the Stokes equation [30] has been extended to the quasi-Newtonian flows [2]. In engineering, the error analysis is often based on an energy norm point of view. In this context, the simple Zienkiewicz-Zhu adaptive procedure [25] based on the smoothness of the stress tensor has been adopted in [18, 17, 19] from which several 2D examples are presented in this paper. From the mechanical point of view, a constitutive relation error analysis has been introduced by Ladeveze and specifically applied to elastoplastic materials [24].

3. ADAPTIVE MESH GENERATION TECHNIQUES

3.1. Mesh generation techniques

The 2D code Forge2 is equipped with a remesher based on the classical Delaunay triangulation technique [21]. It has been extended in order to work under the constraint of a mesh size map for 2D forging applications [18]. This 2D remeshing procedure must be understood as a rebuilding process. The mesh and its boundary are entirely built at each remeshing stage.

The three-dimensional meshing routine is based on a different idea, the remeshing being an improvement process [10]. The advantage of this technique is evident in the context of a moving mesh method, because it makes it possible to progressively change the mesh topology rather to build a complete new one. In fact, it has been especially designed to solve the remeshing stage in forming calculations [13, 10]. In the forging code Forge3, the remeshing is required both to avoid the element distortions and to dynamically adapt the mesh to the contact between the moving dies and the material. Another advantage of this approach, is its ability to work in parallel without any modification. The complete remeshing procedure can be applied on each subdomain of a partition, independently running on each processor of a parallel platform [11], as shown on Figure 3.1.

3.2. Adaptive remeshing by a local optimization principle

In 2D and in 3D, the meshing topics can be revisited through a local optimization principle [12]. The general mesh improvement method [7] is the basis of the mesher used in the presented 3D applications. It is an extension the mesh improvement technique described in [29]. This method allows to deeply improve a mesh and includes mesh refinement, mesh coarsening and node repositioning [8]. Moreover, it efficient enough to be used as a mesh generation technique, enables to locally control the mesh size in an isotropic or anisotropic way (Figure 3.2).

The improvement process is controlled by a minimal volume principle which guarantees the validity of the resulting mesh as plotted on Figure 3.2. Indeed, on this Figure, an initial mesh topology is obtained by connecting one particular node to every boundary nodes. The final mesh (without internal node in this case) is derived from this initial guess by successive local improvements. The change are chosen to minimize the surface covered by the triangles. The improvement process is also controlled by the mesh quality criterion.

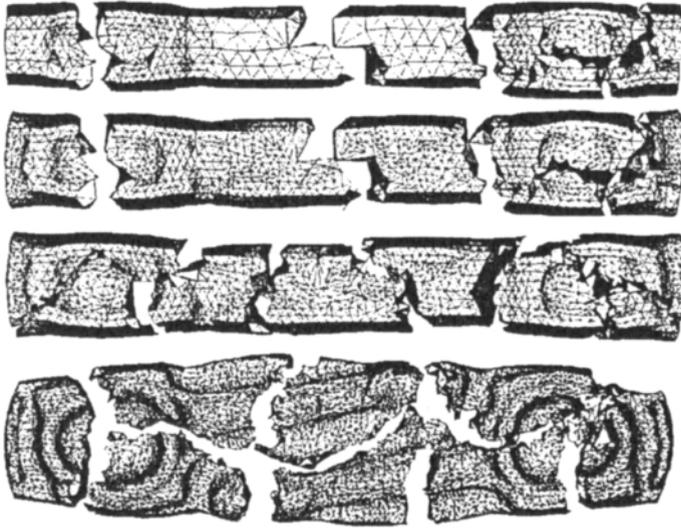


Figure 1: Parallel adaptive remeshing by independent remeshing with interface constraining and repartitioning

The shape of a simplex (triangle or tetrahedron) can be evaluated by the a -dimensional ratio between its volume and diameter :

$$c(T) = \frac{v(T)}{h(T)^n},$$

where $n = 2, 3$ is the dimension, $v(T)$ denotes the volume of the element T and $h(T)$ its diameter. An objective function, for instance the averaged value of the above shape factor can be locally optimized:

$$\phi(\mathcal{T}) = \left(\sum_{T \in \mathcal{T}} C(T) \right) / |\mathcal{T}|,$$

where \mathcal{T} is the set of simplex and $|\mathcal{T}|$ its number.

The local mesh size control is performed just by changing the shape factor :

$$c(T) = \frac{\min(\hat{v}(x), v(T))}{h(T)^n},$$

where $\hat{v}(x)$ is the required element volume at position x . $\hat{v}(x)$ can be derived from $h(x)$, the required mesh size at point coordinate x . Finally, the extension to anisotropic meshing is obtained by the calculation of the shape factor in the local metric (a volume and a length).

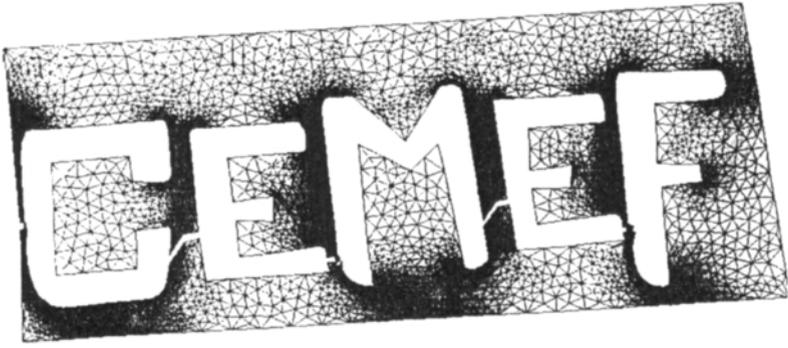


Figure 2: Automatic mesh generation example

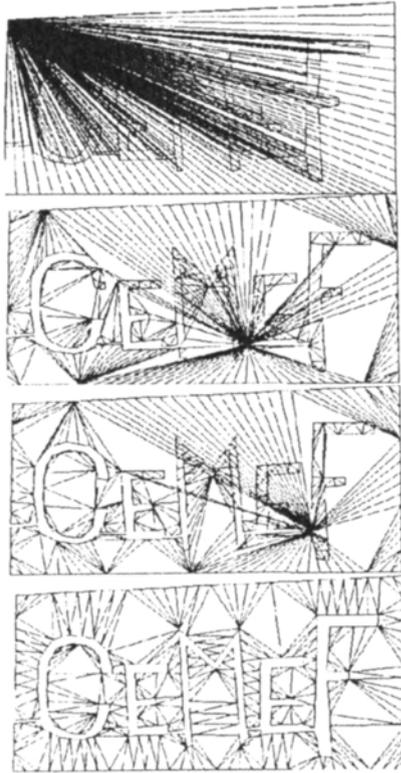


Figure 3: Mesh generation by the minimal volume principle

4. EXAMPLES OF ADAPTIVE SOLUTION IN LARGE DEFORMATION

4.1. 2D orthogonal metal cutting

Figure 4.1 shows the chip formation in orthogonal cutting. The cutting tool can be seen on the right part of the figures, and the initial workpiece is a simple rectangle. In order to accurately model the flow in the two main shear zones, an adaptive meshing technique is used. At each time step, the error is estimated. If it is larger than a prescribed value of 10%, a new adapted mesh is generated and the computations start again. It makes it possible to dynamically track the shear zones. The error estimate used in this

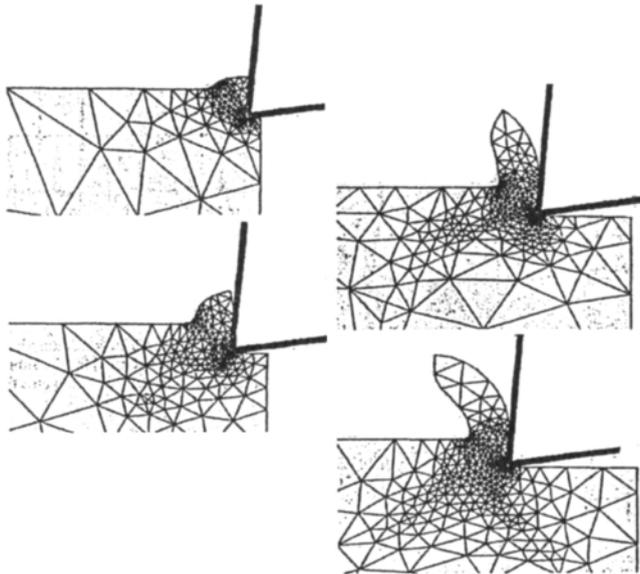


Figure 4: Adapted meshes in orthogonal metal cutting process simulation. Different stages of the chip formation.

adaptive scheme is based on the Zienkiewicz-Zhu [25] error estimators. The Liska/Orkisz finite difference method is used as smoothing operator. It improves the efficiency of the estimation [17]. Being a second order smoothing technique, it can be compared to the superconvergence patch recovery techniques, and provides a robust and reliable estimation. The error of the estimation is less than 5% for elasticity problems and less than 10% for viscoplasticity problems.

The discretization error is defined in term of an energy norm for both the elasticity

and plasticity problems:

$$\|v - v^h\|_E = \left(\int_{\Omega} (S - S^h) : (\varepsilon(v) - \varepsilon(v^h)) \, d\Omega \right)^{\frac{1}{2}} \quad (3)$$

In the context of a viscoplastic material modeled by a power law, this can be related to the $W^{1,m+1}$ (semi)norm used in [2].

Using classical C^0 approximation for the velocity, the discrete deviatoric stress tensor S^h defined by the constitutive equation is discontinuous. A simple idea, used in the example of Figure 4.1, is to build a continuous interpolation of the stress from the natural discontinuous one and to compute the gap between these two value of the stress field.

The natural discrete space for the stress field is the one spanned by the symmetric gradient operator : $S^h = \varepsilon(\mathcal{V}^h)$, \mathcal{V}^h being the velocity discrete space. For a continuous polynomial interpolation of degree k :

$$\mathcal{S}^h = \{ \text{tensor field } S, S \in P^{k-1}(\Omega_e)^{d \times d} \}$$

A more regular discrete space is then introduced :

$$\tilde{\mathcal{S}}^h = \{ \text{tensor field } S \in C^0(\Omega)^{d \times d}, S \in P^l(\Omega_e)^{d \times d} \}$$

where l is generally greater than $k - 1$ and the continuity is enforced. The last stage is to build a projection operator from \mathcal{S}^h to $\tilde{\mathcal{S}}^h$ (for instance using the the superconvergence patch recovery techniques, or a second order local finite difference scheme as proposed by Liska and Orkisz) and finally to define the estimate by :

$$E = \left(\int_{\Omega} \eta(|\varepsilon(v)|)^{-1} (\tilde{S} - S^h) : (\tilde{S} - S^h) \, d\Omega \right)^{\frac{1}{2}} \quad (4)$$

Adaptive refinement is based on error indicators, which can be the contribution of the element to the global error (4). For 2D problems, a Delaunay remeshing procedure provides six node quadratic triangles. The quality of the finite element mesh is checked at each time increment. Then, a new mesh is generated when the estimated error is too high . The example of Figure 4.1 and the details of the method is given in [18].

4.2 3D polymer cutting

The example of Figure 4.2 (obtained in [28]) is probably among the more difficult trial in numerical simulation involving remeshing. Figures 4.2 and 4.2 show the polymer granular formation by cutting a cylinder extrudate. The aim of the simulation is to predict the shape aspect of the granular, which depends on the cutting stage and more particularly on the relative value of the extrusion flow rate and the knife velocity.

This example combines several difficulties, as it is a problem mixing fluid and solid dynamic. As in the 2D metal example, the cutting is obtained by a local deformation. The unilateral contact condition is enforced by changing the inequality to an equality and using a penalty method :

$$(\vec{v}^{\dagger} - \vec{v}^{\ddagger}) \cdot \vec{n}^{\dagger} \geq 0 \iff [(\vec{v}^{\dagger} - \vec{v}^{\ddagger}) \cdot \vec{n}^{\dagger}]^{-} = 0$$

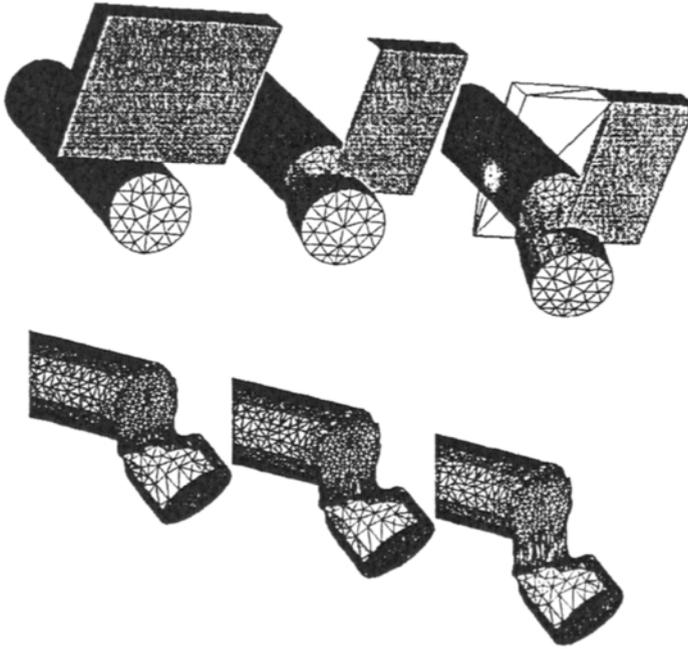


Figure 5: 3D cutting simulation in a mixed extrusion deformation application. Flow velocity/knife velocity = $\frac{1}{10}$

where $[x]^-$ denotes the negative part of x . Although in this case, the remeshing is not controlled by a volume error estimate, the mesh size is locally enforced both by the contact adaptive procedure and by an a priori local mesh size map.

In Figure 4.2, the knife velocity is slower, giving the time to the formation of a material jet before it leaves. This shows the possibilities in terms of numerical simulation offer by a dynamic remeshing scheme. The meshing becomes a full part of the calculation and the results will depend both on the ability to compute such a deformation and to provide the mesh fitting it.

5. MESHING IN EXTRUSION

5.1. 2D extrusion

The error control procedure and the corresponding adaptive scheme are illustrated in Figure 5.1, for a metal die extrusion problem. In this example, the adaptive process tends to exacerbate the refinement at the neighboring of the singularities of the inner corner and of stick/slip transition. If the estimated error is greater than a prescribed value, a totally new optimal mesh is generated. This is due to the mesh generator based on the standard Delaunay triangulation algorithm. The situation is different in our 3D examples

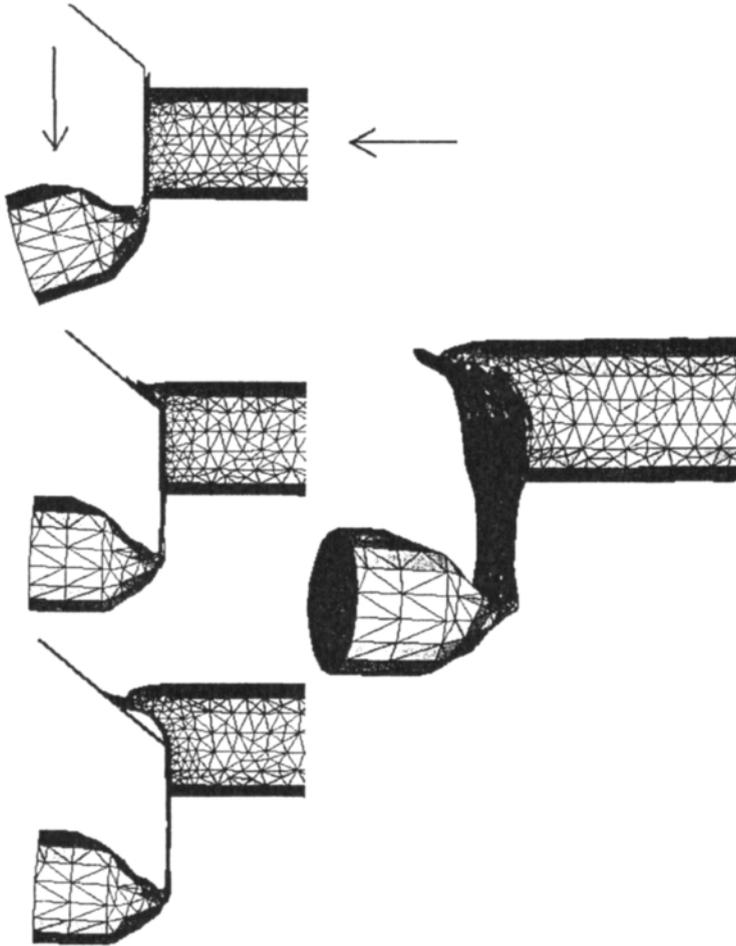


Figure 6: 3D extrusion-cutting. Flow velocity/knife velocity = $\frac{1}{3}$

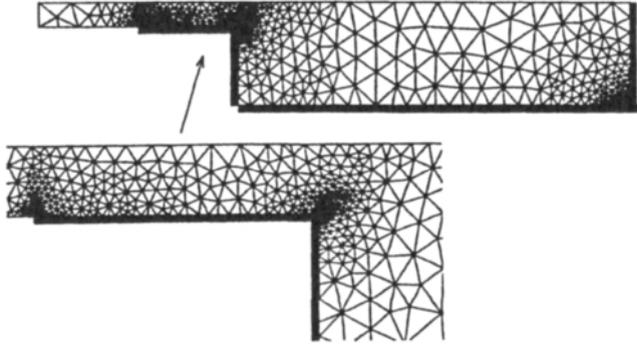


Figure 7: Adaptive remeshing in a 2D (axisymmetric) metal extrusion application

since the mesh generator is based on the general mesh improvement principle and can use the old mesh as a starting point.

This new mesh is said to be optimal if the error is uniformly distributed over all the elements. Under the hypothesis of a uniform convergence rate p of the finite element method, the predicted size of each new element is a function of the present size of the element and its contribution to the error.

The size of a quadratic triangle is defined as the maximum length of its three linear sides. For a viscoplastic material this value depends both on the value of m and on the degree of the interpolation functions. For a value of m around 0.1 which is the common value in metal application, the a priori convergence rate of the finite element method can decrease to 1.1 for a quadratic interpolation. A new mesh is subsequently generated and the size criterion is imposed over all the elements. At each time step the error is estimated in the workpiece (and possibly in the tools). If at any time the prescribed value is not respected, a new adapted mesh is generated. If the problem exhibits a singularity or if boundary conditions change very quickly, several iterations of this remeshing procedure are required to reach the prescribed accuracy for a given time step.

5.2. Anisotropic mesh in 3D die extrusion problem

The 3D meshing problem of the 3D die extrusion problem is not only due to the multiple singularities encountered in the realistic geometries plotted on Figure 8, but also to the shape ratio between the entrance and the exit which exhibits a factor of about 1/100.

The only way to do any computation with such a geometry is to reduce the number of elements by using an anisotropic meshing technique [23]. By this way it possible to guarantee several elements in the thickness at the die exit. However it is not yet theoretically well established, the use of such meshes showing stretched tetrahedra can be a posteriori justified by numerical experiments. The finite element convergence can even be reached by means of an asymptotically mesh independent solution.

An error estimate based on the residual method has been implemented in [22] for the heat transfer equation showing clearly the anisotropic character of the error.

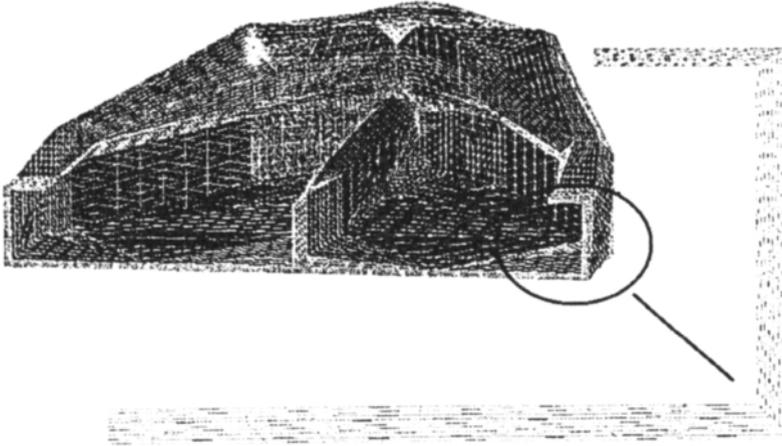


Figure 8: boundary of the mesh of a die extrusion geometry and detail of the mesh in the exit thickness

The temperature field is solution of :

$$\rho C \left(\frac{\partial T}{\partial t} + \nabla T \cdot v \right) = k \Delta T + \dot{w} \quad (5)$$

The thermal and mechanical equations are coupled by the dissipative term $\dot{w} = \sigma : \dot{\epsilon}$.

Assuming that a standard Galerkin method is used to solve the above equation, on each element the error indicator is given by :

$$I_e^2 = |\Omega_e| \int_{\Omega_e} (\rho C \nabla T \cdot v - \dot{w})^2 d\Omega_e + \sum_{F \in \partial\Omega_e} |F| \int_F (k[\nabla T \cdot n])^2 dS \quad (6)$$

where $[\nabla T \cdot n]$ is the jump through the face F of tetrahedron Ω_e . This jump gives rise to an error indicator in the three spatial directions, indeed :

$$\int_F (k[\nabla T \cdot n])^2 dS \leq 3 \times \sum_{i=1,3} Err_i^2$$

with

$$Err_i = \left(\int_F (k[(\nabla T)_{,i}])^2 dS \right)^{\frac{1}{2}}$$

for $i = 1, 3$.

In table 1, the first column contains the number of nodes of the different tested meshes, showing that the error in the z-axis is about 10 times greater than the error in the other directions.

Although the construction of a mesh with anisotropic properties, by means of different mesh length requirement in different directions at space points is a difficult task in 3D, the

nodes	I_e^2 max	Err-x max	Err-y max	Err-z max
45	11263	282	448	8090
225	8005	193	341	3714
1377	1356	83	95	1149
9537	427	18	25	414

Table 1: Error estimate on the complete heat transfer equation with respect to the mesh refinement

building of the Riemann metrics from the computed solution is also not obvious. From the numerical analysis point of view, the background theory of the finite element method needs still to be extended.

A first approach is based on the second derivative of a scalar field [4], providing a tensor from which is deduced the metric tensor. This is a natural idea when one uses simple $P1$ finite element. Indeed in this case, the a priori error depends on the second derivative of solution. This technique requires to built a second derivative from the discrete solution which not exists. Moreover each scalar field provides a metric tensor and each component of a vector field (in particular the velocity) must independently analyzed, giving rise to different metrics which must be combined.

A another idea has been proposed in [19]. The error on the stress field $\Delta S = \tilde{S} - S^h$ of Equation 4 can be rewritten by using its eigenvalues :

$$Error = (\Delta S : \Delta S)_E = \left(\int_{\Omega} \eta^{-1} (\Delta S_I^2 + \Delta S_{II}^2 + \Delta S_{III}^2) d\Omega \right)^{\frac{1}{2}}$$

where ΔS_I are the principal value of ΔS in the principal direction.

Assuming the validity of the error estimate, it is clear that the error will be optimally minimized in the eigenvector basis. Consequently, the metric M can be defined by :

$$M = \eta^{-\frac{1}{2}} \mathcal{R} \begin{pmatrix} |\Delta S_I| & 0 & 0 \\ 0 & |\Delta S_{II}| & 0 \\ 0 & 0 & |\Delta S_{III}| \end{pmatrix} \mathcal{R}^{-1} \quad (7)$$

6. L^2 ERROR ESTIMATION

6.1. Domain interpolation

Generally in finite element method the domain of calculation is assumed to be known and it is supposed that the mesh represents it exactly. However for certain fluid dynamic application it is easier to adopt an Eulerian approach with a fixed mesh. Then the moving domain is not any longer represented by a deformable mesh.

We set the following problem : let Ω be a bounded domain in \mathbb{R}^3 . The characteristic function of this domain $\mathbb{1}_{\Omega}$ is defined by :

$$\mathbb{1}_{\Omega}(x) = \begin{cases} 1 & \text{if } x \in \Omega \\ 0 & \text{elsewhere} \end{cases} \quad (8)$$

The following development is based on the fact that it is possible to approximate a domain by interpolating its characteristic function. However this function is discontinuous. Let us introduce the following finite element interpolation (in the sense of the Clement operator [6]) using discontinuous piecewise constant function .

$$\mathbb{1}_{\Omega}^h = \sum_{e \in \mathcal{E}} \mathbb{1}^e \mathbb{1}_{\Omega^e}$$

where \mathcal{E} in the set of geometric elements

Theorem 1 *If $\mathbb{1}_{\Omega}^h$ is the best approximation of $\mathbb{1}_{\Omega}$ in the L^2 norm (the projection onto the piecewise function finite space) then :*

$$\mathbb{1}^e = \frac{|\Omega \cap \Omega^e|}{|\Omega^e|}$$

and the error

$$\|\mathbb{1}_{\Omega} - \mathbb{1}_{\Omega}^h\|_{L^2(\Omega)} = \left(\sum_{e \in \mathcal{E}} \mathbb{1}^e (1 - \mathbb{1}^e) |\Omega^e| \right)^{\frac{1}{2}}$$

and in \mathbb{R}^2 :

$$\|\mathbb{1}_{\Omega} - \mathbb{1}_{\Omega}^h\|_{L^2(\Omega)} \leq C(\Gamma) h^{\frac{1}{2}}$$

$C(\Gamma)$ is a constant depending on the regularity of Γ the boundary of the domain. Although, it is still unclear, it is expected that the error evaluation is valid in 3D.

One important point is that the error estimate (1) depends only on the local values $\mathbb{1}^e$. When these value are given otherwise, it can be used as an a posteriori error estimate. That is the case in the following where the approximate function $\mathbb{1}_{\Omega}^h$ is also solution of a transport equation.

6.2. Dynamic a posteriori estimation of a flow front position

The 3D mold filling can be solved by coupling an extended Stokes solver and the solution of the following transport equation [26] :

$$\frac{d}{dt} \mathbb{1}_{\Omega}^h = 0.$$

It is solved at element level by using a Taylor discontinuous Galerkin method. The front position is given as an average surface in Figure 9 and it corresponds to the region of the maximum value of the above error estimate.

7. CONCLUSION

Examples have been presented showing the new possibilities in the numerical simulation of forming processes provided by adaptive solutions. The 2D examples show the complete adaptive procedure with a prediction of the numerical error for nonlinear behavior material

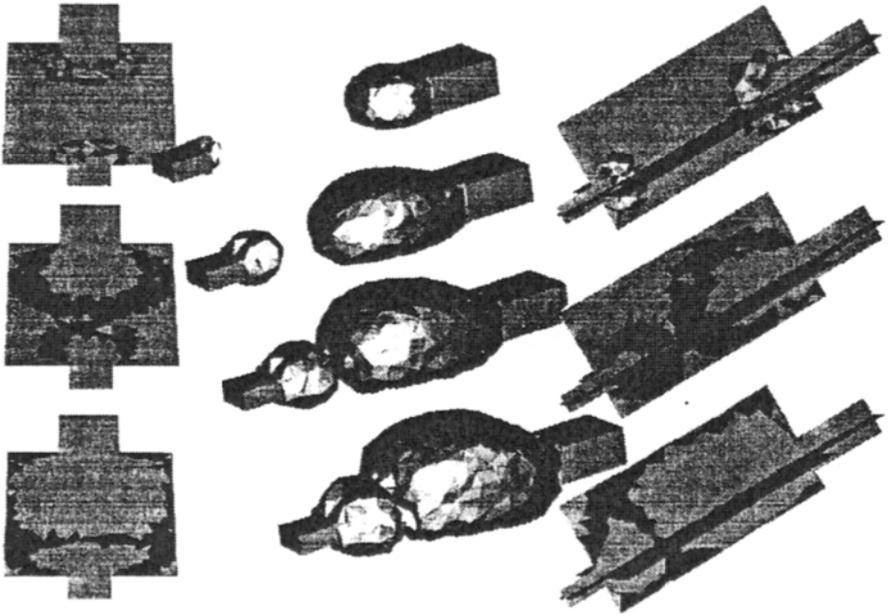


Figure 9: Calculation of a two colliding jets in 3D : the average front surfaces and the a posteriori error estimate at successive time in a vertical and horizontal cutting plans

and the rebuilding of the mesh under the constraint of the calculated mesh size map, all in the context of dynamic simulation with time dependent moving domain. Examples where the local mesh size control is essential are given for cutting application both in 2D and in 3D. The extrusion problem with multiple singularities can be treated adaptively. In 3D, for the die extrusion problem, the new emerging technique of anisotropic meshing has shown to be unavoidable in order to carry out any calculation. Finally, an L^2 error estimate has been used to quantify the error in the fluid domain representation for 3D mold filling application.

REFERENCES

- [1] D. N. Arnold, F. Brezzi, and M. Fortin. A stable finite element for Stokes equations. *Calcolo*, 21:337–344, 1984.
- [2] J. Baranger and H. El. Amri. Estimateurs a posteriori d'erreur pour le calcul adaptatif d'écoulements quasi-newtonien. *R.A.I.R.O. mod. math. and anal. numer.*, 25(1):31–48, 1991.
- [3] J. Baranger and K. Najib. Numerical analysis of quasi-newtonian flow obeying the power law or the carreau law. *Numer. Math.*, 58:35–49, 1990.
- [4] H. Borouchaki, M.J. Castro-Diaz, P.L. George, F. Hecht, and B. Mohammadi. Anisotropic adaptive mesh generation in two dimensions for cfd. In B.K. Soni et al., editor, *Numerical Grid Generation in Computational Field Simulations*, volume 1, pages 197–206. Mississippi State University, 1996.
- [5] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, 1991.
- [6] Ph. Clement. Approximation by finite element functions using local regularization. *R.A.I.R.O.*, R-2(77–84), 1975.
- [7] T. Coupez. *Grandes transformations et remaillage automatique*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris, CEMEF, Sophia Antipolis, 1991.
- [8] T. Coupez. A mesh improvement method for 3d automatic remeshing. In N.P. Weatherill et al., editor, *Numerical Grid Generation in Computational Fluid Dynamics and Related Fields*, pages 615–626. Pineridge Press, 1994.
- [9] T. Coupez. Stable/stabilized finite element for 3d calculation of viscoplastic/pseudoplastic material. CEMEF, en preparation, 1994.
- [10] T. Coupez. Automatic remeshing in three-dimensional moving mesh finite element analysis of industrial forming. In Shan-Fu Shen et al., editor, *Numerical Methods in Industrial Forming Processes - NUMIFORM 95*, pages 407–412. A.A. Balkema, 1995.

- [11] T. Coupez. Parallel adaptive remeshing in 3d moving mesh finite element. In B.K. Soni et al., editor, *Numerical Grid Generation in Computational Field Simulations*, volume 1, pages 783–792. Mississippi State University, 1996.
- [12] T. Coupez. Mesh generation and adaptive remeshing by a local optimisation principle. In *NAFEMS world congress 97*, pages 1051–1060. NAFEMS Ltd, Glasgow, 1997.
- [13] T. Coupez and J. L. Chenot. Mesh topology for mesh generation problems. application to three-dimensional remeshing. In O.C. Zienkiewicz J. L. Chenot, R.D. Wood, editor, *Numerical Methods in Industrial Forming Processes (NUMIFORM)*. A.A. Balkema, 1992.
- [14] T. Coupez and S. Marie. From a direct solver to a parallel iterative solver in 3d forming simulation. *Int. J. of Supercomputer Applications*, 11(4):205–211, 1997.
- [15] Maitre J. F. and Wabo E. Stabilized formulations and mini-element for the n-dimensional stokes equations: Properties and solution procedure. *lecture notes in pure and applied mathematics*, 164:285–298, 1994.
- [16] M. Fortin and A. Fortin. Experiments with several elements for viscous incompressible flows. *Int. J. for Num. Meth. in Fluids*, 5:911–928, 1985.
- [17] L. Fourment and J. L. Chenot. Error estimators for viscoplastic materials: application to forming processes. *Engrg. Comp.*, 12(5):469–490, 1995.
- [18] L. Fourment and J.L. Chenot. Adaptive remeshing and error control for forming processes. *R. europ. des éléments finis*, 3(2):247–280, 1994.
- [19] L. Fourment, M. P. Miles, F. Bay, D. Carpentier, and J. L. Chenot. Adaptive strategies for the simulation of the 2d and 3d forming processes. In E. Onate D.J.R. Owen, editor, *Computational Plasticity*, volume 1, pages 395–406. Pineridge Press, Swansea, 1995.
- [20] L.P. Franca, S.L. Frey, and T.J.R. Hughes. Stabilized finite element methods : I. application to the advective-diffusive model. *Comput. Meth. Appl. Mech. Engrg.*, 95:253–276, 1992.
- [21] P.L. George. *Automatic mesh generation. Applications to finite element methods*. Wiley, 1991.
- [22] J. F. Gobeau. *Etude expérimentale et simulation numérique 3D par éléments finis de l'écoulement dans les filières d'extrusion de profilés PVC*. thèse de doctorat, Ecole des Mines de Paris - CEMEF, Sophia Antipolis, 1996.
- [23] Gobeau J.F., Coupez T., Vergnes B., and Agassant J.F. Computation of profile dies for thermoplastic polymers using anisotropic meshing. In Paul R. Dawson Shan-Fu Shen, editor, *Numerical Methods in Industrial Forming Processes - NUMIFORM 95*, pages 59–66. A.A. Balkema, 1995.

- [24] P. Ladevèze, G. Coffignal, and J. P. Pelle. *Accuracy estimates and adaptivity for finite elements*, chapter 10, pages 181–203. Wiley, Chichester, 1986.
- [25] J.Z. Zhu O.C. Zienkiewicz. A simple error estimator and adaptive procedures for practical engineering analysis. *Int. J. Num. Meth. Eng.*, 24:337–357, 1987.
- [26] E. Pichelin and T. Coupez. Finite element solution of the 3d mold filling problem for viscous incompressible fluid. submitted to *Comp. Meth. Appl. Mech. Engrg.*, 1997.
- [27] R. Pierre. Optimal selection of the bubble function in the stabilization of the p1-p1 element for the Stokes problem. *SIAM J. Numer. Anal.*, 32(4):1210–1224, August 1995.
- [28] J. Renaud. Master's thesis, Ecole des Mines, CEMEF, 1997.
- [29] J. Y. Talon. Algorithme d amelioration de maillage. Rapport technique, Artemis Imag, 25, 1987.
- [30] R. Verfurth. A posteriori error estimat for the stokes equations. *Numer. Mathematik*, 55:255–309, 1989.

This Page Intentionally Left Blank

Recovery procedures in error estimation and adaptivity : Adaptivity in non-linear problems of elasto-plasticity behaviour

B. Boroomand^{*}, O.C. Zienkiewicz^{*}

^{*}University of Wales, Swansea, U.K.

ABSTRACT

The first part of this paper [1] discusses in some detail the use of recovery procedures for stress, strain or gradients for efficient error estimator and the methods of arriving at adaptive, h , refinement. In this paper we extend such procedure to deal with of elasto-plasticity and discuss in detail some special features of such computation.

1. INTRODUCTION

Error estimation procedures and adaptive mesh design based on recovery of stresses , strains other gradients have been discussed in some detail on Part I of this paper for linear problems. Indeed application of such methods to practical engineering analysis is today widely available. The extension of the methodology to non-linear problems appears at first glance straightforward as the solution is generally obtained by successive linearization, however several difficulties arise which we shall discuss in this part of the paper. Much research has been devoted to the subject in the last decade as is evident from reference [2]- [17].

The applications in this paper are focused on static problems of elasto-plasticity, common in engineering, but the methodology as equally application of large strain problems and in dynamic phenomena.

Generally, due to path dependency of the solution in materially non-linear analyses, the loading is applied in an incremental manner with an iterative linearization, using the Newton-Raphson process in each increment. The iteration is usually continued until a suitable norm of variables falls within a prescribed tolerance. Depending on the tolerance used in the termination of Newton-Raphson iteration, an additional error to that due to discretization will occur in the computational process. This paper mainly focuses on the discretization error and we assume that the magnitudes of the increments and tolerance used are sufficiently small to make the iteration errors secondary.

The magnitude of error due to spatial discretization itself, is a function of time (or pseudo time defined by the load history in static problems). The procedure usually starts with the selection of a reasonable element distribution over the mesh. The decision of remeshing while the non-linear computation is being carried out, requires an important engineering judgement. An appropriate norm of variables must be devised to monitor the accuracy of the solution helping to recognize the necessity of remeshing. The norm must be capable of describing the error and depending on the importance and computation cost of the problem the desired bounds of permissible error itself may vary.

Various measures of error are adapted by investigators. The majority use either the L_2 norm [6] of displacements or the energy norm [6, 8, 12]. Some however use local indicators for refinement. Here local values of plastic strain [7] and more usefully gradients (or curvatures) of total displacement [4, 5, 10, 11] have been used for this purpose. In this paper we shall use energy norm in the same sense as in linear cases. As the recovery based error estimates have proved to be the best in linear problems, we shall adapt them in non-linear ones to construct the energy error norm. Obviously the performance of the recovery method used plays a predominant role in the asymptotic exactness of the error estimates and therefore we shall take the advantage of two recovery methods SPR (Superconvergent Patch Recovery [18]) and REP (Recovery by Equilibrium in Patches [19, 20]) which have shown excellent performance in linear problems.

If the error is controlled up to n th increment, the procedure will reduce to control the error occurring in the next increment. Therefore instead of defining the energy norm for the whole process up to increment n we use an incremental energy norm. The procedure will therefore be simply that of solving the non-linear problem with the old mesh for increment $n + 1$ and checking the incremental energy norm. If the relative error exceeds a prescribed limit, the analysis will be repeated for the same increment with a new mesh. The new element sizes will be calculated by introducing a-priori assumptions about the convergence rate and using the optimal mesh concept in which a uniform distribution of error is considered for the new mesh configuration.

The history of the solution is usually stored at integration points and after generating a new mesh, a corresponding history must be constructed for the new integration points. Various transfer operators can be found in the literature [4, 7, 13]. The state variables at the end of increment n satisfy the constitutive relations of material and the stresses are in equilibrium state in finite element sense. However, there is no guarantee that the new state variables satisfy such conditions. It seems to be impossible to satisfy both consistency and equilibrium (in FEM sense) conditions on the new mesh. Here engineering judgement is required for selection of one of these conditions. Satisfaction of the consistency condition and the neglect of the equilibrium condition is, sometimes, preferred [7, 13].

In this paper, after a brief overview of general formulation of plastic problems, we explain the error norm used in error estimator. The problem of updating the state variable will be discussed and here we shall introduce a new transfer operator through which we transfer the information directly from old integration points to the new ones. The authors have proposed in another paper [17] a procedure in which the SPR method

is used to derive nodal values and to obtain integration values by interpolation but occasionally the procedure appears to be divergent. Finally we shall conclude our paper by some numerical examples of adaptive refinement.

2. GENERAL FORMULATION OF ELASTO-PLASTICITY [21]

For a brief review of general plasticity procedure, for infinitesimal deformation, we start from the differential equilibrium equation as:

$$\mathbf{S}\sigma = \mathbf{b} \quad \text{in } \Omega \quad (1)$$

with σ being Cauchy stress tensor. The constitutive equation results from the following assumptions:

$$\begin{aligned} \epsilon(\mathbf{u}) &= \epsilon_e + \epsilon_p & \epsilon_p &= \sum \delta\epsilon_p \\ \sigma &= \mathbf{D}\epsilon_e \\ \delta\epsilon_p &= \delta\lambda(\partial F/\partial\sigma) \end{aligned}$$

where \mathbf{u} , ϵ , ϵ_e , ϵ_p and \mathbf{D} denote the displacement field, total strain, elastic strain, plastic strain and elastic modulus of material, respectively, and λ is a proportionality parameter. Further, F represents the yield surface defined in terms of stresses and some internal variables κ (hardening etc) as:

$$F(\sigma, \kappa) = 0 \quad (2)$$

The equilibrium equation (1) in the finite element sense is written as:

$$\psi(\mathbf{u}) \equiv \int_{\Omega} \mathbf{B}^T \sigma d\Omega - \mathbf{f} = \mathbf{0} \quad (3)$$

Here $\mathbf{B} \equiv \mathbf{S}\mathbf{N}$ where \mathbf{N} is 'shape' function for \mathbf{u} and \mathbf{f} represents the nodal loads due to the body forces \mathbf{b} and tractions \mathbf{t} as a vector. In incremental schemes for non-linear problems, in which we define the loading incrementally, the above equation should be satisfied in each increment:

$$\psi_n(\mathbf{u}) \equiv \int_{\Omega} \mathbf{B}^T \sigma_n d\Omega - \mathbf{f}_n = \mathbf{0} \quad (4)$$

where

$$\mathbf{f}_n = \int_{\Omega} \mathbf{N}^T \mathbf{b}_n d\Omega + \int_{\Gamma} \mathbf{N}^T \mathbf{t}_n d\Gamma \quad (5)$$

Equation (4) together with the constitutive plastic relations must be satisfied simultaneously. This can be achieved by a Newton-Raphson iterative scheme within the load increment writing:

$$\psi_n \approx \psi_n^i + \frac{\partial \psi_n^i}{\partial \bar{u}} \delta \bar{u}_n^i = 0 \quad (6)$$

This leads to

$$\int_{\Omega} \mathbf{B}^T \sigma_n^i d\Omega - \mathbf{f}_n + \mathbf{K}_n^i \delta \bar{u}_n^i = 0 \quad (7)$$

where i is the iteration number, $\delta \bar{u}_n^i$ denotes the incremental nodal displacements and \mathbf{K}_n^i is an appropriate global tangent matrix evaluated from the application of constitutive plastic relations:

$$\mathbf{K}_n^i = \int_{\Omega} \mathbf{B}^T \mathbf{D}_{ep}^i \mathbf{B} d\Omega \quad (8)$$

The incremental nodal displacements can now be determined by:

$$\delta \bar{u}_n^i = \mathbf{K}_n^{i-1} \{\psi_n^i\} \quad (9)$$

from which the total displacement at increment i can be obtained as:

$$\bar{u}_n^i = \bar{u}_{n-1} + \sum_{i=0}^i \delta \bar{u}_n^i \quad (10)$$

Having the former situations of history dependent parameters, such as stress components and hardening parameters, and using the constitutive plastic relations leads to a new set of values for the parameters:

$$(\sigma_n^i, \kappa_n^i, \delta \epsilon_n^i) \Rightarrow (\sigma_n^{i+1}, \kappa_n^{i+1}) \Rightarrow \mathbf{D}_{ep}^{i+1} \quad (11)$$

By returning to Equation (7) the procedure will continue until reasonable convergence is achieved.

The state of material at each integration point is recognized by using the incremental stresses and hardening values in the yield surface relation:

$$\begin{cases} F(\sigma_n^{i+1}, \kappa_n^{i+1}) < 0 & \text{Elastic} \\ F(\sigma_n^{i+1}, \kappa_n^{i+1}) \geq 0 & \text{Plastic} \end{cases} \quad (12)$$

At the start of each iteration new trial values for stresses are computed from the strain changes assuming that the material is elastic:

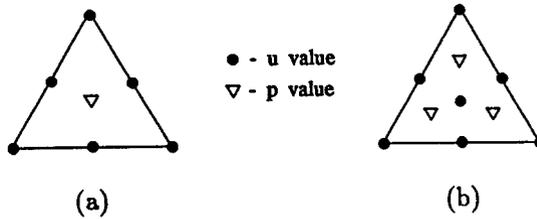


Figure 1: Two triangular elements used in adaptive analyses with nearly incompressible behaviour (a) T6/1 - quadratic displacement, constant pressure (b) T6/3B - quadratic displacement using bubble function and linear discontinuous pressure field

$$\sigma_n^{i+1} = \sigma_n^i + \mathbf{D}\delta\epsilon_n^i \quad (13)$$

Equation (13) is called the *elastic predictor*. If the material remains elastic the computation will be continued by restoring the new values of stresses and hardening values and using the elastic modulus $\mathbf{D}^{i+1} = \mathbf{D}$. But if the material is plastic a new set of stresses and hardening parameters should be found so that the yield surface relation of Equation (2), is satisfied.

A dimensionless norm of variables is usually selected as:

$$\xi = \eta_n^i = \frac{\|\psi_n^i\| - \|\psi_n^{i-1}\|}{\|f\|} \times 100 \leq \bar{\xi} \quad (14)$$

to terminate the Newton-Raphson iteration. If the tolerance $\bar{\xi}$ is chosen to be very small, the errors due to the linearization process will be negligible compared with the error of discretization.

Recent advances in computational plasticity have led to a new algorithm for some classes of flow rules such as Von Mises criterion, called return mapping algorithm [22]-[23] (for Von Mises it is called radial return mapping). In this algorithm, after evaluating elastic predictor, a new state of stresses is usually found on the yield surface by projecting the stresses from their new position along the unit normal to the surface. Based on this new state of stresses an algorithmic elasto-plastic tangent modulus "consistent tangent modulus" is computed. The main advantage of the latter algorithm is the increase of rate of convergence in the Newton-Raphson iteration.

The convergence of plasticity procedure is also affected by other factors such as the possibility of incompressibility. For Von Mises criterion, when the material is fully plastic, the behavior of the material is nearly incompressible. In such situations a 'locking' phenomenon may happen which can lead to divergence. There are several remedies for solving problems near incompressibility. In this paper we have used the so called 'B-bar' method [24, 25]. This is essentially a mixed procedure of analysis using

as variables the displacement u and the mean stress (or pressure p) and is described in some detail in text by Zienkiewicz and Taylor [21]. For success of the method only certain forms of shape functions can be used. In the examples dealt with in the paper two triangular elements shown in Figure 1 were used. These elements are introduced and discussed in reference [26].

3. ERROR ESTIMATION AND ADAPTIVE REFINEMENT

In this section we shall explain the procedure of adaptive refinement in a typical plasticity problem. The aim is to perform adaptive procedure incrementally refining the mesh at any increment in which the error has exceeded a prescribed value, and to continue the procedure without restarting from the first load step.

3.1. ERROR ESTIMATION

Several refinement criteria have been so far suggested. Two main categories can be identified in literature for refinement criterion in non-linear adaptivity:

- (1) Refinement criterion based on gradients (or curvature) of displacement. This kind of criterion is usually used in localization problems [4, 5, 10] and can only indicate the existence of error but not its magnitude.
- (2) Refinement criterion based on evaluation of errors in each element. The error is usually evaluated through computing an improved values of stress/strain (or other parameters) by suitable recovery method in the manner described in Part I of this paper.

In this paper we shall be exclusively concerned with error evaluation using recovery procedures and thus using the second method of refinement. The concept of using recovery procedures for error estimation was first introduced by Zienkiewicz and Zhu in 1987 [27]. In that paper simple stress averaging and L_2 projection were used as recovery procedures. In non-linear application many authors followed precisely this early approach. Belytschko and Tabbara [6] found the recovery by L_2 projection performing well (though they also have shown that another recovery suggested by Daiz *et al* [28]). Gallimard *et al* [12] use an incremental form of error estimation and a recovery based on achieving equilibrating stresses between elements [29].

Lee and Bathe [7] use a pointwise error in strains to guide the refinement and a recovery procedure introduced by Hinton and Campbell [30].

In reference [8], Peric *et al* have used some element-wise error estimators based on incremental energy norms though they also introduced an alternative norm based alone on the elastic modulus (also see [14]). Here they compare the behavior of two classes of material models, those of Classical plasticity and of Cosserat continuum. The improved fields of stress or strains are computed by L_2 smoothing method (see [21]). It has been concluded that the performance of the error estimators lies in the superconvergence property of the recovery method.

New methods of recovery such as the SPR procedure [18] and the REP procedure [19, 20] result, in linear cases, in much improved error estimation and their use for non-linear problems of plasticity is made for the first time in this paper.

The error occurring in a typical increment will be measured by an energy norm defined as:

$$\|e\| = \left[\int_{\Omega} |(\sigma - \sigma_h)^T (\Delta \epsilon - \Delta \epsilon_h)| d\Omega \right]^{\frac{1}{2}} \quad (15)$$

In above σ and σ_h denote the exact and finite element stress fields, respectively.

This kind of norm may be considered as related to an average stress (or strain) error over the domain. But what makes it attractive is that it also represents the accuracy of the displacement field since energy can be considered as a product of the displacement and applied force. Therefore energy error norm represents average error of displacements at points with external forces (or average error of displacements over the whole domain in the presence of body forces) or average error of reactions at points with prescribed displacements.

In n th increment the above norm may be written as:

$$\|e\| = \left[\int_{\Omega} |(\sigma_n - \sigma_{h_n})^T (\Delta \epsilon_n - \Delta \epsilon_{h_n})| d\Omega \right]^{\frac{1}{2}} \quad (16)$$

in which

$$\Delta \epsilon_n = \epsilon_n - \epsilon_{n-1} \quad \Delta \epsilon_{h_n} = \epsilon_{h_n} - \epsilon_{h_{n-1}} \quad (17)$$

If we define also:

$$\Delta \sigma_n = \sigma_n - \sigma_{n-1} \quad \Delta \sigma_{h_n} = \sigma_{h_n} - \sigma_{h_{n-1}} \quad (18)$$

then using triangle inequality we have:

$$\|e\| \leq \left[\int_{\Omega} |(\sigma_{n-1} - \sigma_{h_{n-1}})^T (\Delta \epsilon_n - \Delta \epsilon_{h_n})| d\Omega \right]^{\frac{1}{2}} + \left[\int_{\Omega} |(\Delta \sigma_n - \Delta \sigma_{h_n})^T (\Delta \epsilon_n - \Delta \epsilon_{h_n})| d\Omega \right]^{\frac{1}{2}} \quad (19)$$

The first conclusion of this decomposition is that in an efficient adaptive procedure where we control the error up to time step $n - 1$, the difference between stress values $(\sigma_{n-1} - \sigma_{h_{n-1}})$ will be sufficiently small therefore the magnitude of $\|e\|$ will mainly be governed by the magnitude of the product of $(\Delta \epsilon_n - \Delta \epsilon_{h_n})$ and $(\Delta \sigma_n - \Delta \sigma_{h_n})$ which belong to time step n .

The second conclusion from (19) is that in time step n the stress values from previous steps act as constants and therefore, supposing that the order of solution is p and a-priori assumptions for convergence of gradients in linear problems are valid for non-linear ones, the convergence of the error norm will be:

$$\|e\| = O(h^{\frac{1}{2}}) + O(h^p) \quad (20)$$

Therefore the convergence of the adaptive procedure will not be the same as that of linear problems. This may affect the number of successive remeshing steps when we use a-priori assumption of $O(h^p)$ for the convergence at element level.

Similarly to linear examples, since there is no information about the exact solution, an improved stress and strain field can be used in the incremental energy error norm (16), thus:

$$\|e\| \approx \|e^*\| = \left[\int_{\Omega} |(\sigma_n^* - \sigma_{h_n})^T (\Delta \epsilon_n^* - \Delta \epsilon_{h_n})| d\Omega \right]^{\frac{1}{2}} = \left[\sum \|e^*\|_{\Omega_e} \right]^{\frac{1}{2}} \quad (21)$$

In order to have a robust error estimate, we recover the stress and strain fields by use of two recovery procedure, REP and SPR. For stress recovery we note that the information is only available at integration points and therefore the REP which is insensitive to the number of integration points can be used. The equilibration used in REP is very consistent with the general procedure of plasticity. Since the procedure is based on a patch-wise equilibrium, we can directly obtain the improved stress field σ_n^* at the end of time step n from the finite element answers σ_{h_n} as:

$$\int_{\Omega_p} \mathbf{B}^T \sigma_{h_n} d\Omega \approx \int_{\Omega_p} \mathbf{B}^T \sigma_n^* d\Omega \quad (22)$$

As we recommended elsewhere [20] the procedure can be performed component by component:

$$\int_{\Omega_p} \mathbf{B}^T \sigma_{i_n}^h d\Omega \approx \int_{\Omega_p} \mathbf{B}^T \sigma_{i_n}^* = \left(\int_{\Omega_p} \mathbf{B}^T (1^i)^T \mathbf{P} d\Omega \right) \mathbf{a}^i \quad (23)$$

Here $\sigma_{i_n}^h$, $\sigma_{i_n}^*$ are components of stress vector σ and \mathbf{a}^i is a vector of polynomial coefficients corresponding to the i th component of stress. The detail of minimization process can be found in reference [20]. The nodal values obtained from this process will then be interpolated to construct the new field of stresses in (21).

For strain recovery, the procedure of SPR can directly be used since derivatives of displacement are available at any arbitrary point inside the elements including sampling points. Thus a polynomial will be fit over the strain values at sampling points with a least square sense for each component:

$$\Pi = \sum (\epsilon_i^* - \epsilon_{s_i}^h)^2 \quad \Pi \rightarrow \text{minimum} \quad (24)$$

and

$$\epsilon_i^* = \mathbf{P} \mathbf{a}_i \quad (25)$$

Since we are dealing with incremental strains $\Delta\epsilon$ the answer of smoothing process of the previous increment step $n - 1$ needs to be stored at nodal points. To avoid this the incremental values of strains can be directly smoothed.

Having found the incremental energy error norm (21) the relative error of the solution can be obtained as:

$$\eta = \frac{\|e^*\|}{E} \times 100 \quad E = \left| \int_{\Omega} \sigma^{*T} \Delta\epsilon^* d\Omega \right|^{\frac{1}{2}} \quad (26)$$

In a successive mesh refinement within an increment, we seek to obtain an accuracy better than a prescribed value i.e.:

$$\eta \leq \bar{\eta} \quad (27)$$

The procedure of obtaining the refinement criterion follows exactly that used in linear problem by assuming equal distribution of error over the elements:

$$\frac{h_{new}}{h_{old}} = \left(\frac{\bar{\eta} E / N^{\frac{1}{2}}}{\|e^*\|_{\Omega_c}} \right)^{\frac{1}{p}} \quad (28)$$

In which N is the number of elements in the old mesh. We note that here we again assumed that the rate of convergence of $\|e^*\|_{\Omega_c}$ is p and this might affect the number of remeshing.

3.2. DATA TRANSFER OPERATOR

3.2.1. GENERAL SURVEY OF PROCEDURE

Having computed the new element sizes, a new mesh is generated which can be completely different from the old mesh. For such a mesh a set of new state variables such as displacements, stresses and hardening parameters must be constructed for the new mesh using the information from the old mesh. There are two kinds of state variables; namely, those which are stored at nodes (i.e displacements) and those which are stored at integration points (i.e stresses, etc).

For those state variables which are available at nodes a direct interpolation from the old nodes to the new ones may be used. However, the transfer of information from old Gauss points to the new ones is quite a challenging problem and has been addressed by many authors [4, 7, 9, 13]. In most cases a new field of nodal values is constructed for a particular variable (existing at Gauss points) after which a simple interpolation, similar to those of the first kind, is followed. Here we shall briefly describe the methods used.

In reference [4] a so called consistent transfer operator, using appropriate Hu-Washizu functional, is proposed. The method uses a discontinuous distribution of history dependent parameters using local interpolation functions. These functions are chosen so that

the interpolated field gives exactly the values at integration points. This can easily be achieved by employing local shape functions having unit values at integration points. In other words the method is based on interpolation of values at integration points when the new Gauss points (of the new mesh) are inside the group of old Gauss points in a particular element and extrapolating the values when the new Gauss point is outside the group of old Gauss points (but not outside the old element)¹. Therefore, clearly, the discontinuity of the field along the boundaries of the elements makes some difficulties for evaluating the values at those new Gauss points laying on the element edges. To solve this problem a specific form of remeshing is used in which the old element is divided into smaller elements so that the new Gauss points will always be inside the group of old ones in the old element. It is also shown that, for displacements, the consistency condition will be satisfied if the new displacement field is evaluated by simple interpolation of the nodal values from the old mesh. The methodology for transferring data from Gauss points fails when a general new mesh is used. In such a case a L_2 projection of variables from old Gauss points to new ones is suggested to evaluate a set of nodal values (viz [15, 16]).

The procedures used in references [7, 9] and [13] are very similar. In these approaches first the history dependent values in the old mesh are projected to nodes. Then the values at the new nodal points in the new mesh are computed by simple interpolation of the old nodal value using the original shape-functions. Construction of the set of nodal values in [7, 9] is performed by considering a patch of elements connected to the node and projecting the variable using a least square scheme in a similar fashion as SPR procedure from integration points (instead of sampling point in SPR). In [13] also, a patch-wise computation of nodal values using average of extrapolated fields of integration points, has been suggested. Similar to reference [4] the new displacement field is evaluated by interpolation of the values from the old nodes to the new ones. In order to have self-consistent parameters, in reference [7], only few parameters are first transferred. The remaining part of these parameters is computed from the constitutive relation. For example for J_2 materials only effective plastic strain e_p and the trial elastic deformation gradients $\delta\epsilon$ are transferred to the new mesh. The same discussion has been made in [13] but instead of transferring trial elastic deformation gradients to the new mesh, the transferred values of the displacements at the end of load step n in the old mesh is considered as a trial solution of the new mesh for this load step.

There can be an argument about the necessity of self-consistency of the history dependent parameters in the last two approaches. The reader can notice that by transferring the information to nodes and recalculating them at the new Gauss points the equilibrium of the system will be violated. In fact even if the mesh is not changed the above procedure will violate the equilibrium of the system. This can easily be observed by performing the procedure in a single mesh while the material is still elastic. Such a procedure, indeed, has been suggested to improve the finite element answers in elastic problems by Loubignac *et al* [31].

In another paper [17] the authors have used the SPR procedure to obtain nodal

¹In this method the stress field resulting from this interpolation will be exactly the same as the conventional stress distribution if the element remains elastic.

stresses before interpolating these to new Gauss points. If this is practiced at every step of the iteration, we recover an SPR-version of the Loubignac method. Unfortunately the process is not always convergent and for this reason we have adapted here the direct transfer (Gauss point of the old mesh to the Gauss points of the new mesh).

In L_2 form of projection of data to nodes, since the integration is taken over the whole domain, the information in a small area will spread over a comparatively larger zone. This effect is smaller for patch-wise approaches. For example if only one Gauss point in the patch becomes plastic, after smoothing, the information will be projected to the patch node and by interpolation of such value the whole patch area will receive a fraction of information of the Gauss point. In the following section a weighted form of the patch-wise approach will be introduced by which we can directly transfer the information from old Gauss points to the new one and since the weight is considered over the area of each Gauss point the data will not spread over a large area.

In summary in this paper we shall use:

a - a direct interpolation of variables available at nodes using interpolation basis:

$$\bar{u}_{new} = N\bar{u}_{old} \quad (29)$$

b - a direct Gauss point to Gauss point transfer operator described in the next section for information at integration points.

3.2.2. DIRECT GAUSS POINT TO GAUSS POINT TRANSFER OPERATOR

In this section we shall propose a suitable scheme which does not need nodes as reference points but still uses them for assembling the patches. We shall try to obtain a continuous field of parameters with the values at the integration points sufficiently close to the original ones. Therefore the produced continuous field of parameters will satisfy the properties proposed in [4].

The idea is very simple and is based on using some appropriate weight functions in passing polynomials over the integration points. Although we do not evaluate the nodal values, we use these points to construct the patches. For this, we use all nodal points and in addition to corner nodes, in quadratic elements for example, the edge nodes and central nodes will be used. Typical forms of such patches for nine node elements are shown in Figure 2.

For simplicity we use σ as a representative of history dependent parameters. The process of evaluating a continuous field of σ is similar to what we had before but with a small difference. In order to recover the values at integration points close enough to their original former values, we shall use an appropriate weight function inside the minimization process:

$$\Pi = \sum \tilde{\omega}(\sigma^* - \sigma^g)^2 \quad \sigma^* = Pa \quad (30)$$

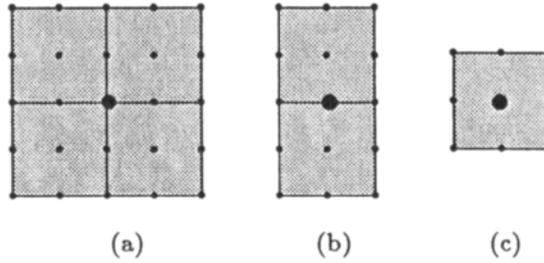


Figure 2: Typical patches required in direct Gauss point to Gauss point transfer operator (a) patch for corner nodes (b) patch for edge nodes (c) patch for inside nodes.

The weight function $\tilde{\omega}$ is such that it has a comparatively large value at the points near the patch node (where the patch is assembled). For this we define a normalized continuous weight function for each patch node having unit value near the patch node and very small value elsewhere. To construct such function we allocate a part of element area to each Gauss point so that $\tilde{\omega} \approx 1$ at the position of Gauss point. The area allocated to the Gauss point can be determined according to the weight of the Gauss point in numerical integration process.

Having such weight function, we minimize Π with respect to \mathbf{a} :

$$\frac{\partial \Pi}{\partial \mathbf{a}} = 0 \quad (31)$$

which leads to a system of equation with series of unknowns as \mathbf{a} . Having solved this system of equation, the new distribution of the parameter can be determined as $\sigma^* = \mathbf{Pa}$ which is suitable for points very close to the patch node.

Now we assume that the new Gauss point is inside element number i . The above mentioned procedure can be performed for all nodes of the element. Therefore we need to determine the local relative position of the new Gauss point to all other old Gauss points inside the element in order to recognize which one of these patches are suitable for computing the new Gauss point value. This seems to be impractical for an adaptive procedure in which we need to perform the transfer procedure for many times. However there is a simple solution for this problem. Suppose that for each node j of the element (element i) we have a set of unknown coefficients \mathbf{a}_j which can define the variation of the parameter near node j , namely:

$$\sigma_j^* = \mathbf{Pa}_j \quad (32)$$

The value at the new Gauss point $\tilde{\sigma}$ can be evaluated by a weight form of averaging:

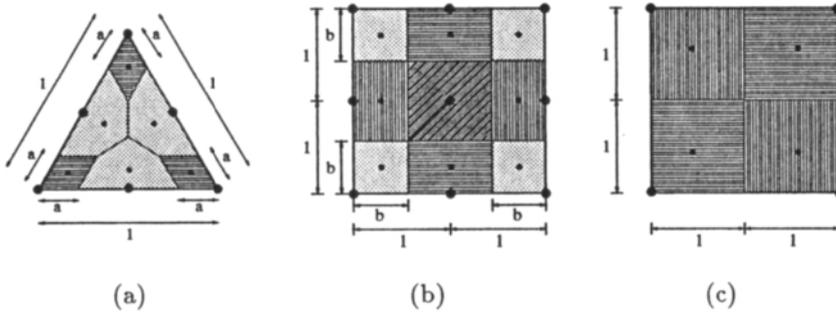


Figure 3: The allocated areas to Gauss points (according to their weights) for different elements. (a) quadratic triangles elements $a = 0.234469341$. (b) quadratic quadrilateral elements $b = 0.555555556$. (c) linear quadrilateral elements

$$\bar{\sigma}(\sum_{j=1}^N \omega_j) = \sum_{j=1}^N \omega_j \sigma_j^* \quad (33)$$

or:

$$\bar{\sigma} = \frac{\sum_{j=1}^N \omega_j \sigma_j^*}{\sum_{j=1}^N \omega_j} \quad (34)$$

Here N is the number of nodes of the element and ω_j is associated weight function of node j varying smoothly from $\omega_j \approx 0$, for points far from the node, to $\omega_j \approx 1$ for points close to the node. Generally this weight ω_j can be different from $\tilde{\omega}$ used in Equation (30). But we can use:

$$\tilde{\omega} = \max(\varepsilon, \omega) \quad (35)$$

with ε being very small value ($\varepsilon \ll 1$). Using $\tilde{\omega}$ as Equation (35) in minimization process of (31) gives the opportunity of using all Gauss points in the patch when $\omega = 0$ for some of them.

The allocated areas of different integration points, according to their weights, for some elements widely used in plasticity are shown in Figure 3. Here we introduce these weight functions at element level for various forms of elements:

(a) For bilinear elements:

$$\begin{aligned} \omega_1 &= \phi_1(\xi + 1, 0, 1, 0, n) \cdot \phi_1(\eta + 1, 0, 1, 0, n) \\ \omega_2 &= \phi_2(\xi + 1, 0, 1, 0, n) \cdot \phi_1(\eta + 1, 0, 1, 0, n) \\ \omega_3 &= \phi_2(\xi + 1, 0, 1, 0, n) \cdot \phi_2(\eta + 1, 0, 1, 0, n) \\ \omega_4 &= \phi_1(\xi + 1, 0, 1, 0, n) \cdot \phi_2(\eta + 1, 0, 1, 0, n) \end{aligned} \quad (36)$$

Here ξ, η are normalized coordinates. Node numbering begins from $\xi = -1, \eta = -1$ with anti-clockwise direction and:

$$\begin{cases} \phi_1(\xi_0, \eta_0, \lambda, \mu, n) = \frac{1}{2}(1 + \chi) & \text{if } 2(\lambda + \mu\eta_0) - \xi_0 \geq 0 \\ \phi_1(\xi_0, \eta_0, \lambda, \mu, n) = 0 & \text{if } 2(\lambda + \mu\eta_0) - \xi_0 < 0 \end{cases} \quad (37)$$

$$\begin{cases} \phi_2(\xi_0, \eta_0, \lambda, \mu, n) = \frac{1}{2}(1 - \chi) & \text{if } 2(\lambda + \mu\eta_0) - \xi_0 \geq 0 \\ \phi_2(\xi_0, \eta_0, \lambda, \mu, n) = 1 & \text{if } 2(\lambda + \mu\eta_0) - \xi_0 < 0 \end{cases} \quad (38)$$

in which

$$\chi(\xi_0, \eta_0, \lambda, \mu, n) = \operatorname{sgn}\left(\cos\left(\frac{\pi\xi_0}{2(\lambda + \mu\eta_0)}\right)\right) \left| \cos\left(\frac{\pi\xi_0}{2(\lambda + \mu\eta_0)}\right) \right|^{\frac{1}{n}} \quad (39)$$

(b) For biquadratic elements:

$$\begin{aligned} \omega_1 &= \phi_1(\xi + 1, 0, \alpha, 0, n) \cdot \phi_1(\eta + 1, 0, \alpha, 0, n) \\ \omega_2 &= \phi_1(|\xi|, 0, \beta, 0, n) \cdot \phi_1(\eta + 1, 0, \alpha, 0, n) \\ \omega_3 &= \phi_1(1 - \xi, 0, \alpha, 0, n) \cdot \phi_1(\eta + 1, 0, \alpha, 0, n) \\ \omega_4 &= \phi_1(1 - \xi, 0, \alpha, 0, n) \cdot \phi_1(|\eta|, 0, \beta, 0, n) \\ \omega_5 &= \phi_1(1 - \xi, 0, \alpha, 0, n) \cdot \phi_1(1 - \eta, 0, \alpha, 0, n) \\ \omega_6 &= \phi_1(|\xi|, 0, \beta, 0, n) \cdot \phi_1(1 - \eta, 0, \alpha, 0, n) \\ \omega_7 &= \phi_1(\xi + 1, 0, \alpha, 0, n) \cdot \phi_1(1 - \eta, 0, \alpha, 0, n) \\ \omega_8 &= \phi_1(\xi + 1, 0, \alpha, 0, n) \cdot \phi_1(|\eta|, 0, \beta, 0, n) \\ \omega_9 &= \phi_1(|\xi|, 0, \beta, 0, n) \cdot \phi_1(|\eta|, 0, \beta, 0, n) \end{aligned} \quad (40)$$

Where $\alpha = 0.555555556, \beta = 1 - \alpha$. Here again node numbering begins from $\xi = -1, \eta = -1$ with anti-clockwise direction. Functions ϕ_1, ϕ_2, \dots are similar to case (a).

(c) For quadratic triangular elements:

$$\begin{aligned} \omega_1 &= \phi_1(\xi, \eta, \gamma, 0, n) \cdot \phi_1(\eta, \xi, \gamma, 0, n) \\ p_2 &= \phi_1(\eta, \zeta, \alpha, \beta, n) \cdot \phi_2(\zeta, \xi, \alpha, \beta, n) \\ q_2 &= \phi_2(\xi, \eta, \alpha, 0, n) \cdot \phi_2(\zeta, \eta, \alpha, 0, n) \\ \omega_2 &= p_2 \cdot q_2 \\ \omega_3 &= \phi_1(\zeta, \eta, \gamma, 0, n) \cdot \phi_1(\eta, \zeta, \gamma, 0, n) \\ p_4 &= \phi_1(\zeta, \xi, \alpha, \beta, n) \cdot \phi_2(\xi, \eta, \alpha, \beta, n) \\ q_4 &= \phi_2(\eta, \zeta, \alpha, 0, n) \cdot \phi_2(\xi, \zeta, \alpha, 0, n) \\ \omega_4 &= p_4 \cdot q_4 \\ \omega_5 &= \phi_1(\xi, \zeta, \gamma, 0, n) \cdot \phi_1(\zeta, \xi, \gamma, 0, n) \\ p_6 &= \phi_1(\xi, \eta, \alpha, \beta, n) \cdot \phi_2(\eta, \zeta, \alpha, \beta, n) \\ q_6 &= \phi_2(\zeta, \xi, \alpha, 0, n) \cdot \phi_2(\eta, \xi, \alpha, 0, n) \\ \omega_6 &= p_6 \cdot q_6 \end{aligned} \quad (41)$$

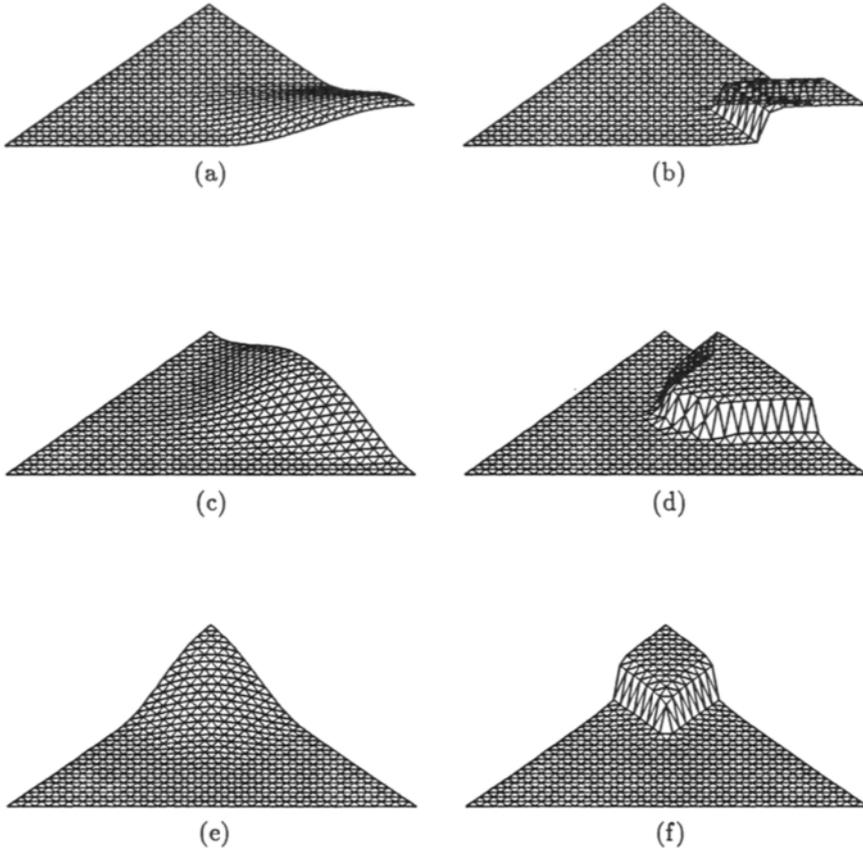


Figure 4: Variation of suggested weight functions at element level in direct Gauss point to Gauss points data transfer for corner and edge nodes of quadratic triangular elements. (a)/(c)/(e) Weight functions with $n=1$, (b)/(d)/(f) Weight functions with $n=10$.

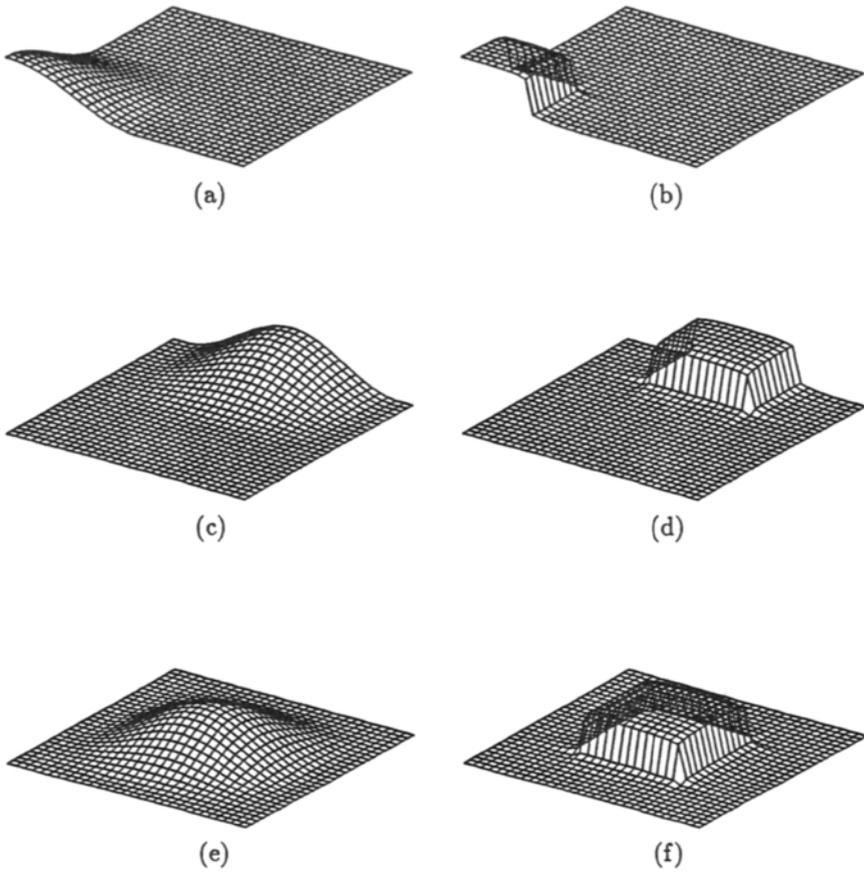


Figure 5: Variation of suggested weight functions at element level in direct Gauss point to Gauss points data transfer for corner, edge, and middle nodes of quadratic quadrilateral elements. (a)/(c)/(e) Weight functions with $n=1$, (b)/(d)/(f) Weight functions with $n=10$.

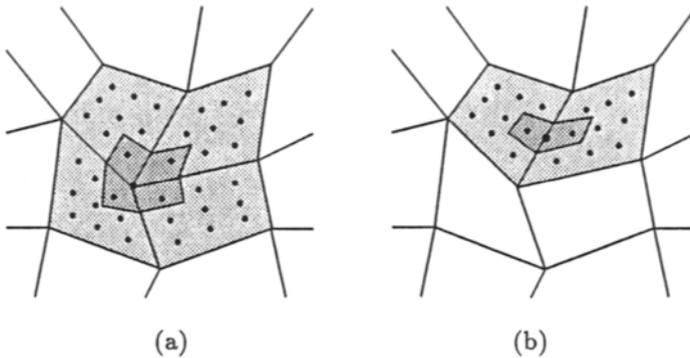


Figure 6: The allocated area to each nodal point (a) typical patches on corner nodes (b) typical patches on edge nodes.

Where $\alpha = 0.5$, $\beta = -0.5$, $\gamma = 0.234469341$ and $\zeta = 1 - \xi - \eta$. Here ξ , η and ζ are area coordinates and node numbering begins from $\xi = 0$, $\eta = 0$ with anti-clockwise direction. Functions ϕ_1 , ϕ_2 , .. are similar to case (a).

Figures (4) to (5) show the variation of proposed weight functions for different elements.

It should be mentioned that there is no need of special treatment for boundary patches and they can consist of even one element. To balance the number of unknowns and equations in (39), the polynomial order can be chosen automatically so that the number of coefficients in the polynomial is always less than that of elements times the number of Gauss points². However, in order to have a cheap operator linear polynomial may be used.

One important point should be mentioned here is that, according to Equation (34) evaluation of information at each new Gauss point needs an averaging process for values from the patches made by all nodes of the old mesh. To avoid repetition of the least square procedure for a node, since there may be several new Gauss points inside a particular element, a three step procedure can be followed:

- 1 - Using an efficient search algorithm, the old element containing a new Gauss point is identified and the local coordinates of the point are computed. This must be performed for all Gauss points so that at the end of this step those new Gauss points falling inside a particular element are identified (and visa versa).

²Some additional points can be considered in each element in order to use a high order polynomial in the least square procedure. In this case the corresponding information of the new point can be obtained by a simple relation as (34) with N being the number of Gauss points in the element.

The reader will notice that, in the case that the new element size is less than (or nearly equal to) the old element size, by identifying the old element of the first Gauss point of the new element, the procedure for the rest of the Gauss points becomes quite fast and the search needs only to be performed on old elements connected to the first element.

- 2 - The least square procedure of (30) and (31) is performed for all nodes of the old mesh and the contribution of each node to the new Gauss points inside the elements connected to the node is computed for both weight ω_j and the state variable $\omega_j \sigma_j^*$ (viz Equation (33)) and added to the current values.
- 3 - The values stored at Gauss points are divided by the corresponding summation of the weights for all new Gauss points (Equation (33)).

In the following section we briefly explain the procedure of element identification (the first step) for a new node or Gauss point.

3.2.3. ELEMENT IDENTIFICATION

Supposing that the local position of a point with global coordinate $\mathbf{X}_p = [x_p, y_p]^T$ is required to be determined in an old mesh, a simple two step scheme can be employed as:

- 1 - We find the closest node to the point in the old mesh and consider all elements connected to the node (the efficient algorithm will be reported in the future publication).
- 2 - By an inverse mapping we find the local coordinates of the point in each element. The search for the element will stop when the local coordinates of the point are within an appropriate range (for instance in quadrilateral elements this range is $[-1, 1]$ and in triangular elements $[0, 1]$). The inverse mapping is very simple and, except for triangular elements with straight edges, the local coordinates can be found through a small Newton-Raphson iteration. For this, in each element, we write;

$$\mathbf{G} = \mathbf{X}_p - \sum \mathbf{N}_i \mathbf{X}_i = 0 \quad (42)$$

Where \mathbf{N} are interpolation functions in terms of local coordinates ξ and η . We can start from $\xi = 0$ and $\eta = 0$ and write linearized form of Equation (42) for k th iteration as:

$$\mathbf{G} = \mathbf{G}_k + \left[\frac{\partial \mathbf{G}}{\partial \mathbf{x}} \right]_k \delta \mathbf{x}_k = 0 \quad (43)$$

or:

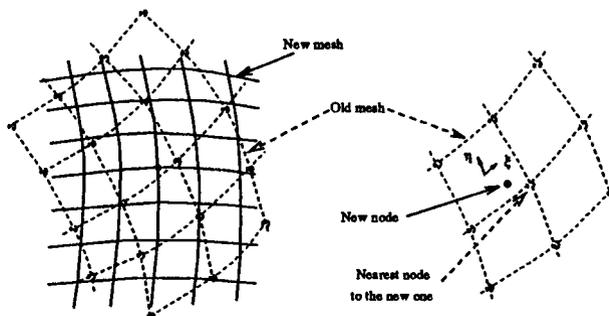


Figure 7: The position of each new node is determined in the old mesh by finding the nearest old node and using inverse mapping to calculate the local coordinates of the point in the elements connected to the old node.

$$\mathbf{G} = \mathbf{G}_k + \mathbf{J}_k \delta \mathbf{x}_k = 0 \quad (44)$$

with $\mathbf{x} = [\xi, \eta]^T$ and \mathbf{J} being the well known Jacobian matrix in the FEM formulation. Solving for \mathbf{x} gives:

$$\delta \mathbf{x}_k = -\mathbf{J}_k^{-1} \mathbf{G}_k \quad (45)$$

Now the local coordinates can be computed as:

$$\mathbf{x} = \sum \delta \mathbf{x}_k \quad (46)$$

Once the local position of the point is obtained the data can be transferred either by interpolation functions, for nodal values, or direct Gauss point to Gauss point operator described in Section 3.2.1.

3.2.4. THE STEP BY STEP ADAPTIVE PROCEDURE

Here we give the general form of step by step adaptive procedure in a typical non-linear analysis.

We consider a generic load step as step number n :

- 1 - At the end of the step, the continuous fields of stresses σ and incremental stress $\Delta \epsilon$ are evaluated using the REP and SPR procedures, respectively.
- 2 - The incremental error is computed by Equation (21) for all elements as well as the whole domain.
- 3 - If the error is within a prescribed limit, the non-linear analysis will be continued for the next increment.

4 - If the error exceeds a prescribed limit then;

- 4-a new mesh sizes are evaluated by relation (28) at each vertex of the old mesh and
- 4-b all state variables for increment $n - 1$ at nodes and integration points are stored as a reference values for subsequent remeshings,
- 4-c a new mesh will be generated by the new size of elements available at nodes,
- 4-d the information at old nodes and Gauss points are transferred to the new ones by the described procedure in the preceding sections,
- 4-e the new mesh and the new information are considered as the starting point of increment n and therefore the corresponding non-linear analysis is performed on the new mesh.

In order to reduce the effect of data transfer operator on the incremental error (note that $\Delta \epsilon = \mathbf{B}[\mathbf{u}_n - \mathbf{u}_{n-1}]$ and that \mathbf{u}_{n-1} has already been transferred from the old mesh to the new one), the new increment can be divided into two parts and the error can be evaluated for the second part. This can be done by dividing the load increment $\delta \mathbf{f}_n$ to $\alpha \delta \mathbf{f}_n$ and $(1 - \alpha) \delta \mathbf{f}_n$. If α is considered to be zero ($\alpha = 0$) the procedure will convert to equilibration of the state variables at the beginning of the load step (after transferring the data) resulting to a new state of these variables which not only satisfy the equilibrium condition but also the constitutive relations.

- 4-f If the error at the end of the n th step, for the new mesh, is less than the prescribed limit, the non-linear analysis will be continued for the next increment, otherwise the procedure will be repeated from 4-c noting that the new state variables can be transferred from the first old mesh (the state variable of which was stored in stage 4-b).

In the following section some numerical examples, solved by above procedure, are presented.

3.3. NUMERICAL EXAMPLES OF ADAPTIVE REFINEMENT

Example 1 A perforated tensile specimen with displacement control is considered as an example. The problem and the material properties are shown in Figure 8. A comparatively fine mesh with quadratic elements has been used as a benchmark and plays the role of exact solution. In order to study the discretization errors and reduce the effect of errors due to magnitude of the increments, in both fine solution and adaptive ones, identical load steps are used.

Assuming that the reactions obtained by the solution of the fine mesh are very close to exact ones, exact incremental energy error can be obtained as:

$$e_{ex} = (|P_{ex}^n \delta - P^n \delta|)^{\frac{1}{2}} \quad \eta_{ex} = \frac{e_{ex}}{|P_{ex}^n \delta|^{\frac{1}{2}}} = \left(\frac{|P_{ex}^n - P^n|}{|P_{ex}^n|} \right)^{\frac{1}{2}} \quad (47)$$

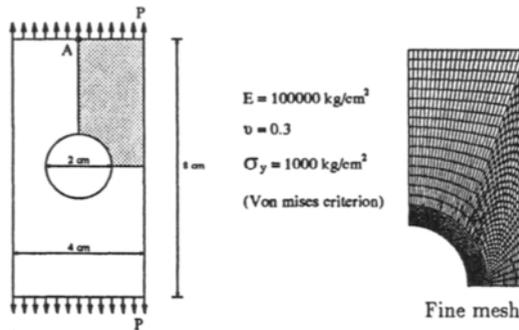


Figure 8: A tensile specimen and the mesh used for so called exact solution.

in which P_{ex}^n , P^n and δ are exact reaction (from the fine mesh), approximate reaction (from a typical mesh in adaptive procedure) and the corresponding deformation at an instant during the increment n , respectively. For convenience the reactions at the end of increment can be used to achieve a rough measure of the exact error.

Figures 9 and 10 demonstrate the adaptive error control procedure starting from a very coarse mesh. In both figures quadratic triangular elements are used and the aim of the adaptive procedures is to achieve 5% relative incremental error. After each step of remeshing the data have been transferred from the old mesh to the new one. These two figures clearly show success of the suggested methodology.

The main difference between Figure 9 and 10 is the number of pressure variables in the plastic solution. In Figure 9, six node elements ($p = 2$) with one pressure variable are used. Although the elements are quadratic, the behavior is very similar to linear elements and this, of course, affects the rate of convergence in the adaptive solution. In Figure 10, six node elements with three pressure variables and bubble function are used. The effect of the number of pressure variables, for those problems that locking phenomenon is predominant, can be considered in the refinement criterion (28) by an engineering judgement.

Example 2 The second example is a slope carrying a rigid footing with a point load and solved by assuming the plane strain condition. The geometry and material properties are shown in Figure 11. The point load is in fact induced by vertical control of displacement of the footing. The elasticity modulus of the footing is considered to be 100 times as much as that of the slope. Forty equal displacement steps are considered to achieve the maximum displacement $u = 1.0$ at the restrained point on the footing. Plastic solution of a very fine mesh with 1080 quadratic (9 node) elements and 4455 node is considered as a benchmark for the comparisons.

Figure 12 shows the result of the proposed adaptive procedure starting from a very coarse mesh (Mesh A). The aim of this adaptive analysis is to achieve less than 15%

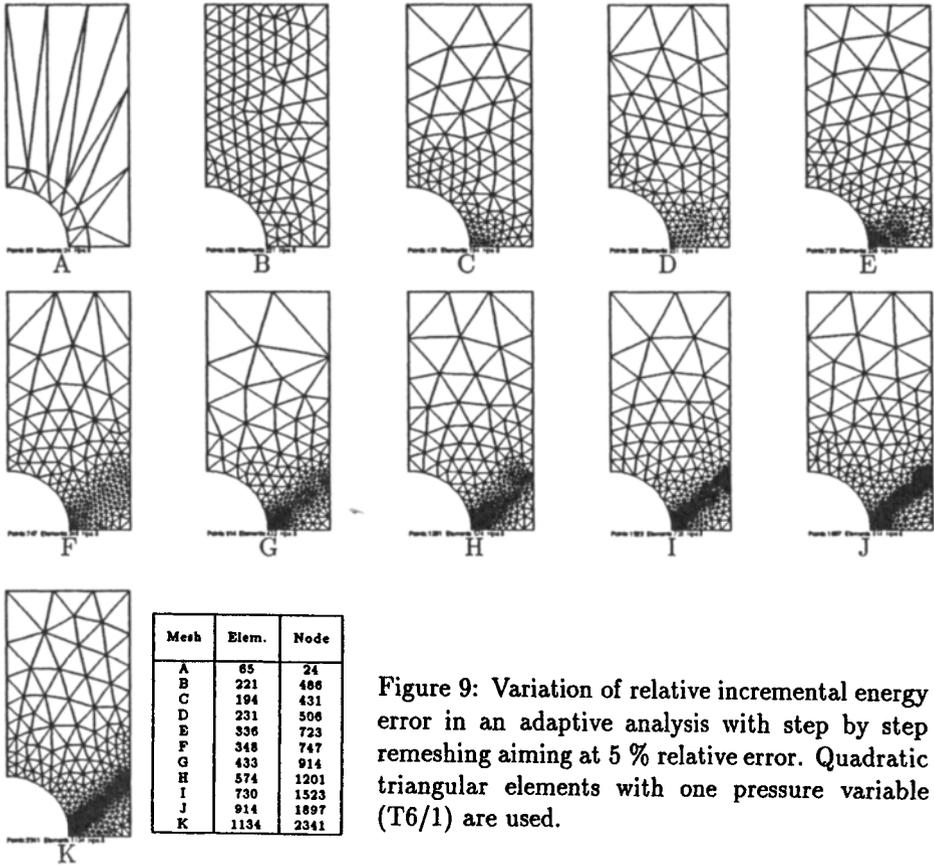
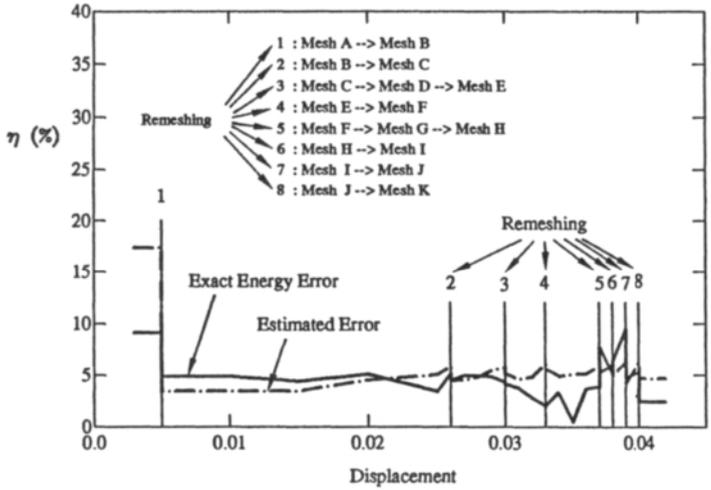


Figure 9: Variation of relative incremental energy error in an adaptive analysis with step by step remeshing aiming at 5 % relative error. Quadratic triangular elements with one pressure variable (T6/1) are used.

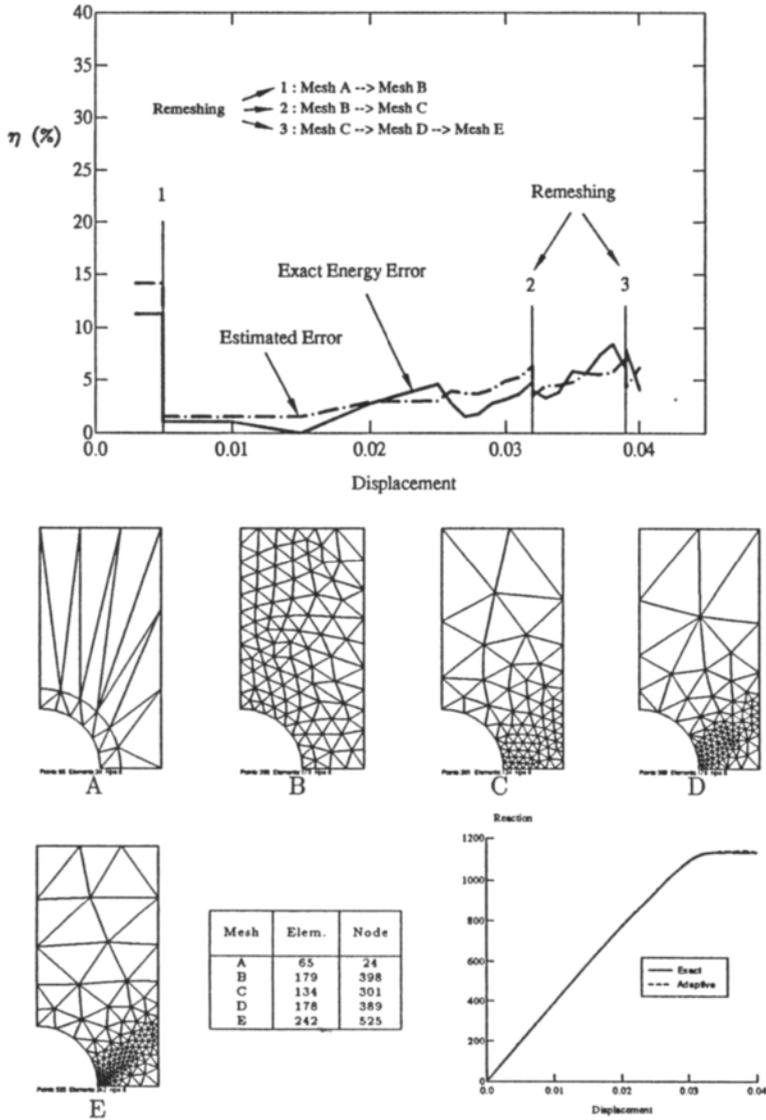


Figure 10: Variation of relative incremental energy error in an adaptive analysis with step by step remeshing aiming at 5 % relative error. Quadratic triangular elements with bubble function and three pressure variables (T6/3B) are used.

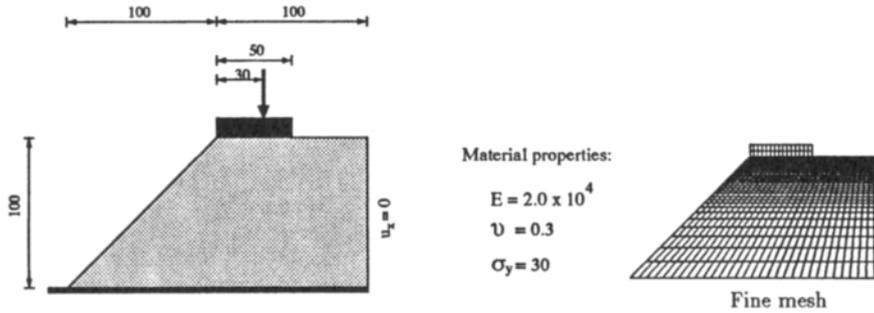


Figure 11: Geometry and material properties of an elasto-plastic slope.

incremental energy error in order to have less than 2.5% error in ultimate load. However, because of the singularity at the corner of the footing, in the first few steps we aimed at 5% energy error in order to localize the singularity effect. The plastic behaviour starts from the first load step in a very small area around the singular point. Meshes B to H shown in this figure are the representatives of the meshes used in the adaptive analysis. The figure also demonstrates the behaviour of the coarse and the fine meshes in a simple plastic solution without adaptive procedure. As expected, the solution of the coarse mesh shows a very stiff behaviour. Although the fine mesh with nine node elements shows a much improved behaviour, it can be seen that this mesh is not an optimal mesh and the solution obtained cannot be considered as an exact solution. The problem is solved again from the first load step using the last mesh (named as 'Optimal Mesh') to assess the final results of the adaptive procedure. It can be seen that this mesh shows even better behaviour than the fine mesh though the number of degrees of freedom is less. Comparison of $P - \delta$ diagram of the 'Optimal mesh' and that of the adaptive analysis clearly show an excellent performance of the proposed adaptive procedure.

4. CONCLUSION

In this paper we have presented, in detail, an adaptive procedure suitable for non-linear elasto-plastic problems using recovery techniques as both error estimation and transfer operator. Although the procedure has been described for plasticity problems, it can be used in general non-linear problems.

An incremental energy error norm has been employed to estimate the error. In principle, the procedure follows the same steps as linear cases described in Part I of this paper. In this way two recovery methods, the REP and SPR, are used to recover the total stresses and incremental strains used in the error definition, respectively. These two recovery methods are proved to be most robust in linear cases.

The problem of transfer of data from the old mesh to the new mesh has been also addressed in this paper and a new transfer operator using weighted form of least square

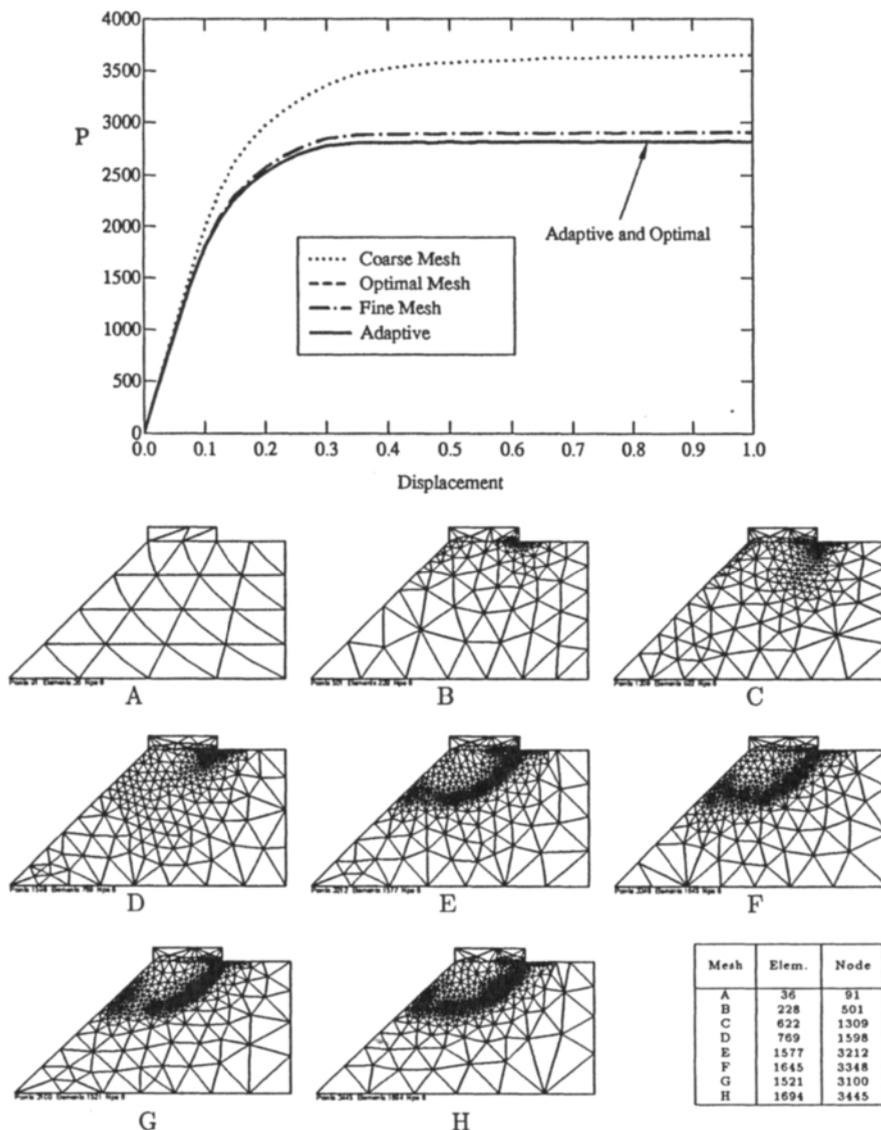


Figure 12: $P - \delta$ diagram of elasto-plastic slope aiming at 2.5 % error in ultimate load (15% incremental energy error) using quadratic triangular elements (T6/3B); Mesh A: $u = 0.0$ (coarse mesh), Mesh B: $u = 0.025$, Mesh C: $u = 0.15$, Mesh D: $u = 0.3$, Mesh E: $u = 0.45$, Mesh F: $u = 0.6$, Mesh G: $u = 0.75$, Mesh H: $u = 0.9$. The last mesh (Mesh H named as 'Optimal Mesh') is used for the solution of the problem from the first load step without further refinement. The Fine structured mesh of Fig. 11 consists of 1080 quadratic (9 node) elements with 4455 nodes.

procedure, in the same sense as the recovery procedure, has been suggested. The weight functions are defined at element level for each Gauss point according to the weight of the Gauss point in the numerical integration process. This makes the procedure capable of transferring the information directly from old Gauss points to the new ones. The main advantage of this method is that it prevents the spread of the local information over a large area when several remeshing is required in the adaptive analysis. In order to use such data transfer operator, compatible algorithms for element identification and inverse mapping have been explained.

Performance of the combination of all suggested procedures has been demonstrated on two elasto-plastic problems from which excellent results have been obtained indicating the high efficiency of the procedure.

REFERENCES

- [1] O.C. Zienkiewicz, B. Boroomand and J.Z. Zhu, Recovery procedures in error estimation and adaptivity, Part I : adaptivity in linear problems.
- [2] H. Jin, L. Bernspang, R. Larsson and N.E. Wiberg, Mesh generation and mesh refinement procedures for computational plasticity, In *Computational Plasticity II: Models, Software and Applications*, (edited by D.R.J Owen et al.), 347-357, Pineridge Press, Swansea (1989)
- [3] A. Samuelsson and N.E. Wiberg, Finite element adaptivity in dynamics and elasto-plasticity, In *The finite element method in the 1990's*, O.C Zienkiewicz's anniversary volume, (edited by E. Onate et al.), 152-162 CIMNE and Springer-Verlag, Barcelona (1991)
- [4] M. Ortiz and J.J. Quigley, IV, Adaptive mesh refinement in strain localization problems, *Com. Meth. Appl. Mech. Eng.*, 90, (1991), 781-804.
- [5] C. Johnson and P. Hansbo, Adaptive finite element methods in computational mechanics, *Com. Meth. Appl. Mech. Eng.*, 101, (1992), 143-181.
- [6] T. Belytschko, M. Tabbara, H-Adaptive finite element methods for dynamic problems, with emphasis on localization, *Int. J. Num. Meth. Eng.*, 36, (1993), 4245-4265.
- [7] N.S. Lee and K.J. Bathe, Error indicators and adaptive remeshing in large deformation finite element analysis, *Finite Elem. Anal. Design.*, 16, (1994), 99-139.
- [8] D. Peric, J. Yu and D.R.J. Owen, On error estimates and adaptivity in elastoplastic solids: application to the numerical simulation of strain localization in classical and Cosserat continua, *Int. J. Num. Meth. Eng.*, 37, (1994), 1351-1379.
- [9] Ch. Hochard, D. Peric, M. Dutko and D.R.J. Owen, Transfer operators for evolving meshes in elasto-plasticity, *Proceeding of Computational plasticity fundamentals and applications*, 2331-2339, Barcelona (1995)

- [10] O.C. Zienkiewicz, M. Huang and M. Pastor, Localization problems in plasticity using finite elements with adaptive remeshing, *Int. J. for Num. & Anal. Meth. in Geomechanics*, 19, (1995), 127-148.
- [11] O.C. Zienkiewicz, M. Pastor and M. Huang, Softening, localization and adaptive remeshing. Capture of discontinuous solution, *Computational Mechanics*, 17, (1995), 98-106.
- [12] L. Gallimard, P. Ladevèze and J.P. Pelle, Error estimation and adaptivity in elastoplasticity, *Int. J. Num. Meth. Eng.*, 39, (1996), 189-217.
- [13] D. Peric, Ch. Hochard, M. Dutko and D.R.J. Owen, Transfer operators for evolving meshes in small strain elasto-plasticity, *Com. Meth. Appl. Mech. Eng.*, 137, (1996), 331-344.
- [14] A. Selman, E. Hinton and N. Bićanić, Adaptive mesh refinement for localised phenomena, *Computers and Structures*, 63, No. 3, (1997), 475-495.
- [15] G.T. Camacho, M. Ortiz, Computational modelling of impact damage in brittle materials, *Int. J. Solids Struct.*, 33, (1996), 2899-2938.
- [16] G.T. Camacho, M. Ortiz, Adaptive lagrangian modelling of ballistic penetration of metallic targets, *Comp. Meth. Appl. Mech. Eng.*, 142, (1997), 269-301.
- [17] O.C. Zienkiewicz, B. Boroomand, Adaptive mesh updating in plasticity problems and an improved solution process using SPR. A preliminary study, *Advances in finite element technology*, N. Wiberg, 1-21 (1995)
- [18] O.C. Zienkiewicz, J.Z. Zhu, The superconvergent patch recovery and *a posteriori* error estimates. Part 1: The recovery technique, *Int. J. Num. Meth. Eng.*, 33, (1992), 1331-1364.
- [19] B. Boroomand, O.C. Zienkiewicz, Recovery by equilibrium in patches (REP), *Int. J. Num. Meth. Eng.*, 40, (1997), 137-164.
- [20] B. Boroomand, O.C. Zienkiewicz, An improved REP recovery and the effectivity robustness test, *Int. J. Num. Meth. Eng.*, 40, (1997), 3247-3277.
- [21] O.C. Zienkiewicz, R.L. Taylor, *The Finite Element Method*, IV Edition, Vol. I (1989), Vol. II (1991)
- [22] R.L. Taylor, J.C. Simo, A return mapping algorithm for plane stress elasto-plasticity, *Int. J. Num. Meth. Eng.*, 22, (1986), 649-670.
- [23] M. Ortiz, J.C. Simo, An analysis of a new class of integration algorithms for elasto-plastic constitutive relations, *Int. J. Num. Meth. Eng.*, 23, (1986), 353-366.
- [24] T.R.J. Hughes, Generalization of selective integration procedures to anisotropic and non-linear media, *Int. J. Num. Meth. Eng.*, 15, (1980), 1413-1418.

- [25] J.C. Simo, R.L. Taylor, K.S. Pister, Variational and projection methods for the volume constraint in finite deformation plasticity, *Comp. Meth. Appl. Mech. Eng.*, 51, (1985), 177-208.
- [26] O.C. Zienkiewicz, Y.C. Liu and G.C. Huang, Error estimates and convergence rates for various incompressible elements, *Int. J. Num. Meth. Eng.*, 28, (1989), 2191-2202.
- [27] O.C. Zienkiewicz, J.Z. Zhu, A simple error estimator and adaptive procedure for practical engineering analysis, *Int. J. Num. Meth. Eng.*, 24, (1987), 337-357.
- [28] A.R. Daiz, P. Kikuchi and J. E. Taylor, Design of an optimal grid for finite element methods, *J. Struct. Mech.*, 11, (1983)
- [29] P. Ladevèze, G. Coffignal, J.P. Pelle, Accuracy of elastoplastic and dynamic analysis, *Accuracy Estimates and Adaptive Refinements in Finite Element Computations* I. Babuška, O.C. Zienkiewicz, J. Gago and E.R. de A. Oliveria. Chapter 11 (1986)
- [30] E. Hinton and J.S. Campbell, Local and global smoothing of discontinuous finite element functions using a least squares method, *Int. J. Num. Meth. Eng.*, 8, (1974), 461-480.
- [31] G. Loubignac, G. Cantin and G. Touzot, Continuous stress fields in finite element analysis, *AIAA J.*, 15, (1977), 1645-1647.

PART 5

ADAPTIVE COMPUTATIONAL METHODS FOR 3D LINEAR PROBLEMS

This Page Intentionally Left Blank

3-D error estimation and mesh adaptation using improved R.E.P. method

Pierre Beckers and Emilio Dufeu

Laboratoire de Techniques Aéronautiques et Spatiales, Université de Liège, 21, rue E. Solvay, 4000 Liège, Belgium

SUMMARY

In 3-D finite element analyses, error estimation and mesh adaptation are absolutely necessary to warranty the reliability and the quality of the results. To achieve this goal, suitable error estimators have to be developed. They must be able to provide a good *a posteriori* estimation of the obtained precision for any kind of models, especially for 3-D structures involving both solid parts and thin regions where the mesh is very coarse in one direction. New techniques have been developed to fulfil these requirements.

1. BASIC METHOD

The first step in the estimation of the precision of a finite element method solution is generally to build a comparison stress field that will be assimilated to the exact one. According to the name given in one of the standard techniques of recovering this field, we will call it the smoothed stress field. The most difficult step of the whole procedure is to obtain a sufficiently good smoothed stress field. For displacement models, the stress components should be continuous and satisfy the equilibrium equations. Focusing the recovering procedure on the second condition we obtain a first family of estimators like the models developed in Cahan [1-2], or on the first condition, and we obtain a S.P.R. like formulation [3]. Recently, two attempts have been performed in order to base the stress recovering on both methods. Their names are respectively R.E.P. (Recovering by Equilibrium in Patches [4]), and I.P.R. (in french recovering by weighted integration [5]).

In this method, some of the weighted equilibrium equations present in a kinematically admissible model and applied to the finite element stresses are used to project the smoothed stress field on the F.E.M. stress field. The idea is to force this smoothed stress field to respect the equilibrium equations in a weak sense. In particular situations such as uniform rectangular first degree meshes, this method is identical to the S.P.R. method where smoothed and F.E.M. fields are identified at the superconvergent points.

This new method can be seen as a generalization of the S.P.R. one, but with the fundamental advantage that it can be applied to any displacement model. In the begin of the finite element software development it was proposed in many programs to use mean values for element stresses. The use of other averages was proposed later [6], but only for postprocessing

purposes and with the drawback that this procedure was increasing the difficulty of interpreting higher order stress averages. The main interest of the new method is to apply this concept for building the smoothed stress field directly on a patch so as to transform averages into local values easier to process later.

Tests performed in 2-D [4] and in 3-D [5, 7] show that the new method is very efficient when it works inside the domain but that it must still be improved close to the boundaries of the model or when the structure is very thin.

The extension of the method toward the loaded boundaries is rather obvious if we apply the same idea of making equal weighted averages of applied surfaces traction and of the smoothed stress field. Here the weight functions are the displacement shape functions used for defining the generalized forces. With respect to method [1], the main advantage of this procedure is that it can deal with any kind of boundary condition on surface traction.

The second ingredient of the procedure is now classical, it consists in choosing suitable patches centred either on nodes or on elements. Some strategies have been set up to have a good compromise between quality and computing cost. The last step is not dependent of the recovering procedure, standard techniques are used to compute the energy norm of the error, the convergence rate of the solution and finally to adapt the mesh [8]. The described methods have been implemented in a finite element method postprocessor and tested in 3-D [7].

2. BACKGROUND

The main advantage of the finite element method is its ability to model a wide variety of structures or components and to adapt easily to any particular situation. With the development of efficient algorithms to compute the solution of linear or non linear systems of equations, the method is virtually able to solve any engineering problem in the field of structural mechanics.

However, the counterpart is the tedious and very time consuming task for defining the input of the model and analyzing the output of the simulation. When the finite element model is very small, an interesting but difficult study of the model before and after process, aided by some engineer judgment, coming from experience, makes possible a very fruitful understanding of the behaviour of the structure and should the mesh be coarse or not, it is possible to obtain useful results.

Today, because of the still increasing power of the computers there is no difficulty to handle very fine meshes, but the problem is to generate these models and to interpret the results.

2.1. Modelling step

For the step of generation, if the designer is concerned with the analysis of homogeneous and regular domains, he can use automatic mesh generation algorithms and obtain very easily a refined mesh. These programs are efficient and reliable in 2-D and they also become sufficiently robust to generate 3-D models.

But the modelling of an object still remains difficult if we must analyze complex structures involving different components with sophisticated joins. Here, specific tools are necessary, they must be able not only to perform the job automatically but also to help the user to check that the mesh fulfil the basic requirements of the finite element theory.

Many research teams and software companies are trying to identify rules and develop new

programs involving as basic tools, classical mesh generation but also C.A.D. mechanisms able to simplify the geometry, to remove details and to check that the essential characteristics of the structure are preserved: mainly the global stiffness, the total mass and their distribution in the whole structure.

The idea is to make sure that the forces are acting in the model in the same way that in the true object and that a first analysis will provide sufficiently reliable results in the local components so as to allow detailed analyses. These complementary simulations have to be made in order to give the information necessary to design or to optimize the small parts of the structure and to eventually detect local accidents that can produce structural damages.

All these developments are assumed to replace or to help the engineer skills in order to produce a suitable analysis model.

2.2. Analysis step

The second difficulty that has to be overcome by using fine meshes is the interpretation of the results and the identification of important information coming from the model.

The recent developments of computer graphics tools help to produce easily and quickly high quality graphics of the output. Many techniques are available to represent all the state variables of the problem, mainly in mechanical engineering: displacement, strains, stresses or equivalent variables. This postprocessing step can however be dangerous because it implicitly involves some transformation of the output, like smoothing, mapping of values into colours, etc.

The designer must then absolutely know, not only how the finite element model was built and how is working the finite element package, but he also must know what type of postprocessing was used to lead to the final results.

And finally he must know what is the reliability of the results and what is the influence of the initial hypotheses made on the model.

It is important at this step to note that when using very simple model, it is easy to repeat analyses, to make small modifications or to test variants of the model. These actions are giving to the user some feedback of the numerical simulation and help him to better know the behaviour of the model. It is the reason why it was possible to obtain useful results with very coarse meshes twenty years ago. At that time, when simulating some mechanical behaviour with a finite element model some effort was made to interpret the results, a very coarse mesh was sufficient and its reliability was strongly depending on the link between modelling hypotheses and interpretation rules.

With very large models this method is no more usable, because with a big amount of output, it is impossible to compare the results and to make a synthesis of the modelling and analysis processes. If we try to make both aspects independent we have to control the quality of the mesh. It means that we have to know the error coming from the discretization.

For this reason, some new formulations or new kind of analyses are necessary. To be useful they must reproduce the quality of the information provided by the previous simple models. These new requirements can be easily summarized; the first important aspect, is to know the influence of a component or of an hypothesis on the global response: this is a sensitivity analysis. It helps to check the validity of modelling hypotheses, for example how to represent a fixation, how to represent a join between two components. The second aspect is the control of the finite element mesh: it is the main concern of this presentation.

When a measure of the error is available different strategies can be used: either we use this information to give safety margins, either we try to improve the model so as to reach a given

level of precision. It will be the second part of the presentation: mesh adaptation. This step is important not only because it allows to improve the results but also because it gives the possibility to save computation time.

3. ERROR ESTIMATION

We assume that the mechanical model is convenient and that the error of the solution is only due to the imperfection of the finite element mesh.

To compute the error, the only way is to compare the results with something else.

The ideal but rarely found situation is to compare with the exact solution. Its existence can help to find good criteria to quantify the error of the approximate solution.

The second method is to compare with another finite element model. Happily, the method itself is providing good rules to indicate how to build the second model.

We have basically two possibilities either we build the second model with completely different hypotheses, for example equilibrium model if the first one is a displacement model. The second possibility is to use the convergence property of the method: if we refine a given finite element model or if we increase the degree of the elements, the solution is improved.

In some way it is necessary to define a distance between the two fields (approximate and exact). A norm has to be used and the most practical one, at least in linear elasticity, seems to be the energy norm of the error. It can be computed at the element level and by adding all the contributions, at the structure level. This method provides a single value that quantify the precision of the solution and allow comparisons between different approximations.

Now if we have two finite element models, we can compute the "distance" between them that is assumed to be a good image of the error. The energy norm of the error field is computed from the difference between the finite element stress and the stress produced by the second analysis.

3.1. Global analyses

In order to compute a second model, let us first compare the solution with the same mesh using a different model. The first proposed method was the so-called dual analysis [9]. It offers three advantages, first the same mesh can be used, secondly this model gives an upper bound on the global error and finally, dual analysis helps to understand the mechanism of the modelling process.

In some situations, the drawback is that it is not obvious to reformulate an equivalent problem with statically admissible elements. The reason is that these models accept only the loads defined inside the model, generally, constant or first and second degree polynomials. But they are more flexible with respect to the displacements boundary conditions even if this situation is less frequent in most engineering problems and if concentrated fixations are not allowed. Finally it is also observed that it is difficult to build 3-D equilibrium models.

The second method to compare finite element solutions is to introduce a more refined model either with more elements (h-extension) or with higher degrees (p-extension). With this second finite element model it is possible to compute a convergence rate and to extrapolate the solution to get a better approximation of the exact one.

A very efficient method often used in p-methods consists in using three levels of approximations (same mesh with three different degrees) and to perform a Richardson

extrapolation [10]. Another method mainly used in h-extension consists in taking into account the results of this comparison to guide the procedure towards more suitable meshes.

3.2. Local analyses

For all the previous methods it is necessary to perform a second finite element analysis. So the idea is to get the results of a comparison mesh without performing a full second analysis. Several procedures are available, all the methods try to perform a local analysis that is giving a good approximation of what should be the second analysis.

The basic idea is to work in a region or a patch. the most simple patches are composed of all the elements surrounding a node or an element. Some smoothing is performed at this level in order to obtain in the central node or in the central element more information about the higher order derivatives of the field. Basically knowing the linear displacement field of two 1-D element sharing a node, it is possible to compute the first derivative at the node by interpolating the slopes.

Coming back to the method [1], it is also possible to perform a local dual analysis, for instance to build a local equilibrium field.

Independently of the chosen method, the new comparison field has to fulfil two requirements: first, the comparison field must correspond to the finite element solution, it is the prolongation condition [1], and secondly, it must be close to the exact solution, so it must avoid at least one of the two defaults and hopefully both of the finite element solution, lack of continuity and lack of equilibrium.

To conform to the finite element solution, one can try to match directly to local values of stresses (S.P.R. method) or we can find a stress field that satisfy the surface traction applied to the patch. Another possibility that will be explored in the following consists in respecting the level of the generalized stresses conjugated to the deformation modes and balanced by the generalized forces. This last procedure conforms perfectly with the finite element method philosophy.

4. STRESS INTERPRETATION

4.1. Energy conjugate stresses

When the body forces are not taken into account, the displacement models are derived from the minimum total energy principle:

$$\delta (U+P) = \delta \left(\frac{1}{2} \int_{\Omega} \varepsilon_h^T H \varepsilon_h - \int_{\Gamma_t} \bar{t}^T u_h d\Gamma \right) = 0 \quad (1)$$

where:

U is the strain energy,

P is the potential energy (here, the body forces are not considered),

ε_h is the column vector of the strain components,

H is a symmetric matrix of elastic coefficients,

\bar{t} is the column vector of surface traction on Γ_t ,

u_h is the displacement vector,

Γ_t is the part of the boundary where surface traction is prescribed. The superscript T denotes transposition and the bar prescribed values. The displacement field in the finite element is discretized in the form

$$u_h = W q \quad (2)$$

where W is the matrix of assumed shape functions and q represents the discrete nodal values of this displacement. The column vector of the strain tensor components is then expressed in terms of generalized displacements by

$$\epsilon_h = \partial W q = B q \quad (3)$$

the stress vector,

$$\sigma_h = H \epsilon_h = H B q \quad (4)$$

The strain energy becomes:

$$U = \frac{1}{2} \int_{\Omega} q^T B^T H B q d\Omega = \frac{1}{2} q^T K q \quad (5)$$

where K is the stiffness matrix. The virtual work (V.W.) is expressed in terms of the prescribed surface traction by:

$$V.W. = \int_{\Gamma_t} u_h^T \bar{t} d\Gamma = \int_{\Gamma_t} q^T W^T \bar{t} d\Gamma = q^T g \quad (6)$$

It allows to define the generalized forces

$$g = \int_{\Gamma_t} W^T \bar{t} d\Gamma \quad (7)$$

conjugated to the displacement modes q .

Expressed in terms of weighted averages it gives by (7) and (4), the equilibrium relations between stresses and surface traction modes:

$$\int_{\Gamma_t} W^T \bar{t} dS = \int_{\Omega} B^T \sigma_h d\Gamma \quad (8)$$

This can be expressed in terms of the generalized displacements in the following way. By (5) and (6), the principle (1) becomes

$$\delta (U+P) = \delta \left(\frac{1}{2} q^T K q - q^T g \right) \quad (9)$$

and after variation of the parameters q , we obtain the well known relations

$$g = K q \quad (10)$$

equivalent to the equilibrium equations (8).

4.2. Practical signification of the energy conjugate stresses

For all the elements which pass the patch test [11], a sufficient condition for convergence is the existence of constant deformation modes. Their shape function are equal to 1 and amongst the weighted average stresses we find the simple averages of the stresses or of their linear combinations. The simplest models like linear membrane triangles provide only these averages. In more sophisticated models, higher order averages are not easily interpreted.

Instead of analyzing the properties of the full set of energy conjugate stresses, we can make some observations on the simple stress averages.

First, their definition is very general and allows to use the same procedure for all the types of elements. Secondly, their computation is cheap because they form a sub-product of the computation of the stiffness matrix. Thirdly they satisfy always the equilibrium equations in the mean sense of relation (8).

However some difficulties remain unsolved; first we cannot get useful information on the boundary of the element and secondly, as the interpretation of higher order conjugate stresses is more difficult, only the simple averages are of direct use.

4.3. Interpretation of the generalized forces

The generalized forces constitute also a special stress output. They are related to the stresses by relations (7) and (8) and can be directly computed from the displacements by relations (10). It is important to note that they are a natural direct output of the computation. They are not independent because they verify the global equilibrium equations of the element. It is well known that the generalized forces are very useful to check the global equilibrium of the structure or to find the loads which are transmitted from a substructure to another.

However at the element level their interpretation is more difficult due to the presence of the corner loads for which it is impossible to separate as seen in expression (7) the effects of the surface traction modes of the edges adjacent to the corner.

For elements of higher degree (2 or more), the interface loads come only from interface surface traction contribution. With a mixed formulation, it is possible to change the connector which is the displacement associated to this load into a equilibrium connector which will provide an easier interpretation of the surface traction [11]. The procedure will consist in building a particular hybrid model [12].

For this purpose the principle (1) is modified by relaxing the compatibility condition on the interface Γ_u . The new principle is written:

$$\left(\frac{1}{2} \int_{\Omega} \varepsilon^T H \varepsilon \, d\Omega - \int_{\Gamma_t} \bar{t}^T u \, d\Gamma - \int_{\Gamma_u} \bar{t}^T (u - \bar{u}) \, d\Gamma \right)_{\min} \quad (11)$$

The Lagrange multipliers t are easily identified as the surface traction. Now the discretized fields are u in the volume and on the boundary Γ_t and the surface traction t on the boundary

Γ_u . The displacement and the surface traction are discretized in the same way as for equilibrium models. The principle becomes:

$$\left(\frac{1}{2} \int_{\Omega} q^T B^T H B q d\Omega - \int_{\Gamma_t} \bar{t}^T W q d\Gamma - \int_{\Gamma_u} g^T V^T (Wq - \bar{u}) d\Gamma \right)_{\min} \quad (12)$$

In this principle the generalized displacements associated to the surface traction modes introduced on Γ_u are defined by the weighted averages:

$$\bar{q} = \int_{\Gamma_u} V^T \bar{u} d\Gamma \quad (13)$$

For example, if the surface traction mode introduced on the interface is constant, the conjugate displacement is simply the average of the displacement on the interface.

When a constant surface traction mode is introduced in an interface of quadratic displacement model it only changes the definition of the connectors. The local displacement becomes an average displacement and the weighted average of the surface traction becomes the intensity of the constant surface traction mode.

In membrane theory, when a surface traction mode of degree n is defined on an interface the continuity properties of the displacement model are not changed if the displacement field is of degree $n+2$.

The advantage of this procedure is that it allows to obtain directly, as in the equilibrium models the surface traction modes of the interface. On the separation surface of two substructures involving many interfaces of displacement elements for which the displacement connectors have been transformed in force connectors, there are two possibilities of connection:

- firstly, to connect the nodal displacements and the interface forces. The solution is exactly the same as for the displacement models but with a modified interpretation of the interface forces and displacements. The picture of surface traction modes is given partially by the interface connectors and partially by the forces conjugated to the nodal displacements.
- secondly, to connect only the interface forces. The interface between the two substructures behave exactly as in equilibrium models and the surface traction modes are directly available.

This last procedure is providing the best information on stresses along chosen boundaries.

5. COMPUTATION OF NODAL VALUES

The recovery procedure is now very simple, it consists in defining a polynomial stress field on a patch centred either on a node or on an element. With this field, it is necessary to compute the generalized stresses according to expression (8) in which the finite element stress is replaced by the smoothed stress of the patch and to identify the two values.

$$\int_{\Omega} B^T (\sigma_p - \sigma_h) d\Gamma = 0 \quad (14)$$

The integral is performed on all the elements pertaining to the patch and σ_p represents the smoothed stress field defined on the patch.

According to expression (8), the inter-element equilibrium equations are automatically satisfied in a weak sense by σ_h while σ_p is continuous by construction.

In many situations it is possible to use simplification. In order to decrease the number of equations it is possible to use only a sub-set of the generalized stresses. A second way used to simplify the problem consists in replacing the stress modes function by simple low degree polynomials in order to avoid the coupling between components. In many tests performed in [5], it has been shown that for low degree elements, the constant modes are sufficient to obtain good results.

In a general situation, the system of equations is over-determined and a least square method has to be used. Note that these conditions are easy to set up because they are only using the conditions defining the element characteristic matrices and vectors. They are also very general and can be applied to any model and formulation without important modifications.

If the patch is connected to a loaded boundary, it is convenient to add other equations ensuring that the smoothed field respects the surface traction condition (7). This condition is very important when the mesh is very coarse in one direction because in this situation the stress field is strongly influenced by the generalized applied loads. However to express these additional conditions it is necessary to know the imposed surface traction.

The condition is:

$$\int_{\Gamma_i} W^T (\bar{t} - t_p) = 0 \quad (15)$$

where t_p is the surface traction computed from σ_p .

The only difficulty is the choice of the polynomial smoothed stress field. Numerous experiments have shown that a good choice is a full polynomial with the same degree as the displacement field. Inside the domain this solution has proven to be very efficient. However in the thin parts of the structure the smoothed field has to include a sufficient number of terms in order to be able to represent the exact stress modes. A not yet investigated possibility could be to use an equilibrium stress field defined on the patch.

The tests also show that for solid structures, patches centred on nodes are sufficient inside the domain what means a significant saving in computation. For 3-D problems the best solution consists in using nodal patches inside the domain and element patches on the boundary.

Moreover, it is also observed that the introduction of surface traction mode is more important than high order stress modes.

After the solution of this system of equations, the value of the smoothed stress field is assigned to the central node and the nodes belonging to its adjacent edges or to the nodes of the central element.

6. INTERPOLATION SCHEME AND MESH ADAPTATION

In order to obtain standard output and to simplify the production of graphics, the nodal values are averaged and the results are displayed using the interpolation scheme of the displacement field. The obtained smoothed stress field is no more compatible, but continuous. It don't satisfy the equilibrium equation except on the loaded boundaries in the sense of least square method. However part of the weak equilibrium conditions are fulfilled at least before the averaging procedure.

If the estimated error of an element is higher than the prescribed one, subdivision of this element is required. In order to know how to refine each element, an estimation of the error sensitivity with respect to the size of the element must be provided. The *a priori* convergence law at element level has the following form:

$$\varepsilon_i = C_i h_i^{\beta_i} \quad (16)$$

where C_i is a constant, h_i the element size and β_i the element convergence rate. In 3-D, if the element size is sufficiently small, β_i is given by formula:

$$\beta_i = \min(p, \lambda) \quad (17)$$

where p is the degree of the polynomial approximation and λ , a parameter characterizing the smoothness of the solution.

Several methods can be used to estimate β_i [13, 14], i.e. methods based on uniform mesh refinement, methods based on local analysis and methods based on numerical adjustment. The last one is chosen in the proposed procedure.

6.1. Mesh optimization

Knowing the sizes of elements, the estimated errors and the convergence rates, we have to build the optimal mesh: it minimizes the total number of elements and ensures a prescribed total error level. The problem of mesh optimization has been solved in [13, 14] by minimizing a Lagrangian function with an equality constraint.

6.2. Adaptation

After a first analysis using the initial mesh supplied by an automatic mesh generator, an error estimation is performed and the optimal mesh densities are computed. All the information needed to improve the mesh is thus available.

This problem of mesh adaptation has been solved for 2-D problems using subdivision and transition techniques [13]. For 3-D geometries the new mesh is preferably obtained by remeshing techniques.

They consist in using an automatic mesh generator to build a new mesh which respects the elements sizes map given as a result of the mesh optimization process.

Once more, it seems here that hexahedral mesh generators (transfinite methods) are in general not adapted to respect local densities. On the other hand, more and more tetrahedral mesh generators based on Delaunay-Voronoi or frontal methods are able to produce the strongly graded meshes required for the adaptive procedures.

In some of them, local densities of internal elements are controlled by the densities of the surface triangles [15]. In more up-to-date techniques a background mesh, usually the current mesh is provided to deliver the necessary density information. For the generation of the new mesh, it is possible to get the size of the elements everywhere in the domain by interpolating values on the previous one. Those remeshing techniques seem to be the most attractive to perform 3-D mesh adaptation [16-19].

7. NUMERICAL EXAMPLE

Many attempts have been made previously to introduce discretization error control in the process of shape optimization, either in 2-D or in 3-D, [16-18].

As explained above, in 3-D, a good control of the adaptive meshing can only be obtained if free mesh generation is used, generally Delaunay or frontal methods. However, in this case, it still remains to overcome the important difficulty of handling sensitivity analysis with unstructured meshes.

In shape optimization, sensitivity analysis consists in computing the influence of shape perturbations on objective function, inequality constraints and structural responses.

This computation is achieved by repeating a significant part of the finite element analysis for all the design variables. The computing cost of this operation is then proportional to the number of design variables and is strongly dependent of the number of D.O.F. The necessary increase in the precision is thus paid by a significant increase in the computing time and fine tuning of the overall procedure has to be set up to limit the cost of the process [17].

The following example is designed to show what should be the benefit or such a procedure. But, it is not yet fully achieved, only a small number of steps have been run to demonstrate the feasibility of the method.

A clamped "I" beam is assumed to be loaded with an uniform pressure $P = 100$ on the upper spar cap (figure 1). The objective function corresponds to the minimization of the weight. The inequality constraints impose that the von Mises stress is less than 1000.

The design variables are the thicknesses of the spar caps $S1$ and $S2$ whose initial value is "a" and the radius of the filled allowed to linearly vary between the fore and the back sections of the beam (variables $R1$ and $R2$, reported in table 1).

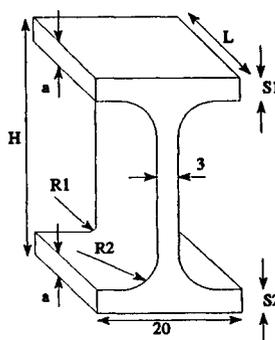


figure 1 : Geometric model

Table 1. Dimensions and optimal values for the tetrahedral mesh

H	L	S1	S2	a	R1	R2
33	20	3	3	3.0	2.0	6.0

Table 2. Error estimation for the tetrahedral mesh

Nodes	Elements	D.O.F.	$\frac{1}{2} \ u_h\ ^2$	$\frac{1}{2} \varepsilon^2$	$\bar{\eta}^2 \%$	σ_h	$\bar{\sigma}_h$	δ
639	2008	1506	9829.36	1617.18	14.128	898	1097	1.301
1839	7213	5175	12263.67	1187.36	8.827	1202	1247	1.718
3293	13842	9417	12773.45	928.33	6.775	1325	1370	1.825
9903	46942	28644	13435.69	515.09	3.692	1624	1655	1.956
17456	86132	50727	13629.35	375.67	2.682	1696	1859	1.999

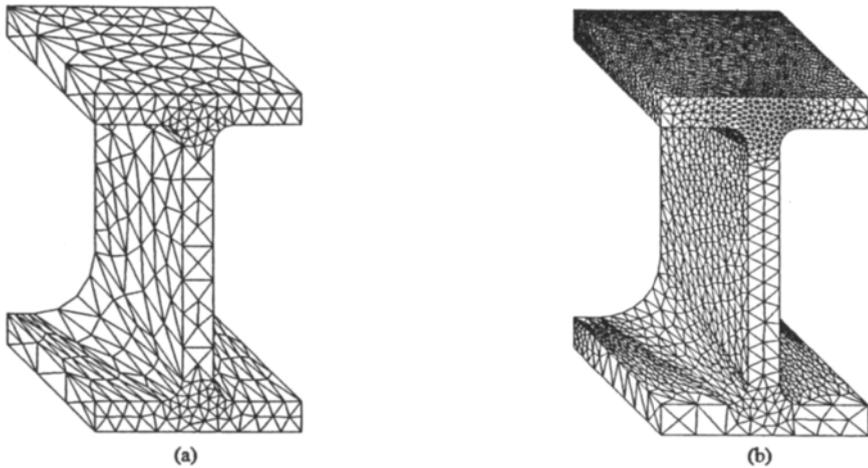


figure 2 : a) Initial mesh, b) final adapted mesh

Elasticity module and Poisson's coefficient are $E = 21000$, $\nu = 0.3$. The optimization has been carried on with a finite element model composed of tetrahedra (figure 2a). The results are shown in table 2 where $\frac{1}{2} \varepsilon^2$ is the absolute error and η^2 the relative error, $\bar{\sigma}$ represents the smoothed stress field. After the optimization for which the relative energy error is equal to 14%, an adaptive study was performed in order to obtain a prescribed level of relative energy error $\eta^2 = 2.5\%$.

As expected, the adaptation is increasing the number of D.O.F., it also demonstrates that the optimization performed with a coarse finite element model yields to a very important violation of the maximum von Mises stress (about 80%). The maximum displacement (δ , table 2) is also significantly growing during the adaptive process proving that the structural stiffness was very badly modelled.

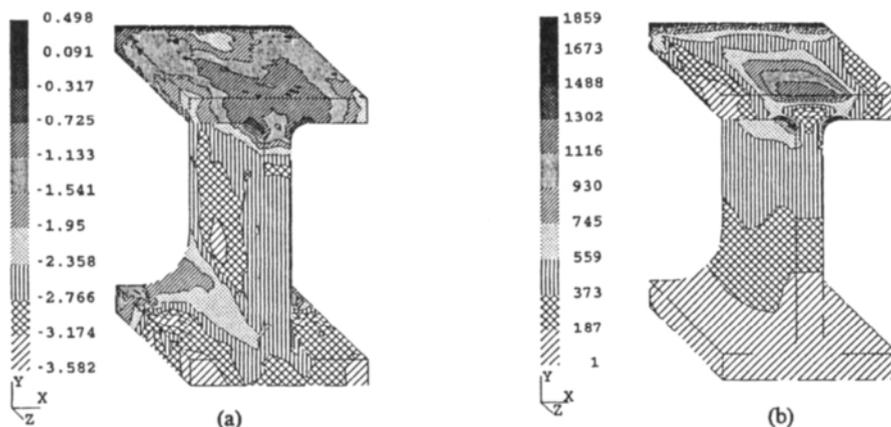


figure 3 : Final mesh a) relative error density b) von Mises stress

8. CONCLUSION

The developed method based on the computation of weighted averages of stresses on patches is very general and easy to implement. Tested on some analytical examples it exhibits a very good behaviour and all the observed effectivity indices are very good and seems to always converge to 1.

In practical 3-D examples, the method is able to provide precise information on optimal mesh densities and allows to control the quality of the finite element solution for static linear elastic problems.

The techniques presented here derive from previous 2-D developments [13] and behave very well. Their 3-D versions have been successfully tested in several numerical examples. Even if a careful examination of the different output of the program allows to conclude that the results are satisfactory, more tests are necessary to verify the exactness of all the numerical values. However the next challenge, in shape optimization, is to produce a fully optimized mesh satisfying exactly the requirements of the background map of element sizes.

Experiments are currently performed to compare the respective efficiencies of Delaunay and frontal methods. Moreover, these procedures of error control are now used in the frame of advanced techniques like shape optimization [17].

REFERENCES

1. Ladevèze P., "Comparaison de modèles de milieux continus", Thèse d'état, Université P & M Curie, (1975).
2. Pelle J.P., "Contrôle des paramètres des calculs éléments finis: application au 3D et au non-linéaire", Journée d'étude C.S.M.A, (1994).
3. Zienkiewicz O.C. & Zhu J.Z., "The superconvergent patch recovery and a posteriori

- error estimates. Part I: the recovery technique". *International Journal for Numerical Methods in Engineering*, vol. 33, 1331-1364, (1992).
4. Boroomand B. & Zienkiewicz O.C., "Recovery by Equilibrium in Patches (R.E.P.)", *International journal for numerical methods in engineering*. Vol.40, 137-164, (1997).
 5. Dufeu E., "Calcul d'erreur et adaptation de maillages en 3 dimensions", Thèse de doctorat, Coll. des Pub. de la F.S.A., n°174, Université de Liège, (1997).
 6. Beckers P., "Stress interpretation in the finite element method", AGARD Conference Proceedings No.248, pp. 8.1 to 8.10, (1978).
 7. Beckers P. & Dufeu E., "Estimation d'erreur basée sur l'équilibre et le lissage des tensions", Version étendue de la communication présentée aux Journées Internationales de "Mecánica Computacional y CAD/CAM". Novembre 22-24, 1995, Université de Concepción, Chili. Rapport SF-215, L.T.A.S., Université de Liège, (1996).
 8. Pelle J.P., Beckers P. & Gallimard L., "Estimation des erreurs de discrétisation et analyses adaptatives. Application à l'automatisation des calculs éléments finis", Institut pour la promotion des sciences de l'ingénieur, IPSI, (1996).
 9. Fraeijs de Veubeke B., "Bending and stretching of plates" AFFDL-TR-66-80, Wright-Patterson A.F.B., Dayton, Ohio, (1966).
 10. Szabo B.A. & Babuška I., "Finite element Analysis", John Wiley & Sons, (1991).
 11. Sander G. and Beckers P., "The influence of the choice of connectors in the finite element method", *Int. Jnl. for num.. meth. in Engng.*, Vol 11, pp 1491-1505, (1977).
 12. Pian T., "Formulations of finite element methods for solid continua" in "Recent advances in matrix methods of structural analysis and design" ed. Gallagher, Yamada, Oden, Univ. of Alabama Press, (1971).
 13. Beckers P. & Zhong H.G., "Mesh adaptation for 2-D stress analysis", 2nd International Conference on Computational Structures Technology, Athens, Greece, August 30 September, LTAS report SA-178, (1994).
 14. Coorevits P., Ladevèze P. & Pelle J-P., "Sur l'automatisation des calculs éléments finis en élasticité 2-D ou axisymétrique", rapport interne n° 142, LMT, E.N.S. de Cachan, France, (1993).
 15. George P.L., "Synthèse de quelques méthodes de génération de maillages éléments finis", *Revue européenne des éléments finis*, vol. 2, 121-153, (1993).
 16. Bueda G. & Oliver J., "A general methodology for structural shape optimization problems using automatic adaptive remeshing", *International Journal for Numerical Methods in Engineering*, Vol.36, 3161-3185, (1993).
 17. Cugnon F. & Beckers P., "Design of a 3-D error control procedure", project BRITE EURAM BRE2-CT94-0590-MODSYSS, LTAS report SA-183, (1995).
 18. Hinton E. & Sieng J., "Aspects of adaptive finite element analysis and structural optimization". 2nd International Conference on Computational Structures Technology, Athens, Greece, Civil Comp Ltd, Edinburgh, Scotland B.H.V. Topping and M. Papadrakakis (Editors), *Advances in Structural Optimization*, 1-25, (1994).
 19. George P-L. & Borouchaki H., "Triangulation de Delaunay et maillage - applications aux éléments finis", Hermes, Paris, (1997).
 20. Dufeu E. & Beckers P., "Estimacion de error basada en esfuerzos generalizados y tracciones en los limites", to be presented at the X ENIEF, congreso sobre Metodos Numericos y sus Aplicaciones, San Carlos de Bariloche, Argentina, 10-14 de noviembre de 1997.

Adaptive methods and related issues from the viewpoint of hybrid equilibrium finite element models

O. J. B. Almeida Pereira^a, J. P. Moitinho de Almeida^a and E. A. W. Maunder^b

^aDepartamento de Engenharia Civil, Instituto Superior Técnico,
Av. Rovisco Pais, 1096 Lisboa, CODEX Portugal

^bSchool of Engineering, University of Exeter,
North Park Road Exeter, EX4 4QF United Kingdom

This paper presents some recent results on the development of adaptive procedures for finite element formulations locally enforcing the equilibrium conditions. The topics covered include the control of spurious kinematic modes, solution techniques, the definition of compatibility defaults, the derivation of compatible displacement fields and possible error estimators and indicators.

These procedures are illustrated with a simple example in 2D elastostatics.

1. INTRODUCTION

Though considerable attention has been given to the development of dual approaches in structural problems [1–4], the approach based on conforming displacement elements, nodal variables, and stiffness methods was bound to become popular due to its inherent mathematical simplicity. Indeed this approach has become the dominant method.

Whilst the displacement approach may be said to have reached a state of maturity, it suffers, from an engineering viewpoint, in placing its first priority on compatibility and relaxing local equilibrium. This is the reverse of the priorities commonly required in structural engineering, and this reversal appears to have made a significant contribution to the catastrophic collapse of the Sleipner offshore platform in 1991 [5].

These priorities are reversed in the alternative approach based on stress elements with locally enforced equilibrium. The earlier proponents and advocates of such an approach also realised the benefits of dual approaches in that they provided the opportunities to complement each other, and bound quantities of interest, e.g. strain energy. This bounding property has become of considerable importance more recently in the context of error analysis and adaptivity in the finite element method [6].

More recent developments [7–12] have tended towards hybrid forms of equilibrium elements with side variables. The hybrid approach has the advantage of allowing the development of new elements with extreme ease, and it is hoped that it will further allow progress into investigating many topics where the equilibrium formulation has not yet had a chance to make an impact, e.g. error estimation and adaptivity in finite element modelling, non-linear analysis (both geometric and material non-linearities), and 3-D

modelling.

The present authors have been involved in developing the hybrid equilibrium approach, and this paper will present some of the results that have been obtained in two relevant areas in this context:

- determination of robust equilibrium finite element models [12];
- development of error estimators and indicators for adaptive strategies [13].

2. TERMINOLOGY

In this paper the following terminology is used:

displacement formulations are normally based on the use of displacement fields as variables, and a form of possibly incomplete compatibility is imposed. This implies that the equilibrium conditions are integrated by parts (explicitly or implicitly) and are meant to be satisfied on average;

compatible formulations are obtained when the displacement fields are constrained to be kinematically admissible, i.e. to satisfy the compatibility equations in an exact (strong) sense;

stress formulations are normally based on the use of stress fields as variables, and a form of possibly incomplete equilibrium is imposed. This formulation implies that the compatibility conditions are integrated by parts (explicitly or implicitly) and are meant to be satisfied on average;

equilibrium formulations are obtained when the stress fields are constrained to be statically admissible, i.e. to satisfy the equilibrium equations in an exact (strong) sense;

hybrid formulations are based on independent approximating functions within an element and on element boundaries, the latter serving to impose interelement conditions, e.g. using the concept of Lagrange multipliers.

In this paper only hybrid equilibrium formulations are considered for linear elastostatic problems in 2-D. These concepts are currently being extended to 3-D elastostatics [13–15].

3. BASIC EQUATIONS

The basic governing equations for linear elastostatic problems, where for the sake of simplicity initial strains are not considered, are in matrix format:

$$\bullet \text{ equilibrium of stress vector } \boldsymbol{\sigma} \text{ with body forces } \mathbf{b}, \partial^i \boldsymbol{\sigma} + \mathbf{b} = \mathbf{0}; \quad (1)$$

$$\bullet \text{ compatibility of strains } \boldsymbol{\varepsilon} \text{ with displacements } \mathbf{u}, \partial \mathbf{u} = \boldsymbol{\varepsilon}; \quad (2)$$

$$\bullet \text{ constitutive relations, } \boldsymbol{\varepsilon} = \mathbf{f} \boldsymbol{\sigma}; \quad (3)$$

$$\bullet \text{ static boundary conditions, } \mathbf{N} \boldsymbol{\sigma} = \mathbf{t}_\Gamma \text{ on } \Gamma_t; \quad (4)$$

$$\bullet \text{ kinematic boundary conditions, } \mathbf{u} = \mathbf{u}_\Gamma \text{ on } \Gamma_u. \quad (5)$$

Matrix N transforms the stress vector to boundary traction components, t_Γ represents the prescribed tractions on Γ_t and u_Γ represents the prescribed displacements on Γ_u .

4. HYBRID EQUILIBRIUM FINITE ELEMENT FORMULATIONS

This paper deals with a hybrid equilibrium formulation that satisfies the local equilibrium conditions of stress within elements and across element interfaces. This formulation belongs to a broader class of hybrid formulations [9,10] based on multi-field approximating functions within each element and on an element boundary. The boundary is discretised into sides. To locally enforce equilibrium between elements, independent functions are defined for each side. Therefore the functions for adjacent sides of an element are not constrained to be continuous ("frame" functions), unlike the hybrid formulations of Pian [16]. Hybrid-Trefftz formulations [11,17] are hybrid formulations for which the element stress fields are both statically and kinematically admissible, i.e. the strains corresponding to the stresses are locally compatible. Trefftz polynomials are not so convenient for equilibrium elements, because they provide fewer functions to form a solution that satisfies local equilibrium on the sides.

The element equations are built up in matrix format as follows:

- equilibrium within an element $\sigma = Ss + \sigma_0$, $\partial^t \sigma_0 + b = 0$, and $\partial^t S = 0$, where the columns of matrix S form a basis for stress fields that are statically admissible with zero body forces, vector s contains n_S stress parameters, and σ_0 represents a particular solution;
- kinematic boundary conditions of side j , $u = V_j v_j$, where the columns of V_j form a basis for assumed displacements for side j ; this equation can be extended to include all sides of an element, $u = Vv$. The vector v contains n_V displacement parameters;
- weak equilibrium on the boundary of an element is imposed by:

$$\oint_{\Gamma} V^t N \sigma \, d\Gamma = \oint_{\Gamma} V^t N S \, d\Gamma \, s + \oint_{\Gamma} V^t N \sigma_0 \, d\Gamma = \oint_{\Gamma} V^t t_\Gamma \, d\Gamma, \quad (6)$$

or $Ds + g_0 = g$, where here t_Γ represents tractions applied to the boundary segments of an element. Vectors g_0 and g can be identified as generalised boundary tractions;

- weak compatibility within an element and its boundary is imposed by:

$$\oint_{\Gamma} (NS)^t u \, d\Gamma = \int_{\Omega} S^t \varepsilon \, d\Omega, \text{ or}$$

$$\oint_{\Gamma} S^t N^t V \, d\Gamma v = \int_{\Omega} S^t f S \, d\Omega \, s + \int_{\Omega} S^t f \sigma_0 \, d\Omega, \quad (7)$$

or $D^t v = F s + e_0$.

Combining equilibrium and compatibility for an element leads to the partitioned matrix equation:

$$\begin{bmatrix} -\mathbf{F} & \mathbf{D}^t \\ \mathbf{D} & \cdot \end{bmatrix} \begin{Bmatrix} \mathbf{s} \\ \mathbf{v} \end{Bmatrix} = \begin{Bmatrix} \mathbf{e}_0 \\ \mathbf{g} - \mathbf{g}_0 \end{Bmatrix}, \text{ or } \begin{Bmatrix} \cdot \\ \mathbf{g} \end{Bmatrix}, \quad (8)$$

in the absence of prescribed body forces.

Since \mathbf{F} is positive definite, \mathbf{s} can be eliminated to yield an alternative stiffness equation, e.g. $\mathbf{D}\mathbf{F}^{-1}\mathbf{D}^t\mathbf{v} = \mathbf{g}$.

Note that the weak form of equilibrium becomes the strong form $\mathbf{N}\boldsymbol{\sigma} = \mathbf{t}_\Gamma$ when the displacement functions in \mathbf{V} and the tractions ($\mathbf{N}\boldsymbol{\sigma} - \mathbf{t}_\Gamma$) are described in terms of the same set of functions. For then, $\oint_\Gamma \mathbf{u}^t \mathbf{t} \, d\Gamma = 0$ for all $\mathbf{u} \Rightarrow \mathbf{t} = 0$. For example this will occur when displacements are based on complete polynomials of degree p and the degree of tractions (or stresses when the sides are straight) does not exceed p .

Generalised displacements \mathbf{v} and tractions \mathbf{g} form dual variables by virtue of the scalar work product:

$$\oint_\Gamma \mathbf{u}^t \mathbf{t} \, d\Gamma = \mathbf{v}^t \oint_\Gamma \mathbf{V}^t \mathbf{t} \, d\Gamma = \mathbf{v}^t \mathbf{g} = \mathbf{v}^t \mathbf{D}\mathbf{s} = \mathbf{s}^t \mathbf{F}\mathbf{s}, \quad (9)$$

when tractions \mathbf{t} are self-balanced in the absence of body forces. The $(n_V \times n_S)$ matrix \mathbf{D} represents the boundary work done by each stress field with each boundary displacement field. Depending on the definition of these fields, it may be possible for non-zero displacements represented by a vector \mathbf{v} to incur zero work for all stress fields in \mathbf{S} , and hence zero stress or strain energy within the element. In this case \mathbf{v} satisfies:

$$\mathbf{D}^t \mathbf{v} = 0, \quad (10)$$

or in other words, \mathbf{v} belongs to the nullspace of \mathbf{D}^t . Clearly this will occur when \mathbf{V} includes the n_R rigid body movements of the complete boundary of an element. However the nullspace of \mathbf{D}^t may also include n_{skm} spurious kinematic modes (SKM's), which can be considered in a similar way to mechanisms,

$$n_{\text{skm}} = n_V - n_R - \text{rank}(\mathbf{D}). \quad (11)$$

Furthermore, the existence of spurious kinematic modes influences the manner in which an element may be loaded with boundary tractions. Tractions are admissible only if they do not excite mechanisms corresponding to the spurious kinematic modes. In terms of dual bases for the vectors, a traction vector \mathbf{g} is admissible if zero work is done with the nullspace of \mathbf{D}^t , and in particular if

$$\mathbf{A}\mathbf{g} = 0,$$

where \mathbf{A}^t represents that subspace of the nullspace which is based on the spurious kinematic modes. Zero work by \mathbf{g} with the rigid body modes is of course a consequence of the overall equilibrium requirement of the tractions. If \mathbf{g} is admissible, then it is also consistent with

$$\mathbf{D}\mathbf{s} = \mathbf{g}, \quad (12)$$

and thus allows for a solution for s . Multiple solutions to this equation are possible when S allows n_{hyp} hyperstatic, or residual, stress fields for which $N\sigma = 0$, or more generally $N\sigma = t$ which is orthogonal to all displacements in V . When strong equilibrium is enforced $t = 0$,

$$n_{\text{hyp}} = n_S - \text{rank}(D), \text{ and thus } n_{\text{skm}} = (n_V - n_R) - (n_S - n_{\text{hyp}}). \quad (13)$$

For an assembly of elements, the system of equations can be expressed as:

$$\begin{bmatrix} -F & \bar{D}^t \\ \bar{D} & \cdot \end{bmatrix} \begin{Bmatrix} s \\ \bar{v} \end{Bmatrix} = \begin{Bmatrix} 0 \\ \bar{g} \end{Bmatrix} \quad (14)$$

For this system two fundamental questions arise:

1. Do spurious kinematic modes of the elements propagate through the system, if so in what form and how many such modes exist for the system?
2. Is the load admissible for the system?

These questions can be answered before attempting to solve the system equations. A possible pre-evaluation of the system can be made by focusing on the matrix \bar{D} . For simplicity of presentation it is now assumed that the sides are (re)ordered so that:

$$\bar{v} = \begin{Bmatrix} \bar{v}_1 \\ \bar{v}_2 \end{Bmatrix}; \quad \bar{g} = \begin{Bmatrix} \bar{g}_1 \\ \bar{g}_2 \end{Bmatrix}; \quad \text{and } \bar{D} = \begin{Bmatrix} \bar{D}_1 \\ \bar{D}_2 \end{Bmatrix};$$

where \bar{v}_1 and \bar{g}_1 refer to the unknown side displacements and specified side tractions respectively; \bar{v}_2 and \bar{g}_2 refer to prescribed displacements and reactive tractions respectively. Then spurious kinematic modes for the supported system are defined by the nullspace of \bar{D}_1^t , represented by \bar{A}_1^t ; and admissible loads satisfy $\bar{A}_1 \bar{g}_1 = 0$.

An alternative, and possibly simpler, way of pre-evaluating the stability of the system comes from focusing on *mesh compatibility*. That is,

1. Determine the spurious kinematic modes for each element, and form

$$A^t = \begin{bmatrix} A^{1t} & & 0 \\ & \ddots & \\ 0 & & A^{nt} \end{bmatrix}.$$

Spurious kinematic modes for the unassembled system are defined by $A^t k$ for arbitrary vectors k .

2. In the assembled system, the above modes propagate when they are compatible between elements. Compatibility is checked with the aid of a system kinematic matrix H^t , i.e. for compatibility $H^t A^t k = 0$, or $C^t k = 0$ where $C = AH$. H^t can be simply formed by appealing to contragradience, and using the statical relations of the statically indeterminate system, i.e. $g = Hp$ where the vector p contains biactions, and g represents self-balancing element tractions within the system.

Then the number and definition of the spurious kinematic modes of the system can be found from the nullspace of matrix C^t . In this case it should be noted that the dimensions of C may be much smaller than those of \bar{D}_1 , but in addition the nullspaces of each elemental matrix D^{et} are also required.

5. SOLUTION TECHNIQUES IN THE PRESENCE OF SKM'S

For meshes without spurious kinematic modes the systems of equations described in the previous section can be solved using any technique appropriate for symmetric matrices.

As the matrices are very sparse an important gain in solution time may be achieved by using appropriate algorithms. Banded and skyline algorithms can be used, but are not optimal. The best results have been obtained by using procedures for general sparse matrices [18].

If spurious kinematic modes are present in the mesh, the system of equations is singular. The solution of such a system may be either undetermined or impossible.

If the solution of the system is impossible a combination of the approximation functions that equilibrates the applied loads cannot be found and the load is inadmissible.

When the solution is possible some of the displacement variables are undetermined.

To solve these systems standard algorithms for non-singular matrices cannot be used. One alternative is to transform the system equations after performing an explicit singular value decomposition of the matrix in equation (14), or \bar{D} . This procedure has the advantages of being very stable, and explicitly calculating the spurious kinematic modes, but in practice the computational cost is likely to be too high. Instead it is preferred to use a Gaussian elimination technique with control of zero pivots.

Good results have been obtained using an adaptation of an "off the shelf" procedure prepared for symmetric, diagonal dominant, sparse matrices [18]. An high performance routine for unsymmetric sparse matrices which can handle singular systems has also been used to solve these systems, with good computational performance [19].

6. CONTROL OF THE SPURIOUS KINEMATIC MODES

The direct solution technique described above must be used with extreme care, because in situations of numerical instability it is difficult to distinguish the numerical zeros from small non-zero values. It is always better to establish *a priori* that there are no SKM's. If this is not possible then it is preferable to know how many exist, so that the solution procedure can be controlled.

One way to control the number of SKM's is to use special assemblies of triangular primitive elements, macro elements, as those presented in Figure 1. Meshes built using

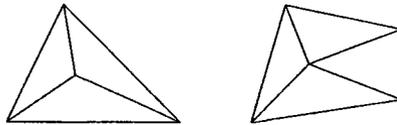


Figure 1. Typical macro elements.

triangular macro elements are free from SKM's. When quadrilateral macro elements are used it is guaranteed [12] that with approximation functions of degree 2 or higher there may exist at most one SKM per macro element, but this is not excited by external side tractions. Based on this characterisation it is possible to eliminate the SKM's in the assembly process or to do a check on the number of SKM's that were found by the solver.

The hybrid finite element formulation used easily allows for the inclusion of irregular elements (triangles with more than three sides/nodes or quadrilaterals with more than four sides/nodes), as those presented in Figures 6 to 8. This is extensively used in the refinement procedures as it eliminates the need for remeshings or mesh regularization procedures.

Numerical experience has shown that for meshes obtained using this type of refinement the number of SKM's remains constant throughout the adaptive process.

7. COMPATIBILITY DEFAULTS

Unless the exact solution can be represented by the approximation functions used in the finite element model, there will be an error associated with the equilibrium finite element solution. This error is due to the fact that the corresponding strains do not satisfy locally the compatibility conditions.

Inside the elements this lack of compatibility can be measured by the residual in the St Venant compatibility equation,

$$r = \frac{\partial^2 \varepsilon_{xx}}{\partial y^2} + \frac{\partial^2 \varepsilon_{yy}}{\partial x^2} - 2 \frac{\partial^2 \varepsilon_{xy}}{\partial x \partial y}. \quad (15)$$

On the sides between the elements and on those that belong to the kinematic boundary there may also exist compatibility defaults. These defaults are associated with the possible existence of different tangential strains and curvatures [13].

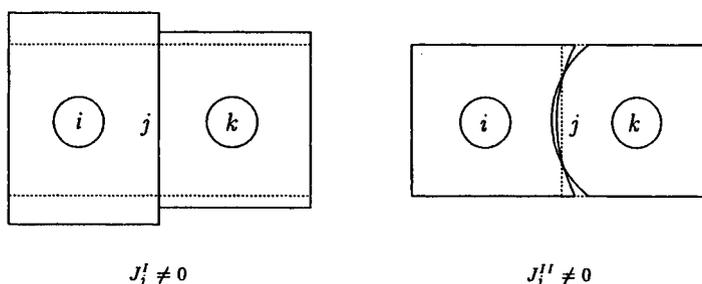


Figure 2. Compatibility defaults of side j between two elements.

As shown in Figure 2, for side j that connects elements i and k a tangential strain jump,

$$J_j^I = [\varepsilon_{tt}]_{j,i} - [\varepsilon_{tt}]_{j,k}, \quad (16)$$

can be defined. A curvature jump may also exist, which for a straight side is defined by

$$J_j^{II} = \left[\frac{\partial^2 u_n}{\partial t^2} \right]_{j,i} + \left[\frac{\partial^2 u_n}{\partial t^2} \right]_{j,k} = \left[2 \frac{\partial \varepsilon_{nt}}{\partial t} - \frac{\partial \varepsilon_{tt}}{\partial n} \right]_{j,i} + \left[2 \frac{\partial \varepsilon_{nt}}{\partial t} - \frac{\partial \varepsilon_{tt}}{\partial n} \right]_{j,k}, \quad (17)$$

where each \vec{n} is the external normal of the corresponding element and $\vec{t} = [-n_y \ n_x]^t$. Similar quantities can be defined for the sides of the elements that belong to the kinematic

boundary. These quantities are similar to the unbalanced body forces and the traction jumps for compatible finite elements [20,21]. Likewise, as will be presented later, they can be used to indicate the error of the solution.

8. DERIVATION OF CONFORMING DISPLACEMENT FIELDS

A finite element solution obtained using hybrid equilibrium elements of degree p yields, for each side j , a displacement field \mathbf{u}_j of degree p .

A compatible displacement field may be constructed from the equilibrium finite element solution and the kinematic boundary conditions. This is achieved by calculating a displacement field that is continuous within each element, and then by making this field compatible in the whole domain and on the kinematic boundary [13]. This is the dual of the Ladeveze approach [24] for displacement formulations.

If $p \leq 1$, the strains computed from the discretised stresses correspond to a displacement field which is continuous within each element but is normally discontinuous across the sides. If $p > 1$, normally it is not possible to obtain a displacement field, within each element, for which the strains will match exactly those computed from the stress field.

Nevertheless, it is always possible, in each element, to calculate a displacement field of degree $p + 1$, \mathbf{u}_e , for which the strains are a projection of those calculated from a stress field of degree p .

This displacement field is given by $\mathbf{u}_e = \Psi \delta = \Psi(\delta_S + \mathbf{R} \delta_R)$, where Ψ is a matrix of approximation functions for displacements of an element with displacement parameters δ , and matrix \mathbf{R} transforms the rigid body mode amplitudes δ_R into displacement parameters.

The deformed shape of an element as described by $\Psi \delta_S$ is computed, to within the rigid body displacement, from the stress solution:

$$\left(\int_{\Omega_i} (\partial \Psi)^t \mathbf{f}^{-1} (\partial \Psi) d\Omega \right) \delta_S = \int_{\Omega_i} (\partial \Psi)^t \sigma d\Omega. \quad (18)$$

This system has three dependencies, corresponding to the rigid body movement. The displaced position of an element may be calculated from the side displacements of the hybrid finite element solution when spurious kinematic modes are not present. In this case a least squares solution to the local overdetermined system of equations,

$$\left(\int_{\Gamma_j} \Psi d\Gamma \right) (\delta_S + \mathbf{R} \delta_R) = \int_{\Gamma_j} \mathbf{u}_j d\Gamma, \quad (19)$$

which apply simultaneously for each side Γ_j of element Ω_i , provides the amplitudes δ_R of the three degrees of freedom for rigid body displacement.

As both σ and \mathbf{u}_j correspond to the same equilibrium finite element solution, three of the equations in (19) are linearly dependent. Thus, for triangular elements, all the equations in (19) can be satisfied exactly.

For triangular elements, a system without redundant equations is formed by equations

$$\left(\int_{\Gamma_j} \bar{\mathbf{t}}^t \Psi d\Gamma \right) (\delta_S + \mathbf{R} \delta_R) + \int_{\Gamma_j} \bar{\mathbf{t}}^t \mathbf{u}_j d\Gamma = 0, \quad (20)$$

for each side Γ_j of element Ω_i .

The displacement field \mathbf{u}_e is normally discontinuous across the sides and does not satisfy the kinematic boundary conditions. A compatible displacement field of degree $p + 1$, \mathbf{u}_c , may be obtained by smoothing the displacements \mathbf{u}_e and enforcing the kinematic boundary conditions. Simple nodal averaging may be used, but several of the more sophisticated methods used for stress smoothing [22] may also be adapted for smoothing displacements.

9. ERROR ESTIMATORS AND ERROR INDICATORS

Both element error indicators and global error estimators, always directed at the energy norm, may be obtained *a posteriori* in a variety of ways for equilibrium finite element solutions.

In the following we will focus on two different approaches. Methods that compute an upper bound of the error of the equilibrated solution using a compatible displacement field and methods that are based on the explicit use of compatibility defaults.

For these approaches the error estimator is always calculated as the sum of element error indicators.

$$\epsilon = \left(\sum_i \epsilon^2 \right)^{\frac{1}{2}}. \quad (21)$$

9.1. DUAL ANALYSIS

Given a compatible displacement field \mathbf{u}_c , which is associated with a (normally non equilibrated) stress field $\boldsymbol{\sigma}_c$, and an equilibrated stress field $\boldsymbol{\sigma}_e$, then the following relation holds [23]:

$$U(\mathbf{e}_e) + U(\mathbf{e}_c) = U(\boldsymbol{\sigma}_e - \boldsymbol{\sigma}_c). \quad (22)$$

Where $U(\cdot)$ is the strain energy, $\mathbf{e}_e = \boldsymbol{\sigma}_e - \boldsymbol{\sigma}_{\text{exact}}$ and $\mathbf{e}_c = \boldsymbol{\sigma}_c - \boldsymbol{\sigma}_{\text{exact}}$.

As $U(\mathbf{e}_c) \geq 0$ and $U(\mathbf{e}_e) \geq 0$, an upper bound of the global error of either solution can be obtained from a pair of dual solutions:

$$\|\mathbf{e}\| \leq \epsilon = \|\boldsymbol{\sigma}_e - \boldsymbol{\sigma}_c\| = (2 U(\boldsymbol{\sigma}_e - \boldsymbol{\sigma}_c))^{\frac{1}{2}}. \quad (23)$$

The value of the error bound can be calculated element-wise and each of the contributions can be used as an indicator of the error within the corresponding element [24].

The compatible displacement field can either be obtained from a compatible finite element model [15] or derived from the equilibrated solution. The quality of the displacement field is crucial as regards the effectivity of the error estimator. When dual solutions of different quality are used, the upper bound will be closer to the error of the worst solution.

9.2. USE OF COMPATIBILITY DEFAULTS

The compatibility defaults r , J^I and J^{II} are related to the errors in the equilibrated stress fields. An element error indicator suitable for estimating the energy error norm is

expressed by:

$$\epsilon_i^2 = a c_i h_i^4 \int_{\Omega_i} r^2 d\Omega + a \sum_j \left(c_2 h_j \int_{\Gamma_j} \left(\frac{1}{2} J_j^I \right)^2 d\Gamma + c_3 h_j^3 \int_{\Gamma_j} \left(\frac{1}{2} J_j^{II} \right)^2 d\Gamma \right), \quad (24)$$

where the summation extends to all the sides of element i . This expression is similar to the one that has been used in [20,21] for compatible elements, and the $\frac{1}{2}$ factors are included to equally divide the contributions from the deformation jumps across an interface between elements. An alternative division of these terms could be based on using the local deformations of the interface, as quantified from \mathbf{u}_j , as a common approximation for the real deformations.

To ensure the correct dimensionality of the expression, coefficient a has the units of a stress, h of a length and the c_i are nondimensional. The values of coefficients c_i , the criteria to define the h for each side and the appropriate value for a have been determined by numerical experimentation [13].

10. ADAPTIVE PROCEDURES

The square cantilever presented in Figure 3 is modelled as a plane stress problem, and is used to show the characteristics of an h -adaptive strategy based on three different procedures for defining element error indicators. Triangular elements are used for all meshes, and the initial mesh for each procedure is also shown in Figure 3. The procedures for defining the error indicators are as follows:

- i) use dual solutions for compatible displacements and equilibrating stresses from the analyses of pairs of finite element models, with primitive elements and quadratic approximation functions;
- ii) use finite element equilibrium models with macro-elements and piecewise linear stress fields, and piecewise quadratic compatible displacement fields derived from the stress fields as in Section 8;
- iii) use finite element equilibrium models with primitive elements and quadratic approximation functions, and derive the error indicators from the compatibility defaults as in Section 9.2.

The general strategy for each procedure is based on:

h -refinement – this is performed by successively dividing each element or macro-element into 4 elements, until the new mesh has the required refinement level at each vertex of the current mesh;

optimality criteria – refinement is organised with the aim of achieving equal error in each element of the new mesh. The required element error is computed taking into account the global convergence rate. The refinement levels are based on the error indicators and local convergence rates;

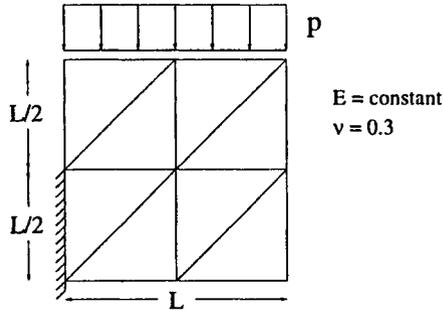


Figure 3. Square cantilever and initial finite element mesh.

iterative process – the target in each stage of refinement is to reduce the global error to one quarter of its current value.

Further details of this adaptive strategy, including the detection of singularities, are given in [13].

Table 1 shows the values of the actual relative errors of the equilibrium solutions $\|e_e\|/\|u\|$ and the corresponding effectivity ratios $\beta = \epsilon/\|e_e\|$, during the adaptive procedures. The actual errors were evaluated with reference to a very accurate solution obtained using procedure i) with elements of fourth degree and then applying a dual extrapolation to the last three pairs of energy values [13].

The numbers of (primitive) elements or macro-elements are indicated by n_E or n_{ME} respectively. For the initial meshes n_E or $n_{ME} = 8$.

Table 1
Refinement steps for the different procedures.

Procedure i)			Procedure ii)			Procedure iii)		
n_E	$\ e_e\ /\ u\ $	β	n_{ME}	$\ e_e\ /\ u\ $	β	n_E	$\ e_e\ /\ u\ $	β
8	0.473727	1.340	8	0.723374	3.322	8	0.473727	1.351
83	0.114939	1.644	26	0.340692	2.848	38	0.157006	1.346
311	0.020929	2.157	104	0.125922	2.664	185	0.036445	1.077
			434	0.036972	3.112			

Figures 4 and 5 present the results graphically as convergence graphs of error energy and effectivity ratio respectively, using log scales for error energy and the numbers of elements. It is apparent that although the behaviours of the effectivity ratios are very different, nevertheless the convergence rates of error for all three procedures are similar. The model with piecewise linear macro-elements starts with the largest error, but converges at a similar rate as for the models with the quadratic primitive elements as proved by Johnson [25]. In this example the effectivity ratio of procedure (iii) is clearly superior to the other procedures. This reflects the phenomenon of greater relevance being attributed to the worse solution when dual solutions are used to generate upper bounds.

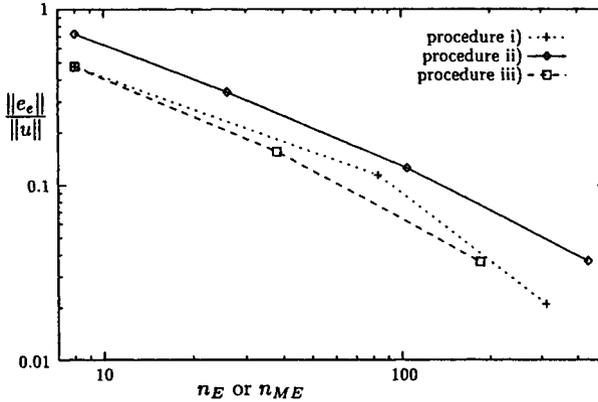


Figure 4. Convergence graphs for the different procedures.

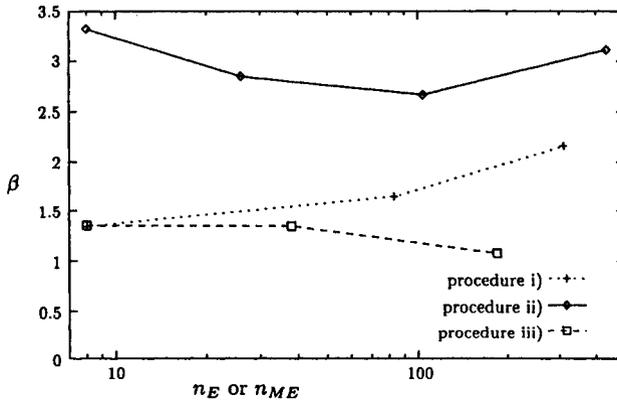


Figure 5. Effectivity of the different procedures.

As expected procedure (i), which is based on a pair of dual finite element solutions, gives better values of effectivity ratios compared with procedure (ii). Nevertheless as the convergence rate of the compatible solution is slightly lower than that of the equilibrium solution, in spite of using the same mesh, the quality of the estimator decreases during the adaptive process. On the contrary for procedure (ii) the quality of the estimator is not subject to such variations.

The initial and final deformed meshes for each procedure are presented in Figures 6 to 8. Figure 6 refers to procedure (i), and the deformations are shown for the hybrid displacement model.

Figure 7 refers to procedure (ii) which involves macro-elements. In this case the stress field in each primitive triangle is linear and corresponding element-wise continuous displacement fields are derived exactly since there are no compatibility defaults within the primitives. Rigid body positioning of the primitives is determined using the method of Section 8. These deformed elements are shown in Figure 7 before smoothing the displacements to obtain compatible displacement fields. It should be noted that procedure (ii)

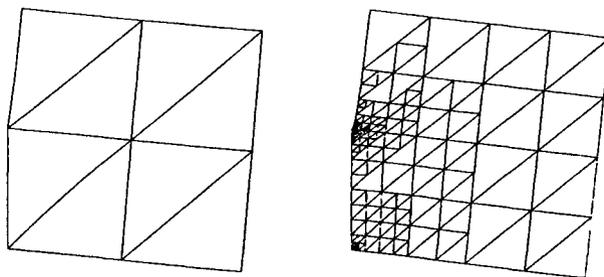


Figure 6. Displacements of the compatible finite element solution for procedure (i).

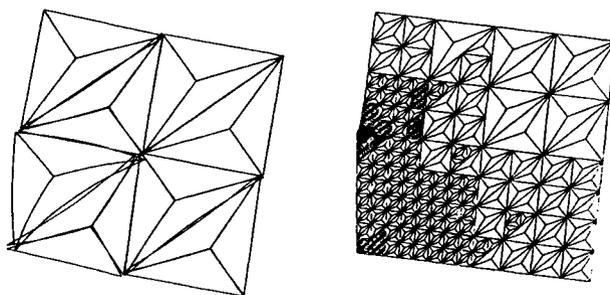


Figure 7. Element displacements, before smoothing, for procedure (ii).

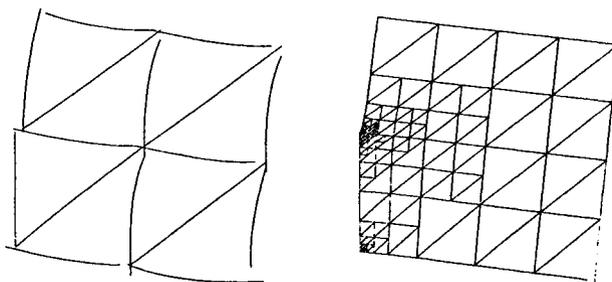


Figure 8. Side displacements for procedure (iii).

leads to unnecessary mesh refinement in some regions. More sophisticated displacement smoothing techniques may reduce this problem.

Figure 8 refers to procedure (iii), and the deformations are shown for the hybrid equilibrium model as discontinuous side displacements. Spurious kinematic modes are in evidence, particularly at the lower right hand corners of the meshes. It should be noted that procedure (iii) results in not only good non-diverging effectivity ratios for the global error, but also the minimum number of elements for the required convergence. Comparing the numbers of elements and the actual global errors for procedures (i) and (iii), it can be seen that procedure (i) converged a little faster than required and procedure (iii) converged a little slower.

11. CONCLUSIONS

For equilibrium finite element formulations three error indicators capable of driving adaptive refinement procedures were presented, all with good performance.

The error indicators obtained from an equilibrium solution only, have the advantage of requiring a smaller computational effort. The approach which uses a derived displacement field provides an upper bound of the solution error (which is often overestimated) to the user, as well as a compatible displacement field.

The error estimator based on the compatibility defaults is simpler, immune to SKM's (as long as the loading is admissible for the initial mesh), but needs a theoretical justification for the coefficients.

The full dual approach is, from a formal viewpoint the most appealing method, it is also immune to SKM's, but is computationally more expensive and may have the tendency to converge towards the error of the worse solution. Nevertheless it should be pointed out that this method is parallel in nature, so that both analyses can be performed independently.

Further investigation and experimentation are required before general conclusions can be made with confidence.

ACKNOWLEDGEMENTS

This work has been partially supported by ICIST, program PRAXIS XXI (2/2.1/CEG/33/94) and by the European Community (CHRX-CT93-0390). Initial collaboration between the two institutions at Exeter and Lisbon was enabled by the EC-STRIDE Portugal program.

REFERENCES

1. BMF deVeubeke, Displacement and equilibrium models in the finite element method, in: OC Zienkiewicz and GS Holister (eds.), *Stress Analysis* (Wiley, 1965).
2. JS Przemieniecki, *Theory of Matrix Structural Analysis* (McGraw-Hill, 1968).
3. J Robinson, *Integrated Theory of Finite Element Methods* (Wiley, 1973).
4. RH Gallagher, *Finite Element Analysis Fundamentals* (Prentice-Hall, 1975).
5. B Jakobsen, The Sleipner accident and its causes, in: J Robinson (ed.), *FEM Today and the Future* (Robinson and Associates, 1993), pages 102-108.

6. P Ladeveze, JP Pelle and PH Rougeot, Error estimation and mesh optimization for classical finite elements, *Engrg. Comput.* 8 (1991) 69–80.
7. RH Gallagher, JC Heinrich and N Sarigul, Complementary energy revisited, in: SN Atluri, RH Gallagher and OC Zienkiewicz (eds.), *Hybrid and Mixed Finite Element Methods* (Wiley, 1983), pages 453–465.
8. J Robinson, The mode-amplitude technique and hierarchical stress elements - a simplified and natural approach, *Int. J. Num. Meth. Eng.* 21 (1985) 487–507.
9. JPM Almeida and JAT Freitas, Alternative approach to the formulation of hybrid equilibrium finite elements, *Computers & Structures* 40(4) (1991) 1043–1047.
10. JPM Almeida and JAT Freitas, Continuity conditions for finite element analysis of solids, *Int. J. Num. Meth. Eng.* 33 (1992) 845–853.
11. J Jirousek and AP Zielinski, Dual hybrid-trefftz element formulation based on independent boundary traction frame, *Int. J. Num. Meth. Eng.* 36 (1993) 2955–2980.
12. EAW Maunder, JPM Almeida and ACA Ramsay, A general formulation of equilibrium macro-elements with control of spurious kinematic modes, *Int. J. Num. Meth. Eng.* 39 (1996) 3175–3194.
13. OJBA Pereira, Utilização de elementos finitos de equilíbrio em refinamento adaptativo, Ph.D. thesis, Technical University of Lisbon, (1996).
14. JPM Almeida and OJBA Pereira, A set of hybrid equilibrium finite element models for the analysis of three-dimensional solids, *Int. J. Num. Meth. Eng.* 39 (1996) 2789–2802.
15. OJBA Pereira and JPM Almeida, Equilibrium finite elements and dual analysis in three-dimensional elastostatics, in: ERA Oliveira and J Bento (eds.), *Education, Practice and Promotion of Computational Methods in Engineering using Small Computers* (Techno-Press, 1995), volume 2, pages 955–960.
16. THH Pian, Derivation of element stiffness matrices by assumed stress distributions, *AIAA Journal* 3 (1967) 1333–1335.
17. JAT Freitas, JPM Almeida and EMBR Pereira, Non-conventional formulations for the finite element method, *Structural Engineering and Mechanics* 4 (1996) 655–678.
18. S Pissanetzky, *Sparse Matrix Technology* (Academic Press, 1984).
19. Advanced Computing Department, Harwell Laboratory, Oxfordshire, Harwell Subroutine Library, (1991).
20. I Babuska and WC Rheinbolt, A posteriori error estimates for the finite element method, *Int. J. Num. Meth. Eng.* 12 (1978) 1597–1615.
21. DW Kelly, JPSR Gago and OC Zienkiewicz, A posteriori error analysis and adaptive processes in the finite element method: Part i – error analysis, *Int. J. Num. Meth. Eng.* 19 (1983) 1593–1619.
22. OC Zienkiewicz and JZ Zhu, The superconvergent patch recovery (spr) and adaptive finite element refinement, *Comp. Methods in Applied Mechanics and Engineering* 101 (1992) 207–224.
23. W Prager and JL Synge, Approximations in elasticity based on the concept of function space, *Quart. Appl. Math.* 5(3) (1947) 241–269.
24. P Ladeveze and D Leguillon, Error estimate procedure in the finite element method and applications, *SIAM J. Num. Anal.* 20(3) (1983) 483–509.
25. C Johnson and B Mercier, Some equilibrium finite element methods for two-dimensional elasticity, *Numer. Math.* 30 (1978) 103–116.

This Page Intentionally Left Blank

Error estimator and adaptivity for three-dimensional finite element analyses

P. Coorevits, J.-P. Dumeau and J.-P. Pelle

Laboratoire de Mécanique et Technologie, ENS de Cachan/CNRS/Université P. et M. Curie
61, avenue du Président Wilson, 94235 Cachan Cedex - FRANCE

1. INTRODUCTION

During the design phase of a structure, it is necessary to conduct several studies of the mechanical behavior. These computations often rely upon finite element analyses within an industrial setting. Each study consists of three steps: mesh generation, finite element analysis and post-processing and control of results. The finite element analysis, strictly speaking, no longer poses any difficulty on account of the numerous finite element software programs available on the market. However, with three-dimensional complex structures, the generation of the mesh is carried out interactively by specialized engineers or, at best, semi-automatically. The cost in human time is much greater than that to derive the finite element results. The post-processing and control of the results also generate sizable costs in term of human time. Moreover, in order to assure the quality of the computed results, the analyses are generally performed on several meshes with different sizes which strongly increases the overall cost of a design. Finally, it should be noted that even with today's powerful computers, it is illusory to think that the mesh just has to be refined in order to obtain a very accurate result. Indeed, a very fine and regular mesh for a three-dimensional structure quickly leads to saturating the available computational resources.

The purpose of this paper is to propose and develop procedures, available within an industrial setting, which allow us to automatically achieve finite element analyses of complex 3D structures while respecting the level of accuracy prescribed by the user and minimizing as much as possible the total cost of computation. Achieving this aim is essential in order for 3D finite element analyses to be used during the phase of design.

Two main difficulties must be overcome to reach this aim. The first difficulty consists of controlling the quality of the computations performed. To do so, we must be able to evaluate the discretization errors and then to define the element sizes necessary to obtain the prescribed accuracy as specified by the user. The second difficulty is to decrease the human time spent in the computation process by automating as much as possible the different steps of the computation and especially the mesh generation.

To overcome the first difficulty, we rely on the studies conducted in the domain of finite element control. Indeed, over the last 15 years, substantial advances have been made in the control of analysis quality and in the development of methods to actually quantify discretization errors. Three main approaches can generally be discerned:

- error estimators using the residuals from the equilibrium equations [1-4],
- error indicators obtained by comparing the finite element stress field with a smoothed stress field [5-7], and
- error estimators based both on the concept of error in constitutive relation and on adapted techniques for building admissible fields [8-10].

In the framework of this study, the error measures developed in the LMT are used and briefly presented in the second section. In particular, we have applied the results of the study in [10] which has led to the development of the error estimator software ESTEREF3D. It should be noted that this software program has been supplemented by a procedure which automatically takes into account the steep gradient areas for computing the optimal size map; the selected method is a 3D extension to the method proposed in 2D [11-12]. Obviously, any method that allows obtaining a global measure ϵ of the discretization errors as the local contributions ϵ_E of the elements to the global error can be used with the proposed procedures.

From an initial finite element analysis with a relatively coarse mesh, these tools allow us to define the size map which must be configured to reach the prescribed accuracy while minimizing the element number of the mesh and hence the computation cost. Thus, the main difficulty herein is to generate the 3D mesh while respecting the imposed sizes. While there are many automatic 2D meshers able to correctly respect a size map, such tools don't actually exist in 3D. Presently, the most efficient automatic meshers on the market (for example, GHS3D [13]) are only able to mesh the volume from the skin mesh while assuring that the tetrahedral elements generated are not overly stretched. However, respecting a 3D size map is still at the research stage. To overcome this difficulty, several procedures for helping to pilot meshers are proposed. These procedures maximize the use of the software programs available within an industrial setting; they are entirely automated and allow a reasonable conformance within a 3D size map, even for relatively complex geometries. With the automatic procedures, naturally coupled to an efficient 3D automatic mesher, generating an optimized mesh, which can necessitate up to ten hours of an engineer's time, requires only several minutes of CPU time on a work station.

The definition of an optimal mesh, the extension to 3D of the automatic method that takes the steep gradient areas into account and the definition of the nodal sizes are all presented in the third section. The different procedures for piloting meshers and the automation procedure are detailed in the fourth and fifth sections respectively. All these procedures are applied to industrial geometries.

2. ERROR IN CONSTITUTIVE RELATION

To set the framework, we consider herein the problem of the analysis of a structure in elasticity [9]. Suppose that \hat{U} is a kinematically-admissible displacement field, i.e. it satisfies the kinematic constraints and that $\hat{\sigma}$ is a statically-admissible stress field, i.e. it satisfies the equilibrium equations. The quantity:

$$\hat{\epsilon} = \hat{\sigma} - K\epsilon(\hat{U}) \quad (1)$$

where K denotes the Hooke's tensor is called the *error in the constitutive relation* associated with the pair $(\hat{U}, \hat{\sigma})$. If $\hat{\epsilon}$ is equal to zero, the pair is the solution to this problem. Otherwise, $\hat{\epsilon}$ allows us to estimate the quality of $(\hat{U}, \hat{\sigma})$ as an approximate solution. To measure the error $\hat{\epsilon}$, we use the standard energy norm over the whole structure:

$$e = \|\hat{\epsilon}\|_{\Omega} = \|\hat{\sigma} - K\epsilon(\hat{U})\|_{\Omega} \quad \text{with} \quad \|\sigma\|_{\Omega} = \left[\int_{\Omega} \sigma^T K^{-1} \sigma d\Omega \right]^{1/2} \quad (2)$$

From the absolute error, we define a relative error:

$$\epsilon = \frac{\|\hat{\sigma} - K\epsilon(\hat{U})\|_{\Omega}}{\|\hat{\sigma} + K\epsilon(\hat{U})\|_{\Omega}} \quad (3)$$

is sufficiently smooth, we then have [16-17]:

- $q=1$ for the 4-node tetrahedra, and
- $q=2$ for the 10-node tetrahedra.

If the exact solution includes a singularity of strength α on the structure, the convergence rate q of the error is defined by [16-17]:

$$q = \min(\alpha, p) \quad (7)$$

In practice, the global convergence rate varies between α and p . Thus, the theoretical result of convergence cannot be used directly. We observe numerically [11] that the convergence rate p_E of the elements connected to the singularity is close to α , whereas it displays a value close to p for the other elements. Hence, we propose a two-step procedure as in the 2D:

- detection of singular areas (and more generally steep gradient areas), and
- if the area is detected, we evaluate the value of the coefficient p_E ; else we use $p_E = p$.

3.2.1 Detection of the steep gradient areas

The detection of the steep gradient regions is enabled by the computation of the local errors $\bar{\epsilon}_E$, defined by:

$$\bar{\epsilon}_E^2 = \frac{|\Omega|}{|E|} \epsilon_E^2 \quad (8)$$

where ϵ_E denotes the contribution to the relative error of the element E , $|E|$ the volume of the element and $|\Omega|$ the volume of the structure.

Indeed, we can observe that the local errors are larger in the singular areas (Figure 5) than in the other areas. Thus, a node i of the mesh will be considered as singular if the average \bar{m}_i of local errors $\bar{\epsilon}_E$ for the elements connected to the node and the average \bar{M}_i of local errors $\bar{\epsilon}_E$ on the whole structure satisfy:

$$\bar{m}_i \geq \delta \bar{M}_i \quad (9)$$

where δ is a coefficient. The numerical experiments in 3D lead to $\delta = 3$.

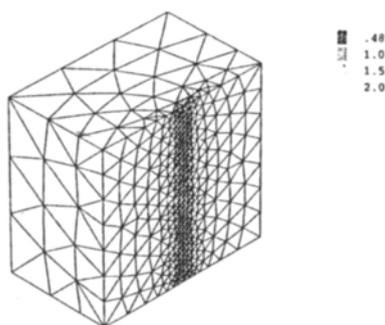


Figure 1. Detection of the steep gradient areas in 3D

The detection of the steep gradient areas in 3D for a problem of a crack in mode I is shown in Figure 1. The above test allows us to define the areas where the coefficients p_E of the

as the contribution to the relative error of an element of the mesh E :

$$\varepsilon_E = \frac{\|\hat{\sigma} - K\varepsilon(\hat{U})\|_E}{\|\hat{\sigma} + K\varepsilon(\hat{U})\|_\Omega} \text{ with } \|\sigma\|_E = \left[\int_E \sigma^T K^{-1} \sigma dE \right]^{1/2} \quad (4)$$

The global measure ε allows us to quantify the global quality of the approximation $(\hat{U}, \hat{\sigma})$ and the local contributions ε_E to localize the errors on the structure.

To apply this process, a post-processing of the finite element solution (U_h, σ_h) must be carried out in order to build an admissible displacement-stress $(\hat{U}_h, \hat{\sigma}_h)$ pair from the data of the reference problem (3) and the solution (U_h, σ_h) .

Within the framework of the finite element method in displacement, the displacement field U_h is kinematically admissible. For the sake of simplicity, we generally choose:

$$\hat{U}_h = U_h \quad (5)$$

On the other hand, the computed stress σ_h is not statically admissible. It is necessary, therefore, to build a stress field $\hat{\sigma}_h$ that satisfies the equilibrium equations. This construction is carried out in two steps:

- building on the edges of force densities in equilibrium with volumic loadings, and
- building of a simple solution for the equilibrium equations in each element.

3. DETERMINATION OF A 3D OPTIMIZED SIZE MAP

3.1. Definition of an optimal mesh

The aim of all adaptive procedures is to provide the user with a mesh that respects the level of accuracy ε_0 at a minimal computational cost. The present study will be restricted to the h-version which is the most commonly used procedure: one modifies the size and the topology of elements while preserving the same type of shape functions for the various meshes. Moreover, tetrahedral elements will be used because automatic meshers only work for this kind of element.

A mesh T^* is considered optimal with respect to a measure of the error ε if [14]:

$$\begin{cases} \varepsilon^* = \varepsilon_0 & \text{prescribed accuracy} \\ N^* & \text{minimal (number of elements of } T^*) \end{cases} \quad (6)$$

This criterion of optimization naturally leads to a minimization of computation costs. To solve problem (6), the following procedure is used:

- (i) an initial analysis is performed on a relatively coarse mesh T ,
- (ii) the global error ε and the local contributions ε_E are computed for this mesh and
- (iii) the characteristics of the optimal mesh T^* are determined.

The optimized mesh T^* is built and a second finite element analysis is carried out.

3.2. Taking into account the singularities

The determination of a size map uses the convergence rate q of the error ε as a function of the element size h . In elasticity, the error in constitutive relation exhibits the same behavior as the classical error between the exact and finite element solutions [15]. If the exact solution

singularities must be estimated.

3.2.2 Computation of the coefficient p_E

The computation of the convergence coefficients p_E is based, as in 2D, on the computation of the finite element energy density \bar{e}_h in the steep gradient areas. In 3D, isolated singularities (however, this case is unusual) and singularities on the edges can appear.

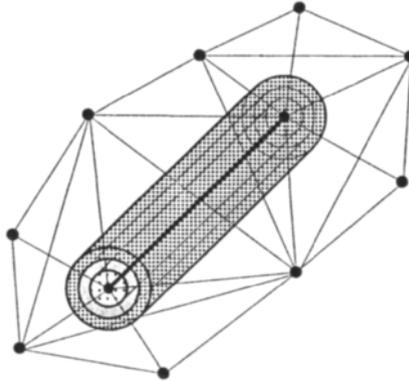


Figure 2. Computation of the coefficients p_E by cylinders

If the method of detection identifies two nodes at the extremities of a same edge as singular, the edge is considered as singular. In this case, the middle energy density \bar{e}_h of the co-axial cylinders C built on the edge (Figure 2) is evaluated:

$$\bar{e}_h = \frac{1}{\text{vol}(C)} \int_C \text{Tr}[\epsilon(U_h) K \epsilon(U_h)] dC \quad (10)$$

The energy density is identified by the least squares method with the theoretical value:

$$e(r) = kr^{2(\alpha-1)} + c \quad (11)$$

where k and c are coefficients which depend on the mechanical problem and α is the strength of the singularity. A coefficient $\bar{\alpha}$ close to the searched value is therefore identified.

In the case of isolated nodes, a similar technique is used with concentric spheres. The identified value of the coefficients p_E for a problem of a clamped cube is shown in Figure 3. The value is close to 0.8 on the singular lines of the clamping.

In practice, this method of identification allows us to account for not only the singularities (clamping, crack tip, etc.), but also the steep gradient areas which are not mathematically singular (for example, in the case of a fillet radius with a strong curvature). In addition, we find that the body loadings are very smooth for most static problems in elasticity, and that the steep gradients are at the edge of the structure. Under these conditions, the computation is performed only for the nodes close to the boundary and consequently requires very little CPU time.

3.3. Determination of the size map

Taking into account the steep gradient areas in the determination of the size map necessitates, as in 2D, both a modification to the problem (6) in order to introduce the

coefficients p_E and a more precise definition of the size coefficients which allows the size to vary more rapidly in the steep gradient areas. To do so, nodal size coefficients are introduced and inside an element of the initial mesh, a hypothesis of the linear distribution of the volume is forwarded. The element number N_E^* can then be evaluated by:

$$N_E^* = \frac{1}{|E|} \int_E \frac{h_E^3}{\sum_{i=1}^n h_i^{*3} \lambda_i} dE \quad (12)$$

and the contribution ϵ_E^* of an element to the global error ϵ^* by:

$$\epsilon_E^{*2} = \frac{1}{|E|} \int_E \frac{\left(\sum_{i=1}^n h_i^{*3} \lambda_i \right)^{\frac{2p_E}{3}}}{h_E^{2p_E}} \epsilon_E^2 dE \quad (13)$$

where n denotes the number of vertices, h_E its size, p_E the computed convergence coefficient and λ_i the barycentric coordinates. Lastly, h_i^* are the prescribed sizes on the mesh \mathbf{T}^* computed at the nodes of the present element.

The problem can be formulated in the following manner:

$$\text{Minimize } N^* = \sum_E N_E^* \text{ with } \epsilon_0^2 = \sum_E \epsilon_E^{*2} \quad (14)$$

This optimization problem is numerically solved; its solution provides at every node of the mesh the size for building the optimized mesh \mathbf{T}^* .

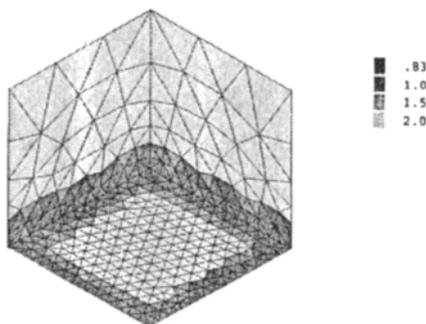


Figure 3. Distribution of the coefficients p_E

3.4. Control of the quality of the size map

The control of optimality is crucial in judging the quality of the results. In some cases, the prescribed accuracy can be obtained without a minimum number of elements. Control is achieved by the following method: on the optimized mesh, if the computed error ϵ^* is near the prescribed error ϵ_0 , we compute again the size map for an imposed error equal to ϵ^* . If the mesh \mathbf{T}^* is perfectly optimal, the size modification coefficient r_E must be equal to one. In practice, such meshes don't exist and the mesh is considered as optimal if:

$$r_E = \frac{h_E^*}{h_E} = \frac{\text{prescribed size}}{\text{present size}} \text{ and } 2/3 \leq r_E \leq 3/2 \quad (15)$$

Results are then visualized on a histogram that represents an overall study of optimality.

4. ASSISTANCE IN PILOTING 3D MESHES

To automate the building of the three-dimensional mesh with respect to the size map, two piloting procedures have been developed:

- The first and simpler method consists of generating a mesh whose boundary respects the size map and then to use an automatic mesher to generate the volumic mesh.
- The second uses possibilities of mesher GHS3D developed by INRIA [13]. Once the skin mesh has been generated, the user can specify the prescribed size at some internal nodes. During the building of the volumic mesh, these nodes are used as vertices of the tetrahedra and the prescribed size is in conformance in the vicinity of the control nodes. An automatic strategy for choosing their position is presented.

The procedures require a good mesh of the surface of the structure. Indeed, most 3D automatic meshers available within an industrial setting usually use this mesh as a starting point.

If the structure's skin to be meshed is constituted of plane surfaces, it is relatively easy to build the mesh with one of the 2D plane meshers which are able to respect a size map. This technique can also be envisioned for sufficiently simple curved surfaces (parts of cylinders, cones and spheres) by using a projection on a plane surface.

However, the situation is more complicated for the complex curved surfaces defined by patches. Meshers which are able to respect a map size for any curved surface don't exist to our knowledge. It must be noted that such a skin mesher is being developed [18], but the adaptation of the mesh still requires some interactive work.

4.1. Method by optimization of the skin mesh

Many automatic volumic meshers use as a starting point a discretization of the skin object. This discretization is analyzed and is useful in determining the sizes assigned to the skin and thus the size of the volumic elements which are connected. By continuity, the sizes of the elements in the volume are determined by predefined criteria (arithmetic or geometric evolution) which only take into account the data on the skin.

An initial method (called METH1) to control the element sizes consists of optimizing the surfaces of the structure with a mesher that is able to respect the size map and allowing the 3D mesh generator to run without any other information. If the variation in sizes is smooth in the interior, we can hope to correctly adapt the 3D mesh. It is obvious that this technique is inadequate when the internal variations are complex. Nevertheless, this very simple technique quickly enables obtaining an initial result.

4.2. Method by the points of size control

The second method uses a specificity of the mesher GHS3D developed in INRIA. In addition to the classical data of the skin mesh, GHS3D accepts the data of internal vertices coupled with sizes which are to be respected in the neighborhood of the vertices. However, the internal vertices are obligatory points of passage, that is to say nodes of the future mesh. Hence, it is necessary to limit the number of these vertices and to verify their relative positions, in order not to create any stretched elements or not to respect the prescribed sizes.

This technique requires the development of an automatic procedure for the choice of the internal control vertices. It uses the initial mesh (or previous one) and is based on the direct analysis of the size map. To do so, we compute for every node n :

$$S_n = \text{Sup}_i \left(\frac{h_i^*}{h_n^*} \right); I_n = \text{Inf}_i \left(\frac{h_i^*}{h_n^*} \right) \quad (16)$$

where h_i^* describe the requested size at node i , with i varying on each node of the elements connected to node n . If the sizes gradients S_n or I_n are more than β and less than $1/\beta$ respectively, the node of the initial mesh is kept as the vertex of control P_i . Thus, the control vertices are positioned in the regions where the size gradients are steep.

In order to both limit the number of control vertices and not constrain too heavily the mesh, some criteria of relative distance and distance to the surface have been introduced. Every point P_i is then tested:

$$d(P_i, P_j) \geq \text{prescribed size in } P_i \quad (17)$$

where P_i denotes a control vertex or a node near the skin mesh. If the relation is not satisfied, then the control vertex is not created.

The coefficient β is a parameter of the method. Different examples show that a reasonable value is near 1.3 to 1.5.

This method is almost as simple to carry out as the method by optimization of the skins, but it does allow for an important improvement of the size respect with, consequently, a significant decrease in the element number.

4.3. Control of the size respect

To estimate the quality of the meshing procedures, an indicator of size respect is defined in comparing real sizes created by the mesher to the sizes actually defined by the map of sizes:

$$I_E = \frac{\text{real size}}{\text{prescribed size}} \quad (18)$$

The ideal ratio is equal to 1; however, the respect of sizes is considered satisfactory if the coefficient is such that: $2/3 \leq I_E \leq 3/2$. Results are visualized on a histogram with respect to coefficients r_E that represent an overall study of the respect of sizes.

This indicator allows us to compare the different methods. The analysis of both size respect (18) and optimality (15) coefficients can enable discerning when the mesh is not optimal, whether this is due to poor respect of the size map or an incorrect computation of the sizes.

4.4. Examples of mesh adaptation

To illustrate the methods discussed above, two examples are presented. The first is a symmetric geometry with a low radius of curvature. We show in Figure 4 the initial mesh as meshed with 4-node tetrahedra and in Figure 5 the meshes obtained both with the method of optimizing the skin mesh (METH1) and with the method of control vertices (METH2). The prescribed error is 20%.

The second example is a simplified structure meshed with 10-node tetrahedra. For symmetry reasons, only an eighth of the structure has been meshed. The optimized mesh (Figure 8) has been obtained from the initial mesh (Figure 7) for a prescribed error of 7%. The size respect and modification coefficients are shown in Figure 10 and 11, respectively.

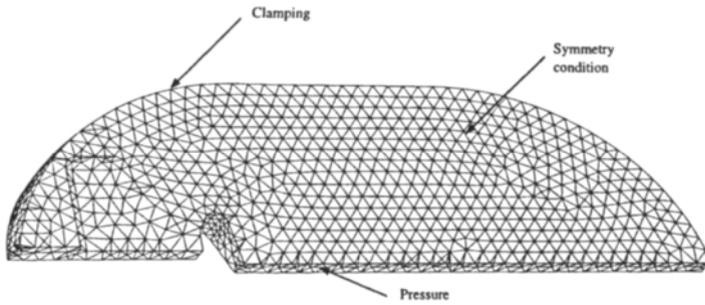


Figure 4. Initial mesh - 17,229 elements - 3,755 nodes - $\epsilon=40.14\%$

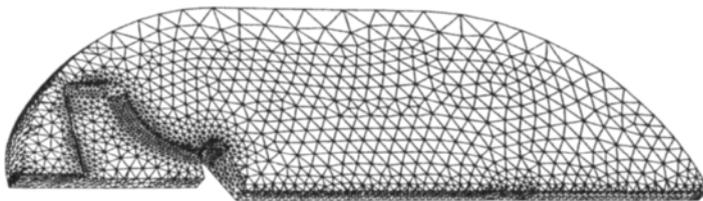


Figure 5. Optimised mesh
 METH1 - 31,510 elements - 6,800 nodes - $\epsilon=21.55\%$
 METH2 - 25,170 elements - 5,757 nodes - $\epsilon=23.13\%$

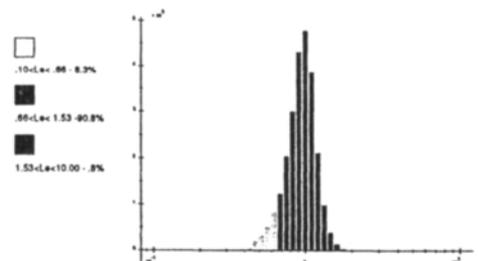
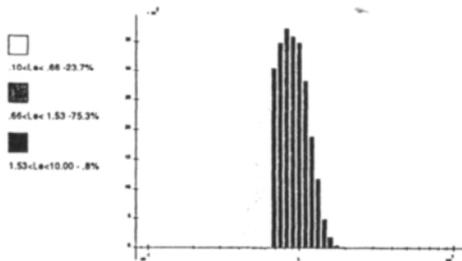


Figure 6. Coefficient of size respect - METH1 Figure 7. Coefficient of size respect - METH2

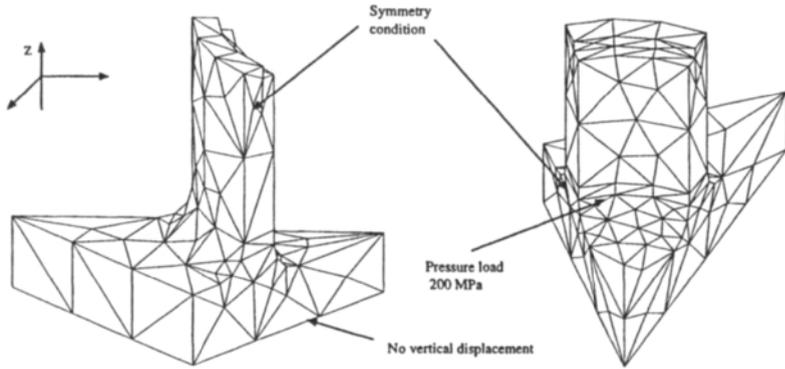


Figure 8. Initial mesh - 5,132 elements - 8,350 nodes - $\epsilon=13.57\%$

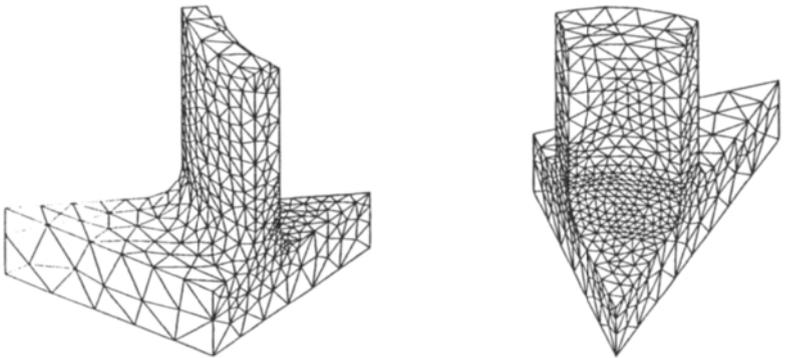


Figure 9. Optimised mesh - 5,008 elements - 8,012 nodes - $\epsilon=6.97\%$

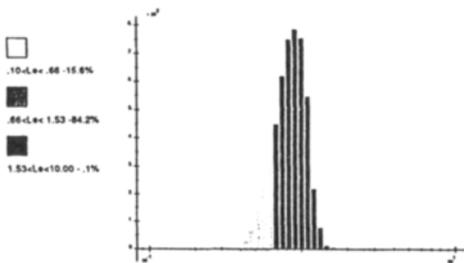


Figure 10. Coefficient of size respect

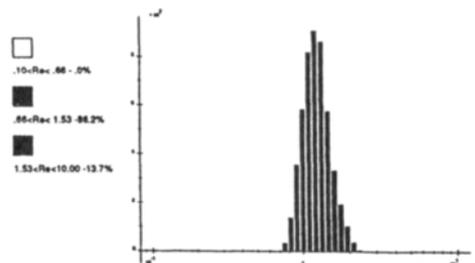


Figure 11. Coefficient of size modification

The first example shows a decrease of 6,300 elements between the method of optimizing the skin mesh and the method of the control vertices with errors close to 20%. The same observation can be made for the size respect where the second method creates 90.8% of the elements in the admissible area (Figure 5), while the first method only creates 75.3% (Figure 6). For the second example, we notice that the optimized mesh contains fewer elements than the initial mesh, while its error has been divided by two.

5. AUTOMATION OF THE COMPUTATIONS

5.1. Algorithm

The aim of automation is to minimize the involvement of the user in the analysis itself. In order to do so, we need the following tools:

- (i) an efficient and conservative measure of the discretization errors,
- (ii) an efficient method for determining the optimal sizes which requires no knowledge of the mechanical problem, and
- (iii) in 2D, an automatic mesh generator that is able to respect the size map and in 3D, an automatic mesh generator for use with the above procedures.

With these tools, it is possible to entirely automate analyses in elasticity. More precisely, the involvement of the user should be limited to the following operations:

1. a description of the geometry by a CAD-CAM software program,
2. a description of the mechanical data (imposed displacements, loadings, coefficients of elasticity of the materials), and
3. an indication of the prescribed accuracy ε_0 .

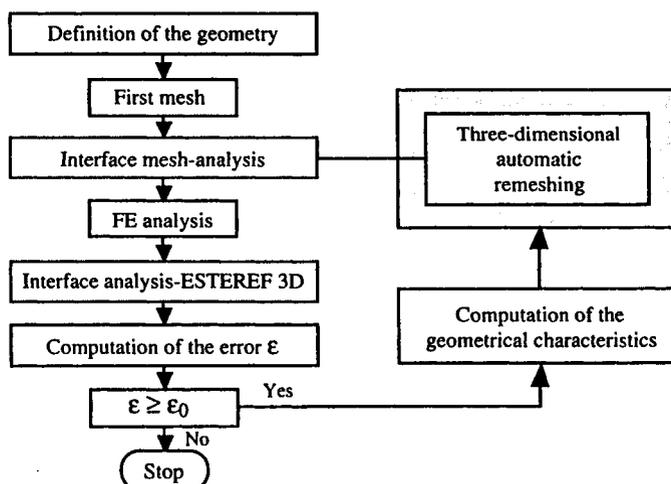


Figure 12. Diagram of the 3D analysis

From a coarse mesh T_0 , the algorithm described in Figure 12 provides in just a few iterations

an optimized mesh for the accuracy ϵ_0 . When the first mesh is built, we conduct an initial finite element analysis, whose accuracy ϵ is computed with our post-processor. Then, this accuracy is compared with the prescribed accuracy ϵ_0 . Since mesh T_0 is coarse, we generally have $\epsilon > \epsilon_0$. Thus, an iteration of optimization is required to attain the accuracy ϵ_0 . If ϵ is not too high compared to ϵ_0 , in practice if:

$$\epsilon \leq \theta \epsilon_0 \quad (19)$$

where θ is a coefficient, we choose as target value:

$$\epsilon_{\text{target}} = \epsilon_0 \quad (20)$$

Otherwise, the target value is imposed by:

$$\epsilon_{\text{target}} = \epsilon / d \quad (21)$$

where d denotes a coefficient. Experimentally, we have observed that the values:

$$\theta = 4 \quad \text{and} \quad d = 3 \quad (22)$$

yield good results for 2D and 3D computations, respectively.

A map of mesh sizes has been computed with ϵ_{target} as an imposed accuracy. The new mesh is built with the above procedures and automatic mesh generators. Then, we begin another iteration by conducting a new finite element analysis. After a few iterations (one to four, depending on the initial mesh and ϵ_0), the prescribed accuracy is reached.

5.2. Examples of automation

A piece of wheel is meshed (Figure 13) with 10-node tetrahedra by using the second method of adaptation and the automation. From the initial mesh, the final mesh (Figure 15) has been built with an intermediate step (Figure 14) for a prescribed error of 10%.

The second example is the simplified structure meshed with 10-node tetrahedra. The optimized mesh (Figure 18) is built from the initial mesh (Figure 16) with an intermediate step (Figure 17). The prescribed error is 5%.

CONCLUSION

The examples presented show that it is possible to generate well-adapted 3D meshes with entirely automated procedures. The methods employed are very simple to develop and their computation costs remain far below those of the finite element analysis. Associated with an efficient technique of optimal size prediction which takes the steep gradient areas automatically into account, these methods allow for a real automation of the 3D finite element analyses, at least in elasticity.

We have presented a procedure that allows obtaining adapted meshes in 3D by the use of an automatic surface mesher that is able to respect a map of sizes, an automatic 3D mesher and a post-processor for the control of finite element analyses.

Naturally, several difficulties remain to be overcome, such as:

- the adaptation of curved surface meshes which is essential for the generation of well-adapted 3D meshes and
- the improvement of a CAD-computation connection; the transfer of a geometry to an automatic mesher is far from being adequately solved.

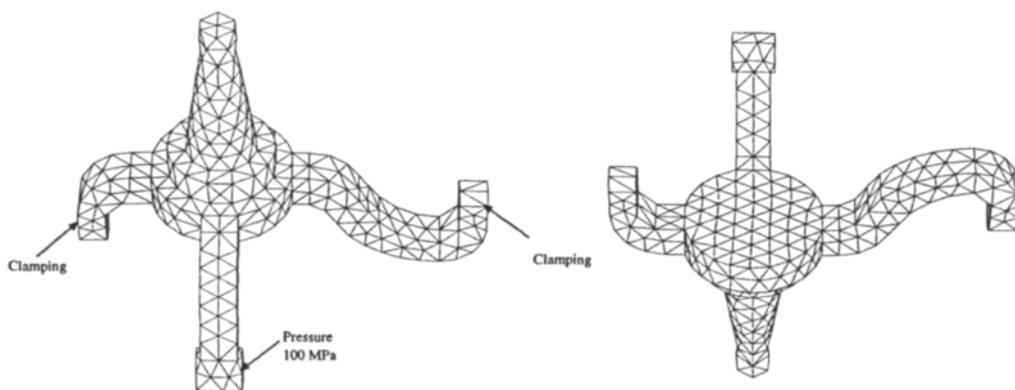


Figure 13. Initial mesh - 1,731 elements - 3,435 nodes - $\epsilon = 36.89\%$

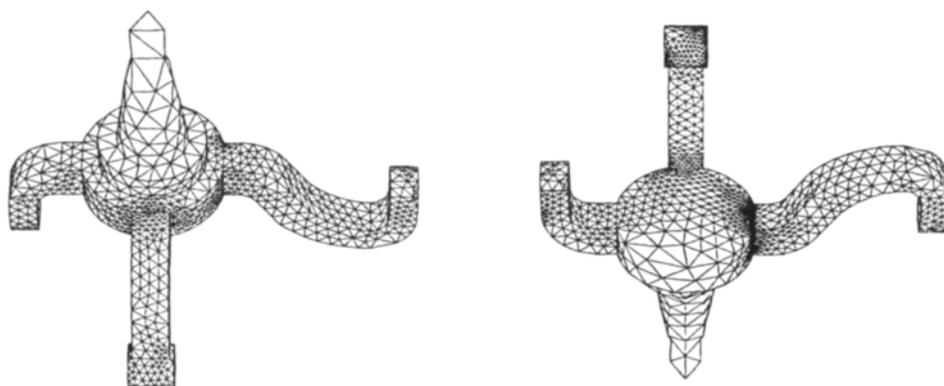


Figure 14. Intermediate mesh - 14,105 elements - 23,268 nodes - $\epsilon = 16.16\%$

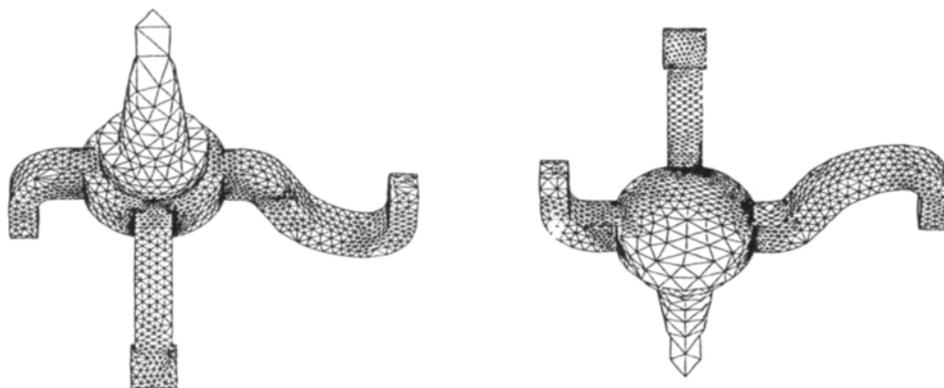


Figure 15. Final mesh - 25,018 elements - 40,880 nodes - $\epsilon = 11.52\%$

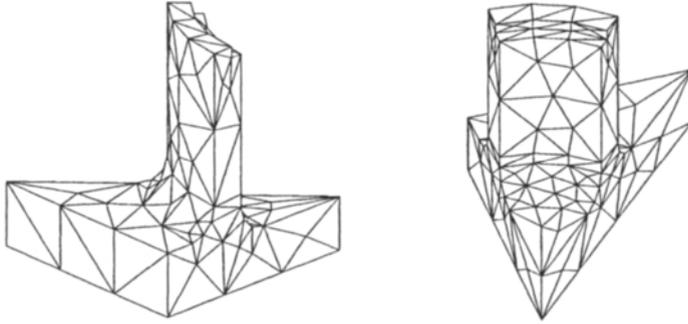


Figure 16. Initial mesh - 355 elements - 777 nodes - $\epsilon=34.30\%$

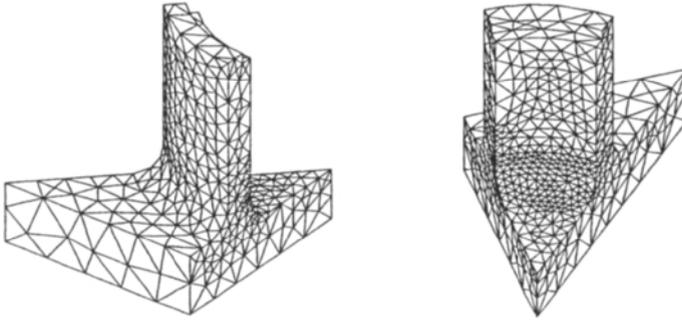


Figure 17. Intermediate mesh - 3,364 elements - 5,673 nodes - $\epsilon=9.95\%$

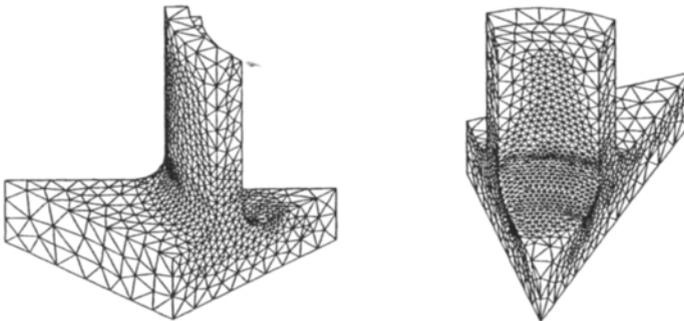


Figure 18. Final mesh - 10,249 elements - 16,346 nodes - $\epsilon=4.88\%$

REFERENCES

1. Babuska I., Rheinboldt W.C., Error estimate for adaptive finite computation, *SIAM J. Num. Anal.*, Vol. 15, N°4, pp. 736-754, 1978.
2. Gago J.P., Kelly D.W., Zienkiewicz O.C., Babuska I., A posteriori error analysis and adaptive processes in the finite element method: part II - Adaptive mesh refinement, *Int. Jour. Num. Meth. Engng.*, Vol. 19, pp. 1621-1656, 1983.
3. Kelly D.W., Gago J.P., Zienkiewicz O.C., Babuska I., A posteriori error analysis and adaptive processes in the finite element method : part I - Error analysis, *Int. Jour. Num. Meth. Engng.*, Vol 19, pp. 1593-1619, 1983.
4. Ohtsubo H., Kitamura M., Element by element a posteriori error estimation of the finite element analysis for three dimensionnal elastic problems, *Int. J. Num. Meth. in Engng.*, Vol. 29, pp. 223-244, 1992.
5. Zienkiewicz O.C., Zhu J.Z., A simple error estimator and adaptive procedure for practical engineering analysis, *Int. J. Num. Meth. Engng.*, Vol. 24, pp. 337-357, 1987.
6. Zienkiewicz O.C, Zhu J.Z., Gong N.G, Effective and practical h-p version adaptive analysis procedures for the finite element method, *Int. Jour. Num. Meth. Engng.*, Vol. 28, pp. 879-891, 1989.
7. Zienkiewicz O.C, Zhu J.Z., The superconvergent patch recovery and a posteriori error estimates: part I - The recovery technique, *Int. Jour. Num. Meth. Engng.*, Vol 33, pp. 1331-1364, 1992.
8. Ladevèze P., Nouvelles procédures d'estimations d'erreur relative à la méthode des éléments finis et applications, *Proc. Journées Elements Finis*, Rennes, 1977.
9. Ladevèze P., Pelle J.P., Rougeot P., Error estimation and mesh optimization for classical Finite Elements, *Engng. Computation*, Vol. 8, pp. 69-80, 1991.
10. Gastine J.L., Ladevèze P., Marin P., Pelle J.P., Accuracy and Optimal Meshes in Finite Element Computation for nearly incompressible Materials, *Comp. Meth. in Appl. Mech. and Engng*, Vol. 94, pp. 303-315, 1992.
11. Coorevits P., Ladevèze P., Pelle J.P., Mesh optimization for problems with steep gradients, *Engineering Computations*, Vol. 11, pp. 129-144, 1994.
12. Coorevits P., Ladevèze P., Pelle J.P., An automatic procedure for finite element analysis in 2D elasticity, *Comp. Meth. in Appl. Mech. and Engng*, Vol. 121, pp. 91-120, 1995.
13. Georges P.L., Automatic mesh generation: application to finite element methods, *Collection RMA*, Ed. Masson, 1993.
14. Ladevèze P., Coffignal G., Pelle J.P., Accuracy of elastoplastic and dynamic analysis, Accuracy estimates and adaptative refinements in Finite Element computations, Babuska, Gago, Oliveira, Zienkiewicz Editors, J. Wiley, Chapter 11, pp. 181-203, 1986.
15. Ladevèze P., Leguillon D., Error estimate procedure in the finite element method and applications, *SIAM J. Num. Anal.*, Vol. 20, n° 3, pp. 485-509, 1983.
16. Ciarlet P., *The finite element method for elliptic problem*, North Holland, 1978.
17. Strang G., Fix G., *An analysis of finite element*, Prentice-Hall, 1976.
18. Noel F., Leon J.C. , Trompette P., A new approach to free form surface mesh control in a CAD environment, *Int. J. for Num. Meth. in Engng.*, Vol. 38, pp. 3121-3142, 1995.

This Page Intentionally Left Blank

PART 6

ERROR ESTIMATORS AND MESH ADAPTIVITY FOR VIBRATION, ACOUSTICS AND ELECTROMAGNETICS PROBLEMS

This Page Intentionally Left Blank

Error estimation and adaptivity for h-version eigenfrequency analysis

Nils-Erik Wiberg and Patrik Hager

Department of Structural Mechanics, Chalmers University of Technology

S-41296, Gothenburg, Sweden

Abstract

The Superconvergent Patch Recovery technique is used to improve the displacement field of vibrating eigenmodes. The improved eigenmodes are used in the Rayleigh quotient to obtain improved eigenfrequencies, which replace the unknown exact solution in the error estimates. The improved eigenmodes are used to determine the error locally as error indicator and is the fundamental ingredient in the adaptive strategy.

KEYWORDS: Adaptivity; Eigenfrequency; Error estimation

1. INTRODUCTION

Quality aspects have received increased interest in recent years in the engineering community. In particular quality assessed computations. The most general tool for engineering computations is recognized to be the Finite Element Method, FEM. Dynamic problems require, in general, eigenfrequencies and eigenmodes with some specified accuracy. Techniques to solve the algebraic eigenproblem efficiently and with a demanded accuracy are well established. However, by using FEM we also introduce a discretization error, which must be controlled for the solution to be acceptable. Engineering needs create large problems, which require adaptive procedures in order to be solvable. The final goals of adaptivity in modal analysis are three fold:

- 1) optimal mesh for one eigenfrequency;
- 2) optimal mesh for a set of eigenfrequencies;
- 3) quality assessed dynamic response.

Work have been conducted on the topic, and this work is a continuation of the work presented in [1] and it meets 1) and partly 2).

A posteriori error estimation is a fundamental ingredient in an adaptive procedure. The basic idea in *a posteriori* error estimation is simply to construct a new higher quality solution to replace the unknown exact solution in the error expression. Zienkiewicz and Zhu, [2], [3], devised the Superconvergent Patch Recovery, SPR, technique as an efficient mean to construct improved stress fields for static problems. The SPR technique has been applied to static problems and dynamic wave problems to improve the stress field and to estimate the spatial discretization error,

[4], [5], [6], and [7]. Ladeveze and Pelle, [8], proposed a technique to obtain upper and lower bounds for the eigenfrequencies of elastic structures. The technique requires global systems of equations to be solved and the computation of a constant characteristic of the discretization subspace. Avrashi and Cook, [9], proposed a technique for C^0 eigenproblems. The procedure starts with smoothing of the stress field and then the smoothed stresses are used to modify the displacement field. An improved eigenfrequency is then obtained by the Rayleigh quotient. The technique requires user tuned parameters. The current approach for eigenproblems is to improve the displacements directly and obtain an improved eigenfrequency by the Rayleigh quotient. We improve the displacements by SPR techniques, henceforth referred to as SPRD techniques, see Wiberg *et al*, [10], [11], [1]. In [12], an h-adaptive eigenfrequency analysis was presented based on the estimated strain energy of the error and an accepted error tolerance of the strain energy, and was not directly coupled to the error in eigenfrequencies. Stephen and Steven, [13], proposed another approach to improve the displacements where the differential equation is solved for local patches and the superconvergent nodal values are used as boundary conditions.

However, the previously mentioned improvement techniques for eigenproblems either suffer from limitations in reliability or a low accuracy combined with a low order of convergence or they require user tuned parameters. The node patch SPRD technique advocated in [1] may fail when applied to general quadrilateral elements due to the fitting of a high degree polynomial to a few points giving a highly distorted behavior between the sampling points. The technique proposed in [13] has a low order of convergence, indicating that the superconvergence properties can be better utilized. The drawback of the SPRD technique can be remedied by decreasing the order of the polynomial used the recovery. We exploit the technique in an h-adaptive algorithm for which the mesh refinement strategy is based on the difference of strain energies of the FE solution and the improved solution, and coupled directly to the accepted error in eigenfrequencies.

Numerical examples show that it is possible to use the node patch strategy to obtain eigenfrequencies to a specified accuracy within a few iterations for complicated problems. We also include numerical examples that show an increase of order of convergence which demonstrates superconvergence. Moreover, the accuracy of the node patch SPRD technique using linear triangular and bilinear quadrilateral elements is compared to the accuracy of the corresponding one order higher finite element solution.

2. BASIC EQUATIONS

Consider elliptic eigenvalue problem of order $2m$ which are governed by the following differential equation

$$-\lambda \rho u(x) + \bar{\nabla}^T D \bar{\nabla} u(x) = 0, \quad \text{in } \Omega \quad (1)$$

with boundary conditions

$$u(x) = u_b \quad \text{on } \Gamma_u, \quad (2)$$

$$\bar{\nabla}_n^T D \bar{\nabla} u(x) = \sigma_b \quad \text{on } \Gamma_\sigma \quad (3)$$

where $\bar{\nabla}$ is the differential operator, $\bar{\nabla}_n$ is the boundary operator, \mathbf{D} is the constitutive matrix and ρ is the mass density. Ω is a spatial domain under consideration with boundary $\Gamma = \Gamma_u \cup \Gamma_\sigma$ ($\Gamma_u \cap \Gamma_\sigma = \emptyset$).

The solutions of equation (1) are the functions which give the Rayleigh quotient

$$R(v) = \frac{a(v, v)}{b(v, v)} \quad (4)$$

a stationary value when taken over all possible functions $v \in \mathcal{V}(\Omega)$, where $\mathcal{V}(\Omega) = \{\mathbf{V}v : v \in \mathcal{H}^1(\Omega), v = 0 \text{ on } \Gamma_u\}$. Here $b(v, w)$ is a kinematic energy scalar product on Ω as

$$b(v, w) = \int_{\Omega} \rho v w dx \quad (5)$$

and $a(v, w)$ is a strain energy scalar product on Ω as

$$a(v, w) = \int_{\Omega} (\bar{\nabla}v)^T \mathbf{D} \bar{\nabla}w dx \quad (6)$$

The stationary values of $R(v)$ are the eigenvalues λ_i and the corresponding functions are the eigenmodes u_i .

When the finite element method is applied, the stationary values of $R(v)$ are determined over a finite dimensional space $\mathcal{V}(\mathcal{T}_h)_p^h = \{v^h \in \mathcal{H}^1(\mathcal{T}_h) : v^h|_{\tau} \in \mathcal{P}_p \ \forall \tau \in \mathcal{T}_h\}$ where \mathcal{T}_h is a spatial finite element mesh space associated with Ω and \mathcal{P}^p is the set of p -th order polynomials.

A function v^h is expressed in the form $v^h = \sum_{j=1}^n N_j(x, y) v_j \equiv \mathbf{N}^T \mathbf{V}$, where $N_j(x, y)$ are local basis functions and v_j are the nodal values of v^h .

On substituting this expression for the eigenmodes u_i in the Rayleigh quotient we obtain

$$\lambda_i^h = R(u_i^h) = \frac{U_i^T \mathbf{K} U_i}{U_i^T \mathbf{M} U_i} \quad (7)$$

where \mathbf{K} is the stiffness matrix, \mathbf{M} is the consistent mass matrix of the structure and U_i is the eigenvector containing the nodal values of u_i^h .

The stationary values of $R(v)$ are found by solving the generalized eigenvalue problem

$$(\mathbf{K} - \lambda^h \mathbf{M})U = 0 \quad (8)$$

The eigenfrequencies are the square roots of the eigenvalues, $\omega_i^h = \sqrt{\lambda_i^h}$.

The eigenfrequency ω_i^h and the associated eigenmode u_i^h for each mode i of interest can be done by any convenient solution procedure of the generalized eigenvalue problem.

3. POST-PROCESSED EIGENFREQUENCIES BY NODE PATCH SPRD

3.1. General

The post-processed eigenfrequencies are calculated using the Rayleigh quotient without solving the generalized eigenvalue problem. A new improved eigenfrequency will be of the form

$$(\omega_i^{**})^2 = \frac{\sum_e \int_{\Omega_e} (\tilde{\nabla} u_i^{**})^T D \tilde{\nabla} u_i^{**} dx}{\sum_e \int_{\Omega_e} (u_i^{**})^T Q u_i^{**} dx} \quad (9)$$

where e is summed over the total number of elements, and u_i^{**} is a displacement field with a higher order of accuracy. The recovered displacement u_i^{**} field of the eigenmode number i will be determined by the SPRD technique, described below. It is important to keep in mind that since the integrands of the integrals in equation (9) includes, as we shall see below, higher order terms than the finite element solution, then the numerical integration must be properly adjusted to be sufficiently accurate.

The SPR technique is based on the fact that for finite element solutions there exist certain points in each element at which the prime variables (displacements) or the derivatives (stresses) have superior accuracy to that found globally. These points are called the superconvergent points of the finite element solution.

From the expression of the Rayleigh quotient it is clear that in order to improve an eigenfrequency it is sufficient to determine a higher accuracy displacement field for the corresponding eigenmode. The exceptional points where the prime variables have higher order accuracy as compared to the global accuracy are known to be the nodal points of the finite element approximation. We anticipate that SPRD based on these points gives a recovered displacement field, which is more accurate and which has a higher order of convergence.

The proposed recovery method is local, i.e. no global system of equations has to be constructed and solved. Let the element for which SPRD currently is performed be called the master element. The index i denoting the eigenpair number is henceforth omitted in order to increase the readability of the equations. The improved displacement field is computed in a three major steps. Step one is to define local patches for the vertex nodes of the master element and for each patch construct a displacement field, u_k^* , by a least squares fit of a local polynomial

$$u_k^*(x) = P(x)a \quad (10)$$

to displacement values at higher accuracy points in the patch. $P(x) = [1 \ x_1 \ x_2 \ x_1 x_2 \ \dots]$ is the basis of the polynomial, and $x = (x_1, x_2)$ is the spatial variable, and k denotes the vertex node number, and a is unknown parameters to be determined.

In order to determine the unknown parameters a we minimize

$$\pi = \sum_{i=1}^{NP} (u_s^i(x_1^i, x_2^i) - P(x_1^i, x_2^i)a)^2 \quad (11)$$

where NP is the number of superconvergent points in the patch and u_s^h denotes the finite element solution in these superconvergent points. This implies

$$Aa = b \tag{12}$$

where

$$A = \sum_{i=1}^{NP} P^T(x_1^i, x_2^i) P(x_1^i, x_2^i) , \quad b = \sum_{i=1}^{NP} P^T(x_1^i, x_2^i) u_s^h(x_1^i, x_2^i) \tag{13}$$

The second step is to weight and sum the node patches

$$\bar{u} = \sum_{k=1}^{NEVN} N_k u_k^* \tag{14}$$

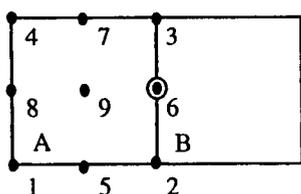
In expression (14) $NEVN$ denotes the number of vertex nodes on the master element and the weight N_k is the local, linear or bilinear, basis function on node number k .

In the third step we change the basis to a finite element basis, where the polynomials have at least one degree higher terms than the complete polynomial appearing in the original finite element basis. The change of basis is simple

$$u^{**} = N^{p+1} \bar{U} \tag{15}$$

$$\bar{U} = [\bar{u}_j]_{j=1}^{NEVN} , \quad \bar{u}_j = \begin{cases} u(x_j, y_j) & \text{if } j \text{ old node} \\ \bar{u}(x_j, y_j) & \text{if } j \text{ new node} \\ (\bar{u}_j^A + \bar{u}_j^B)/2 & \text{if } (x_j, y_j) \in \Omega_A \cup \Omega_B \end{cases} \tag{16}$$

where $NEVN$ denotes the the new number of nodes on the element and A and B are two adjacent elements. In words: corner node values of the original finite element solution are kept and additional nodal values are calculated by (14). It will be clear, from the patch selecting strategy presented below, that the displacement fields for adjacent elements not are identical over inter-element boundaries. In case the new node lies on an inter-element side the new value is calculated by averaging the two displacements for the node. See Figure 1 for an example.



⊙ New node on interelement edge

$$\bar{u}_6 = 1/2(\bar{u}_6^A + \bar{u}_6^B)$$

$$u^{**} = N^{p+1} \bar{U}$$

Figure 1. Recovered displacement field for element A, expressed by nodal values and finite element basis functions of degree $p+1$. The new node, 6, belongs to both element A and B and the final value is the average of the values obtained for each of element A and B by equation (14).

3.2. Patch selecting strategy

The strategy is based on the following arguments: local information have a greater importance than distant information for a curve fit (c.f. a Taylor expansion); a large number of sampling points in a least squares fit decreases accuracy; a small number of sampling points may cause distortions.

Strategy for quadrilateral elements.. We select one patch for each element and vertex node on that element, called master element and assembly point. Thus, a patch will be identified by a pair of numbers—an element number and a vertex node number. The patch includes the master element and those elements which share a side with the master element and which also have the assembly point as a vertex node, see Figure 2. In most cases this will be three elements, except for boundary vertex nodes where the patch then will have less than three elements, for these cases some more element/s have to be included for the patch to have optimum number of sampling points. The rule is that elements to be included should lie on as a small distance as possible from the assembly point and that the patch ideally has three elements.

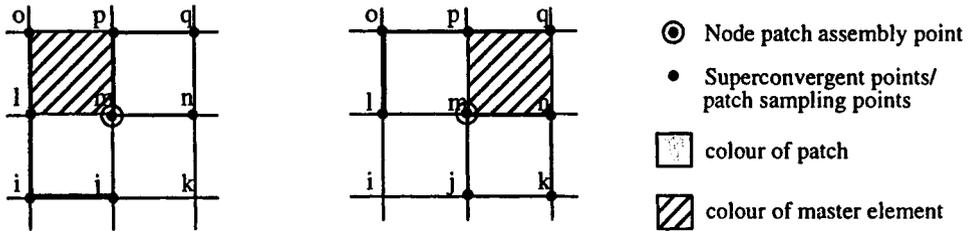


Figure 2. Selecting node patches on a quadrilateral mesh.

Strategy for triangular elements. Also for triangular elements we let a pair of element number and element vertex node number define each patch as for quadrilateral elements. The patch includes all elements with the assembly point as a vertex node, see Figure 3. When it is the case that the patch has too few sampling points we add some more elements to the patch, so that the number of sampling points will be equal to or one or two more than the number of terms in the polynomial P .

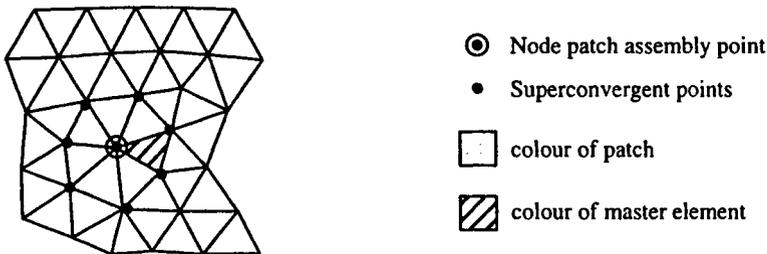


Figure 3. Typical scenario when selecting a node patch on a triangular mesh subject to SPRD.

3.3. Remarks

If there are more nodes than necessary in the patch one must ensure that the polynomial equals the value of the finite element solution in the node patch assembly point, which can be accomplished by weighting in the least squares fit.

4. ERROR ESTIMATION

The absolute and relative error in eigenfrequencies of the original finite element solution is expressed as

$$\Delta\omega_i^h = \omega_i^h - \omega_i \quad \eta_i^\omega = \frac{\Delta\omega_i^h}{\omega_i} \cdot 100 \quad (17)$$

of the postprocessed solution as

$$\Delta\omega_i^{**} = \omega_i^{**} - \omega_i \quad \eta_i^{**} = \frac{\Delta\omega_i^{**}}{\omega_i} \cdot 100 \quad (18)$$

and the estimated error of the finite element solution

$$\Delta\bar{\omega}_i^h = \omega_i^h - \omega_i^{**} \quad \bar{\eta}^\omega = \frac{\Delta\bar{\omega}_i^h}{\omega_i^{**}} \cdot 100 \quad (19)$$

The quality of any error estimator is measured by an effectivity index which gives the ratio of the estimated errors to the actual ones as follows

$$\theta_i = \frac{\Delta\bar{\omega}_i^h}{\Delta\omega_i^h} \quad (20)$$

The error estimator is said to be asymptotically exact if the estimated errors approach the exact errors as the finite element mesh is refined. This is equivalent to the effectivity index converging to unity as the finite element mesh is refined.

5. ADAPTIVITY

Let (ω^h, U) denote a finite element solution eigenpair and let (ω^{**}, U^{**}) denote an improved eigenpair (improved by any technique). Then we want the estimated relative error to satisfy

$$\bar{\eta}^\omega = \frac{\omega^h - \omega^{**}}{\omega^{**}} \leq \bar{\eta}_{TOL}^\omega \quad (21)$$

Manipulation of the left hand side of equation (21) reveals that

$$\frac{\omega^h - \omega^{**}}{\omega^{**}} = \frac{\omega^h - \omega^{**}}{\omega^{**}} \times \frac{\omega^h + \omega^{**}}{\omega^h + \omega^{**}} = \frac{\lambda^h - \lambda^{**}}{\omega^h \omega^{**} + \lambda^{**}} \quad (22)$$

and

$$\frac{\lambda^h - \lambda^{**}}{\omega^h \omega^{**} + \lambda^{**}} \leq \frac{\lambda^h - \lambda^{**}}{2\lambda^{**}}, \text{ since } \omega^{**} \leq \omega^h \quad (23)$$

Clearly, (21) will hold if the following holds

$$\frac{\lambda^h - \lambda^{**}}{2\lambda^{**}} \leq \bar{\eta}_{TOL}^\omega \Rightarrow \lambda^h - \lambda^{**} \leq 2\lambda^{**}\bar{\eta}_{TOL}^\omega \quad (24)$$

Restrict the refinement technique so that we by using the Rayleigh quotient, equation (7), can write

$$\lambda^h - \lambda^{**} = \frac{\sum_e U^T K^e U}{\sum_e U^T M^e U} - \frac{\sum_e U^{**T} K^{**e} U^{**}}{\sum_e U^{**T} M^{**e} U^{**}} \quad (25)$$

The matrices $K^{**} = \sum_e K^{**e}$ and $M^{**} = \sum_e M^{**e}$ are the stiffness and mass matrices of some finite dimensional space which allows improved solutions. The summation is over all elements, NEL , in the mesh. Equation (25) simplifies if the eigenvectors are normalized such that

$$\sum_e U^T M^e U = 1 \text{ and } \sum_e U^{**T} M^{**e} U^{**} = 1 \text{ to}$$

$$\lambda^h - \lambda^{**} = \sum_e (U^T K^e U - U^{**T} K^{**e} U^{**}) \equiv \sum_e \Delta^e \quad (26)$$

where Δ^e is the error indicator.

We wish to predict a new mesh size so that equation (24) holds, which we write

$$\lambda^{h_{new}} - \lambda^{**_{new}} \leq 2\lambda^{**_{new}}\bar{\eta}_{TOL}^\omega \quad (27)$$

Then (25) and (27) yields

$$\sum_e \Delta_{new}^e \leq 2\lambda^{**_{new}}\bar{\eta}_{TOL}^\omega \quad (28)$$

The optimality criterion, [14], on a mesh is that the error is equally distributed, thus we require

$$\Delta_{new}^e = \Delta_{new}, \quad \forall e \quad (29)$$

Then equation (28) can be written

$$NEL_{new}\Delta_{new} \leq 2\lambda^{**_{new}}\bar{\eta}_{TOL}^\omega \quad (30)$$

Since the eigenvectors minimizes the Rayleigh quotient, [15], we know that $\lambda \leq \lambda^h$, because λ^h is the minimum over a possible smaller set than λ . Therefore we also know for λ^{**} to be an improved estimate of λ , it must hold that $\lambda \leq \lambda^{**} \leq \lambda^h$. The eigenvalues will be greater than zero if the matrices are positive definite, which is the case that we consider, we have $\lambda^h - \lambda^{**} \leq \lambda^h - \lambda$. Then basic a priori error estimates,[15], yields

$$\lambda^h - \lambda^{**} \leq \lambda^h - \lambda \leq 2\delta h^{2(p+1-m)}\lambda^{(p+1)/m} \quad (31)$$

where δ is constant with respect to the characteristic mesh size h .

For a sequence of optimal meshes we assume that equation (31) can be stated in terms of the error indicator and the local mesh size.

$$\Delta^e \approx Ch^{2(p+1-m)}\lambda^{(p+1)/m} \quad (32)$$

or with the notion of a new mesh and an old mesh

$$\Delta_{new} \approx C_{new}h_{new}^{2(p+1-m)}\lambda^{(p+1)/m} \quad (33)$$

$$\Delta^e \approx Ch_{old}^{2(p+1-m)}\lambda^{(p+1)/m}$$

Equation (33) gives

$$\frac{h_{new}^{2(p+1-m)}}{h_{old}^{2(p+1-m)}} \approx C' \frac{\Delta_{new}}{\Delta^e} \quad (34)$$

Equation (34) and equation (30) yields that we require

$$h_{new}^{2(p+1-m)} \leq C' \frac{2\lambda^{**}{}_{new}\bar{\eta}_{TOL}^\omega}{NEL_{new}\Delta^e} h_{old}^{2(p+1-m)} \quad (35)$$

We do not know $\lambda^{**}{}_{new}$ and NEL_{new} so we replace them with the known quantities of the old mesh, λ^{**} and NEL and make the assumption that $NEL_{new} \approx NEL(\frac{h_{new}}{h_{old}})^D$ for a D -dimensional problem. We assume that $C' = 1$. Hence, the mesh refinement is performed according to

$$h_{new}^{2(p+1-m)+D} \leq \frac{2\lambda^{**}}{NEL\Delta^e} \bar{\eta}_{TOL}^\omega h_{old}^{2(p+1-m)+D} \quad (36)$$

If the old mesh is optimal then $\frac{2\lambda^{**}}{NEL\Delta^e} = \frac{2\lambda^{**}}{\lambda^h - \lambda^{**}}$.

6. NUMERICAL EXAMPLES

6.1. Rectangular plate, regular mesh

In order to study convergence rate and accuracy of the proposed method, we considered a rectangular plate subject to plane stress for which the exact solution is known. The plate has uniform thickness. See Figure 4 for boundary conditions and dimensions of the plate. The vibrations are in-plane. We used nondimensionalized eigenfrequencies $\omega \sqrt{\rho/E}$.

In order to study convergence rate and accuracy of the proposed technique, we used a sequence of five meshes with 4x4, 8x8, 10x10, 15x15 and 30x30 equally sized rectangular elements. Figure 5 shows the effectivity indices of the error estimator for the 1-st and 8-th eigenfrequency. Figure 6 shows the convergence rate for the exact error using 4-node elements and 8-node rectangular elements, it also shows the convergence rate of the SPRD improved 4-node solution and the estimated rate of convergence for the 4-node element.

We make the following observations from the regular computation. The recovered solution demonstrates superconvergence properties if regular elements are used. Effectivity index shows that the error estimate is asymptotic exact.

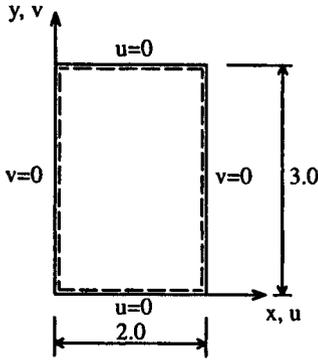


Figure 4. Dimensions and boundary conditions of rectangular plate.

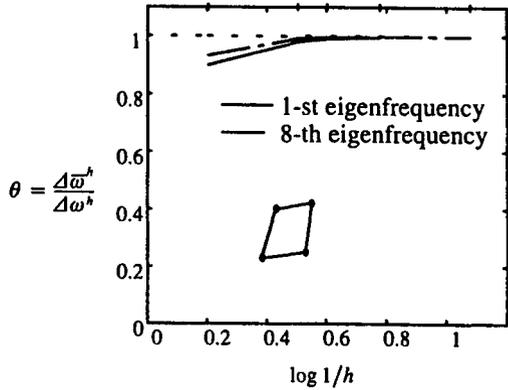


Figure 5. Effectivity indices for rectangular plate using 4-node quadrilateral elements.

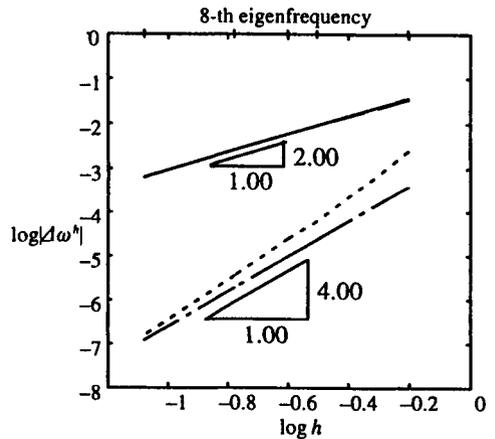
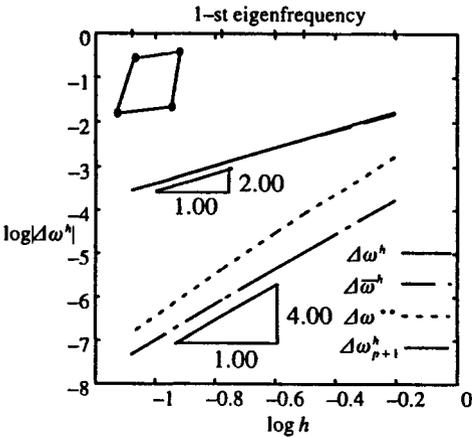


Figure 6. Convergence rates for rectangular plate using 4-node quadrilateral elements.

6.2. Rectangular plate, adaptive computation

In order to study the behavior of the error indicator equation (36), the eigenfrequencies of the rectangular plate was computed adaptively. We used 4-node quadrilateral elements. The tolerance of the estimated relative error, c.f. equation (21), was set to $\bar{\eta}_{TOL}^h = 0.05\%$. We made two adaptive sequences: In the first the mesh was refined according to the error indicator of the 1-st eigenfrequency; In the second the mesh was refined according to the error indicator of the 8-th eigenfrequency. The maximum number of adaptive iterations was limited to three. The mesh generation was performed using advancing front technique, Peraire *et al.* [16].

In both the adaptive sequences the error tolerance was met for the eigenfrequency considered. It can also be seen that if the demanded accuracy of a higher frequency is met, then also (often) the accuracy of the lower frequencies meet the same demands.

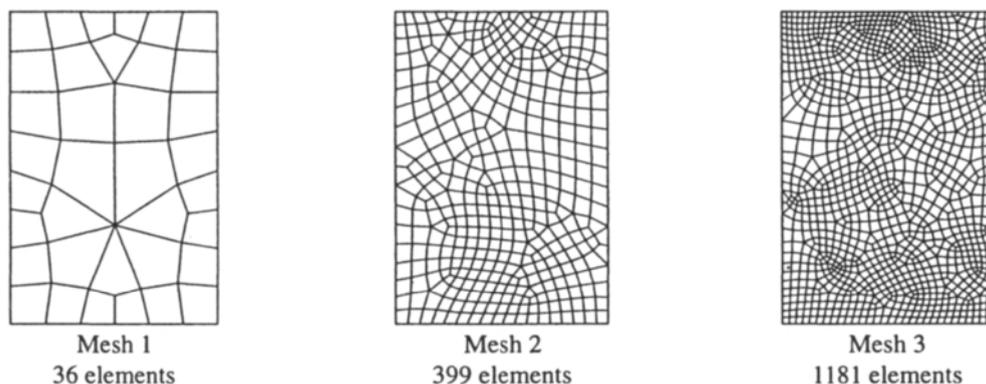


Figure 7. Sequence of meshes for adaptive computation of the 1-st eigenfrequency.

Table 1.

Results obtained for adaptive computation of the 1-st eigenfrequency

Mesh	ω_1	ω_1^h	ω_1^{**}	$\bar{\eta}_1^\omega$	ω_8	ω_8^h	ω_8^{**}	$\bar{\eta}_8^\omega$
1	0.6494	0.6575	0.6514	0.940	1.9790	2.1576	1.9860	7.740
2	0.6494	0.6502	0.6495	0.104	1.9790	1.9824	1.9792	0.164
3	0.6494	0.6497	0.6495	0.040	1.9790	1.9804	1.9791	0.065

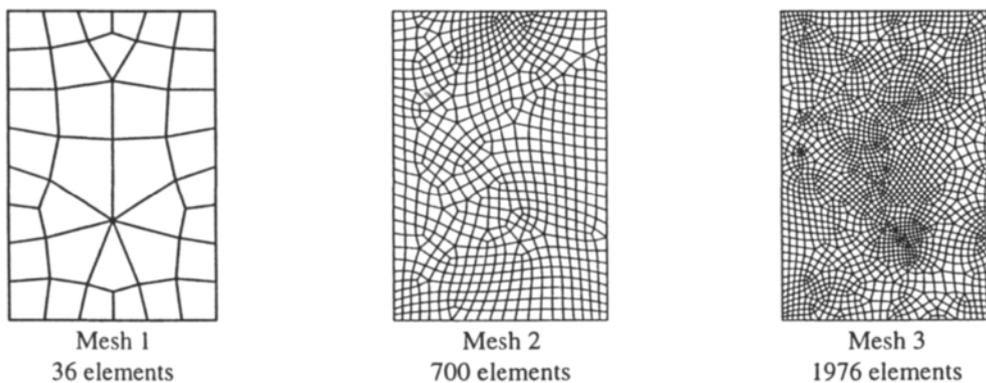


Figure 8. Sequence of meshes for adaptive computation of the 8-th eigenfrequency.

Table 2.

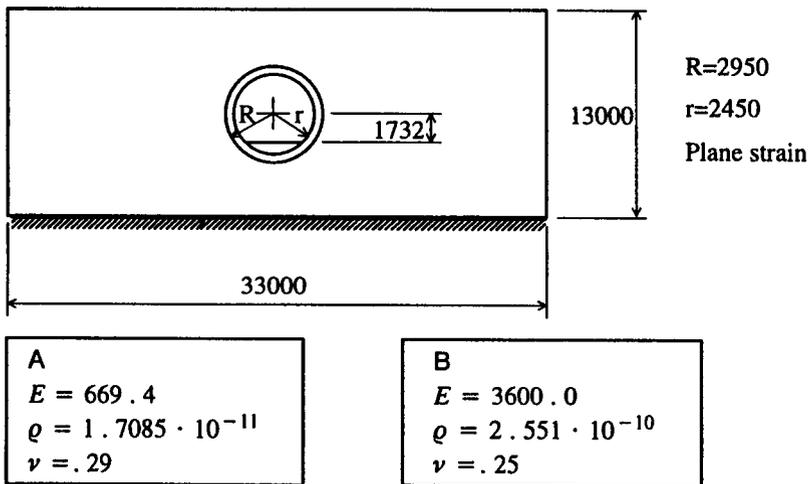
Results obtained for adaptive computation of the 8-th eigenfrequency

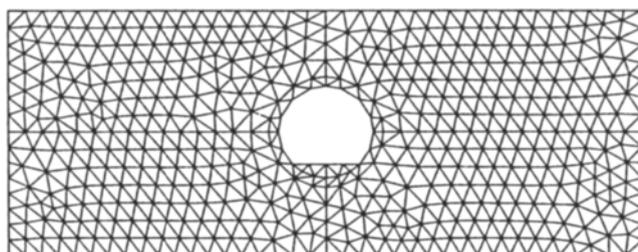
Mesh	ω_1	ω_1^h	ω_1^{**}	$\bar{\eta}_1^\omega$	ω_8	ω_8^h	ω_8^{**}	$\bar{\eta}_8^\omega$
1	0.6494	0.6575	0.6514	0.940	1.9790	2.1576	2.0025	7.740
2	0.6494	0.6498	0.6495	0.053	1.9790	1.9806	1.9791	0.080
3	0.6494	0.6496	0.6495	0.018	1.9790	1.9798	1.9791	0.038

6.3. Tunnel

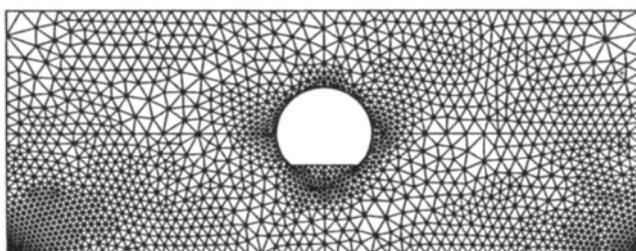
In order to study the applicability of the method to more complicated engineering problems the eigenfrequencies of a vibrating tunnel, Figure 9, was computed adaptively for the 1-st and the 8-th eigenfrequency. The error tolerance was $\bar{\eta}_{TOL}^h = 0.3\%$. We used linear triangular elements.

In both cases convergence was reached within one adaptive refinement of the same starting mesh, Figure 10 Mesh 1. Adaptive refinement according to the 1-st eigenfrequency resulted in Mesh 2a in Figure 10. Adaptive refinement according to the 8-th eigenfrequency resulted in Mesh 2b in Figure 10. The results are listed in table 3.

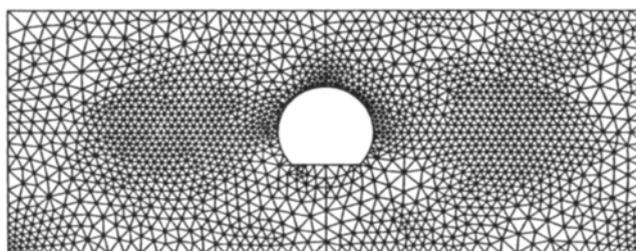
**Figure 9.** Geometry, material properties and boundary conditions of tunnel structure.



Mesh 1
974 elements



Mesh 2a
3302 elements



Mesh 2b
3327 elements

Figure 10. Sequence of meshes for adaptive computation.

Table 3.

Results obtained for adaptive computation of the 1-st (2a) and 8-th (2b) eigenfrequency

Mesh	ω_1^h	ω_1^{**}	$\bar{\eta}_1^\omega$	ω_8^h	ω_8^{**}	$\bar{\eta}_8^\omega$
1	117.01	116.09	0.788	376.81	374.05	0.736
2a	115.27	115.10	0.150	373.10	372.31	0.213
2b	115.55	115.41	0.122	373.06	372.56	0.133

7. CONCLUDING REMARKS

The local updating technique provides an effective and reliable mean to improve eigenfrequencies without any user tuned coefficients. The improved eigenfrequencies exhibit superconvergence so that the order of convergence of the improved values are at least two orders higher compared to the finite element solution. An h -adaptive scheme converges for a group of eigenvalues to a demanded accuracy within one or two steps.

REFERENCES

- [1] N.-E. Wiberg, P. Hager and R. Bausys, Accuracy assessed eigenfrequency analysis by node patch recovery, in: *Proc. of NAFEMS WORLD CONGRESS '97*, Vol1, (Stuttgart, Germany, 9–11 April 1997) 13–24.
- [2] O. C. Zienkiewicz and J. Z. Zhu, The superconvergent patch recovery and a posteriori error estimators, Part 1 The recovery technique. *Int. J. numer. Meth. Engrg*, **33** (1992) 1331–1364.
- [3] O. C. Zienkiewicz and J. Z. Zhu, The superconvergent patch recovery and a posteriori error estimators, Part 2 Error estimates and adaptivity. *Int. J. numer. Meth. Engrg*, **33** (1992) 1365–1382.
- [4] N.-E. Wiberg and F. Abdulwahab, Patch recovery based on superconvergent derivatives and equilibrium, *Int. J. numer. Meth. Engrg*, **36** (1993) 2703–2724.
- [5] N.-E. Wiberg, F. Abdulwahab and S. Ziukas, Enhanced superconvergent patch recovery incorporating equilibrium and boundary conditions, *Int. J. numer. Meth. Engrg*, **37** (1994) 3417–3440.
- [6] N.-E. Wiberg and X. D. LI, A postprocessed error estimate and an adaptive procedure for the semidiscrete finite element method in dynamic analysis, *Int. J. numer. Meth. Engrg*, **37** (1994) 3585–3603.
- [7] N.-E. Wiberg, F. Abdulwahab and S. Ziukas, Improved element stresses for node and element patches using superconvergent patch recovery, *Comput. Meth. appl. Mech. Engrg*, **11** (1995) 619–627.
- [8] P. Ladeveze and J. P. Pelle, Accuracy in finite element computation for eigenfrequencies, *Int. J. numer. Meth. Engrg*, **28** (1989) 1929–1949.
- [9] J. Avrashi and R. D. Cook, New error estimation for eigenproblems in finite element analysis, *Engrg. Comput.* **10** (1993) 243–256.
- [10] N.-E. Wiberg and R. Bausys, Error estimation for eigenfrequencies and eigenmodes in dynamics, in: J. Pal and H. Werner, ed., *Computing in Civil and Building Engineering*, (Balke-ma, 1995).

- [11] N.-E. Wiberg, R. Bausys and P. Hager, Improved eigenfrequencies and eigenmodes in free vibration analysis, in: B. H. V. Topping, ed., *Advances in finite element technology*, (Civil-Comp Press, 1996).
- [12] N.-E. Wiberg, R. Bausys and P. Hager, Adaptive h-version eigenfrequency analysis, Publ. 96:11, Dept of Structural Mechanics, Chalmers University of Technology, (1996).
- [13] D. B. Stephen and G. P. Steven, Error estimation for finite element eigenproblems, in: *Proc. of NAFEMS WORLD CONGRESS '97*, Vol2, (Stuttgart, Germany, 9–11 April 1997) 707–717.
- [14] I. Babuska, Feedback, Adaptivity, and a posteriori estimates in finite elements: aims, theory, and experience, in: Babuska *et al*, ed., *Accuracy estimates and adaptive refinements in finite element computations*, (John Wiley and sons, 1986) 3–23.
- [15] G. Strang and G. J. Fix, *An analysis of the finite element method* (Prentice-Hall, 1973).
- [16] J. Peraire, M. Vahdati, K. Morgan and O. C. Zienkiewicz, *Adaptive remeshing for compressible flow computations*, *J. of Computational Physics*, **72**, (1987) 449–466.

This Page Intentionally Left Blank

Error estimation and adaptivity for the finite element method in acoustics

Ph. Bouillard^a and F. Ihlenburg^{b,*}

^aDepartment of Continuum Mechanics, Université Libre de Bruxelles, CP 194/5,
Av. F. D. Roosevelt 50, 1050 Brussels, Belgium, e-mail: pbouilla@smc.ulb.ac.be

^bUniversity Rostock, FB Bauingenieurwesen, PF1410, D-23952 Wismar, Germany

1. INTRODUCTION

The acoustic wave propagation in an inviscid medium is described by the Helmholtz equation $\Delta p + k^2 p = 0$ where p is the spatial distribution of a small harmonic perturbation of pressure $p'(\mathbf{x}, t) = p(\mathbf{x}) e^{j\omega t}$ around a steady state and $k = \omega c^{-1}$ is a given physical parameter. Classically, a distinction is made between interior and exterior problems, coupled or uncoupled to a structure.

Let $\Omega \in \mathcal{R}^3$ be the domain of a closed thin-walled shell. If Ω is assumed as rigid, only the incidental and diffracted waves are propagating. In most applications however, Ω is elastic and the dynamic vibrations then add a radiated sound field to the fields above. Acoustic fluid-solid interaction is modelled by a coupled system of equations for dynamo-acoustic solid-fluid interaction. If the sound propagates into an unbounded exterior medium, the assumption that no waves are reflected from the far field is formalised by the Sommerfeld boundary condition $|\partial_{\Gamma} p - jk p| = o(R^{-1})$ where R is the radius of a fictitious sphere in the far field. For exterior solid-fluid interaction, this condition introduces the effect of radiation damping into the boundary value problem, that is, the vibrations of the elastic structure are damped if the acoustic medium fills an unbounded domain. If the acoustic medium (fluid or gas) fills the interior of a cavity, the damping effect is not significant. If no other damping conditions such as structural damping are assumed, the interior problem becomes singular at the eigenfrequencies of the cavity. Detection and prediction of the eigenmodes is thus an important issue of computational simulation.

As in other application fields, error control is an important issue in acoustic computations. It is now clear that the numerical parameters (mesh size h and degree of approximation p) must be adapted to the physical parameter k , or better to the non-dimensional wave number κ scaled by a characteristic length ($\kappa = kL$). The well known rule for the h version with $p=1$ is to resolve a wavelength $\lambda = 2\pi\kappa^{-1}$ by six elements (that is $kh \approx 1$). This 'rule of the thumb' was shown for *interpolation* of a given sinusoidal wave, i.e. it characterises the *approximability* of a given finite element mesh for the Helmholtz equation. If the numerical discretisation is stable, the quality of the numerical solution is entirely controlled by the approximability of the finite element mesh. The situation is quite different in the presence of singularities. In that case, *stability* (or the lack thereof) is equally (sometimes more) important.

Adaptive approximation of singularities has become a well researched field with many implementations in engineering practice, especially in the traditional field of structural analysis

* Present address: Germanischer Lloyd, Vorstzen 32, D-20459 Hamburg, ihl@hamburg.germanlloyd.de

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 the numerical approximation.

The situation is different in wave propagation with high wave numbers where the solutions are 'rough', i.e. highly oscillatory. For instance, the function $p=\sin(kx)$ is a solution of the Helmholtz equation in one dimension. The higher wave number k , the more oscillatory the solution becomes. Its gradient $p'=k\cos(kx)$ is large everywhere, i.e. *globally*. This is a singularity inherent to the differential operator rather than to the domain or to the boundary conditions and is called *k-singularity* [1-3].

Similarly, the discrete operator (stiffness matrix) becomes singular at eigenvalues of the discretised interior problem (or nearly singular at damped eigenvalues in solid-fluid interaction) defining the *λ-singularities*.

Both singularities are of global character and without adaptive correction, their destabilizing effect generally leads to large error of the finite element results, even if the finite element mesh has sufficient approximability. For example, it is well known that the 'rule of thumb' does not guarantee small error if wave number k grows. This is due to the *k-singularity* which causes an increasing phase shift of the discrete solution (even with h respecting the 'rule of the thumb' $kh = \text{constant}$). In the presence of *λ-singularity*, a large error can be observed also for small wave numbers. Due to these specific phenomena of dynamo-acoustic computations, *error control cannot, in general, be accomplished by just 'transplanting' methods that worked well in static computations*. As a first step towards reliable computations, the nature of the error has to be understood.

The paper is organised as follows: Section 2 establishes the finite element formulation for the acoustic problem and defines the discretisation error. A short review of *a priori* error estimations for the Helmholtz equation and rules for *k*-adaptive mesh design is given in the Section 3. First investigations on the effect of *k- and λ-singularities* for elastic scattering and for the interior problem are also discussed. Section 4 is dedicated to the reliability of two *a posteriori* error estimators. Finally, Section 5 illustrates on a real-life problem the specific singularities of the Helmholtz operator and shows that, with low wave numbers, *h*-adaptive refinements are necessary in order to control the accuracy.

2. THE FINITE ELEMENT METHOD FOR ACOUSTICS

2.1. Strong formulation

The acoustic wave propagation, under the assumption that the wave is a small harmonic perturbation of pressure around a steady uniform state in an inviscid fluid Ω , is addressed by the Helmholtz equation

$$\Delta p + k^2 p = 0 \quad (1)$$

and the appropriate boundary conditions of three types :

$$\text{Dirichlet on } \Gamma_D \quad p = \bar{p} \quad (2)$$

$$\text{Neumann on } \Gamma_N \quad v_n = \bar{v}_n \quad \text{or} \quad \nabla_n p = -j\rho ck \bar{v}_n \quad (3)$$

$$\text{Robin on } \Gamma_R \quad \mathbf{v}_n = A_n p \quad \text{or} \quad \nabla_n p = -j\rho c k A_n p \quad (4)$$

where ρ is the mass density. The particle velocity \mathbf{v} is linked to the gradient of the pressure by the harmonic equation of motion

$$\mathbf{v} = \frac{-1}{j\rho c k} \nabla p \quad (5)$$

2.2. Finite element discretisation

The formulation and the model problems will be defined using the following non-dimensional variables

$$\xi = \frac{x}{L} \quad \kappa = kL \quad \mathbf{h} = \frac{h}{L} \quad (6)$$

The acoustic problem (Eqs. 1-4) can be cast in a weak form and discretised by a standard Galerkin procedure resulting in a linear algebraic system of equations

$$(\mathbf{K} + jk \mathbf{C} - k^2 \mathbf{M}) \mathbf{q}_p = -jk \mathbf{f} \quad (7)$$

where \mathbf{K} , \mathbf{C} and \mathbf{M} are the classical stiffness, damping and mass matrices analogous to those arising in structural dynamics.

$$\mathbf{K} = \int_{\Omega} (\nabla \mathbf{N})^t (\nabla \mathbf{N}) d\Omega \quad \mathbf{C} = \frac{1}{\rho c} \int_{\Gamma_R} \mathbf{N}^t \mathbf{N} A_n d\Gamma \quad \mathbf{M} = \int_{\Omega} \mathbf{N}^t \mathbf{N} d\Omega \quad (8)$$

with \mathbf{N} denoting usual shape functions of order p . The \mathbf{q}_p vector contains the unknown nodal admissible values of pressure. The right-hand side of Eq. (7) contains normal velocity boundary conditions (Eq. 3)

$$\mathbf{f} = \frac{1}{\rho c} \int_{\Gamma_N} \mathbf{N}^t \bar{v}_n d\Gamma \quad (9)$$

2.3. Discretisation error

The discretisation error considered throughout this paper is defined in the H^1 -seminorm for complex variables

$$|p - p^h|_1^2 = \int_{\Omega} (\bar{v} - \bar{v}^h)^t (\mathbf{v} - \mathbf{v}^h) d\Omega \quad (10)$$

where $\bar{\cdot}$ denotes the complex conjugate. In the following, this error will be equivalently denoted as the error in L^2 norm on the velocity

$$|p - p^h|_1 = \|\mathbf{v} - \mathbf{v}^h\|_0 \quad (11)$$

2.4. Model problem I

Throughout the paper, two one dimensional model problems will be used in order to illustrate the specific singularities of the Helmholtz operator. Model problem I is defined by

$$\frac{d^2 p}{d\xi^2} + \kappa^2 p = 0 \quad \text{in } (0, 1) \quad (12)$$

$$p(0) = 1 \quad \frac{dp}{d\xi}(1) - j\kappa p(1) = 0 \quad (13)$$

and has the exact solution, corresponding in fact to an exterior problem with a propagating one dimensional wave

$$\begin{aligned} p(\xi) &= e^{j\kappa\xi} \\ &= \cos(\kappa\xi) + j \sin(\kappa\xi) \end{aligned} \quad (14)$$

Indeed, consider the exterior scattering problem in three dimensions. The physical process is formulated as an exterior Helmholtz boundary value problem

$$\Delta p + k^2 p = 0 \quad \text{in } \Omega^+ \quad (15)$$

$$\nabla_n p = w \quad \text{on } \Gamma, \quad \lim_{R \rightarrow \infty} R \left| \frac{\partial p}{\partial r} - jkp \right|_{r=R} = 0 \quad (16)$$

For application of the finite element method, the exterior domain is decomposed by introducing an artificial boundary Γ_a enclosing obstacle Ω . A usual finite element partition is then introduced in the near field between Γ and Γ_a . In the remaining exterior domain outside Γ_a , the boundary value problem is discretised by other means (e.g. boundary elements, infinite elements, absorbing or non reflecting boundary conditions). This exterior discretisation is mapped into the near field by a so-called Dirichlet-to-Neumann map. The most simple procedure is to impose the Sommerfeld condition on a spherical boundary Γ_a rather than in the far field. Model problem I corresponds to this approach in one dimension and was first studied in [1-2]. Unlike applications in two and three dimensions, no spurious reflections are caused by the far field approximation at Γ_a , hence the numerical error is only due to the finite element discretisation.

2.5. Model problem II

Model problem II is defined by

$$\frac{d^2 p}{d\xi^2} + \kappa^2 p = 0 \quad \text{in } (0, 1) \quad (17)$$

$$\frac{dp}{d\xi}(0) = j\rho c v_0 \quad \frac{dp}{d\xi}(1) = 0 \quad (18)$$

and has the exact solution

$$p = -j\rho c v_0 \frac{\cos[\kappa(1-\xi)]}{\sin(\kappa)} \quad (19)$$

Model problem II defines an interior undamped problem and has eigenmodes corresponding to the non-dimensional wave numbers

$$\kappa = m\pi \quad m \in \mathcal{N} \quad (20)$$

Note that the eigenfrequencies depend on the length of the tube by definition of the non-dimensional wave number κ (Eq. 6).

3. A PRIORI ESTIMATE

3.1. Rule of the thumb

Most acoustic finite element analyses are computed by keeping the product kh constant. This is called 'rule of the thumb' and is only based on the simple assumption that a wavelength should always be resolved by a constant number of elements (Fig. 1). According to the user's manual of SYSNOISE [4], five or six elements are sufficient for linear elements ($p=1$) which corresponds to $kh \approx 1$.

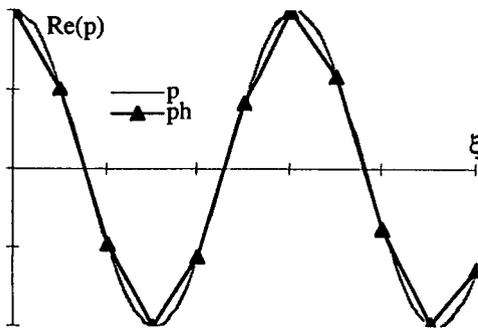


Figure 1. Rule of the thumb: the resolution of a wavelength is usually kept constant ($kh \approx 1$).

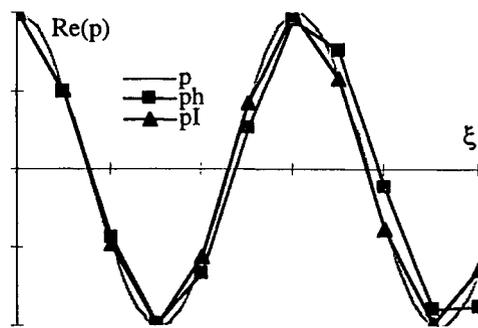


Figure 2. Splitting of the error between the error on the interpolant pI (approximation error) and on the finite element wave ph (pollution error).

3.2. k -singularity

The influence of the k -singularity is well highlighted on model problem I. The theoretical analysis reveals:

1. First, both the exact and discrete Helmholtz operators are increasingly unstable as wave number k grows. This effect is quantified by the (continuous and discrete) inf-sup constants γ, γ^h which are of order k^{-1} [1].
2. Second, the general estimate for the relative error in H^1 -seminorm of the hp version can be rewritten as [2]

$$\frac{|p - p^h|_1}{|p|_1} \leq C_1 \Theta + C_2 \kappa \Theta^2 \tag{21}$$

where $\Theta = (\kappa h_p)^p$ is the scale of the finite element mesh. Estimate (21) splits the error in two terms: the first one corresponds to the approximation error between the exact solution and the interpolant. The second one estimates the numerical pollution due to the k -singularity. This term can be, for large wave numbers, larger than the approximation error (Fig. 2).

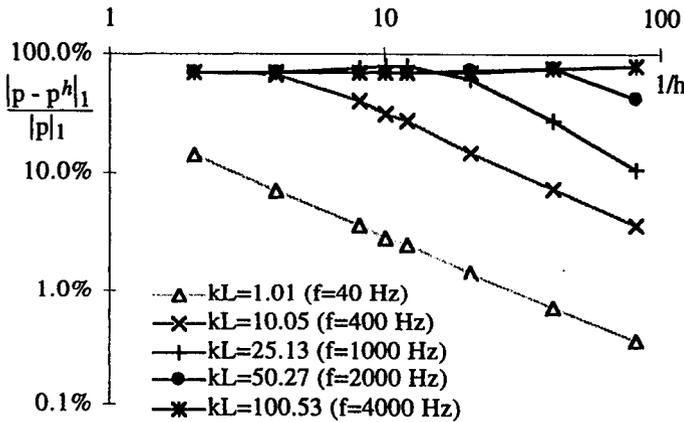


Figure 3. Model problem I: Convergence in H^1 -seminorm for several non-dimensional wave numbers κ .

Fig. 3 gives, for model problem I with a length of 1 m, the convergence in H^1 -seminorm for several non-dimensional wave numbers showing that:

1. for low wave numbers ($\kappa \leq 1$), the asymptotic convergence is immediately reached,
2. for high wave numbers ($\kappa > 1$), there is a preasymptotic range before the asymptotic convergence.

The non-dimensional wave number plays thus an important role. For instance, Fig. 4 shows that the error grows if the length of the tube in model problem I increases. This is remarkable since the rule of the thumb typically does not incorporate this phenomenon.

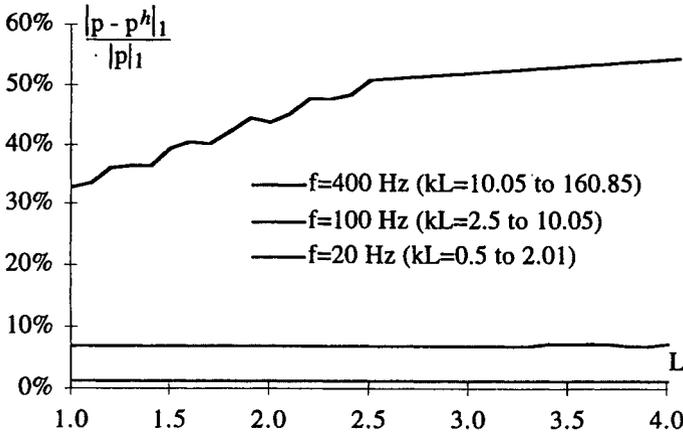


Figure 4. Model problem I: Error in H^1 -seminorm is a growing function of the tube length.

Estimate (22) shows the advantage of the hp method, as compared to the traditional h method, since the magnitude of the scale decreases when p grows. Hence, the pollution of the error by large κ is smaller for $p=2,3,4$ compared to the h version with $p=1$. The resulting *a priori* rule for k -adaptive mesh design is now

$$\kappa \left(\kappa \frac{h}{p} \right)^{2p} \leq C(\bar{\eta}) \quad (22)$$

where $C(\bar{\eta})$ is a constant depending on the prescribed tolerance. The 'rule of thumb' guarantees a small error only asymptotically, when Θ is small enough so that the pollution term can be neglected with respect to the approximation error.

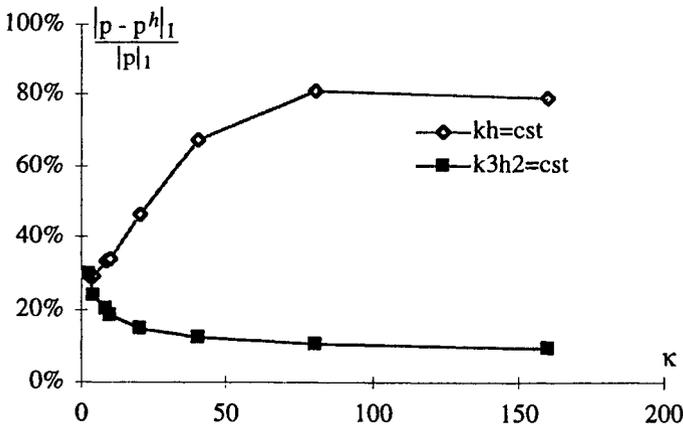


Figure 5. Model problem I: Error in H^1 -seminorm: keeping $\kappa^3 h^2$ constant vs. keeping κh constant.

Fig. 5 gives the error in H^1 -seminorm as a function of the non-dimensional wave number κ with rule (22) vs. rule of the thumb κh constant ($p=1$). It shows that the error is well controlled by keeping $\kappa^3 h^2$ constant while the error grows with the wave number when only keeping κh constant.

3.3. λ -singularity

Consider now the undamped interior model problem II. When computing the acoustic forced response of discretised Helmholtz equation (7), there exist eigenfrequencies for which the system matrix is singular and the corresponding discretisation error in H^1 -seminorm is infinite. Fig. 6 shows the effect of the λ -singularity at the wave numbers corresponding to the finite element eigenvalues λ^h .

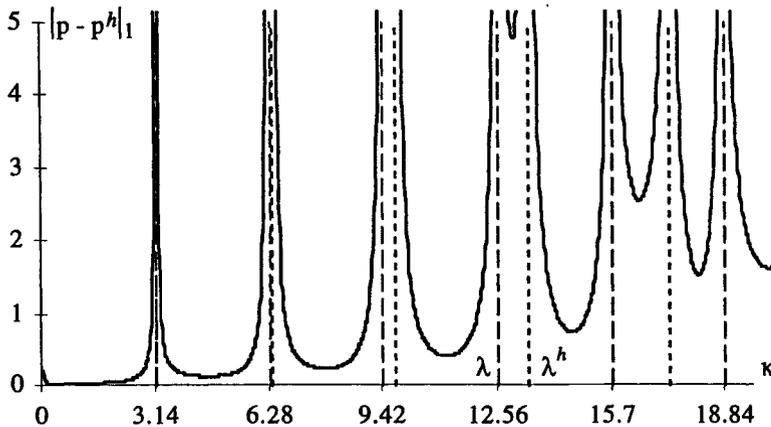


Figure 6. Model problem II: error in H^1 -seminorm as a function of the wave number. The error is infinite for the wave numbers corresponding either to the exact eigenvalues λ or to the finite element eigenvalues λ^h .

4. A POSTERIORI ERROR ESTIMATION

Now that the specific singularities of acoustics have been described and the nature of the error is well understood, it is necessary to develop the tools that can control the discretisation error in acoustics. Two error estimators that have been initially developed for elliptic problems are investigated here showing that they can not control either the influence of the k -singularity nor the λ -singularity.

4.1. Error estimators

Two error estimators, based on completely different principles, have been applied to the acoustic problem (Eqs. 1-5):

1. Error estimation based on the *Superconvergent Patch Recovery* technique due to O. C. Zienkiewicz and J. Z. Zhu [5]. In this case, one computes from the finite element solution \mathbf{v}^h a recovered velocity field \mathbf{v}^* using a local superconvergent smoothening

technique. The error is then simply estimated by substituting the exact velocity field \mathbf{v} by the recovered one \mathbf{v}^*

$$e^{\text{SPR}} = \|\mathbf{v}^* - \mathbf{v}^h\|_0 \quad (23)$$

2. Error in *admissible fields* based on the construction of fields satisfying strongly, element by element, the equations of acoustics (Eqs. 1-4) except the link between the pressure and the velocity (Eq. 5). This method is known in structural mechanics as the error in constitutive law and is due to P. Ladevèze. Here, the error in admissible field is defined by

$$(e^{\text{ADM}})^2 = \int_{\Omega} \left(\hat{\mathbf{v}}^t - \frac{1}{j\rho ck} \nabla^t \hat{p} \right) \left(\hat{\mathbf{v}} + \frac{1}{j\rho ck} \nabla \hat{p} \right) d\Omega \quad (24)$$

We can prove [6] that with an admissible pressure \hat{p} and an admissible velocity $\hat{\mathbf{v}}$, it holds

$$\left\| \hat{\mathbf{v}} + \frac{1}{j\rho ck} \nabla \hat{p} \right\|_0^2 = \|\hat{\mathbf{v}} - \mathbf{v}\|_0^2 + \|\mathbf{v} - \mathbf{v}^h\|_0^2 - \frac{2}{\rho^2 c^2} \|p - p^h\|_0^2 \quad (25)$$

Unlike elastostatics, this theorem shows that the error in admissible field is an upper bound of the exact error *only* asymptotically because the error in L^2 norm (last term in Eq. 26) can not be neglected for practical engineering computations.

4.2. Quality of the *a posteriori* error estimation vs. *k*-singularity

In order to assess the efficiency of the error estimation, we use the effectivity index defined by

$$\theta = \frac{e^*}{\|\mathbf{v} - \mathbf{v}^h\|_0} \quad (26)$$

where e^* stands either for e^{SPR} or for e^{ADM} . The estimation is efficient and convergent if, respectively,

$$0.8 \leq \theta \leq 1.2, \quad \lim_{h \rightarrow 0} \theta = 1 \quad (27)$$

For a Babuska-Miller residual estimator, convergence is proved in [7]. There, also a lower bound for the efficiency is shown. This bound does not satisfy the first condition of (27) if κ is large and κh constant.

Fig. 7 gives the effectivity index as a function of the wave number. It shows that the estimation is efficient for meshes satisfying rule (22) but deteriorates for meshes satisfying rule of the thumb, κh constant. The definition of the *k*-singularity (Eq. 21) shows that this is related to the phase lag between the exact and numerical waves and also to the error on the amplitude. It is then expected that an error estimator based on a local procedure (smoothing like in the Superconvergent Patch Recovery or solving local problems like in the construction of admissible fields) is only able to estimate the approximation error *but* not the pollution term.

On the same model problem, Fig. 8 shows that the estimated error is convergent for all wave numbers. As a matter of fact, estimate (21) shows that the pollution term is asymptotically negligible so that only the approximation error remains which is properly estimated.

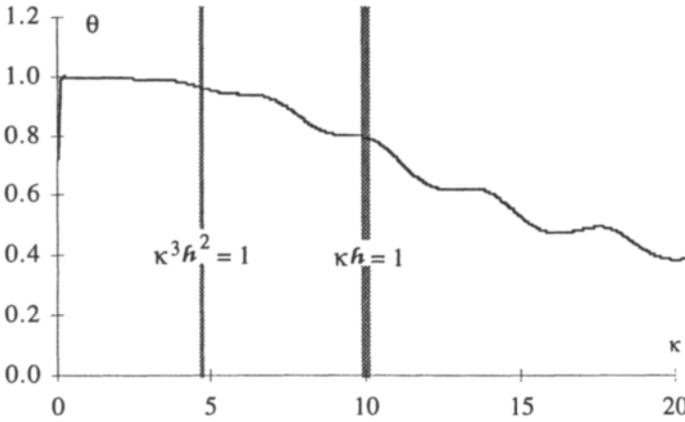


Figure 7. Model problem I: effectivity index of the SPR estimated error as a function of the non-dimensional wave number.

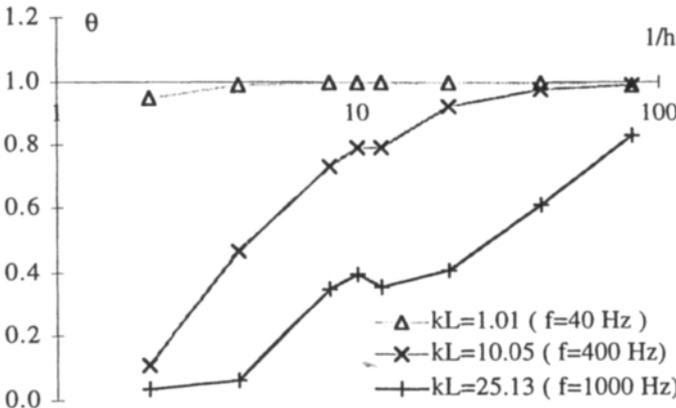


Figure 8. Model problem I: Convergence of the effectivity index of the SPR estimated error for several non-dimensional wave numbers.

4.3. Quality of the *a posteriori* error estimation vs. λ -singularity

Consider now model problem II. Error estimation based on the post-processing of the finite element forced response cannot give any information on the exact eigenvalue. Then, as shown in Fig. 9 plotting the exact and estimated SPR errors as functions of the non-dimensional wave number κ , the estimated error only recognises the existence of the finite element eigenvalues.

For industrial purpose, it is clear here that one will be interested in the error on the eigenvalue itself more than on its amplitude which is infinite.

Fig. 10, giving the effectivity index θ as a function of the wave number, confirms this conclusion but also shows that, between two eigenvalues, the estimated error deteriorates very fast due to the k -singularity.

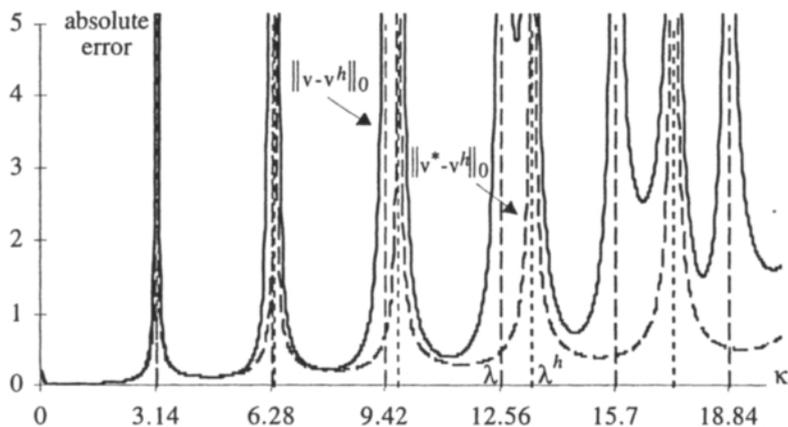


Figure 9. Model problem II: Absolute error in H^1 -seminorm as a function of the non-dimensional wave number κ (exact and SPR).

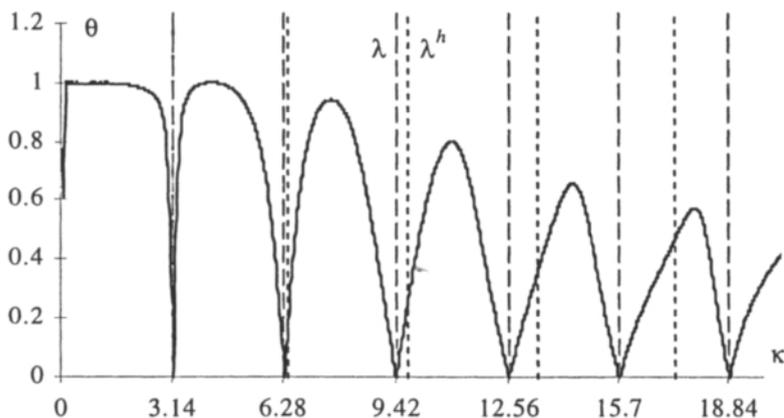


Figure 10. Model problem II: Effectivity index as a function of the non-dimensional wave number κ (SPR).

5. APPLICATION

5.1. Finite element analysis with accuracy control

Consider now a two dimensional section within the cabin of a car, excited by the vibrations of the front panel and damped by Robin boundary conditions. This analysis has been first performed by J. Nefske [8]. Fig. 11 shows the model and its discretisation by finite elements. The mesh consists of 264 elements (198 quadrilaterals and 66 triangles) of degree $p=1$ and non-dimensional size $h=0.034$. With the application of the rule of the thumb κh constant, the user expects confident results up to $\kappa=29.38$ ($f=600$ Hz).

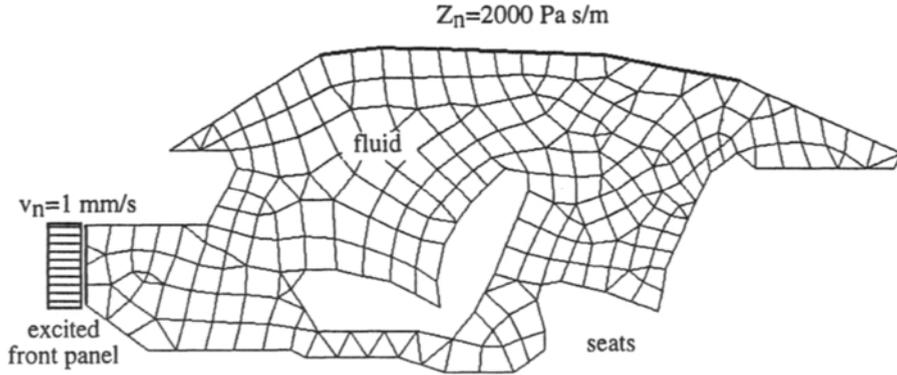


Figure 11. 2D section of a car: mathematical model and initial mesh ($h=0.034$, $p=1$).

Our purpose here is double:

1. To show that, due to the k -singularity, the confidence in the results should be restricted under the limit of rule (Eq. 22 with $C(\bar{\eta})=1$ for instance), i.e. $\kappa=9.79$ ($f=200$ Hz).
2. To show that, for low wave numbers ($\kappa \leq 1$), it is necessary to control the mesh according to the geometric (sharp corners) or to the physical (discontinuity in the boundary conditions) singularities.

First, Fig. 12 gives the convergence of the *estimated error SPR* in H^1 -seminorm for uniform successive refinements. From results of the previous paragraph, we expect that this estimation is not reliable for $\kappa > 9.79$. Yet, the examination of the figure shows:

1. $\kappa \leq 9.79$: the curves are already in the asymptotic range. The quality of the meshes can be controlled by a posteriori error estimators.
2. $9.79 < \kappa \leq 29.38$: the curves exhibit a preasymptotic behaviour for the coarse meshes and an asymptotic behaviour for the refined meshes, i.e. the coarse meshes are polluted. No confidence can be given to the finite element results nor to the error estimation for the coarse meshes.
3. $\kappa > 29.38$: finite element results are completely misleading for the coarse meshes which are under the limit of resolution of rule of the thumb κh constant.

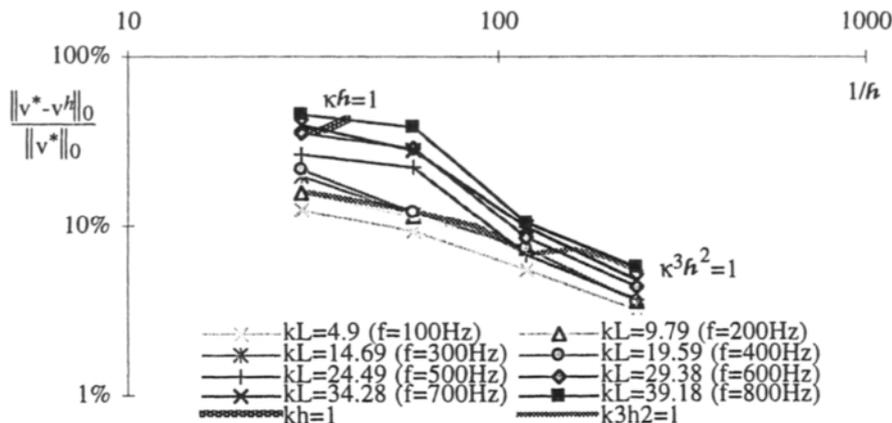


Figure 12. Convergence in H^1 -seminorm for the estimated error SPR for several non-dimensional wave numbers κ .

For low wave numbers, the comparison between both error estimators defined in the previous section shows a very good agreement and suggests that the error is now dominated by the pollution due to the geometric singularities (mainly near the seats) or the physical singularities (discontinuity of the boundary conditions near the front panel for instance). Fig. 13 (a-b) gives the distribution of the absolute error in seminorm estimated by the SPR estimator (enhanced by the incorporation of the boundary conditions) and by the error in admissible fields, respectively.

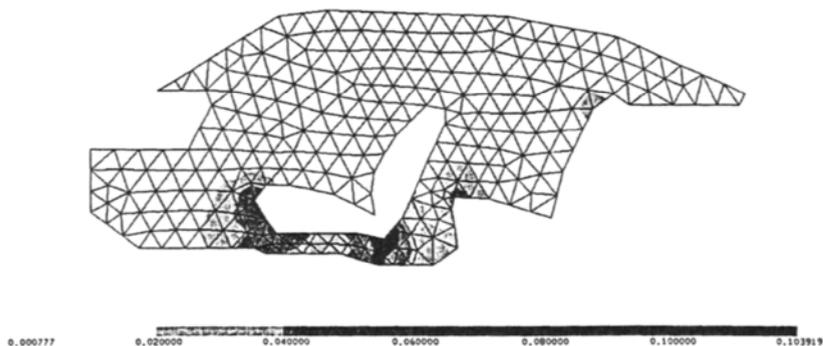


Figure 13 (a). Distribution of the estimated error in H^1 -seminorm by the SPR method ($\kappa=49$, $f=100$ Hz).

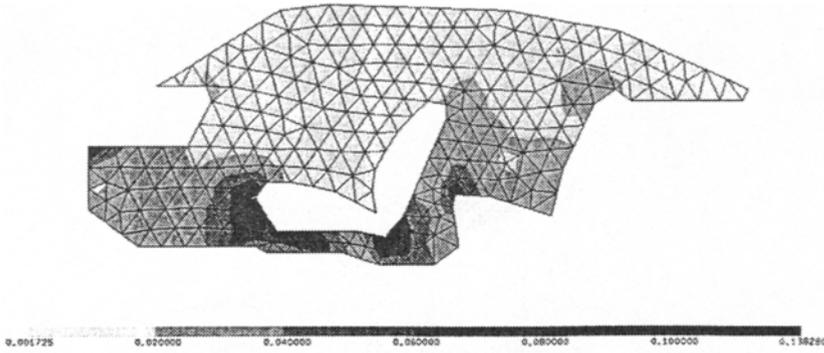


Figure 13' (b). Distribution of the error in admissible fields ($\kappa=4.9$, $f=100$ Hz).

5.2. *h*-adaptive refinements

Consider now the acoustic analysis with a low wave number, e. g. $\kappa=2.45$ ($f=50$ Hz). The error in H^1 -seminorm estimated by the smoothening of the velocity field by the SPR method is 10.9 %. Suppose that the user prescribes a maximum tolerance of 5 %, it is then possible to define an optimal size ratio by

$$\xi_{\tau}(\bar{\eta}) = \frac{h^{old}}{h^{new}} \tag{29}$$

where h^{new} is predicted under the assumptions that the finite element solution already converges asymptotically and that we are able to generate a mesh distributing equally the absolute error. Fig. 14 gives the distribution of the optimal size ratio (29) $\xi_{\tau}(\bar{\eta}=5\%)$ for the initial mesh (element τ must be refined if $\xi_{\tau} > 1$). By applying successive refinement procedures, the optimal *h*-adapted mesh is reached after two steps and is given in Fig 16 ($\eta^{SPR}=4\%$). The advantage of the *h*-adaptive procedure vs. the uniform refinements is shown in Fig. 18 where the *h*-adaptive curve exhibits a better order of convergence.

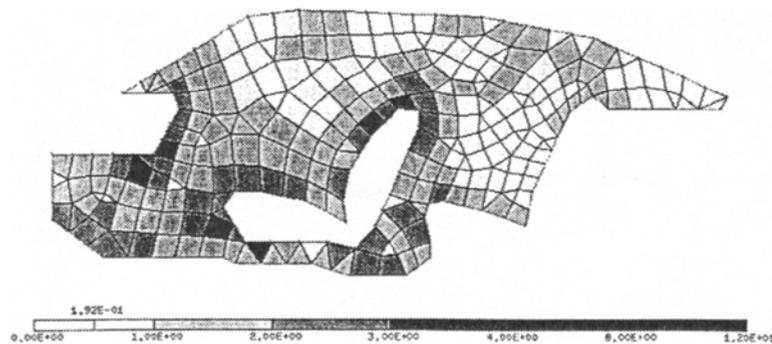


Figure 14. Distribution of the optimal size ratio $\xi_{\tau}(\bar{\eta}=5\%)$ for $\kappa=2.45$ ($f=50$ Hz).

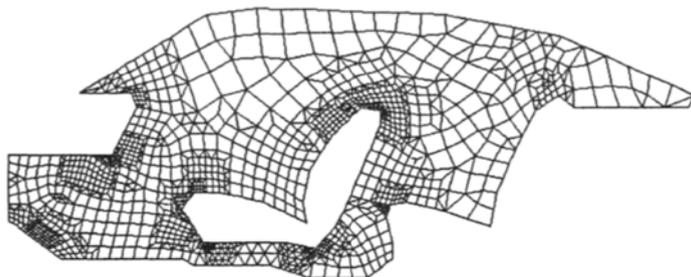


Figure 15. Optimal mesh step 2 - $\eta = 4\%$ ($\kappa=2.45$, $f=50$ Hz)

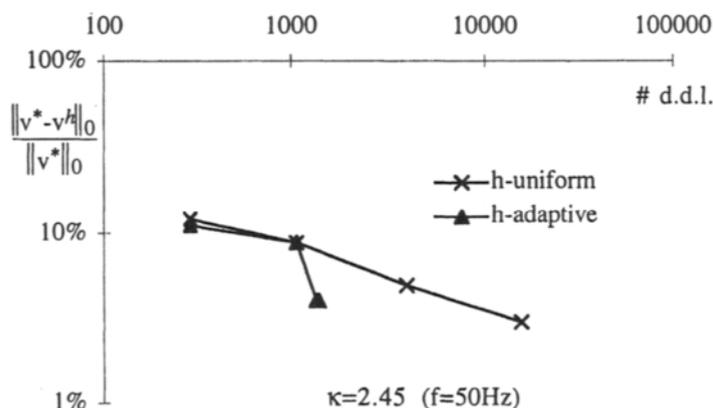


Figure 16. Convergence in H^1 -seminorm for the estimated error SPR uniform refinement vs. adaptive refinement for $\kappa=2.45$ ($f=50$ Hz).

6. CONCLUDING REMARKS

The finite element solutions of Helmholtz equations exhibit two specific singularities inherent to the variational problem for the Helmholtz equation arising from time-harmonic problems (as well as elasto- or electro-dynamics): the k -singularity is defined by the error between the interpolant wave and the finite element wave. Its influence grows with the non-dimensional wave number κ and is not controlled by the traditional rule of the thumb $\kappa h = \text{constant}$. The λ -singularity is defined, for undamped interior problems, by the singularity of the discretised Helmholtz system of equations at the natural frequencies.

Efficient accuracy control in acoustic finite element computations cannot, in general, be achieved by 'transplanting' methods developed for other application fields, e.g. elastostatics. The reason is the specific character of the k - and λ -singularities illustrated in this paper both on one dimensional problems and on a real-life application. In this last case, we also show that the adaptivity is required for low wave numbers, in the absence of specific Helmholtz singularities and present a successful h -adaptive procedure.

ACKNOWLEDGEMENT

The second author was supported by grant Ih-23 of Deutsche Forschungsgemeinschaft.

REFERENCES

1. F. Ihlenburg and I. Babuska, 'Finite Element Solution of the Helmholtz Equation with High Wave Number. Part 1: The h -Version of the FEM', *Computers Math. Applic.*, 38(9) (1995) 9-37.
2. F. Ihlenburg and I. Babuska, 'Finite Element Solution of the Helmholtz Equation with High Wave Number. Part 2: The hp -Version of the FEM', *SIAM Numer. Anal.*, 34(1) (1995) 315-358.
3. F. Ihlenburg and I. Babuska, 'Dispersion Analysis and Error Estimation of Galerkin Finite Element Methods for the Helmholtz Equation', *Int. j. numer. methods eng.*, 38 (1995) 3745-3774.
4. LMS Numerical Technologies, *SYSNOISE Rev. 5.3 users manual* (1997) Interleuvenlaan, 3001 Heverlee, Belgium.
5. Ph. Bouillard, J.-F. Allard and G. Warzée, 'Superconvergent Patch Recovery Technique for the Finite Element Method in Acoustics', *Commun. appl. numer. methods*, 12 (1996) 581-594.
6. Ph. Bouillard, 'Error estimation in constitutive law with low wavenumbers', submitted to *Computers and Structures* (1997).
7. I. Babuska, F. Ihlenburg, T. Strouboulis and S. K. Gangaraj, 'A Posteriori Error Estimation for FEM Solutions of Helmholtz's Equation. Part I: The Quality of Local Indicators and Estimators', *Int. j. numer. methods eng.*, 40 (1997) 3443-3462.
8. D. J. Nefske, 'Sound in small enclosures', in *Noise and Vibration Control Engineering. Principles and Applications*, L. Beranek and I. VÉR (Editors), 1st edition, J. Wiley & Sons, ISBN 0-471-61751-2, London (1992).

Error through the constitutive relation for beam or C° plate finite elements: static and vibration analyses

P. Boisse and G. Coffignal

Laboratoire de Modélisation et Mécanique des Structures
 URA CNRS 1776, ENSAM Paris, Université Paris 6, ENS Cachan
 151, Boulevard de l'Hôpital, 75013 Paris

ABSTRACT

An extension of the error through the constitutive relation is proposed in the case of beams and plates, for static and modal analyses. The error associated to the performed computation is estimated from the lack of verification, by two admissible fields, of the constitutive relation. One is an admissible stress resultant field and the other is the finite element one. Admissible stress fields are obtained by local calculations at the node or element level. A set of examples is presented where the estimated error is compared to the error obtained from the exact solution. These examples show that beam and plate analyses are areas well adapted to the method. In the case of normal mode analyses an error estimation is made for each eigenvector which has an associated frequency within a given range. An example of beam structure shows that it is possible to define a mesh that automatically insures a given accuracy for all modes within a given range.

1. NOTATIONS, DEFINITIONS

$\mathbf{u}_h, \mathbf{U}_h$	finite element displacement and generalised displacement
$\boldsymbol{\varepsilon}(\mathbf{u}_h)$	deformation related to \mathbf{u}_h ($\boldsymbol{\varepsilon}(\mathbf{u}_h) = \frac{1}{2}(\text{grad } \mathbf{u} + \text{grad}^T \mathbf{u})$)
\mathbf{e}_h	generalised deformation $\mathbf{e}_h = \mathbf{L} \mathbf{U}_h$ (\mathbf{L} convenient differential operator)
$\boldsymbol{\sigma}_h, \mathbf{S}_h$	stress and generalised finite element stress
$\hat{\mathbf{u}}$	kinematically admissible (KA) displacement field derived from \mathbf{u}_h
$\hat{\boldsymbol{\sigma}}, \hat{\mathbf{S}}$	statically admissible (SA) stress and generalised stress
\mathbf{C}	elastic Hooke's tensor or elastic matrix for beams
r_Ω, r_E	error through the constitutive equation on domain Ω , on element E
$\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$	natural orthogonal basis in plate analyses, \mathbf{e}_3 is normal to the plate
\mathbf{n}_p	unit vector in plane $\mathbf{e}_1, \mathbf{e}_2$

- α, β indices taking the value 1 or 2
- σ plane stress tensor in the case of plates, $\sigma = \sigma_{\alpha\beta} \mathbf{e}_\alpha \otimes \mathbf{e}_\beta$
- T, M, Q** generalised stress: membrane forces, bending moments, shear force
- T** $\mathbf{T} = T_{\alpha\beta} \mathbf{e}_\alpha \otimes \mathbf{e}_\beta$; $T_{\alpha\beta} = \int_{-h/2}^{+h/2} \sigma_{\alpha\beta} dx_3$ membrane tensor (plate)
- M** $\mathbf{M} = M_{\alpha\beta} \mathbf{e}_\alpha \otimes \mathbf{e}_\beta$; $M_{\alpha\beta} = \int_{-h/2}^{+h/2} x_3 \sigma_{\alpha\beta} dx_3$ bending moment tensor (plate)
- Q** $\mathbf{Q} = Q_\alpha \mathbf{e}_\alpha$; $Q_\alpha = \int_{-h/2}^{+h/2} \sigma_{\alpha 3} dx_3$ shear force vector (plate)
- S_m mid surface of a plate
- $\bar{\mathbf{p}}, p_3$ column of inplane load components and normal load on S_m (plate)
- $\partial S_{m\nu} \partial S_u$ part of the boundary of S_m with prescribed applied loads, displacements
- $\bar{\mathcal{F}}, \bar{\mathcal{M}}$ prescribed in plane load vector, moment vector on ∂S_m
- $\bar{q}\mathbf{e}_3$ prescribed transverse load vector on ∂S_m
- $[\bar{\mathbf{t}}]$ column of prescribed load components on ∂S_m^E , $[\bar{\mathbf{t}}] = [\bar{\mathcal{F}} \ \bar{q} \ \bar{\mathcal{M}}]^T$
- $[\mathbf{p}]$ column of prescribed load components on S_m^E , $[\mathbf{p}] = [\bar{\mathbf{p}}^T \ p_3 \ 0^T]^T$
- Γ segment, edge an element or common two adjacent elements
- $\hat{\mathcal{F}}_\Gamma$ $\hat{\mathcal{F}}_\Gamma = \hat{C}_\Gamma^1 \mathbf{e}_1 + \hat{C}_\Gamma^2 \mathbf{e}_2 + \hat{C}_\Gamma^3 \mathbf{e}_3$ load vector on Γ
- $\hat{\mathcal{M}}_\Gamma$ $\hat{\mathcal{M}}_\Gamma = \hat{C}_\Gamma^4 \mathbf{e}_1 + \hat{C}_\Gamma^5 \mathbf{e}_2$ moment vector on Γ
- N^i linear interpolation function (scalar) related to node i
- \mathbf{N} displacement interpolation matrix of an element
- \mathbf{B} strain interpolation matrix of an element
- $\int_\Delta^i P(C)$ $= \int_\Delta N^i C d\Gamma$ weighted value of a scalar density C along line Δ , at node i
- γ_E, γ_i set of edges Γ of element E , set of edges Γ connected to node i
- $\eta, [\eta]$ virtual displacement vector and related components verifying $\eta = 0$ on ∂S_m . δ being the mid-surface virtual displacement, and θ the virtual rotation, we have $\eta = \delta + \theta x_3$ for plates.
- $[\delta_n^e], [\theta_n^e]$ column of mid-surface nodal virtual displacement components and column of nodal virtual rotation components
- $\beta_{\Gamma E}, \beta_{\Gamma E'}$ orientation functions of two adjacent elements E and E' with common edge Γ . For instance, $\beta_{\Gamma E} = 1$ and $\beta_{\Gamma E'} = -1$ if E' has a greater index than E in the global numbering. β_Γ have opposite signs otherwise.
- l_Γ length of edge Γ
- U_i, ω_i exact mode shape i and its natural angular frequency

S_i	generalised stress associated to mode i for the theory (beam or plate)
Y_i, ω_{hi}^2	finite element eigenvector and eigenvalue associated to mode i
m	mass operator depending on the theory (beam or plate)
F_{li}	$F_{li} = \omega_i^2 m U_i$ generalised inertia forces associated to mode i
K, M	stiffness matrix and consistent mass matrix
A_E	dof localisation operator for element E
U_{hi}	generalised finite element displacement for mode i : $U_{hi} = N Y_{Ei}$
Y_{Ei}	column of dof associated to mode i for element E : $Y_{Ei} = A_E Y_i$

2. INTRODUCTION

Numerical tools able to give a bound or to estimate the error of a finite element computation have become essential for practical analyses. In addition to the quality of a calculation, they allow an automatic definition of so-called 'adaptative meshes' after a first coarse analyse, using a mesh generator and the map of the error over a structure. Several methods or families of methods have been proposed in order to estimate *a posteriori* errors. The most commonly used error estimators have been proposed by Zienkiewicz and Zhu [1-2] and are based on recovery methods and especially on a superconvergent patch recovery technique [3-6]. Another family of indicators is based on the element equilibrium residuals and has been initiated by Babuska [7-8]. The concept of "error through the constitutive relation" has been introduced by Ladevèze [9]. An error estimation is calculated from strictly admissible static and kinematic fields built from the finite element results. Their lack of verification of the constitutive relation defines the error which is calculated as a strain energy. Its strong mechanical foundation is the main interest of the method since the exact solution of the problem is strictly admissible. Especially, the lack verification of the boundary conditions is taken into account. The method has been adapted to a large class of 2D and 3D linear and non-linear analyses. In particular for linear 2D analyses [10-13], linear 3D analyses [14-15], elastoplasticity [16-18], viscoplasticity [19], vibrations [16], [20] or large strains [21].

In practical finite elements analyses, a large class of structures present one or two dimensions that are small with respect to others. They can be respectively analysed within the beam, plate or shell theories. The finite elements including the kinematic assumptions of these theories are numerous, efficient and very much used in practice. The objective of this paper is to present an application of the error through the constitutive relation method to 2D and 3D beam [22-23] and Mindlin plates [24] static and free vibration analyses. It also aims to show the efficiency of the method in these kinds of applications.

In a first section, the calculation of the error from the construction of an optional statically admissible stress field is presented for static Mindlin plate

analyses. The good efficiency of the calculation is shown on a set of classical examples performed using a three node finite element. Then the calculation of the error through the constitutive equation is presented for nodal analyses of 2D and 3D beam structures. On the example of a "clamped beams" truss, it is shown that the error estimation allows the automatic definition of a mesh insuring a given quality over a prescribed frequency range. The complexity of the mode shapes under consideration renders an empirical approach quasi impossible. At last the error estimation method for eigen frequency calculations is extended to plate analyses. For all the presented results, the comparison of the estimated error with the error obtained from the exact solution (analytical one or solution obtained from a very fine mesh) is emphasised.

3. ERROR THROUGH THE CONSTITUTIVE RELATION

Notations and definitions are given in section 1. The study is restricted to geometric and material linearity, so the constitutive relation is:

$$\sigma_h = C : \varepsilon(u_h) \quad (1)$$

Within a classical displacement approach, u_h is kinematically admissible and we can choose to take:

$$\hat{u} = u_h \quad (2)$$

The error through the constitutive relation on a domain Ω is defined from the lack of verification of the constitutive relation by the couple of admissible fields \hat{u} and $\hat{\sigma}$ [12],[25]

$$r_\Omega = \frac{\|\hat{\sigma} - C : \varepsilon(\hat{u})\|_\Omega}{\|\hat{\sigma} + C : \varepsilon(\hat{u})\|_\Omega} \quad \text{where} \quad \|\sigma\|_\Omega = \left(\int_\Omega \sigma : C^{-1} : \sigma dV \right)^{1/2} \quad (3)$$

An adaptative meshing procedure needs a map of the error over the structure. The contribution from each element is considered

$$r_E = \frac{\|\hat{\sigma} - C : \varepsilon(\hat{u})\|_E}{\|\hat{\sigma} + C : \varepsilon(\hat{u})\|_\Omega} \quad \text{with} \quad r_\Omega^2 = \sum_{\text{elements}} r_E^2 \quad (4)$$

From the Prager-Synge's theorem, it has been shown in [10] that, in statics, the error through the constitutive relation provides an upper bound of u_h error.

4. STATIC ANALYSES OF MINDLIN PLATES

4.1. Error for Mindlin plates

A statically admissible stress field $\hat{\sigma}$ corresponds to generalised stress

$\hat{\mathbf{S}} = [\hat{\mathbf{T}}, \hat{\mathbf{Q}}, \hat{\mathbf{M}}]^T$ verifying local equilibrium equations (5) for the Mindlin plate model in S_m and static boundary conditions on ∂S_m (6).

$$\begin{cases} \operatorname{div} \hat{\mathbf{T}} + \mathbf{p} = 0 \\ \operatorname{div} \hat{\mathbf{Q}} + \mathbf{p}_3 = 0 \\ \operatorname{div} \hat{\mathbf{M}} - \hat{\mathbf{Q}} = 0 \end{cases} \quad \text{which can be written more concisely} \quad \mathbf{D} \hat{\mathbf{S}} + \mathbf{F} = 0 \quad (5)$$

$$\begin{cases} \hat{\mathbf{T}} \cdot \mathbf{n}_p = \bar{\mathcal{F}} \\ \hat{\mathbf{Q}} \cdot \mathbf{n}_p = \bar{\mathcal{q}} \\ \hat{\mathbf{M}} \cdot \mathbf{n}_p = \bar{\mathcal{M}} \end{cases} \quad (6)$$

The error through the constitutive relation r_Ω is obtained from $\hat{\mathbf{S}} = [\hat{\mathbf{T}}, \hat{\mathbf{Q}}, \hat{\mathbf{M}}]^T$, and the finite element result $S_h = [T_h, M_h, Q_h]^T$. It is defined by:

$$r_\Omega = \frac{\left(\int_S (\hat{\mathbf{M}} - \mathbf{M}_h) : \mathbf{D}^{b^{-1}} : (\hat{\mathbf{M}} - \mathbf{M}_h) + (\hat{\mathbf{Q}} - \mathbf{Q}_h) : \mathbf{D}^{s^{-1}} : (\hat{\mathbf{Q}} - \mathbf{Q}_h) + (\hat{\mathbf{T}} - \mathbf{T}_h) : \mathbf{D}^{m^{-1}} : (\hat{\mathbf{T}} - \mathbf{T}_h) dS \right)^{1/2}}{\left(\int_S (\hat{\mathbf{M}} + \mathbf{M}_h) : \mathbf{D}^{b^{-1}} : (\hat{\mathbf{M}} + \mathbf{M}_h) + (\hat{\mathbf{Q}} + \mathbf{Q}_h) : \mathbf{D}^{s^{-1}} : (\hat{\mathbf{Q}} + \mathbf{Q}_h) + (\hat{\mathbf{T}} + \mathbf{T}_h) : \mathbf{D}^{m^{-1}} : (\hat{\mathbf{T}} + \mathbf{T}_h) dS \right)^{1/2}} \quad (7)$$

Bending, membrane and shear parts of the error can be separated limiting the numerator to the corresponding energy in (7).

4.2. C° plate finite elements

Three node C° plate finite elements with linear interpolations are considered here. The triangular shell element proposed in [26-27] will be used in the examples. Other plate elements with linear interpolations such as Belytshcko's 3 node plate element proposed in [28] can be used as well. In so far the deformation interpolation matrix \mathbf{B} used in the error calculation is the one used in the finite element calculation, the method is unchanged.

4.3. Bases of the statically admissible stress resultants construction

The statically admissible stress resultants $\hat{\mathbf{T}}, \hat{\mathbf{Q}}, \hat{\mathbf{M}}$ (the knowledge of which is equivalent to the one of $\hat{\sigma}$) have to verify (5) and (6). Statically admissible stress resultant are not unique. Their calculation will be made in order to obtain an error estimation as efficient as possible. In that goal, two ideas lead the determination of the admissible stress resultants.

Calculations are chosen to be local in order to render efficient, in term of numerical cost, the error estimation for large number of degrees of freedom analyses. To this aim, and as it has been done for 2D and 3D problems [10], [16], [17] the calculation of $\hat{\mathbf{T}}, \hat{\mathbf{Q}}, \hat{\mathbf{M}}$ verifying (5) and (6) is done by means of $\hat{\mathbf{T}}_E, \hat{\mathbf{Q}}_E, \hat{\mathbf{M}}_E$, at the element level. These generalised stresses have to verify: i) local equilibrium within element E , ii) boundary conditions on the edges Γ of element E . To manage these boundary condition, loads and bending moments densities $\hat{\mathcal{F}}_\Gamma$ and

$\hat{\mathcal{M}}_\Gamma$ due to the neighbouring elements (or exterior loads) are introduced. They insure the equilibrium of each element and are associated to each edge Γ .

As Prager-Syngé's theorem shows that the error through the constitutive relation provides an upper bound of the error, the error estimation will be minimised by choosing SA stress resultants as close as possible to the finite element field. This last choice plays a major role in the determination of an efficient admissible stress field and in practical definition of the different stages described below.

4.4. In equilibrium and reciprocal load and moment densities on element edges

In a first stage loads $\hat{\mathcal{F}}_\Gamma$ and bending moments $\hat{\mathcal{M}}_\Gamma$ are calculated on each element edge Γ in order to insure the equilibrium of element E submitted to $\beta_{\Gamma E} \hat{\mathcal{F}}_\Gamma$ and $\beta_{\Gamma E} \hat{\mathcal{M}}_\Gamma$ and eventually exterior loads. These densities $\beta_{\Gamma E} \hat{\mathcal{F}}_\Gamma$ and $\beta_{\Gamma E} \hat{\mathcal{M}}_\Gamma$ represent the actions of its neighbours on element E. They insure to get opposite values for two elements connected to the same edge. The components of $\hat{\mathcal{F}}_\Gamma$, $\hat{\mathcal{M}}_\Gamma$ are equal to exterior loads on the part ∂S_m of the boundary. Densities $\hat{\mathcal{F}}_\Gamma$, $\hat{\mathcal{M}}_\Gamma$ will be chosen as close as possible to the finite element result $\mathcal{F}_{\Gamma m}$, $\mathcal{M}_{\Gamma m}$ defined by the finite element stress solution on the edge Γ relating nodes i and j:

$$\mathcal{F}_{\Gamma m} = \int_{-b/2}^{+b/2} (N^i \sigma_m^i + N^j \sigma_m^j) \cdot \mathbf{n}_\Gamma dx_3 \quad \mathcal{M}_{\Gamma m} = \int_{-b/2}^{+b/2} x_3 (N^i \sigma_m^i + N^j \sigma_m^j) \cdot \mathbf{n}_\Gamma dx_3 \quad (8)$$

σ_m^i is the nodal average stress usually calculated by finite element post-processors.

$\hat{\mathcal{F}}_\Gamma$ and $\hat{\mathcal{M}}_\Gamma$ are defined by (9) and have a linear evolution along each edge Γ

$$\hat{\mathcal{F}}_\Gamma = \hat{\mathcal{F}}_{\Gamma a} + \hat{\mathcal{F}}_{\Gamma b} \quad \text{and} \quad \hat{\mathcal{M}}_\Gamma = \hat{\mathcal{M}}_{\Gamma a} + \hat{\mathcal{M}}_{\Gamma b} \quad (9)$$

First densities $\hat{\mathcal{F}}_{\Gamma a}$, $\hat{\mathcal{M}}_{\Gamma a}$ are directly calculated from the finite element solution. The goal is to make a simple calculation and to get back some qualities of the finite element solution concerning the "on average" proximity with the exact solution. Second densities $\hat{\mathcal{F}}_{\Gamma b}$, $\hat{\mathcal{M}}_{\Gamma b}$ are used to improve the first ones and render $\hat{\mathcal{F}}_\Gamma$, $\hat{\mathcal{M}}_\Gamma$ as close as possible to $\mathcal{F}_{\Gamma m}$, $\mathcal{M}_{\Gamma m}$. For plate analyses, restricting the test displacement $\eta = \delta + \theta x_3$ to its interpolated form

$$\delta = \delta_k e_k \quad , \quad \theta = \theta_\alpha e_\alpha \quad , \quad [\delta] = \mathbf{N} [\delta_n^e] \quad , \quad [\theta] = \mathbf{N} [\theta_n^e] \quad (10)$$

Virtual work principle applied to the whole structure leads to the finite element solution such that:

$$\sum_{n=1}^{\substack{\text{n = number} \\ \text{of elements}}} [(\delta_n^e)^T (\theta_n^e)^T] \left(\int_{\Omega^e} \mathbf{B}^T [\hat{\sigma}_h] dV - \int_{S_m^e} \mathbf{N}^T [\mathbf{p}] dS - \int_{\partial S_m^e} \mathbf{N}^T [\mathbf{t}] d\Gamma \right) = 0 \quad (11)$$

Considering the particular case where δ_n^e and θ_n^e are equal to zero for any node

to which element E is not connected, (11) leads to:

$$[\mathbf{F}_{\text{neighbour}}] + [\mathbf{F}_E] = 0 \quad (12)$$

$[\mathbf{F}_{\text{neighbour}}]$ is a column (13) containing nodal load components. It corresponds to the actions of neighbouring elements on element E. The finite element solution insures the equilibrium of any element E submitted to these actions with other exterior loads applied on E.

$$[\mathbf{F}_{\text{neighbour}}] = \sum_{p=1}^{\text{p=number of neighbouring elements}} \left(\int_{\Omega^p} \mathbf{B}^T [\hat{\sigma}_h] dV - \int_{S_m^p} \mathbf{N}^T [\mathbf{p}] dS - \int_{\partial S_m^p} \mathbf{N}^T [\mathbf{t}] d\Gamma \right) \quad (13)$$

$$[\mathbf{F}_E] = \int_{\Omega^E} \mathbf{B}^T [\hat{\sigma}_h] dV - \int_{S_m^E} \mathbf{N}^T [\mathbf{p}] dS - \int_{\partial S_m^E} \mathbf{N}^T [\mathbf{t}] d\Gamma \quad (14)$$

As stated before, $\hat{\mathcal{F}}_{\Gamma_a}$, and $\hat{\mathcal{M}}_{\Gamma_a}$ are chosen linear along each edge Γ , such as the oriented actions $\beta_{\Gamma E} \hat{\mathcal{F}}_{\Gamma_a}$ and $\beta_{\Gamma E} \hat{\mathcal{M}}_{\Gamma_a}$ have nodal components equal to $[\mathbf{F}_{\text{neighbour}}]$. Introducing components of $\hat{\mathcal{F}}_{\Gamma_a}$, and $\hat{\mathcal{M}}_{\Gamma_a}$, equations (12-14) lead to:

$$\int_{\partial E} \mathbf{N}^T [\beta_{\Gamma E} \hat{\mathcal{C}}_{\Gamma_a}^k] dS = \left[\begin{matrix} i \\ \partial E \end{matrix} P(\beta_{\Gamma E} \hat{\mathcal{C}}_{\Gamma_a}^k) \right] = \int_{S_m^E} \mathbf{B}^T [\hat{\sigma}_h] dS - \int_{S_m^E} \mathbf{N}^T [\mathbf{p}] dS - \int_{\partial S_m^E} \mathbf{N}^T [\mathbf{t}] dS \quad (15)$$

Consequently, $\begin{matrix} i \\ \mathcal{E} \end{matrix} P(\beta_{\Gamma E} \hat{\mathcal{C}}_{\Gamma_a}^k)$ are quantities determined from the finite element calculation. Considering a node i , nodal components $\begin{matrix} i \\ \mathcal{E} \end{matrix} P(\beta_{\Gamma E} \hat{\mathcal{C}}_{\Gamma_a}^k)$ are related to components on each edge

$$\begin{matrix} i \\ \partial E \end{matrix} P(\beta_{\Gamma E} \hat{\mathcal{C}}_{\Gamma_a}^k) = \sum_{\Gamma \in \gamma_E \cap \gamma_i} \beta_{\Gamma E} \int_{\Gamma} \mathbf{N}^i \hat{\mathcal{C}}_{\Gamma_a}^k d\Gamma \quad (16)$$

A local and very small system (17) is obtained for each node i :

$$\sum_{\Gamma \in \gamma_E \cap \gamma_i} \beta_{\Gamma E} \begin{matrix} i \\ \Gamma \end{matrix} P(\hat{\mathcal{C}}_{\Gamma_a}^k) = \begin{matrix} i \\ \mathcal{E} \end{matrix} P(\beta_{\Gamma E} \hat{\mathcal{C}}_{\Gamma_a}^k) \quad \forall E \text{ containing node } i. \quad (17)$$

Nodal load components $\begin{matrix} i \\ \mathcal{E} \end{matrix} P(\hat{\mathcal{C}}_{\Gamma_a}^k)$ are the unknowns.

Their number is equal to the number of edges related to node i . The number n of equations in (17) is equal to the number of element connected to i . When node i is interior, the rank of system (17) is $n - 1$. In this case, a supplementary equation is obtained choosing to minimise of the distance between the unknown nodal components and those of \mathcal{F}_{Γ_m} and \mathcal{M}_{Γ_m} *i. e.*:

$$\sum_{\Gamma \in \gamma_E \cap \gamma_i} \left[\begin{matrix} i \\ \Gamma \end{matrix} P(\hat{\mathcal{C}}_{\Gamma_a}^k) - \begin{matrix} i \\ \Gamma \end{matrix} P(\mathcal{C}_{\Gamma_m}^k) \right]^2 = \text{minimum} \quad (18)$$

(17) has a unique solution if $\hat{\mathcal{C}}_{\Gamma_a}^k$ is imposed on one or two edges of γ_i .

Densities $\hat{\mathcal{F}}_{\Gamma_a}$, $\hat{\mathcal{M}}_{\Gamma_a}$ insure the equilibrium of each element and the global equilibrium of the structure. As stated before, densities $\hat{\mathcal{F}}_{\Gamma_b}$, $\hat{\mathcal{M}}_{\Gamma_b}$ aim improve the quality of $\hat{\mathcal{F}}_{\Gamma}$, $\hat{\mathcal{M}}_{\Gamma}$ in the sense they are chosen as close as possible to \mathcal{F}_{Γ_m} , \mathcal{M}_{Γ_m} . In order to preserve the equilibrium properties, the calculation of $\hat{\mathcal{F}}_{\Gamma_b}$, $\hat{\mathcal{M}}_{\Gamma_b}$ is made prescribing the following conditions on any edge Γ

$$e_{\Gamma_b} = \int_{\Gamma} \hat{\mathcal{F}}_{\Gamma_b} d\Gamma = 0 \quad (19)$$

$$\mu_{\Gamma_b}^o = \int_{\Gamma} \hat{\mathcal{M}}_{\Gamma_b} d\Gamma + \int_{\Gamma} \mathbf{OM} \wedge \hat{\mathcal{F}}_{\Gamma_b} d\Gamma = 0 \quad (20)$$

The calculations of shear load and bending moment components $\hat{C}_{\Gamma_b}^3$, $\hat{C}_{\Gamma_b}^4$, $\hat{C}_{\Gamma_b}^5$ in one hand and membrane loads $\hat{C}_{\Gamma_b}^1$, $\hat{C}_{\Gamma_b}^2$ on the other hand are independent.

Shear and bending components are calculated in order to render $(e_{\Gamma_b} \cdot e_3) e_3$ and $(\mu_{\Gamma_b}^o \cdot e_\alpha) e_\alpha$ equal to zero and to obtain three bending components $\hat{C}_{\Gamma_b}^3$, $\hat{C}_{\Gamma_b}^4$, $\hat{C}_{\Gamma_b}^5$ with slopes equal to those of the components $\hat{C}_{\Gamma_m}^3$, $\hat{C}_{\Gamma_m}^4$, $\hat{C}_{\Gamma_m}^5$ respectively. This leads to six conditions determining $\hat{C}_{\Gamma_b}^3$, $\hat{C}_{\Gamma_b}^4$, $\hat{C}_{\Gamma_b}^5$, due to their linear nature:

$$\hat{C}_{\Gamma_b}^3 = -\frac{1}{\Gamma} \int_{\Gamma} (C_{\Gamma_m}^3 - \hat{C}_{\Gamma_a}^3) d\Gamma + C_{\Gamma_m}^3 - \hat{C}_{\Gamma_a}^3 \quad (21)$$

$$\hat{C}_{\Gamma_b}^4 = -\frac{1}{\Gamma} \int_{\Gamma} (C_{\Gamma_m}^4 - \hat{C}_{\Gamma_a}^4 + x_2 \hat{C}_{\Gamma_b}^3) d\Gamma + C_{\Gamma_m}^4 - \hat{C}_{\Gamma_a}^4 \quad (22)$$

$$\hat{C}_{\Gamma_b}^5 = -\frac{1}{\Gamma} \int_{\Gamma} (C_{\Gamma_m}^5 - \hat{C}_{\Gamma_a}^5 - x_1 \hat{C}_{\Gamma_b}^3) d\Gamma + C_{\Gamma_m}^5 - \hat{C}_{\Gamma_a}^5 \quad (23)$$

Membrane components $\hat{C}_{\Gamma_b}^1$ and $\hat{C}_{\Gamma_b}^2$ are calculated in order to render $e_{\Gamma_b}^1$, $e_{\Gamma_b}^2$, $\mu_{\Gamma_b}^3$ equal to zero. Their determination with slopes equal to those of $C_{\Gamma_m}^1$, $C_{\Gamma_m}^2$ is not possible exactly. A least square minimisation of their distance is made.

4.5. Statically admissible stress resultants within the element

Once loads $\beta_{\Gamma_E} \hat{\mathcal{F}}_{\Gamma}$ and moments $\beta_{\Gamma_E} \hat{\mathcal{M}}_{\Gamma}$ on the edges of each element E have been calculated, they are introduced in boundary conditions (6). Stress resultants \hat{T}_E , \hat{Q}_E , \hat{M}_E verifying local equilibrium equations (5) can thus be calculated. This second stage, as the previous one is local and concerns each element separately.

Determination of the shear force \hat{Q}_E and bending moment tensor \hat{M}_E on the one hand and normal tensors \hat{T}_E on the other hand can be done separately, equilibrium equations and boundary conditions being independent. Determination of \hat{T}_E is formally identical to those proposed for the stress tensor in case of 2D plane stress problems [17]. We only present the determination of the couple \hat{Q}_E , \hat{M}_E . Nodal value computation for these fields only needs boundary conditions. However, these constraints connected with each element edge do not allow the construction of a regular field \hat{M} over the triangular plate element. At node i, boundary conditions on the two edges Γ_E and Γ_E' of element E connected to node i imposes:

$$\hat{\mathbf{M}}_{\mathbf{E} \cdot \mathbf{n}_{\Gamma_E}} = \beta_{\Gamma_E} \hat{\mathcal{M}}_{\Gamma} \quad \text{and} \quad \hat{\mathbf{M}}_{\mathbf{E} \cdot \mathbf{n}_{\Gamma'_E}} = \beta_{\Gamma'_E} \hat{\mathcal{M}}_{\Gamma'} \quad (24)$$

which does not have any solution is the general case. A subdivision of Watwood [29] in three subtriangles, based on the centre G of the element, is used. This subdivision has been used in the case of plane stress [16-17] for equivalent reasons. Moment boundary conditions and continuity conditions at each vertex node of the element determines nodal values of $\hat{\mathbf{M}}$ at the corresponding nodes for the two connected subtriangles.

Different choices are possible for the variation of $\hat{\mathbf{Q}}$ and $\hat{\mathbf{M}}$ over each subtriangle. The approach we follow here and that has proved to be efficient from the effectivity point of view, is to determine the stress resultants within a set which is large enough and where the solutions verifying equilibrium equations and boundary conditions are not unique. This allows to add supplementary conditions aiming an error minimisation. This gives the best solution among the set of admissible ones. In the case of plate static analyses, for instance, $\hat{\mathbf{Q}}$ is assumed to be linear and $\hat{\mathbf{M}}$ quadratic over each subtriangle. Under these assumptions, equilibrium equations, boundary conditions and continuity conditions within an element leads to a system with thirty-nine equations and thirty-nine unknowns. However, due to the fact that the external loading on element E is already in equilibrium, the rank of this system is only thirty-six. This system is solved in order to get an optimal stress field leading to a minimum for the error through the constitutive relation. Denoting $H(\hat{\mathbf{M}}, \hat{\mathbf{Q}})$ the part of the error through constitutive relation associated to bending and transverse shear, a new function (25) is introduced, where f_i for $i \in [1, \dots, 36]$ represent the independent constraint equations and λ_i are Lagrange's multipliers.

$$R(\hat{\mathbf{M}}, \hat{\mathbf{Q}}, \lambda_i) = H(\hat{\mathbf{M}}, \hat{\mathbf{Q}}) + \sum_{i=1}^{36} \lambda_i f_i \quad (25)$$

The unknowns $[\hat{\mathbf{M}}^T \hat{\mathbf{Q}}^T \lambda^T]$ are solution of

$$\frac{\partial R}{\partial \hat{\mathbf{M}}} = 0 \quad \frac{\partial R}{\partial \hat{\mathbf{Q}}} = 0 \quad \frac{\partial R}{\partial \lambda_i} = 0 \quad \forall i \in [1, \dots, 36] \quad (26)$$

5. NORMAL MODES OF BEAMS OR PLATES

We consider here undamped vibration of general structures in the case of the displacement approach. Then, we particularise to beams and plates. The natural vibrations are the particular solutions of the problem, in the absence of applied external loads:

$$\mathbf{U}_i^*(M, t) = \mathbf{U}_i(M) \cos \omega_i t \quad (27)$$

$$\mathbf{S}_i^*(M, t) = \mathbf{S}_i(M) \cos \omega_i t \quad (28)$$

In the case of free vibrations, the equilibrium equation becomes:

$$\forall [\eta] \int_{\Omega} (-[\eta]^T F_{II} + (L[\eta])^T S_i) d\Omega = 0 \quad (29)$$

Once the problem has been discretized, the displacement approach leads to the usual equation:

$$(K - \omega_{hi}^2 M) Y_i = 0 \quad (30)$$

5.1. Error indication for a frequency range.

For dynamic analyses, finding a good mesh is a tricky job, even when focusing only on the first mode shapes and frequencies. As a matter of fact, mass and stiffness distributions complicate very deeply the behaviour of the structure.

The indicator α we propose is still based on the concept of error through the constitutive relation defined in statics. The main difference is that an estimator α_i has to be used for each natural mode i .

α_i has been introduced by Pelle [16], [20] and we have shown its application to beams in [22-23].

It is important to emphasise that, as in statics, a zero value for α_i implies that mode i as described by U_{hi} and ω_{hi} is the exact one, with respect to the underlying beam or plate theory. Indicator α is defined by:

$$\alpha = \max_{i \leq i_{\max}} \alpha_i \quad (31)$$

where i_{\max} must be carefully chosen due to the following difficulties which appear when dealing with approximate modes.

A first difficulty, when considering natural modes of vibration, comes from the fact that, using a coarse mesh, it is not possible to know the actual exact number of modes existing within a given range of frequency $[0, f_{\max}]$ ($f_{\max} = \omega_{\max}/2\pi$). However, this information is very important in engineering practice.

A second difficulty, comes from the fact that, in structures, the accuracy of modes is not a decreasing function of index i for a given mesh. This can be observed, for instance, on a simple beam when longitudinal modes and bending modes are both considered. An example of procedure to derive i_{\max} is given in the section devoted to the automatic refinement strategy.

5.2. Dynamically admissible stress field for mode i .

Another difficulty arises when considering natural modes of vibration: in opposition to the static case, loads that generalised stresses have to equilibrate, for exact mode i , are the inertia forces F_{II} which are not known. Their definition shows explicitly their dependence upon the unknown exact mode shape.

In order to construct the estimator α_i associated to FE mode i , we utilise the

expression F_{Ihi} available in the analysis and defined by

$$F_{Ihi} = \omega_{hi}^2 \mathbf{m} U_{hi} \quad (32)$$

So we define a dynamically admissible stress field for mode i as a stress field which is in equilibrium with zero applied external load but under inertia loads F_{Ihi} .

The main difference with statics is that, using this approach, we do not preserve the upper bound property of the estimator. This is a consequence of the fact that F_{Ihi} depends upon the mesh and that F_{Ii} cannot be known *a priori*.

5.3. Beam structures.

For the sake of simplicity, we consider plane (2D) straight Euler beam elements. The generalisation to the corresponding 3D beam element including torsion is straightforward.

We consider here the usual C^0 linear element for longitudinal displacement u and C^1 cubic element for transverse displacement v . $\theta = v'$ is the associated rotation of cross section. This element gives directly the exact the solution in statics, when loads are only applied on nodes.

In local coordinate system of element E :

$$U_h = (u_h, v_h, v'_h)^T = \mathbf{N} U_E \quad (33)$$

$$e_h = (u'_h, v''_h)^T = (\text{longitudinal strain, bending curvature})^T = \mathbf{B} U_E \quad (34)$$

$$S_h = (T_h, M_h)^T = (\text{membrane force, bending moment})^T \quad (35)$$

$$S_h = \mathbf{C} e_h \quad \text{with} \quad \mathbf{C} = \text{diag}(EA, EI) \quad (36)$$

when taking into account rotational inertia:

$$\mathbf{m} = \text{diag}(\rho A, \rho A, \rho I) \quad (37)$$

if rotational inertia is neglected, \mathbf{m} and U_h reduce to:

$$U_h = (u_h, v_h)^T \quad \text{and} \quad \mathbf{m} = \text{diag}(\rho A, \rho A) \quad (38)$$

We adopt this assumption in the following. It can easily be removed as done for plates. We have thus for FE mode i , in element E :

$$F_{Ihi} = \omega_{hi}^2 (u_{hi}, v_{hi})^T = \omega_{hi}^2 \rho A \mathbf{N} Y_{Ei} \quad (39)$$

$$Y_{Ei} = (u_{1i}, v_{1i}, \theta_{1i}, u_{2i}, v_{2i}, \theta_{2i})^T \quad (40)$$

For beams, the construction of a dynamically stress field $\hat{\mathbf{S}}$ becomes simple due to the fact that element edges Γ are reduced to the two beam sections which are connected to the nodes. In this case, a dynamically stress field is such that:

- 1) any node is in equilibrium with forces applied by connected beams,

2) in any element, \hat{S}_i satisfies the equilibrium equation (41) where D is the convenient differential operator,

$$D\hat{S}_i + F_{Ihi} = 0 \quad (41)$$

for the considered beam element, this can be written:

$$\hat{T}_i' = -\omega_{hi}^2 \rho A u_{hi} \quad (42)$$

$$\hat{M}_i'' = \omega_{hi}^2 \rho A v_{hi} \quad (43)$$

denoting by $\xi = x/L$ the reduced abscissa on E , L its length, integration leads to :

$$\hat{T}_i(x) = \hat{T}_i(\xi L) = -\omega_{hi}^2 \rho A \left(\left(\xi - \frac{\xi^2}{2} \right) u_{1hi} + \frac{\xi^2}{2} u_{2hi} \right) + \hat{T}_i(0) \quad (44)$$

$$\hat{M}_i(x) = \omega_{hi}^2 \rho A \hat{M}_C + \hat{M}_i'(0) x + \hat{M}_i(0) \quad (45)$$

$$\hat{M}_C(x) = \left(\frac{\xi^2}{2} - \frac{\xi^4}{4} + \frac{\xi^5}{10} \right) v_{1i} + L \left(\frac{\xi^3}{6} - \frac{\xi^4}{6} + \frac{\xi^5}{20} \right) q_{1i} + \left(\frac{\xi^4}{4} - \frac{\xi^5}{10} \right) v_{2i} + L \left(-\frac{\xi^4}{12} + \frac{\xi^5}{20} \right) q_{2i} \quad (46)$$

Quantities $\hat{T}_i(0)$, $\hat{M}_i'(0)$ and $\hat{M}_i(0)$ appearing in above expressions are quantities which are extracted from the FE results, as shown in next section.

5.4. Determination of \hat{S}_i in the case of beams

In the case of beams, determination of \hat{S}_i is a very particular case of the process used for other elements and presented in previous sections for plates. Let R_{Ei} be the generalised forces applied for mode i by the two nodes to which E is connected. Principle of virtual work applied to beam element E in the approximate displacements space leads to:

$$\forall [\eta], [\tilde{\eta}] = N V_{E'} - V_E^T \int_0^L N^T F_{Ii} dx + V_E^T \int_0^L B^T S_{hi} dx = V_E^T R_{Ei} \quad (47)$$

If N contains rigid body modes, and if a consistent mass matrix has been used, writing (47) for rigid body motions gives 3 equations which show that components of R_{Ei} are in equilibrium with the inertia force F_{Ii} . Moreover, the assembly of the FE equations (47) corresponds to write the equilibrium of each node. It must be emphasised that, in dynamics, this is only true if no reduction has been applied to the FE model.

This allow to take values from R_{Ei} to build \hat{S}_i because R_{Ei} insure equilibrium of nodes and are in equilibrium with inertia loads on each element.

$\hat{T}_i(0)$, $\hat{M}_i'(0)$ and $\hat{M}_i(0)$ are extracted from values associated to node 1 in R_{Ei} . Values $\hat{T}_i(L)$, $\hat{M}_i'(L)$ and $\hat{M}_i(L)$ are then compatible with corresponding values for node 2 in R_{Ei} due to their equilibrium properties. In the case of beams, the solution obtained by this approach is unique.

Here, the contribution of element E to the error through the constitutive

relation for mode i is chosen as :

$$r_{Ei} = \frac{\int_0^L \left((T_{hi} - \hat{T}_i)^2 / EA + (M_{hi} - \hat{M}_i)^2 / EI \right) dx}{\int_{\Omega} \left(T_{hi}^2 / EA + M_{hi}^2 / EI \right) dx} \quad (48)$$

5.5. Automatic refinement strategy for beams.

In the case of beams, an automatic strategy has effectively been implemented for 2D beams [22-23]. The user prescribes a tolerance α_{tol} and a frequency range $[0, f_{max}]$ in which he desires an accuracy α_{tol} in other words $\alpha < \alpha_{tol}$.

The choice of i_{max} cannot be done without some precision from the user, who must give information concerning density of modes beyond f_{max} . The fact that, for coarse meshes, some exact modes in $[0, f_{max}]$ are not visible is thus taken into account. A safety coefficient ζ and an additional number of modes to check i_{add} have to be given. Denoting i_1 the number of modes in $[0, f_{max}]$ and i_{ζ} the number of modes in $[0, \zeta f_{max}]$ for the FE model, i_{max} is defined by:

$$i_{max} = \max(i_1 + i_{add}, i_{\zeta}) \quad (49)$$

This definition of i_{max} results of observations made on numerical results when the current FE element model does not predict the good position of last modes in the range of interest: in this case, the presence of an important error estimation α_i for $i > i_1$ but close to i_1 was a good indication of such a situation. Automatic subdivision of elements is then based on a modified expression of r_{Ei} . In order to take into account very bad higher mode definitions for coarse meshes, and not get an overrefined new mesh in this case, a bound is also set to values leading to mesh generation in the new mesh definition.

5.6. Plate structures.

For plate structures, as before, the only difference with statics is that we have to consider inertia forces which are given by F_{hi} instead of applied loads. So a dynamically generalised stress field associated to mode i has to satisfy:

$$\begin{cases} \operatorname{div} \hat{T}_i + \omega_{hi}^2 \rho h (u_i e_1 + v_i e_2) = 0 \\ \operatorname{div} \hat{Q}_i + \omega_{hi}^2 \rho h w_i = 0 \\ \operatorname{div} \hat{M}_i - \hat{Q}_i + \omega_{hi}^2 \rho \frac{h^3}{12} (q_{1i} e_1 + q_{2i} e_2) = 0 \end{cases} \quad \text{or} \quad D \hat{S}_i + F_{hi} = 0 \quad (50)$$

By inspection of these equations, it can be seen that the minimum degree needed for generalised stress \hat{S}_i must be compatible with the choice of the weighting functions introduced in N for the element. For the element presented here, membrane components in \hat{T}_i must be at least quadratic functions over each sub-triangle. The same choice has to be done for shear components \hat{Q}_i , while for

bending components in \hat{M}_i cubic interpolations are needed. This choice allows local optimisations to be done for \hat{T}_i and \hat{Q}_i .

5.7. Modifications when a static condensation is done.

In the case where a reduction is made for the dynamic solution, the global column of dof q is reduced to a smaller one q_f , after the choice of a convenient set of selected dof. q is related to q_f by means of a transformation involving a known matrix G

$$q = G q_f \quad (51)$$

G is a constant matrix. Equilibrium of individual nodes no longer occurs. After reduction, we get:

$$M_{ff} \ddot{q}_f + K_{ff} q_f = Q_f \text{ where } M_{ff} = G^T M G, K = G^T K G \text{ and } Q_f = G^T Q \quad (52)$$

approximate modes satisfy

$$K_{ff} Y_{fi} = \omega_{ai}^2 M_{ff} Y_{fi}, \text{ and } Y_{ai} = G Y_{fi} \quad (53)$$

This is an application of the classical modal acceleration method. It must be noted that (53) does not imply (54)

$$K Y_{ai} = \omega_{ai}^2 M Y_{ai} \quad (54)$$

It is then sufficient, for each mode i for which indicator α_i is needed to solve a static problem on the whole uncondensed model, considering the structure submitted to inertia loads $\omega_{ai}^2 m U_{hi}$. We then get:

$$\hat{Y}_i = K^{-1} \omega_{ai}^2 m G Y_{fi} \quad (55)$$

Associated element loads \hat{R}_{Ei} have the properties of the R_{Ei} used when no condensation is made, which allows to construct, in the same way a dynamically stress field \hat{S}_i .

In the case where the condensation leads to a mode of poor quality, this can come from a too coarse mesh or from a bad choice of final dof q_f .

Estimation α_i measures simultaneously these two aspects. Introducing an indicator β_i :

$$\beta_i = (\hat{Y}_i - Y_{ai})^T K (\hat{Y}_i - Y_{ai}) / Y_{ai}^T K Y_{ai} \quad (56)$$

β_i allows to evaluate the error through the constitutive relation of Y_i obtained by reduction. If β_i is high, a better choice of the set of dof in q_f must be made, if not, the mesh as to be refined first.

6. NUMERICAL EXAMPLES

6.1. Circular plate with a hole submitted to a constant pressure

A holed circular plate is subjected to a shear load on the interior edge. The exterior edge is simply supported and clamped (figure 1). The error on constitutive relation and the error given by the Kirchhoff analytical solution are in good agreement (figures 2, 3)

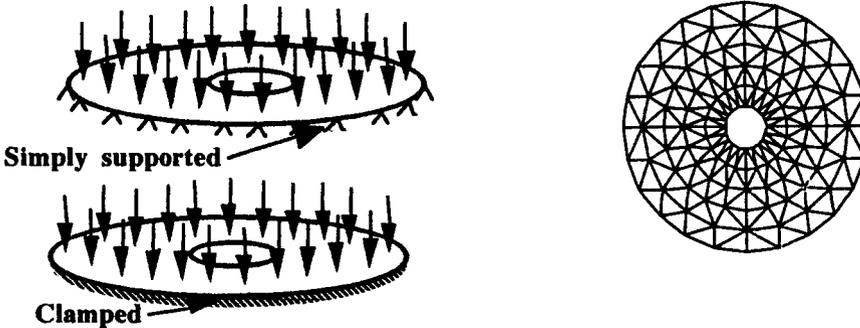


Figure 1. Circular plate under constant pressure

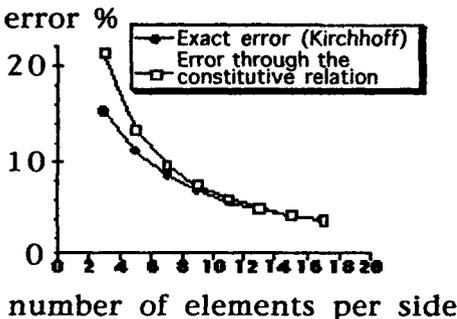


Figure 2. Simply supported circular plate under constant pressure; Error through the constitutive relation

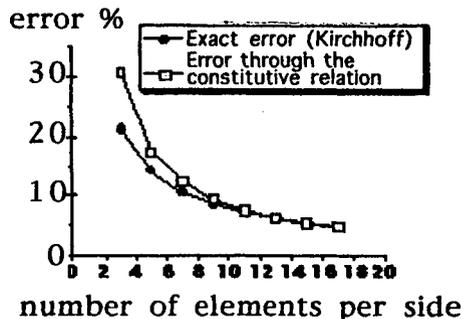


Figure 3. Clamped circular plate under constant pressure; Error through the constitutive relation

6.2. Square plate under constant pressure

A square plate with simply supported edges or clamped edges is submitted to a constant pressure (figure 4). The error through the constitutive relation is compared to the error given by the Kirchhoff analytical solution (figures 5, 6). Because the solution of this problem is more complicated, especially near the

corner of the plate, the number of elements needed to get a very efficient estimation is larger than for the previous examples. This example is the most important because two bending moments appear in principal directions of the plate. That is why it is a strong test for our error estimator.

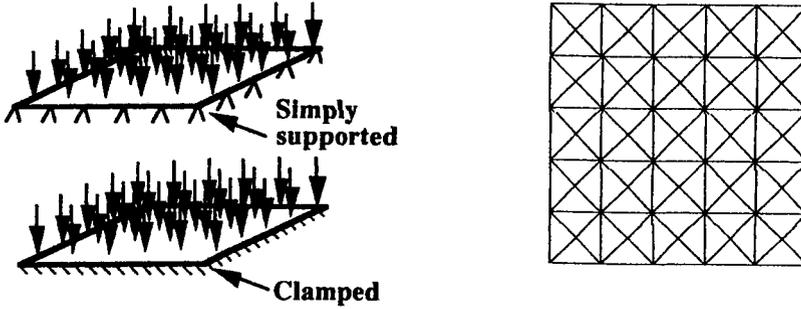


Figure 4. Square plate under constant pressure

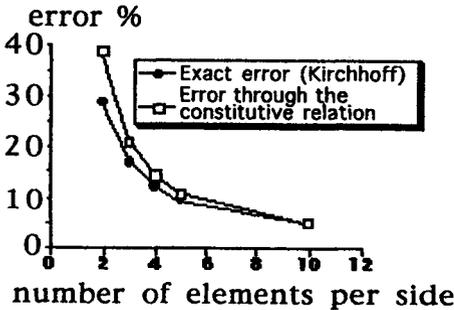


Figure 5. Simply supported square plate under constant pressure; Error through the constitutive relation

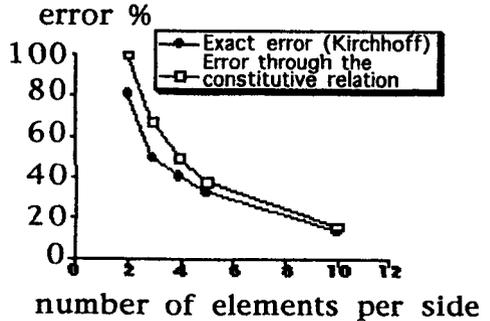


Figure 6. Clamped square plate under constant pressure; Error through the constitutive relation

6.3. Free vibrations of beams

The presented example concerns a plane, free-free, flexible frame. The adopted frequency range corresponds to $i_{max} = 45$ modes, and $\alpha_{tol} = \bar{\alpha} = 1\%$.

The minimal mesh h_1 shown figure 7 is used to start the automatic analysis (regular mesh, one element per beam). It can be seen, for this mesh, that despite the physical good looking of the second mode, it is completely wrong.

The asked accuracy is reached in two iterations. The second computation gives a 3.45% error and the indicator is inferior to 1% for the first 45 modes. The mesh c_2 used in this second computation is regular with 3 elements per beam ($c_2 = h_3$).

This is due to limitations which are introduced in the subdivision process, when the indication is too large. Information about too wrong modes is only partially taken into account. This leads to an error indication which can be too large during the first steps. The third computation (c3) leads to a 0.3% error widely inferior the asked accuracy with only 510 DOF due to the consideration of additional modes in the process for avoiding the missing of modes in the range.

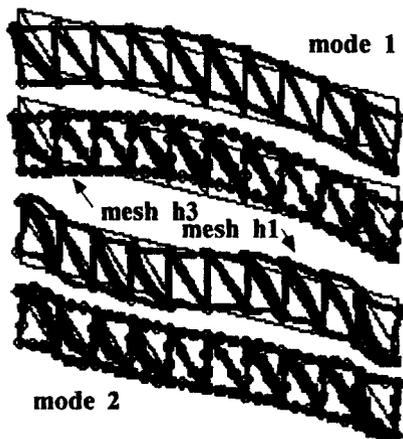


Figure 7. - First two elastic modes shapes.

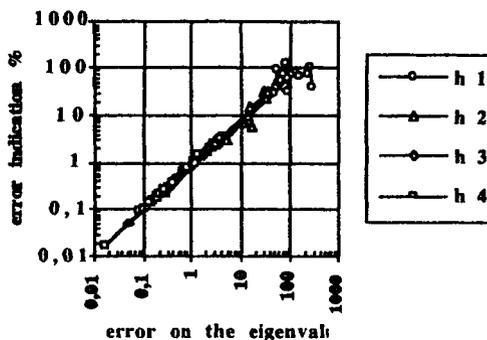


Figure 8. - Effectivity of the error estimator α_i with respect to the eigenvalue error

The automatically reached optimal mesh is as follows: 4 elements in the horizontal and vertical beams; diagonals: (6, 7, 6, 6, 7, 7, 6, 6, 7, 6). The effectivity of the error indicator is shown figure 8 for regular meshes h1, h2, h3, h4, by comparison with the error given using the best results of the analyses for the eigenvalue. A good agreement is shown especially for low errors and a very strong warning for errors larger than 10%. It must be noted that, when large errors are considered, we do not make the comparison between the exact mode and the closest approximated one: we compare the modes in sequence, which can explain the apparent discrepancy between the error indication and the exact one.

6.4. Free vibrations of a plate strip

The modal analysis of a plate strip is performed using different mesh refinements. In order to get a reference solution, the analysis has been performed with a very fine mesh and high order elements. For the first three modes (figure 9), we present the computed error α_i in comparison with the quasi exact one obtained from reference solution. Figure 10 to 12, show that this agreement is good. That is true for higher order modes too.

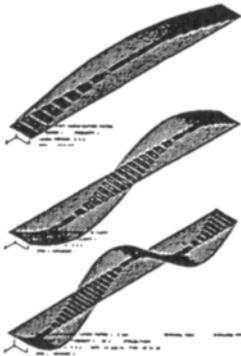


Figure 9. Three first mode shapes for a plate strip analysis with two simply supported edges

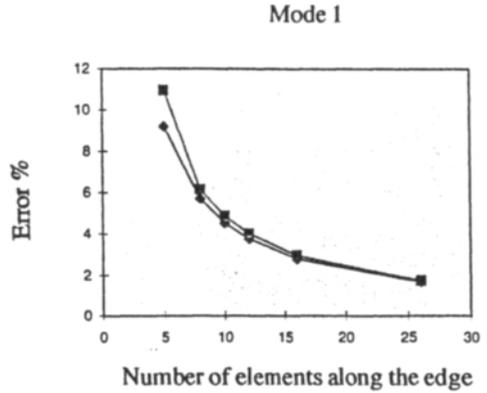


Figure 10. Plate strip with two simply supported edges; First mode; ■ Estimated error and error from the reference solution ◆

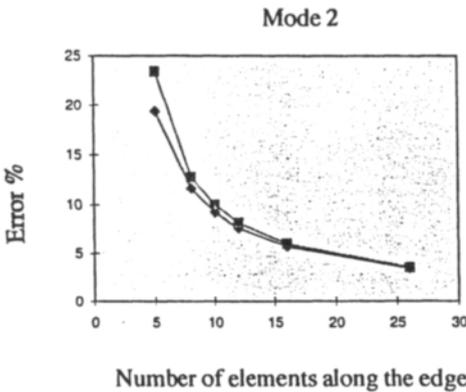


Figure 11. Plate strip with two simply supported edges; second mode; ■ Estimated error and error from the reference solution ◆

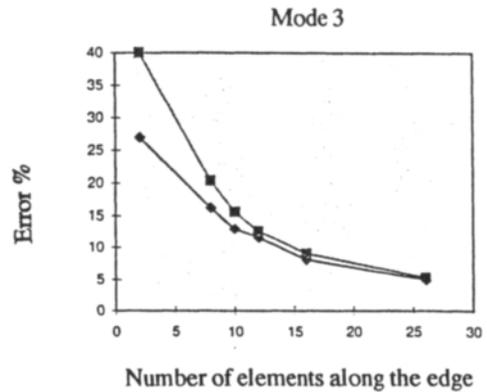


Figure 12. Plate strip with two simply supported edges; third mode; ■ Estimated error and error from the reference solution ◆

7. CONCLUSION

Extensions of the so called "error through the constitutive relation" error estimation method has been presented in case of beam and plate static and modal analyses. It is based on the building of statically or dynamically admissible stress resultant fields. In the case of plate analysis these fields are obtained in two stages. In a first one load and moment densities along the edges of the element, insuring

the equilibrium of each element are obtained as the sum of two densities rendering the result as close as possible to finite element results. Computations are local, at the node level or at the element edge level. Computation of the stress resultant fields within each element consider the solution in a large set and uses an error minimisation to keep the best one. Calculations are local there too.

The method is also presented in case of modal analyses of beam and plates. It has been shown on a structure made of beams that it is possible to define a mesh that insures a given accuracy for any mode within a frequency range.

A set of examples has been proposed for the different analyses under consideration where the error estimated has been compared to the error with respect to the exact solution. The examples show that the beam and plate analyses appears to be a field where the method is well adapted.

REFERENCES

1. O.C. Zienkiewicz and J.Z. Zhu, A simple error estimator and adaptative procedure for practical engineering analysis. *Int. J. Num. Meth. Eng.* **24**, 337-357 (1987).
2. O.C. Zienkiewicz and J.Z. Zhu, The superconvergent patch recovery and a posteriori error estimates. Part 1: The recovery technique; Part 2: Error estimates and adaptativity. *Int. J. Num. Meth. Eng.* **33**, 1331-1382 (1992).
3. N.E. Wiberg and F. Abdulwahab, A posteriori error estimation based on superconvergent derivatives and equilibrium. *Int. J. Num. Meth. Eng.* **36**, 2703-2724 (1993).
4. O.C. Zienkiewicz and J.Z. Zhu, The superconvergent patch recovery (SPR) and adaptative finite element refinement. *Comp. Meth. in Appl. Mech. Eng.* **101**, 207-224 (1992).
5. O.C. Zienkiewicz, J.Z. Zhu and J. Wu, Superconvergent patch recovery techniques. Some further tests. *Com. Num. Meth. Eng.* **9**, 251-258 (1993).
6. R. Rodriguez, Some remarks on Zienkiewicz-Zhu estimators *Int. J. Num. Meth. in PDEs* to appear.
7. I. Babushka and W.C. Reinboldt, Error estimates for adaptative finite element computations. *SIAM J. Num. Anal.* **15**, n°4, 736-754 (1978).
8. I. Babushka and W.C. Reinboldt, A posteriori error estimates for the finite element method. *Int. J. Num. Meth. Eng.* **12**, 1597-1659 (1978).
9. P. Ladevèze, Comparaison de modèles de milieux continus, *Thèse d'Etat, Université P. et M. Curie*. Paris, (1975).
10. P. Ladevèze and D. Leguillon, Error estimate procedure in the finite element method and applications. *SIAM J. Num. Anal.* **20**, n°3, 485-509 (1983).
11. P. Ladevèze, J.P. Pelle and P. Rougeot, Error estimation and mesh optimisation for classical finite elements. *Engr. Computations* **8**, 69-80 (1991).
12. P. Coorevits, P. Ladevèze and J.P. Pelle, Mesh optimisation for problems with steep gradients. *Engineering Computations*. **11**, 129-144 (1994).
13. P. Coorevits, P. Ladevèze and J.P. Pelle, An automatic procedure with a control of accuracy for finite element analysis in 2D elasticity, *Comp. Meth. in Appl.*

Mech. Eng. 121, 91-120 (1995).

14. P. Coorevits, J.P. Dumeau and J.P. Pelle, Analyse éléments finis adaptative pour les structures tridimensionnelles en élasticité, *Revue Européenne des éléments finis*, 5, n°3, 341-373 (1996).
15. P. Coorevits, J.P. Dumeau and J.P. Pelle, Adaptation et automatisation des maillages 3D, *Actes du troisième colloque national en calcul des structures*, Presses académiques de l'Ouest, 1, 367-372 (1997).
16. P. Ladevèze, G. Coffignal and J.P. Pelle, Accuracy of elastoplastic and dynamic analysis In *Babuska, I., Zienkiewicz, O.C., Cago, J.P. and Oliovera, ER de A. eds. Accuracy Estimates and Adaptativity for Finite Elements.*, Wiley, J. New-York. 181-203. (1986).
17. G. Coffignal, Fiabilité et optimisation des calculs élastoplastiques par éléments finis. *Thèse d'état . Université P.et M. Curie. Paris* (1987).
18. L. Gallimard, P. Ladevèze and J.P. Pelle, Error estimation and adaptativity in elastoplasticity, *Int. J. Num. Meth. Eng.* 39, 189-217 (1996).
19. J.P. Pelle and D. Ryckelynck, Sur une stratégie adaptative pour les calculs non linéaires; application aux calculs en viscoplasticité, *Actes du troisième colloque national en calcul des structures*, Presses académiques de l'Ouest, 1, 343-348 (1997).
20. P. Ladevèze and J.P. Pelle, Accuracy in finite element computations for eigen frequencies. *Int. J. Num. Meth. Eng.* 29, 1929-1949 (1989).
21. P. Bussy, Optimisation et fiabilité des calculs par éléments finis en non linéarité géométrique, *Thèse Docteur-Ingénieur , Université PARIS VI*, 1984.
22. G. Coffignal, Indicateur d'erreur pour les modes propres des modèles éléments finis, *Proceedings STRUCOME*, pp 227-241, ed. DATAID AS & I. Paris nov 91.
23. G. Coffignal, Indicateur d'erreur pour les modes propres des éléments finis de poutres, *Actes du premier colloque national en calcul des structures*, Hermès, 115-122 (Paris 1993).
24. P. Boisse and S. Perrin, Calcul de l'erreur en relation de comportement pour les modélisations par des éléments finis de plaques C°, C.R. *Académie des Sciences*, Paris, t.320, série II b, 289-294, (1995).
25. J.P. Pelle, Sur la maîtrise des calculs éléments finis: état actuel et perspectives, *Actes du deuxième colloque national en calcul des structures*, Hermès, 1, 43-54 (Paris 1995).
26. P. Boisse, Modèles mécaniques et numériques pour l'analyse non-linéaires des structures minces. *Mémoire d'habilitation à diriger des recherches*, Université de Franche Comté, (1994).
27. P. Boisse, J.L. Daniel and J.C. Gelin, A C⁰ three node shell finite element for non-linear analyses. *Int. J. Num. Meth. Eng.* 37, 2239-2364 (1994).
28. T. Belytschko, H. Stolarski and N. Carpenter, A C⁰ triangular plate element with one-point quadrature, *Int. J. Num. Meth. Engng.*, 20, 787-802.(1984).
29. V.B Watwood and B.J. Hartz , An equilibrium stress field model for finite element solutions of two dimensional elastoplastic problems, *Int. Jour. Solids Struct.*, vol 4, pp 857-873, 1968.

A posteriori error analysis for steady-state Maxwell's equations

L. Demkowicz

The Texas Institute for Computational and Applied Mathematics, The University of Texas at Austin, Taylor Hall 2.400, Austin, Texas 78712, USA

1. INTRODUCTION

This paper presents a simple version of the *Implicit Element Residual Method* for steady-state Maxwell's equations, and an *hp* FE approximation presented in [1, 2]. The idea originated from a series of papers by Babuška and Rheinboldt, see e.g. [3], and was formulated in [4, 5, 6]. The possibility of residual equilibration, using the technique of Ladevèze or Ainsworth and Oden (see [7, 8] and references therein) allows for the use of a local Neumann problem, significantly improving the quality of the method.

It turns out that the relaxed continuity assumptions for $\mathbf{H}(\text{curl})$ -conforming approximations (only the tangential component of the approximated vector field must be continuous) allow for a straightforward approximation of fluxes and residual equilibration for 2D Maxwell's equations. A generalization to 3D involves solution of "patch problems" as in [7, 8].

The plan of the contribution is as follows. We start with a short presentation of the *hp* FE method for Maxwell's equations. The idea of the error estimation is explained next, followed by a couple of simple numerical experiments. A comparison with the Laplace equation and 3D Maxwell's equations concludes the paper.

2. A ROBUST (STABLE) MIXED *hp* FE APPROXIMATION FOR MAXWELL'S EQUATIONS

We shall consider the following model problem. Given a *bounded* domain $\Omega \subset \mathbb{R}^2(\mathbb{R}^3)$ with boundary $\partial\Omega$ consisting of two disjoint parts Γ_1 and Γ_2 , we wish to determine electric

field $\mathbf{E}(\mathbf{x})$ satisfying the following boundary-value problem:

$$\begin{cases} \nabla \times (\nabla \times \mathbf{E}) - \omega^2 \mathbf{E} = \mathbf{J} & \text{in } \Omega \\ \mathbf{n} \times \mathbf{E} = \mathbf{0} & \text{on } \Gamma_1 \\ \mathbf{n} \times (\nabla \times \mathbf{E}) = \mathbf{0} & \text{on } \Gamma_2. \end{cases} \quad (2.1)$$

Here \mathbf{J} denotes a given field, related to an impressed current. For the sake of simplicity of the presentation, the material constants are normalized to unity, and all quantities are assumed to be real-valued. Multiplying the reduced wave equation by a test function, integrating over domain Ω , and integrating by parts, with the use of natural boundary condition (2.1)₃, we arrive at the standard variational formulation:

$$\begin{cases} \mathbf{E} \in \mathbf{W} \\ \int_{\Omega} (\nabla \times \mathbf{E}) \circ (\nabla \times \mathbf{F}) - \omega^2 \int_{\Omega} \mathbf{E} \circ \mathbf{F} = \int_{\Omega} \mathbf{J} \circ \mathbf{F} \quad \forall \mathbf{F} \in \mathbf{W} \end{cases} \quad (2.2)$$

where \mathbf{W} is the space of admissible solutions,

$$\mathbf{W} = \{\mathbf{E} \in L^2(\Omega) : \nabla \times \mathbf{E} \in L^2(\Omega), \quad \mathbf{n} \times \mathbf{E} = \mathbf{0} \text{ on } \Gamma_1\}. \quad (2.3)$$

Variational formulation (2.2) gives rise to standard finite element (FE) discretizations. Approximations of this type, in general, are unstable and lack convergence. One way to see the deficiency of formulation (2.2) is to introduce a spectral decomposition corresponding to the curl-curl operator present in the reduced wave equation,

$$\mathbf{E} = \mathbf{E}_0 + \sum_{i=1}^{\infty} E_i \mathbf{e}_i. \quad (2.4)$$

Here \mathbf{E}_0 denotes the curl-free component corresponding to the zero eigenvalue of the operator (the corresponding eigenspace is *infinite-dimensional*), and \mathbf{e}_i are eigenvectors with corresponding positive eigenvalues λ_i forming a sequence converging to infinity,

$$\nabla \times \mathbf{E}_0 = \mathbf{0}, \quad \nabla \times (\nabla \times \mathbf{e}_i) = \lambda_i \mathbf{e}_i, \quad \lambda_i \rightarrow \infty. \quad (2.5)$$

The eigenvectors and the eigenspace of curl-free components are L^2 -orthogonal. Setting $\mathbf{F} = \mathbf{E}_0$ in (2.2), we obtain the following stability estimate for the curl-free component:

$$\|\mathbf{E}_0\|_{L^2(\Omega)} \leq \omega^{-2} \|\mathbf{J}_0\|_{L^2(\Omega)} \quad (2.6)$$

where \mathbf{J}_0 denotes the curl-free component of the impressed current \mathbf{J} . In other words, the stability of \mathbf{E}_0 is controlled *only* by the zero-order term and it deteriorates as the frequency approaches zero.

A remedy to this problem was proposed by Kikuchi [10] and recently by Demkowicz and Vardapetyan [1, 2]. We begin by introducing a scalar space (of Lagrange multipliers):

$$V = \{q \in H^1(\Omega) : p = 0 \text{ on } \Gamma_1\}. \quad (2.7)$$

We shall assume that the following *compatibility condition* between spaces \mathbf{W} and V holds:

$$\mathbf{E} \in \mathbf{W}, \nabla \times \mathbf{E} = \mathbf{0} \Leftrightarrow \exists \phi \in V, \mathbf{E} = \nabla \phi \quad (2.8)$$

Upon substituting $\mathbf{F} = \nabla q$, for $q \in V$, we observe that the solution to variational problem (2.2) satisfies automatically the constraint:

$$\int_{\Omega} \mathbf{E} \circ \nabla q = \int_{\Omega} \mathbf{J} \circ \nabla q \quad \forall q \in V \quad (2.9)$$

The condition is equivalent to a divergence equation in Ω (continuity equation) and a boundary condition on Γ_2 . Thus, any solution to the original problem satisfies the extra equation and boundary condition. This, in general, *does not* hold at the discrete level. The idea now is to reimpose (2.9) as an additional equation at the cost of introducing a corresponding Lagrange multiplier p . We end up with the following *mixed variational formulation*:

$$\left\{ \begin{array}{l} \mathbf{E} \in \mathbf{W}, p \in V \\ \int_{\Omega} (\nabla \times \mathbf{E}) \circ (\nabla \times \mathbf{F}) - \omega^2 \int_{\Omega} (\mathbf{E} + \nabla p) \mathbf{F} = \int_{\Omega} \mathbf{J} \circ \mathbf{F} \quad \forall \mathbf{F} \in \mathbf{W} \\ \int_{\Omega} \mathbf{E} \circ \nabla q = \int_{\Omega} \mathbf{J} \circ \nabla q \quad \forall q \in V. \end{array} \right. \quad (2.10)$$

Two simple observations follow. First of all, the explicitly imposed constraint implies now a frequency-independent stability of the curl-free component.

$$\|\mathbf{E}_0\|_{L^2(\Omega)} \leq \|\mathbf{J}_0\|_{L^2(\Omega)} \quad (2.11)$$

Secondly, upon substituting $\mathbf{F} = \nabla q$, $q \in V$ in (2.10)₁ and using (2.10)₂, we obtain a weak form of the Laplace equation for the Lagrange multiplier p . Consequently, p must vanish.

Based on the compatibility condition (2.8), one can prove [1] the following lower bound for the inf-sup constant γ associated with the mixed variational formulation:

$$\max\left\{1 + \omega^2, \frac{1 + \lambda_i}{|\lambda_i - \omega^2|}, i = 1, \dots\right\} \leq \gamma. \quad (2.12)$$

We are ready now to introduce a stable FE discretization with a variable order of approximation. As an example, we shall present the concept for triangular elements only, although the idea holds for quads in 2D, and tetrahedra, bricks and prisms in 3D, as well.

The scalar element space of shape functions is identified as the space of polynomials of order $p + 1$ whose restrictions to element sides reduce to polynomials of *lower* order $p_1 + 1, p_2 + 1, p_3 + 1$, respectively. That is, we assume that

$$p \geq \max\{p_1, p_2, p_3\}. \quad (2.13)$$

The key in understanding the construction of the vector element lies in compatibility condition (2.8) that we postulate to hold for the discrete spaces as well. This implies that we *must* employ different orders of approximation for the tangential and normal components of the E -field. More precisely, the vector space of element shape functions consists of (vector-valued) polynomials of order p whose tangential components on the element sides reduce to polynomials of orders p_1, p_2, p_3 , respectively. One can prove [1, 2] that the FE discretization admits optimal convergence error estimates.

The elements are depicted in Fig. 1. We do not discuss here the choice of particu-

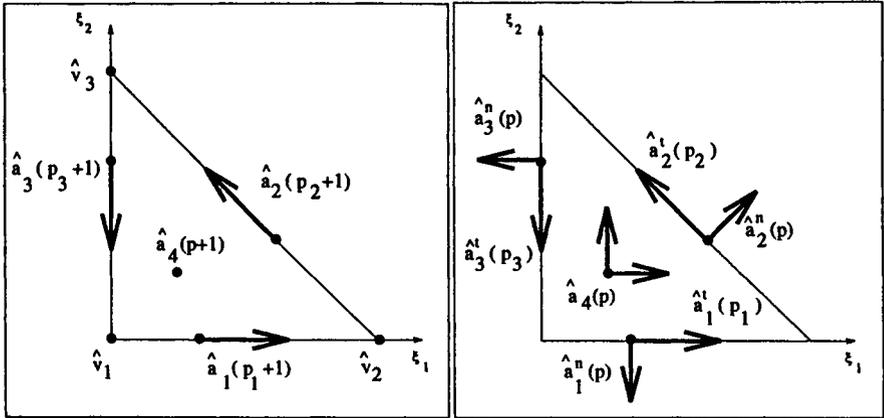


Figure 1: Scalar (left) and corresponding vector (right) triangular element with variable order of approximation

lar degrees of freedom which results in the definition of the corresponding scalar shape functions $\hat{\chi}_k(\xi)$ and vector shape functions $\hat{e}_k(\xi)$. Neglecting the round-off error, such a choice has no effect on our discussion. For details on a current implementation we refer to [11].

As a simple illustration, we present a solution to the model problem in a rectangular domain. The exact solution is a polynomial of order 5. Fig. 2 presents a sample mesh consisting of four linear and four quadratic elements with the corresponding contour plots of the first component of the electric field E . Notice the continuity of the tangential components only.

Finally, Fig. 3 displays typical h -convergence rates for the same rectangular domain problem, using uniform meshes of quadratic and cubic elements.

We also mention that the concept of the approximation generalizes in a standard way to curvilinear elements. With

$$\mathbf{x} = \mathbf{x}(\xi) \tag{2.14}$$

denoting the map from the master triangular element onto a curved element, the deformed

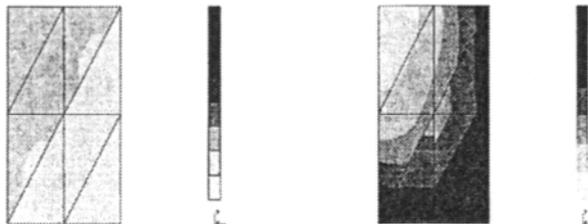


Figure 2: Solution of the model problem (a) an hp FE mesh of linear and quadratic elements (b) x component of the electric field

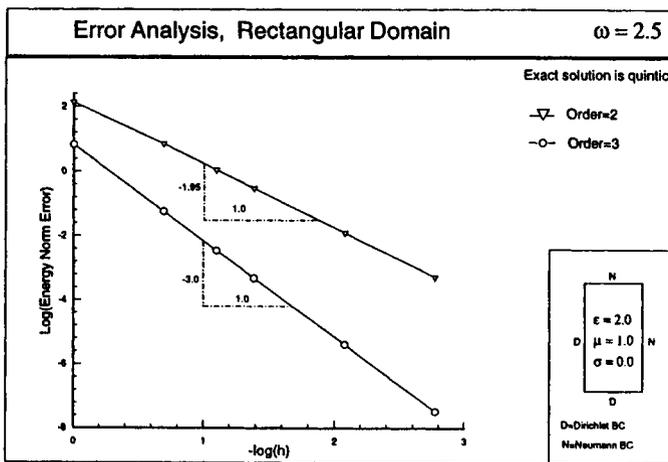


Figure 3: Experimental convergence rates for quadratic and cubic elements

element shape functions $\chi_k(\mathbf{x})$ and $\hat{e}_k(\mathbf{x})$ are defined as follows:

$$\chi_k(\mathbf{x}) = \hat{\chi}_k(\boldsymbol{\xi})$$

$$\mathbf{e}_k = (e_k^1, e_k^2), \quad e_k^j(\mathbf{x}) = \sum_{l=1}^2 \hat{e}_k^l(\boldsymbol{\xi}) \frac{\partial \xi_l}{\partial x_j}. \tag{2.15}$$

The definition above is a key to satisfy again the compatibility condition (2.8) and prove convergence with optimal rates.

REMARK 1 If we neglect the round off error, and assume that there is no quadrature error, then, due to the compatibility condition (2.8), the discrete Lagrange multiplier must vanish as well. Thus, for the proposed FE discretization, the solutions to discrete equivalents of (2.2) and (2.9) coincide with each other! From that point of view, the mixed formulation might be understood only as a key to constructing the stable element and proving convergence. In practice, however, we never integrate exactly. Another essential difference between the two discrete problems is the conditioning of the corresponding

stiffness matrices.

3. A POSTERIORI ERROR ESTIMATION FOR THE MAXWELL EQUATION IN 2D

The element implicit residual estimate

We recall first the idea of the error estimation using the standard abstract variational formulation:

$$\begin{cases} u \in V \\ b(u, v) = l(v) \quad \forall v \in V. \end{cases} \quad (3.16)$$

Introducing a finite element space $V_H \subset V$, we calculate the corresponding FE solution u_H by solving the approximate problem:

$$\begin{cases} u_H \in V_H \\ b(u_H, v_H) = l(v_H) \quad \forall v_H \in V_H. \end{cases} \quad (3.17)$$

Similarly, we introduce a *fine FE mesh space* V_h , and the corresponding fine mesh solution u_h .

The goal is now to estimate the fine mesh residual:

$$\sup_{\|v_h\|=1} \frac{|b(u_H, v_h) - l(v_h)|}{\|v_h\|} \quad (3.18)$$

where $\|\cdot\|$ is the energy norm:

$$\|u\|^2 = a(u, u). \quad (3.19)$$

Here $a(u, v)$ denotes a symmetric, V -coercive bilinear form, and $b(u, v)$ is the actual bilinear form (possibly neither symmetric nor coercive) corresponding to the problem of interest.

The methodology consists in splitting the residual into the corresponding contributions over elements K and introducing the *element flux functionals* λ_K

$$\begin{aligned} r(u_H; v_h) &:= b(u_H, v_h) - l(v_h) \\ &= \sum_K \{b_K(u_H, v_h) - l_K(v_h)\} \\ &= \sum_K \left\{ b_K(u_H, v_h) - l_K(v_h) - \underbrace{\lambda_K(v_h)}_{\text{element flux functional}} \right\} \end{aligned} \quad (3.20)$$

We shall postulate now two main assumptions:

- the element residuals are in equilibrium with respect to the coarse mesh:

$$r_K(u_H; v_H) := b_K(u_H, v_H) - l_K(v_H) - \lambda_K(v_H) = 0 \quad \forall v_H \in V_H(K); \quad (3.21)$$

- Consistency:

$$\sum_K \lambda_K(v_h) = 0 \quad v_h \in V_h. \quad (3.22)$$

Here $V_H(K)$ denotes the space of element K shape functions, possibly incorporating Dirichlet boundary conditions, if element K is adjacent to Γ_1 part of the boundary.

Introducing the local element Neumann problems:

$$\begin{cases} \phi_{h,K} \in V_h(K) \\ a_K(\phi_{h,K}, \psi_h) = r_K(u_H; \psi_h) \quad \forall \psi_h \in V_h(K), \end{cases} \quad (3.23)$$

we can conveniently now express the fine mesh residual in terms of the element error indicator functions $\phi_{h,K}$,

$$\begin{aligned} |r(u_H; v_h)| &= \left| \sum_K a_K(\phi_{h,K}, v_h) \right| \\ &\leq \left(\sum_K \|\phi_{h,K}\|_K^2 \right)^{\frac{1}{2}} \|v_h\|. \end{aligned} \quad (3.24)$$

This leads to the final estimate:

$$\|r\|_{V_h^*} \leq \left(\sum_K \|\phi_{h,K}\|_K^2 \right)^{\frac{1}{2}} \quad (3.25)$$

We make now the following observations.

- The local problems can be solved equivalently on the *bubble spaces* $V_{h,H}(K)$,

$$V_h(K) = V_H(K) \oplus V_{h,H}(K) \quad (a_K(\cdot, \cdot) - \text{orthogonal decomposition}). \quad (3.26)$$

- For $b(u, v) = a(u, v)$ (see e.g. [9]),

$$\|r\|_{V_h^*} = \|u_h - u_H\|. \quad (3.27)$$

In particular, if we identify the bilinear form a with the curl-curl form in the reduced wave equation (2.1), the proposed residual estimate should approximate well the fine mesh error for small frequency ω .

- The element flux functional λ_K is specified *uniquely* on the coarse element space $V_H(K)$.

The big question thus is:

How to extend functionals λ_K to the fine element space $V_h(K)$?

This is precisely the place where the information about the actual, continuous problem comes into the picture.

Determination of fluxes for 2D Maxwell's equations

The procedure will consist of three steps.

Step 1: We introduce the space of *coarse mesh traces of tangential components*:

$$Y_H(\partial K) = \{F_t|_{\partial K} : F \in W_H(K)\} \quad (3.28)$$

equipped with the $L^2(\partial K)$ product $(\cdot, \cdot)_{\partial K}$.

Step 2: The functional λ_K restricted to the coarse space $W_H(K)$ depends *only* upon the tangential components of test function F . Consequently, we can find its Riesz representation $f_{H,K}$ using the $L^2(\partial K)$ product,

$$\begin{cases} f_{H,K} \in Y_H(\partial K) \\ (f_{H,K}, F_t|_{\partial K}) = b_K(\mathbf{E}_H, F_H) - l_K(F_H) \quad \forall F_H \in W_H(K). \end{cases} \quad (3.29)$$

Step 3: We define the extension to the fine space W_h as:

$$\lambda_K(F_h) = (f_{H,K}, F_{ht})_{\partial K}.$$

A few important observations are in place.

- Due to a decoupling between the element edges, determination of fluxes f_H is done edge-wise, $f_{H,K} = \{f_{H,K}^e\}$, $e = 1, 2, 3$.
- For two adjacent elements K_1 and K_2 with a common edge e ,

$$f_{H,K_1}^e = -f_{H,K_2}^e. \quad (3.30)$$

This guarantees the consistency!

- In order to determine λ_K we need data from element K only !

Definition of the local problem for the Maxwell equations

We identify the bilinear form $a(\mathbf{E}, \mathbf{F})$ inducing the energy norm as:

$$a(\mathbf{E}, \mathbf{F}) = \int_{\Omega} (\nabla \times \mathbf{E}) \circ (\nabla \times \mathbf{F}) + \mathbf{E} \circ \mathbf{F} \quad (3.31)$$

and solve the local problems using the *mixed formulation*:

$$\begin{cases} \phi_{h,K} \in W_h(K), \quad p_{h,K} \in V_h(K) \\ \int_K (\nabla \times \phi_{h,K}) \circ (\nabla \times \psi_h) + \int_K (\nabla p_{h,K} + \psi_{h,K}) \circ \psi_h = r_K(\mathbf{E}_H, \psi_h) \quad \forall \psi_h \in W_h(K) \\ \int_K \psi_{h,K} \circ \nabla q_h = 0 \quad \forall q_h \in V_h(K) \end{cases} \quad (3.32)$$

where $r_K(\mathbf{E}_H, \psi_h)$ denotes the element K residual (3.21) including the element flux functional defined above. Imposing *local* divergence condition in the local problem should be interpreted as an extra step in extending the element flux functional $\lambda_K(\mathbf{F}_H)$ to the fine space, with a simultaneous enforcement of the orthogonality condition:

$$b_K(\mathbf{E}_H, \nabla q_h) - l_K(\nabla q_h) - \lambda_K(\nabla q_h) = 0 \quad \forall q_h \in V_h(K) \quad (3.33)$$

A numerical experiment

We illustrate the proposed technique with results of the a-posteriori error estimation for the rectangular domain problem, and the polynomial exact solution discussed earlier, for a small frequency $\omega = .01$. According to the remark above, for small ω , the fine mesh residual should coincide with the relative error, i.e. $\|u_H - u_h\|$. Neglecting the difference between the exact and fine mesh solutions, we will compare the error estimate directly with the actual error.

Table 1 compares the elemental contributions to the residual estimate, i.e.

$$\int_K \{(\nabla \times \phi_{h,K})^2 + \phi_{h,K}^2\} dK \quad (3.34)$$

with $\phi_{h,K}$ being the solution to the local problem, with the element contributions to the error measured in the energy norm, i.e.

$$\int_K \{(\nabla \times (\mathbf{E} - \mathbf{E}_H))^2 + (\mathbf{E} - \mathbf{E}_H)^2\} dK. \quad (3.35)$$

The comparison has been done for four meshes with order of approximation $p = 1, 2, 3$ and the mesh with linear and quadratic elements displayed in Fig. 2. The corresponding global estimates are compared with the global error at the bottom of the table. The global effectivity indices are consistently close to unity.

For an arbitrary frequency ω , the simple relation between the residual and the error is lost. Table 2 displays the comparison of the estimate of the residual with the error for a number of frequencies including those close to the first resonant frequency of the operator. To our best knowledge, the a-posteriori error estimation for "forced vibrations" problems remains an open research issue. However, we would like to emphasize that, in our opinion, the residual estimate may still be used as a meaningful basis for mesh adaptation.

4. A DISCUSSION FOR THE LAPLACE EQUATION, MAXWELL'S EQUATIONS IN 3D

The essential point in the construction of the error estimate for the 2D Maxwell equations is the fact that the interaction between neighboring elements takes place only through

Element number	mesh $p=1$ estimate/error	mesh $p=1/2$ estimate/error	mesh $p=2$ estimate/error	mesh $p=3$ estimate/error
1	1.94/2.00	1.94/1.95	.32E-2/.42E-2	.33E-3/.34E-3
2	1.41/1.49	1.41/1.46	.56E-2/.61E-2	.19E-3/.19E-3
3	1.55/1.73	1.55/1.66	.34E-2/.32E-2	.19E-3/.20E-3
4	0.53/0.62	0.53/0.60	.20E-1/.20E-1	.36E-4/.39E-4
5	2.06/2.19	.63E-2/.43E-1	.75E-2/.78E-2	.34E-3/.34E-3
6	1.76/1.90	.12E-1/.15E-1	.12E-1/.12E-1	.10E-3/.11E-3
7	0.62/0.64	.72E-2/.26E-1	.20E-1/.21E-1	.40E-4/.39E-4
8	2.37/2.30	.12E-1/.14E-1	.12E-1/.13E-1	.12E-3/.12E-3
Global	3.50/3.60	2.34/2.40	.294/.298	.371E-1/.374E-1
Eff. index	0.97	0.98	0.98	0.99

Table 1: Comparison of the a-posteriori error estimate with the error. Frequency $\omega = 0.01$

element edges. This manifests itself in the observation that there are no degrees-of-freedom to approximate electric field \mathbf{E} that would be associated with vertices (except for the approximation of geometry and the Lagrange multiplier, which is a different story). The degrees of freedom are either interior (local) to an element or correspond to element edges, hence the name of the *edge elements* is frequently used. The edge degrees of freedom are shared by *only two elements* and, consequently, the edge fluxes determined for two neighboring elements are identical (modulo sign) which implies the consistency condition.

Another explanation comes from the Lagrange multipliers method and the idea of the hybrid finite elements [12, 13]. Enforcing the continuity of the tangential component of the electric field through Lagrange multipliers (interpreted as the magnetic flux across the interelement boundaries), we end up with a method that is *exactly equivalent* to the approximation described here, provided the Lagrange multipliers (fluxes) are approximated with the same number of degrees of freedom as the tangential component of the \mathbf{E} -field shared by two adjacent elements. In other words, the electric fields obtained using either method are identical, and the described determination of fluxes can be interpreted as a reconstruction of the Lagrange multipliers.

The situation is more difficult in three dimensions where the \mathbf{E} -field degrees of freedom are associated with elements interiors, sides, and edges. The edge degrees of freedom are shared, in general, by more than two elements, and the reconstruction of fluxes is more difficult. Similarly as for the Laplace equation in 2D, we run into the problem of non-uniqueness of the representation for the Lagrange multipliers. As the methodology for both 2D Laplace and the 3D Maxwell equations is practically identical, we take the freedom of presenting the idea of the method using the 2D Laplace equation only. The resulting procedure reduces then just to a slight reinterpretation of the method of Ladev ze

frequency ω	residual estimate	error
1.00	3.32	3.66
1.82	29.8	133.5
1.825	102.8	427.5
1.8275	1101.6	4427.8
1.83	92.3	359.3
1.84	22.4	77.3
1.90	6.74	14.4
2.00	4.80	7.18
3.00	3.36	4.31
10.00	5.66	8.27

Table 2: Comparison of the residual estimate with the error for various frequencies ω

and Maunder [7].

For simplicity we shall restrict ourselves to a mesh of linear (Courant) triangles only. Let K be a typical element in the mesh with vertices $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$. The determination of the fluxes will be done in the following four steps.

Step 1: Representation of λ_K with vertex nodal forces (Dirac's deltas, see Fig. 4),

$$\lambda_K(v_H) = \sum_{i=1,2,3} P_{i,K} v_H(\underbrace{\mathbf{a}_i}_{i\text{-th vertex}}) \quad (4.36)$$

where

$$P_{i,K} = b_K(u_H, \underbrace{\chi_{i,K}}_{i\text{-th vertex shape function}}) - l_K(\chi_{i,K}) \quad (4.37)$$

The representation of the fluxes in terms of the Dirac functionals reflects the continuity of the restriction (trace) of FE solution to the element boundary. This representation, however, cannot be used to extend the flux functional to the fine space (or perhaps even all the way to the continuous space), as it has little to do with the actual fluxes (normal derivatives across the interelement boundaries) which are, in particular, discontinuous at the element vertices. So the task now is to switch from the Dirac deltas representation to a representation which would correspond to the actual fluxes and, *at the same time*, would be consistent.

Step 2: Representation of λ_K with edge nodal fluxes (Lagrange multipliers, see Fig. 5),

$$\lambda_K(v_H) = \sum_{e=1}^3 \pm \{R_{e1} v_H(\mathbf{a}_e) + R_{e2} v_H(\mathbf{a}_{e+1})\} \quad (4.38)$$

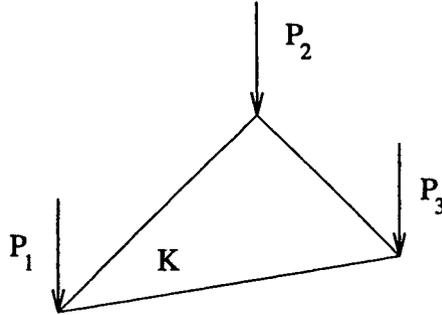


Figure 4: Flux representation in terms of "nodal forces"

Given a global orientation for all the edges in the mesh (e.g. determined by the vertex node numbers), we use in (4.38) the plus sign if the global and the local (counterclockwise) orientations of edge e are consistent, or the minus sign otherwise. The $+/-$ sign convention guarantees the consistency as the fluxes now are represented edge-wise.

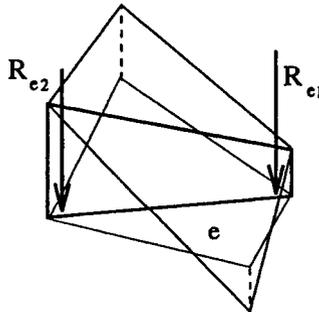


Figure 5: Flux representation on terms of "edge nodal forces"

Step3: Patch equilibrium.

Obviously, the two representations have to be identical. In particular, the two functionals have to take the same values for each nodal basis function. That leads to the following patch equilibrium problem:

$$R_i - R_{i-1} = P_i \quad i = 1, \dots, \text{number of elements in the patch} \quad (4.39)$$

Here, the "nodal forces" P_i are known from Step 1, and the "nodal edge forces" R_i are to be determined. Due to the global equilibrium equation corresponding to the considered node, system (4.39) *does have* a solution but the R_i 's can be determined only up to a (the node related) constant. An extra *normalizing condition* per node is needed! The situation is illustrated in Fig. 6.

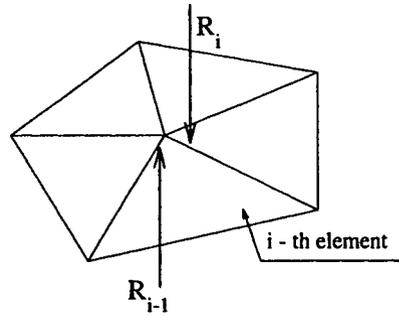
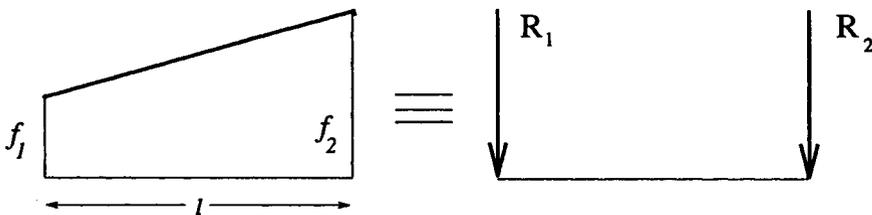


Figure 6: Nodal patch problem

Step 4: L^2 -edge representation of the Lagrange multipliers.

Finally, in the last step we need to switch to a continuous representation for the edge fluxes which will be more appropriate to define the extension to the fine space, see Fig. 7.

Figure 7: Switching to the L^2 representation of the edge fluxes

$$\begin{pmatrix} \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix} \quad (4.40)$$

Consequently,

$$\begin{aligned} f_1 &= \frac{4}{7}R_1 - \frac{2}{7}R_2 \\ f_2 &= -\frac{2}{7}R_1 + \frac{4}{7}R_2 \end{aligned} \quad (4.41)$$

The choice of an optimal normalizing condition, remains a separate issue [7].

REFERENCES

- [1] L. Demkowicz, L. Vardapetyan, "Modeling of Electromagnetic Absorption/Scattering Problems Using hp -adaptive Finite Elements", *Computer Methods in Applied Mechanics and Engineering*, **152**, 1-2, 103-124, 1998.

- [2] L. Vardapetyan, L. Demkowicz, "*hp*-Adaptive Finite Elements in Electromagnetics", *TICAM Report 97-18, Computer Methods in Applied Mechanics and Engineering*, in review.
- [3] I. Babuška, W.C. Rheinboldt, "Error Estimates for Adaptive Finite Element Computations", *SIAM J. Numer. Anal.*, **18**, 736-754, 1978.
- [4] L. Demkowicz, J.T. Oden and T. Strouboulis, "Adaptive Finite Elements for Flow Problems with Moving Boundaries. Part 1: Variational Principles and A-Posteriori Error Estimates", *Computer Methods in Applied Mechanics and Engineering*, **46**, 217-151, 1984.
- [5] R.E. Bank, A. Weiser, "Some A-Posteriori Error Estimators for Elliptic Partial Differential Equations", *Math. Comput.*, **44**, 283-301, 1985.
- [6] J.T. Oden, L. Demkowicz, T. Strouboulis and Ph. Devloo, "Adaptive Methods for Problems in Solid and Fluid Mechanics", in *Accuracy Estimates and Adaptive Refinements in Finite Element Computations*, I. Babuška et al., eds., 249-280, Wiley, New York 1986.
- [7] P. Ladevéze, E.A.W. Maunder, "A General Method for Recovering Equilibrating Element Traction", *Computer Methods in Applied Mechanics and Engineering*, **137**, 111-151, 1996.
- [8] M. Ainsworth, J.T. Oden, "A Posteriori Error Estimation in Finite Element Analysis", *Computer Methods in Applied Mechanics and Engineering*, **142**, 1-88, 1997.
- [9] J. T. Oden, L. Demkowicz, W. Rachowicz and T. A. Westermann, "Towards a Universal *h-p* Adaptive Finite Element Strategy, Part 2. A Posteriori Error Estimation," *Comp. Meth. Appl. Meth. Eng.*, **77**, 113-180, 1989.
- [10] F.Kikuchi, "Mixed and Penalty Formulations for Finite Element Analysis of an Eigenvalue Problem in Electromagnetism", *Computer Methods in Applied Mechanics and Engineering*, **64**, 509-521, 1987.
- [11] W. Rachowicz, L. Demkowicz, "2Dhp90-EM, a Two-Dimensional *hp*-Adaptive FE Package for Electromagnetics", *TICAM Report*, in preparation.
- [12] Nicholas Moes, *Private communication*, June 1997.
- [13] P.A. Raviart, J.M. Thomas, "Primal Hybrid Finite Element Methods for 2nd Order Elliptic Equations", *Mathematics of Computation*, **31**, 138, 391-413, 1977.

AUTHOR INDEX

Ainsworth, M.	101
Almeida Pereira, O.J.B.	427
Arnold, A.	101
Aubry, D.	325
Babuska, I.	155
Barthold, F.J.	81
Beckers, P.	413
Belytschko, T.	217
Boisse, Ph.	493
Boroomand, B.	3, 383
Bouillard, Ph.	477
Chenot, J.L.	365
Coffignal, G.	493
Collot, M.	65
Coorevits, P.	443
Copps, K.	155
Coupez, T.	365
Cramer, H.	293
Datta, D.K.	155
Defeü, E.	413
Demkowicz, L.	513
Destuynder, P.	65
Diez, P.	333
Dumeau, J.P.	443
Dutko, M.	349
Fourment, L.	365
Gangaraj, S.K.	155
Garcia, J.	309
Hager, P.	461
Huerta, A.	333
Idelsohn, S.	309
Ihlenburg, F.	477
Ladevèze, P.	135, 231
Liu, W.K.	217
Lucas, D.	325
Maunder, E.A.W.	427

Moës, N.	115, 231
Moitinho De Almeida, J.P.	427
Oden, J.T.	43, 115
Ohnimus, S.	81
Oñate, E.	309
Owen, D.R.J.	349
Patera, A.T.	199
Pelle, J.P.	443
Peraire, J.	199
Perić, D.	349
Pijaudier-Cabot, G.	333
Prudhomme, S.	43
Rannacher, R.	275
Rodriguez-Ferran, A.	333
Salaün, M.	65
Schmidt, M.	81
Singer, M.	217
Stein, E.	81
Steinl, G.	293
Strouboulis, T.	155
Suttmeier, F.T.	275
Tie, B.	325
Vemaganti, K.	115
Verfürth, R.	257
Wiberg, N.E.	461
Wunderlich, W.	293
Zhang, Z.	25
Zhu, J.Z.	3, 25
Zienkiewicz, O.C.	3, 383