Error Estimation and Adaptive Spatial Discretisation for Quasi-Brittle Failure

Tanyada Pannachet

Error Estimation and
Adaptive Spatial Discretisation
for Quasi-Brittle Failure

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. J. T. Fokkema,
voorzitter van het College van Promoties,
in het openbaar te verdedigen op donderdag 19 oktober 2006 om 10.00 uur
door

Tanyada PANNACHET

Master of Engineering, Asian Institute of Technology
geboren te Khon Kaen, Thailand

Dit proefschrift is goedgekeurd door de promotoren:

Prof. dr. ir. L. J. Sluys
Prof. dr. ir. H. Askes

Samenstelling promotiecommissie:

Rector Magnificus	Voorzitter
Prof. dr. ir. L. J. Sluys	Technische Universiteit Delft, promotor
Prof. dr. ir. H. Askes	University of Sheffield, The United Kingdom, promotor
Prof. dr. K. Runesson	Chalmers Tekniska Högskola, Sweden
Prof. dr. ir. A. van Keulen	Technische Universiteit Delft
Dr. P. Díez	Universitat Politècnica de Catalunya, Spain
Dr. ir. R. H. J. Peerlings	Technische Universiteit Eindhoven
Dr. G. N. Wells	Technische Universiteit Delft
Prof. dr. ir. J. G. Rots	Technische Universiteit Delft , reservelid

Contents

Overview

The finite element method is a numerical tool to approximate solutions to partial differential equations, for instance those describing physical phenomena in engineering. Accuracy of a finite element solution depends mainly on the discretisation of the problem domain. Certainly, a more refined/enriched discretisation improves the ability of the finite element analysis to approximate the exact solution.

However, some questions arise, for example, whether the mesh used in the computation is good enough to output an acceptably accurate result and, if not, how fine it should be. Using a finer mesh also means an increased number of unknowns that must be solved in the finite element computation. And, even though the capability of computers nowadays is much improved, the numerical models are also becoming more complicated as the knowledge about the physical phenomena has become much clearer than in the past.

The measurement of error information is the basis for an answer to the above questions. Error information is an objective measure to assess whether the used finite element mesh is of sufficient quality. Moreover, local error information and the corresponding local criteria give the user some hints where in the mesh the discretisation should be improved. This procedure of discretisation improvement is known as *mesh adaptivity*. It can enhance the efficiency of the discretisation enormously, especially in problems whose solutions need very fine discretisation only in a small part, whereas coarse discretisation may be applied in the rest of the problem domain. A typical example of such a problem, to which this dissertation is devoted, is the analysis of cracks. For quasi-brittle materials, cracks constitute small zones where the mechanical nonlinear activity is concentrated, while the rest of the structure behaves elastically. The cracking zones, normally not known *a priori*, require a fine discretisation whereas the remainder of the structure can be analysed with a coarser discretisation. Thus, crack analysis can benefit from mesh adaptivity.

The aim of this chapter is to give a brief introduction to the whole dissertation. We will start with defining three levels of problems in preparation for the finite element

analysis as well as the corresponding errors that emerge during transitions from one level to another. Next, as the finite element solution relies essentially on how the problem domain is discretised, remarks about mesh discretisation in finite element analysis will be addressed. Essentials about mesh adaptivity and error estimation, as well as its applications in crack modelling, will end this chapter.

1.1 Physical, model and discretised problems

Generally, there are three defined problems in numerical computation. In practice, the *physical problem* to analyse must be defined as the first step. Due to the complexity of the real physical problem, normally some assumptions are made. These assumptions may be, for instance, a 2D representation of the real 3D problem being under plane stress/plane strain conditions with the assumed material behaviour during a loading process described by a certain constitutive relation. With those assumptions, the physical problem is now transformed into the *model problem*. In finite element modelling, the problem domain must then be discretised so that it can be analysed numerically. At this stage, the problem becomes the *discretised problem*. The boundary conditions are projected to the discretised domain and the forces are distributed corresponding to the discretisation, resulting in so-called consistent nodal forces.

Progressing from one problem to another leads to different types of error. The assumptions made in the model problem to represent the physical problem cause the so-called *modelling error*, while the mapping of the model into the discretised domain brings about the *discretisation error*. While the modelling error indicates how accurate the mathematical model is in representing the real physical problem, the discretisation error indicates how accurate the discretisation is in approximating the solution to the mathematical model[*]. While the modelling error is measured by comparing the mathematical model (model problem) with the experimental data (physical problem), the discretisation error can be estimated by comparing the solutions of the discretised problem with those of the model problem represented by a very refined/enriched discretisation[†]. Even so, in real practice, it is not simple at all to distinguish between the modelling error and the discretisation error since the answer to the constitutive relation (model problem) can generally not be determined analytically but only numerically. And, via the finite element concept, the discretisation of the model problem is unavoidable, whereby it becomes impossible to separate modelling errors from discretisation errors. However, this constitutes a dilemma in the transition from physical problems to model problems. In this dissertation, we are concerned with the transition from model problems to discretised problems.

---

[*]Here, we denote the solution to the mathematical model as the *exact* solution of the model problem.
[†]A *refined* discretisation is defined as a discretisation with an improvement regarding element sizes, whereas an *enriched* discretisation denotes a discretisation with an improvement regarding interpolation capability.

In particular, our main goal is to assess the error in finite element discretisation. Thus, one of our assumptions is that the constitutive relations of material models used throughout this thesis are perfectly correct, i.e. they are a perfect representation of the underlying physical processes. The discretisation error, resulting from the projection of the model quantities to the discretised domain, originates from two sources, namely the inability to reproduce the geometric boundary of the model problem and the inability to reproduce the exact solution of the model problem. The error from the first source is actually a source of error that, for not too complicated boundary geometries, is avoidable by carefully selecting suitable type of finite elements. Thus, our main focus will be on the second source of discretisation error.

1.2 Quality of a finite element mesh

As mentioned earlier, accuracy of the finite element solution depends on how the model problem is discretised. Two main factors of the standard finite element discretisation are

- size of the finite elements (the h-factor), and

- characteristic of the interpolation functions, for instance the polynomial order (the p-factor).

Obviously, a smaller element size may provide a better resolution of the exact solution. However, the approximation also depends on how suited the interpolation function, often based on piecewise polynomials, is for describing the exact solution.

Figure 1.1 shows how the finite element analysis approximates the exact solution of an ordinary differential equation, which here is a quartic polynomial. Keeping the interpolation function in linear form, a better resolution to the exact solution can be obtained via the uniform refinement of the finite element mesh, the so-called h-version finite element method. Each smaller element has to model a smaller segment of the exact solution. On the other hand, in the p-version finite element framework, the approximation is improved by enrichment of the interpolation functions. Without changing the element size, the resolution of the exact solution is well-fitted, especially when the order of interpolation polynomial approaches that of the analytical solution. Another important observation from this problem is that, with the same number of degrees of freedom, the p-version provides a better approximation to the exact solution than the h-version. This holds in particular for higher values of the interpolation orders.

In this research, we focus on the p-version (cf. Chapter 2) as well as the h-version finite element methods. Although not as popular, the p-version has some outstanding advantages. Firstly, it provides accuracy improvement without changing the
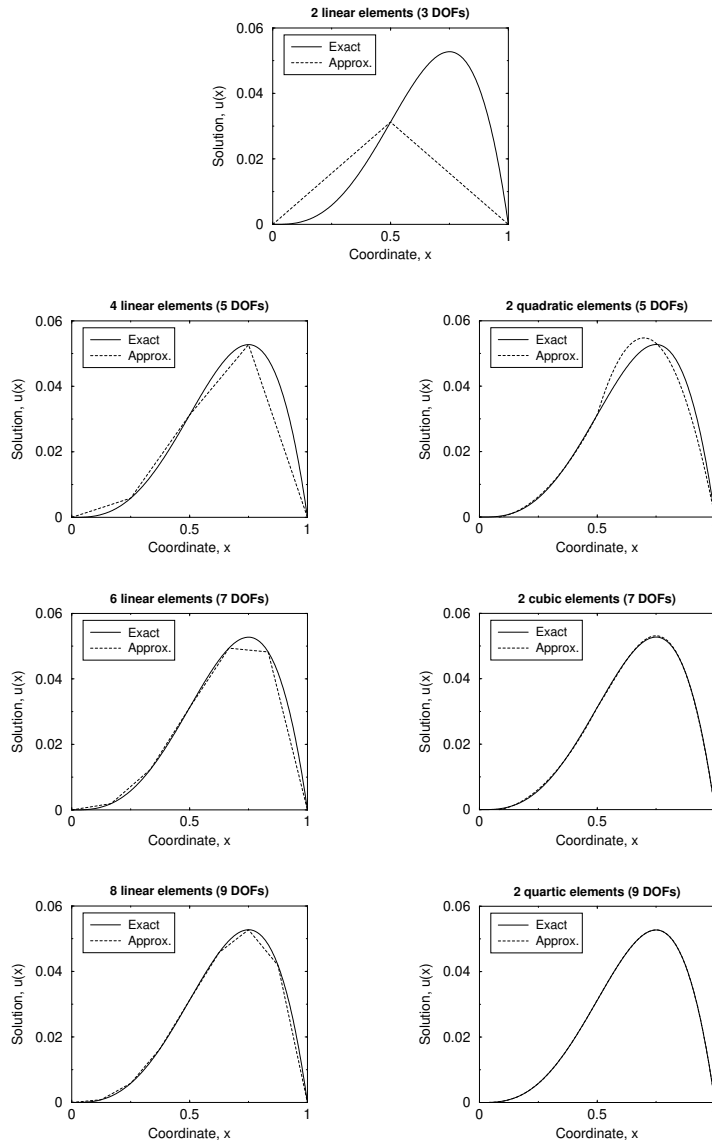
Figure 1.1 Comparison between *h*-extension (left column) and *p*-extension (right column) of the problem $-\frac{\partial^2 u}{\partial x^2} = 6x^2 - 3x$, with boundary conditions $u(0) = 0$ and $u(1) = 0$.

mesh configuration[‡]. Secondly, for problems with smooth solutions, the p-version provides a higher rate of convergence [93], i.e. the approximate solution becomes more accurate by increasing the polynomial degree than by adding the same number of degrees of freedom via the h-version. Thirdly, when the *hierarchical p*-version (for example, [93,94]) is employed, each additional higher-order contribution does not change any of the interpolation functions used in the previous contribution. As such, the stiffness matrix for order p is embedded in the stiffness matrix for order $p + 1$, reducing computational effort and improving the conditioning of the stiffness matrix.

1.3 Error control and mesh adaptivity

Due to limitation of computer capacity, not all information describing the actual continuum model can be included in the finite element computation. And even though a more refined/enriched discretisation is a better representation of the continuum model, it requires higher computational cost accordingly. As a solution to this problem, one should set a balance between accuracy and computational cost. An acceptably accurate solution that does not require outrageous computation should be the rule for practical applications.

Types of error assessment

To measure the accuracy of the finite element solution, it is necessary to assess an error quantity, which results from the finite element discretisation. Basically, there are two types of error estimation procedures available, namely *a priori* and *a posteriori* error estimators. The *a priori* estimate provides general information on the asymptotic behaviour of the discretisation errors but is not designed to give an actual error estimate for a specific given mesh, geometry and loading conditions. On the other hand, the *a posteriori* estimate measures the actual error at the end of a specific computation and can be exploited to drive a subsequent mesh adaptivity procedure.

In this context, following [43], we distinguish between *error estimation* and *error indication* based on objectivity of the output quantity. The error indication does not provide objective information about the exact error, but gives some hints where the solution may need a more refined/enriched discretisation. Based on heuristic observations, we can actually predict in which regions of the problem domain errors are likely to occur based on the problem geometry and the solution itself. For example, errors are always concentrated at sharp corners of the problem domain, where point loads are prescribed, and where there is an abrupt change in boundary conditions; in other words, errors concentrate where high gradients of the solution occur.

[‡]The geometric representation of a problem with complex geometry may change slightly the mesh configuration during the p-extension. However, this type of problems will not be studied in this thesis.

Original

Mesh enrichment
(p–adaptivity)

Mesh refinement
(h–adaptivity)

Mesh gradation
(r–adaptivity)
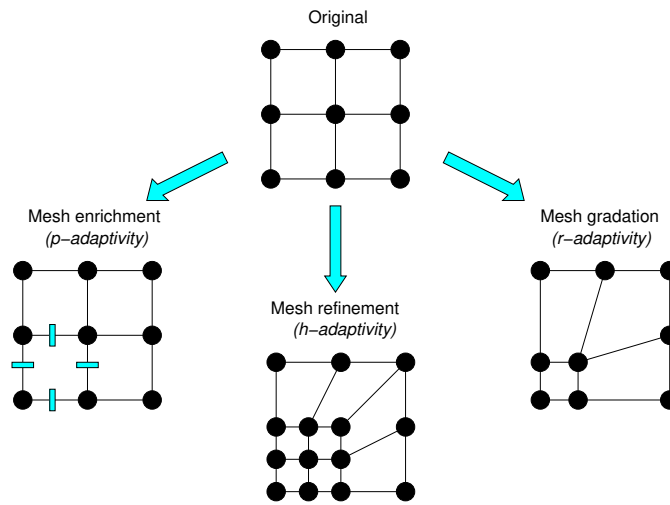
Figure 1.2 Some mesh adaptive schemes used in this research.

However, as the error indication directly links available quantities to error information, it needs to be derived for each material model and is rather restricted to the assumptions based on types of the problem to be analysed. In contrast, the standardised and mathematically founded error estimation can be applied to every problem for any material model without any (major) reformulation. In spite of being computationally more expensive than the indication, the error information obtained from the error estimation is objective and can be exploited with *optimality criteria* in designing an optimal mesh. We employ, by such reasons, an *error estimator* in this research study.

Error estimators

Basically, the *a posteriori* error estimators can be categorised into two main classes. The *recovery-type* error estimators (for example, [106,107]) measure the smoothness of stresses between adjacent elements. Since the methods do not require solving the error equations, they are simple and more preferable in many practical problems. However, there are not so many cases reported in [106,107] that show superconvergence[§]. On anisotropic meshes or those with mixed element meshes, the analysis is hindered by an apparent lack of superconvergence properties. Also the recovery-type estimator is not proven to converge in nonlinear problems.

In contrast, the *residual-type* estimators, although related somehow to the recovery-type [103], do not depend on the superconvergence properties. Thus, they

---

[§]Superconvergence property belongs to some points where a very accurate solution can be obtained. They are usually the quadrature points [105].
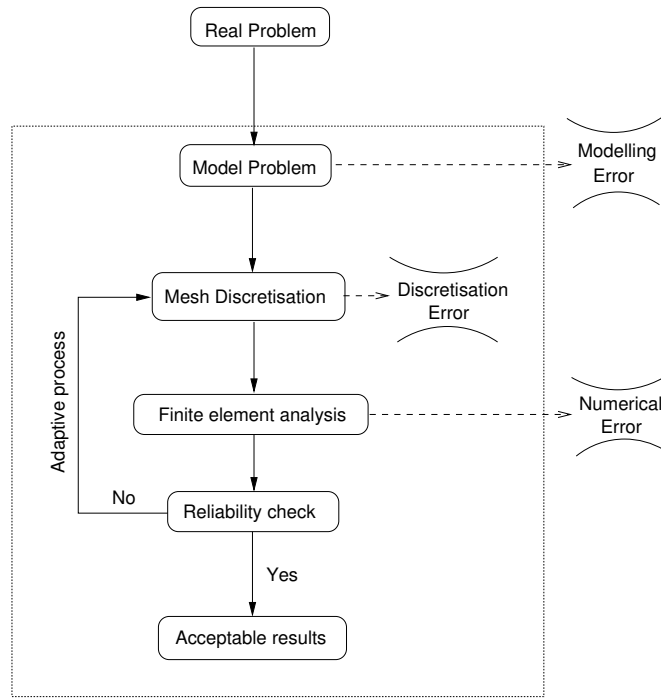
Figure 1.3 Standard procedure for adaptive finite element computation.

can be applied to a wider variety of problems. The methods (for example, [12]) determine the error by calculating the residual of the finite element solutions in each local space. We have chosen a residual-type error estimator in this study. Following the idea in [29], homogeneous Dirichlet conditions are imposed in the error equation defined by forming patches of several elements. The method is applied for estimating the error in energy norm (cf. Chapter 3), as well as the error in a local quantity of interest (cf. Chapter 4).

Mesh adaptivity

Once the error information is obtained, the finite element mesh can be adapted accordingly. The mesh should be improved where the local error exceeds the acceptable limit (controlled by *refinement criteria* – also known as *adaptive criteria*). There are many techniques for local mesh improvement, for instance, *mesh refinement* (*h*-adaptivity), *mesh enrichment* (*p*-adaptivity), *mesh gradation* (*r*-adaptivity), *mesh superposition* (*s*-adaptivity) or combinations of any two.

In this dissertation, we consider only three adaptive techniques (cf. Figure 1.2). By applying *h*-adaptivity, it is possible to design an optimal mesh based on an *op-*

timality criterion, which is formulated from the *a priori* convergence assumption (cf. Chapter 5). On the other hand, finding a precise balance between acceptable error levels and computational costs via *p*-adaptivity may imply that *fractional* polynomial degrees must be used, which is not a feasible option. Hence, in *p*-adaptivity, the interpolation is enriched hierarchically by one order at a time. Without adding any extra degrees of freedom, *r*-adaptivity can be a compromising alternative to *h*-adaptivity. And without solving any equation, a smoothing based *r*-adaptive technique based on the weighted Laplace smoothing is introduced and investigated in Chapter 5.

Figure 1.3 shows the standard adaptive procedure in the finite element analysis. In the same figure, the dashed box roughly indicates the scope of this research, wherein the discretisation error is the only error under consideration. Although it is difficult to neglect involvement of the numerical error (e.g. floating point error) in this study, its contribution is assumed to be very marginal as compared to the discretisation error. All detailed information about mesh adaptive aspects, including the *transfer of state variables* for nonlinear analysis, is addressed in Chapter 5.

1.4 Adaptive modelling of quasi-brittle failure

In this dissertation, error estimation and mesh adaptivity are applied to problems involving stationary and propagating cracks. The focus is on materials such as concrete, rock, ceramics and some matrix composites, which show so-called *quasi-brittle* behaviour. Unlike perfectly brittle materials, quasi-brittle materials do not lose their entire strength immediately after the maximum strength is exceeded but instead gradually lose their material strength and show the so-called *strain-softening* phenomenon (cf. Figure 1.4). Softening stress-strain relations show a drop of stress after the applied load exceeds the material strength (peak point). In fact, microcracks are initiated in the material before the stress in the material reaches its maximum strength [89]. However, the material is still able to carry loads to an extent. Upon further loading, these microcracks will then join together to form a dominant crack line which will lead to failure of the specimen.

Another phenomenon that occurs during the fracture process is *strain localisation* (cf. Figure 1.5). When the material loses its ability to carry load, the affected part shows increasing displacement gradients. Ultimately, when a complete rupture has occurred and the material is separated into distinct pieces, the displacement gradient has transformed into a displacement jump.

Basically, there are two main assumptions to model the fracture mechanism occurring in these quasi-brittle materials. The first class consists of *continuous crack models*, in which the material deterioration is accounted for in a *smeared* way. The stress field and strain field remain continuous during the entire fracture process resulting from a gradual degradation in material properties.

Discontinuous crack models can be regarded as the second class. In these models, the failure mechanism is presented by means of geometrical discontinuities in the
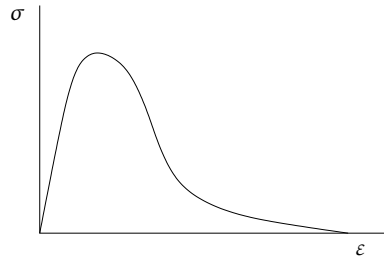
Figure 1.4 Softening phenomenon.

material domain. Cracking takes place when stresses in the materials, in any direction, exceed the maximum quantity that the material can resist in that direction. Such discontinuities imply that materials have separated parts and a jump in the displacement field can be found in the zone where discontinuities exist. Figure 1.5 shows the difference between crack representations of the two assumptions in the context of a three-point bending test.

In standard finite element computation, in order to deal with complicated material models, the finite element mesh must be properly designed *a priori*. Such mesh design has to rely on information before the computation. The mesh may be designed based on information such as the regions wherein the stresses may concentrate or where the material/geometrical imperfections are. These guidelines are not always obvious in practice and the designed mesh does not always guarantee appropriate results during cracking processes. Apparently, the need of mesh adaptivity becomes of great importance in crack propagation analyses.

1.4.1 The continuous crack model

Continuous crack models can be implemented using either the concept of *plasticity* or the concept of *continuum damage mechanics*. In this research, the *gradient-enhanced damage model* [70] is used for mesh adaptivity in a continuous crack concept. Damage occurs in the part of material domain where the stress cannot be sustained fully anymore. As a regularised continuum, the gradient-enhanced damage model converges properly upon refinement of the finite element discretisation.

Error estimation, as well as error indication, has been applied in problems with softening phenomena. Some outstanding works, employing the residual-type error estimation and *h*-adaptivity in softening media such as viscoplastic or nonlocal damage models, can be found in [28,78]. In these works, the error estimation takes place at the end of the analysis. Thus, the mesh is adapted based on the final error distribution. This refined mesh is then used to restart the whole analysis from scratch. As the error is not measured *during* computation, there is a possibility that the failure mechanism obtained is incorrect. It was shown in [6] that crack paths may be different for diffent meshes and the adaptive process must be updated dur-
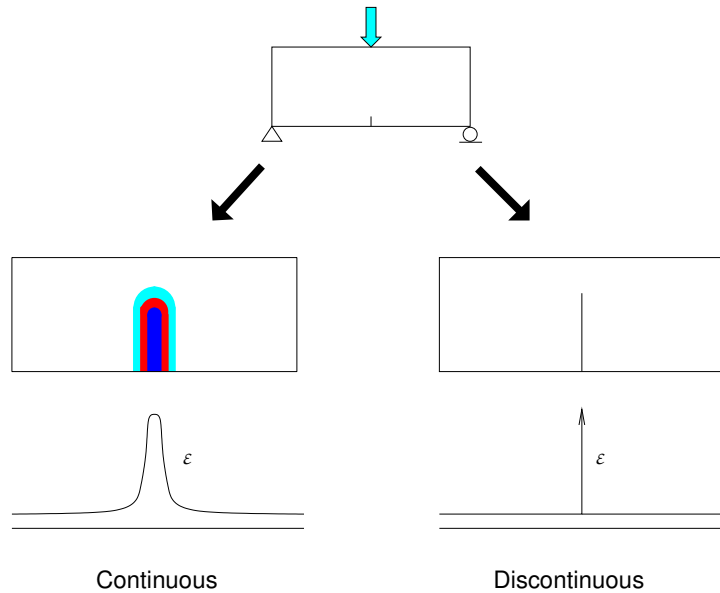
Figure 1.5 Representation of crack and corresponding strain localisation. Strain localisation occurs when the tensile strength is exceeded in the bottom part of the beam (top). The resulting crack can be modelled with a continuous (bottom left) or a discontinuous (bottom right) crack concept.

ing the nonlinear computation in order to make sure that the solution path is correct.

An alternative to the use of expensive error estimation in driving mesh adaptivity is the use of inexpensive error indication. In [6, 90], an error indicator is derived from the critical wave length in the damage model. The desired element sizes are defined as functions of the damage level and are successfully applied in *h-*, *r-* and *hr*-adaptivity. However, in [10,11,69], it is suggested that it is as important to assess the error both in the linear regime (where no damage exists) and the nonlinear regime (where there exists damage). Without damage and localised strain fields, error estimation may be a suitable choice to drive the adaptive process in the earlier computational steps (the linear elastic part), whereas the error indicator is used when the solution presents nonlinearity. To support this idea, it is claimed in [25] that the error estimate [49] becomes less significant in the localisation region as damage grows and stresses tend to vanish. By applying the error estimation during the whole computation, we can verify these statements.

As *h*-adaptivity leads to changes in mesh configuration, the finite element analysis needs reformulating the shape functions, stiffness matrix and force vectors. A challenging alternative is the use of richer interpolation, or *p*-adaptivity. In this contribution, we investigate performances of *p*-adaptivity in combination with simple mesh gradation applied to problems with strain localisation. A slightly modified

version of the error estimation in [29] is chosen in this study as it can be easily applied to problems with non-uniform higher-order interpolations while still being well-integrated with the optimality criterion in designing the element sizes. Performance of the adaptive models will be investigated in Chapter 6.

1.4.2 The discontinuous crack model

As the terms *cracking* and *rupture* already imply, introduction of discontinuity as a result of material failure seems to be natural. Unlike the continuous modelling, the fracture criterion of this concept is defined separately from the constitutive relations. Discontinuities in the material domain are modelled by introducing a jump either in the displacement field (the so-called strong discontinuity) or the strain field (the so-called weak discontinuity).

A classical approach to model a crack is to adapt the finite element mesh according to geometrical change due to crack propagation. It then requires a continuous change of the topology of discretisation (i.e., remeshing process), which is computationally laborious and complicated. An alternative approach is to place interface elements of zero width in the finite element mesh [79]. However, since the direction of crack growth is not known a priori, small elements are needed to allow a jump in the displacement field in a range of possible directions of cracking, resulting in an expensive computation.

Without restriction to mesh alignment, the crack can be modelled in a much simpler way. It is shown in [5,63,85] that modelling cracks within elements is possible by both weak and strong discontinuity assumptions. Via the introduction of internal degrees of freedom, the discontinuous contribution is solved on the element level and the displacement jump can be modelled without being restricted to the underlying mesh. The method is known as the *embedded discontinuity approach*. Another recent development is to model the displacement discontinuity by simply adding extra nodal degrees of freedom via the partition of unity (PU) [14], which is a basic property of the finite element interpolation. This *PU-based finite element method*, also known as the *extended finite element method* (XFEM) [21,32,58,101], is more robust in implementation than the embedded discontinuity approach. As extra degrees of freedom, the enhanced functions are solved at the global level and do not involve modification at the element level, thus preserving symmetry of the
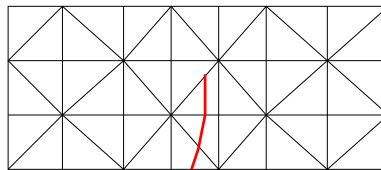
Figure 1.6 Discontinuity modelling based on enrichment via the partition of unity.

global stiffness matrix.

Although, via the PU concept, the jump in the displacement field can be modelled without any restriction to the underlying finite element mesh (cf. Figure 1.6), the resolution of the discretisation along the cracked element still needs to be ensured. It has been observed in [99] that a too coarse discretisation may lead to a rough global response. Even without the oscillations, the response and the resulting crack path may not be sufficiently accurate. So far, without any research investigating mesh requirement in the PU-based discontinuity model, it is hardly certain that the propagation of a discontinuity leads to an acceptable level of accuracy. We will investigate intensively the discretisation aspect of the discontinuous crack model in Chapter 7.

Finite element interpolation

The finite element method is a numerical tool for approximating solutions of boundary value problems, which are usually too complicated to be solved by analytical techniques. As its name implies, the method employs the concept of subdividing the model problem into a series of finite elements over which variational formulations are set to construct an approximation of the solution.

The finite element approximation relies mainly on the interpolation via piecewise polynomials over a set of finite elements. As mentioned before, the introduction of higher-order interpolation functions (also called *shape functions*) is one technique to achieve a better approximation to the solutions of the problem and is our main motivation for this study.

The higher-order interpolation can be constructed either based on the so-called Lagrange (*non-hierarchical*) elements, or based on adding *hierarchical* counterparts. Two types of hierarchical shape functions, formulated based on elements and nodes, as well as some critical aspects are presented in this chapter.

In the first part of this chapter, we attempt to give a short introduction of standard finite element analysis, and move on to the formulations of higher-order shape functions in the second part, as this concept will subsequently be used in so-called p-elements in the rest of the thesis.

2.1 Basic settings

Let Ω be a bounded domain with the boundary $\partial\Omega$. The boundary consists of the Dirichlet boundary Γ_d and the Neumann boundary Γ_n for which $\Gamma_d \cap \Gamma_n = \emptyset$ and $\overline{\Gamma_d \cup \Gamma_n} = \partial\Omega$. For a problem in statics, we try to find the unknown solution $\boldsymbol{u}$ of the variational boundary value problem

$$\int_{\Omega} \boldsymbol{\varepsilon}(\boldsymbol{v}) : \boldsymbol{\sigma}(\boldsymbol{u}) \, \mathrm{d}\Omega = \int_{\Gamma_n} \boldsymbol{v} \cdot \boldsymbol{g} \, \mathrm{d}\Gamma + \int_{\Omega} \boldsymbol{v} \cdot \boldsymbol{q} \, \mathrm{d}\Omega \tag{2.1}$$

which can be written in terms of derivatives of trial and test functions, u and v, as

$$\int_{\Omega} (\boldsymbol{\nabla} \boldsymbol{v}) : \boldsymbol{D} : (\boldsymbol{\nabla} \boldsymbol{u}) \, \mathrm{d}\Omega = \int_{\Gamma_n} \boldsymbol{v} \cdot \boldsymbol{g} \, \mathrm{d}\Gamma + \int_{\Omega} \boldsymbol{v} \cdot \boldsymbol{q} \, \mathrm{d}\Omega \tag{2.2}$$

The test function v is any arbitrary function in the Sobolev space $\mathcal{V}$, which is defined by $\mathcal{V} := \{v \in H^1(\Omega); v = 0 \text{ on } \Gamma_d\}$. Moreover, $\boldsymbol{\varepsilon}(\boldsymbol{v}) := \boldsymbol{\nabla} \boldsymbol{v}$ and $\boldsymbol{\sigma}(\boldsymbol{u}) := \boldsymbol{D} : \boldsymbol{\nabla} \boldsymbol{u}$ represent strains and stresses, $\boldsymbol{g}$ represents the traction forces along the boundary Γ_n and $\boldsymbol{q}$ denotes the body forces in the domain Ω. The Galerkin weak form of a linear problem can also be written as

$$\mathcal{B}(\boldsymbol{u}, \boldsymbol{v}) = \mathcal{F}^{(\Gamma_n)}(\boldsymbol{v}) + \mathcal{F}^{(\Omega)}(\boldsymbol{v}) = \mathcal{F}(\boldsymbol{v}), \quad \forall \boldsymbol{v} \in \mathcal{V} \tag{2.3}$$

where the term $\mathcal{B}(\cdot, \cdot)$ is a symmetric positive-definite bilinear form, corresponding to the left-hand-side of Eq. (2.2), while $\mathcal{F}^{(\Gamma_n)}$ and $\mathcal{F}^{(\Omega)}$ refer to the first and the second terms of the right-hand-side, respectively.

In order to approximate the continuous variable $\boldsymbol{u}$, a numerical computation must be performed. The discretised system of equations

$$\mathcal{B}(\boldsymbol{u}_{(h,p)}, \boldsymbol{v}_{(h,p)}) = \mathcal{F}(\boldsymbol{v}_{(h,p)}), \quad \forall \boldsymbol{v}_{(h,p)} \in \mathcal{V}_{(h,p)} \tag{2.4}$$

is solved in the finite element space $\mathcal{V}_{(h,p)}$, where $\mathcal{V}_{(h,p)} \subset \mathcal{V}$. The subscripts h and p denote the finite element analysis using element size h and polynomial order p. As a result, the solution $\boldsymbol{u}_{(h,p)}$ is an approximation to the exact function, $\boldsymbol{u}$. The approximate solution $\boldsymbol{u}_{(h,p)} \in \mathcal{V}_{(h,p)}$ and the test function $\boldsymbol{v}_{(h,p)} \in \mathcal{V}_{(h,p)}$ are discretised as

$$\boldsymbol{u}_{(h,p)} = \sum_{i=1}^{n} \boldsymbol{\phi}_i \, \boldsymbol{a}_i = \boldsymbol{\phi} \cdot \boldsymbol{a}, \quad \boldsymbol{v}_{(h,p)} = \sum_{j=1}^{n} \boldsymbol{\psi}_j \, \boldsymbol{c}_j = \boldsymbol{\psi} \cdot \boldsymbol{c} \tag{2.5}$$

via the use of basis functions (also known as shape functions) $\boldsymbol{\phi}_i$ and $\boldsymbol{\psi}_j$ of the trial (unknown solution) and the test functions, respectively. Substituting the discretised fields $\boldsymbol{u}_{(h,p)}$ and $\boldsymbol{v}_{(h,p)}$ back into Eq. (2.4) results in a system of discretised equations

$$\sum_{j=1}^{n} \boldsymbol{K}_{ij} \boldsymbol{a}_j = \boldsymbol{f}_i, \quad i = 1, 2, .., n, \quad n := \text{number of nodes} \tag{2.6}$$

where $\boldsymbol{K}_{ij} := \mathcal{B}(\boldsymbol{\phi}_j, \boldsymbol{\psi}_i)$, $\boldsymbol{f}_i := \mathcal{F}(\boldsymbol{\psi}_i)$, and $\boldsymbol{a}_j$ denotes the approximate solutions corresponding to the shape function $\boldsymbol{\phi}_j$. Eq. (2.6) can be rewritten in a matrix form as

$$\boldsymbol{K}\boldsymbol{a} = \boldsymbol{f} \tag{2.7}$$

where $\boldsymbol{K}$ denotes the stiffness matrix of the linear system, $\boldsymbol{a}$ represents the vector containing the unknowns and $\boldsymbol{f}$ denotes the force vector.
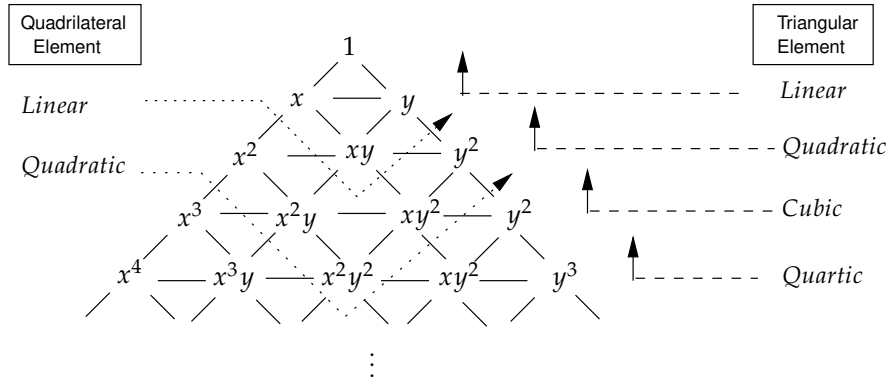
Figure 2.1 Complete 2D polynomial terms described by Pascal's triangle [18].

2.2 Element-based finite element shape functions

The finite element shape functions ϕ are characterised by two basic features, as suggested in [105], which are the continuity requirement[*] and the so-called *partition-of-unity* property. The latter property suggests that

$$\sum_{i=1}^{n} \phi_i(x) = 1, \forall x \in \Omega \tag{2.8}$$

allowing the description of rigid body motions. Importantly, the shape functions should not permit straining of an element when nodal displacements are caused by a rigid body displacement.

The finite element interpolation is fundamentally set in a piecewise polynomial format. To ensure the convergence of the approximation, it has been suggested that the shape functions should contain complete polynomials, which can be described in the Pascal triangle shown in Figure 2.1[†]. Basically, there are two categories of polynomial-based interpolation functions, namely the *non-hierarchical* functions and the *hierarchical* functions. The key difference between the two schemes is how the polynomial bases are upgraded to higher-order levels. While higher-order shape functions in the non-hierarchical scheme are completely different from the lower-order bases, the hierarchical scheme hierarchically adds the higher-order contributions and retains the lower-order bases without any reformulation. Details of the two versions will be described in this section.

---

[*]The differential equations studied here are all second-order. Hence, $\mathcal{C}^0$-continuity is required (that is, interelement continuity of the unknowns but not of derivatives of the unknowns).

[†]The quadrilateral elements referred to in Figure 2.1 are the so-called *Lagrangian* elements. The quadrilateral *serendipity* elements use a subset of the Lagrangian elements' polynomials.
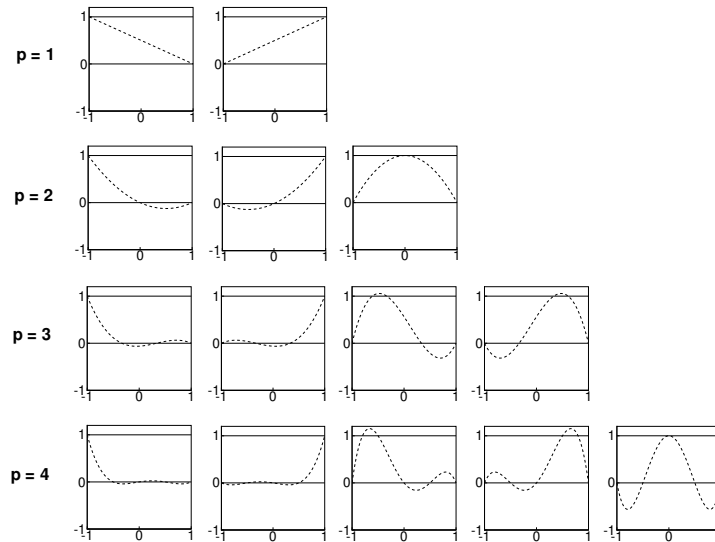
Figure 2.2 One dimensional shape functions for a non-hierarchical element.

2.2.1 Non-hierarchical (classical) shape functions

The classical finite element approach employs the so-called *Lagrange polynomials* introducing the local interpolation function by prescribing values at nodal points. The approach is a direct extension of the classical Lagrange interpolation. The interpolation is based on fitting values at nodal points. For one-dimensional problems, the shape functions containing polynomials of degree p are of the general form

$$\phi_i^{(1D,p)}(\xi) = \frac{\displaystyle\prod_{j=1}^{p}(\xi - \xi_j)}{\displaystyle\prod_{j=1;i\neq j}^{p}(\xi_i - \xi_j)} \tag{2.9}$$

where ξ_i, $i = 1, 2, ..., p + 1$, denotes a set of nodal coordinates in the finite element model. With some manipulations, the one-dimensional functions can be extended to generate higher-dimensional functions such as 2D quadrilateral and 3D brick elements[‡].

The computation of the shape functions in the given form obviously requires the reconstruction of the shape functions once it is upgraded to higher orders, which implies that the stiffness matrix must be completely recomputed. As our mesh

---

[‡]It is noted here that the shape functions for triangular and pyramid elements can also be formed differently, in terms of area coordinates or barycentric coordinates. (See, for example, [93].)
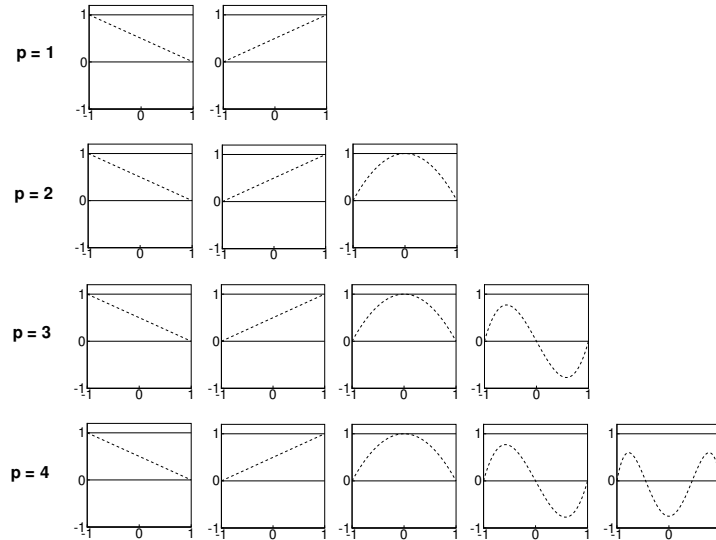
Figure 2.3 One dimensional shape functions for a hierarchical element based on Legendre polynomials.

adaptive technique includes *p*-adaptivity, having to recompute all stiffness matrix components everytime the mesh is upgraded can be an unpreferable feature.

2.2.2 Hierarchical shape functions

Unlike in the classical version, higher-order shape functions can be extended by adding an extra set of functions while the existing functions are preserved, i.e. span of $\phi_i^{(p)}$ is contained in span of $\phi_i^{(p+1)}$, in the hierarchical approach. Some examples of hierarchical interpolations are those based on *Legendre polynomials* (for example, [93]), *Chebychev polynomials* (for example, [98]) and *Hermite polynomials* [61]. Also, Lagrange shape functions can be reformulated in the hierarchical form (for example, [27]), where the hierarchical degrees of freedom can be referred to as tangential derivatives of various orders at the midside nodes.

In this work, we focus on the use of Legendre polynomials since they possess orthogonality implying no linear dependence between the polynomial functions [105] and, hence, a sparse data structure as compared to the use of classical shape functions. The Legendre basis also provides consistent element conditioning number when the polynomial order is increased, thus leading to a smaller numerical round-off error and a faster convergence in nonlinear analysis, as compared to other bases [35].

The Legendre interpolation function is based on the Legendre polynomials, which originally are solutions to Legendre's differential equations. The polynomial

of degree p may be expressed using *Rodrigue's formula*

$$P_p(\xi) = (2^p p!)^{-1} \frac{d^p}{d\xi^p} \left[(\xi^2 - 1)^p \right]. \tag{2.10}$$

In addition to the standard linear shape functions (vertex modes), the hierarchical enrichment including edge and internal modes are defined in the interval $-1 \leq \xi \leq 1$ as

$$\phi_i^{1D,p}(\xi) := \sqrt{\frac{2p-1}{2}} \int_{-1}^{\xi} P_{p-1}(t) \, dt = \frac{1}{\sqrt{2(2p-1)}} \left[P_p(\xi) - P_{p-2}(\xi) \right] \tag{2.11}$$

for $p \geq 2$. The main difference between the standard finite element shape functions and Legendre shape functions, given in Figures 2.2 and 2.3, can be clearly observed.

Similar to the standard finite element interpolation, the higher-dimensional functions are based on products of one-dimensional functions and Legendre polynomials [93]. Another set of combinations, in forming the higher-dimensional set of shape functions, has been suggested in [24], with the improvement of sparsity and conditioning of the stiffness matrix.

2.2.3 Comparison

It is noted that, in the hierarchical approach, the higher-order degrees of freedom, known as edge modes and internal modes (also known as bubble modes), are not based on nodes. Figure 2.4 compares how the two interpolation schemes work. While the non-hierarchical version (based on Lagrange polynomials) interpolates values at nodes, the hierarchical version (here, based on Legendre polynomials) interpolates values at the primary nodes as well as values corresponding to additional higher-order interpolation functions. Due to such characteristic, the following difficulties obviously emerge:

(A) Enforcement of constraints

The standard element shape functions have a superiority over the hierarchical functions when it comes to constraint enforcement. Possessing the Kronecker delta property (i.e. $\phi_i(x_j) = \delta_{ij}$, where i and j refer to nodes), either external constraints (i.e. prescribed values of primary variables) or internal constraints (i.e. the relationship between different degrees of freedom) can be simply imposed at nodes. In contrast, the enforcement of constraints in the hierarchical approach causes some difficulties due to the obsence of nodes on edges. Direct imposition can be applied only in case of constant or linear constraints ($p \leq 1$). In that case the edge shape functions at the corresponding edge are dropped out (i.e. zero-value prescribed) and the linearly varying constraints are directly prescribed at nodes, which exist only at the vertices of an element in the hierarchical approach.

In real applications, there hardly exist problems with constraints of higher-order functions. However, if necessary, special techniques such as Lagrange multipliers
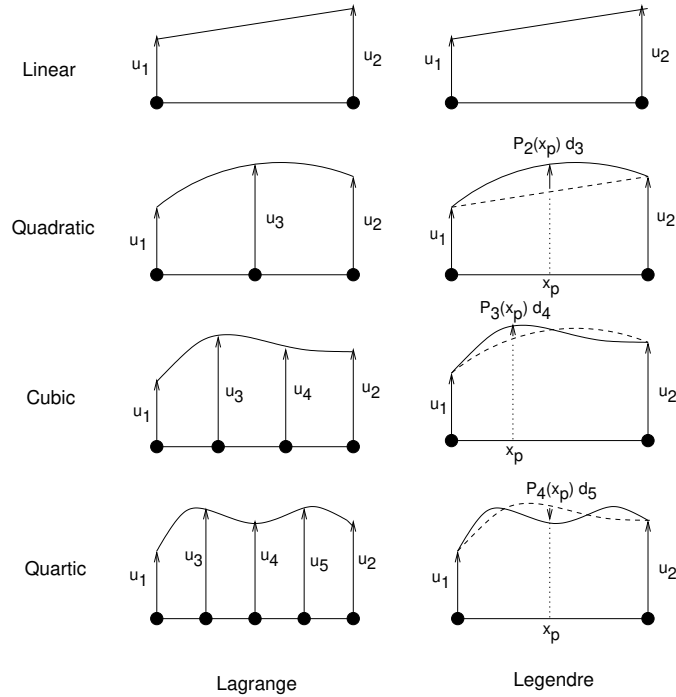
Figure 2.4 Higher-order interpolation based on standard isoparametric element and hierarchical element based on Legendre polynomials.

or a Penalty formulation may be applied (for example, [65]), leading to a modified Galerkin weak form.

(B) Modelling of geometrical data

The standard *p*-elements are able to describe the model geometry via the higher-order shape functions by relocating the edge nodes. A complex geometry, however, brings some complications in the hierarchical *p*-version as the edge nodes do not exist. As a remedy, geometrical mapping via linear/quadratic parametric mapping functions [93] or the so-called *blending functions* [36] is suggested. The blending functions can be flexibly selected, thus allowing an accurate representation of various configurations.

(C) Compatibility of the hierarchical modes between adjacent elements

Due to the $\mathcal{C}^0$-continuity requirement, it is necessary that the interpolation functions (shape functions) between adjacent elements are compatible. Using the standard shape functions, nodes at shared edges (in case of 2D problem) and shared faces (in case of 3D problem) have identical values of the primary unknown, ensuring compatibility of the corresponding shape functions. In contrast, the hierarchical
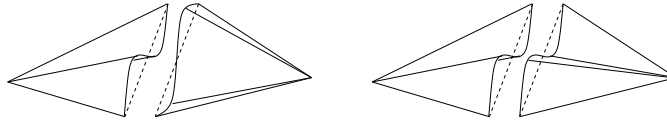
Figure 2.5 Example of interelement compatibility of the hierarchical edge mode: (left) wrong definition and (right) correct definition.

version does not have a physical definition of the modes at shared parts of elements. A problem of incompatibility may occur. In Figure 2.5, we illustrate this problem using the edge mode that is added for upgrading an element from quadratic order to cubic order. Obviously, the edge shape functions may not be continuous over the interelement boundary if the edge mode is separately defined for each element (cf. Figure 2.5 (left)). This is due to the fact that asymmetric shape functions do not appear in pairs, as in Figure 2.2. Nevertheless, with careful consideration, the edge mode can be properly defined using an appropriate node ordering rule as in Figure 2.5 (right).

2.3 Node-based hierarchical enhancement

In the last decade, the so-called meshless methods [22] have gained popularity due to their ability in avoiding a complicated remeshing procedures in adaptive finite element analysis. However, they possess some limitations. For example, the meshless shape functions (e.g. the moving Least-Squares (MLS) approximation [50]) are normally much more computionally expensive than the conventional finite element interpolation. Furthermore, most meshless shape functions do not possess the Kronecker delta property, implying that the approximation function does not pass through data points, thus leading to difficulties in imposing essential boundary conditions [46,59,65]. And, since they are *meshless*, difficulties due to numerical integration arise [22]. Instead of being specified on elements, the quadrature points are then located in the newly created background cells, which may not conform the domain (or subdomain) geometry. Such treatment brings about quadrature errors and a background of integration cells destroys the *meshless* nature of these interpolations.

By such shortcomings, attention has been concentrated on how to improve the existing finite element models with the strong points of the meshless methods. Examples of some attempts are the cloud-based finite element method [62], the partition of unity finite element method [57], the generalised finite element method [91,92], the special finite element method [13] and the new hierarchical finite element method [94], which are all based on the same concept: nodal enrichment via the partition of unity property of the finite element shape functions (cf. Eq. (2.8)). Based on the finite element *hat* functions[§], computational cost is much reduced as

---

[§]Due to its shape, linear finite element shape functions are also known as *hat* functions.

compared to the use of MLS shape functions.

The node-based enrichment technique inherits the strong points of meshless techniques while it retains the strong points of the FEM. For instance, the choice of enrichment functions is much more flexible as compared to the element-based hierarchical enrichment. The technique also concentrates hierarchical degrees of freedom at nodes thus providing a sparser band structure of the stiffness matrix than the one in the traditional approach. Moreover, the Kronecker delta property of the finite element interpolation introduces straightforward imposition of boundary constraints, while numerical integration based on element structure is automatic and conformed to the element domains resulting in better accuracy in the numerical analysis. The technique can be implemented easily and efficiently without introducing any complicated arrangement of hierarchical modes at edges and inside elements as in the element-based hierarchical *p*-version finite element method.

2.3.1 Enhancement technique

The enhancement of the finite element shape functions to higher-order polynomials being added through the partition-of-unity property has been applied in many studies (for example, [54, 91, 92, 94, 102]). The scheme avoids the use of additional nodes in the domain to enrich the polynomial order of the shape function and can be considered as a hierarchical class of *p*-enrichment. In particular, for the approximant $\boldsymbol{u}$, it is written that

$$\boldsymbol{u} = \sum_{i=1}^{n} \boldsymbol{\phi}_i \left(\sum_{j=1}^{m_i+1} \vartheta_j^{(i)} \tilde{\boldsymbol{a}}_j^{(i)} \right) \tag{2.12}$$

where, at node i, $\vartheta_1^{(i)} = 1$ always and the corresponding degree of freedom $\tilde{\boldsymbol{a}}_1^{(i)}$ represents rigid-body movement. The enhancement can be added hierarchically. To reveal this property, Eq. (2.12) can also be written as

$$\boldsymbol{u} = \sum_{i=1}^{n} \boldsymbol{\phi}_i \left(\boldsymbol{a}_i + \sum_{j=1}^{m_i} \gamma_j^{(i)} \boldsymbol{b}_j^{(i)} \right) \tag{2.13}$$

which distinguishes between the existing interpolation functions (corresponding to $\boldsymbol{a}_i$ with $i = 1, 2, .., n$) and the additional enrichment functions (corresponding to $\boldsymbol{b}_j^{(i)}$ with $i = 1, 2, .., n$ and $j = 1, 2, .., m_i$). Here, n and m_i are number of nodes and extra (non-unity) terms for node i, $\boldsymbol{\phi}$ denotes the interpolation function of the discrete primary unknown $\boldsymbol{a}$, $\boldsymbol{\gamma}$ contains enhancement terms and $\boldsymbol{b}$ refers to a set of extra degrees of freedom that is introduced through the partition of unity property of the finite element shape function. Note that the interpolation of the degrees of freedom $\boldsymbol{b}$ is not set by $\boldsymbol{\gamma}$ alone but rather through the product $\boldsymbol{\phi}_i \gamma_j^{(i)}$.

It is noted here that Eq. (2.13) is equivalent to Eq. (2.12) assuming that $\{\boldsymbol{\vartheta}\} = \{\boldsymbol{\gamma}\} + \{1\}$, i.e. span of the enrichment function $\boldsymbol{\vartheta}(\boldsymbol{x})$ (cf. Eq. (2.12)) comprises one

extra component representing rigid body movement, i.e. the unity component, in addition to the span of the enrichment function $\gamma(x)$(cf. Eq. (2.13)).

2.3.2 Choices of polynomial enrichment functions

In order to construct a set of shape functions based on higher-order interpolation, one should realise that the resulting shape functions should possess the complete polynomial property to guarantee convergence of the finite element solutions and satisfy the continuity requirement. Generally, the polynomial enrichment functions are added through nodal shape functions, which are basically of linear order (i.e. only vertex shape functions exist). In such a case, to obtain higher-order shape functions, one may specify a set of enrichment functions at node i, according to the polynomial terms in Pascal's triangle (cf. Figure 2.1) as

$$\gamma_{(1\to2)}^{(i)} = \{\xi_i^2, \xi_i\eta_i, \eta_i^2\} \tag{2.14}$$

$$\gamma_{(1\to3)}^{(i)} = \{\xi_i^2, \xi_i\eta_i, \eta_i^2, \xi_i^3, \xi_i^2\eta_i, \xi_i\eta_i^2, \eta_i^3\} \tag{2.15}$$

$$\gamma_{(1\to4)}^{(i)} = \{\xi_i^2, \xi_i\eta_i, \eta_i^2, \xi_i^3, \xi_i^2\eta_i, \xi_i\eta_i^2, \eta_i^3, \xi_i^4, \xi_i^3\eta_i, \xi_i^2\eta_i^2, \xi_i\eta_i^3, \eta_i^4\} \tag{2.16}$$

to upgrade linear shape function to quadratic, cubic and quartic order, respectively. The subscript $(j \to k)$ here refers to an upgrade from polynomial degree j to polynomial degree k, and the superscript (i) refers to the enrichment function associated with node i. It is worth noting that the enrichment functions are added hierarchically, i.e.

$$\gamma_{(1\to p+1)}^{(i)} \subset \gamma_{(1\to p+2)}^{(i)} \subset \gamma_{(1\to p+3)}^{(i)} \subset \ldots \subset \gamma_{(1\to p+\infty)}^{(i)} \tag{2.17}$$

As such, the resulting shape functions can be viewed as a specific type of hierarchical shape functions.

The functions $\xi = \xi(x)$ and $\eta = \eta(y)$ must be chosen such that the aforementioned continuity requirement is satisfied. An example is the one proposed in [94], i.e.

$$\xi_i = (x - x_i) \text{ and } \eta_i = (y - y_i) \tag{2.18}$$

where x and y represents the global Cartesian coordinates of a point in the domain. Corresponding to the enrichment at node i, x_i and y_i denote the global coordinates of the nodal point. The choice represents the distance of any point to the to-be-enriched node, providing the continuity of the enrichment functions throughout the domain. The enrichment functions $\gamma_j^{(i)}$ increase in magnitude with increasing distance from the associated node i. However, the enrichment is cut off at the end of the element edge due to the multiplication with the existing finite element shape function ϕ_i, which equals zero in all elements not adjacent to node i.

In [33], the choice

$$\xi_i = \frac{(x - x_i)}{h_i} \ \text{ and } \ \eta_i = \frac{(y - y_i)}{h_i} \tag{2.19}$$

was chosen to weigh the enrichment function with h_i, the diameter of the largest finite element sharing node *i*. Obviously, this format provides an improved version of the format in Eq. (2.18) as a better conditioning number of the resulting stiffness matrix is obtained.

In addition to the enrichment functions presented above, it is also possible to select other types of polynomials such as *harmonic* polynomials as presented in [91] and [57]. This set of polynomials has the advantage that its dimension grows linearly with polynomial order, whereas the set of full polynomials from FEM grows quadratically.

2.4 Remarks

In this chapter, finite element shape functions have been introduced in various forms. In spite of their non-physical meaning, the hierarchical shape functions gain more popularity in the mesh adaptive studies as the existing shape functions are preserved thus avoiding that the whole stiffness matrix system must be recalculated. This attractive feature facilitates *p*-adaptive analysis to be carried out in this thesis.

The hierarchical enhancement of the finite element shape functions can be introduced in an element-based fashion or a node-based fashion. Despite the attractive features of the node-based hierarchical extension, one serious problem that makes the method unattractive is *linear dependence* of the resulting shape functions, which leads to unsolvability of the discretised equations. This is discussed in detail in Appendix A. Although there are some techniques to overcome such shortcoming, we do not wish to complicate the present work unnecessarily. Therefore, in this thesis, we will employ only the element-based hierarchical shape functions.

A posteriori error estimation

An important component of finite element adaptive analysis is how to assess the local error accurately. This error information normally gives a clue where and to which extent some parts of the mesh should be enhanced so that the finite element analysis can provide acceptably accurate and cost effective results. As such, the so-called *a posteriori* error estimators, which approximate the actual error at the end of the calculation step, play an important role in ensuring reliability of finite element models. The error information, which is the focus in this research work, refers to the error that is caused by inadequate discretisation in the finite element analysis, and it is also known as the *discretisation error*.

This chapter starts with a mathematical definition of the discretisation error in the finite element method, which is usually measured in terms of an energy norm. Then, we address some basic ideas about the standard *residual-type* error estimation, which later leads to the formulation of the simple error estimator used in this research. The chapter ends with some investigations about performances and some critical comments about the method.

3.1 Discretisation error

The discretisation error, e, is defined as

$$e := u - u_{(h,p)} \tag{3.1}$$

i.e. the difference between the exact solution to the mathematical model, u, and the finite element solution, $u_{(h,p)}$. Here, we assume that the error that comes from the numerical process, known as the *numerical error*, is marginal in comparison to the error in the discretisation part, and thus can be neglected.

Apparently, the error e in Eq. (3.1) cannot be computed directly since the exact solution u is generally unknown. Nevertheless, as a more refined/enriched discreti-

sation gives a better approximation to the actual solution $\boldsymbol{u}$, we can closely represent the actual solution $\boldsymbol{u}$ by a very fine discretisation (so-called *reference mesh*), via h-extension and/or p-extension*, for example.

The finite element solution from the refined/enriched system $\boldsymbol{u}_{(\tilde{h},\tilde{p})}$, obtained from solving the reference discretised problem

$$\mathcal{B}(\boldsymbol{u}_{(\tilde{h},\tilde{p})}, \boldsymbol{v}_{(\tilde{h},\tilde{p})}) = \mathcal{F}(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) \qquad \forall \boldsymbol{v}_{(\tilde{h},\tilde{p})} \in \mathcal{V}_{(\tilde{h},\tilde{p})} \tag{3.2}$$

is now denoted as a reference to the actual solution u. As a consequence, the discretisation error, defined in Eq. (3.1), is approximated by

$$\boldsymbol{e} \approx \boldsymbol{u}_{(\tilde{h},\tilde{p})} - \boldsymbol{u}_{(h,p)} =: \boldsymbol{e}_{(\tilde{h},\tilde{p})} \tag{3.3}$$

The approximation involved in Eq. (3.3) is sufficiently accurate because the actual solution $\boldsymbol{u}$ is much closer to the solution from the refined system $\boldsymbol{u}_{(\tilde{h},\tilde{p})}$ than to the primary solution $\boldsymbol{u}_{(h,p)}$.

In order to provide a proper measurement of global and elemental error, the discrete error should be measured in a well-defined norm. A classical option, also employed in this contribution, is the measurement of error in an *energy norm* defined as

$$\|\boldsymbol{e}\| := \sqrt{\mathcal{B}(\boldsymbol{e}, \boldsymbol{e})} = \sqrt{\sum_k \mathcal{B}_k(\boldsymbol{e}, \boldsymbol{e})} = \sqrt{\sum_k \|\boldsymbol{e}\|_k^2} \tag{3.4}$$

where the subscript k denotes the error contribution obtained from the elemental level. The global estimation is obtained by summing up the elemental contributions. The *global error measure* $\|\boldsymbol{e}\|$ is used in consideration whether or not the finite element solution is acceptably accurate. As well, the *elemental error measure* of the element k,

$$\|\boldsymbol{e}\|_k := \sqrt{\mathcal{B}_k(\boldsymbol{e}, \boldsymbol{e})} \tag{3.5}$$

is necessary in driving the mesh adaptive process (See Chapter 5).

3.2 Standard residual-type error estimation

Basicallly, *a posteriori* error estimators can be categorised in two main groups namely the *recovery type* and the *residual type*. As aforementioned in Chapter 1, the residual-type error estimators are employed in this research. The methods, pioneered by the work of Babuška and Rheinboldt [12], determine the error by calculating the residual of the finite element solutions in each local space. Without

---

*The mesh may be either refined (*h*-extension) or enriched (*p*-extension). It is not necessary that both factors are enhanced to form the reference solution.
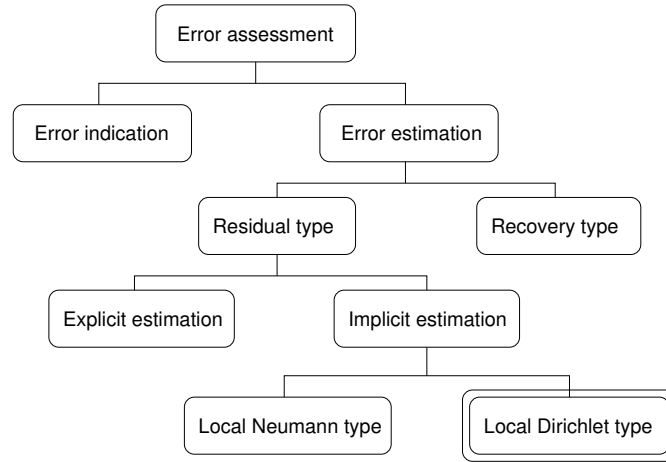
Figure 3.1 Error assessment techniques in finite element analysis. Note that the double-bounded box refers to the type used in this research.

relying on the superconvergence property of some sample points in the problem domain as in the recovery type, the residual-type error estimators can be applied to a wider variety of problems, including non-homogeneous higher-order interpolation or even nonlinear solution control, which are in the scope of this research.

The standard residual-type error estimation can be formulated either explicitly or implicitly. Whereas the *explicit version* employs the residuals in the current approximation directly, the *implicit version* uses the residuals indirectly via a set of local algebraic equations. Obviously, the implicit version, in comparison to the explicit version, requires more computational effort in solving an additional set of equations. The bigger effort, however, pays for the approximate error function, which is subsequently measured in a quantified norm. This error estimate provides more accurate information than those from the explicit version that relies on the inequality setting [4, 97]. Figure 3.1 shows an overview of error assessment techniques used in finite element analysis.

In this research, we concentrate on the implicit error estimation. The method consists of three components, i.e.

- a set of error equations,
- a reference discretisation, and
- a local computational framework.

Basically, the set of *error equations* is formulated based on residuals in a global computational framework. Without the known exact solutions, the residuals are estimated by setting the *reference discretisation* via either *h*-extension, *p*-extension, or any other mesh improvement approaches. Finally, the computational costs in-

volved with the reference discretisation can be reduced importantly by replacing the solution of a global system with the solution of a series of *local problems*.

Setting of error equations based on residuals

The residual-type error estimator, as its name implies, approximates the error based on residuals, i.e. the amount by which the finite element solution fails to satisfy pointwise the equilibrium equation in the mathematical model. The finite element solution $u_{(h,p)} \in \mathcal{V}_{(h,p)}$ is obtained by solving the set of equations

$$\mathcal{B}(u_{(h,p)}, v_{(h,p)}) = \mathcal{F}(v_{(h,p)}) \qquad \forall v_{(h,p)} \in \mathcal{V}_{(h,p)} \tag{3.6}$$

In order to estimate the error of the finite element solution, we recall the set of equations of the reference system[†]. That is,

$$\mathcal{B}(u_{(\tilde{h},\tilde{p})}, v_{(\tilde{h},\tilde{p})}) = \mathcal{F}(v_{(\tilde{h},\tilde{p})}) \qquad \forall v_{(\tilde{h},\tilde{p})} \in \mathcal{V}_{(\tilde{h},\tilde{p})} \tag{3.7}$$

is to be solved and used as a close representative to the actual model.

The difference between Eq. (3.6) and Eq. (3.7), and using Eq. (3.3) leads to a set of error equations

$$\mathcal{B}(e_{(\tilde{h},\tilde{p})}, v_{(\tilde{h},\tilde{p})}) = \mathcal{R}^u(v_{(\tilde{h},\tilde{p})}) = \mathcal{F}(v_{(\tilde{h},\tilde{p})}) - \mathcal{B}(u_{(h,p)}, v_{(\tilde{h},\tilde{p})})$$
$$\forall v_{(\tilde{h},\tilde{p})} \in \mathcal{V}_{(\tilde{h},\tilde{p})} \tag{3.8}$$

with the boundary condition that $e = 0$ on Γ_d. The residual $\mathcal{R}^u$, which is based on the primary unknown u, can be interpreted as a fictitious load by which the approximate solution deviates from the actual solution.

Setting of local computational framework

In fact, one can estimate the error of a finite element model by comparing the finite element solutions obtained from the original mesh to those from the enhanced mesh. This however requires a large amount of computation and makes no sense. There is obviously little value in estimating the error of a coarse discretisation by solving a global system of equations according to an enhanced discretisation. The computational costs involved with the error estimation would far outweigh those involved with solving for $u_{(h,p)}$, while at the same time an improved solution $u_{(\tilde{h},\tilde{p})}$ is already provided. By virtue of $u_{(\tilde{h},\tilde{p})}$, the solution $u_{(h,p)}$ has become redundant, and so has $e_{(\tilde{h},\tilde{p})}$. In contrast, an *efficient* calculation of $e_{(\tilde{h},\tilde{p})}$ should involve *local* (rather than global) solutions of $u_{(\tilde{h},\tilde{p})}$.

Since Eq. (3.8) is defined globally, it requires a large amount of computer resources. In order to avoid this, the local spaces $\mathcal{V}_k$, $k = 1, 2, ..., n$ and $\mathcal{V}_k \subset \mathcal{V}$,

---

[†]Again, it is not necessary that both *h* and *p* factors are enhanced to form the reference solution. However, at least one factor needs to be upgraded to form the reference system of equations.

are defined and the residual-based error is computed in each local space. That is, instead of solving Eq. (3.8), we, instead, solve a set of local equations

$$\mathcal{B}_k(\boldsymbol{e}_{(\tilde{h},\tilde{p})}, \boldsymbol{v}_{(\tilde{h},\tilde{p})}) = \mathcal{R}_k^u(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) \qquad \forall \boldsymbol{v}_{(\tilde{h},\tilde{p})} \in \mathcal{V}_{k(\tilde{h},\tilde{p})} \qquad (3.9)$$

where the local residual is defined as

$$\mathcal{R}_k^u(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) = \mathcal{F}_k(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) - \mathcal{B}_k(\boldsymbol{u}_{(h,p)}, \boldsymbol{v}_{(\tilde{h},\tilde{p})}) + \int\limits_{\partial\Omega_k \setminus (\partial\Omega_k \cap \Gamma_n)} \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{n}_k} \boldsymbol{v}_{(\tilde{h},\tilde{p})} \, \mathrm{d}\Gamma \quad (3.10)$$

As a result from the integration by parts on each local domain, the additional contribution, which is the last term on the right-hand side of Eq. (3.10), represents the normal derivatives (or flux) on the interelement boundary $\partial\Omega_k$ as well as on Γ_d and cancels in the global system of equations. Note that the contribution of the normal derivative on element edges on Γ_n, i.e. $\int_{\partial\Omega_k \cap \Gamma_n} \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{n}_k} \boldsymbol{v} \, \mathrm{d}\Gamma$ is included in $\mathcal{F}_k(\boldsymbol{v})$ as defined earlier.

To obtain the error associated with the primary unknowns $\boldsymbol{u}_{(h,p)}$, the local error equations (cf. Eq. (3.9)) must be solved. It is then necessary to define a proper set of boundary conditions of these local problems. We will address this subject in the next section.

3.3 Boundary conditions of the local error equations

As mentioned in the last section, a key ingredient in solving local error equations is setting the boundary conditions to be prescribed in Eq. (3.9). Taken from the global finite element setting, the only Dirichlet boundary condition defined in each local space $\Omega_k \subset \Omega$ is

$$\boldsymbol{e} = \boldsymbol{0} \quad \text{on } \partial\Omega_k \cap \Gamma_d \qquad (3.11)$$

This is because the primary unknown $\boldsymbol{u}$ is exactly prescribed on the Dirichlet boundary Γ_d. Obviously, additional boundary conditions for the local problems are needed.

Basically, there are two subclasses of the implicit residual error estimation, depending on how the boundary conditions are defined in the local problems. While the *Neumann-type* error estimation prescribes the Neumann conditions in the local problems, the *Dirichlet-type* error estimation imposes the local Dirichlet conditions, see also the overview in Figure 3.1. Some basic ideas about the two approaches will be presented in this section.

3.3.1 Local Neumann conditions

The imposition of the non-homogeneous flux boundary conditions (local Neumann conditions), represented by the last term of Eq. (3.10), may be set via the simple *flux*

averaging technique as

$$\frac{\partial \boldsymbol{u}_{(h,p)}}{\partial \boldsymbol{n}_k} \approx \left\langle \frac{\partial \boldsymbol{u}_{(h,p)}}{\partial \boldsymbol{n}_k} \right\rangle = \frac{1}{2} \boldsymbol{n}_k \cdot \left\{ (\boldsymbol{\nabla} \boldsymbol{u}_{(h,p)})_k + (\boldsymbol{\nabla} \boldsymbol{u}_{(h,p)})_{k'} \right\} \ \text{ on } \ \partial \Omega_k \cap \partial \Omega_{k'} \quad (3.12)$$

The considered edge of an element k is shared by another (adjacent) element denoted as k'. The introduction of two distinct indices k and k' allows to describe jumps of the normal fluxes at the interelement boundary.

   The simple averaging has been criticised for being ad-hoc and fails to respect the basic requirement for the local problem to be well-posed. Some researchers [3, 48] have proposed a new modification, the so-called *equilibrated flux* approach by setting the equilibration condition

$$\mathcal{F}_k(\boldsymbol{v}) - \mathcal{B}_k(\boldsymbol{u}_{(h,p)}, \boldsymbol{v}) + \int\limits_{\partial \Omega_k \backslash (\partial \Omega_k \cap \Gamma_n)} \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{n}_k} \boldsymbol{v} \, \mathrm{d}\Gamma = 0 \qquad (3.13)$$

where $\boldsymbol{v} = 1$ and $\boldsymbol{v} = \boldsymbol{\phi}$ are selected for zeroth-order equilibration and first-order equilibration conditions, respectively. And with the consistency condition

$$\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{n}_k} + \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{n}_{k'}} = \boldsymbol{0} \quad \text{ on } \ \partial \Omega_k \cap \partial \Omega_{k'} \qquad (3.14)$$

the error equations are well-posed on the regular subspace and the resulting error estimator will provide a guaranteed upper bound of the exact error.

   It should be noted that imposing only Neumann boundary conditions in the local problems is not sufficient. It is necessary to impose a proper set of Dirichlet conditions to eliminate the zero energy modes (rigid body modes), leading to solvability of the equations. Obviously, the Dirichlet conditions described in Eq.(3.11) are not sufficient for solving the local problems that are not attached to the Dirichlet boundary. To overcome this problem, one may reformulate the local problem over a reduced subspace where the zero energy modes have been factored out [1,2,16].

3.3.2 Local Dirichlet conditions

Modelling of the equilibrated residual fluxes at the interelement boundaries generally requires high computational effort. To avoid such complicated computation, the local Neumann boundary conditions in conventional element residual method may be replaced by a set of local Dirichlet conditions. The method approximates local errors without the necessity to compute the flux jump, thus the computational cost can be significantly diminished. However, this assumption leads to a lower bound estimate that is often not of a good quality.

   An improvement of the approach has been proposed by Díez et al. [29]. In their approach, an additional set of local error equations is introduced to help improving quality of the error estimate computed based on the elemental basis. The error function can be approximated by solving a set of local problems whose spaces overlap. These local (patch) spaces must be selected in such a way that
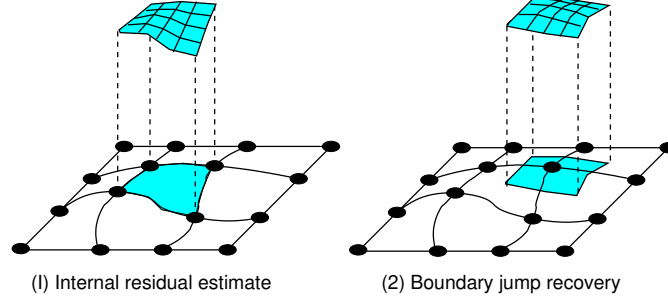
Figure 3.2 Two steps of residual-based error computation in the local Dirichlet-type framework based on local *h*-refinement [29].

- $\overline{\Omega} = \bigcup_k \overline{\Omega}_k^{\text{pat}}$, i.e. all patches together cover the whole problem domain, and
- $\Omega_i^{\text{pat}} \cap \Omega_j^{\text{pat}} \neq \emptyset$, i.e. a patch Ω_i^{pat} must overlap partly at least one other patch Ω_j^{pat} that is in the neighbourhood. The overlapping of patches depends on how the reference mesh is chosen.

For a local space (or patch) Ω_k^{pat}, a set of homogeneous boundary conditions is defined by suppressing error components as zero on the local boundary $\partial\Omega_k^{\text{pat}} \setminus (\partial\Omega_k^{\text{pat}} \cap \Gamma_n)$. The error estimate can then be obtained by finding $\boldsymbol{\xi} \in \mathcal{V}_k^*$ where $\mathcal{V}_k^* := \{ \boldsymbol{v} \in H^1(\Omega_k^{\text{pat}}); \boldsymbol{\xi} = \boldsymbol{0} \text{ on } \Gamma_d \cup \partial\Omega_k^{\text{pat}} \}$ from

$$\mathcal{B}_k(\boldsymbol{\xi}_{(\tilde{h},\tilde{p})}, \boldsymbol{v}_{(\tilde{h},\tilde{p})}) = \mathcal{R}_k^u(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) = \mathcal{F}_k(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) - \mathcal{B}_k(\boldsymbol{u}_{(h,p)}, \boldsymbol{v}_{(\tilde{h},\tilde{p})})$$
$$\forall \boldsymbol{v}_{(\tilde{h},\tilde{p})} \in \mathcal{V}_{k(\tilde{h},\tilde{p})} \qquad (3.15)$$

temporarily neglecting the last term appearing in Eq.(3.10). The space $\mathcal{V}_k = \text{supp}(\Omega_k^{\text{pat}})$, thus $\mathcal{V}_k \subset \mathcal{V}$. In the original work [29], this first estimate $\boldsymbol{\xi}$ to $\boldsymbol{e}_{(\tilde{h},\tilde{p})}$ is computed elementwise (i.e. the local space is based on one element) and denoted as the *interior estimate*.

Since the estimated error is suppressed to zero on the inter-patch boundaries, the obtained error solution is a poor approximation to the exact error. It is then necessary to enrich the first patch solutions by a set of patches overlapping the local space. Let Λ_l be the local space that overlaps Ω_k, find another error estimate $\boldsymbol{\eta} \in \mathcal{U}_l^*$, where $\mathcal{U}_l^* := \{ \boldsymbol{v} \in H^1(\Lambda_l^{\text{pat}}); \boldsymbol{\eta} = \boldsymbol{0} \text{ on } \Gamma_d \cup \partial\Lambda_l^{\text{pat}} \}$ from another boundary value problem

$$\mathcal{B}_l(\boldsymbol{\eta}_{(\tilde{h},\tilde{p})}, \boldsymbol{v}_{(\tilde{h},\tilde{p})}) = \mathcal{R}_l^u(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) = \mathcal{F}_l(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) - \mathcal{B}_l(\boldsymbol{u}_{(h,p)}, \boldsymbol{v}_{(\tilde{h},\tilde{p})})$$
$$\forall \boldsymbol{v}_{(\tilde{h},\tilde{p})} \in \mathcal{U}_{l(\tilde{h},\tilde{p})} \qquad (3.16)$$

The second estimate $\boldsymbol{\eta}$ is based on the collection of parts in surrounding elements to form each *patch* overlapping the element domain (thus interior domain) and is
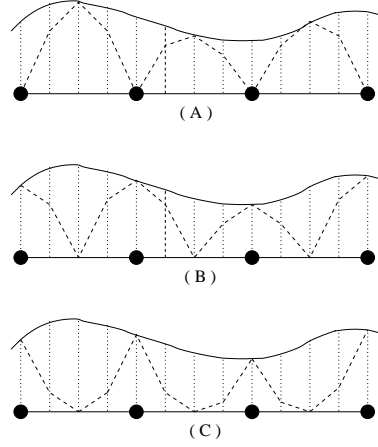
Figure 3.3 How the local Dirichlet method works: (A) interior estimation ξ, (B) patch estimation before orthogonality setting η and (C) patch estimation after the orthogonality setting η^*. The exact errors and the estimated errors are shown in solid and dashed lines, respectively. The filled circles denote nodal points in the one-dimensional problem domain.

called the *patch estimate* [29]. Similar to the interior estimate ξ, the local errors on $\partial \Lambda_l \backslash (\partial \Lambda_l \cap \Gamma_n)$ are prescribed to zero. This patch estimate provides information of the error caused by the residual fluxes on the elemental boundaries. See Figure 3.2 for an illustration of the two-step error computation.

To combine components from different patches, the contributions ξ and η must be adjusted to satisfy the Galerkin orthogonality property, that is $\mathcal{B}_l(\xi_{(\tilde{h},\tilde{p})}, \eta_{(\tilde{h},\tilde{p})}) = 0$ on Λ_l. Retrieving the interior estimate which is projected onto Λ_l, the patch solution in each Λ_l can be recalculated as

$$\eta^*_{(\tilde{h},\tilde{p})} = \eta_{(\tilde{h},\tilde{p})} - \frac{\mathcal{B}_l(\eta_{(\tilde{h},\tilde{p})}, \xi^*_{(\tilde{h},\tilde{p})})}{\mathcal{B}_l(\xi^*_{(\tilde{h},\tilde{p})}, \xi^*_{(\tilde{h},\tilde{p})})} \, \xi^*_{(\tilde{h},\tilde{p})} \quad \text{in } \Omega_l \tag{3.17}$$

where ξ^* is the projection of the interior estimate ξ on Λ_l which is adjusted as zero on $\partial \Lambda_l \backslash (\partial \Lambda_l \cap \Gamma_n)$ by

$$\mathcal{B}^*_l(\xi^*_{(\tilde{h},\tilde{p})}, v_{(\tilde{h},\tilde{p})}) = \mathcal{B}_l(\xi_{(\tilde{h},\tilde{p})}, v_{(\tilde{h},\tilde{p})}) \qquad \forall v_{(\tilde{h},\tilde{p})} \in \mathcal{U}_{l(\tilde{h},\tilde{p})} \tag{3.18}$$

where $\mathcal{B}^*_l(\cdot, \cdot)$ represents $\mathcal{B}_l(\cdot, \cdot)$ with the prescription of zero error on $\partial \Lambda_l \backslash (\partial \Lambda_l \cap \Gamma_n)$. The use of ξ^* instead of ξ in Eq. (3.17) ensures continuity of the estimated error function after setting the orthogonality to the interior values, which subsequently guarantees the continuity of the complete solution[‡].

[‡]It should be noted that the orthogonality setting procedure unfortunately causes some blind points on the Neumann's boundary where two patches meet. At those points on Γ_n, the estimation of zero error is obtained.

For each patch of elements

1. Retrieve elements in the patch:

    a) subdivide the patch/enrich the shape functions,

    b) compute the refined/enriched stiffness matrix $K^{\mathrm{el}}_{(\tilde{h},\tilde{p})}$,

    c) compute the refined/enriched load vector $f^{\mathrm{el}}_{(\tilde{h},\tilde{p})}$,

    d) interpolate solution vectors to the refined/enriched system $u^{\mathrm{el}}_{\mathrm{interp}}$, and

    e) add to the patch stiffness matrix $K^{\mathrm{pat}}_{(\tilde{h},\tilde{p})}$, patch load vector $f^{\mathrm{pat}}_{(\tilde{h},\tilde{p})}$ and patch solution vector $u^{\mathrm{pat}}_{\mathrm{interp}}$.

2. Impose the error boundary conditions for the patch problem, i.e. $e^{\mathrm{pat}} = \mathbf{0}$ on $\partial\Omega^{\mathrm{pat}}\backslash(\Omega^{\mathrm{pat}} \cup \Gamma_n)$.

3. Solve for patch error vector e^{pat}.

4. Retrieve existing patch error vector $e^{\mathrm{pat},0}$.

5. Set the orthogonality of the existing error vector and the new error vector $e^{\mathrm{pat}}_{\mathrm{orth}}$.

6. Add the orthogonal new error vector to the global numbering $e^{\mathrm{glob},0} = e^{\mathrm{glob},0} + e^{\mathrm{pat}}_{\mathrm{orth}}$.

7. Continue to next patch.

Table 3.1 Flow diagram for the error estimation [29].

By adding the two components, the complete estimate becomes

$$e_{(\tilde{h},\tilde{p})} \approx \boldsymbol{\xi}_{(\tilde{h},\tilde{p})} + \boldsymbol{\eta}^{*}_{(\tilde{h},\tilde{p})} \tag{3.19}$$

which shows the automatic recovery of the errors along the inter-patch boundaries.

The estimated error can be measured in the energy norm by

$$\|e_{(\tilde{h},\tilde{p})}\|^2_k \approx \|\boldsymbol{\xi}_{(\tilde{h},\tilde{p})}\|^2_k + \|\boldsymbol{\eta}^{*}_{(\tilde{h},\tilde{p})}\|^2_k \tag{3.20}$$

and summed up to get the global error in the energy norm (cf. Eq. (3.4)). Figure 3.3 shows how this local Dirichlet error estimation works in one-dimensional setting. The method can be implemented following the flow chart suggested in Table 3.1.

3.4 Error estimation for non-uniform interpolation

In the original version [29], two sets of local problems are formulated. First, the interior residual-based error is computed by setting zero error on the edge of the elements which do not belong to the Neumann's boundary. The second set is to recover the residual on the boundary of the element by setting the patches, based on nodes or edges, overlapping the neighbouring elements.

Originally, the local Dirichlet type error estimator was formulated and successfully applied with a local *h*-extension[§]. To provide a proper reference discretisation in the case of non-uniform interpolation, a *p*-extension is more suitable. Instead of subdividing each local space (element), the degree of polynomial interpolation (p) is upgraded to one higher order ($p + 1$). The error with respect to the reference mesh, in comparison to the actual error, is illustrated in Figure 3.4.

Here, we have implemented two error estimators for use with non-uniform interpolation, i.e.

- based on the element-based hierarchical enhancement (cf. Section 2.2), and

- based on the node-based hierarchical enhancement (cf. Section 2.3).

We have found that, although the node-based approach [67] facilitates the implementation, it requires a tedious element selection procedure to avoid linear dependency problems usually found in the local model containing a small number of elements. As a consequence, this can diminish the robustness of the nodal *p*-enrichment method since the required minimum number of elements may keep growing if the polynomial order grows (see also Appendix A). The element-based approach [68], however, does not show such a shortcoming. For this reason, our *p*-version error estimation will only be based on the element-based *p*-extension.

One problem remains. Unlike the *h*-version error estimation [29], it is impossible to use a *portion* of a certain finite element within a patch. Instead, the *whole* element must be taken. Automatically set based on each node, a patch can be constructed by a set of surrounding elements, as illustrated in Figures 3.5 and 3.6. By this scheme, the global number of patches is fixed by the number of nodes. Each element is enriched by the number of patches that corresponds to the element type. For instance, contributions from three patches are combined to recover errors in a linear triangular element and contributions from four patches are used for a bilinear quadrilateral element.

3.5 Error assessment in nonlinear analysis

It has been proven that the Dirichlet-type error estimator [29] can be easily integrated in a conventional finite element program and it has been successfully applied to linear as well as nonlinear problems. In [28,42,78], the estimator has shown

---

[§]Note that *h*-version and *p*-version refer here to the error estimation procedure, not to a possible global enhancement of the discretisation that may follow the error estimation.
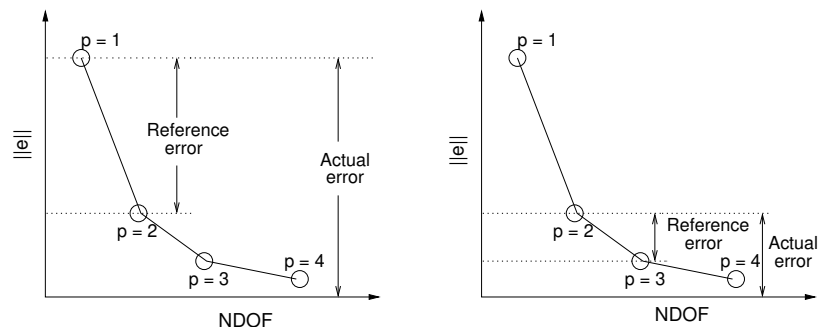
Figure 3.4 Reference error and actual error in the local *p*-enrichment scheme: the reference for linear elements (left) and for quadratic elements (right).
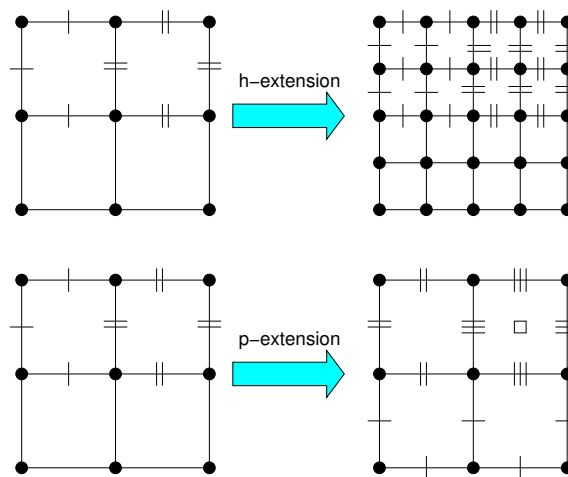

Figure 3.5 Reference mesh in the framework of *h*- and *p*- extensions for the problem with non-uniform higher-order elements. The short lines on element edges denote the edge modes and the square symbol denotes the internal mode.
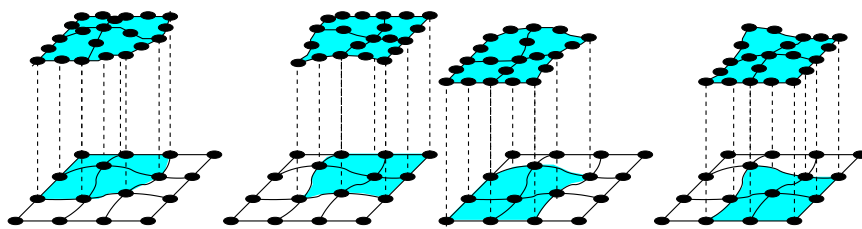

Figure 3.6 Node-based patches for error estimation based on local *p*-enrichment.

to provide a good lower-bound estimate in problems that exhibit *softening* either in the viscoplastic setting or in the nonlocal damage setting.

In nonlinear analysis, $\mathcal{B}(\cdot, \cdot)$ is a nonlinear form. As such, the Newton-Raphson iterative procedure is generally applied. Furthermore, the load is normally applied in a number of increments. Within an incremental-iterative scheme, solution of a nonlinear problem pertains to the solution of a sequence of linear problems. At the computational step (time step) $t-1$ to t, we can write the solution of the linearised problem as

$$\mathcal{B}^{\text{tang}}(\Delta u_{(h,p)}, v_{(h,p)})|_{(t-1:t)} = \Delta \mathcal{F}(v_{(h,p)})|_{(t-1:t)} \tag{3.21}$$

The correction based on nonlinear solution residuals is formed as

$$\mathcal{B}^{\text{tang}}(\delta u_{(h,p)}^{(i)}, v_{(h,p)})|_{(t)} = \mathcal{F}(v_{(h,p)})|_{(t)} - \mathcal{B}(u_{(h,p)}^{(i-1)}, v_{(h,p)})|_{(t)} \tag{3.22}$$

where $i = 1, 2, ..,$ number of iterations and $u_{(h,p)}$ is updated after each iteration

$$u_{(h,p)}|_{(t)} = u_{(h,p)}|_{(t-1)} + \Delta u_{(h,p)}|_{(t-1:t)} + \sum_i \delta u_{(h,p)}^{(i)}|_{(t)} \tag{3.23}$$

Upon convergence, at the end of a loading step, the error according to the introduction of Newton-Raphson procedure approaches zero, which implies

$$\mathcal{F}(v_{(h,p)})|_{(t)} \approx \mathcal{B}(u_{(h,p)}, v_{(h,p)})|_{(t)} \tag{3.24}$$

That is to say, the external force (i.e. the left-hand side of Eq. (3.24)) is at this stage balanced with the internal force (i.e. the right-hand side of Eq. (3.24)). At this equilibrium stage, we assume that the error from the nonlinear solution control is marginal. However, the discretisation error still needs to be measured. Again, the discretisation error is estimated by upgrading the existing discretisation and the upgraded discretisation is referred to as the reference mesh.

The discretisation error equation in nonlinear analysis can be formulated as

$$\mathcal{B}^{\text{tang}}(\delta e_{(\tilde{h},\tilde{p})}^{(i)}, v_{(\tilde{h},\tilde{p})})|_{(t)} = \mathcal{F}(v_{(\tilde{h},\tilde{p})})|_{(t)} - \mathcal{B}(u_{(h,p)}^{(i-1)}, v_{(\tilde{h},\tilde{p})})|_{(t)} \tag{3.25}$$

where $u_{(h,p)}^{(i)}$ is updated during the iteration

$$u_{(h,p)}^{(i)}|_{(t)} = u_{(h,p)}^{(i-1)}|_{(t)} + \sum_i \delta e_{(\tilde{h},\tilde{p})}^{(i)}|_{(t)} \tag{3.26}$$

and the error is accumulated as

$$e_{(\tilde{h},\tilde{p})}|_{(t)} = \sum_i \delta e_{(\tilde{h},\tilde{p})}^{(i)}|_{(t)} \tag{3.27}$$

The error computation is terminated as the tolerance is reached, that is, $\delta e \approx 0$. For simplicity, we compute our error quantity only once at the end of each computational step in this study, as it is assumed that the error quantity is much smaller than the unknown itself.

3.6 Some implementational aspects

3.6.1 Solution mapping

Discretising the continuous test function v with the finite element shape functions, the error can be obtained by solving the local discretised equations

$$\mathcal{B}_k(e_{(\tilde{h},\tilde{p})}, \boldsymbol{\phi}_{(\tilde{h},\tilde{p})}) = \mathcal{F}_k(\boldsymbol{\phi}_{(\tilde{h},\tilde{p})}) - \mathcal{B}_k(u_{(h,p)}, \boldsymbol{\phi}_{(\tilde{h},\tilde{p})}) \tag{3.28}$$

which may be better known in matrix form

$$\boldsymbol{K}^{(k)}_{(\tilde{h},\tilde{p})} \, \mathbf{e}^{(k)}_{(\tilde{h},\tilde{p})} = \boldsymbol{f}^{(k)}_{(\tilde{h},\tilde{p})} - \check{\boldsymbol{K}}^{(k)}_{(\tilde{h},\tilde{p})} \, \boldsymbol{a}^{(k)}_{(h,p)} \tag{3.29}$$

where

$$\boldsymbol{K}^{(k)}_{(\tilde{h},\tilde{p})} := \mathcal{B}_k(\boldsymbol{\phi}_{(\tilde{h},\tilde{p})}, \boldsymbol{\phi}_{(\tilde{h},\tilde{p})}) = \int\limits_{\Omega_k} (\boldsymbol{\nabla}\boldsymbol{\phi}_{(\tilde{h},\tilde{p})}) : \boldsymbol{D} : (\boldsymbol{\nabla}\boldsymbol{\phi}_{(\tilde{h},\tilde{p})}) \, \mathrm{d}\Omega_k \tag{3.30}$$

$$\boldsymbol{f}^{(k)}_{(\tilde{h},\tilde{p})} := \mathcal{F}_k(\boldsymbol{\phi}_{(\tilde{h},\tilde{p})}) = \int\limits_{\Omega_k} \boldsymbol{\phi}_{(\tilde{h},\tilde{p})} \cdot \boldsymbol{q} \, \mathrm{d}\Omega_k + \int\limits_{\Gamma_n \cap \partial\Omega_k} \boldsymbol{\phi}_{(\tilde{h},\tilde{p})} \cdot \boldsymbol{g} \, \mathrm{d}\Omega_k \tag{3.31}$$

and

$$\check{\boldsymbol{K}}^{(k)}_{(\tilde{h},\tilde{p})} := \mathcal{B}_k(\boldsymbol{\phi}_{(h,p)}, \boldsymbol{\phi}_{(\tilde{h},\tilde{p})}) = \int\limits_{\Omega_k} (\boldsymbol{\nabla}\boldsymbol{\phi}_{(\tilde{h},\tilde{p})}) : \boldsymbol{D} : (\boldsymbol{\nabla}\boldsymbol{\phi}_{(h,p)}) \, \mathrm{d}\Omega_k \tag{3.32}$$

Alternatively, to make use of the existing stiffness matrix **K** for error computation, one can use an equivalent form as

$$\boldsymbol{K}^{(k)}_{(\tilde{h},\tilde{p})} \, \mathbf{e}^{(k)}_{(\tilde{h},\tilde{p})} = \boldsymbol{f}^{(k)}_{(\tilde{h},\tilde{p})} - \boldsymbol{K}^{(k)}_{(\tilde{h},\tilde{p})} \, \check{\boldsymbol{a}}^{(k)}_{(\tilde{h},\tilde{p})} \tag{3.33}$$

where $\check{\boldsymbol{a}}^{(k)}_{(\tilde{h},\tilde{p})}$ is interpolated using the original shape functions $\boldsymbol{\phi}_{(h,p)}$ that correspond to the new set of nodal positions $\boldsymbol{x}_{(\tilde{h},\tilde{p})}$ in the enriched local problem, yielding

$$\check{\boldsymbol{u}}^{(k)}_{(\tilde{h},\tilde{p})} := \boldsymbol{\phi}_{(h,p)}(\boldsymbol{x}_{(\tilde{h},\tilde{p})}) \cdot \boldsymbol{a}^{(k)}_{(h,p)} \tag{3.34}$$

The form in Eq. (3.33) is better than that in Eq. (3.29) in terms of data storage. Also, computation of another integral for a modified stiffness matrix costs more than interpolation of the displacement field $\boldsymbol{u}$. However, the form in Eq. (3.29) provides more flexibility in selecting the reference mesh. For the hierarchical approach, use of the form in Eq. (3.33) is natural. This is because $\mathcal{V}_{(h,p)} \in \mathcal{V}_{(\tilde{h},\tilde{p})}$, i.e. the existing shape functions are preserved and completely separate themselves from the set of additional degrees of freedom. Therefore, at the additional degrees of freedom, a zero contribution is simply added to form the interpolated set of the primary unknowns $\check{\boldsymbol{a}}^{(k)}_{(\tilde{h},\tilde{p})}$.
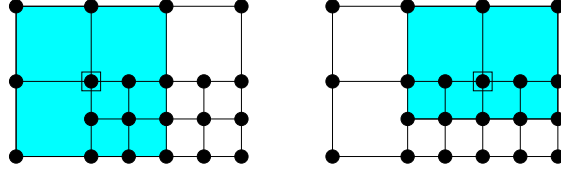
Figure 3.7 Examples of patch selection in existence of hanging nodes.

3.6.2 Irregular element connectivity

The requirement of overlapping patches may complicate the patch selection process. An example is when *hanging* nodes exist. In such a case, the finite element interpolation requires special constraints at the hanging nodes so that the shape functions over the irregular partitioning are compatible.

It is obvious in the original *h*-version, where a patch consists of a part of each element in the neighbourhood, that it is merely impossible to obtain the patching scheme described in Figure 3.2. However, as an alternative, the patches which constitute all neighbouring elements (as in the *p*-version, cf. Figure 3.6) may be selected to include all involved parent-child relations (cf. Figure 3.7), i.e. the local problems have to allow the element connecting to the hanging node to be in the part as well.

3.7 Performance analyses

One-dimensional problem

The first numerical example is a one-dimensional problem described by the ordinary differential equation

$$-\frac{d^2u}{dx^2}(x) = 6x^2 - 3x \tag{3.35}$$

in $\Omega =]0,1[$ and the Dirichlet boundary conditions are prescribed as $u(0) = u(1) = 0$ (see also Figure 1.1). The analytical solution for this problem is

$$u(x) = \frac{(x^3 - x^4)}{2} \tag{3.36}$$

In order to investigate performances of the error estimator, we start with measuring the capability of the local *p*-version in estimating the error in linear elements, in comparison with the local *h*-version. In this problem, each patch is selected based on one element, and, for the same reason as in [29], we assume that there exist only interior residuals due to the superconvergence property of the problem.

In Figure 3.8, the local reference discretisation is varied from order 2 to order 6. Based on *h*-extension, this means that the patch is subdivided into 2 to 6 subelements. On the other hand, in the *p*-scheme, the order of interpolation is varied from
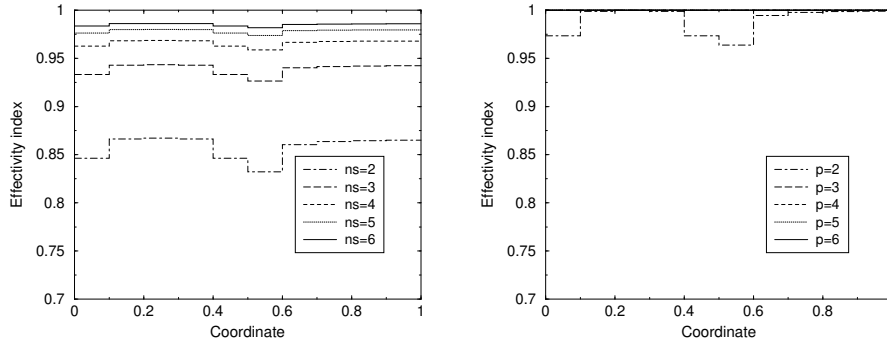
Figure 3.8 Distribution of local effectivity index in the 1D problem with a 10-element discretisation, based on h-version (left) and p-version (right) error estimates and varying order of local refinement from 2 to 6.

$p = 2$ (quadratic) to $p = 6$ (hexic) interpolations. The quality of the estimates is measured through an *effectivity index*, which is defined as

$$\theta_e := \frac{\|\boldsymbol{e}_{(\tilde{h},\tilde{p})}\|}{\|\boldsymbol{e}\|} \tag{3.37}$$

i.e. the ratio of the estimated and exact error, $e := u - u_{(h,p)}$, measured in the energy norm.

In Figure 3.8, the high effectivity index in all elements in the local *p*-version, especially when the local enrichment order is higher than two, reveals that the local *p*-version performs better than the local *h*-version. This is partly due to the fact that the exact solution is in the polynomial form and thus the higher-order polynomial interpolation can capture the solution better than the element subdivision. It is shown in Figure 3.9 that the estimators perform better in the more refined finite element discretisation as, at the same degree of local refinement, the local reference discretisation provides a closer representation of the exact solution.

One can notice that in Figure 3.10, where we use an odd number of elements, the local effectivity index appearing in the middle elements shows how poorly the estimation performs, in both *h*-version and *p*-version, especially when applying the second order of local refinement. The improvement in effectivity index when going to third order local refinement is significant. Moreover, Figure 3.11 reveals that the error estimate of the middle element can become worse as the number of degrees of freedom increases. We observe that, at the middle point (i.e. $x = 0.5$), the curvature of the solution function changes from being convex to being concave and that should be the main reason for the poor error estimates in the element including such a transition point. Apparently, the smaller the middle element is, the bigger the relative effect of the curve transition is on the error estimate. However, the failure in the local error estimation does not affect the global measurement significantly.
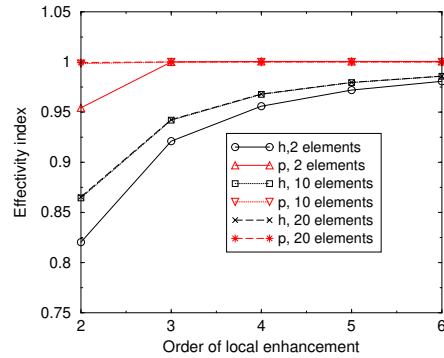
Figure 3.9 The global effectivity index obtained by different orders of h- and p-local refinement schemes in the 1D example with 2-element, 10-element and 20-element discretisations.
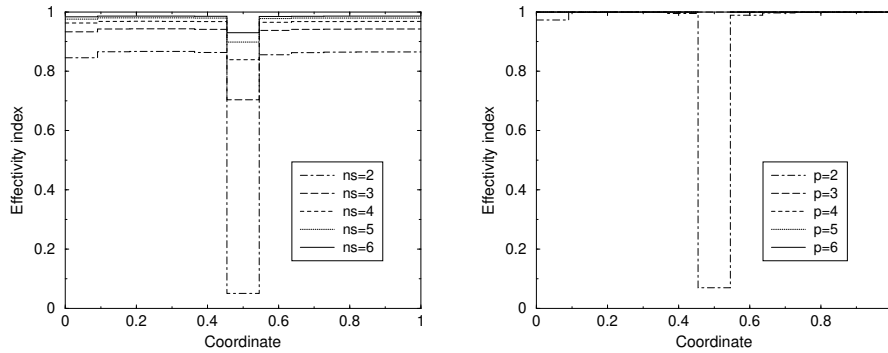

Figure 3.10 Distribution of local effectivity index in the 1D problem with an 11-element discretisation, based on h-version (left) and p-version (right) error estimates and varying order of local refinement from 2 to 6. The number of subelements is denoted as "ns".
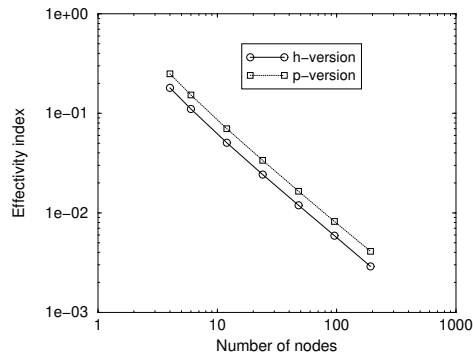

Figure 3.11 Failure of the error estimates due to inadequate representation (i.e. local refinement of order 2) of the actual solutions in the 1D problem modelled by an odd number of elements. The effectivity indices shown are those of the middle elements only.

Patch test

The error estimation is now applied to a two-dimensional problem domain. The quadratic stress patch is modelled as in Figure 3.12 (left). A patch of 1 mm × 1 mm is modelled in the plane stress condition with Young's modulus $E = 1$ MPa and Poisson's ratio $\nu = 0.25$. We model this problem by setting the non-homogeneous Dirichlet conditions of the exact solution

$$u_x(x, y) = \frac{x^2 y}{2E} + \frac{\nu y^3}{6E} - \frac{(1 + \nu)y^3}{3E} \tag{3.38}$$

$$u_y(x, y) = -\frac{\nu x y^2}{2E} - \frac{x^3}{6E} \tag{3.39}$$

on the whole boundary. The finite element solution is obtained by a linear interpolation.

We start our investigation with the local *h*-extension scheme. The error functions plotted in Figure 3.13 apparently depend on how the patches overlap in the local *h*-refinement scheme [29]. In terms of performance, Figure 3.14 (left) reveals a better estimation and effectivity in case of all local *p*-enrichments. In this example, a triangular element is subdivided into 27 subelements, adding 68 degrees of freedom for the elemental computation in the *h*-version. Conversely, a set of quadratic polynomials is added to each element in the *p*-scheme, adding 3 degrees of freedom per element. If considering a patch of maximum 6 elements, we need to solve only 38 equations in each local computation. Yet, the global error estimate based on the local *p*-enrichment, in comparison to the local *h*-refinement, shows a closer estimation to the exact error computed based on the analytical solution in Eqs. (3.38) and (3.39).

The quadratic enrichment is, of course, not the most accurate choice for estimating the error in linear elements. It is, however, sufficient since the errors are much smaller in comparison to the original linear interpolation. Furthermore, enriching to a higher-order polynomial does not provide a great improvement to the error estimation, considering the fact that the number of local equations is greatly increased. However, it still provides the possibility to set the local reference solutions
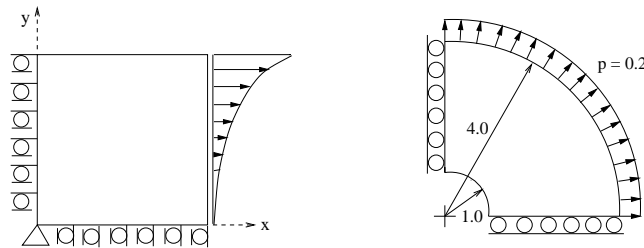

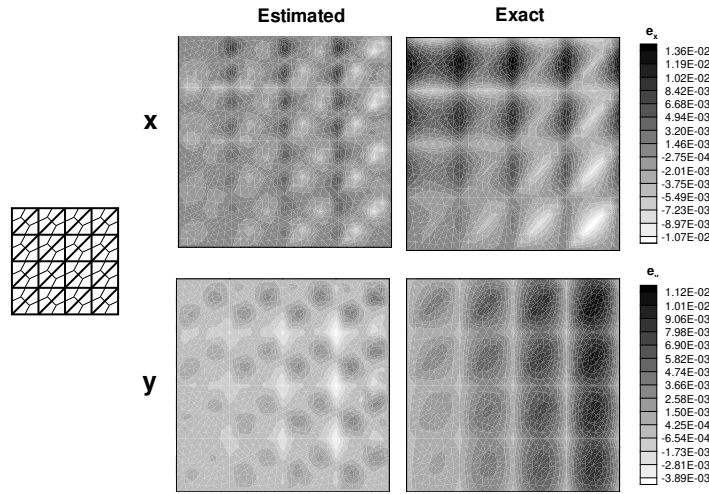**Figure 3.12** The patch test (left) and the disk test (right).

Figure 3.13 The estimated and exact error functions in x- and y-directions of the quadratic patch test. The mesh (thick line) and the overlapping patches (thin line) are shown on the left.

from higher-order polynomials, if highly accurate error solutions are necessary. In Figure 3.14 (right), the global error estimates based on different orders of interpolation are compared.

Before moving further to investigate the performance of error estimation in higher-order elements, let us consider the actual error trend of the quadratic stress patch test, as shown in Figure 3.15. Evidently, the exact solution can be obtained by applying cubic interpolation in the finite element modelling, but not by the same number of equations in the element subdivision scheme. Comparing the same number of degrees of freedom, the polynomial enrichment shows a more efficient trend. For this reason, the local *p*-enrichment scheme is more efficient than the local *h*-subdivision. However, if cubic polynomials, which represent the exact solution, are employed, why can the estimator not exactly predict the actual error? (See Figure 3.14 (right)).

To provide an explanation, we refer back to the basic concept in Subsection 3.3.2, i.e. the error is prescribed as zero on $\partial\Omega_k \backslash (\partial\Omega_k \cap \Gamma_n)$. This means, the error on the boundary, which is linearly-interpolated, is assumed null, while the real error should in fact be cubically-interpolated. A good proof for this point is shown by the local error plot in Figure 3.16, where the estimated and the actual error functions along the boundary differ. Based on this interpretation, the estimator can never recover the exact error unless the primary solution on the boundary contains the polynomial order that interpolates the exact solution.

We investigate further the error estimate in higher-order elements. Following the
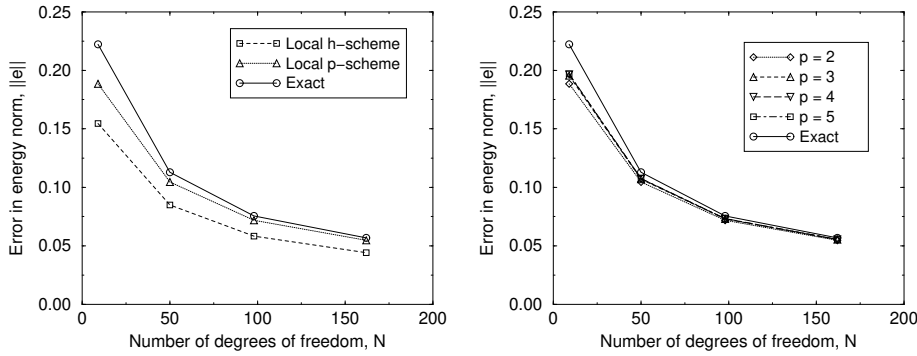
Figure 3.14 Comparison of estimated error in energy norm by different local enrichment schemes in the quadratic stress patch test.
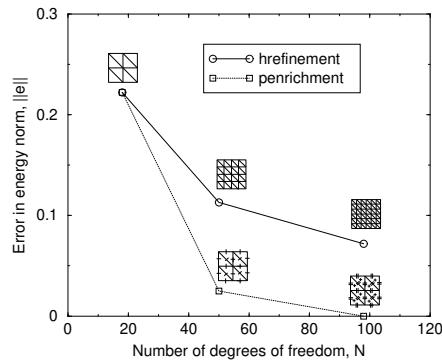

Figure 3.15 Convergence analysis of quadratic stress patch problem based on global *h*-refinement and global *p*-enrichment.

same strategy, a set of higher-order terms is added for each displacement mode. As mentioned earlier, for each polynomial order of the displacements, the error is estimated with one higher polynomial order. The quadratic stress patch is again tested and the results are shown in Table 3.2, where the global performance is measured in terms of the effectivity index. As shown, the effectivity index is greatly reduced when the element polynomial order increases. This is not a surprising phenomenon in this problem, which consists of non-homogeneous Dirichlet boundary conditions. In this case, it is rather difficult to directly impose the exact displacements along the boundary. The higher-order polynomials along the boundary edges are dropped out and the global boundary conditions are, instead, represented by linear interpolation. Obviously, the reference model, in this case, cannot resemble the actual model, thus leading to failure of the error estimation.

If the prescribed displacements are imposed exactly, we should be able to obtain

Estimated   Exact

Figure 3.16 Comparison between the estimated error function(left) and the exact error function(right), based on analytical solutions, in the x-(top) and y-(bottom) direction of the linearly-interpolated patch.

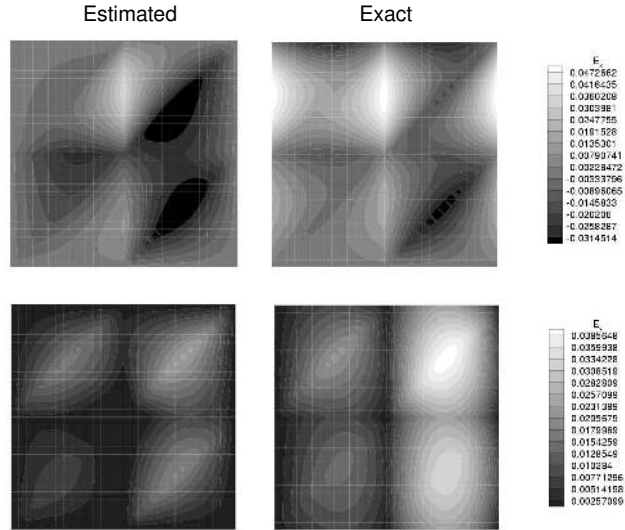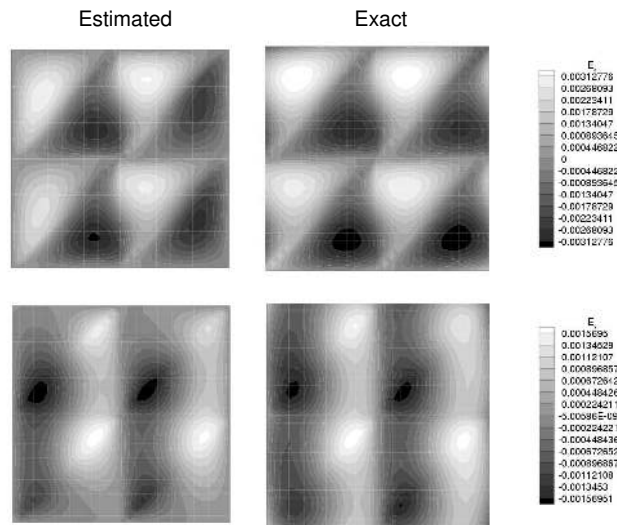Estimated   Exact

Figure 3.17 Comparison between the estimated error function(left) and the exact error function(right), based on analytical solutions, in the x-(top) and y-(bottom) direction of the quadratically-interpolated patch.

a good error estimate in the patch problem. The proper enforcement of boundary conditions is made possible by adding Lagrange multipliers or penalty terms in the variational formulation. As a result, Eq. (3.6) is replaced by

$$\mathcal{B}^*(\boldsymbol{u}_{(h,p)}, \boldsymbol{v}_{(h,p)}) = \mathcal{L}(\boldsymbol{v}_{(h,p)}) \qquad \forall \boldsymbol{v}_{(h,p)} \in \mathcal{V}_{(h,p)} \tag{3.40}$$

where $\mathcal{B}^*(\cdot, \cdot)$ is equivalent to the original $\mathcal{B}(\cdot, \cdot)$, further including the already pre-scribed Dirichlet boundary conditions. The displacements on the boundary edges can then be properly imposed. As a consequence, the error estimation with penalty functions provides a better estimate than the error estimation with the direct imposition model, as shown in Table 3.3. The errors are better estimated in the higher-order elements because, firstly, the reference model contains a better solution, i.e. the reference error becomes closer to zero. Secondly, the errors on the global Dirichlet boundary become closer to zero, thus closer to the basic assumption of this error estimation.

Figures 3.16 and 3.17 present the local error distribution for the 8-element patch. The comparison between the estimated errors (left) and the actual errors (right) yields very similar profiles, especially in the case of higher-order elements. It can

model	DOFs	reference	$\|e\|$	$\|e\|^{\text{ex}}$	θ_e
p = 1	18	p = 2	0.1885	0.2223	84.8%
p = 2	50	p = 3	0.0522	0.1176	44.4%
p = 3	98	p = 4	0.0199	0.1051	18.9%

Table 3.2 Global error in energy norm obtained in quadratic stress patch test. Dirichlet boundary conditions are imposed by direct imposition scheme in the standard *p*-finite element modelling.

model	DOFs	reference	$\|e\|$	$\|e\|^{\text{ex}}$	θ_e
p = 1	18	p = 2	0.1885	0.2223	84.8%
p = 2	50	p = 3	0.0234	0.0249	94.0%
p = 3	98	p = 4	0.0000	0.0000	100.0%

Table 3.3 Global error in energy norm obtained in quadratic stress patch test. Dirichlet boundary conditions are imposed by penalty formulation in the standard *p*-finite element modelling.

model	DOFs	reference	$\|e\|^{\text{loc}}$	$\|e\|^{\text{glob}}$
p = 1	50	p = 2	0.0872	0.0897
p = 2	162	p = 3	0.0176	0.0176
p = 3	338	p = 4	0.0101	0.0101
p = 4	578	p = 5	0.0087	0.0087

Table 3.4 Global error in energy norm obtained in circular disk problem.

be seen that, since the displacements are prescribed along the whole domain, the errors are forced to zero. This, of course, affects the error distributions in the boundary zones. The error contour in the case of quadratic elements in Figure 3.17 apparently reveals a closer estimate to the actual error distribution. Even better, the estimate in the cubic elements shows zero error in the whole domain, which is the same as the exact error plot.

Disk problem

Finally, consider a disk being pulled uniformly at the outer boundary, as shown in Figure 3.12 (right). We neglect the discretisation error due to the geometrical modelling by assuming that the disk is not circular but composed of linear boundary segments. In this example, we would like to study the effect of the missing global part in local error estimation scheme. To this end, we compare the estimated error ($\|e\|^{\text{loc}}$) and the error obtained by the difference between two sets of global finite element solutions, the original and the reference, ($\|e\|^{\text{glob}}$), defined as

$$
\|e\|^{\text{glob}} = \left[\int_{\Omega} \left(\varepsilon_{(p+1)} - \varepsilon_{(p)} \right)^{T} : \boldsymbol{D} : \left(\varepsilon_{(p+1)} - \varepsilon_{(p)} \right) \mathrm{d}\Omega \right]^{\frac{1}{2}}
\tag{3.41}
$$

where $\varepsilon_{(p)}$ and $\varepsilon_{(p+1)}$ represent the set of strain components computed using (p)-order and $(p+1)$-order interpolation, respectively. In other words, Eq. (3.41) provides the error computed via the global stiffness matrix, and this *reference error* is used to calibrate the locally computed error estimate.

The global results in Table 3.4 exhibit good agreement between the errors based on the local computation $\|e\|^{\text{loc}}$ and those based on the global computation $\|e\|^{\text{glob}}$, especially in the case with higher-order elements. This suggests that, in this problem, error solutions from a coarse mesh may be sufficiently recovered by the local computation, which is definitely much cheaper than the global computation underlying Eq. (3.41). Also, this implies that the missing global part is of a marginal magnitude. The local contour plots in Figure 3.18 again show a good agreement.

It should be noted here that, in order to obtain a good error estimate, a refined local patch must resemble the (reference) refined global system as much as possible. For example, a circular disk can be modelled by either prescribing displacements or prescribing forces along its outer boundary. The displacement control model (Figure 3.19 (B)) provides different error patterns than the force control model (Figure 3.19 (C)), since we assume exact displacements along the global Dirichlet boundary, i.e. $e(\mathbf{x}) = 0, \forall \mathbf{x} \in \Gamma_d$. This means, displacements or forces must be correctly prescribed on the local boundary and must resemble the real globally-refined model. In this example, if one fails to correctly prescribe the interpolated forces along the outer boundary, instead of Figure 3.19(C), an obviously wrong estimation can be obtained, as shown in Figure 3.19 (D).
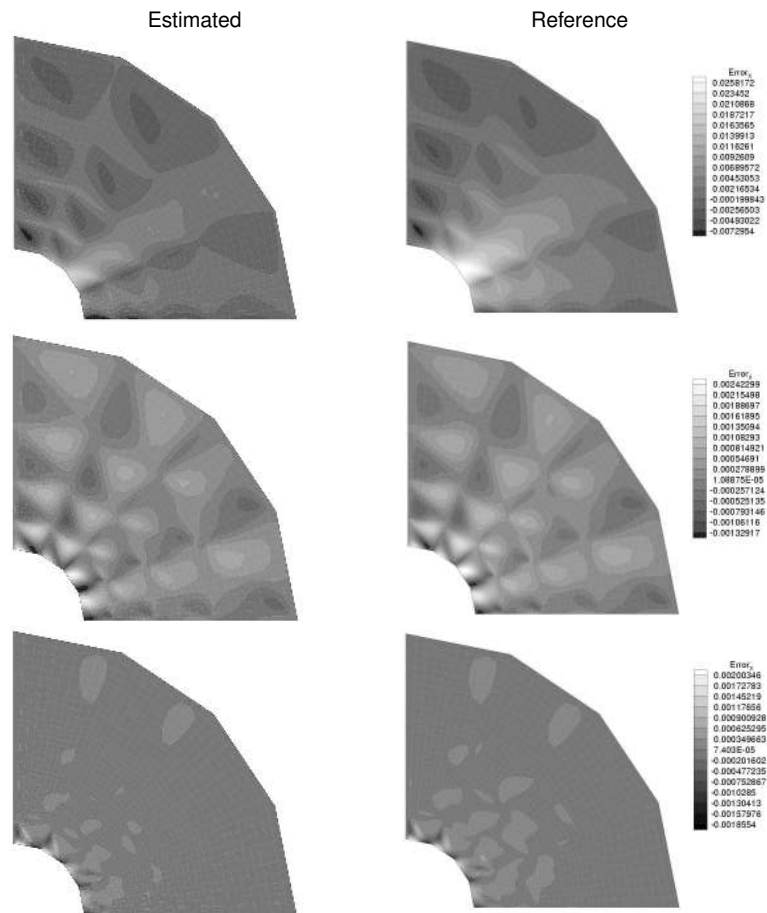
Figure 3.18 Comparison between the estimated error function(left) and the reference error function(right), based on solutions from a higher-order interpolation scheme, in the x-direction of the circular disk. The polynomial basis is varied from p = 1 to p = 3 as shown in uppermost to lowermost subfigures.
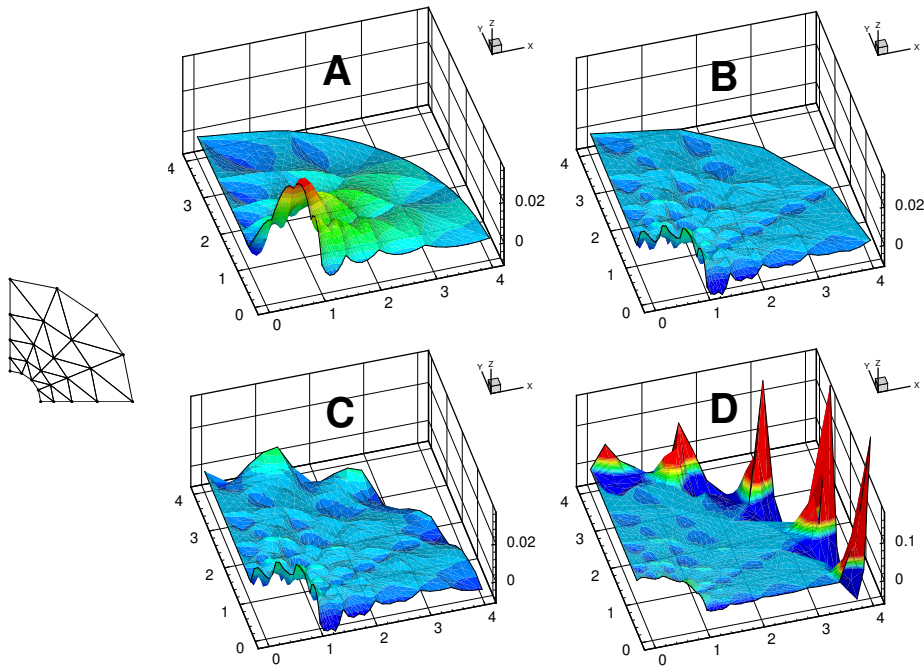
Figure 3.19 Functions of error in x-displacement of the circular disk problem: (A) expected error, (B) estimated error in the prescribed displacement model, (C) estimated error based on the prescribed force model and (D) estimated error based on prescribed force model (improper modelling).

3.8 Remarks

In this chapter, the error estimator to be used in this thesis has been presented. The method, based on enhancing each local domain with a higher-order polynomial interpolation, has shown a good performance in estimating the error. Even then, the quality of the error estimate depends on the finite element discretisation. The error estimate of the element where there is a change of the solution curvature can be very poor, and a sufficiently-high order of local refinement/enrichment should be applied. It has also been found that the estimate is very sensitive to how the reference mesh is set, especially the description of the boundary conditions. Without a correct reference discretisation, the estimated error can be far from the exact error.

The performance analyses carried out in this chapter do deliberately not include nonlinear problems. Our intention is to examine the quality of our *p*-version error estimator (cf. Section 3.4) to be used throughout this research work. Thus the test problems are chosen to be simple, as they can provide a basic understanding about the error estimators, in implementational as well as performance aspects.

Error estimation for specific goals

In the previous chapter, we computed the error measured in the energy norm. Although this gives a good indication of the overall error, it may also be relatively insensitive to certain local values of the state variables and their accuracy. In other words, a slight error in energy norm does not always guarantee that the local quantities of interest, such as stresses or damage profile in a critical region, are sufficiently accurate. For the problems where there are some specific goals in mind, estimation of error of such specific quantities can provide more relevant information for the adaptive finite element process.

Therefore, in this chapter, we concentrate on estimating the error in chosen quantities of interest in any local domain. At some extra cost, the so-called *goal-oriented error estimation* can be set and applied at the end of any computational steps in the finite element analysis. Techniques for the goal-oriented error measurement make use of the concept of *influence function*, which in the context of error estimation indicates how the discretisation error affects the specific quantity of interest. That is, to measure such an error quantity, an additional boundary value problem needs to be solved. This so-called *dual (adjoint) problem* is constructed in a dual argument setting and makes use of Maxwell's *reciprocal theorem*. Once the dual problem is solved, the discretisation residuals can be redistributed according to the influence function to provide a proper error measure to the selected goal quantity.

In this chapter, the above mentioned ingredients will be presented to provide an introduction to the goal-oriented error estimation in linear and nonlinear analyses. In addition, some error measures will be studied and selected for use in the next chapter, which involves setting of adaptive and optimality criteria to be used in the rest of the thesis.

4.1 Quantities of interest

As mentioned earlier, accuracy of the conventional energy norm does not always guarantee accuracy of a local quantity of engineering significance. Especially in problems with some specific goals, relevant quantities must be accurately achieved. For instance, in linear elastic fracture mechanics, the most important variable may be the stress intensity factor, which heavily relies upon the accurate computation of the $\mathcal{J}$-integral [77] in the vicinity of the crack tip. In many occasions, we also find that engineers would prefer to have an accurate calculation of stress in some critical regions, rather than the accurate overall energy norm measurement. These quantities can be considered as the *quantities of interest* and, hence, achieving high accuracy of these quantities is a *specific goal* in finite element analysis.

Basic rules

The most crucial point of the goal-oriented error measurement is how to choose the quantity of interest such that it can be analysed straightforwardly. Basically, the choice depends on the model problem to be analysed and may be chosen in the forms of domain integrals, contour integrals or pointwise quantities. Two important features of the quantity of interest to be used in a duality argument are as follows.

- The quantity of interest must be a function of the primary unknown u, such that it can be defined as $Q(u)$.

  It is convenient to choose the quantity of interest as an as-simple-as-possible function of u. For instance, in linear elastic fracture analysis, it is efficient to express the stress intensity factor $\mathcal{K}$ as $\mathcal{K}(u) = \mathcal{K}(\mathcal{J}(u))$ and use the $\mathcal{J}$-integral $\mathcal{J}(u)$ as the quantity of interest [38, 39, 81, 82] – any error measurement in $\mathcal{J}(u)$ can then be translated straightforwardly in error measurement of $\mathcal{K}(u)$.

- Ideally, $Q(u)$ is a linear functional; if not, one of the linearization techniques [51, 80, 83] should be applied. Essentially, the linear functional $Q(u)$ must be bounded [4, 75]. The boundedness of a functional guarantees convergence of the solution, thus providing a meaningful norm measurement.

Some examples

- *Primary unknown*

  The simplest form of the quantity of interest is a function of the primary unknown itself, not its derivatives. For instance, one may choose the quantity of interest in the form

$$Q(u) = \int_{\Omega_s} \varphi(\boldsymbol{x})\, u(\boldsymbol{x})\, \mathrm{d}\Omega \tag{4.1}$$

which is the integral of the weighted primary unknown u in the local region Ω_s. Choosing such a quantity as the quantity of interest implies that we control, in the local region, the primary unknown in an averaged sense. In other words, the quantity at each point in the selected domain Ω_s may not be of interest but the average of all is. The weight function $\varphi(\boldsymbol{x})$ may be constant or spatially varied.

For a pointwise measure at a point x_0, where the quantity of interest

$$Q(u) = u(x)|_{x=x_0} = u(x_0) \tag{4.2}$$

is not bounded on the space of admissible displacements (H_1-space), the mollification technique may be applied to slightly smear out the quantity of interest in a small region surrounding the point of interest instead of the single point itself [75]. The quantity of interest (cf. Eq. (4.2)) may then be written in an averaged/mollified format. Considering Eq. (4.1), $\varphi(\boldsymbol{x})$ can be chosen as the mollifying function [75] to ensure the goal quantity of a bounded linear functional.

- *Derivatives of the primary unknown*

  In strength computations, it is important to ensure accuracy of the stresses in the specific region of interest. This quantity may be infinite in some regions, such as at the crack tip or at the re-entrant corner. Hence, the quantity of interest in a weighted average form (in Ω_s) of the stress $\sigma(u)$

  $$Q(u) = \int_{\Omega_s} \varphi(\boldsymbol{x}) \, \boldsymbol{\sigma}(u(\boldsymbol{x})) \, \mathrm{d}\Omega \tag{4.3}$$

  may be chosen.

4.2 Setting of duality argument

Measurement of the error in a quantity of interest requires solving another set of equations, besides those from the primal problem mentioned earlier in the previous chapters. This so-called dual or adjoint problem is different from the standard primal problem as it is defined in terms of influence function (Green's function), instead of the primary unknown (e.g., displacement). Basically, the influence function measures the influence of any external loading to the local quantity of interest. However, if the function is applied in the framework of error estimation, it is possible to determine the influence of the discretisation residuals to the quantity of interest.

$f(x_q)$

$u(x_p; f(x_q))$

$f(x_p)$

$u(x_q; f(x_p))$

Figure 4.1 Illustration of the reciprocal theorem

4.2.1 The influence function

To explain the basic idea about the dual problem, it is important to understand what the influence function is. Let us consider the quantity of interest

$$Q(u) = u(x_p) \tag{4.4}$$

defined as the pointwise displacement at point x_p. To measure the magnitude of the displacement at x_p when the system is subjected to a unit point load at x_i, we can define the value of the influence function of the quantity $Q(u)$ at point x_i as

$$w(x_i) := u(x_p; \delta_{x_i}) \tag{4.5}$$

i.e., the displacement at point x_p subjected to a unit force at point x_i, denoted as δ_{x_i}. If there exists only one point load at x_i acting on the structure, Eq. (4.4) can be written as

$$Q(u) = u(x_p) = f(x_i)\, w(x_i) \tag{4.6}$$

where the influence function w has a dimension of length per unit force.

In order to determine the influence function at any point $w(x)$ in a much simpler way, the *Maxwell's reciprocal theorem* may be applied. Based on the equivalent work done, the theorem states that

> *"The work done by one set of forces in undergoing the corresponding displacements caused by the second set of forces is equal to the work done by the second set of forces in undergoing the corresponding displacements caused by the first set of forces".*

This is explained in Figure 4.1, where we can set

$$f(x_p)\, u(x_p; f(x_q)) = f(x_q)\, u(x_q; f(x_p)) \tag{4.7}$$

If the forces are of a unit value, Eq. (4.7) becomes

$$u(x_p; \delta_{x_q}) = u(x_q; \delta_{x_p}) \tag{4.8}$$

That is, the displacement at coordinate x_p due to a unit force at coordinate x_q is equal to the displacement at coordinate x_q due to a unit force at coordinate x_p.

Let us go back to our quantity of interest $Q(u) = u(x_p)$ (cf. Eq. (4.4)). Applying the reciprocal theorem to Eq. (4.5), the influence function at any point x in the problem domain reads

$$w(x) = u(x; \delta_{x_p}) \tag{4.9}$$

That implies solving the problem only once in order to obtain the influence function of the quantity of interest $Q(u)$.

4.2.2 The dual problem

Generally, the quantity of interest may be written in terms of the influence function, denoted as w for a multi-dimensional problem. Subjected to a set of body forces q and tractions g, the quantity of interest is represented in a general form

$$Q(u) = \int_\Omega q \cdot w \, d\Omega + \int_{\Gamma_n} g \cdot w \, d\Gamma \tag{4.10}$$

Recalling the right-hand side of the primal problem (cf. Eq.(2.3)), we can write

$$Q(u) = \mathcal{F}(w) = \mathcal{B}(u, w) \tag{4.11}$$

Thus the influence function w can be determined by solving the *dual problem*

$$\mathcal{B}(v, w) = Q(v) \qquad \forall v \in \mathcal{V} \tag{4.12}$$

where $\mathcal{V} := \{v \in H^1(\Omega); v = 0 \text{ on } \Gamma_d\}$, and the boundary condition is set as $w = 0$ on Γ_d.

4.3 Goal-oriented error estimation

The influence function obtained from solving the dual problem (cf. Eq. (4.12)) can be regarded as a weight function to transfer the *loading* to a specific quantity of interest and is also known as the *extractor*. In this context, this *loading* may refer to a contribution of the external forces as aforementioned in the previous chapter, as well as a contribution of the discretisation residuals from the finite element computation. The latter is what we focus on in this section.

4.3.1 Setting of error in the goal quantity

As our goal is to assess the error in a specific quantity, it is necessary to give its clear definition. Considering the quantity of interest $Q(u)$, we can define the discretisation error of this quantity as
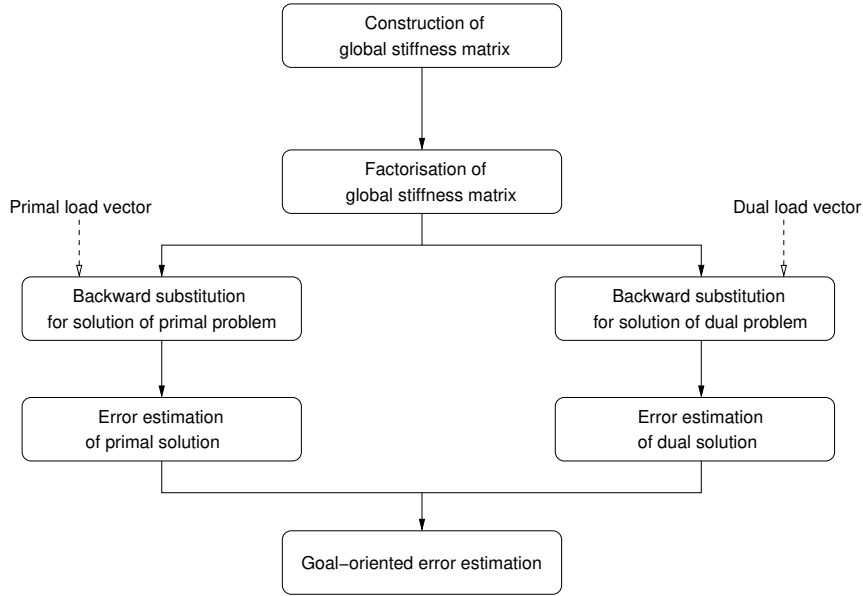
$$\mathcal{E} := Q(u) - Q(u_{(h,p)}) \tag{4.13}$$

Figure 4.2 Goal oriented error estimation procedure.

If $Q(\boldsymbol{u})$ is a linear functional, the error of the goal-oriented quantity can be rewritten as

$$
\begin{aligned}
\mathcal{E} = Q(\boldsymbol{u}) - Q(\boldsymbol{u}_{(h,p)}) &= Q(\boldsymbol{u} - \boldsymbol{u}_{(h,p)}) \\
&= Q(\boldsymbol{e})
\end{aligned}
\tag{4.14}
$$

where $Q(\boldsymbol{e})$ denotes the discretisation error of the finite element analysis (primal problem) measured in the quantity of interest. Recalling Eq. (4.12), $Q(\boldsymbol{e})$ can be obtained by selecting $\boldsymbol{v} = \boldsymbol{e}$. The equation then becomes

$$
Q(\boldsymbol{e}) = \mathcal{B}(\boldsymbol{e}, \boldsymbol{w})
\tag{4.15}
$$

which implies that the influence function acts as a weight function for distribution of the discretization error $\boldsymbol{e}$ to the quantity Q in an energy norm measure.

In finite element analysis, Eq. (4.15) is replaced by the discretised system of equations. The quantity of the error, where $\boldsymbol{e} \approx \boldsymbol{e}_{(\tilde{h},\tilde{p})}$, becomes

$$
\begin{aligned}
Q(\boldsymbol{e}_{(\tilde{h},\tilde{p})}) &= \mathcal{B}(\boldsymbol{e}_{(\tilde{h},\tilde{p})}, \boldsymbol{w}_{(\tilde{h},\tilde{p})}) \\
&= \mathcal{B}(\boldsymbol{e}_{(\tilde{h},\tilde{p})}, \boldsymbol{w}_{(h,p)} + \boldsymbol{\epsilon}_{(\tilde{h},\tilde{p})}) \\
&= \mathcal{B}(\boldsymbol{e}_{(\tilde{h},\tilde{p})}, \boldsymbol{w}_{(h,p)}) + \mathcal{B}(\boldsymbol{e}_{(\tilde{h},\tilde{p})}, \boldsymbol{\epsilon}_{(\tilde{h},\tilde{p})}) \\
&= \mathcal{B}(\boldsymbol{e}_{(\tilde{h},\tilde{p})}, \boldsymbol{\epsilon}_{(\tilde{h},\tilde{p})})
\end{aligned}
\tag{4.16}
$$

Here, $\boldsymbol{\epsilon}$ denotes the discretisation error in the dual problem. It is worth noting that $\mathcal{B}(\boldsymbol{e}, \boldsymbol{w}_{(h,p)}) = 0, \forall \boldsymbol{w}_{(h,p)} \in \mathcal{V}_{(h,p)}$ in order to satisfy the Galerkin's orthogonality property of the finite element analysis[*].

Figure 4.2 shows the procedure in the goal-oriented error estimation. Two global sets of equations, namely the primal and the dual problems, are solved. However, one can utilise the factorised global stiffness matrix, which is formed during the solving of the primal problem, also in the dual problem. This factorisation, in fact, is the main computation in solving the global equations. Thus, back-substitution to the factorised matrix to obtain the solution of the dual problem can save considerable computational time.

4.3.2 Error assessment in the dual problem

As an ingredient in the goal-oriented error estimation, the error from solving the dual problem

$$\boldsymbol{\epsilon} := \boldsymbol{w} - \boldsymbol{w}_{(h,p)} \tag{4.17}$$

must be estimated. Consider the discretised dual problem

$$\mathcal{B}(\boldsymbol{v}_{(h,p)}, \boldsymbol{w}_{(h,p)}) = Q(\boldsymbol{v}_{(h,p)}) \qquad\qquad \forall \boldsymbol{v}_{(h,p)} \in \mathcal{V}_{(h,p)} \tag{4.18}$$

the error of the discretised dual problem can be approximated as

$$\boldsymbol{\epsilon} \approx \boldsymbol{\epsilon}_{(\tilde{h},\tilde{p})} = \boldsymbol{w}_{(\tilde{h},\tilde{p})} - \boldsymbol{w}_{(h,p)} \tag{4.19}$$

Similar to the error estimation in the physical (primal) framework, the error of the dual problem can be computed from

$$\begin{aligned} \mathcal{B}(\boldsymbol{v}_{(\tilde{h},\tilde{p})}, \boldsymbol{\epsilon}_{(\tilde{h},\tilde{p})}) &= \mathcal{R}^w(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) \\ &= Q(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) - \mathcal{B}(\boldsymbol{v}_{(\tilde{h},\tilde{p})}, \boldsymbol{w}_{(h,p)}) \quad \forall \boldsymbol{v}_{(\tilde{h},\tilde{p})} \in \mathcal{V}_{(\tilde{h},\tilde{p})} \end{aligned} \tag{4.20}$$

Again, instead of solving the equation globally, the local error equations set as

$$\mathcal{B}_k(\boldsymbol{v}_{(\tilde{h},\tilde{p})}, \boldsymbol{\epsilon}_{(\tilde{h},\tilde{p})}) = Q_k(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) - \mathcal{B}_k(\boldsymbol{v}_{(\tilde{h},\tilde{p})}, \boldsymbol{w}_{(h,p)}) \quad \forall \boldsymbol{v}_{(\tilde{h},\tilde{p})} \in \mathcal{V}_{k(\tilde{h},\tilde{p})} \tag{4.21}$$

are solved. The detailed procedure of the error estimation follows the error estimation described in Subsection 3.3.2.

---

[*]The orthogonality $\mathcal{B}(\boldsymbol{e}_{(\tilde{h},\tilde{p})}, \boldsymbol{w}_{(h,p)}) = 0$ does not always hold, for instance, due to inexact numerical integration. Nevertheless, the measured quantity $B(\boldsymbol{e}, \boldsymbol{\epsilon})$ is still widely used as it efficiently indicates the error in the quantity of interest.

4.3.3 Choices of error measures in local domains

The globally-defined error quantity $Q(e_{(\tilde{h},\tilde{p})})$ can be split into elemental contributions as

$$Q(e_{(\tilde{h},\tilde{p})}) = \sum_k \left(Q_k(e_{(\tilde{h},\tilde{p})}) \right) \tag{4.22}$$

where

$$Q_k(e_{(\tilde{h},\tilde{p})}) = \mathcal{B}_k(e_{(\tilde{h},\tilde{p})}, w_{(h,p)}) + \mathcal{B}_k(e_{(\tilde{h},\tilde{p})}, \epsilon_{(\tilde{h},\tilde{p})}) \tag{4.23}$$

It should be noted that the first right-hand-side term of Eq. (4.23) does not necessarily cancel since the Galerkin's orthogonality property does not hold in this local setting. Nevertheless, it has been proven that the second term can represent the distribution of the error in the local region effectively [31].

To simplify the adaptive criteria, the non-negative value of the term is employed. We have chosen the first measure as

$$\mathcal{E}_k^I := |\mathcal{B}_k(e_{(\tilde{h},\tilde{p})}, \epsilon_{(\tilde{h},\tilde{p})})| \tag{4.24}$$

which is chosen as our first alternative for representing the error in the elemental region Ω_k.

The second alternative is related to the first one by the Cauchy-Schwartz's inequality[†], i.e.

$$|\mathcal{B}_k(e_{(\tilde{h},\tilde{p})}, \epsilon_{(\tilde{h},\tilde{p})})| \le \|e_{(\tilde{h},\tilde{p})}\|_k \|\epsilon_{(\tilde{h},\tilde{p})}\|_k \tag{4.25}$$

leading to the local error measure [4, 31]

$$\mathcal{E}_k^{II} := \|e_{(\tilde{h},\tilde{p})}\|_k \|\epsilon_{(\tilde{h},\tilde{p})}\|_k \tag{4.26}$$

Replacing the use of Eq. (4.22), two corresponding global measures may be set by summing up the local measures as

$$\mathcal{E}^I = \sum_k \mathcal{E}_k^I \tag{4.27}$$

and

$$\mathcal{E}^{II} = \sum_k \mathcal{E}_k^{II} \tag{4.28}$$

These two estimates (cf. Eq. (4.27)) can provide an easy setting for the adaptive and optimality criteria, to be formulated in Chapter 5, as the global measures are computed by directly summing up the local measures. The performance of both choices will be studied later in this chapter.

[†]The Cauchy-Schwartz's inequality holds if $\mathcal{B}_k(e,e) \ge 0$. In this case, the stiffness matrix must be positive-definite.

4.3.4 Nonlinear finite element analysis

In nonlinear finite element analysis, the Newton-Raphson iterative procedure is generally applied to obtain the solution of the physical (primal) problem. This incremental loading procedure is, however, not needed in the dual framework, as the solution of the dual problem indicates the influence of the primal solution at a loading step to the quantity of interest. In other words, a linear solution control should be sufficient for analysing the dual problem.

During the nonlinear (primal) solution control, a set of discretised dual equations may be set based on the tangent representation at the time of computation. The influence function is obtained by solving

$$\mathcal{B}^{\text{tang}}(\boldsymbol{v}_{(h,p)}, \boldsymbol{w}_{(h,p)}) = Q(\boldsymbol{v}_{(h,p)}) \qquad \forall \boldsymbol{v}_{(h,p)} \in \mathcal{V}_{(h,p)} \tag{4.29}$$

Following the linear-elastic case, error analysis in the dual framework is then straightforwardly carried out by means of the error equation

$$\mathcal{B}^{\text{tang}}(\boldsymbol{v}_{(\tilde{h},\tilde{p})}, \boldsymbol{\epsilon}_{(\tilde{h},\tilde{p})}) = \mathcal{R}^w(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) = Q(\boldsymbol{v}_{(\tilde{h},\tilde{p})}) - \mathcal{B}^{\text{tang}}(\boldsymbol{v}_{(\tilde{h},\tilde{p})}, \boldsymbol{w}_{(h,p)})$$
$$\forall \boldsymbol{v}_{(\tilde{h},\tilde{p})} \in \mathcal{V}_{(\tilde{h},\tilde{p})} \tag{4.30}$$

Again, the error is computed based on patches of elements and the computational procedure follows the same procedure as described in Subsection 3.3.2.

4.4 Numerical examples

One-dimensional problem

The one-dimensional problem (cf. Section 3.7) is revisited within a goal-oriented framework. We select our quantity of interest as

$$Q(u) = \int_{\Omega_s} \varphi(x)u(x)dx \tag{4.31}$$

where Ω_s denotes the subdomain of the range $[x_0 - r, x_0 + r]$ around the center point of interest x_0. The weight function $\varphi(x)$ is defined as

$$\varphi(x) = \frac{(r - |x - x_0|)^2}{r^2} \tag{4.32}$$

providing more attention at the finite element solution near the center point of interest x_0. In this problem, we select the center point of interest to be at $x_0 = 0.8$ and four sizes of the radius of influence to be investigated, namely $r = 0.00$, $r = 0.05$, $r = 0.10$ and $r = 0.20$. Note that the first case with the radius of zero implies our choice of a pointwise quantity. Figure 4.3 shows the weight functions used in all cases, the corresponding influence functions and the finite element solution of the physical (primal) analysis.
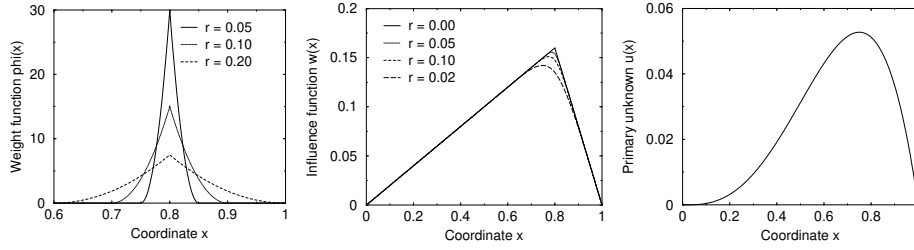
Figure 4.3 Weight functions (left, $\varphi(x)$), the influence functions (middle, $w(x)$) and the exact solution of the primary unknown (right, $u(x)$).
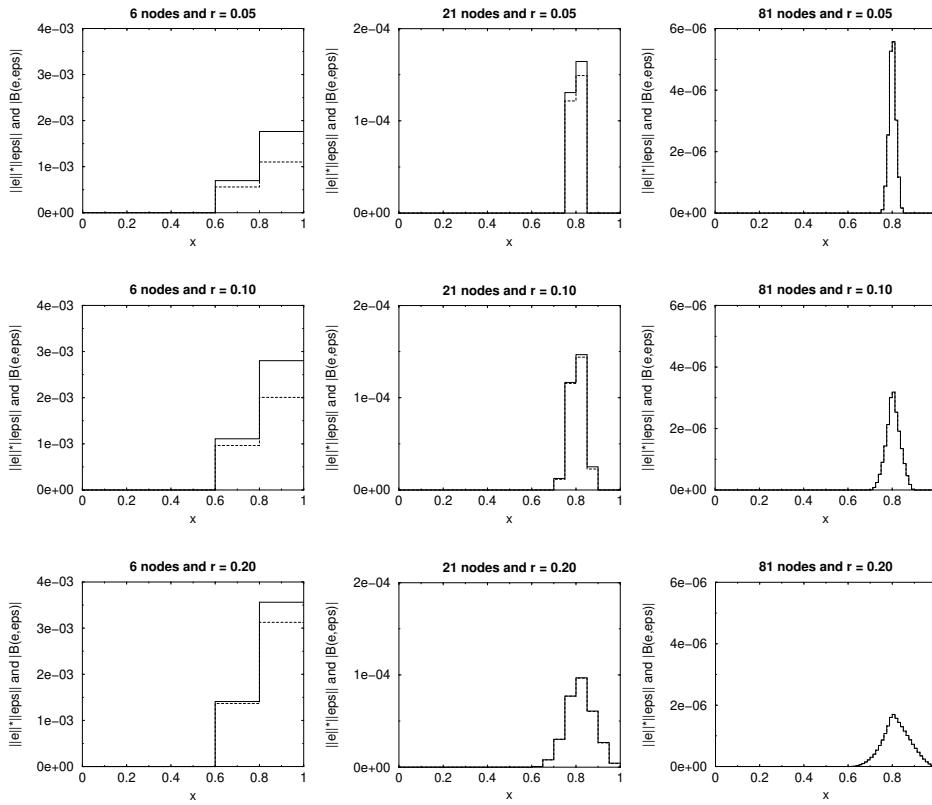

Figure 4.4 Elemental distribution of error in goal-oriented measures, i.e. $\|e\|_k\|\epsilon\|_k$ (solid line) and $|B_k(e,\epsilon)|$ (dashed line).
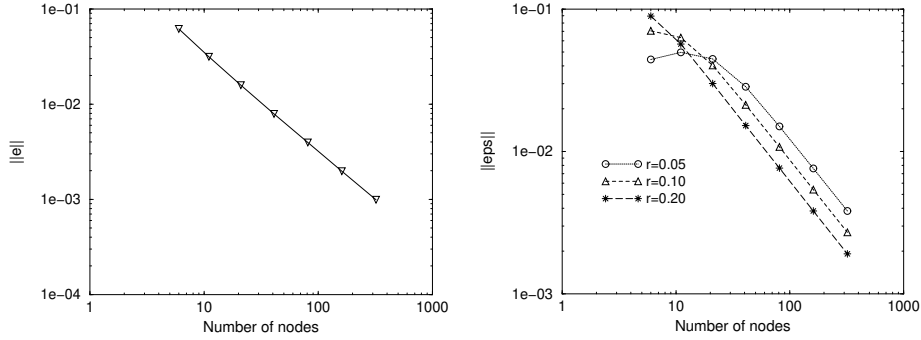
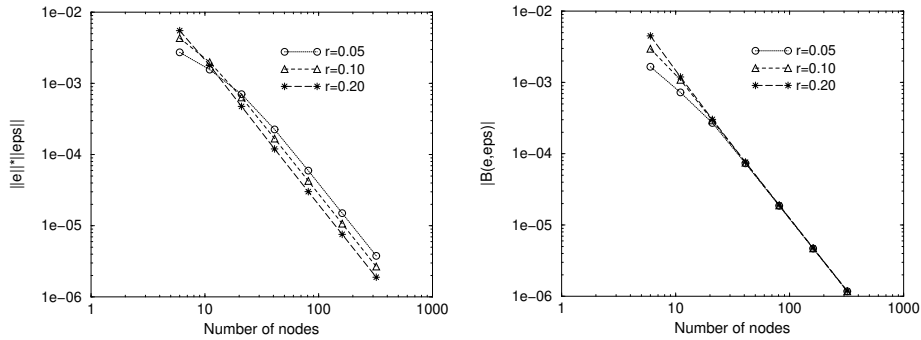**Figure 4.5** Convergence of error in primal problem (left) and dual problem (right).

Figure 4.6 Convergence of the goal-oriented error measures.

Figure 4.4 shows the distribution of elemental errors, i.e. $|B_k(e, \epsilon)|$ and $\|e\|_k\|\epsilon\|_k$, in the goal-oriented framework. The results agree with the Cauchy-Schwartz inequality that $|B_k(e, \epsilon)| \leq \|e\|_k\|\epsilon\|_k$. However, it is observed that the two quantities get closer upon mesh refinement.

The case of $r = 0$ is not shown in Figure 4.4. For this pointwise quantity, the goal-oriented error estimation does not reveal any error in the quantity of interest. It appears that $w = w_{(h,p)} = w_{(\tilde{h},\tilde{p})}$, and consequently $B(e, w) = B(e, \epsilon) = 0$. In fact, one can notice in this problem that the finite element analysis gives the exact solution at the nodes. Thus the error of the solution u at $x = 0.8$ does not exist. If we refer back to Section 3.7, the energy norm error in the primal problem appears to be nonzero in the vicinity of $x = 0.8$. This goal-oriented error estimate clearly gives the information that is more relevant to the quantity of interest which is not pointed out in the norm measurement.

In this example, we employ linear elements in the finite element analysis. To

capture the influence function accurately, three meshes of different discretisations are considered. Clearly illustrated in Figure 4.4, using different mesh resolutions, the error in the goal quantity can be very different, especially when $r \to 0$ is applied in the mesh of a low resolution. The global convergence plots, cf. Figures 4.5 and 4.6, also reveal slow convergence or even divergence of the estimated error when an inadequate mesh resolution is used. Obviously, the element size should not be smaller than the zone of interest, otherwise it will lead to a wrong representation of the equivalent nodal load in the dual problem.

The global measures in Figure 4.6 (left) and (right) give similar trends according to the error in energy norm (cf. Figure 4.5). However, it can be noticed that when the mesh is sufficiently fine, the computed quantities $\mathcal{B}(e, \epsilon)$ obtained from different sizes of domain of interest tend to converge to the same value, unlike the case of $\|e\| \|\epsilon\|$. This phenomenon implies the superior control of the primal error e over the dual error ϵ in the global computation of $\mathcal{B}(e, \epsilon)$.

Punch problem

The Prandtl's punch test is investigated as the second example. A rigid plate is pushed into a confined linear elastic material. The detailed description of the model
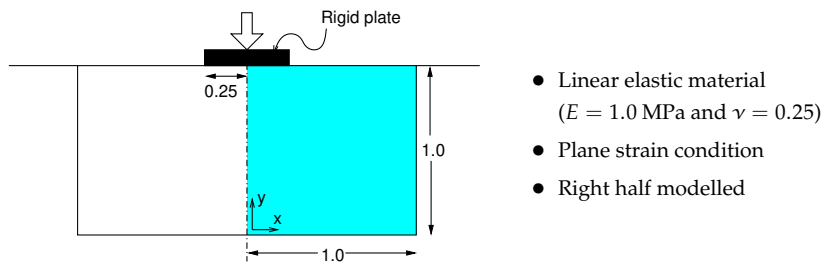

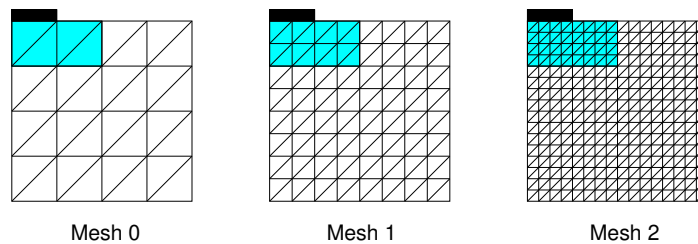**Figure 4.7** The Prandtl's punch test. Dimensions are in mm.

Figure 4.8 Finite element meshes used in the computation. The shaded area is the area of interest Ω_s.

(a) Displacement function u_x

(b) Displacement function u_y

(c) Influence function w_x
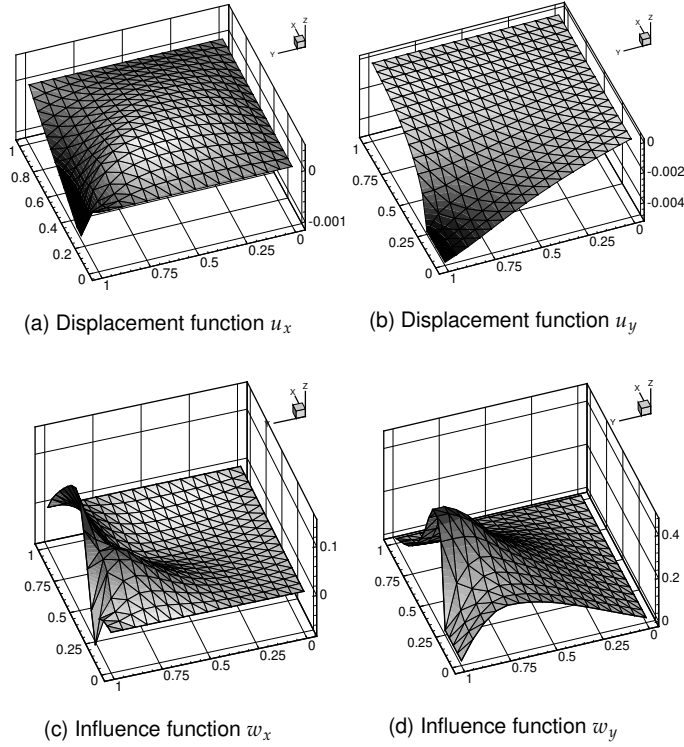
(d) Influence function w_y

Figure 4.9 The displacement function and the influence functions, obtained using the reference mesh, i.e. Mesh 2 with the quartic interpolation ($p = 4$).

problem is given in Figure 4.7[‡]. Three selected meshes, shown in Figure 4.8, are used in this finite element analysis.

For this problem, we choose our goal quantity to be

$$Q(\boldsymbol{u}) = \int_{\Omega_s} u_y(\boldsymbol{x})\, d\Omega \tag{4.33}$$

where $u_y(x)$ denotes the displacement in the y direction at any point x, and the area of interest Ω_s is defined as the shaded area in Figure 4.8. A contant weight function (cf. Eq. (4.1)) $\varphi(x) = 1$ in the subdomain Ω_s is employed here. The solutions of the primal problem, as well as those of the dual problem, are plotted in Figure 4.9.

Considering the elemental distribution of the estimated error in the quantity of interest shown in Figures 4.10 and 4.11, it is observed that both elemental measures, namely $B_k(\boldsymbol{e}_{(\tilde{h},\tilde{p})}, \boldsymbol{\epsilon}_{(\tilde{h},\tilde{p})})$ and $\|\boldsymbol{e}_{(\tilde{h},\tilde{p})}\|_k \, \|\boldsymbol{\epsilon}_{(\tilde{h},\tilde{p})}\|_k$, are distributed in similar fashions. Nevertheless, the latter choice gives a smoother distribution of error as clearly seen

---

[‡]Linear interpolation is not very good when $\nu \to 0.5$ because of locking problems; however in this case of $\nu = 0.25$, locking is not a problem.

in the case of a linear mesh. The error distributions agree well with the profile of the influence function shown in Figure 4.9, suggesting the residuals be distributed towards the end point of the plate where the boundary conditions change abruptly.

In this example, both h-factor and p-factor of the finite element discretisation are investigated. We examine three meshes, each of which is applied with four orders of polynomial interpolation, ranging from linear interpolation ($p = 1$) to quartic interpolation ($p = 4$). The convergence trends of the norms of the displacement, the influence function and the quantity of interest are shown in Figure 4.12. Apparently, all measures converge faster upon the p-extension than the h-extension. With no surprise, the error estimates of the primal and the dual problem also follow the same trends, as shown in Figure 4.13.

For goal-oriented convergence, four error measures, namely $|\mathcal{B}(e, \epsilon)|$, $\|e\|\|\epsilon\|$, $\sum_k |\mathcal{B}_k(e, \epsilon)|$ and $\sum_k (\|e\|_k \|\epsilon\|_k)$, are investigated. In Figure 4.14, all error measures show the same trends of convergence. It is found that the sums of our elemental contributions (Subfigures (c) and (d)) are suitable representations of the global measures (Subfigures (a) and (b)). Providing straightforward contribution from the elemental error data, we trust that the newly proposed global quantities can bring about a good adaptive mesh discretisation.

4.5 Remarks

In this chapter, the basic concept of the goal-oriented error estimation has been presented for use in linear and nonlinear finite element analyses. The performance of the error estimation has been examined only in linear problems as these provide more in-depth understanding of the approach. Also, since the goal-oriented error estimation is set based on the tangent representation at any computational step, without involving the nonlinear counterparts, the estimation in the nonlinear analysis employs the same concept as in the linear framework. Hence, investigating the performance of the goal-oriented error estimation in the linear setting is sufficient.

In order to set a proper error measure in the goal-oriented framework, besides the fundamental error measures, we have also investigated two simple error measures that can be straightforwardly inserted in the adaptivity settings. It has been found that both error measures can give a good indication of the error in the specific quantity and thus are suitable for setting of the adaptive and the optimality criteria, which will be presented in the next chapter.
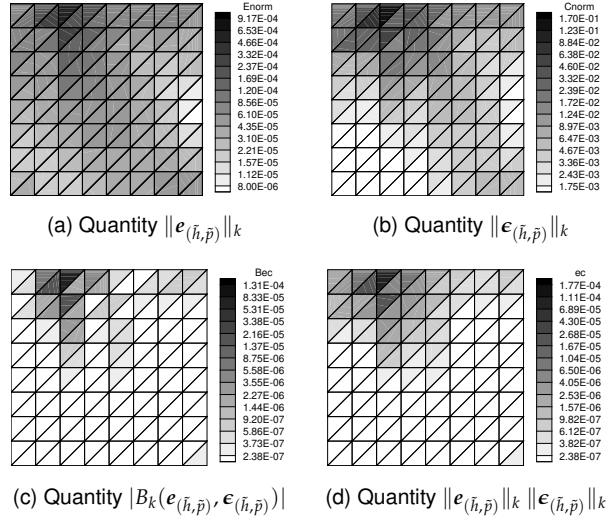
(a) Quantity $\|e_{(\tilde{h},\tilde{p})}\|_k$

(b) Quantity $\|\epsilon_{(\tilde{h},\tilde{p})}\|_k$

(c) Quantity $|B_k(e_{(\tilde{h},\tilde{p})}, \epsilon_{(\tilde{h},\tilde{p})})|$

(d) Quantity $\|e_{(\tilde{h},\tilde{p})}\|_k \|\epsilon_{(\tilde{h},\tilde{p})}\|_k$

Figure 4.10 Distribution of the elemental error measures in the framework of primal and dual problem, as well as the goal-oriented framework. The linear interpolation in Mesh 1 is employed. Exponential scaling is used.
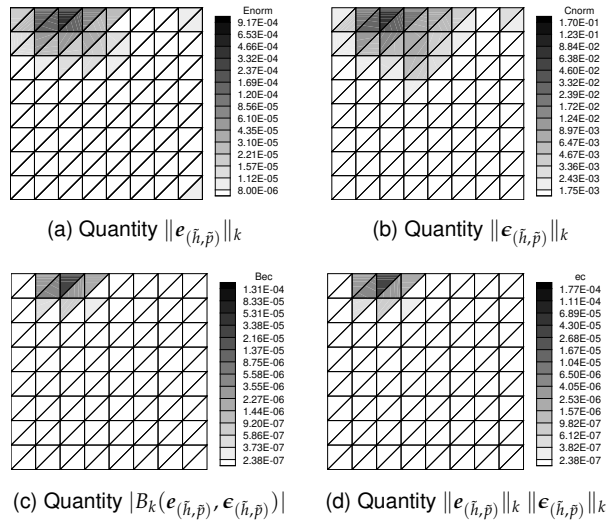

(a) Quantity $\|e_{(\tilde{h},\tilde{p})}\|_k$

(b) Quantity $\|\epsilon_{(\tilde{h},\tilde{p})}\|_k$

(c) Quantity $|B_k(e_{(\tilde{h},\tilde{p})}, \epsilon_{(\tilde{h},\tilde{p})})|$

(d) Quantity $\|e_{(\tilde{h},\tilde{p})}\|_k \|\epsilon_{(\tilde{h},\tilde{p})}\|_k$

Figure 4.11 Distribution of the elemental error measures in the framework of primal and dual problem, as well as the goal-oriented framework. The quadratic interpolation in Mesh 1 is employed. Exponential scaling is used.
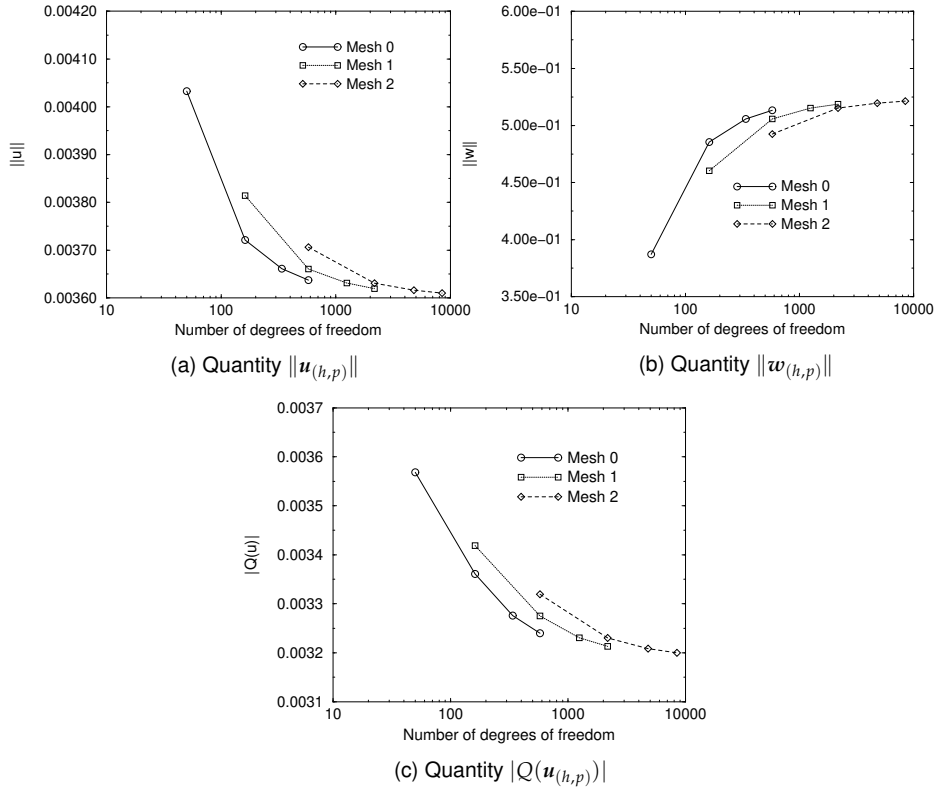
(a) Quantity $\|\boldsymbol{u}_{(h,p)}\|$

(b) Quantity $\|\boldsymbol{w}_{(h,p)}\|$

(c) Quantity $|Q(\boldsymbol{u}_{(h,p)})|$

Figure 4.12 Global convergence of the solutions in the primal problem, the solutions in the dual problem and the goal quantity.

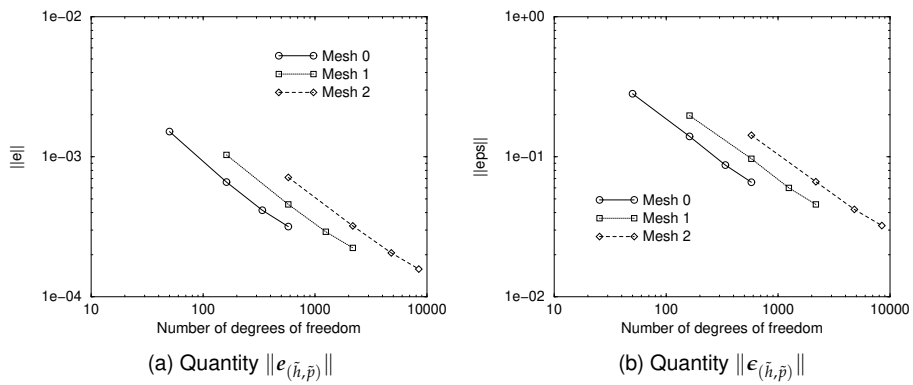(a) Quantity $\|\boldsymbol{e}_{(\tilde{h},\tilde{p})}\|$

(b) Quantity $\|\boldsymbol{\epsilon}_{(\tilde{h},\tilde{p})}\|$

Figure 4.13 Energy norm of error in primal and dual problem.

(a) Quantity $|B(e_{(\tilde{h},\tilde{p})}, \epsilon_{(\tilde{h},\tilde{p})})|$

(b) Quantity $\|e_{(\tilde{h},\tilde{p})}\| \, \|\epsilon_{(\tilde{h},\tilde{p})}\|$

(c) Quantity $\sum_k |B_k(e_{(\tilde{h},\tilde{p})} \epsilon_{(\tilde{h},\tilde{p})})|$

(d) Quantity $\sum_k \|e_{(\tilde{h},\tilde{p})}\|_k \, \|\epsilon_{(\tilde{h},\tilde{p})}\|_k$
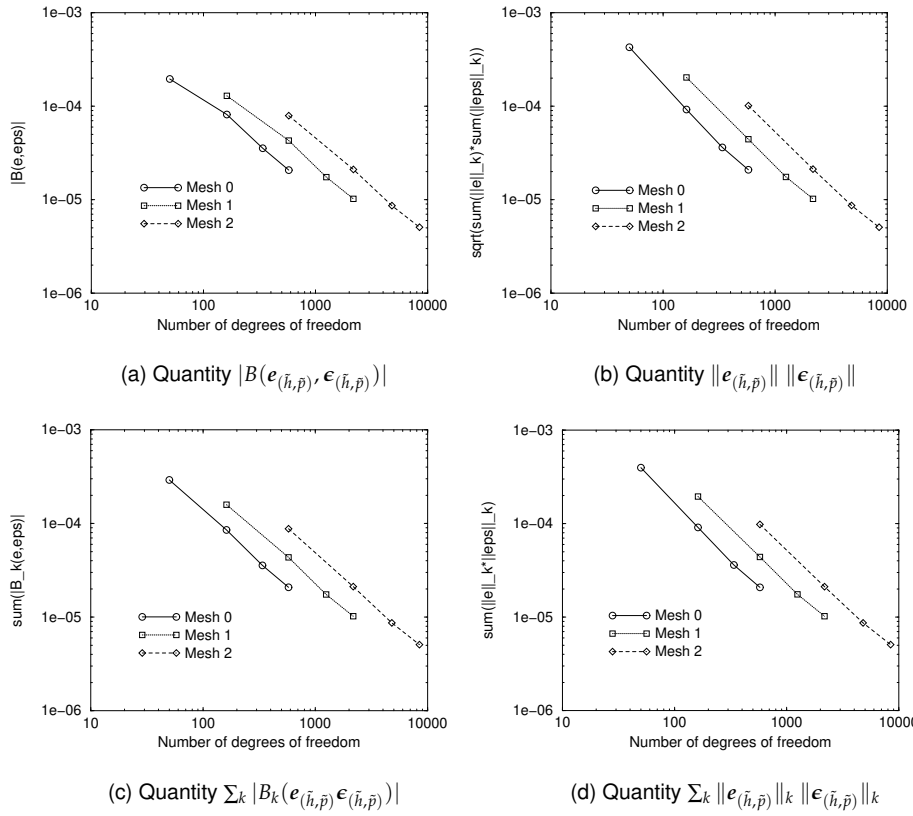
Figure 4.14 Comparison of various goal-oriented measures.

Mesh adaptive strategies

Balancing maximum accuracy and minimum computational effort is the key theme in mesh adaptive strategies. Clearly, one can get a very accurate result by using a very high mesh resolution and a much enriched interpolation. This, however, leads to a larger system of equations and consequently requires a high computational effort. As a compromise, the mesh is considered *optimally designed* once it can provide an acceptably accurate result, while keeping the computational effort as low as possible.

As the finite element solution is unknown *a priori*, designing an optimal mesh at the beginning of the computation becomes a difficult if not impossible task, especially in a nonlinear analysis. An alternative to the optimal *a priori* mesh design is the use of *mesh adaptivity* techniques that may be inserted in order to improve the finite element solutions during the computation.

We make use of the error information described in Chapter 3 and Chapter 4, and continue in this chapter with some criteria to consider if the finite element mesh needs an enhancement and in which regions these enhancements should take place. These adaptive criteria are defined for measuring the accuracy of the finite element solution in a global sense. Besides this *global criterion*, its combination with the *local criterion* defines the problem area where the mesh should be improved via adaptive techniques, such as mesh refinement (*h*-adaptivity), mesh enrichment (*p*-adaptivity) and mesh gradation (*r*-adaptivity). Since the connectivity of element nodes is preserved in the second and the third techniques, these two adaptive schemes simplify the adaptive process.

Another important issue to be presented in this chapter is how variables are transferred from the previous to the new discretisation. Obviously, techniques of *variable transfer* are necessary during nonlinear solution control so that the solution path can be continued without restarting the computation. After each mesh change, the primary unknowns as well as the state variables at the integration points must be accurately transferred to the new system. This issue will be presented in the final

section of this chapter.

5.1 Mesh quality and enhancement strategies

To objectively adapt the mesh, it is necessary to know to which extent the mesh should be refined. The knowledge about *a priori* error estimation supplies information about the convergence rate, i.e. how fast the error is reduced according to changes of the finite element mesh, either by mesh refinement or mesh enrichment.

5.1.1 *A priori* error estimates

For a smooth solution, the error in displacement converges by order $\mathcal{O}(h^{p+1})$ and thus the error in the first derivatives (i.e. strains and stresses) has convergence of $\mathcal{O}(h^p)$ [44, 93, 105]. The error in the energy norm $\|e\|$, defined as the square root of the product of stress error and strain error, is convergent in the order $\mathcal{O}(h^p)$. Thus the *a priori* error estimate for the standard ($\mathcal{C}^0$-continuous) finite element approximation can be described in the form

$$\|e\| \leq \mathcal{C}h^p \tag{5.1}$$

where p is the polynomial order applied in the mesh, h is a characteristic size of the element, $\mathcal{C}$ is a constant depending on distortion of the elements and is proportional to the order of derivatives of the function u. It is worth noting that h is not the real size of the elements since it must be less than one to validate the convergence criteria. This characteristic size is generally used in a relative sense, for example, in measurement of the desired element size with respect to the current element size.

The local error includes the element area of order $\mathcal{O}(h^d)$, cf.

$$\|e\|_k^2 = B_k(e, e) = \int_{\Omega_k} (\varepsilon - \varepsilon_{ex})^T D (\varepsilon - \varepsilon_{ex}) \, \mathrm{d}\Omega_k \tag{5.2}$$

that is, the element area contribution to $\|e\|_k$ is of order $\mathcal{O}(h^{\frac{d}{2}})$. Here, d denotes the number of spatial dimensions of the problem. Consequently, the local error converges as

$$\|e\|_k \leq \mathcal{C}_k h_k^{p+\frac{d}{2}} \tag{5.3}$$

In case of a non-smooth solution, the above assumptions are no longer valid [93]. In existence of singularities, the a priori estimate becomes

$$\|e\| \leq \mathcal{C}^\star(u, p) h^{\min(p,q)} \tag{5.4}$$

where the constant $\mathcal{C}^\star$ depends on the displacement and the order of interpolation. Convergence of the finite element solutions indeed depends on the polynomial order of the interpolation (p) and the solution smoothness ($q > 0$). The higher q, the smoother the solution.

5.1.2 Remarks on mesh adaptive algorithms

The *a priori* estimates give some clues which type of mesh enhancement should be applied. The convergence of the solutions, as pointed out earlier, depends mainly on the element size, the interpolation degree and the smoothness of the actual solution.

In the following, we will discuss the expected performance of the adaptive techniques. As aforementioned in Chapter 1, the adaptive techniques that are in our scope include *h*-adaptity, *p*-adaptivity and *r*-adaptivity. The first two methods, which are the main enhancement techniques in standard adaptive finite element analysis, improve the solution by adding extra degrees of freedom to the system of equations. The third method, i.e. *r*-adaptivity, provides mesh improvement at no extra cost*.

h-adaptivity

The *h*-extension improves the finite element solution by reducing the element size, while keeping the interpolation at the same order. The mesh refinement increases the local smoothness of the solution, hence it is often assumed that the smoothness q (cf. Eq. (5.4)) has no effect on the *h*-convergence [93]. That is, the error converges in the rate described in Equation 5.1.

By *h*-extension, one can choose either to refine the mesh elementwise (i.e., only the problem elements are refined) or to refine the mesh totally (i.e., the whole mesh is redesigned). Some comments for both methods are as follows.

- By hierarchical (elemental) refinement, a selected element is subdivided into smaller sub-elements. If the subdivision is from edge to edge, there will be some irregular connectivities between the subdivided element and some adjacent elements. In existence of the so-called *hanging nodes*, it requires imposition of some special constraints to guarantee continuity of the finite element solutions over the edges (where hanging nodes exist). Alternatively, one can avoid the creation of hanging nodes by extending the subdivision to adjacent elements and ending the subdivision at the vertex nodes. See Figure 5.1.

- By total mesh refinement, the mesh can be reconstructed totally by means of a mesh generator. An obvious advantage of this version of *h*-adaptivity over other adaptive methods is that the element size is not restricted by the previous mesh. By the smooth solution assumptions, it is simple to redesign the desired element size based on optimality criteria, which assemble error information, adaptive criteria and the *a priori* convergence assumptions. (See Section 5.3 for the setting of the optimality criteria.) To construct an optimally designed mesh, both refining and coarsening of the mesh may be applied.

---

*The *extra cost*, here, means that the number of degrees of freedom is not increased so the computational costs involved with solving the global system of equations remain the same. It is assumed here that the costs of designing a new finite element mesh by means of *r*-adaptivity are negligible.
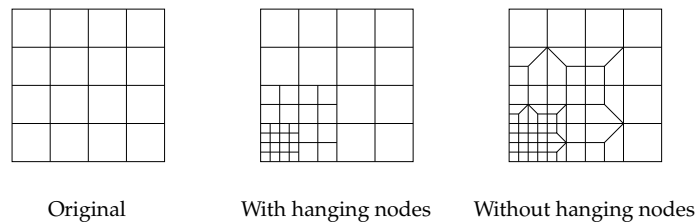
Original   With hanging nodes   Without hanging nodes

Figure 5.1 Examples of hierarchical *h*-adaptivity

p-adaptivity

p-adaptivity improves the mesh by enriching the degrees of interpolation without changing mesh size and configuration. The *a priori* estimate defined in the last subsection (cf. Eq. (5.4)) implies that increasing the polynomial degrees of interpolation (*p*-extension) is better than mesh refinement (*h*-extension) when the solution is smooth (large *q*). On the contrary, the *h*-extension is more suitable in regions where the solution becomes singular as its convergence does not depend on the smoothness of the solution. It, however, depends on how well the mesh is graded. With singularities, performance of the *p*-version finite element method depends on the mesh. The performance will be much better when the mesh has been graded near the singularity [93], increasing the local smoothness of the solution at the singular region and allowing higher-order interpolation to work more effectively. Its combination to the *h*-adaptivity (i.e., *hp*-adaptivity) provides very high convergence rates.

By *p*-extension, the interpolation function is generally enriched to one higher order at a time. Designing the optimal polynomial degree *p* is theoretically possible but not very useful as the value of *p* needs to be a positive integer[†].

By hierarchical enrichment, the higher-order modes can be added to the existing counterpart easily. The enrichment can be different in each spatial direction [27] to optimise the use of higher-order interpolation. However, to reduce the complexity of the data structure in the *p*-approach, we choose to enrich the interpolation by the same degrees in both directions as in [74]. The compatibility of the edge modes of the adjacent elements will be satisfied by choosing the higher-order shape functions between two adjacent elements sharing the same edge. This is for ease of compatibility enforcement.

r- adaptivity

r-adaptivity relocates the spatial nodal points toward the critical region without changing mesh connectivity and interpolation order. In case of smooth solutions, the *r*-adaptive technique alone can only improve the accuracy of the solution

---

[†]This situation is similar for the hierarchical *h*-refinement, where the number of element subdivision must be a positive integer.

slightly [7, 47], in comparison to *h*-adaptivity and *p*-adaptivity. However, the adaptive scheme is very efficient when applied to a non-smooth solution, for example, when there exists a high stress gradient or strain localisation in a small region. Without adding any extra degrees of freedom and complicated remeshing[‡], the mesh gradation efficiently improves the local smoothness of the finite element solution.

To relocate the nodes, sophisticated techniques such as the moving finite element formulation [15] or the arbitrarily Lagrangian-Eulerian techniques [6, 73] may be applied to optimally locate the existing finite element nodes. These methods are based on a kinematic framework and require solving a global system of equations. As an alternative, a simple technique based on local weighted smoothing may be applied. This smoothing-based local approach is employed in this research and will be addressed in Section 5.4.

As the *r*-adaptivity improves the mesh by moving the nodes according to relative error information, use of error indication, which provides information in a relative sense, is actually adequate [43]. The scheme must be carefully applied as it may construct non-proportioned elements (i.e. element with large aspect ratio) that bring about bad quality of the finite element mesh, and occasionally the problem of *mesh tangling* arises. For increased performances in adaptivity, the method may be combined as an auxiliary adaptive technique to the *h*- and *p*- approaches. The method can be more effective when properly combined with *h*-adaptivity (*hr*-adaptivity [9]) or with *p*-adaptivity (*rp*-adaptivity [60]).

5.2 Adaptive criteria

An objective of adaptive strategies is to provide a finite element solution with an acceptable level of discretisation error. As such, some criteria to judge whether or not the mesh is sufficiently discretised are needed. The criteria must be set such that the global error that is allowed to occur should be less than an acceptable value to guarantee an acceptable solution, and the local error should be well distributed throughout the problem domain. Combinations of both global and local criteria, resulting in the so-called *adaptive criteria*, will be addressed in this section.

5.2.1 Energy norm based adaptive criteria

By global consideration, the mesh needs an enhancement when

$$\|e\| > \zeta_{\mathrm{prim}} \|u\| \tag{5.5}$$

where ζ_{prim} is the amount of error allowed in comparison to the norm of the primary unknown u. It can be chosen, for example, as low as 1% or as high as 20%, depending on the judgement of the user.

[‡]Remeshing in this context refers to the change in mesh configuration such as in *h*-adaptivity

If the global criterion is satisfied, the error in some regions of the problem domain may be much higher than the rest. Thus, it is suggested to consider the criteria to properly distribute the error in each local region throughout the domain. Basically, there are two distributing criteria;

- Based on *uniform error* distribution [53, 106]

$$\|e\|_k^2 = \frac{\|e\|^2}{N} \qquad \forall\, k \tag{5.6}$$

where N denotes number of elements in the mesh. The global error, $\|e\|^2 = \sum_{k=1}^{N} \|e\|_k^2$, should be equally distributed over each element. Combining this local criterion (cf. Eq. (5.6)) with the global criterion (cf. Eq. (5.5)), the mesh in the local region k must be enhanced when

$$\|e\|_k > \frac{\zeta_{\mathrm{prim}} \|u\|}{\sqrt{N}} \tag{5.7}$$

- Based on *uniform error density* distribution [40, 64]

$$\frac{\|e\|_k^2}{\Omega_k} = \frac{\|e\|^2}{\Omega} \qquad \forall\, k \tag{5.8}$$

where Ω_k refers to the area of the element k and $\sum_k \Omega_k = \Omega$. This condition suggests that, for smaller elements, a smaller error is allowed. Again, combination of local criterion (cf. Eq. (5.8)) and global criterion (cf. Eq. (5.5)) results in mesh enhancement when

$$\|e\|_k > \zeta_{\mathrm{prim}} \|u\| \left(\frac{\Omega_k}{\Omega} \right)^{\frac{1}{2}} \tag{5.9}$$

It has been found that using the uniform error density distribution leads to a much more expensive mesh than the uniform error distribution does [23, 30, 64]. An observation is that, for the region containing a singularity, the designed element size can be very small. As the uniform error distribution rule leads to acceptable results, in this study we will use the uniform error distribution.

5.2.2 Goal-oriented adaptive criteria

To decide whether or not the mesh should be enhanced, we may consider directly the error in a quantity of interest

$$|Q(e)| > \zeta_{\mathrm{goal}} |Q(u)| \tag{5.10}$$

or

$$\|e\| \|\epsilon\| > \zeta_{\mathrm{goal}} \|u\| \|w\| \tag{5.11}$$

In this research, to facilitate the combination of global and local refinement criteria, we have selected two global refinement criteria.

Adaptive criterion 1

Whenever

$$\sum_{k=1}^{N} |B_k(e, \epsilon)| > \zeta_{\text{goal}} \sum_{k=1}^{N} |B_k(u, w)| \tag{5.12}$$

the mesh needs to be improved. It is noted that the measure $|B_k(u, w)|$ is not equal to zero as in the global measure $|B(u, w)|$ in the displacement control algorithm. Similar to the distribution of error in the energy norm, the error measure can be distributed uniformly as

$$|B_k(e, \epsilon)| = \frac{\sum\limits_{j=1}^{N} |B_j(e, \epsilon)|}{N} \qquad \forall\, k \tag{5.13}$$

And by combining this local criterion with the global criterion, the mesh in the local region k must be enhanced when

$$|B_k(e, \epsilon)| > \frac{\zeta_{\text{goal}} \sum\limits_{j=1}^{N} |B_j(u, w)|}{N} \tag{5.14}$$

Adaptive criterion 2

The mesh needs to be improved whenever

$$\sum_{k=1}^{N} \left(\|e\|_k \|\epsilon\|_k \right) > \zeta_{\text{goal}} \sum_{k=1}^{N} \left(\|u\|_k \|w\|_k \right) \tag{5.15}$$

Again, the error measure can be distributed by uniform error distribution

$$\|e\|_k \|\epsilon\|_k = \frac{\sum\limits_{j=1}^{N} \left(\|e\|_j \|\epsilon\|_j \right)}{N} \qquad \forall\, k \tag{5.16}$$

Combining this local criterion with the global criterion, the mesh in the local region k must be enhanced when

$$\|e\|_k \|\epsilon\|_k > \frac{\zeta_{\text{goal}} \sum\limits_{j=1}^{N} \left(\|u\|_j \|w\|_j \right)}{N} \tag{5.17}$$

Apparently, the global allowance in the framework of goal-oriented error estimation ζ_{goal} can be set as

$$\zeta_{\text{goal}} \approx \zeta_{\text{prim}} \zeta_{\text{dual}} \tag{5.18}$$

where ζ_{prim} and ζ_{dual} are global allowances for the error in the primal problem and for the error in the dual problem, respectively.

5.3 Optimality criteria

From the last section, adaptivity takes place when the error exceeds the critical value. How much the mesh should be refined, via the so-called *optimality criteria*, is the main issue in this section.

Specially designed for an *h*-adaptive scheme, the mesh can be optimally designed by mathematical derivation of the error in terms of its convergence with respect to the element sizes. The first work has been proposed by Zienkiewicz and Zhu [106] in the late 80s. Their optimality criterion is based on the global *a priori* estimate in a generally smooth space with an assumption of uniform distribution of elemental error throughout the problem domain. A modification to the above criterion has been suggested in [64], where the elemental error measure is distributed via a uniform error density setting. Another distinguished criterion is formulated in a relative manner [30] with two user-specified parameters for adjusting the error distribution in the relative and absolute manner.

In this research, we choose to study the criterion introduced in [52, 53] as it can produce the cheapest mesh via the *a priori* local error estimate (cf. Eq. (5.3)) and the uniform error distribution. Whereas the criterion is originally set for use with the energy norm error estimates, it can slightly be adjusted for use in the goal-oriented setting as well.

5.3.1 Energy norm based optimality criteria

The critical value of the error in energy norm is the maximum of $\|e\|$ described in Eq. (5.3), i.e.

$$\|e\|_k = \mathcal{C}_{\mathrm{prim}} h_k^{\frac{2p+d}{2}} \tag{5.19}$$

Let us define the desired value of $\|e\|_k$ as $\{\|e\|_k\}_{\mathrm{des}}$, thus we obtain

$$\{\|e\|_k\}_{\mathrm{des}} = \mathcal{C}_{\mathrm{prim}}\{h_k\}_{\mathrm{des}}^{\frac{2p+d}{2}} \tag{5.20}$$

where $\{h_k\}_{\mathrm{des}}$ denotes the desired characteristic size of element k corresponding to the local error quantity $\{\|e\|_k\}_{\mathrm{des}}$.

Compare the two norms, we obtain

$$\frac{\|e\|_k}{\{\|e\|_k\}_{\mathrm{des}}} = \left(\frac{h_k}{\{h_k\}_{\mathrm{des}}}\right)^{\frac{2p+d}{2}} \tag{5.21}$$

Applying local error requirements as earlier mentioned, the uniform error criterion leads to

$$\frac{\{h_k\}_{\mathrm{des}}}{h_k} = \left(\frac{\{\|e\|_k\}_{\mathrm{des}}}{\|e\|_k}\right)^{\frac{2}{2p+d}} = \left(\frac{\zeta_{\mathrm{prim}}\|u\|}{\sqrt{N}\|e\|_k}\right)^{\frac{2}{2p+d}} \tag{5.22}$$

Note that the characteristic size h is always measured in a relative sense.

As the number of elements may be changed during the mesh refinement, it was suggested to reformulate Eq. (5.22) by replacing N with N_{new}, the new number of elements [53]. The new number of elements can be estimated from

$$N_{\text{new}} = \sum_{k=1}^{N} \left(\frac{h_k}{\{h_k\}_{\text{des}}} \right)^2 = \left[\sum_{k=1}^{N} \left(\frac{\|e\|_k}{\zeta_{\text{prim}}\|u\|} \right)^{\frac{4}{2p+d}} \right]^{\frac{2p+d}{2p+d-2}} \tag{5.23}$$

5.3.2 Goal-oriented optimality criteria

In the goal-oriented estimation, following the error norm in the primal problem, we can write the error in the dual problem as

$$\|\epsilon\|_k = \mathcal{C}_{\text{dual}} h_k^{\frac{2p+d}{2}} \tag{5.24}$$

and

$$\{\|\epsilon\|_k\}_{\text{des}} = \mathcal{C}_{\text{dual}}\{h_k\}_{\text{des}}^{\frac{2p+d}{2}} \tag{5.25}$$

It thus leads to

$$\|e\|_k\|\epsilon\|_k = \mathcal{C}_{\text{prim}}\mathcal{C}_{\text{dual}} h_k^{2p+d} \tag{5.26}$$

and

$$\{\|e\|_k\|\epsilon\|_k\}_{\text{des}} = \mathcal{C}_{\text{prim}}\mathcal{C}_{\text{dual}}\{h_k\}_{\text{des}}^{2p+d} \tag{5.27}$$

Recall the Cauchy-Schwartz's inequality

$$|B_k(e,\epsilon)| \leq \|e\|_k\|\epsilon\|_k \tag{5.28}$$

the local quantity $|B_k(e,\epsilon)|$ can be related to $\|e\|_k\|\epsilon\|_k$ by a factor λ in the range of 0 to 1. And thus the convergence rate of $|B_k(e,\epsilon)|$ can be written as

$$|B_k(e,\epsilon)| = \lambda \|e\|_k\|\epsilon\|_k = \lambda \, \mathcal{C}_{\text{prim}}\mathcal{C}_{\text{dual}} h_k^{2p+d} \tag{5.29}$$

Assuming that λ attains the same value for every element k, it follows that

$$\{|B_k(e,\epsilon)|\}_{\text{des}} = \lambda \, \{\|e\|_k\|\epsilon\|_k\}_{\text{des}} = \lambda \, \mathcal{C}_{\text{prim}}\mathcal{C}_{\text{dual}}\{h_k\}_{\text{des}}^{2p+d} \tag{5.30}$$

Following [53], we can set the optimality criterion for goal-oriented measurement, based on uniform error distribution, as follows.

Based on adaptive criterion 1

From Eqs. (5.14), (5.29) and (5.30), the relative element size can be set as

$$\frac{\{h_k\}_{\text{des}}}{h_k} = \left(\frac{\{|B_k(e,\epsilon)|\}_{\text{des}}}{|B_k(e,\epsilon)|} \right)^{\frac{1}{2p+d}} = \left(\frac{\zeta_{\text{goal}} \sum_{k=1}^{N} |B_k(u,w)|}{N_{\text{new}} |B_k(e,\epsilon)|} \right)^{\frac{1}{2p+d}} \tag{5.31}$$

with the new number of elements N_{new} defined as

$$N_{\text{new}} = \sum_{k=1}^{N} \left(\frac{h_k}{\{h_k\}_{\text{des}}} \right)^2 = \left[\sum_{k=1}^{N} \left(\frac{|B_k(e,\epsilon)|}{\zeta_{\text{goal}} \sum_{k=1}^{N} |B_k(u,w)|} \right)^{\frac{2}{2p+d}} \right]^{\frac{2p+d}{2p+d-2}} \tag{5.32}$$

Based on adaptive criterion 2

From Eqs. (5.17), (5.26) and (5.27), the relative element size is cast as

$$\frac{\{h_k\}_{\text{des}}}{h_k} = \left(\frac{\{\|e\|_k \|\epsilon\|_k\}_{\text{des}}}{\|e\|_k \|\epsilon\|_k} \right)^{\frac{1}{2p+d}} = \left(\frac{\zeta_{\text{goal}} \sum_{k=1}^{N} (\|u\|_k \|w\|_k)}{N_{\text{new}} \|e\|_k \|\epsilon\|_k} \right)^{\frac{1}{2p+d}} \tag{5.33}$$

where the new number of elements N_{new} is defined as

$$N_{\text{new}} = \sum_{k=1}^{N} \left(\frac{h_k}{\{h_k\}_{\text{des}}} \right)^2 = \left[\sum_{k=1}^{N} \left(\frac{\|e\|_k \|\epsilon\|_k}{\zeta_{\text{goal}} \sum_{k=1}^{N} (\|u\|_k \|w\|_k)} \right)^{\frac{2}{2p+d}} \right]^{\frac{2p+d}{2p+d-2}} \tag{5.34}$$

5.4 Smoothing-based mesh gradation

The simple *r*-adaptivity applied in this research is modified from a mesh smoothing concept. Error information stored at nodes in the neighbourhood is used to drive the mesh movement algorithm so that the mesh is graded towards the critical region. As the technique does not require solving any global system of equations, it requires less computational effort than the global techniques (for example, [15,73]). Moreover, mesh tangling, which can often occur during conventional *r*-adaptivity, does not appear in this simple relocation technique. This is because each node is forced to move, at a time, within the limited distance in its corresponding patch (see Figure 5.2). The details of the method will be addressed further in this section.

5.4.1 Mesh gradation strategy

The mesh gradation is modified from *Laplace smoothing technique*, which is

$$\mathbf{x}_i = \frac{\sum_{j=1}^{N_i} \mathbf{x}_j}{N_i} \tag{5.35}$$

For each grading loop

1. Compute weight factors, defined at nodes

2. Move node i, starting $i = 1$;

    a) set patch, elements connected to node i,

    b) move node i according to nodal weight factors, and

    c) update weight factors due to change of nodal position.

3. Continue;

    a) if $i <$ number of nodes, $i = i + 1$, go to Step 2,

    b) otherwise, $i =$ number of nodes, go to next grading loop.

Table 5.1 Mesh gradation scheme based on weighted Laplace smoothing.

where N_i is number of neighbouring nodes j connecting node i. That is, each node is relocated into the center of the corresponding patch of elements surrounding the node. The node-by-node process goes on until all nodes do not move further, resulting in a set of well proportioned elements (or a smooth mesh).

In order to apply the mesh smoothing technique to grade the mesh towards critical regions in the domain, grading weight factors containing error information must be supplied. The resulting *weighted Laplace smoothing technique* can be set as

$$\mathbf{x}_i = \frac{\sum_{j=1}^{N_i} \varpi_j \mathbf{x}_j}{\sum_{j=1}^{N_i} \varpi_j} \tag{5.36}$$

where ϖ_j is the grading weight factor for node j. Obviously, the quality of the gradation depends heavily on the choice of the weight factors. Basically, the selected weight factor must not move the nodes too much nor too little, and should be related to error information. Table 5.1 shows the grading procedure based on the weighted Laplace smoothing.

Choice of grading weight factors

The grading weight factor provides information how much and to which direction the mesh should be graded. Certainly, the gradation should be towards the regions of larger error, where a higher mesh resolution is needed, than the rest of the problem domain. In [43], it was suggested that the use of an error indicator is sufficient and appropriate for driving mesh gradation (*r*-adaptivity) since the error information is employed only in a relative sense. Yet, as our main focus in this study is on error estimation, we select the ratio of the current element size over the desired
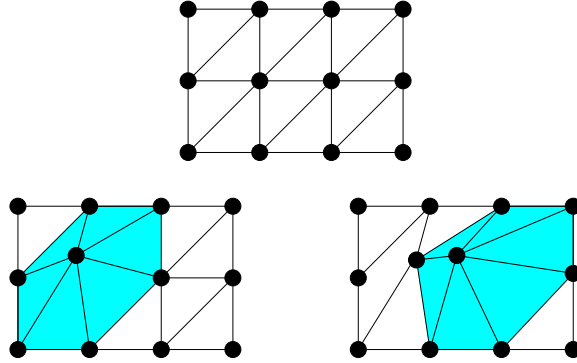
Figure 5.2 A local approach to mesh gradation: original mesh (upper), after the first node moved (left) and after the second node moved (right).

element size (here, named as *refinement ratio* [6])

$$\Upsilon_k^{\text{elem}} := \frac{h_k}{\{h_k\}_{\text{des}}} \tag{5.37}$$

to set the grading weight factor for this gradation.

According to the optimality criterion (cf. Section 5.3), the refinement ratio Υ based on energy norm error estimate (cf. Eq. (5.22)) reads

$$\Upsilon_k^{\text{elem}} = \left(\frac{\|e\|_k \sqrt{N}}{\zeta_{\text{prim}} \|u\|} \right)^{\frac{2}{2p+d}} \tag{5.38}$$

Similarly, based on goal-oriented error estimates (cf. Eqs. (5.31) and (5.33)), we can set

- *Criterion 1*

$$\Upsilon_k^{\text{elem}} = \left(\frac{|B_k(e, \epsilon)| \sqrt{N}}{\zeta_{\text{goal}} \sum_{k=1}^{N} |B_k(u, w)|} \right)^{\frac{1}{2p+d}} \tag{5.39}$$

- *Criterion 2*

$$\Upsilon_k^{\text{elem}} = \left(\frac{\|e\|_k \|\epsilon\|_k \sqrt{N}}{\zeta_{\text{goal}} \sum_{k=1}^{N} (\|u\|_k \|w\|_k)} \right)^{\frac{1}{2p+d}} \tag{5.40}$$

It is noted here that using refinement ratios to form the grading weight factors is a very reasonable choice as it accounts for the existing mesh order (i.e. order

of polynomial interpolation). For an initial mesh (i.e. prior to any other adaptive steps), a strong mesh gradation may be needed, whereas a mild gradation often suffices if adaptivity has already been activated earlier in the analysis.

To provide the information at nodes, the element refinement ratios must be extrapolated to the corresponding nodes. Based on finite element extrapolation and nodal averaging, the elemental values are transferred to node i via the use of nodal shape function $\phi_i(x)$ by

$$\Upsilon_i^{\text{node}} = \frac{\int\limits_{\Omega} \phi_i(x)\Upsilon(x)\,d\Omega}{\int\limits_{\Omega} \phi_i(x)\,d\Omega} \tag{5.41}$$

It is assumed that the element refinement ratio is constant within each element. Hence,

$$\Upsilon(x) = \Upsilon_k^{\text{elem}} \quad \text{if } x \text{ is in } \Omega_k \tag{5.42}$$

Note that $\Upsilon > 1$ in zones that need an improved discretisation and $\Upsilon < 1$ if a coarser discretisation is permitted. Ideally, $\Upsilon = 1$ everywhere in the domain. The rate of mesh gradation can be controlled by taking a power of Υ, by which the nodal weight factor is defined as

$$\varpi_i = \left(\Upsilon_i^{\text{node}}\right)^m \tag{5.43}$$

where m controls the mesh gradation rate and is user-specified. A small value of m indicates a small move of the mesh in each grading step. For more significant relocation, a large value may be used. A large value grades the mesh toward the critical region better than a small value, however, it may also create finite elements of large aspect ratio, resulting in a global stiffness matrix of bad conditioning. Furthermore, too large movement can bring about an instability caused by the nodes being moved back and forth in order to find their optimal positions. It is recommended here to use the value between 0.5 to 2.0 depending on how much mesh movement is preferred.

Weight update scheme

It is necessary to update the grading weight factors during mesh gradation. Every time a node is moved from its old position, the refinement information is obviously changed. Unless updating is included in the algorithm, the gradation may continue forever without the nodes reaching their optimal positions.

The main idea is to keep the set of elemental refinement ratios Υ^{elem} prior to the mesh gradation during the gradation. To update the elemental refinement ratios, the stored refinement ratios are transferred to the new elements according to the overlapping area. As these overlapped regions can be of any polygon forms, it requires complicated element partitioning and thus should be avoided.
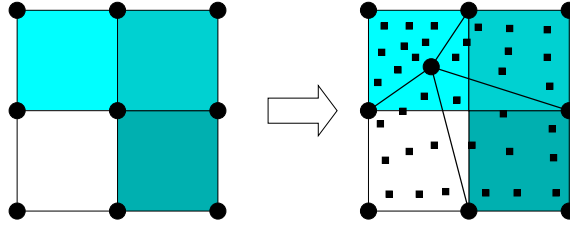
Figure 5.3 The weight update scheme.

An alternative way is to set integration points in the relocated mesh. The refinement ratios are then transferred from the old mesh to the new integration points. As more integration points are used, a more accurate transfer is expected. The values in the new set of integration points are again transferred to nodes by the finite element extrapolation (cf. Eq. (5.41)). The scheme is illustrated in Figure 5.3.

5.4.2 Auxiliary techniques

Mesh quality check

It may happen that the mesh movement does not improve the error solution. Two alternatives can be applied to detect the error trend before and after the mesh movement. The first scheme is based on error estimation. Once the error is computed, the mesh is temporarily moved to the new location. Then the error is estimated again. The algorithm implies that for every *r*-adaptive step, the computational cost for error computation is doubled.

An alternative scheme is to make a rough estimate of the error in the newly-designed mesh[§]. This algorithm is based on rearranging Eq. (5.38). Once the elemental refinement ratio Υ^{elem} is updated corresponding to the new mesh, the error of this new mesh is roughly estimated as

$$\{\|e\|_k\}_{\text{approx}} = \frac{(\Upsilon_k^{\text{elem}})^{\frac{2p+d}{2}} \zeta \|u\|}{\sqrt{N}} \tag{5.44}$$

for each element *k*. Combining the elemental errors together, we find

$$\{\|e\|\}_{\text{approx}}^2 = \sum_{k=1}^{N} \{\|e\|_k\}_{\text{approx}}^2 \tag{5.45}$$

And the nodes will be relocated only if

$$\{\|e\|\}_{\text{approx}} < \|e\| \tag{5.46}$$

[§]This estimate takes place before the mesh is actually graded.

This is to avoid bad mesh gradation, which in our experience can lead to instability of the solution during a nonlinear analysis. This rough estimation of the error norm does not detect the possible error that is caused by a too large aspect ratio.

The checking in case of goal-oriented measurement can be set in the same fashion as

- *Criterion 1*

$$\{\|e\|_k\|\epsilon\|_k\}_{\text{approx}} = \frac{(\Upsilon_k^{\text{elem}})^{2p+d}\zeta_{\text{goal}}\sum_{k=1}^{N}(\|u\|_k\|w\|_k)}{N} \tag{5.47}$$

  for each element k. Combining the elemental error together, the nodes will be relocated only if

$$\sum_{k=1}^{N}\{\|e\|_k\|\epsilon\|_k\}_{\text{approx}} < \sum_{k=1}^{N}(\|e\|_k\|\epsilon\|_k) \tag{5.48}$$

  and

- *Criterion 2*

$$\{|B_k(e,\epsilon)|\}_{\text{approx}} = \frac{(\Upsilon_k^{\text{elem}})^{2p+d}\zeta_{\text{goal}}\sum_{k=1}^{N}|B_k(u,w)|}{N} \tag{5.49}$$

  for each element k. Combining the elemental error, the nodes will be relocated only if

$$\sum_{k=1}^{N}\{|B_k(e,\epsilon)|\}_{\text{approx}} < \sum_{k=1}^{N}|B_k(e,\epsilon)| \tag{5.50}$$

Control of element aspect ratio

After mesh gradation, a problem that is often found is that of non-proportioned elements. Following [6], we define the aspect ratio Λ as

$$\Lambda := \frac{(\max_i h_i^{\text{side}})^2}{A} \tag{5.51}$$

where h_i^{side} refers to length of the element side i. A large aspect ratio implies a thin element which does not perform well in the finite element analysis, especially when linear interpolation applies. It is suggested that the aspect ratio should be well controlled. A threshold aspect ratio of four was used in [6]. However, in higher-order meshes, a higher ratio may be used as the performance of the thin elements is improved. Here, for our convenience, we control the element aspect ratio not to be larger than 6 for all cases.

When the gradation shows the aspect ratio equal to the threshold value, the gradation algorithm should stop. Actually, the element aspect ratio can be improved, for example, via a directional element subdivision or an element edge swapping scheme (see Figure 5.4). However, the analysis of such improvement techniques is not in the scope of this study.
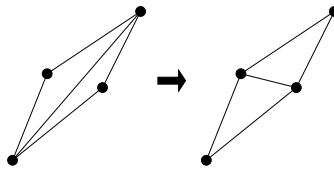
Figure 5.4 Edge swapping

5.4.3 Examples

Figure 5.5 shows some results from the mesh gradation technique applied to a linear elastic central crack problem. It is obvious that the choice of parameter *m* affects patterns of the graded mesh. In all cases, most of the mesh improvement occurs with the first gradation of the mesh, whereas further reduction of the error in subsequent steps is less significant. This example proves that, in case of singularity, the mesh gradation can even cut the error by half.

In another example, the *r*-adaptive scheme is applied to the Prandtl's punch test (Figure 5.6). As the solution is smoother, the finite element result does not show great improvement upon *r*-adaptivity as in the first example. With $m > 2.0$, the graded mesh exhibits large aspect ratios in some elements. This problem can be slightly improved by allowing boundary nodes to move freely along the boundary.

5.5 Variable transfer algorithms

During nonlinear solution control, the discretisation error may be measured and the finite element mesh can be improved accordingly. After each mesh change, one can choose either

- to continue the computation in the next step
- to recompute the solution at the current step and continue to the next step
- to restart the whole computation.

Generally, the third choice is not popular in nonlinear analysis as the computational effort applied in the earlier steps of computation is lost. Choosing to continue the computation, either by the first or the second choice requires an accurate *variable transfer*.

Values of the state variables, such as stresses, are generally stored in the integration points. Once the mesh is upgraded, either by mesh refinement, mesh enrichment or mesh gradation, the integration points are increased in number and/or relocated from their old positions. The values stored at these integration points need to be transferred to the new integration points, so that the nonlinear computation can be continued without restarting the whole computation.

Original (95.55%)

m=0.5

Step 1 (74.42%)   Step 2 (68.57%)   Step 3 (66.57%)   Step 4 (66.00%)

m=1.0

Step 1 (62.98%)   Step 2 (54.78%)   Step 3 (53.52%)   Step 4 (53.27%)

m=2.0

Step 1 (55.25%)   Step 2 (47.88%)   Step 3 (44.56%)   Step 4 (44.56%)
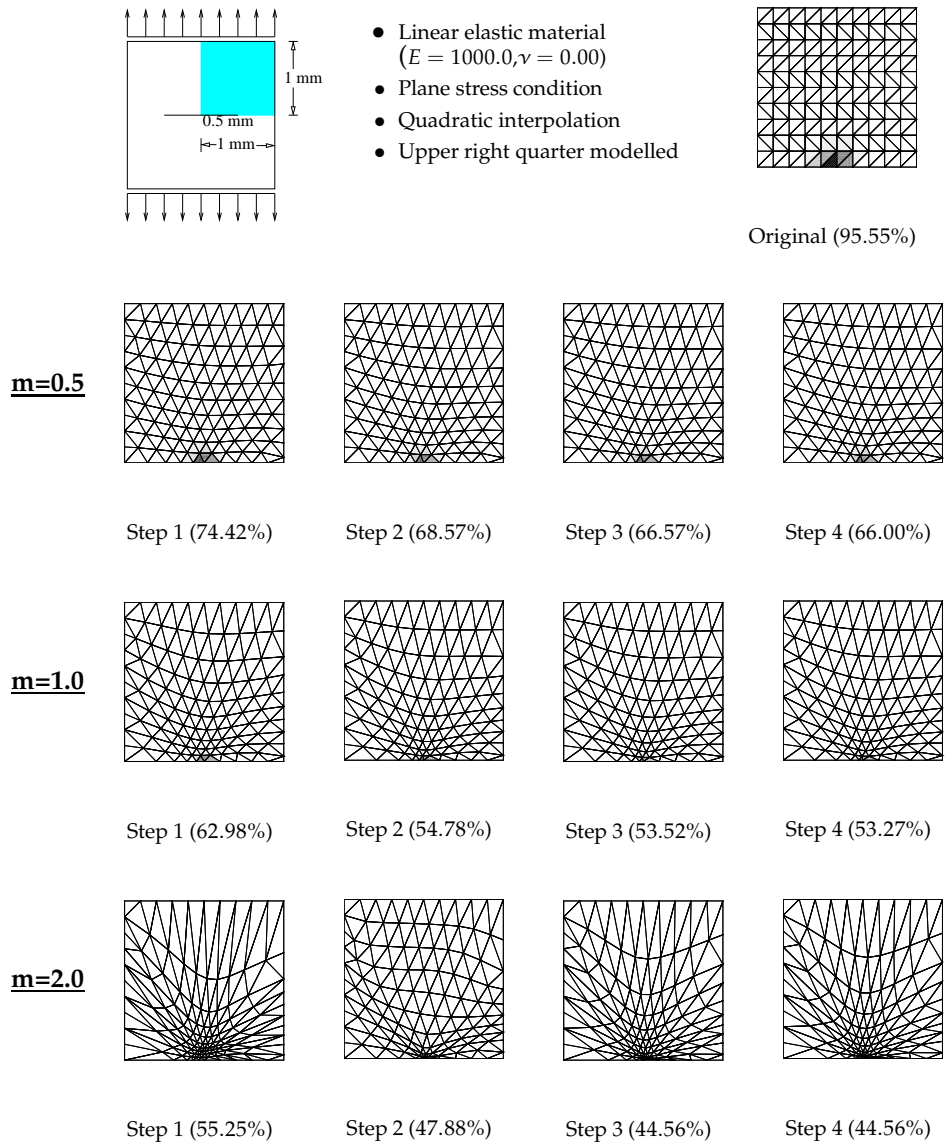
Figure 5.5 The graded mesh with parameters of $m = 0.5$, $m = 1.0$ and $m = 2.0$. Numbers in the brackets refer to the ratio $\|e\|/\|u\|$. Levels of the elemental error are indicated: darker color refers to higher level of error.

Original
(27.06%)

| $m = 0.5$ | $m = 1.0$ | $m = 2.0$ | $m = 3.0$ | $m = 4.0$ |
| (25.74%) | (24.64%) | (24.72%) | (28.46%) | (31.73%) |

| $m = 0.5$ | $m = 1.0$ | $m = 2.0$ | $m = 3.0$ | $m = 4.0$ |
| (24.48%) | (22.46%) | (20.62%) | (23.49%) | (29.17%) |
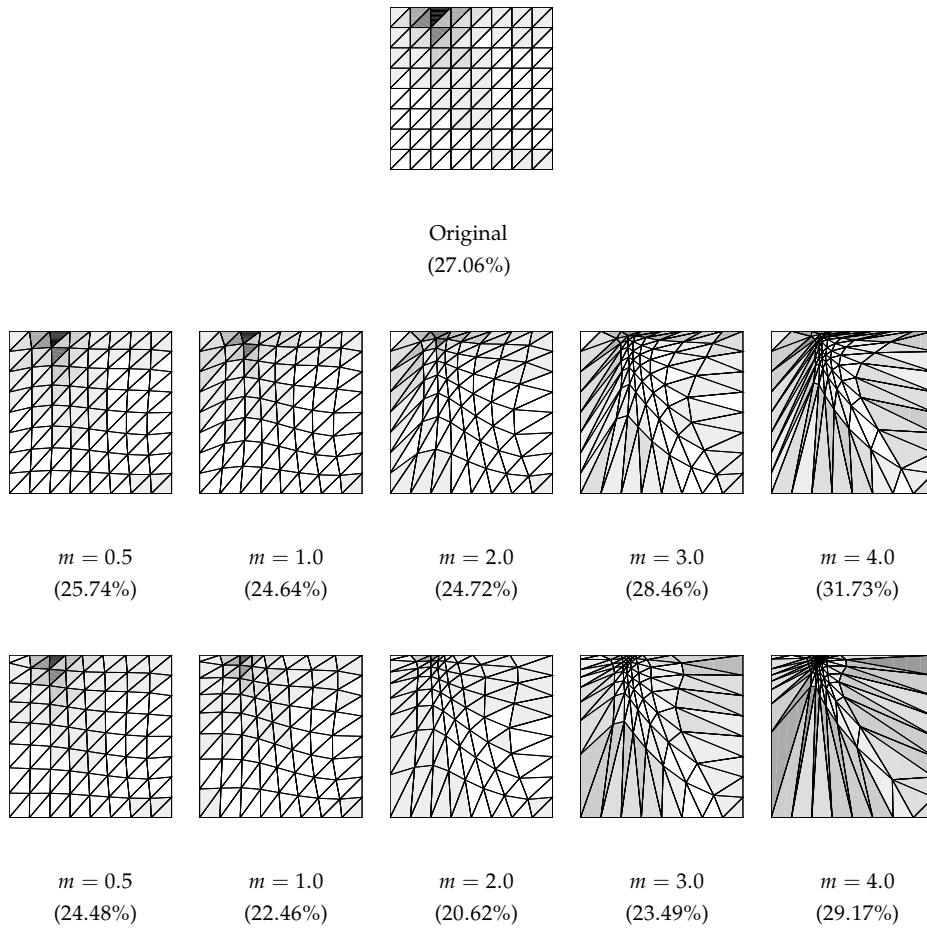
Figure 5.6 The Prandtl's punch test (cf. Figure 4.7) with uniform linear interpolation. The second row restricts the boundary nodes whereas the third row allows nodes on the left and upper edges to move freely. Only the first remeshing step is considered. Numbers in the brackets refer to the ratio $\|e\|/\|u\|$. Levels of the elemental error are indicated: darker color refers to higher level of error. Note that the element aspect ratio is not controlled in this example.
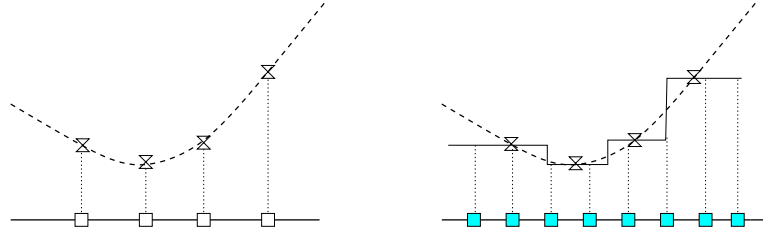
Figure 5.7 The closest point transfer algorithm [69].

5.5.1 Transfer of state variables

There are many techniques for variable transfer, two of which will be investigated in this research as they are simple and widely used.

Direct transfer

In a selected subdomain surrounding the target point, the data from an integration point in the old mesh (source, $f(x_j)$) can be transferred to an integration point in the new mesh (target, $f(x_{\text{target}})$) by

$$f(x_{\text{target}}) = \frac{\sum_{j=1}^{n} \varpi(x_j, x_{\text{target}}) f(x_j)\, d\Omega}{\sum_{j=1}^{N} \varpi(x_j, x_{\text{target}})\, d\Omega}. \tag{5.52}$$

where n refers to a number of the source points in the neighboorhood Ω_s. The scaling factor ϖ may be selected as a function of the distance from the source point x_j to the target point x_{target}

$$\varpi(x_j, x_{\text{target}}) = \frac{r - |x_{\text{target}} - x_j|}{r} \tag{5.53}$$

The local domain Ω_s is defined by a user-specified distance (or a radius) from the target point r.

It is worth noting that this type of transfer is independent of both old and new element connectivity and requires very low computational effort.

As an alternative, the data from the closest source point may simply be transferred to the target point with no scaling, i.e.

$$f(x_{\text{target}}) = f(x_{\text{closest}}) \tag{5.54}$$

where x_{closest} refers to the point which is the closest to the target point x_{target}. This method, known as the *closest point transfer* [69], can provide an acceptably accurate transfer by its very low diffusion property, which may be very important in capturing narrow regions with steep gradients, such as where strain localisation takes

place. Moreover, with its low computational cost, it is preferred to the Least-Squares fitting which requires solving the solution of small system of equations [69]. Figure 5.7 shows how the data is transferred by this method.

Shape function remapping transfer

As a typical method in adaptive finite element analysis, the state variables can be transferred between the meshes via the use of the finite element shape functions [56, 69]. First, data at the source points are mapped to nodes by extrapolation combined with nodal averaging. Then, the nodal data in each element are interpolated to the new integration points (See Figure 5.8 – path A).

The technique of extrapolation and nodal averaging, in fact, follows a similar idea as in the formation of consistent nodal forces, except that the data is scaled at nodes. The data at each integration point x are transferred to nodes (or modes in hierarchical concept) x_I via the use of nodal shape function $\phi_I(x)$ by

$$f_I = \frac{\int\limits_{\Omega} \phi_I(x) f(x) \, d\Omega}{\int\limits_{\Omega} \phi_I(x) \, d\Omega} \tag{5.55}$$

The technique is frequently used for contour plots which are usually based on the values at nodes. However, the extrapolated values from this nodal averaging are generally not as accurate as the values stored at the integration points. By this *smoothing-out* technique, the transfer to nodes may lead to a considerable loss in accuracy in the presence of high stress/strain gradients.

There is, indeed, a complication due to a need of an inverse mapping algorithm in case of quadrilateral elements. The inverse mapping can then be avoided by interpolation of the variables from the old nodes to the new nodes and subsequently from the new nodes to the new set of integration points. This procedure is considered as the second alternative to the one aforementioned, and is illustrated in Figure 5.8 – Path B.

5.5.2 Transfer of primary variables

Not only variables at integration points need to be remapped, the values corresponding to each degree of freedom must also be transported. As compared to the Lagrange elements where nodal locations are well-defined, the Legendre shape function can add another level of complication.

There are two techniques to deal with this problem, namely Lagrange-equivalence transfer and Least-squares fitting.

The first technique utilises the advantage of the Lagrange elements (isoparametric elements), which is missing in the Legendre elements (hierarchical shape functions). As the higher-order *modes* in the hierarchical framework cannot be located
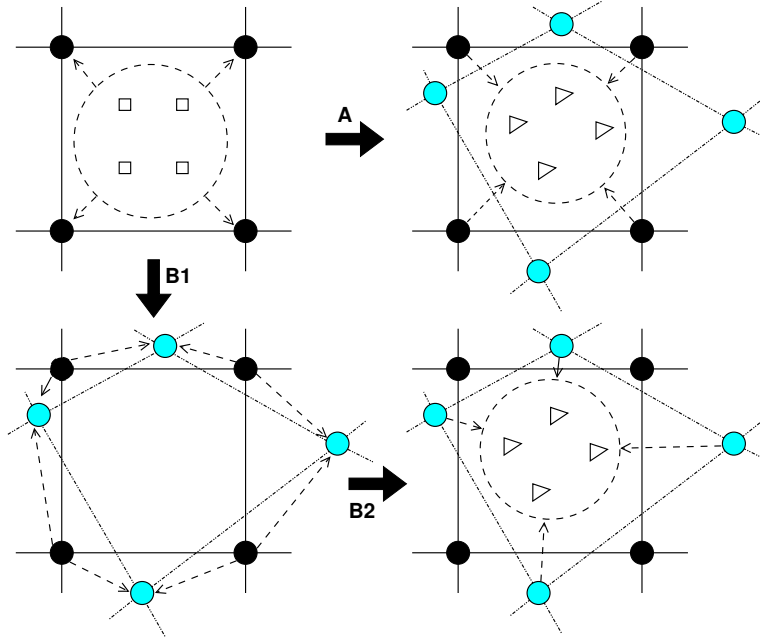
Figure 5.8 Transfer of variables at the integration points by shape function remapping algorithm.

at fixed points, the interpolation of these shape functions can be very difficult, if possible at all.

To be able to remap the hierarchical modes to the new mesh, a set of sample points is selected. These sample points are on the edges and the interior of each element and are picked according to the standard Lagrange elements of the same polynomial order. Once those *Lagrange-equivalence* points are set, the interpolation from the old mesh to the new mesh can be applied naturally (See Figure 5.9).

The values stored at the sampling points must then be mapped to the corresponding hierarchical modes. Since

$$u(x) = \sum_{i=1}^{n} \phi_i(x) a_i, \tag{5.56}$$

it is possible to remap the Lagrangian modes (at sample points) back to the Legendre scheme by seeking $a_i, i = 1, .., n$, where n denotes the number of nodes, from

$$\begin{bmatrix} \phi_1(x_1) & \phi_2(x_1) & \cdots & \phi_n(x_1) \\ \phi_1(x_2) & \phi_2(x_2) & \cdots & \phi_n(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ \phi_1(x_n) & \phi_2(x_n) & \cdots & \phi_n(x_n) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} u(x_1) \\ u(x_2) \\ \vdots \\ u(x_n) \end{bmatrix} \tag{5.57}$$
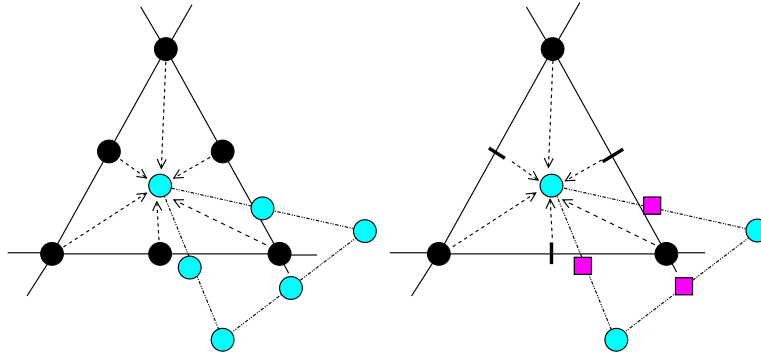
Figure 5.9 Transfer of nodal variables by mapping to equivalent Lagrange elements. The left subfigure refers to interpolation from the nodes in the old mesh to those in the new mesh in the standard Lagrange elements. The right subfigure shows the equivalent transfer from hierachical modes to the Lagrange-equivalent points (indicated by the squares).

where n refers to the number of Lagrange-equivalence points, which is also equal to the number of the shape functions. Only the element shape functions in each direction are considered in the matrix inversion. Thus, the computational cost for the remapping is considerably low.

The accuracy of the transfer via the second technique, i.e. Least-squares fitting, depends on the number of sampling points, which may be based on integration points or Lagrange-equivalence nodes. If the latter choice is considered, the optimal answer would be the same as the first technique where simple interpolation and the equivalence transfer are applied.

Of course, the more sampling points we use, the more computational effort is required in the matrix inversion in the Least-Squares fitting. Also, employing a number of sampling points higher than the number of shape functions may be beyond necessity. In this research, the sampling points are set equal to the number of displacement modes and thus Lagrange-equivalence transfer is our choice.

5.6 Remarks

In this chapter, the adaptive techniques to be used in this thesis have been introduced. The *a posteriori* error estimates, i.e. the energy-norm based error measure as well as the two goal-oriented error measures presented in the previous chapters, have been employed in setting of adaptive and optimality criteria. Also based on the *a priori* error estimates and uniform error distribution assumptions, the criteria can be formulated straightforwardly. Particular attention has been paid to the formulation of mesh gradation schemes in which the aim is to concentrate elements in zones of large error while at the same time controlling the element aspect ratio.

During a nonlinear analysis, the finite element computation may be continued or restarted from the beginning after a mesh change. If the former choice is applied, it is necessary to correctly transfer the variables including those stored at nodes and at the integration points. To provide a considerably accurate and reasonably inexpensive transfer, the variable transfer techniques that will be studied in the next chapter include all techniques mentioned in this chapter.

Mesh adaptivity for continuous failure

Quasi-brittle materials such as plain concrete exhibit so-called *strain softening behaviour*. Once damage is initiated, these materials are still able to carry load while gradually losing their strength. During such processes, deformation tends to concentrate in some parts of the material, subsequently forming cracks which finally leads to failure.

The phenomenon can be suitably modelled by means of *damage mechanics*. The formation and growth of a microstructural crack is modelled via continuous damage variables, such that failure can be simulated entirely within a continuum mechanics framework. During damage growth, the material gradually loses its integrity and its stored energy is dissipated. Unfortunately, a straightforward inclusion of a damage-driven dissipation results in mathematical ill-posedness in the post-peak regime of the structural response, causing a zero width of the localisation zone and subsequently zero energy dissipation. As a result, the finite element size controls the localisation width, leading to so-called *mesh dependence* [72] in the sense that the numerical results do not converge when the discretisation is refined in finite element modelling. Error estimation and adaptivity would consequently not give meaningful results.

Information on microscopic material behaviour must be taken into account in the continuum model. This can be achieved by enhancing the continuum model with an intrinsic length scale to avoid the loss of ellipticity. As a regularised model, the implicit gradient-enhanced damage model [70] is chosen for this study. Even though the numerical results converge upon refinement of discretisation, the finite element modelling requires an adequate mesh discretisation in order to accurately describe the fracture processes. We will show in this chapter the importance of a *good* finite element mesh in damage modelling.

6.1 The gradient-enhanced damage model

Within the context of continuum damage mechanics, material gradually loses its load-carrying capacity as a result of the appearance of microstructural cracks. This material degradation process, described here in the continuum damage mechanics concept by the introduction of a scalar damage parameter ω, is cast in a stress-strain relation written as

$$\boldsymbol{\sigma} = (1 - \omega)\boldsymbol{D}^e : \boldsymbol{\varepsilon} \tag{6.1}$$

where $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ denote stresses and strains. As the model is formulated based on elasticity, the linear-elastic constitutive tensor $\boldsymbol{D}^e$ is employed. The damage parameter ω ranges from 0 (for virgin condition) to 1 (for fully damaged condition) and is defined as a function of a history parameter κ, i.e. $\omega := \omega(\kappa)$.

Representing the largest value of the deformation in the loading history, κ is obtained from

$$\kappa = \max(\kappa_0, \varepsilon_{\text{eq}}) \tag{6.2}$$

where κ_0 is a user-specified damage threshold and ε_{eq} refers to an equivalent strain, which is a scalar invariant representing the strains. Some definitions of this equivalent strain are, for example,

- *Mazars definition* [55]

$$\varepsilon_{\text{eq}} = \sqrt{\sum_{i=1}^{3} \langle \varepsilon_i \rangle^2} \tag{6.3}$$

  where ε_i denotes the principal strain and the positive principal strain $\langle \varepsilon_i \rangle$ is defined as

$$\langle \varepsilon_i \rangle = \frac{\varepsilon_i + |\varepsilon_i|}{2} \tag{6.4}$$

- *Modified von Mises definition* [26]

$$\varepsilon_{\text{eq}} = \frac{k-1}{2k(1-2\nu)} I_1 + \frac{1}{2k}\sqrt{\frac{(k-1)^2}{(1-2\nu)^2}{I_1}^2 + \frac{12k}{(1+\nu)^2}J_2} \tag{6.5}$$

  where k is the ratio of the compressive and tensile strength, and the strain invariants I_1 and J_2 are defined as

$$I_1 = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz} \tag{6.6}$$

$$J_2 = \frac{(\varepsilon_{xx}^2 + \varepsilon_{yy}^2 + \varepsilon_{zz}^2 - \varepsilon_{xx}\varepsilon_{yy} - \varepsilon_{yy}\varepsilon_{zz} - \varepsilon_{zz}\varepsilon_{xx})}{3} + \varepsilon_{xy}^2 + \varepsilon_{yz}^2 + \varepsilon_{zx}^2 \tag{6.7}$$

Damage evolves when the Kuhn-Tucker conditions

$$f \leq 0, \ \dot{\kappa} \geq 0, \ \dot{\kappa} f = 0 \tag{6.8}$$

are satisfied. The loading function f is defined as

$$f = \varepsilon_{eq} - \kappa \tag{6.9}$$

By means of softening laws, the damage growth can be described. Two examples of such laws are

- *Linear softening law*

$$\omega = 1 - \frac{\kappa_0}{\kappa} \frac{\kappa_c - \kappa}{\kappa_c - \kappa_0} \qquad \text{if} \quad \kappa_0 \leq \kappa \leq \kappa_c \tag{6.10}$$

  where the damage is fully developed ($\omega = 1$) when $\kappa > \kappa_c$ (ultimate strain).

- *Exponential softening law* [71]

$$\omega = 1 - \frac{\kappa_0}{\kappa}(1 - \alpha + \alpha \exp(-\beta(\kappa - \kappa_0))) \qquad \text{if} \quad \kappa \geq \kappa_0 \tag{6.11}$$

  where α and β are material parameters controlling the residual stress and the damage growth rate, respectively.

Implicit gradient formulation

The above formulation can be referred to as the standard local damage model. As mentioned above, due to a lack of microstructural information, the localisation zone tends to have a zero width. The above model suffers from mathematical ill-posedness and, consequently, a severe mesh dependence [72]. To overcome these problems, some techniques have been introduced. In this study, we employ regularisation based on replacing the local equivalent strain ε_{eq} by the nonlocal equivalent strain $\check{\varepsilon}_{eq}$ in Eqs. (6.2) and (6.9).

By averaging the local equivalent strain in the gradient form, the nonlocal equivalent strain can be defined as an implicit gradient enhancement form

$$\check{\varepsilon}_{eq} - c_2 \nabla^2 \check{\varepsilon}_{eq} + c_4 \nabla^4 \check{\varepsilon}_{eq} - c_6 \nabla^6 \check{\varepsilon}_{eq} + c_8 \nabla^8 \check{\varepsilon}_{eq} - \ldots = \varepsilon_{eq} \tag{6.12}$$

where c_i are material parameters based on the intrinsic length scale l_{int} and are defined as

$$c_{2n} = \frac{1}{n!} \left(\frac{l_{int}}{2} \right)^{2n} \qquad n = 1, 2, 3, \ldots \tag{6.13}$$

The intrinsic length scale has the dimension of length and it is a representation of the underlying microstructure of the material. Inclusion of an internal length scale

as done in Equation (6.12) ensures that the failure zone has a finite width, which in turns guarantees a non-zero energy dissipation upon mesh refinement. Thus, the problems of the local damage model are overcome [70, 71].

Using a truncation after the second order term (cf. Eq. (6.12)), the implicit gradient-enhanced model is formulated by adding another set of equations

$$\breve{\varepsilon}_{\text{eq}} - c_2 \nabla^2 \breve{\varepsilon}_{\text{eq}} = \varepsilon_{\text{eq}} \tag{6.14}$$

to the standard finite element formulation. Applying integration by parts, the implicit gradient enhancement formulation can be cast in a weak form as

$$\int_\Omega \delta \breve{\varepsilon}_{\text{eq}} \breve{\varepsilon}_{\text{eq}} \, d\Omega + \int_\Omega \nabla \delta \breve{\varepsilon}_{\text{eq}} c_2 \nabla \breve{\varepsilon}_{\text{eq}} \, d\Omega = \int_\Omega \delta \breve{\varepsilon}_{\text{eq}} \varepsilon_{\text{eq}} \, d\Omega \tag{6.15}$$

with a set of homogeneous natural boundary conditions [70]

$$\nabla \breve{\varepsilon}_{eq} \cdot \boldsymbol{n} = 0 \text{ on } \partial \Omega \tag{6.16}$$

where $\boldsymbol{n}$ is the outward unit normal on the boundary Γ. Note that to improve the conditioning of the stiffness matrix, using Young's modulus (E) as a multiplying (scaling) factor for both sides of Eq. (6.15) may be considered. For a discussion on the finite element implementation of the model, the reader is referred to [86, 88].

6.2 Error analyses

The gradient-enhanced damage model formulated in the last section requires two sets of unknowns, namely the displacement field and the nonlocal equivalent strain field. For the second-order implicit gradient formulation, cf. Eq. (6.14), $\mathcal{C}^0$ shape functions are required. We define here the set of unknowns

$$\breve{\boldsymbol{u}} = \{\boldsymbol{u}, \breve{\varepsilon}_{\text{eq}}\} \tag{6.17}$$

which is the solution of the system of equations

$$\mathcal{B}^{\text{tang}}(\Delta \breve{\boldsymbol{u}}_{(h,p)}, \boldsymbol{v}_{(h,p)})|_{(t-1:t)} = \Delta \mathcal{F}(\boldsymbol{v}_{(h,p)})|_{(t-1:t)} \tag{6.18}$$

and $\breve{\boldsymbol{u}}$ is updated during the Newton-Raphson iterative scheme, i.e.

$$\breve{\boldsymbol{u}}_{(h,p)}|_{(t)} = \breve{\boldsymbol{u}}_{(h,p)}|_{(t-1)} + \Delta \breve{\boldsymbol{u}}_{(h,p)}|_{(t-1:t)} \tag{6.19}$$

It is natural to include all unknown degrees of freedom existing in the formulation in the error analysis, as all of them are primary unknowns in the finite element computation[*]. We seek the error solution $\breve{\boldsymbol{e}} = \{\boldsymbol{e}_u, \boldsymbol{e}_{\breve{\varepsilon}}\}$ by solving a set of error

---

[*]The system of error equations is different from the case of nonlocal integral-based damage formulation [20, 72] and the explicit gradient-enhanced damage formulation [19]. In those schemes, the displacement field is the only primary unknown in the finite element computation.

equations

$$
\begin{aligned}
\mathcal{B}^{\text{tang}}(\breve{e}_{(\tilde{h},\tilde{p})}, v)|_{(t)} &= \mathcal{R}^{\breve{u}}(v_{(\tilde{h},\tilde{p})})|_{(t)} \\
&= \mathcal{F}(v_{(\tilde{h},\tilde{p})})|_{(t)} - \mathcal{B}(\breve{u}_{(h,p)}, v_{(\tilde{h},\tilde{p})})|_{(t)}
\end{aligned}
\tag{6.20}
$$

The error computation follows the process described in Chapter 3, whereby a series of patch-based computations is solved instead of a global computation.

Setting a proper norm

The finite element discretisation leads to a consistent tangent stiffness matrix of the form

$$
K^{\text{tang}} = \mathcal{B}^{\text{tang}}(\breve{\phi}, \breve{\phi}) = \begin{bmatrix} K_{uu} & K_{u\breve{\xi}} \\ K_{\breve{\xi}u} & K_{\breve{\xi}\breve{\xi}} \end{bmatrix}
\tag{6.21}
$$

The components of the stiffness matrix, addressed in Eq. (6.21), are defined as

$$
K_{uu} = \int_{\Omega} \nabla\phi_u : (1-\omega)D^e : \nabla\phi_u \, d\Omega
\tag{6.22}
$$

$$
K_{u\breve{\xi}} = -\int_{\Omega} \nabla\phi_u : s_{u\breve{\xi}} : \phi_{\breve{\xi}} \, d\Omega
\tag{6.23}
$$

$$
K_{\breve{\xi}u} = -\int_{\Omega} \phi_{\breve{\xi}} : s_{\breve{\xi}u} : \nabla\phi_u \, d\Omega
\tag{6.24}
$$

$$
K_{\breve{\xi}\breve{\xi}} = \int_{\Omega} \left(\phi_{\breve{\xi}} : \phi_{\breve{\xi}} + \nabla\phi_{\breve{\xi}} : c_2\nabla\phi_{\breve{\xi}} \right) \, d\Omega
\tag{6.25}
$$

where

$$
s_{\breve{\xi}u} = \frac{\partial \varepsilon_{\text{eq}}}{\partial \varepsilon} \quad \text{and} \quad s_{u\breve{\xi}} = \frac{\partial \omega}{\partial \breve{\xi}} D^e \varepsilon
\tag{6.26}
$$

Note that different shape functions ϕ_u and $\phi_{\breve{\xi}}$ are used for the two sets of unknowns.

By using a softening model, it is possible that the computation of the error norm via the use of the consistent tangent stiffness matrix K^{tang} leads to a negative value and thus the norm defined in Chapter 3 becomes meaningless. To avoid such problems, we employ here only those parts of the stiffness matrix which include K_{uu} and $K_{\breve{\xi}\breve{\xi}}$. It is noticeable from the discretised equations that the terms $K_{u\breve{\xi}}$ and $K_{\breve{\xi}u}$ do not always provide positive-definite contributions to the global stiffness matrix. On the other hand, K_{uu} and $K_{\breve{\xi}\breve{\xi}}$ remain positive definite in the whole loading process. To maintain a mathematically meaningful norm for the error, the interaction between the two sets of degrees of freedom (i.e. u and $\breve{\xi}_{\text{eq}}$) is neglected, thus avoiding the occurrence of a negative-definite matrix in the error norm computation. The modified stiffness matrix for the norm computation reads

$$
K^{\text{tang}^+} = \mathcal{B}^{\text{tang}^+}(\breve{\phi}, \breve{\phi}) = \begin{bmatrix} K_{uu} & 0 \\ 0 & K_{\breve{\xi}\breve{\xi}} \end{bmatrix}
\tag{6.27}
$$

Note that K_{uu} is also the *secant stiffness matrix*. The energy norm of solution and error can then be written, respectively, as

$$
\begin{aligned}
(\|\breve{u}\|^{\text{tang}^+})^2 \quad &= \mathcal{B}^{\text{tang}^+}(\breve{u}, \breve{u}) \quad = (\|u\|^{\text{tang}^+})^2 + (\|\breve{\varepsilon}_{\text{eq}}\|^{\text{tang}^+})^2 \\
&\qquad\qquad\qquad\quad = u : K_{uu} : u + \breve{\varepsilon}_{\text{eq}} : K_{\breve{\varepsilon}\breve{\varepsilon}} : \breve{\varepsilon}_{\text{eq}}
\end{aligned}
\tag{6.28}
$$

$$
\begin{aligned}
(\|\breve{e}\|^{\text{tang}^+})^2 \quad &= \mathcal{B}^{\text{tang}^+}(\breve{e}, \breve{e}) \quad = (\|e_u\|^{\text{tang}^+})^2 + (\|e_{\breve{\varepsilon}}\|^{\text{tang}^+})^2 \\
&\qquad\qquad\qquad\quad = e_u : K_{uu} : e_u + e_{\breve{\varepsilon}} : K_{\breve{\varepsilon}\breve{\varepsilon}} : e_{\breve{\varepsilon}}
\end{aligned}
\tag{6.29}
$$

These norms will be used in this chapter, in setting the adaptive criteria and driving mesh adaptivity.

6.3 Central transverse crack test

The first numerical example is the central transverse crack test, described in Figure 6.1. Due to symmetry, only the upper right quarter is modelled under a plane stress condition. We apply the displacement control algorithm with a proportionally prescribed displacement of $\overline{u} = 0.00001$ mm for each incremental step in the full Newton-Raphson iterative scheme.

To investigate the *h*-factor of the finite element discretisation, we select three uniform triangular meshes, namely Mesh 0, Mesh 1 and Mesh 2 (cf. Figure 6.2). For investigation of the *p*-factor, four orders of interpolation, ranging from linear order ($p = 1$) to quartic order ($p = 4$), are applied in a hierarchical manner. Details of these reference meshes are given in Table 6.1.

6.3.1 Preliminary investigation

Effects of mesh discretisation on FE solutions

To examine the effects of mesh discretisation on finite element results, preliminary tests are carried out based on uniform meshes with uniform orders of interpolation. It is shown in Figure 6.3 (left) that the load-displacement relation for the coarsest mesh (Mesh 0) with linear interpolation ($p = 1$) is significantly different from the
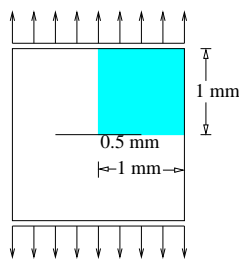
- Gradient-enhanced damage model [70]
  ($E = 1000.0$ MPa, $\nu = 0.00$)
- Plane stress condition
- Thickness = 1 mm
- Exponential softening damage law (cf. Eq.(6.11))
  ($\kappa_i = 0.0003, \alpha = 0.99, \beta = 1000.0$)
- Mazars equivalent strain definition (cf. Eq.(6.3))
- Upper right quarter modelled

Figure 6.1 The central transverse crack test.
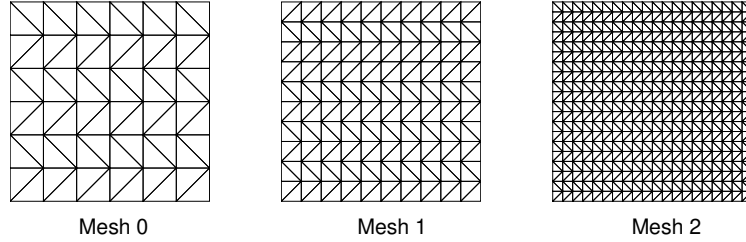
Figure 6.2 Meshes used in the finite element analysis of the central transverse crack test.

Mesh	No. of Nodes	No. of Elements	p-order	NDOFs
0	49	72	1	147
			2	507
			3	1083
			4	1875
1	121	200	1	363
			2	1323
			3	2883
			4	5043
2	441	800	1	1323
			2	5043
			3	11163
			4	19683

Table 6.1 Information of fixed meshes used in the central transverse crack test.
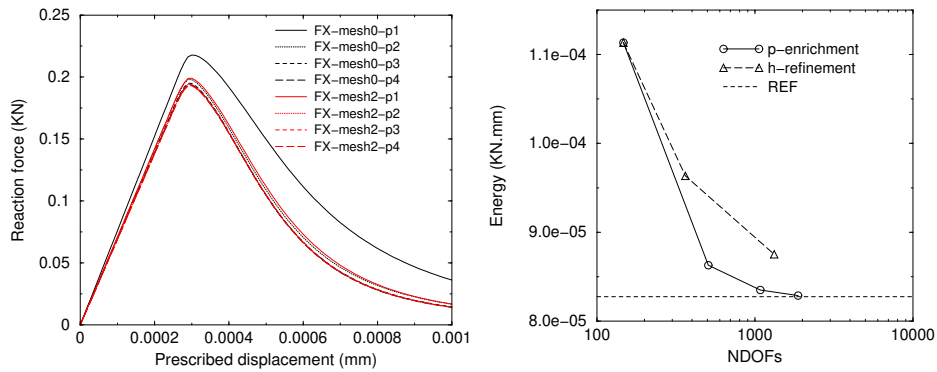
Figure 6.3 Load-displacement relations (left) and corresponding dissipated energy (right) for the central transverse crack test. The abbreviation FX refers to the fixed computational mesh, i.e. no mesh adaptivity is activated. The abbreviation REF denotes the result obtained from the reference mesh, i.e. Mesh 2 with quartic interpolation.
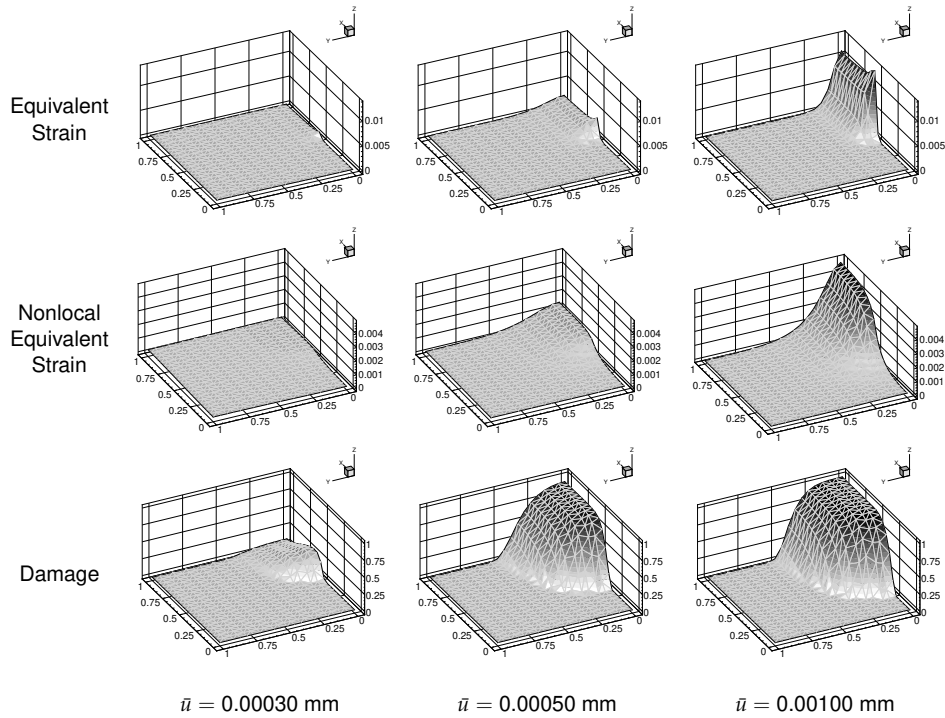
Figure 6.4 Evolution of local equivalent strain, nonlocal equivalent strain and damage obtained from Mesh 2 with quartic interpolation.

rest of the results. We observe that the finite element computation is likely to over-estimate the reaction forces corresponding to the prescribed displacements. Both *h*-factor and *p*-factor can, indeed, improve the accuracy of the solution in this global sense. As an additional observation in Figure 6.3 (right), the response obtained by *p*-extension appears to improve the finite element solution slightly faster than the *h*-extension, if considered at the same computational costs.

We select Mesh 2 with quartic interpolation to describe evolutions of the damage parameter and the equivalent strain during the softening process. It is evident that damage appears initially before the peak load is reached (precisely, incremental step 26 or $\overline{u} = 0.00026$ mm). The material is still able to carry more load up to Step 30 ($\overline{u} = 0.00030$ mm), followed by a global softening. In Figure 6.4, it can be seen that, in the post peak regime, the strain localises ahead of the crack tip in a small zone and the damage grows accordingly.

Profiles of the damage and the equivalent strain at a cut section $x = 1.0$ mm, shown in Figure 6.5, reveal the influence of mesh discretisation in the damage analysis. The meshes under investigation are Mesh 1 and Mesh 2 with linear and quadratic interpolations. We compare two mesh improvement approaches by up-
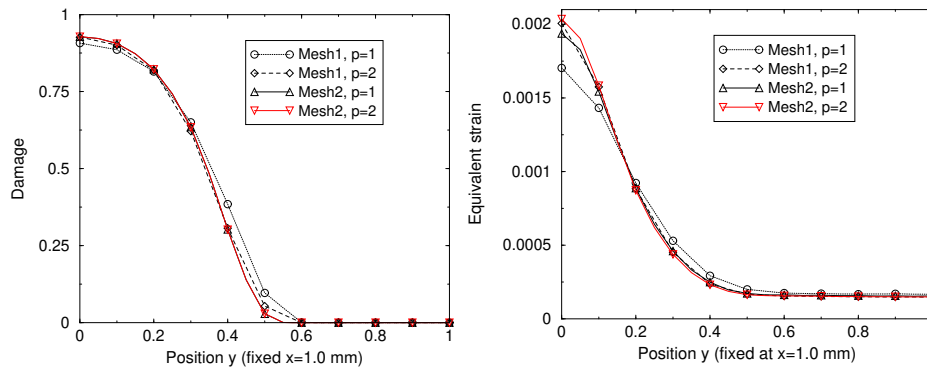
Figure 6.5 Profiles of damage parameter and equivalent strain measured along the $x = 1.0$ mm when $\bar{u} = 0.00050$ mm.
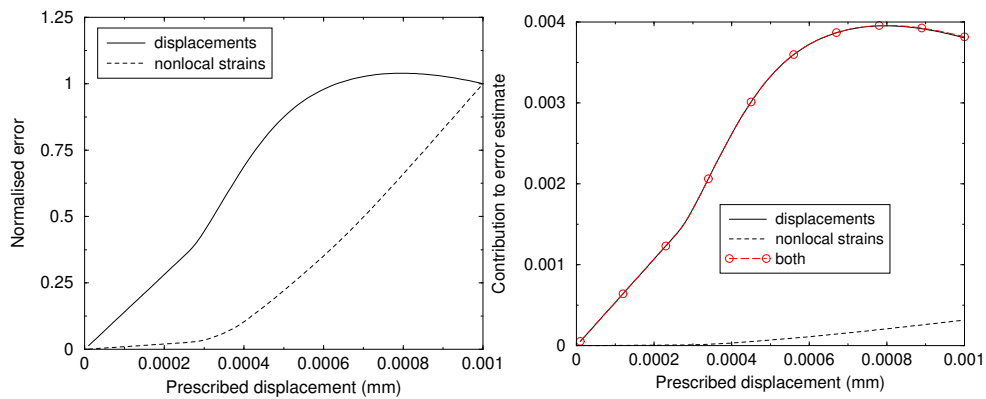
Figure 6.6 Contribution of all degrees of freedom to norm measure, using Mesh 1 with linear interpolation.

grading Mesh 1 with linear interpolation using similar additional numbers of degrees of freedom (i.e. Mesh 1 with quadratic interpolation and Mesh 2 with linear interpolation). As can be seen in Figure 6.3 (left), both the h-factor and the p-factor can improve the damage and equivalent strain profiles at the selected cut section. Nevertheless, improving the mesh by applying higher-order interpolation depends largely on the topology of the mesh. If the mesh possesses an insufficient resolution at the region of high-gradient solutions, using higher order interpolation (p-factor) can only improve the local solution up to a limited extent.

Analysis of error information

As mentioned in the last section, we estimate the errors of all primary unknowns in the discretised equations. Recalling Eq. (6.29), the total error estimate $\|e\|^{\text{tang}+}$ consists of two independent contributions, namely the error in the displacement field $\|e_u\|^{\text{tang}+}$ and the error in the nonlocal equivalent strain field $\|e_{\tilde{\varepsilon}_{\text{eq}}}\|^{\text{tang}+}$. These two contributions to the error estimate do not have the same dimension, thus the computed error norm is no longer an energy measure. However, by the fact that both unknown fields are discretised and solved via the finite element analysis, both two contributions should be taken into account in the error analysis.

Figure 6.6 shows individual contributions of the error estimates obtained at the end of each loading step in the finite element computation. To observe the trend of both error contributions, in Figure 6.6 (left), we present each contribution by normalising it with respect to the value at the end of the loading process ($\overline{u} = 0.001$ mm). It appears that both contributions increase at higher rates after the peak in the load-displacement response is reached. This does not hold anymore towards the end, where the estimated error in the displacement field is decreasing while the estimated error in the nonlocal equivalent strain field grows increasingly till the end.

It is, however, observed in Figure 6.6 (right) that the error in the nonlocal equivalent strain field provides much less contribution to the total error measure than the one in the displacement field[†]. Thus, an economical alternative of this error estimation would be to assess the error in the displacement field only, disregarding the error in the nonlocal equivalent strain field. Despite such observation, we employ in this study the contributions of both fields in estimating the error of the finite element computation, as the programming for the error estimation part should be general for use in various applications.

6.3.2 Mesh adaptive tests

In the central transverse crack test, the discretisation error is approximated at the end of each numerical step. Filtered by the adaptive criteria (cf. Section 5.2), some regions in the finite element mesh may be improved via *h*-adaptivity, *r*-adaptivity or *p*-adaptivity. Comparisons between the three methods, in terms of performance, will be discussed in this subsection. We control the error not to be beyond 15% of the solution norm (cf. Eq. (6.28)) for all adaptive tests[‡].

[†]Note that this is not caused by the different dimensions of the two solutions since each is normalised separately prior to this summation.

[‡]However, for *r*-adaptivity, this critical value is merely a control parameter for mesh movement. Whether or not the mesh is sufficiently improved to reach the critical value depends on the original mesh arrangement.
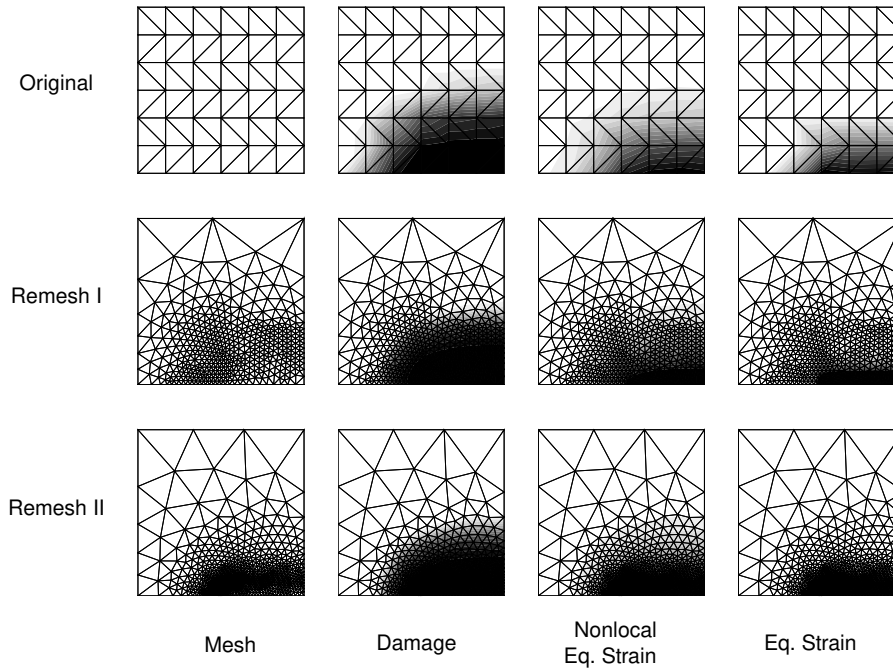
| | Mesh | Damage | Nonlocal Eq. Strain | Eq. Strain |

Figure 6.7 Meshes used in the *h*-adaptivity Scheme 3 and the corresponding state variables obtained at the end of the computation ($\bar{u} = 0.001$ mm). The computation is restarted from the beginning for every mesh change.

h-adaptivity

In *h*-adaptivity, the element sizes can be designed via the optimality criterion relying on the *a priori* convergence assumption (cf. Section 5.3). However, the capabilities of the used mesh generator determine the quality of the resulting mesh. Different meshes are obtained via different generators. Here, we use the mesh generator NeGe[§], which is based on an advancing front method, and insert it as an interface to the finite element computation. Some important points, regarding *h*-adaptivity in the gradient-enhanced damage model, are studied in the following.

(A) Comparisons of h-adaptive schemes after mesh change

As mentioned in Section 5.5, the subsequent process after a mesh change can either be to continue the computation in the next step (Scheme 1), to recompute the solution in the current step (Scheme 2), or to restart the whole computation (Scheme 3). For simplicity, we do not consider the second choice, as its results are expected to be quite similar to those of Scheme 1, and compare the other two. We measure the error in the first scheme at the end of each computational step, while

---

[§]The program was originally developed by Johann Sienz, University of Wales Swansea, United Kingdom.

$\bar{u} = 0.00020$ mm

$\bar{u} = 0.00040$ mm

$\bar{u} = 0.00060$ mm

$\bar{u} = 0.00100$ mm

| Mesh | Damage | Nonlocal Eq. Strain | Eq. Strain |
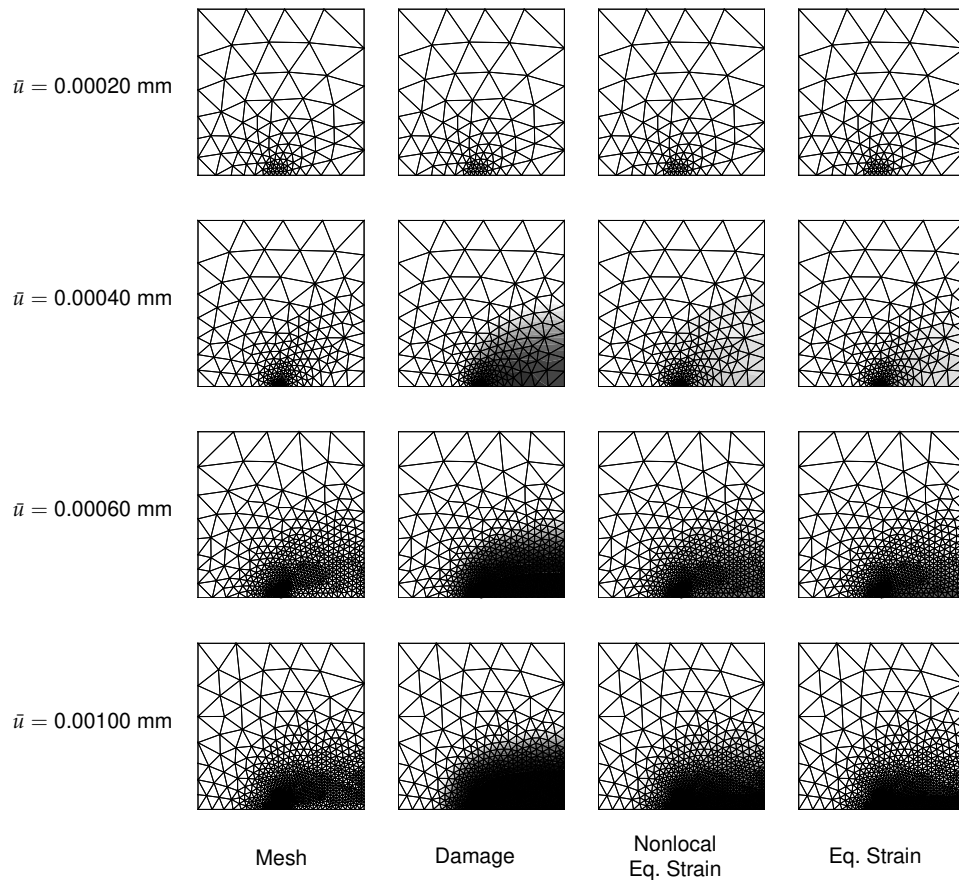
Figure 6.8 Meshes used during the *h*-adaptivity Scheme 1 and the corresponding state variables during the computation. The computation is continued after the variable transfer by the closest point approach. This test is referred to as H-CPT.

the error is measured once at the end of the computation in the third scheme. The results from both strategies are shown in Figure 6.7 and Figure 6.8, where the final meshes at the final step of the computation show great similarity. Note that, via using the third adaptive scheme (cf. Figure 6.7), the original mesh (denoted as H-Original) is re-designed to H-Remesh I and subsequently to H-Remesh II after two whole computations.

Of course, an obvious advantage of Scheme 3 is that it does not require history variable transfer during the computation, thus avoiding *another* error[¶]. A disadvantage of the strategy, however, is that the solution path depends greatly on the designed mesh that relies upon only one error analysis. If the initial finite element

---

[¶]Here, we refer to the error due to the transfer of history variables during the nonlinear computation.
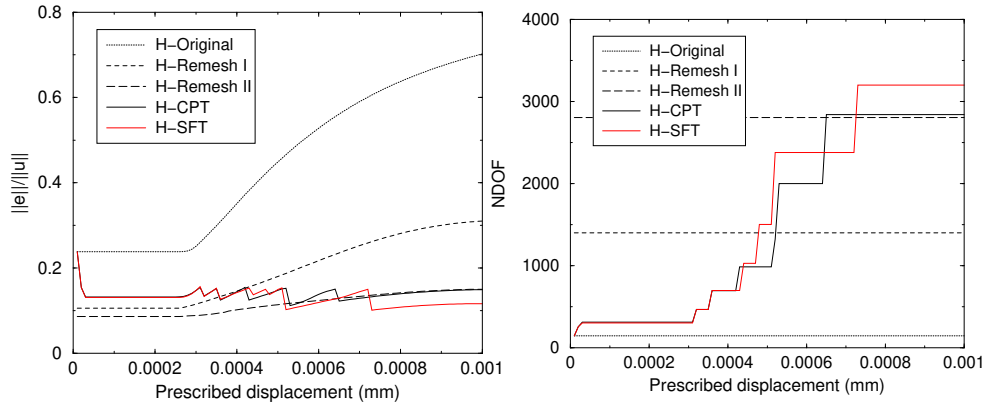
Figure 6.9 Comparison of the estimated error (left) and the number of degrees of freedom used (right) between different *h*-adaptive schemes during the computation. Linear interpolation ($p = 1$) is used.
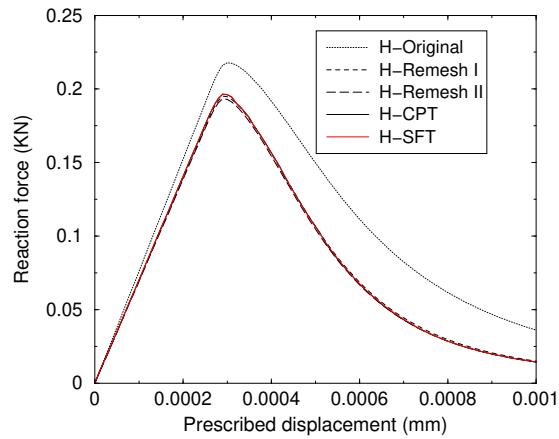
Figure 6.10 Comparison of the load-displacement relations between different *h*-adaptive schemes during the computation. Linear interpolation ($p = 1$) is used.

discretisation is inadequate, a completely wrong failure mechanism may be triggered, and the error at the end of the analysis, dominated by the steep gradients in the failure zone, may lead to completely wrong adaptive information. Such a problem, fortunately, does not appear in this very test. Superiorly however, measuring the error during the computation (Scheme 1) makes sure that the error information is more relevant to where the discretisation should be more needed at that stage of the computation.
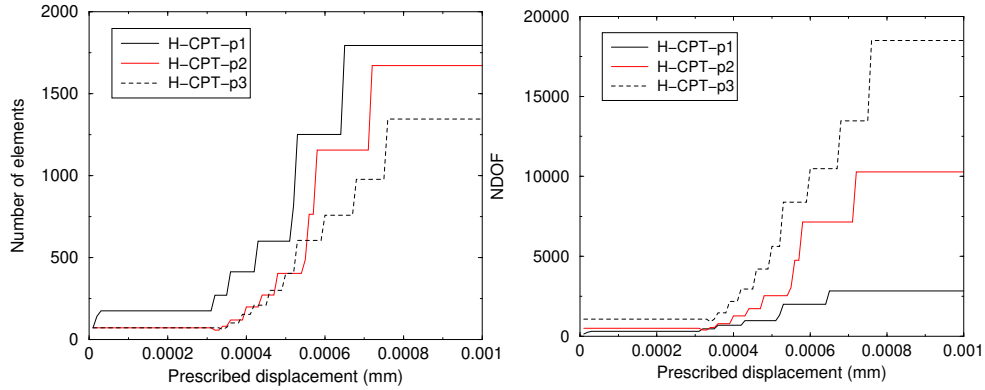
Figure 6.11 Comparison of the number of elements (left) and the number of degrees of freedom (right) used between meshes of different orders during the computation with *h*-adaptivity scheme 1.

(B) Efficiency of the h-adaptive schemes

A global performance of the two *h*-adaptive schemes (i.e., Scheme 1 versus Scheme 3) can be comparatively considered in terms of the number of degrees of freedom. Via different adaptive schemes, Figure 6.9 shows how the discretisation errors are reduced and the corresponding number of degrees of freedom used during the computation. Note that H-Remesh I and H-Remesh II refer to the two meshes used during Scheme 3. The adaptive meshes H-CPT and H-SFT, respectively based on the closest point transfer (CPT) and the shape function remapping transfer (SFT), refer to Scheme 1.

Evidently in this test, both schemes finally employ almost the same number of degrees of freedom at the end of the computation. The load-displacement curves in Figure 6.10 also show similar responses. The mesh providing the most accurate response, regardless of the number of degrees of freedom used, in comparison to the reference mesh[‖], is probably the final mesh obtained from Scheme 3, where the mesh is rearranged prior to each computation to capture the damage path obtained at the final stage of computation.

(C) Interpolation order of the mesh

The examples shown above are based on a linear interpolation. We investigate in Figure 6.11 the efficiency of the *h*-adaptive technique (Scheme 1) as higher-order interpolations are applied. As expected, the number of elements used in the *h*-adaptivity in order to maintain the error below 15% becomes less as the higher-order mesh is used. However, it appears that, considering the number of degrees of freedom, the higher-order meshes require higher computational cost for the analysis, especially when strain localisation becomes dominant. This is

---

[‖] We refer here the response obtained by using Mesh 2 (cf. Figure 6.2) with quartic interpolation.
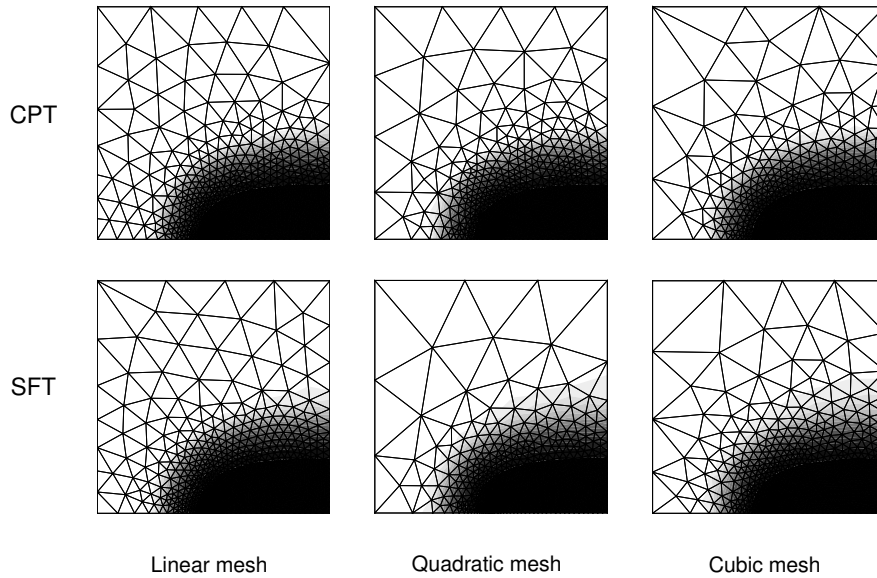
Figure 6.12 Comparison of damage profiles between different techniques of variable transfer used in *h*-adaptivity during the nonlinear computation. The plots are of the end of the loading process (Step 100, $\overline{u} = 0.001$).

understandable as we found in the preliminary tests, which also agreed with the discussion in [69], that a sufficient mesh resolution is still needed to capture the localisation zones, even though the higher-order interpolation is used.

(D) Performance of history transfer operations

Figure 6.12 shows the final damage profiles obtained after *h*-adaptivity based on linear, quadratic and cubic meshes. Comparing different transfer techniques, the damage seems to expand slightly more when applying the shape function remapping transfer (SFT) than the closest point transfer (CPT). The higher the mesh order, the more expanded the damage profile is.

We can simply explain this observation by the smearing-out feature of the SFT approach, which becomes more apparent when the element size of the variable-transfer zone is larger. Of course, by applying higher-order interpolation, the element sizes are allowed to be slightly bigger, thus the damaged zone is dispersed accordingly.

r-adaptivity

As an alternative to *h*-adaptivity, *r*-adaptivity based on a weighted Laplace smoothing algorithm is investigated. Being the simplest version of all improvement tech-
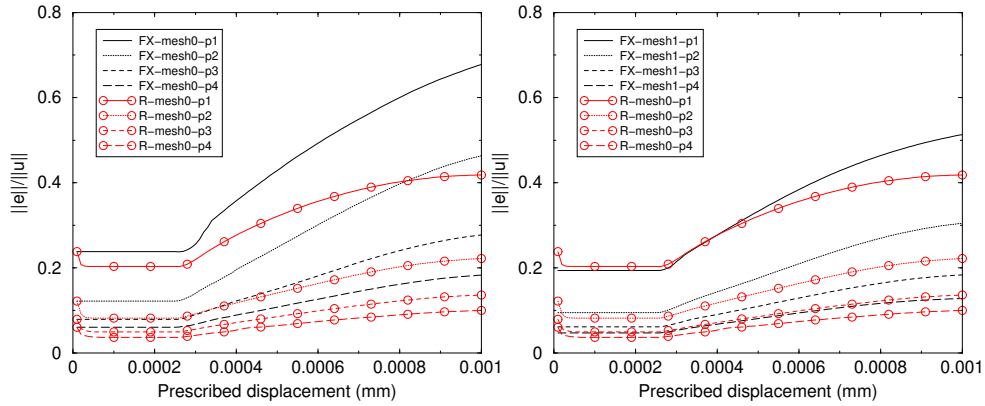
Figure 6.13 Comparison between the estimated errors obtained during *r*-adaptivity and those obtained in the fixed reference meshes.

niques, the number of degrees of freedom, element connectivities, as well as order of interpolation are fixed during the application of an *r*-adaptive scheme. We expect an improvement of what is already available by relocating the nodes to where they are needed most.

As introduced in Section 5.4, we control the speed of mesh gradation by the user-specified parameter m (cf. Eq. (5.43)). Here, the default value of $m = 1$ will be used in this analysis. The nodes on the edges $x = 1.0$ mm and $y = 1.0$ mm are allowed to move freely in the vertical and the horizontal directions, respectively. Exceptions are those nodes located at the corners and the crack tip, which must be fixed to maintain the geometry of the test.

For this test, we choose to continue the computation after each mesh movement, according to the error measure at the end of each loading step. Some interesting observations are made as follows.

(A) Efficiency of weighted-smoothing-based r-adaptivity

The improved results after applying *r*-adaptivity can be measured relatively by comparing them with the results from our reference meshes. In Figure 6.13, it is shown how the results are improved by their estimated error during the computation. There, we activate the adaptivity in the initial Mesh 0, and two reference meshes, namely Mesh 0 and Mesh 1, are for comparisons. Uniform linear interpolation is applied in this test. It is evident that, by only relocating the nodes, the mesh can be improved from the original mesh (cf. Figure 6.13 (left)), and the movement is so efficient that the mesh even becomes of better quality than our reference Mesh 1 (cf. Figure 6.13 (right)). The nodes are relocated towards the damage region, as can be seen in Figure 6.14.

$\bar{u} =$ 0.00030 mm   0.00040 mm   0.00060 mm   0.00080 mm   0.00100 mm
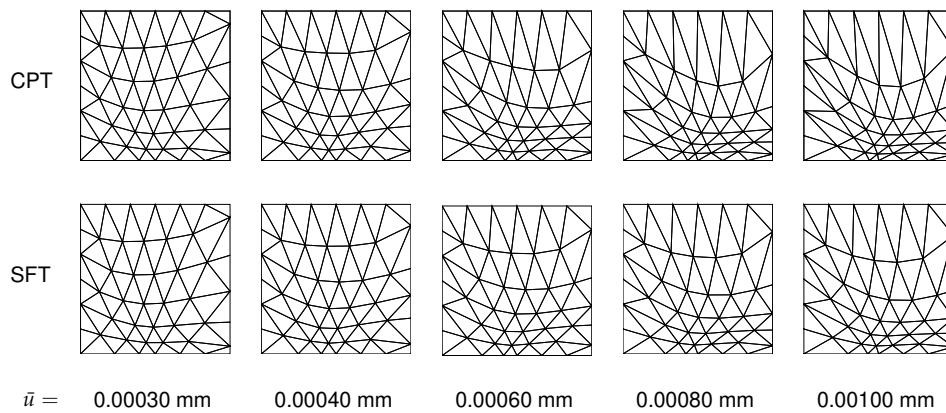
Figure 6.14 Evolution of the linear mesh during *r*-adaptivity, based on different variable transfer techniques.

$\bar{u} =$ 0.00030 mm   0.00040 mm   0.00060 mm   0.00080 mm   0.00100 mm
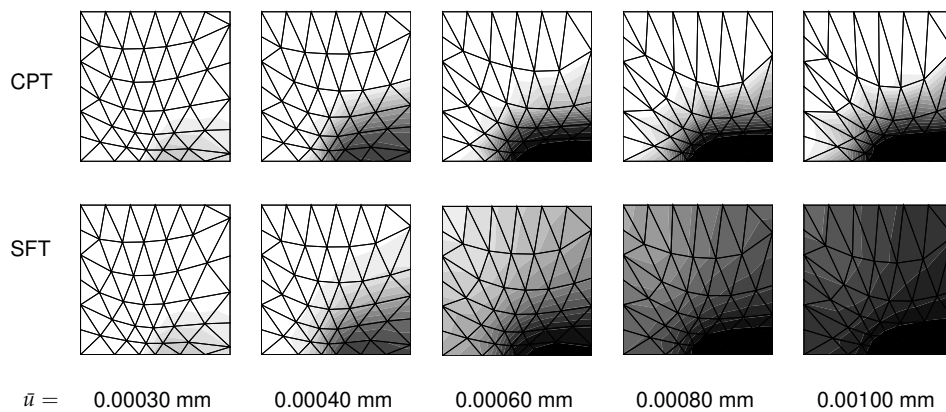
Figure 6.15 Comparison between the damage profiles obtained by using different data transfer techniques during *r*-adaptivity.

(B) Performance of history transfer operations

As the nodal points are relocated, the history variables again need to be transferred to the new discretisation in order to enable continuation of the finite element computation. The mesh movements in Figure 6.14, resulting from the CPT and the SFT approaches, are shown to be quite similar. Nevertheless, we can still notice a denser set of elements at the lower edge when applying the CPT technique than when applying the SFT technique. This is what we can expect when the SFT approach slightly spreads out the history variables for every mesh change. With continuous use of this transfer technique, the spreaded values continue spreading out and thus the strain is not as localised as it should be. This cumulative transfer error
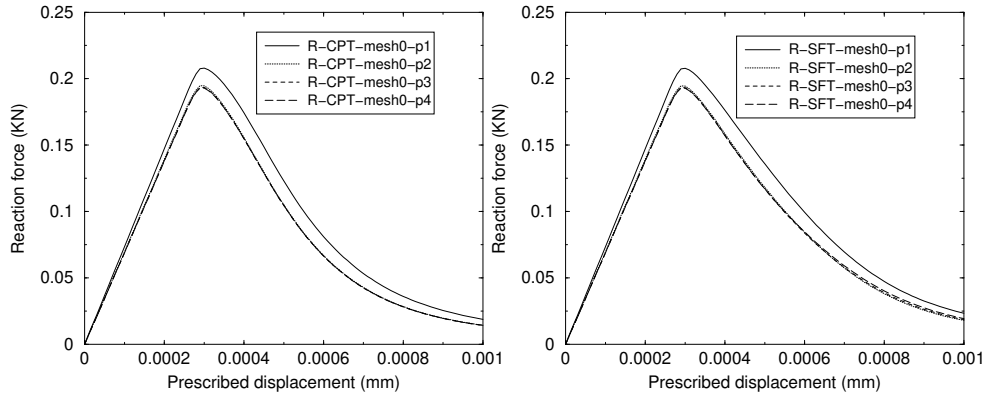
Figure 6.16 Comparison between the load-displacement diagrams obtained by applying different variable transfer techniques.

can result in a very different final damage profile shown in Figure 6.15. As an undesired consequence, the less localised solutions lead to a more ductile responses in the post peak regime. This phenomenon is observed in Figure 6.16 (right), in comparison to Figure 6.16 (left).

The choice of transfer technique affects the mesh undergoing *r*-adaptivity significantly, as the element sizes in the vicinity of the critical regions can be rather large, as compared to the ones from the use of *h*-adaptivity. Thus application of the SFT technique during *r*-adaptivity should completely be avoided.

p-adaptivity

The basic idea of *p*-adaptivity differs from other mesh improvement techniques, such as *h*-adaptivity or *r*-adaptivity. Instead of reducing element sizes, *p*-adaptivity improves finite element solutions by upgrading the finite element shape functions, such that they can better capture the finite element solution, while the mesh geometry is unaltered.

For simplicity, we allow the degrees of interpolation for this test not to exceed 4, i.e. quartic interpolation. *p*-adaptivity is activated whenever the critical error threshold is exceeded. The interpolation function is upgraded to one higher polynomial order at a time and no degrading of the mesh is allowed. Applying *p*-adaptivity on three different base meshes leads to the following remarks.

(A) Efficiency of p-adaptivity

We investigate efficiency of *p*-adaptivity in terms of computational cost and accuracy. For modelling of strain localisation phenomena, upgrading the order of interpolation seems to provide a high global accuracy, as examined in the preliminary tests, while the ability to capture the localised quantities relies very much on the base mesh.
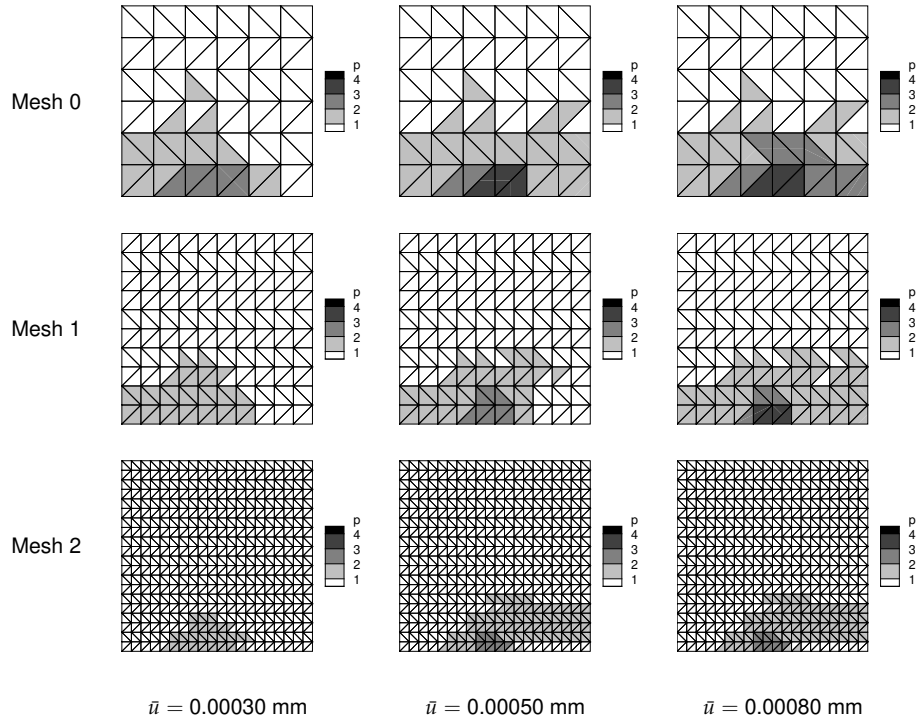
Figure 6.17 Evolution of the interpolation orders during *p*-adaptivity, based on three different base meshes.
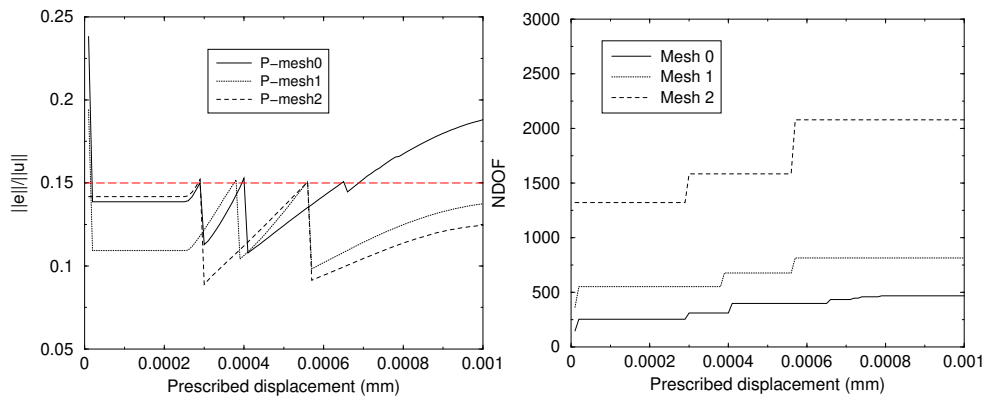
Figure 6.18 The normalised error estimates and the number of degrees of freedom used during *p*-adaptivity: comparisons between three base meshes.
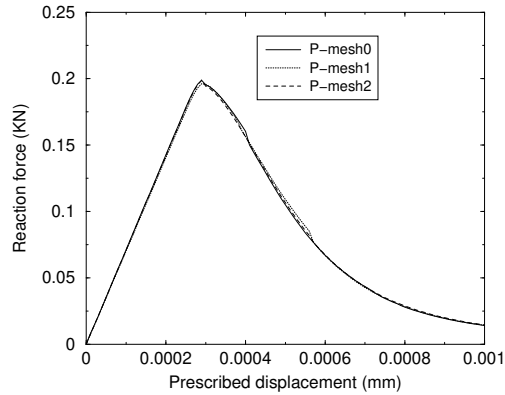
Figure 6.19 The load-displacement relations during *p*-adaptivity: a comparison between three base meshes.

Here, we apply *p*-adaptivity in the three selected base meshes. In Figure 6.17, evolutions of the local polynomial upgrade in those meshes reveal that, even in the coarsest mesh of all three (Mesh 0), the lowest order of interpolation is sufficient in most of the domain. It appears in Mesh 2 that higher-order interpolation is actually needed the most where high strain gradients exist, not where the highest damage level appears. This observation is less apparent when using *h*-adaptivity (cf. Figures 6.7 and 6.8) and visually absent when using *r*-adaptive scheme (cf. Figure 6.14).

Controlled by the 15% error allowance, Figure 6.18 (left) reveals that a higher order of interpolation than 4 is required in order to achieve the control value based on Mesh 0. In contrast, for the other two (finer) meshes the maximum quartic interpolation is adequate, or even too much in Mesh 2 where the cubic interpolation is the maximum order needed. It is unfortunate that the polynomial degree must be an integer, as upgrading one order higher may be excessive in reducing the discretisation error at a time. However, considering the number of degrees of freedom in Figure 6.18 (right), *p*-adaptivity can save a lot of computational cost comparing to that employed in the uniform higher-order meshes (cf. Table 6.1). Even though the computational cost is much reduced, the finite element approximation can still provide highly accurate global responses, as shown in Figure 6.19.

(B) Effect of the initial mesh resolution

Although a high mesh resolution is recommended for a better capturing of state variables, it is only needed in some regions in the problem domain, in particular where the strain localisation takes place. As a result, the underlying mesh for the rest of the domain (beyond those so-called *process zones*) can be constructed by elements as big as those in Mesh 0 (cf. Figure 6.17), where linear interpolation suffices. Figure 6.18 (right) actually shows an increasing trend of the number of degrees of freedom exploited in the analysis upon the increasing resolution of the underlying

mesh. Those increased number of degrees of freedom are due to an excessive mesh resolution (implying too small element sizes) used in the areas outside the process zone, rather than capturing the strain localisation phenomena inside the process zone.

In fact, the finite element analysis can benefit greatly from using p-adaptivity if the various element sizes are designed according to the needs. This, of course, refers to mixed adaptive approaches such as rp-adaptivity or hp-adaptivity, which are beyond the scope of this research.

(C) Transfer of history variables

When some parts of the mesh are upgraded, the number of integration points used in the elements in those parts generally needs changing. The higher the order of interpolation is, the more integration points are needed. For those upgraded elements, the history variables must be transferred from the old to the new sets of integration points located in the same element, so that the computation can go on without having to restart it.

To avoid complications, we employ a fixed set of integration points in each element, implying that our p-adaptivity can go up to the highest interpolation orders without any need to change the set of integration points. However, even if the transfer is needed, the direct transfer, especially the CPT approach, seems to be natural as it can be done elementwise. Considering the results from the other adaptive approaches[**], it is expected that the use of the SFT technique during the p-adaptivity will result in a slightly larger expansion of the localisation zone, as the element sizes are not reduced, than in the case of the h-adaptivity. However, the expansion will not be as pronounced as in the case of r-adaptivity where the element sizes in the vicinity of the process zones are growing.

Goal-oriented mesh adaptivity

The adaptive studies addressed earlier are based on the error measures in the solution norms. However, if one has a specific local quantity of interest, the adaptivity should be driven by the error of such a quantity instead of the error in the global solution. In this study, the performance of the goal-oriented adaptive computation is investigated via h-adaptivity as it is the easiest, among the three approaches, to visualise the adaptive level during the finite element computation.

We choose here the crack mouth opening displacement (CMOD) as our quantity of interest. A half of this quantity can be measured as the displacement in the y-direction at the lower left corner of the problem domain. Some observations can be made as follows.

(A) Analysis of influence functions

We have learned from Chapter 4 that influence functions give information on

---

[**]We refer back to the results from h-adaptivity (pp. 101–105) and r-adaptivity (pp. 105–108).

w_x w_y w_{ξ}
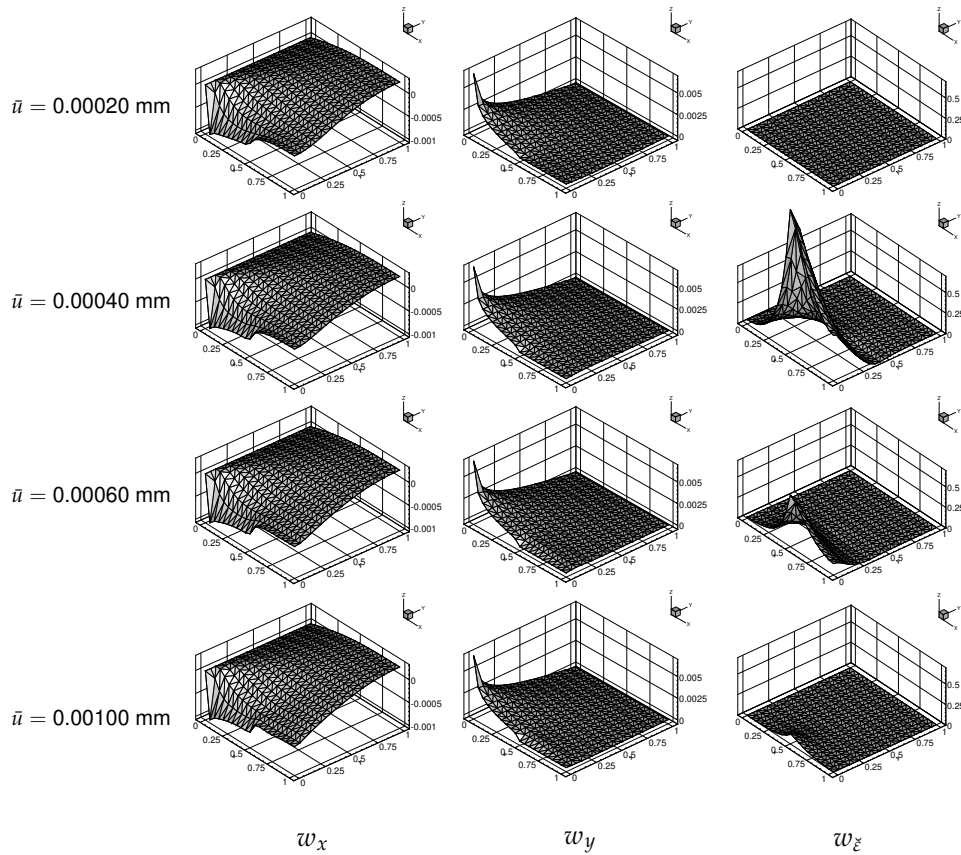
Figure 6.20 Influence functions corresponding to the crack opening displacement, obtained from Mesh 2 with quadratic interpolation.

how a unit change of *loading* at each degree of freedom in a finite element model affects a specific quantity of interest. In this context, this loading refers to the residuals resulting from the finite element discretisation. Thus, in order to measure how the residuals affect our quantity of interest, the corresponding influence functions need to be determined.

As earlier addressed in Section 4.2, the influence functions can be obtained by solving a dual problem. In nonlinear analysis, a question arises on how this dual problem should be set, since the finite element solution depends very much on the loading path. As a reasonable assumption, the dual problem is set as a stationary linear problem at any computational step. At some selected computational steps, three sets of influence functions[††] of the crack mouth opening displacement are

---

[††]This corresponds to 3 degrees of freedom used in the implicit gradient-enhanced damage formulation.
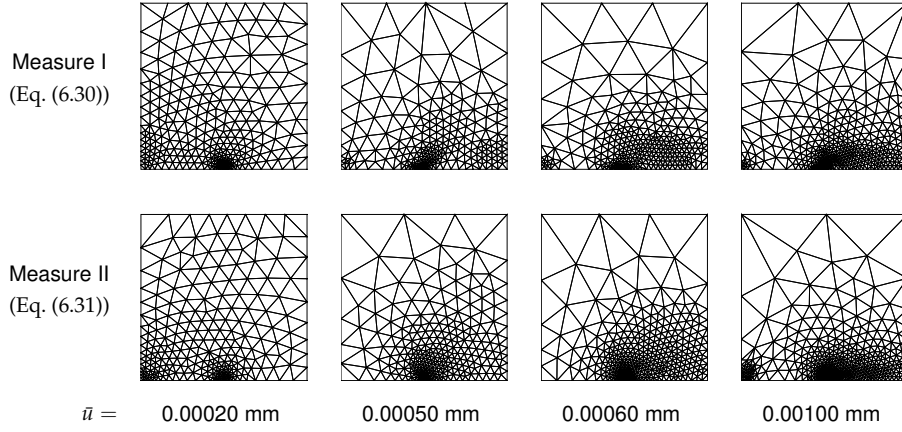
Figure 6.21 Meshes obtained during the goal-oriented *h*-adaptivity.

shown in Figure 6.20.

If we separate the discretisation residual into three parts, corresponding to each degree of freedom, i.e.

- the residual component r_x corresponds to the degree of freedom u_x
- the residual component r_y corresponds to the degree of freedom u_y
- the residual component r_{ξ} corresponds to the degrees of freedom $\check{\varepsilon}_{\text{eq}}$

It is found that a unit change of r_x and r_y consistently affects the CMOD, while a unit change of r_{ξ} varies its influence on the CMOD during the loading process. A unit change in r_{ξ} before the peak (i.e. at $\bar{u} = 0.00020$ mm) does not have any influnce on the CMOD. However, once the strain starts to localise, the influence of r_{ξ} concentrates at the crack tip (i.e. at $\bar{u} = 0.00040$ mm) and distributes in a more uniform fashion in the damaged region in the later stage (i.e. at $\bar{u} = 0.00060$ mm and $\bar{u} = 0.00100$ mm).

(B) Influence of discretisation error on the quantity of interest

After the influence functions are obtained, the error in the specific quantity (i.e. CMOD) is measured. Figure 6.21 shows the meshes used during the goal-oriented *h*-adaptivity, where two error measures

$$\text{Measure I} \qquad \mathcal{E}_k^{I} := |\mathcal{B}_k^{\text{tang}+}(\check{\mathbf{e}}, \check{\boldsymbol{\epsilon}})| \tag{6.30}$$

$$\text{Measure II} \qquad \mathcal{E}_k^{II} := \|\check{\mathbf{e}}\|_k^{\text{tang}+} \|\check{\boldsymbol{\epsilon}}\|_k^{\text{tang}+} \tag{6.31}$$

are compared. Though different, the mesh designs based on both measures also show some similarities. It is evident that, in the linear elastic regime, accuracy of the crack mouth opening displacement (as our goal quantity) depends greatly on
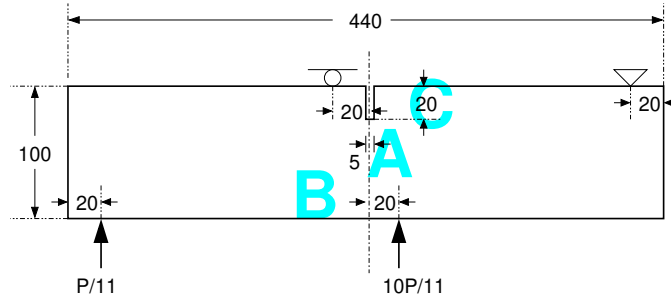
Figure 6.22 The single edge notched (SEN) beam, with dimension in millimeters.

the discretisation at the point of interest and at the crack tip where there exists a stress singularity. On the other hand, the discretisation in the process zone (damage zone) becomes more demanding in the later stage when damage emerges.

The only difference clearly seen between the two measures is that, in the softening regime, Measure I detects an outstanding amount of error at the crack mouth while the error in that region does not appear significant with Measure II. This may be explained by the mathematical definitions of the two measures: there is a *mixed* contribution[‡‡] of the error of both problems in Measure I, while Measure II separates the error of the primal problem and the error of the dual problem.

Note that the goal oriented adaptivity is controlled by threshold value ζ_{goal} (cf. Subsection 5.2.2) for both measures not to be more than 2.25%.

6.4 Single-edge-notched (SEN) beam test

Our second test is the single-edge-notched beam [84] whose geometrical details are given in Figure 6.22. The material parameters used in this analysis are: Young's modulus $E = 30000$ MPa, Poisson's ratio $\nu = 0.2$, gradient parameter $c = 0.3$ mm^2, modified von Mises equivalent strain definition (cf. Eq. (6.5)) with $k = 13.55$ and exponential softening law (cf. Eq. (6.11)) with $\kappa_0 = 0.000115, \alpha = 0.96$ and $\beta = 100$. The beam, with a specified thickness of 100 mm, is analysed under a plane stress condition.

The beam is subjected to a skew-symmetric four-point shear loading, which is applied by means of an indirect displacement control. As the control parameter, an incremental crack mouth sliding displacement of 0.001 mm is applied per computational step in the full Newton-Raphson iterative scheme.

To investigate the *h*-factor and the *p*-factor of the finite element discretisation, we select three uniform triangular meshes, namely Mesh 0, Mesh 1 and Mesh 2 (cf.

[‡‡]By the word *mixed contribution*, we mean that the errors of both problems are multiplied in a componentwise fashion.

Mesh 0

Mesh 1

Mesh 2
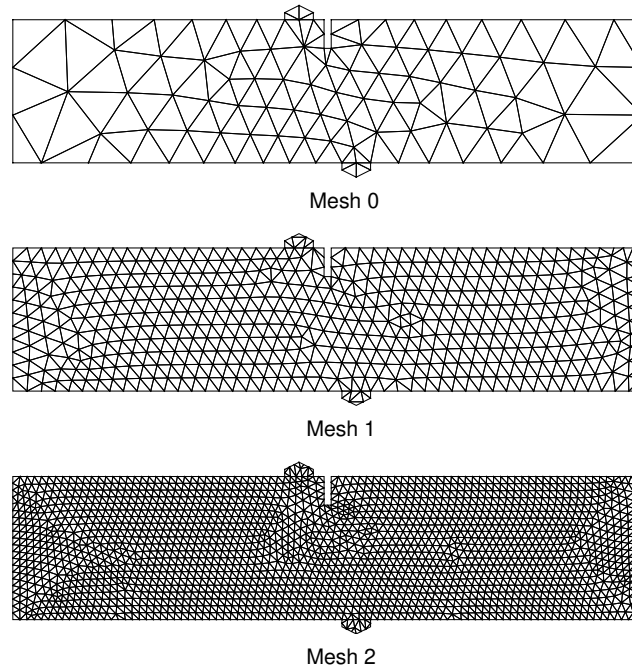
Figure 6.23 Meshes used in the finite element analysis of the SEN beam.

Figure 6.23). Again, each of the four orders of interpolation (p=1 to 4) are applied. Table 6.2 exhibits some details of these reference meshes.

6.4.1 Preliminary investigation

The reference meshes are investigated first. The tests are for examining effects of the mesh resolution (h-factor) and the interpolation degree (p-factor) on the finite element solutions of the SEN beam modelling.

The load-displacement responses obtained from those reference meshes are plotted, comparatively, in Figure 6.24. The finite element computations obviously fail when applying Mesh 0 and Mesh 1 with the interpolation order of less than 4 (quartic interpolation). It is also found that, with linear interpolation, even the finest mesh in the test (Mesh 2) leads to failure of the computation. The computational failure occurs, apparently, before the softening process starts and cannot be avoided by reducing size of the incremental load. These inadequate discretisations trigger incorrect failure mechanisms.

From Figure 6.24, some further remarks are observed at two different stages.

- In the pre-peak stage, a less stiff response and subsequently a smaller load car-

Mesh	No. of Nodes	No. of Elements	p-order	NDOFs
0	107	159	1	321
			2	1116
			3	2388
			4	4137
1	534	938	1	1602
			2	6015
			3	13242
			4	23283
2	1978	3709	1	5934
			2	22992
			3	51177
			4	90489

Table 6.2 Information of fixed meshes used in the SEN beam computation.
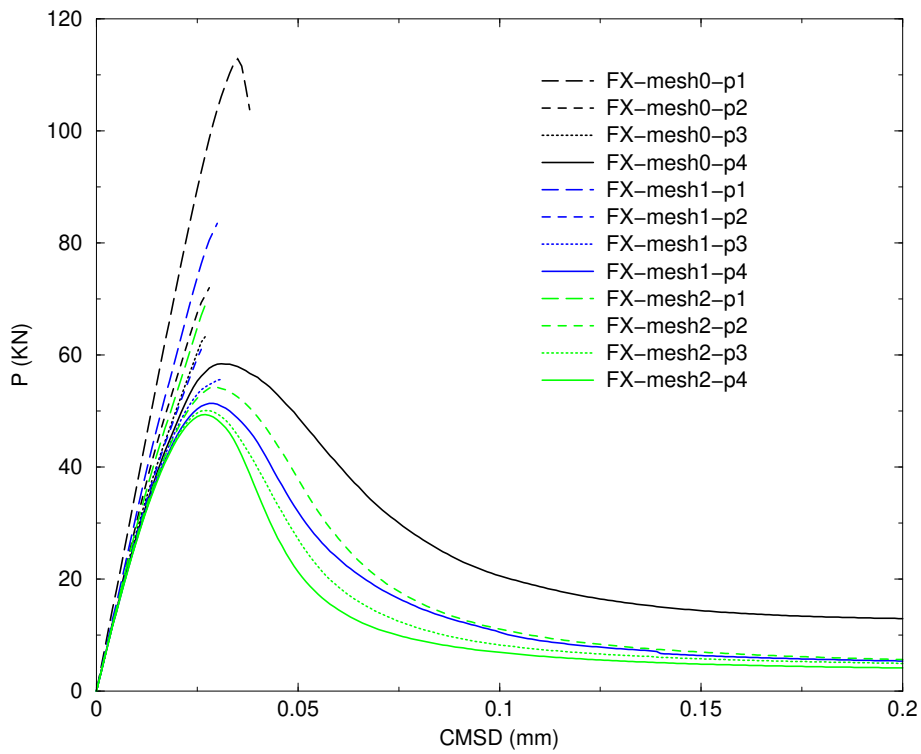
Figure 6.24 Load-displacement relations for the SEN beam.

(a) Damage

(b) Error in energy norm
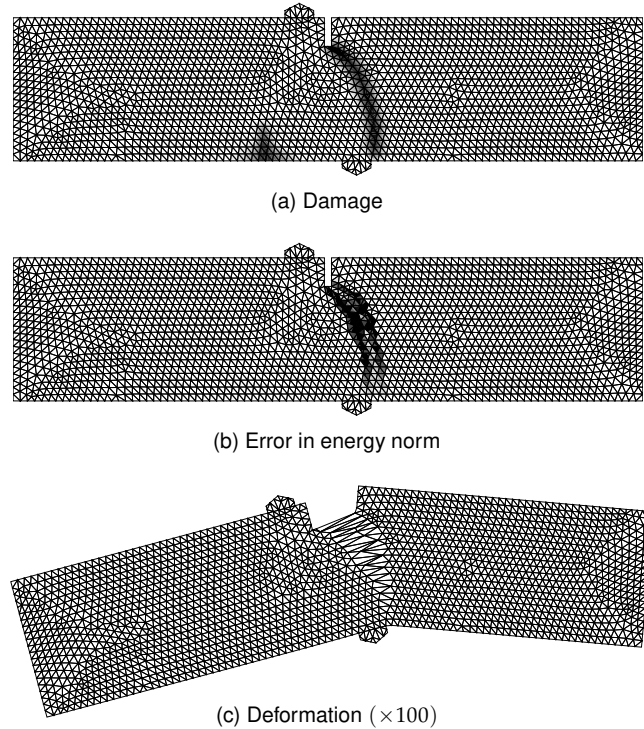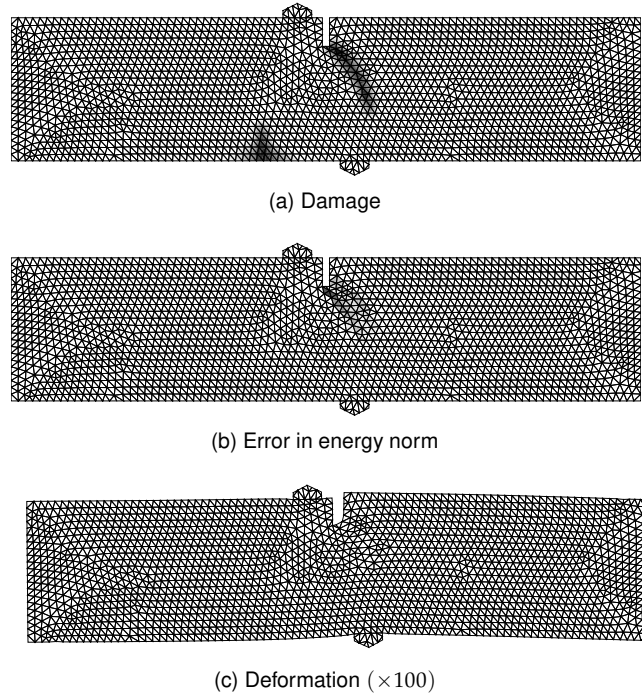

(c) Deformation ($\times 100$)

Figure 6.25 The damage profile, the error distribution and the deformation obtained by using Mesh 2 with quartic interpolation in Step 125 (CMSD = 0.125 mm).

rying capacity (ultimate load) are noticed, upon mesh refinement and mesh enrichment. Considering the same number of degrees of freedom, increasing the interpolation order (mesh enrichment) provides faster convergence than refining the mesh.

- In the post-peak stage, a more brittle softening response is obtained upon mesh refinement and mesh enrichment. In contrast to the early stage, however, reducing element sizes results in a better performance than enriching the interpolation. The result is not surprising since it agrees with what we have found in the last example and also in literature [6,69]. A sufficient mesh resolution is clearly needed when the strain is more localised.

It is expected that damage may appear at three possible zones, namely

- *Zone A*, where the stress singularity is expected at the notch

- *Zone B*, where the maximum bending stress is expected

(a) Damage

(b) Error in energy norm
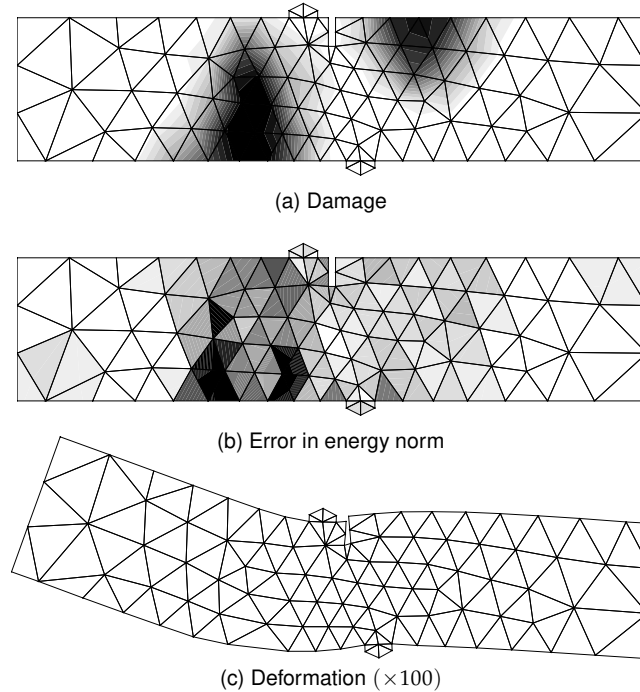
(c) Deformation ($\times 100$)

Figure 6.26 The damage profile, the error distribution and the deformation obtained by using Mesh 0 with linear interpolation in Step 38 (CMSD=0.038 mm).

- *Zone C*, where a high bending stress is expected

as marked in Figure 6.22. With the material parameters set in this test, we find that damage at the central zone (Zone A) is dominant and leads to failure of the continuum, whereas damage at Zone B only grows to a limited extent. We do not find any damage in Zone C in our reference Mesh 2 with quartic interpolation. Plotted at Step 125 (CMSD = 0.125 mm), Figure 6.25 shows the damage profile, the deformation and the error that seems to concentrate on the boundary of the primary damaged zone where the high strain gradients exist.

The situation is different when the discretisation is not sufficient. Using Mesh 0 with linear interpolation, cracks at Zone B and Zone C appear to be dominant, whereas no damage is detected at Zone A. This wrong result subsequently leads to the appearance of the discretisation error at Zone B where there exist high strain gradients, as shown Figure 6.26 for step 38 (CMSD = 0.038 mm). Now, imagine if we use the results obtained from the coarse mesh at this final step to consider where adaptivity should take place. Obviously, the error leads to wrong information (see Figure 6.26(b)) and the adapted mesh is completely useless. To obtain the right

(a) Damage

(b) Error in energy norm

(c) Deformation $(\times 100)$

Figure 6.27 The damage profile, the error distribution and the deformation obtained by using Mesh 2 with quartic interpolation in Step 38 (CMSD = 0.038 mm).

information about adaptivity, it is clear that the mesh should be updated during the loading process. See Figure 6.27 for a comparison with the reference discretisation (Mesh 2 with quadratic interpolation) at the same step.

6.4.2 Mesh adaptive tests

Similar to the central transverse crack test, we allow 15% of the error measure to be present in the analysis. However, since the test is complicated due to very high gradients in the solution, in order to limit the computational effort, we will put some limits in the adaptive level. These limits will be mentioned later in this subsection.

The discretisation error is estimated at the end of some selected computation steps. All three adaptive approaches are investigated. Filtered by the adaptive criteria (cf. Section 5.2), some regions in the finite element mesh may be improved via *h*-adaptivity, *r*-adaptivity or *p*-adaptivity. Comparisons between the three methods, in terms of performance, will be discussed in this subsection.
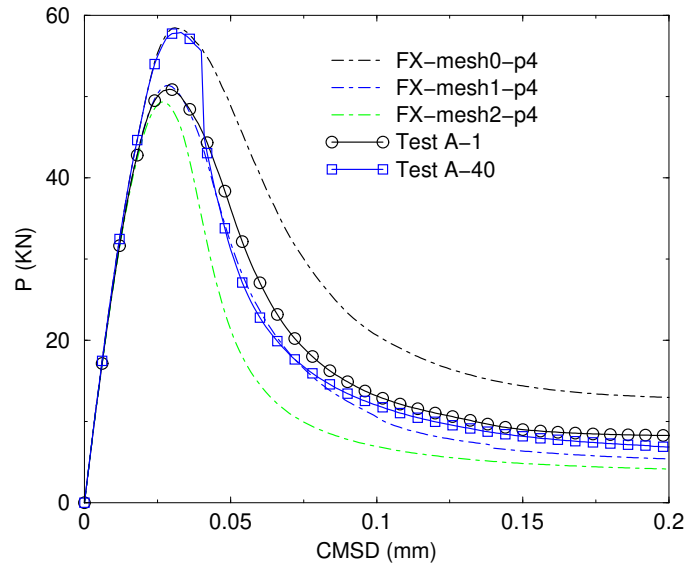
Figure 6.28 Relations between load and crack mouth sliding displacement for the tests based on one activation of *h*-adaptivity, in comparison to those based on reference fixed meshes.

h-adaptivity

(A) Effect of adaptive timing on FE solutions

The preliminary tests in the last subsection show one obvious disadvantage of using error estimates at the end of the computation in driving mesh adaptivity. Certainly, the adaptive process activated during the FE computation gives the information that is more relevant to the updated situation at that step of the computation. However, we would like to focus on how adaptivity helps in improving the solution at different stages during the computation.

To investigate this, we activate, only once during the computation, the adaptive process and see how it affects the FE solution. The starting mesh is chosen to be Mesh 0 with quartic interpolation, and two tests based on *h*-adaptivity are put under investigation. That is,

- Test A-1: the adaptivity is activated only once in the linear elastic regime (Step 1, CMSD = 0.001 mm)

- Test A-40: the adaptivity is activated only once in the softening regime (Step 40, CMSD = 0.040 mm)

If adaptivity is activated at Step 1 only (Test A-1) and the computation is continued further, we get a load-CMSD response that is very similar to the one using Mesh 1, although the response in the softening part is more ductile (cf. Figure 6.28). The FE solution has improved greatly after only one adaptive process at the beginning of the computation. The modified mesh and the damage at CMSD=0.100 mm
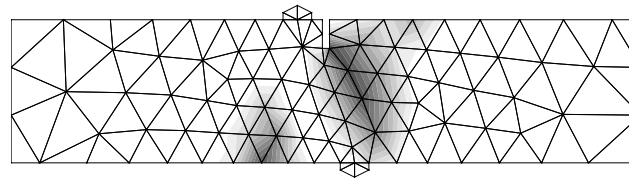
(a) Mesh 0 (Original Mesh)
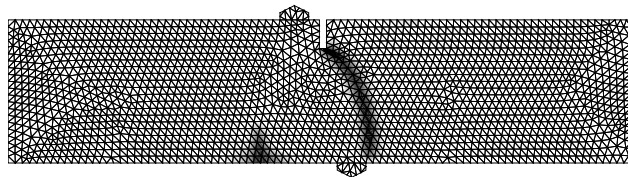
(b) Mesh 2 (Reference Mesh)
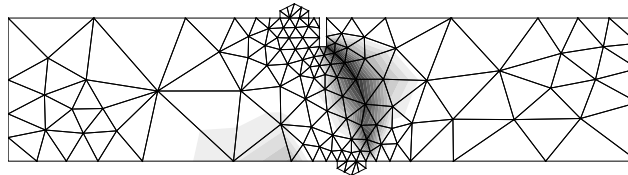
(c) *h*-adaptivity at CMSD = 0.001 mm only (Test A-1)
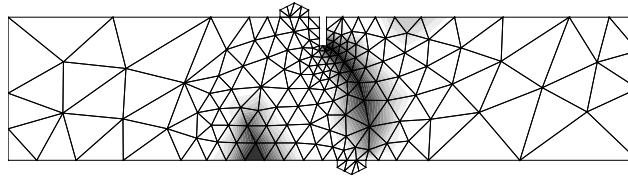
(d) *h*-adaptivity (+CPT) at CMSD = 0.040 mm only (Test A-40)

Figure 6.29 The damage profiles at CMSD = 0.100 mm for the tests based on one activation of *h*-adaptivity, in comparison to those based on reference fixed meshes.

in Figure 6.29 (c) shows needs of using smaller elements at the loading points, especially at the inner-pair loading points, and around the notch, whereas relatively big elements (roughly, twice as big as those of the initial mesh) can be used for the rest of the domain.

Of course, the adaptive process activated at Step 40 (Test A-40) requires the variable transfer algorithm and is thus more complicated. Specifically, we have chosen the closest point transfer technique for this test. Figure 6.28 reveals that after the remeshing, the response jumps from the original path to a new path, which is

CMSD = 0.011 mm

CMSD = 0.021 mm

CMSD = 0.031 mm

CMSD = 0.041 mm

CMSD = 0.051 mm

CMSD = 0.061 mm

CMSD = 0.071 mm
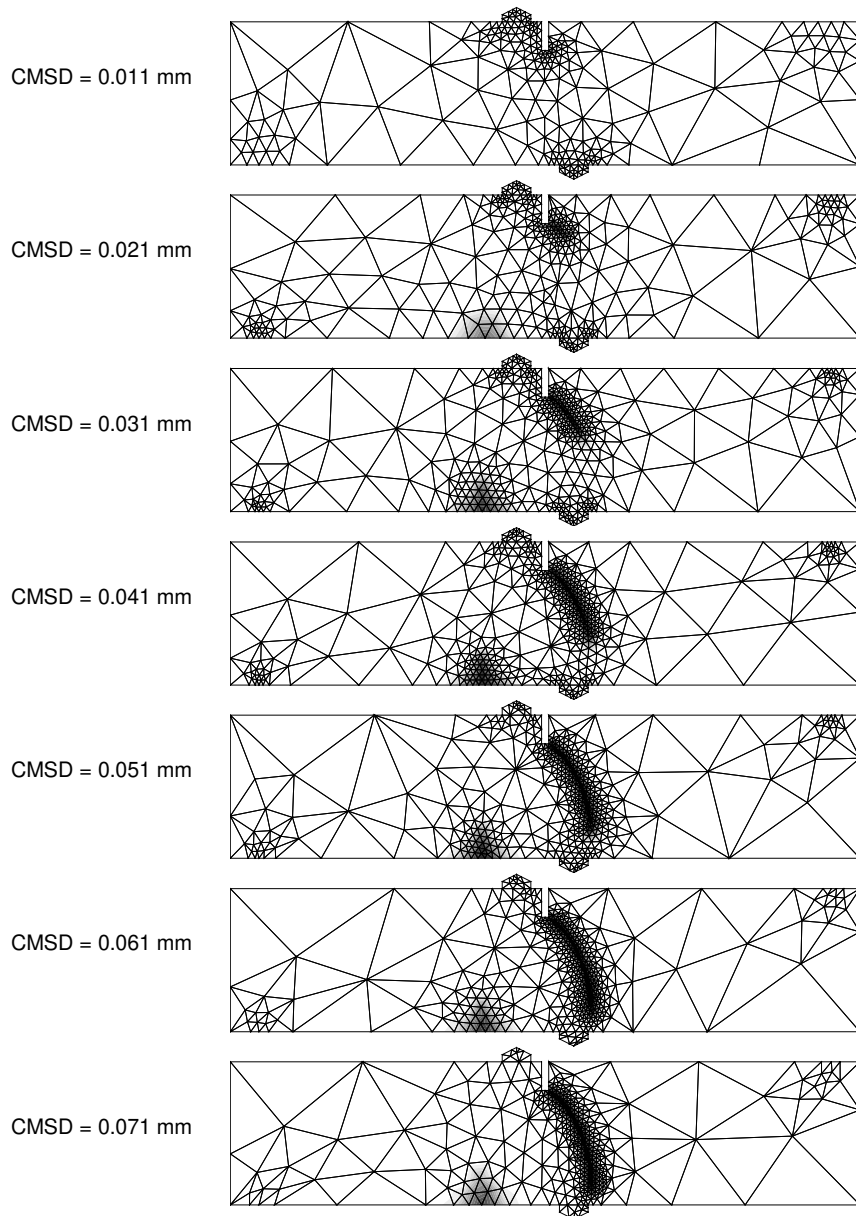
Figure 6.30 Damage evolution and corresponding h-adaptivity in the SEN beam.
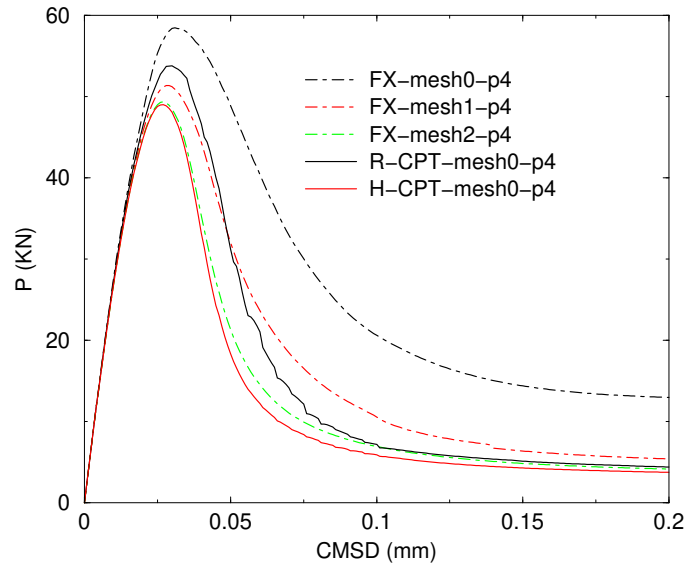
Figure 6.31 Load-CMSD relations for the SEN test with *h*-adaptivity.

closer to the result obtained by the reference mesh. Apparently, using the error information at Step 40, a more intense mesh discretisation is desired in the central part where the damage takes place compared to the outer-pair loading supports (cf. Figure 6.29 (d)).

This test reveals the importance of using mesh adaptivity in both the pre-peak and post-peak regimes. For the best result, it should be activated a sufficient number of times during the computation to make sure that the cracking process follows a correct path with a realistic width of the process zone.

(B) Analysis of mesh adaptivity

Due to the complexity of the SEN beam problem, the FE analysis requires more computational cost than it does in the previous example. Instead of computing error and activating adaptivity at every step of computation, we activate the processes every 10 steps, in order to get reasonable updates of the mesh balanced with the demands on computational efforts. With the total CMSD subdivided in 200 increments, this implies 20 possible adaptive processes during the entire analysis.

We start the *h*-adaptivity with Mesh 0 with quartic interpolation. We prevent excessive refinement by specifying the possible element size not to be smaller than the minimum value, which in this test is half of the element size of Mesh 2. Some remarks can be drawn from the mesh evolution and the corresponding damage profiles in Figure 6.30. The error that appears at the loading points in the linear elastic regime tends to reduce in the later stages (especially in the softening regime). The error at the zones of localised strains, rather than that at the loading points,
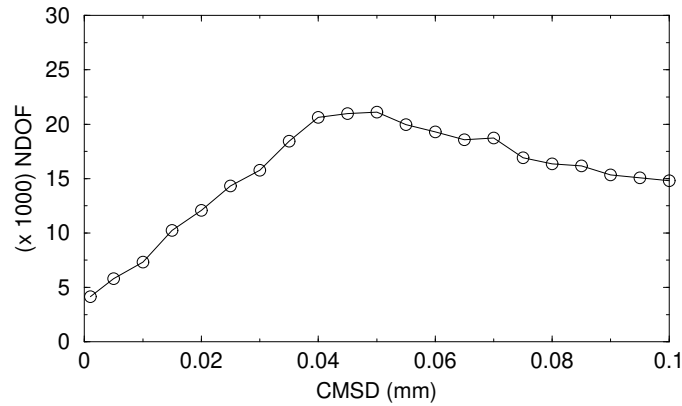
Figure 6.32 Number of degrees of freedom used in the SEN beam with *h*-adaptivity.

simply takes more control of the adaptive process.

Compared with the results from the reference meshes, it is shown in Figure 6.31 that the result from the *h*-adapted mesh almost duplicates the result of Mesh 2. Actually, when smaller element sizes than those in Mesh 2 are applied at the critical zones, a more softened behavior is obtained in the adapted mesh. Considering the computational cost, the number of degrees of freedom used in the adapted case varies during the loading process. Figure 6.32 reveals that the maximum number of degrees of freedom used in the computation is 21099, while in Mesh 2 without any adaptivity, this number is more than four times as much (cf. Table 6.2).

The decrease in number of the degrees of freedom, in Figure 6.32, is partly due to our restriction of element sizes. As the element size cannot be smaller than the minimum specification, the number of degrees of freedom at the process region is restricted. In the meanwhile, the element size requirement at the loading point regions is more relaxed during crack propagation and larger element sizes are applied, which reduces the total number of degrees of freedom.

r-adaptivity

Although improvement via *h*-adaptivity is proven to be very efficient as the mesh can be optimally designed according to the *a priori* estimate, it is always interesting to see how the FE solutions are improved by means of *r*-adaptivity. Without extra costs, the mesh gradation can help in adjusting the mesh size to capture the problem area to enable the optimal usage of the available resource of finite elements.

We allow the internal nodes to move freely, whereas the boundary nodes can move along their corresponding edges. An exception is made for those at the corners, which need to be fixed in order to maintain the geometrical descriptions. Here, the previous test based on Mesh 0 with quartic interpolation is re-investigated via *r*-adaptivity. A default value of the grading parameter, i.e. $m = 1$ (cf. Eq. (5.43)), is
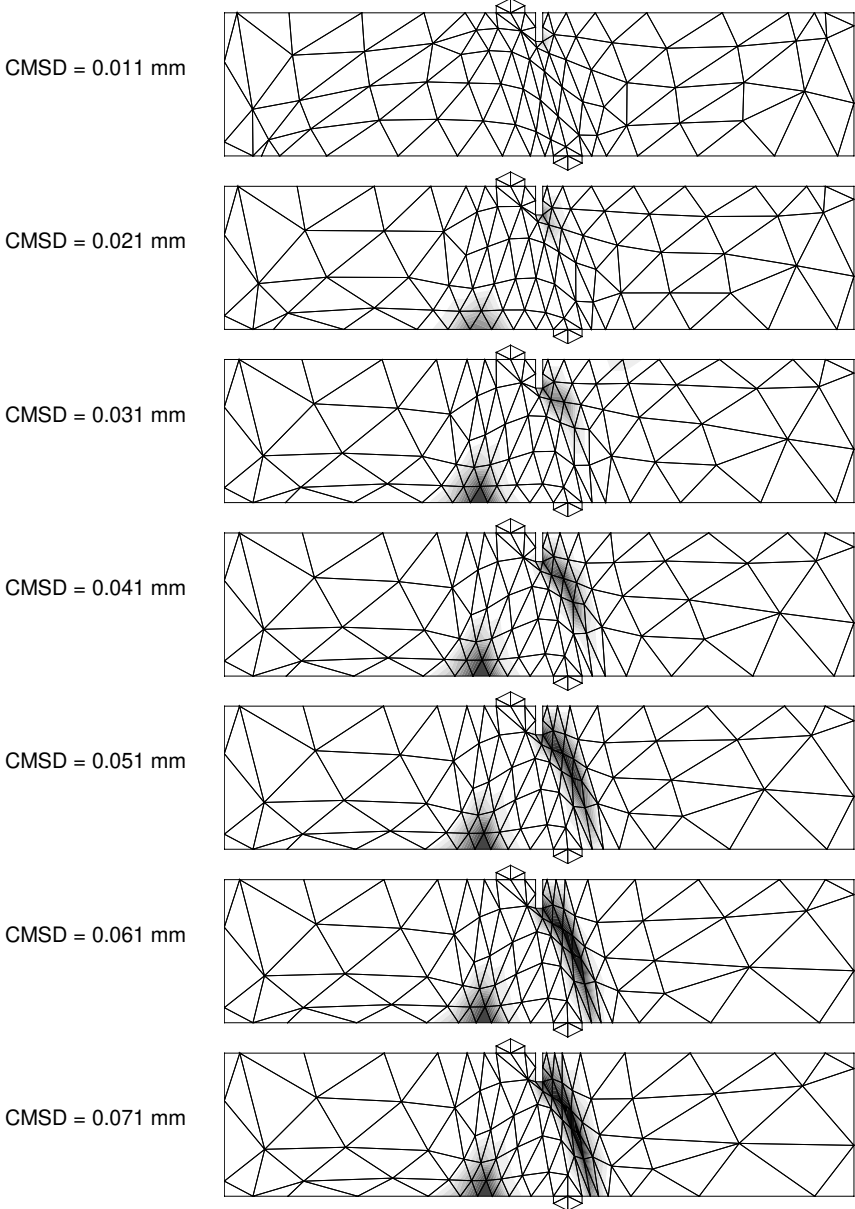
CMSD = 0.011 mm

CMSD = 0.021 mm

CMSD = 0.031 mm

CMSD = 0.041 mm

CMSD = 0.051 mm

CMSD = 0.061 mm

CMSD = 0.071 mm

Figure 6.33 Damage evolution and corresponding *r*-adaptivity in the SEN beam.
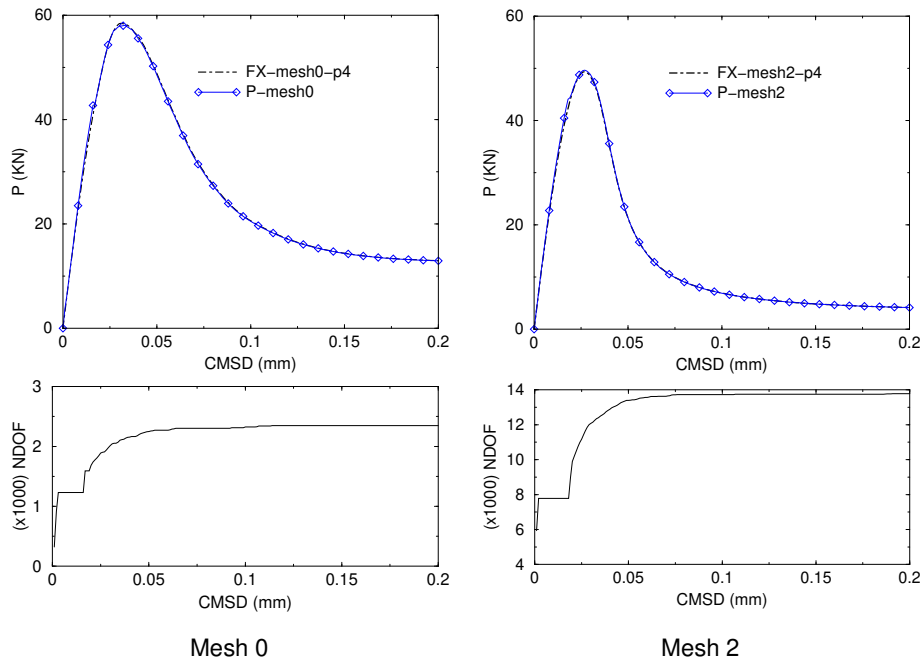
Figure 6.34 Load-CMSD relations and the corresponding number of degrees of freedom used in the SEN beam with *p*-adaptivity.

employed.

The load-CMSD response, plotted in Figure 6.31 in comparison to those of the reference meshes and the *h*-adapted mesh, reveals a very high level of improvement, especially in the softening regime. Despite such great improvement, we do not get the pre-peak response (and correspondingly the peak loading) as accurate as the one using Mesh 1.

Figure 6.33 shows how the mesh evolves during the computation. The adaptivity is apparently able to capture damage profiles that look similar to the results obtained by using *h*-adaptivity, although not as accurate. However, we can notice that the approach lacks the capability to allocate sufficient mesh resolution at the loading points and at the notch tip, possibly due to the mesh restriction and/or the insufficient mesh movement. This observation explains why a more accurate pre-peak response is not obtained. To overcome this, one may consider using a higher value of the *m* parameter to enhance the gradation step. Certainly, higher resolution of the initial mesh should be considered. Otherwise, other means of improvement should be used instead to provide a better result in the pre-peak stage.

p-adaptivity

To complete our study on mesh adaptive modelling of the SEN beam, *p*-adaptivity is also investigated. In this study, the interpolation function is upgraded to one order higher at a time, whenever needed. Degrading is prohibited and quartic order of the interpolation is set as the maximum allowance in our *p*-adaptive study.

As mentioned earlier, the basic idea of mesh enrichment is that, instead of decreasing mesh size, providing a less demanding work via the lower-quality interpolation, quality of the interpolation itself has to be improved. Nevertheless, with no change in mesh configuration, the capability of the *p*-adaptive scheme in the FE modelling seems to be limited to the maximum order of polynomial interpolation used. With up to quartic order, the FE computation is expected to give results that are of comparable accuracy as those using the same mesh with uniform quartic interpolation.

We examine *p*-adaptivity based on Mesh 0 and Mesh 2, in order to find effects of the base mesh on the FE solution. In Figure 6.34, it is found that the load-CMSD response obtained by using Mesh 0 with varying order of interpolation (*p*-adaptivity) resembles the one obtained by using the same mesh with the fixed uniform quartic interpolation. The same situation is also observed in case of Mesh 2. By this test, the remark mentioned in the last paragraph is once more proven.

Despite the fact that *p*-adaptivity provides a comparable result to the one based on the fixed higher-order mesh, costs required during the computation for both cases are quite different. The number of degrees of freedom, which reflects the computational cost, can be greatly reduced via *p*-adaptivity. It is shown in Figure 6.34 that, considering Mesh 0, the number of degrees of freedom that is necessary during the computation varies from 321 to 2349, in order to get results similar to those of the fixed mesh with 4137 degrees of freedom. A greater saving is observed in Mesh 2, where the same figure shows only 5934 to 13779 degrees of freedom needed in the computation, instead of 90489 degrees of freedom needed when employing the uniform quartic interpolation.

Figure 6.35 and Figure 6.36 show how the elemental order of the interpolation evolves during the computation. It is evident that intense mesh enrichment is needed at the loading points and where damage takes place. Comparing the results from the two base meshes, the *p*-extension is able to capture the damage function better when the resolution of the base mesh is sufficiently high. However, outside the critical regions, the mesh resolution can be more relaxed as the *p*-adaptivity alone should be able to take care the process without any problem.

6.5 Remarks

In this chapter, the basic knowledge described in Chapter 2 to 5 has been applied to the gradient-enhanced damage continuum. The error measures, which are set based on the positive-definite part of stiffness matrix, include those in a solution
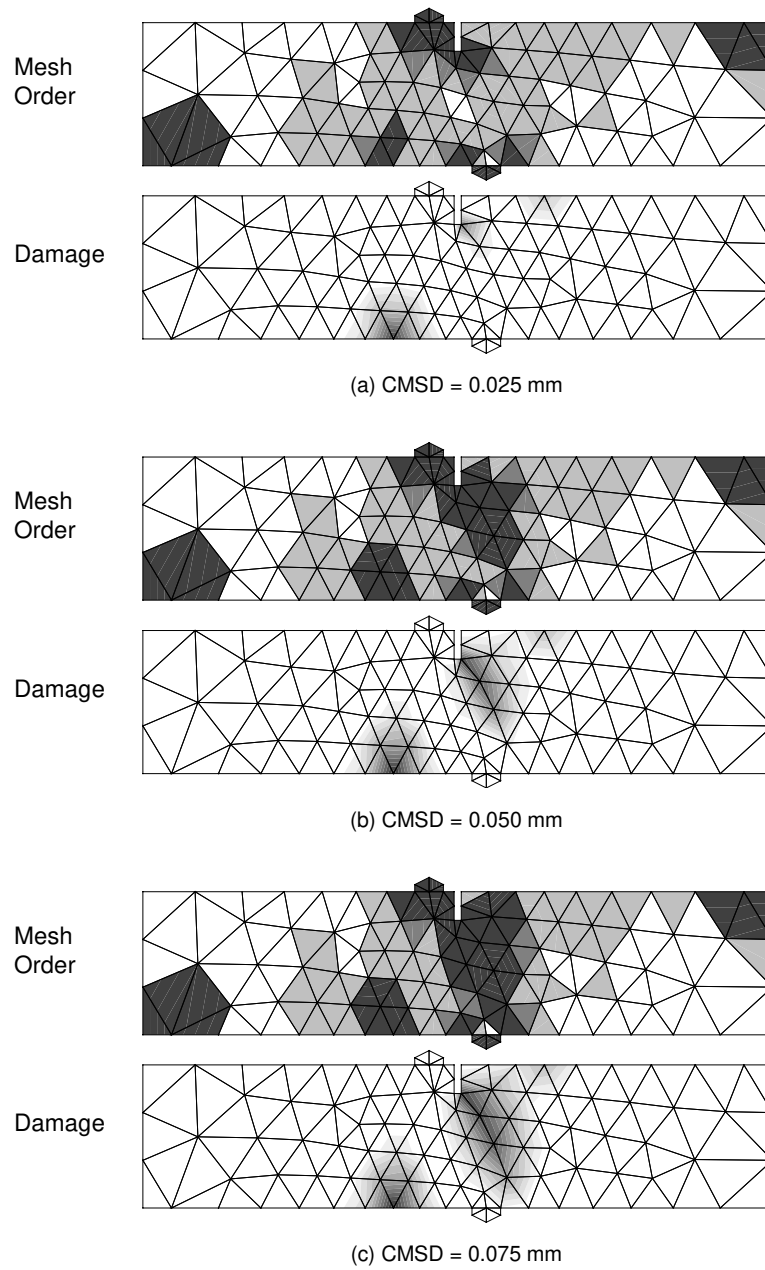
(a) CMSD = 0.025 mm

(b) CMSD = 0.050 mm

(c) CMSD = 0.075 mm

Figure 6.35 Damage evolution and the corresponding p-adaptivity in the SEN beam, based on Mesh 0.

(a) CMSD = 0.025 mm

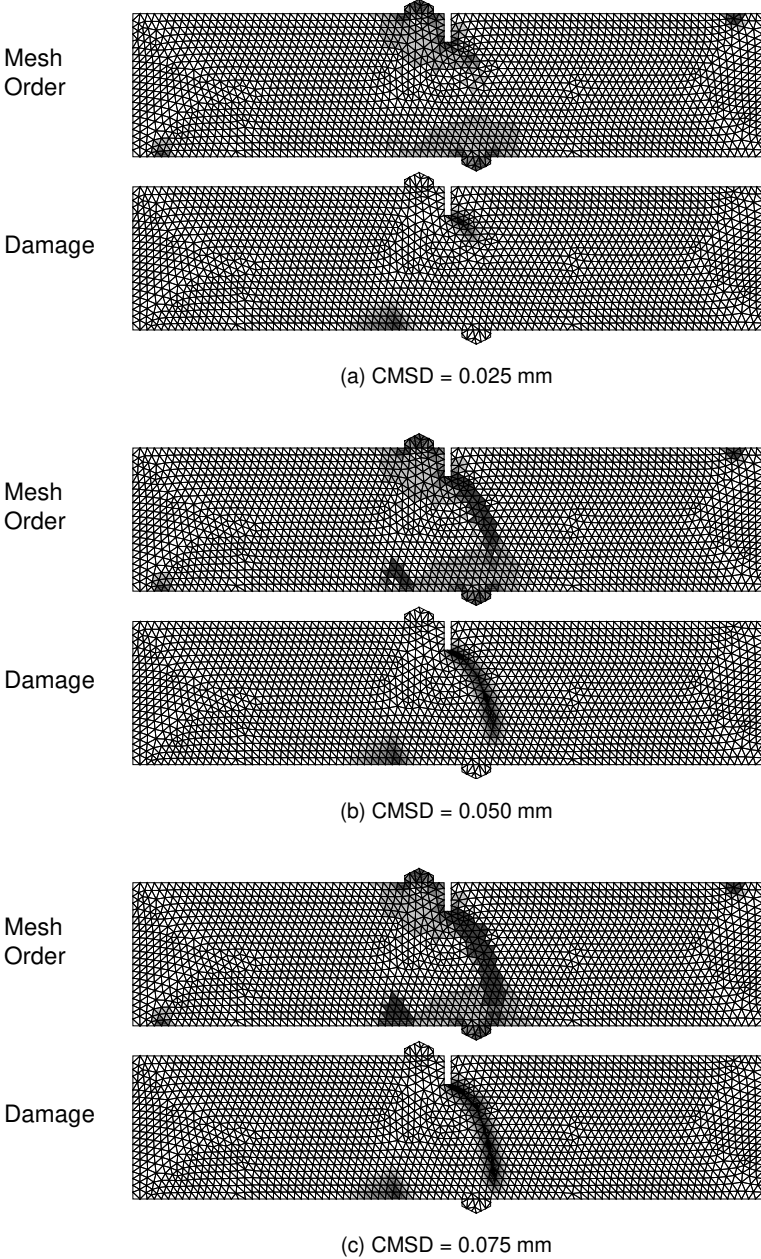(b) CMSD = 0.050 mm

(c) CMSD = 0.075 mm

Figure 6.36 Damage evolution and corresponding *p*-adaptivity in the SEN beam, based on Mesh 2.

norm and in a specific quantity, i.e. the crack mouth opening displacement. To seek the most suitable mesh discretisation for the model, the h-factor and the p-factor have been studied. Moreover, the mesh adaptive strategies, including h-adaptivity, p-adaptivity and r-adaptivity, have been examined for comparisons.

In the study, we have found that the discretisation error in linear elastic regime appears to concentrate at the supports and the notches (where there exists singularity). Once the damage is initiated, the error in the damage zone appears to take more control. During the loading process, a sufficiently refined or enriched discretisation is obviously needed in the zone of high strain gradient, which is usually the zone where damage appears. However, containing the highest damage level does not mean that the zone needs the most refined/enriched discretisation. As a good alternative to the error estimation, an error indication, after the damage is initiated, should be the gradient of the equivalent strain or of its nonlocal counterpart, rather than the damage level. Via goal-oriented adaptivity based on crack mouth opening displacement, we have also found in the example that, once the damage appears, correct modelling of the damage and the strain profiles also guarantees accuracy of the specific quantity.

The study of history transfer techniques, namely shape function remapping transfer (SFT) and the closest point transfer (CPT), have agreed well with the literature [69]. That is, the transfer with low diffusion properties (i.e. the CPT) is capable of retaining the width of the damage/localisation zone. However, if the number of integration points is small, the transfer may also create an expansion of the damage/localisation zone or a less accurate history transfer, especially at the boundary of the history variable profiles. On the other hand, although the SFT technique has less capability of retaining the damage/localisation zone width, it seems to provide fairly accurate transfer in case of h-adaptivity. For r-adaptivity, we strongly recommend not to use the SFT technique, as its expansion effect can be much more induced.

We have studied two examples in this chapter. In the central transverse crack test, it has been shown that the various adaptive strategies and error measures are all capable of attaining user-prescribed error levels. Generally, however, the adapted mesh strongly depends on whether the global error or the error in quantity of interest is used.

Based on the results of the SEN beam example, we conclude that the best adaptive approach for the damage modelling should be p-adaptivity in the pre-peak regime as it can provide the fastest convergence of all methods, especially when the mesh is well-graded toward points of supports and re-entrant corners. However, p-adaptivity alone is not very efficient in capturing the strain localisation of the post-peak regime, resulting in a too ductile response. Thus, h-adaptivity and r-adaptivity may be more preferred, although the best choice would be the combination of the p-adaptivity with h-adaptivity or r-adaptivity. Nevertheless, a survey of the combined approaches is, unfortunately, beyond the scope of this work.

Mesh adaptivity for discontinuous failure

Whereas the previous chapter dealt with the *continuous modelling of failure*, in this chapter the *discontinuous modelling* will be investigated. The basic idea behind this approach is opposite to the continuum concept where the displacement field and the strain field remain continuous throughout the entire cracking process. Instead, displacement/strain discontinuities are introduced in the problem domain to represent cracks.

In modelling of failure, two main discontinuity concepts have been proposed. As a classical approach, the *linear elastic fracture mechanics* (LEFM) model [37,45,76,77] considers sharp cracks in elastic bodies. The fracture process is assumed to occur at the tip of these cracks, which focuses in small regions, whereas the rest of the material domain remains elastic. The energy dissipation is governed by the energy release rate computed from the stress field around the crack, which is singular at the tip of the crack. Normally, the resulting failure process is brittle. For the modelling of quasi-brittle failure, a more suitable model is based on the *cohesive zone concept* [17,34,41]. To simulate nonlinear material behaviour in the crack tip region, inelastic deformations ahead of a pre-existent crack tip are modelled as cohesive tractions transferred from a fictitious face to the other, representing a so-called *cohesive crack*. In the region of the newly-formed crack tip (the cohesive crack tip), the stresses are bounded.

The cohesive zone model employed in this chapter is based on the partition of unity (PU) principle. Using the concept proposed in [21, 58, 100], a displacement discontinuity (also known as strong discontinuity) can pass through a finite element without the necessity to remesh. Two sets of unknowns, i.e. the regular and the enhanced degrees of freedom, are introduced to model the discontinuity. The concept that the enhanced degrees of freedom are gradually added in the finite element computation during crack extensions can be regarded as an *adaptive feature* in the PU-based discontinuity model. However, despite this feature, how the finite element mesh is discretised is still crucial, as the criteria for extending a disconti-
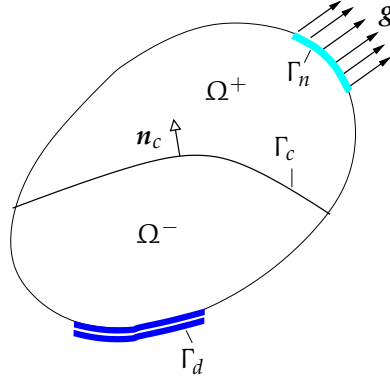
Figure 7.1 Definition of subdomains and boundaries of a body crossed by a discontinuity.

nuity relies upon a correct computation of stresses. Some discretisation aspects will be presented and discussed in this chapter.

7.1 PU-based cohesive zone model

In Chapter 2, we have presented the concept of the hierarchical interpolation enhancement via partitions of unity. Applying a similar concept to the modelling of the discontinuity, the displacement field reads

$$\boldsymbol{u} = \boldsymbol{u}^{\text{reg}} + \underbrace{\mathcal{H}_{\Gamma_c}\boldsymbol{u}^{\text{enh}}}_{\boldsymbol{u}^{\text{dis}}} \tag{7.1}$$

where the two continuous functions $\boldsymbol{u}^{\text{reg}}$ and $\boldsymbol{u}^{\text{enh}}$ represent the regular part and the enhanced part of the complete displacement field $\boldsymbol{u}$. Note that the second right-hand-side term in Eq. (7.1) contributes a discontinuous field in the standard displacement field and is herein referred to as $\boldsymbol{u}^{\text{dis}}$. The Heaviside function is defined, depending on the location of a point $\mathbf{x}$ with respect to the discontinuity (crack) boundary Γ_c (cf. Figure 7.1), as

$$\mathcal{H}_{\Gamma_c} = \begin{cases} 1 & \text{where} \quad \boldsymbol{x} \in \Omega^+ \\ 0 & \text{where} \quad \boldsymbol{x} \in \Omega^- \end{cases} \tag{7.2}$$

At the discontinuity, the displacement jump is defined as

$$[\![\boldsymbol{u}]\!] = \boldsymbol{u}^{\text{enh}}|_{\Gamma_c} \tag{7.3}$$

The kinematics at a discontinuity is based on a local coordinate system (n, s) for a two-dimensional problem, where n denotes the component in the normal direction

to the discontinuity and s denotes the component in the shear (tangential) direction to the discontinuity. Correspondingly, the discontinuity jump $[\![u]\!]$ consists of two local components, i.e. $[\![u]\!]_n$ and $[\![u]\!]_s$.

Since, in the cohesive zone concept, failure is controlled mainly by Mode I fracture, the normal component (involving the normal opening of the crack) dominates the softening behaviour at the discontinuity interface. We define an equivalent displacement jump, which is a scalar measure of the displacement jump, as

$$[\![u]\!]^{\mathrm{eq}} = [\![u]\!]_n \tag{7.4}$$

which will be used as a parameter in the loading function f defined as

$$f([\![u]\!]^{\mathrm{eq}}, \kappa) = [\![u]\!]^{\mathrm{eq}} - \kappa \tag{7.5}$$

where the history parameter κ is defined as the highest value of $[\![u]\!]^{\mathrm{eq}}$ ever achieved.

The normal traction at the discontinuity surface Γ_c can be defined in terms of the tensile strength f_t and the fracture energy G_f of the material [100] as

$$t_n = f_t \exp\left(-\frac{f_t}{G_f}\kappa\right) \tag{7.6}$$

The traction in the other direction (i.e. s-direction) in the local coordinate system (n, s) is defined differently. However, for simplicity, it will not be considered in this study (i.e., $t_s = 0$ here). For the complete definition of the local interface tractions, the reader is referred to [99].

Activation of enhanced degrees of freedom

Once a crack propagates through an element, a set of enhanced degrees of freedom along the crack path is activated to enable modelling of the discontinuous field.
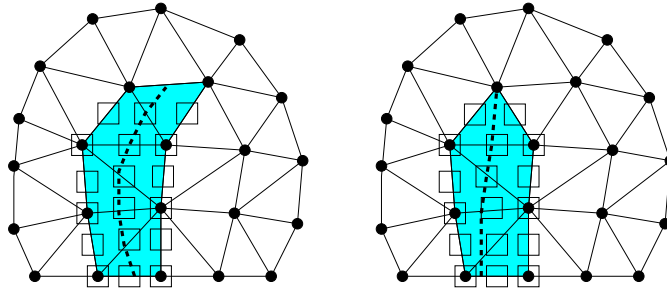
Figure 7.2 Activation of enhanced degrees of freedom in higher-order elements. On nodes and edges, the corresponding enhanced degrees of freedom are activated where the square symbols position. Also, if internal displacement modes exist, the enhanced degrees of freedom in the shaded elements are activated.

For those affected elements, the enhanced degrees of freedom at their nodes (vertex displacement modes), at their edges (edge displacement modes) and inside the elements (internal modes) are activated. The only exception is at the discontinuity tip, as it is important to fix the displacement jump at the discontinuity tip to zero. Therefore, if the discontinuity ends at an edge, the enhanced *edge* degrees of freedom will not be activated. This also applies when the discontinuity ends at a node, where the enhanced degrees of freedom of that particular node must not be activated. See Figure 7.2 for an example.

Discontinuity extension and orientation

There are some criteria to drive the extension of the discontinuity. Judging from the tensile strength of the material, a principal stress criterion [87, 99] may be considered. The criterion suggests that the discontinuity be extended when the maximum principal tensile stress at all integration points in the element ahead of the crack tip exceeds the tensile strength of the material. When it happens, the discontinuity is extended, in a straight line, through the whole element and ends at an element edge or a vertex node.

The discontinuity is extended in the direction normal to the direction of maximum principal stress. It is suggested in [87, 99] that, to prevent the incorrect direction determined from the local stress field, a non-local weighted average of stresses in a local domain (interaction domain) Ω_s may be used. The normal direction to the crack line, defined by a directional vector $\boldsymbol{d}_{\Gamma_c}$, is computed from

$$\boldsymbol{d}_{\Gamma_c} = \sum_{i=1}^{nq} \sigma_i^{(1)} V_i \, w_i \, \boldsymbol{d}_i^{(1)} \tag{7.7}$$

where 'nq' represents the number of the quadrature points (integration points) in the interaction domain. $\sigma_i^{(1)}$, V_i and w_i denote the first principal stress, the gaussian weight and the interaction weight, corresponding to the integration point i. $\boldsymbol{d}_i^{(1)}$ is the unit vector indicating the principal stress direction. The interaction weight w_i is computed from

$$w_i = \frac{1}{(2\pi)^{3/2}(r_{\max})^3} \exp\left(-\frac{r_i^2}{2(r_{\max})^2}\right) \tag{7.8}$$

where r_i is the distance of the integration point i from the discontinuity tip. Note that it is recommended to use the interaction radius $r_{\max}$ about three times the average element size ahead of the crack tip [99].

Numerical integration

When elements are crossed by a discontinuity, a special integration rule must be applied to ensure sufficient integration of the crossed elements. It is suggested in

[87,99] to subdivide each element portion to create small subdomains of a standard element shape, such as quadrilateral or triangular subdomains. Doing so allows the classical Gauss quadrature rule to be straight-forwardly applied in each subdomain. An example of this subdomain integration is illustrated in Figure 7.3.
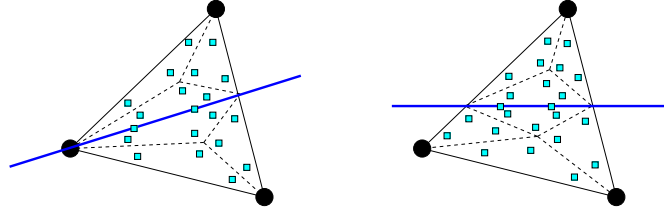

Figure 7.3 Subdivision of the element domain for numerical integration in enhanced quadratic elements.

7.2 Error analyses

As mentioned in the previous section, the PU-based cohesive crack model requires two sets of unknowns, namely the regular displacement field and the enhanced displacement field, i.e. $u = \{u^{\mathrm{reg}}, u^{\mathrm{enh}}\}$, to describe discontinuities. Both fields appear in the standard finite element computation given by

$$\mathcal{B}^{\mathrm{tang}}(\Delta u_{(h,p)}, v_{(h,p)})|_{(t-1:t)} = \Delta \mathcal{F}(v_{(h,p)})|_{(t-1:t)} \tag{7.9}$$

where u is updated during the Newton-Raphson iterative scheme, i.e.

$$u_{(h,p)}|_{(t)} = u_{(h,p)}|_{(t-1)} + \Delta u_{(h,p)}|_{(t-1:t)} \tag{7.10}$$

As both sets of unknowns contribute to model discontinuities, they must also be included in the error analysis. We seek the error solution $e = \{e^{\mathrm{reg}}, e^{\mathrm{enh}}\}$ by solving a set of error equations

$$\begin{aligned} \mathcal{B}^{\mathrm{tang}}(e_{(\tilde{h},\tilde{p})}, v)|_{(t)} &= \mathcal{R}^u(v_{(\tilde{h},\tilde{p})})|_{(t)} \\ &= \mathcal{F}(v_{(\tilde{h},\tilde{p})})|_{(t)} - \mathcal{B}(u_{(h,p)}, v_{(\tilde{h},\tilde{p})})|_{(t)} \end{aligned} \tag{7.11}$$

The error computation follows the process described in Chapter 3, whereby a series of patch-based error equations is solved instead of a global computation.

Setting a proper norm

The finite element discretisation leads to a consistent tangent stiffness matrix of the form

$$K^{\mathrm{tang}} = \mathcal{B}^{\mathrm{tang}}(\phi, \phi) = \begin{bmatrix} K_{rr} & K_{re} \\ K_{er} & K_{ee} \end{bmatrix} \tag{7.12}$$

where subscripts r and e corresponds to the regular degrees of freedom and the enhanced degrees of freedom, respectively. The components of the stiffness matrix, addressed in Eq. (6.21), are defined as

$$K_{rr} = \int_{\Omega} \boldsymbol{\nabla \phi} : \boldsymbol{D}^e : \boldsymbol{\nabla \phi} \, \mathrm{d}\Omega \tag{7.13}$$

$$K_{re} = K_{er} = \int_{\Omega^+} \boldsymbol{\nabla \phi} : \boldsymbol{D}^e : \boldsymbol{\nabla \phi} \, \mathrm{d}\Omega \tag{7.14}$$

$$K_{ee} = \underbrace{\int_{\Omega^+} \boldsymbol{\nabla \phi} : \boldsymbol{D}^e : \boldsymbol{\nabla \phi} \, \mathrm{d}\Omega}_{K_{ee}^{(1)}} + \underbrace{\int_{\Gamma_d} \boldsymbol{\phi} : \boldsymbol{T} : \boldsymbol{\phi} \, \mathrm{d}\Gamma}_{K_{ee}^{(2)}} \tag{7.15}$$

where $\boldsymbol{T}$ is defined as a matrix obtained from the relation

$$\dot{\boldsymbol{t}} = \boldsymbol{T}[\![\dot{\boldsymbol{u}}]\!] \tag{7.16}$$

$$\left\{ \begin{array}{c} \dot{t}_n \\ \dot{t}_s \end{array} \right\} = \left[\begin{array}{cc} T_{nn} & T_{ns} \\ T_{sn} & T_{ss} \end{array} \right] \left\{ \begin{array}{c} [\![\dot{u}_n]\!] \\ [\![\dot{u}_s]\!] \end{array} \right\} \tag{7.17}$$

Consider the case with no shear traction, the components T_{ns}, T_{sn} and T_{ss} are equal to zero. The only non-zero component reads

$$T_{nn} = -\frac{f_t^2}{G_f} \exp\left(-\frac{f_t}{G_f} \kappa \right) \tag{7.18}$$

It is worth noting that, without shear traction components, $\boldsymbol{T}$ is automatically symmetric. This subsequently leads to a symmetric tangent stiffness matrix $\boldsymbol{K}^{\mathrm{tang}}$.

Similar to Chapter 6, the constitive model includes softening and it is possible that computation of the error norm via use of the consistent tangent stiffness matrix $\boldsymbol{K}^{\mathrm{tang}}$ leads to a negative value, by which the norm defined in Chapter 3 becomes meaningless. Apparently, the only term that induces the negative definiteness to the stiffness matrix $\boldsymbol{K}^{\mathrm{tang}}$ is the second term in Eq. (7.15). Thus, to ensure positive definiteness, the norm to be used for this model is defined by neglecting this term. The modified stiffness matrix for the norm computation then reads

$$\boldsymbol{K}^{\mathrm{mod}} = \mathcal{B}^{\mathrm{mod}}(\boldsymbol{\phi}, \boldsymbol{\phi}) = \left[\begin{array}{cc} \boldsymbol{K}_{rr} & \boldsymbol{K}_{re} \\ \boldsymbol{K}_{er} & \boldsymbol{K}_{ee}^{(1)} \end{array} \right] \tag{7.19}$$

Note that $\boldsymbol{K}^{\mathrm{mod}}$ can be regarded as the *linear elastic stiffness matrix*. The energy norms of the solution and the error can then be written, respectively, as

$$(\|\boldsymbol{u}\|^{\mathrm{mod}})^2 \quad = \mathcal{B}^{\mathrm{mod}}(\boldsymbol{u}, \boldsymbol{u}) \quad = (\|\boldsymbol{u}\|^{\mathrm{mod}})^2 \quad = \boldsymbol{u} : \boldsymbol{K}^{\mathrm{mod}} : \boldsymbol{u} \tag{7.20}$$

$$(\|\boldsymbol{e}\|^{\mathrm{mod}})^2 \quad = \mathcal{B}^{\mathrm{mod}}(\boldsymbol{e}, \boldsymbol{e}) \quad = (\|\boldsymbol{e}\|^{\mathrm{mod}})^2 \quad = \boldsymbol{e} : \boldsymbol{K}^{\mathrm{mod}} : \boldsymbol{e} \tag{7.21}$$

These norms will be used in this chapter in setting the adaptive criteria and driving mesh adaptivity.
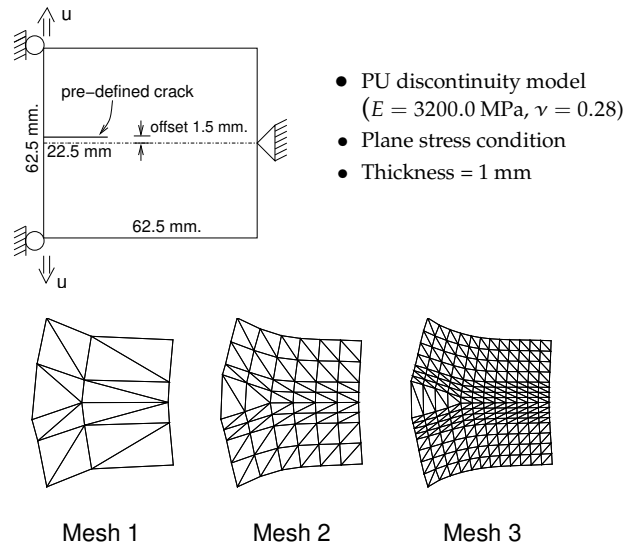
Figure 7.4 Descriptions of the patch test and the deformed configurations ($\times$ 100) of the linear meshes used.

Performance of error estimator for PU discontinuity model

The error estimation in the PU discontinuity model can be implemented straight-forwardly by the concept addressed in Chapter 3. However, since we also include the enhanced set of degrees of freedom, it is interesting to analyse how the error estimator, based on solving a series of local problems, performs.

A simple patch test is selected for this explorative investigation. Described in Figure 7.4, the patch with an eccentric pre-defined traction-free crack is pulled in the vertical direction. Three meshes are used, each of which employs different orders of interpolation ranging from linear order ($p = 1$) to quartic order ($p = 4$).

Since there is no analytical solution to this problem, we compare the estimated error to the error obtained from solving a global system of error equations. Since this concerns a linear elastic test, the stresses are singular at the crack tip and we expect its effect on the missing global components [42]. It is shown in Table 7.1 that the capability to replace a global computation by a series of local computations (measured through the index $\theta_g := \|e\|^{\text{loc}}/\|e\|^{\text{glob}}$) is decreased with an increasing number of elements and increased with an increasing mesh order. The first observation can be explained by the fact that in a fine mesh (i.e. smaller elements, larger number of elements) there are more local problems to be solved instead of a global problem and thus the interaction between distant patches is decreased. To explain the second observation, we must bear in mind the fact that the smoothness of the solution is increased when the interpolation order is higher. The smoother solution may also imply an increase of the interaction between distant patches. As a result,

Mesh	Order	DOFs	Reference	$\|e\|^{\text{loc}}$	$\|e\|^{\text{glob}}$	θ_g
1	p = 1	60	p = 2	2.40	2.49	96.4%
1	p = 2	180	p = 3	1.10	1.13	97.4%
1	p = 3	364	p = 4	0.66	0.67	98.5%
2	p = 1	252	p = 2	1.45	1.56	92.9%
2	p = 2	884	p = 3	0.69	0.73	94.5%
2	p = 3	1900	p = 4	0.51	0.53	96.2%
3	p = 1	884	p = 2	1.01	1.16	87.1%
3	p = 2	3300	p = 3	0.55	0.61	90.2%
3	p = 3	7252	p = 4	0.44	0.46	95.6%

Table 7.1 Comparison of the estimated and the reference global error obtained in the patch test.
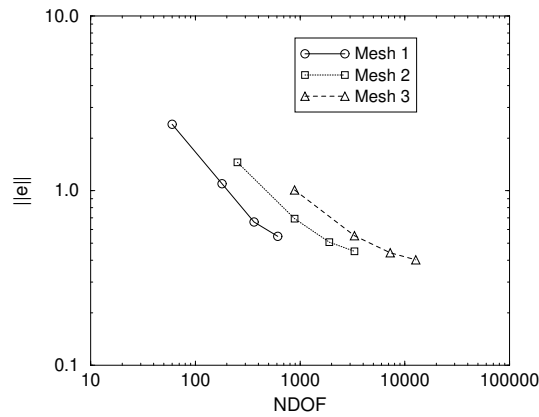
Figure 7.5 Convergence analysis of the patch test. Uniform orders of interpolation ranging from $p = 1$ to 4 are applied.

the error estimation based on solving local problems performs better for a higher-order mesh.

The error estimates provided in Table 7.1 are plotted in Figure 7.5 to study convergence of the finite element solution. Since we restrict the pre-defined crack not to propagate, this problem is merely a linear elastic analysis. As expected, the error decreases in a faster rate via *p*-extension than it does via *h*-extension.

7.3 Crossed crack test

For a propagating discontinuity based on the cohesive zone concept, the crossed crack test (cf. Figure 7.6) is selected as our first test. The problem domain is analysed under a plane stress condition. The displacement control algorithm is applied

with an incremental prescribed displacement at the top edge of $\bar{u} = 0.0001$ mm and the full Newton-Raphson iterative scheme is applied. The vertical crack causes singularities but does not propagate, hence can be modelled via appropriate boundary conditions; whereas horizontal crack does propagate and is modelled in its entirety. For these reasons, only the right half of the test is analysed.

7.3.1 Preliminary investigation

Although, via the PU concept, a crack can propagate through the mesh without the necessity to remesh, the crack extension criterion based on variables such as stresses may depend heavily on the initial mesh discretisation. To investigate the effect of the *h*-factor and the *p*-factor of the finite element discretisation, we select three uniform triangular meshes, namely Mesh 0, Mesh 1 and Mesh 2 (cf. Figure 7.7), to each of which four different orders of interpolation ($p = 1$ to 4) are applied. Basic details of these reference meshes are given in Table 7.2.

Effect of mesh discretisation on FE solutions

The PU-based discontinuity model activates the elemental sets of the enhanced degrees of freedom (i.e. u^{enh}) during crack propagation. This results in an increasing number of unknowns during the analysis. It is shown in Figure 7.8 that, upon mesh refinement (*h*-extension), the number of unknowns increases in a more gradual way. Smaller elements imply smaller distances for crack extension at a time. On the other hand, since the element size is not altered upon mesh enrichment (*p*-extension), the number of unknowns keeps progressing in the same-size steps for a fixed mesh with different mesh orders. In both extensions, we do observe earlier crack extension processes upon mesh improvement, either by *h*-extension or *p*-extension.

The load-displacement relations in Figure 7.9 reveal that the finite element computation tends to overestimate the reaction forces corresponding to the prescribed displacements. The response obtained by using Mesh 0 with linear interpolation is significantly different from the others (cf. Figure 7.9 (upper-left)). However, upon *h*-extension and/or *p*-extension, the solution is greatly improved. In terms of smooth-

- PU discontinuity model
  based on cohesive crack concept [99]
  ($E = 1.0$ MPa, $\nu = 0.20$)
- Plane stress condition
- Thickness = 1 mm
- Evolving discontinuity law
  ($f_t = 0.01$ MPa and $G_f = 0.0001$ N/mm)
- Right half modelled
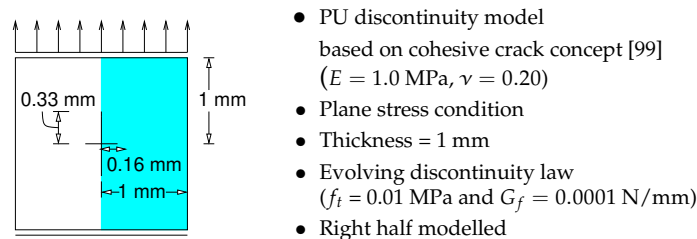
Figure 7.6 The crossed crack test.

Mesh 0   Mesh 1   Mesh 2
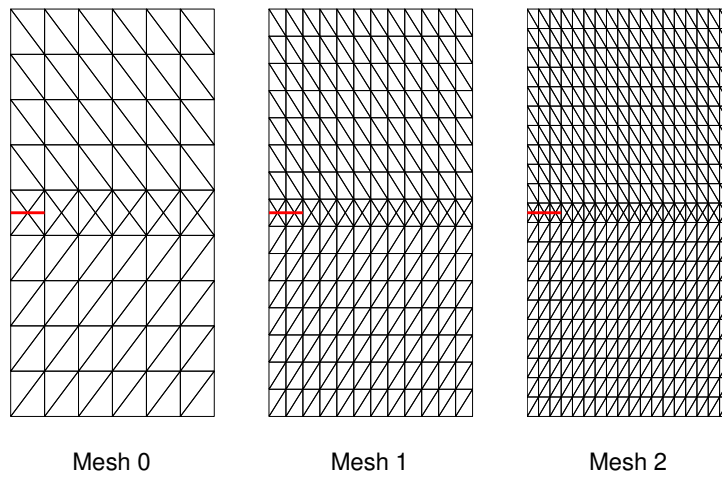
Figure 7.7 Meshes used in the finite element modelling of the crossed crack test.

Mesh	No. of Nodes	No. of Elements	p-order	No. of basic DOF
0	76	120	1	152
			2	542
			3	1172
			4	2042
1	220	384	1	440
			2	1646
			3	3620
			4	6362
2	436	792	1	872
			2	3326
			3	7364
			4	12986

Table 7.2 Information of fixed meshes used in the crossed crack test. The basic degrees of freedom includes the basic displacement modes without any contributions of the enhanced degrees of freedom, i.e. only the first and the second degrees of freedom are counted.
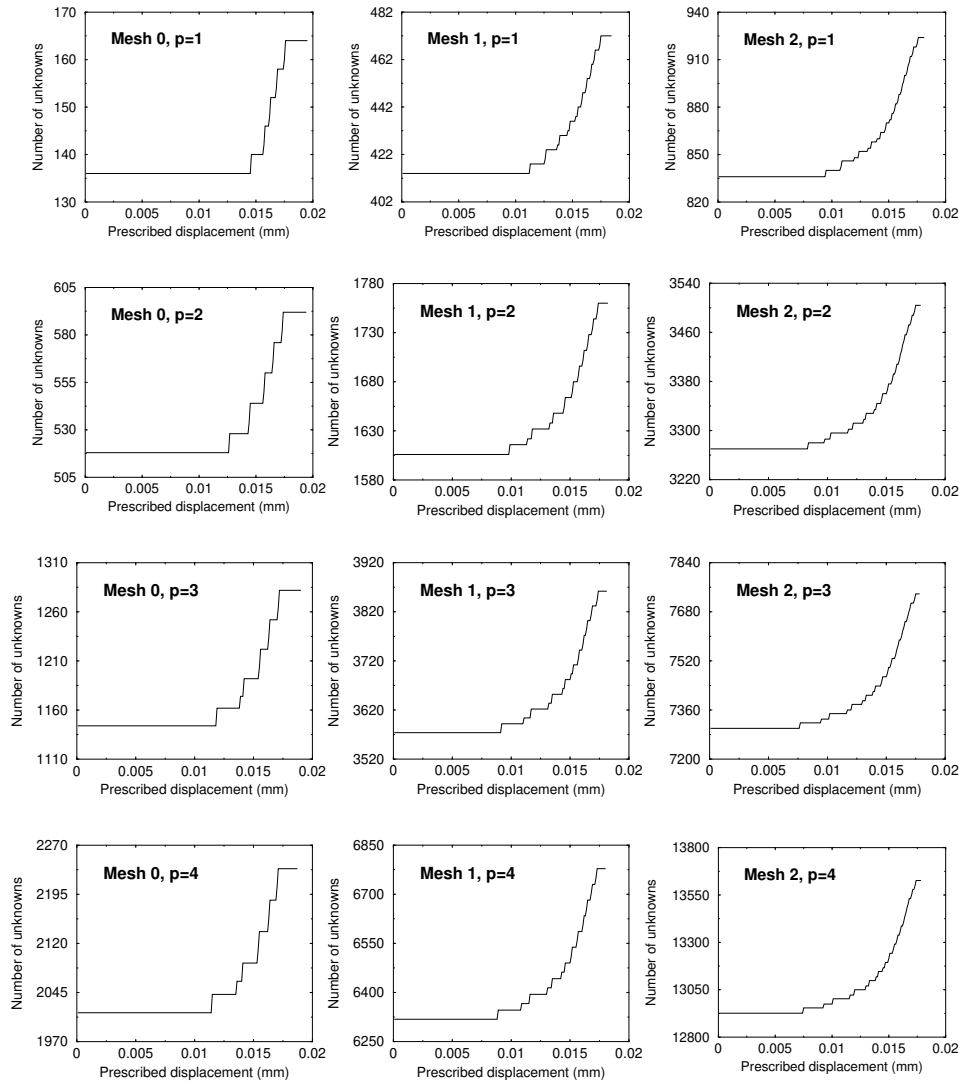
Figure 7.8 The increased number of unknowns during the PU-based discontinuity analysis in the crossed crack test. Three meshes with four uniform interpolations are investigated.
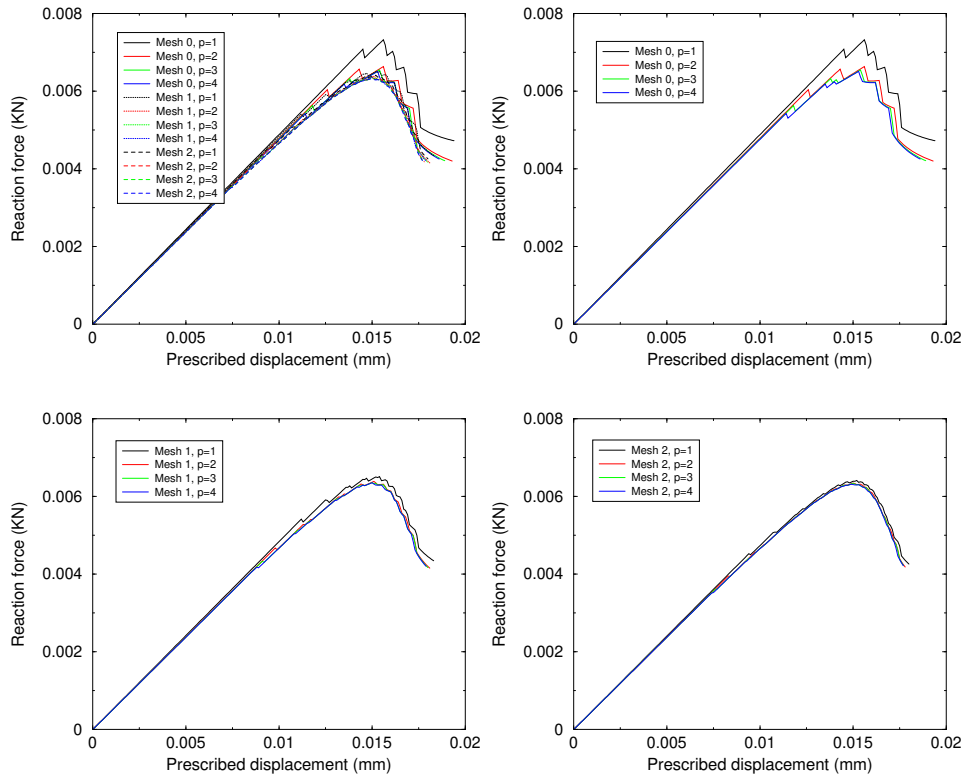
Figure 7.9 Load-displacement relations of the crossed crack test based on Mesh 0, Mesh 1 and Mesh 2. The uniform interpolation ranges from linear order to quartic order ($p = 1$ to 4).

ness of the global response, only h-extension can provide a considerable improvement, since a finer mesh allows the discontinuity to be extended in smaller steps.

Despite the above-mentioned superiority of h-extension over p-extension in the discontinuity analysis, the former approach is not suitable for adaptive analysis as its implementation can become very complicated due to the altered mesh configuration after remeshing. All information of the discontinuity must be transferred with care. In fact, if only the global responses obtained from the coarse mesh are fitted to provide smooth relations, p-adaptivity can be an excellent alternative choice in improving the finite element results without any change in mesh topology.

Analysis of error information

In the previous section, we have argued the use of the positive-definite part of the stiffness matrix (i.e. the linear elastic stiffness matrix), instead of the tangent stiff-

Mesh 0 Mesh 2
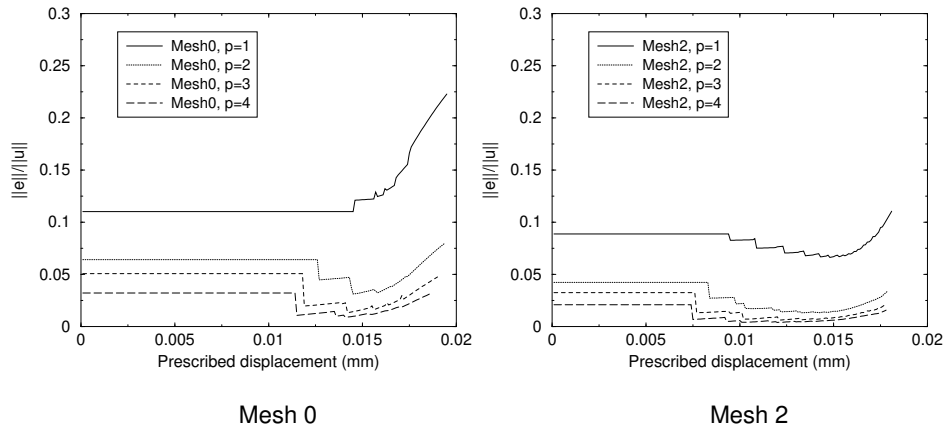
Figure 7.10 Normalised error estimated during the finite element analysis of the crossed crack test.

ness matrix, to be employed in the norm computation. This also means that the same stiffness matrix is used for all of the error computations during the finite element analysis. Based on Mesh 0 and Mesh 2, the normalised errors assessed at the end of each loading increment are shown in Figure 7.10. Upon mesh improvement, the error tends to decrease in the same fashion for all cases.

The elemental error distributions, obtained in the linear meshes, are shown in Figure 7.11. Three stages are selected; the beginning, the intermediate (i.e. when the crack is about to propagate across half the specimen), and the end. It is observed that large errors are located at the supports and at the crack tip, especially at the traction-free pre-defined crack tip. As we always expect large errors in the area where there exist high gradients of stresses (and also strains), it is interesting to see how the stress evolves during the discontinuity analysis to get a good explanation for the findings.

The evolution of the first principal stress during crack propagation is shown in Figure 7.12. It is observed that, when the crack is not yet extended, this principal stress concentrates in the region around the tip of the pre-defined crack, which is traction-free. Once the crack propagates, the stress profile becomes slightly more uniform. This is because of the cohesive zone concept by which stress singularity at the crack tip is translated to inelastic deformation during crack propagation. The evolving stress contours correspond to the error distributions shown in Figure 7.11.

7.3.2 Mesh adaptive tests

In the crossed crack test, the discretisation error is approximated at the end of each numerical step. Filtered by the adaptive criteria (cf. Section 5.2), some regions in the finite element mesh may be improved via various techniques of mesh adaptiv-

Beginning   Intermediate   End
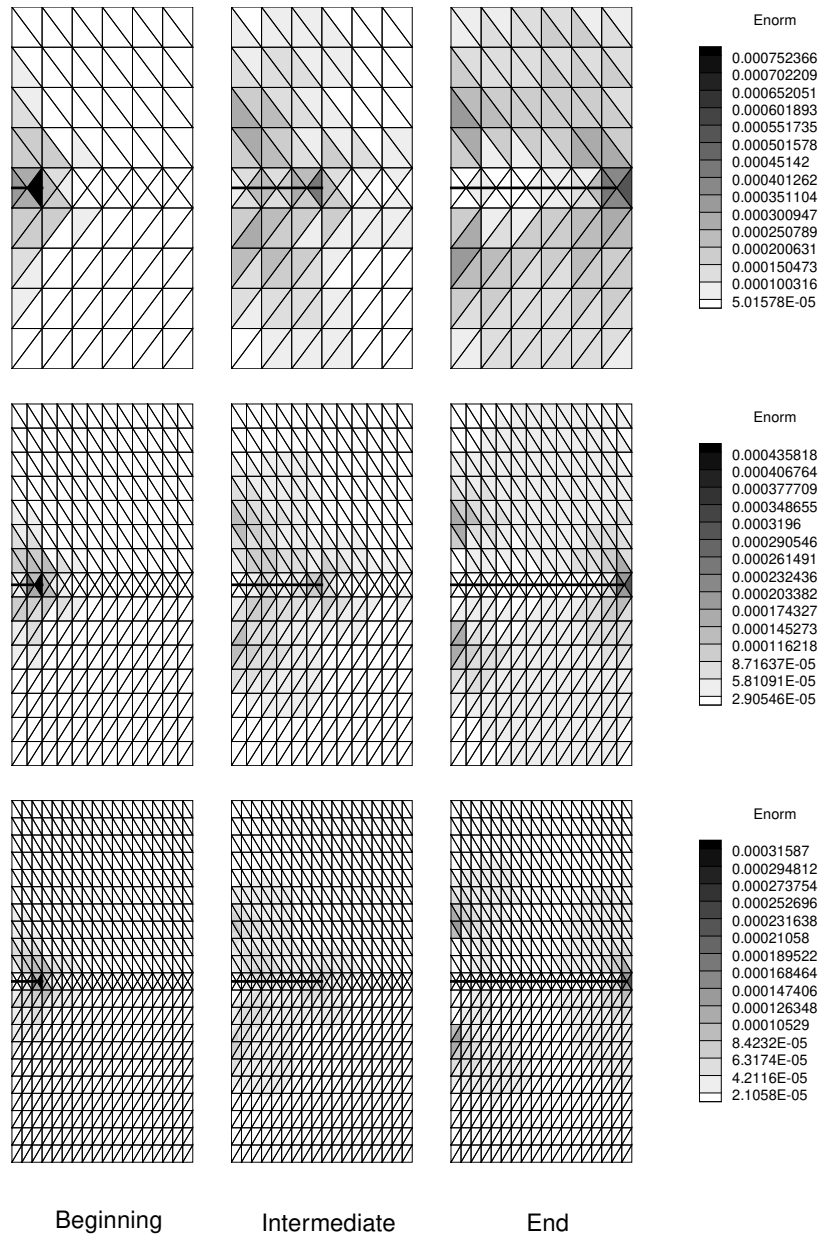
Figure 7.11 Distribution of the estimated elemental error obtained in the crossed crack test at the beginning of the computation, at the half way of the specimen and at the end of the computation. Three uniform linear meshes are investigated.
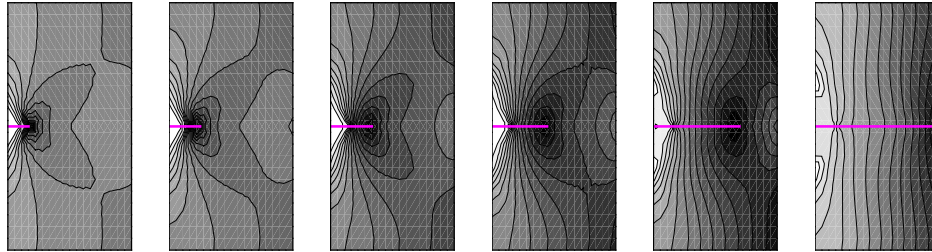
Figure 7.12 Contours of the first principal stress during crack propagation in the crossed crack test.

ity. However, there are some implementational complexities that may arise in the adaptive PU-based discontinuity model due to the change of mesh topology. First of all, detailed information of the discontinuity path that is stored for each element must be transferred. Correspondingly, the special integration scheme applied in the enhanced elements must be reset. These implementational complexities are not impossible to overcome, but they pose enough difficulties that make the adaptive scheme not worth to be carried out.

For this reason, it is best to avoid complicating mesh adaptivity by changing the mesh configuration via *h*-adaptivity or *r*-adaptivity. In addition, as we notice from the uniform mesh improvement in the last subsection, the *p*-factor can greatly improve the result, regardless of the non-smooth responses observed. The performance of *p*-adaptivity will be investigated and discussed in this subsection. As it is a simple problem, we employ a stricter control criterion by requiring that the error will not be beyond 3% of the solution norm (cf. Eq. (6.28)).

p-adaptivity

By upgrading the mesh order in a hierarchical manner, *p*-adaptivity avoids the complicated transfer of discontinuity information. Since the mesh configuration is fixed, extra degrees of freedom added to improve the finite element solution can be straightforwardly inserted in the formulation during computation. Considering the results from the preliminary investigations in Section 7.3.1, we could benefit from *p*-adaptivity in terms of accuracy by relatively low additional cost.

For simplicity, we require the degrees of interpolation not to go beyond quartic order of interpolation. The adaptive technique is activated whenever the critical error threshold is exceeded. The interpolation function is upgraded to one higher polynomial order at a time and no degrading of the mesh is allowed. Applying *p*-adaptivity on three different base meshes leads to the following remarks.
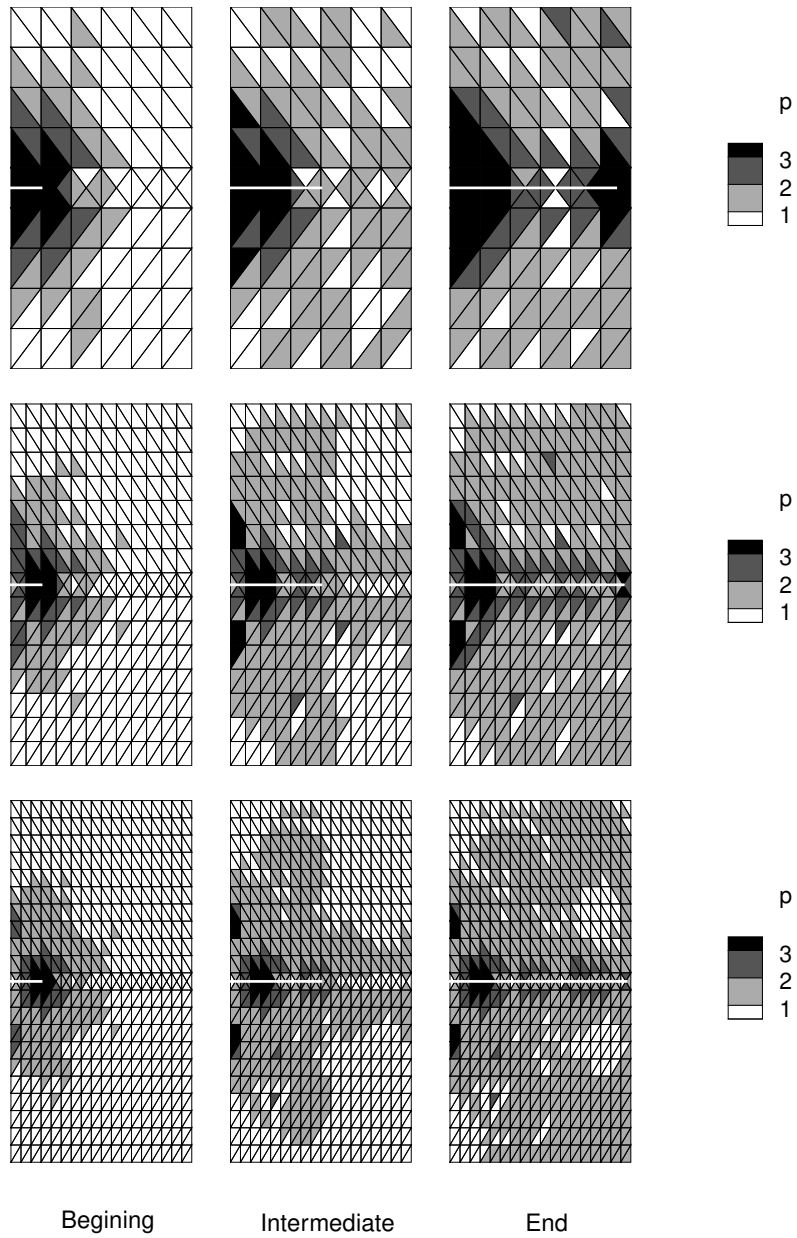
Figure 7.13 Distribution of the interpolation orders in the crossed crack test at the beginning of the computation, half-way the computation and at the end of the computation. Three uniform linear meshes are investigated.
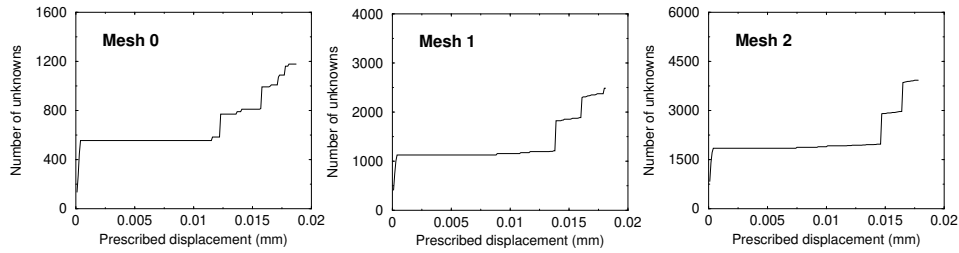
Figure 7.14 The increased number of unknowns during the *p*-adaptive PU-based discontinuity analysis.
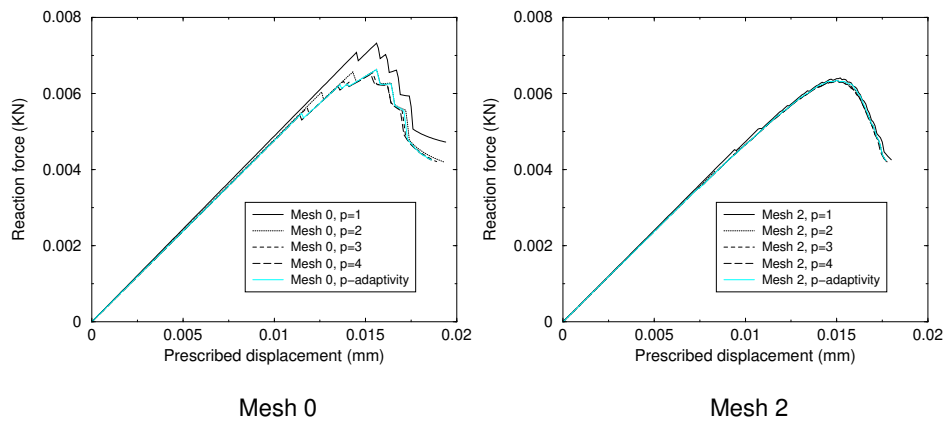

Figure 7.15 Comparison of load-displacement relations obtained in the uniform meshes and the p-adapted mesh.

(A) Efficiency of p-adaptivity

Based on the elemental error distribution shown in Figure 7.11, mesh improvement may be needed only in some small zones in the problem domain. In these zones, including the zones where the supports change type and where high stress gradients exist, higher order interpolation may be needed, while lower order interpolation should be sufficient in most of the domain.

As expected, Figure 7.13 reveals that quartic interpolation is only required in the region of the pre-defined crack tips*. In comparison with Figure 7.8, the number of unknowns used during the *p*-adaptive crack propagation is much less than the number used in the uniform quartic mesh. For instance, at the end of the analysis, the number of unknowns is reduced by a factor of 1.86 for Mesh 0, a factor of 2.71 for Mesh 1 and a factor of 3.50 for Mesh 2. Even though the computational costs are much reduced, the finite element approximation can still provide responses that are comparable to those using uniform quartic meshes, as shown in Figure 7.15.

---

*The pre-defined crack tips include the horizontal crack tip as well as the vertical crack tip which is modelled as an abrupt change in the boundary conditions.

(B) Degrading of mesh order: an investigation

Generally, when mentioning *p*-adaptivity, only upgrading of the mesh order is expected as the technique helps improving the finite element result. However, as the computation continues, it is possible that some enriched regions may change their situation. The higher-order interpolation in those regions may not be necessary anymore. In such cases, degrading of the mesh order may be an interesting technique worth considering.

From the error contour in Figure 7.11, it is observed that the error seems to concentrate at the crack tip especially at the pre-defined crack tip. The error concentration moves according to the updated crack tip, and error in the elements behind the crack tip decreases. The higher-order interpolation used in those regions may be degraded to a lower order. It is worth noting that degrading the mesh order reduces not only the number of basic degrees of freedom, but also the enhanced degrees of freedom, if any.

Since degrading is not the main subject of this study, we consider a simple degrading criterion. If the refinement ratio of any element Υ_k (cf. Chapter 5) is (far) less than the user-specified value ρ, the mesh order of that element may be degraded. That is to say, the degrading will take place whenever

$$\Upsilon_k < \rho \quad \text{where} \quad 0.0 < \rho < 1.0 \tag{7.22}$$

with a rule that the upgrading and the degrading must not take place at the same

$$\rho = 0.30 \qquad\qquad\qquad \rho = 0.10$$
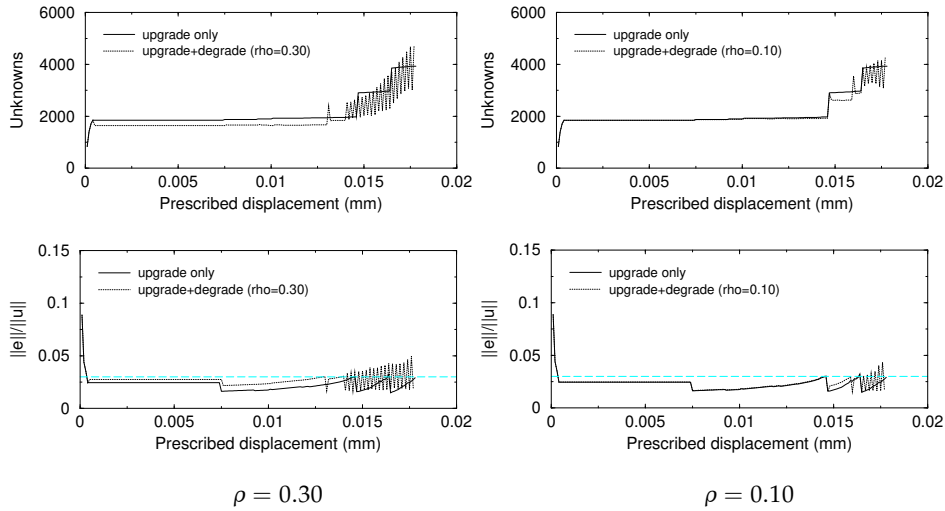
Figure 7.16 Number of unknowns and the corresponding error quantity during the loading process in the crossed crack test. Based on Mesh 2, the degrading is examined by using $\rho = 0.30$ and $\rho = 0.10$ (cf. Eq. (7.22)).

time step and only one order of interpolation must be reduced at a time. As an auxilliary cost-saving technique, the degrading should not disturb the normal mesh improvement procedure. Whenever the mesh does not need an upgrading (that is, after the mesh passes the refinement criterion), the degrading criterion judges whether quality of any part of the mesh can be reduced.

Here, we use two values of ρ for investigation: (a) $\rho = 0.30$ and (b) $\rho = 0.10$. The results are shown in Figure 7.16, where we can notice some fluctuations of the error and thus the number of unknowns during the finite element computation. The fluctuations seem to be more serious when the larger value of ρ (i.e. $\rho = 0.30$) is applied. This can be explained by the fact that a too large ρ offers the possibility of over-degrading the mesh and thus the mesh needs to be re-upgraded again in the next computational step. The process can go on with no mesh order consistency. Comparing the upgrading-degrading technique with the pure upgrading technique of the standard p-adaptivity, the upgrading-degrading technique seems to perform worse despite a more complicated implementation and thus is not recommended. Nevertheless, the degrading seems to be suitable in the linear-elastic regime, when the crack has not yet propagated.

Goal-oriented *p*-adaptivity

In this study, the criterion for crack extension and the computation of the crack orientation are based on the first principal stress in the region ahead of the crack tip. This stress quantity may be regarded as our quantity of interest. Instead of the error measures employed earlier, it may be more relevant to use the error in this local quantity of interest to drive the adaptive process. Here, performance of the goal-oriented adaptive computation is investigated via p-adaptivity during the finite element computation.

Since the crack propagates in the x-direction all the way through, the use of the first principal stress as a quantity of interest can be simply replaced by the stress component σ_{yy} to make a simpler choice[†]. We choose here the average of the stress component σ_{yy} in the element ahead of the crack tip k as our quantity of interest. That is,

$$Q(\boldsymbol{u}) = \int_{\Omega_k} \sigma_{yy}(\boldsymbol{u}(\boldsymbol{x})) \, d\Omega \tag{7.23}$$

Some observations can be made as follows.

(A) Analysis of primary unknowns and influence functions

As addressed in Chapter 4, the discretisation error affects not only the global finite element solution but also any corresponding local quantity of interest. How much the discretisation error influences the local quantity can be approximately

---

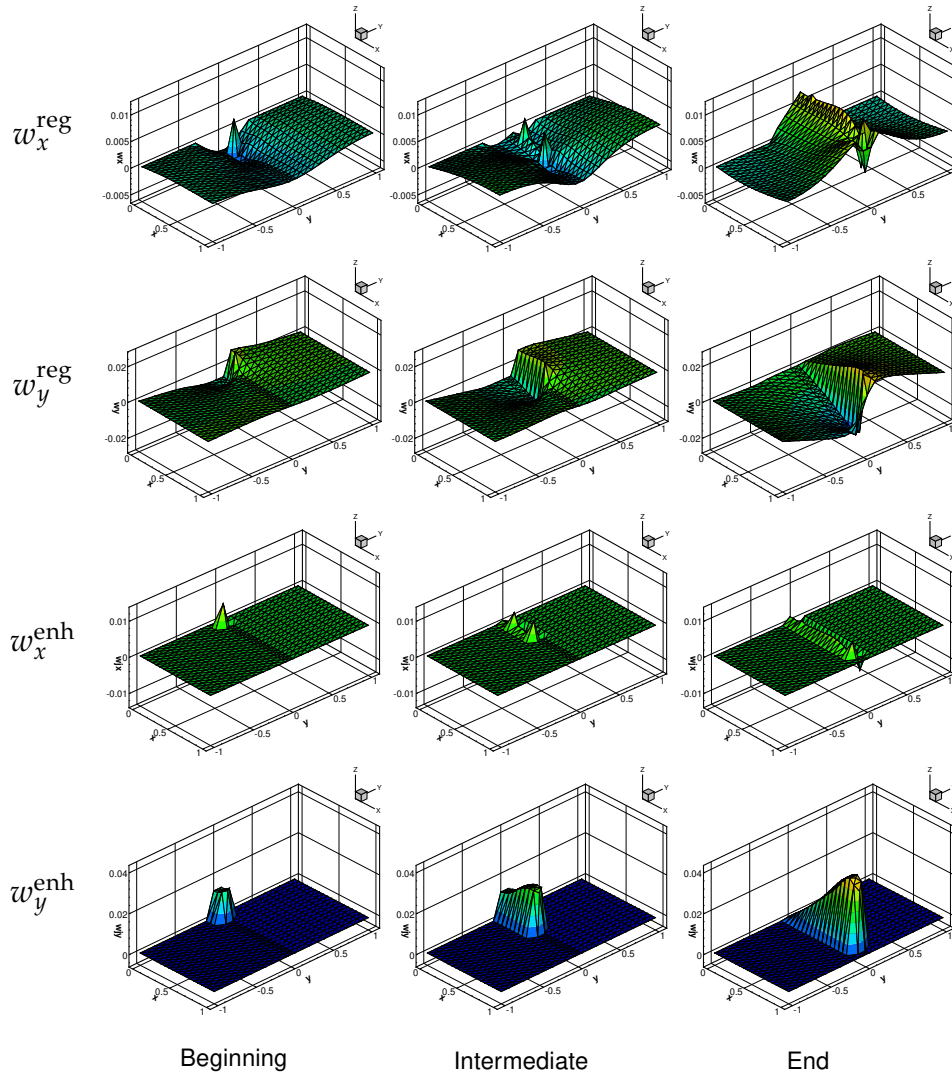[†]This is based on the assumption that a crack propagates in the direction normal to the maximum principal stress.

Figure 7.17 Evolution of the influence functions (w_x^{reg}, w_y^{reg}, w_x^{enh}, and w_y^{enh}) at different stage of crack propagation in the crossed crack test.
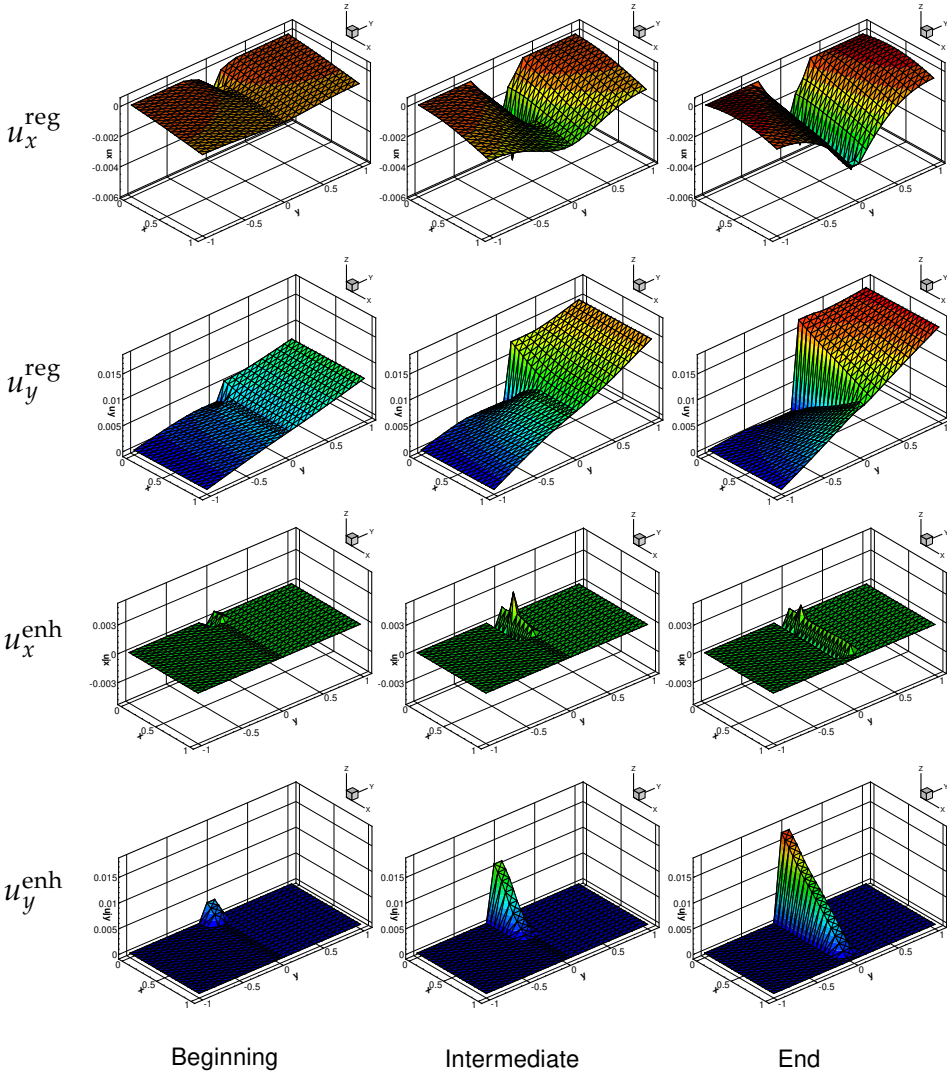
Figure 7.18 Evolution of the primary unknowns (u_x^{reg}, u_y^{reg}, u_x^{enh}, and u_y^{enh}) at different stage of crack propagation in the crossed crack test.
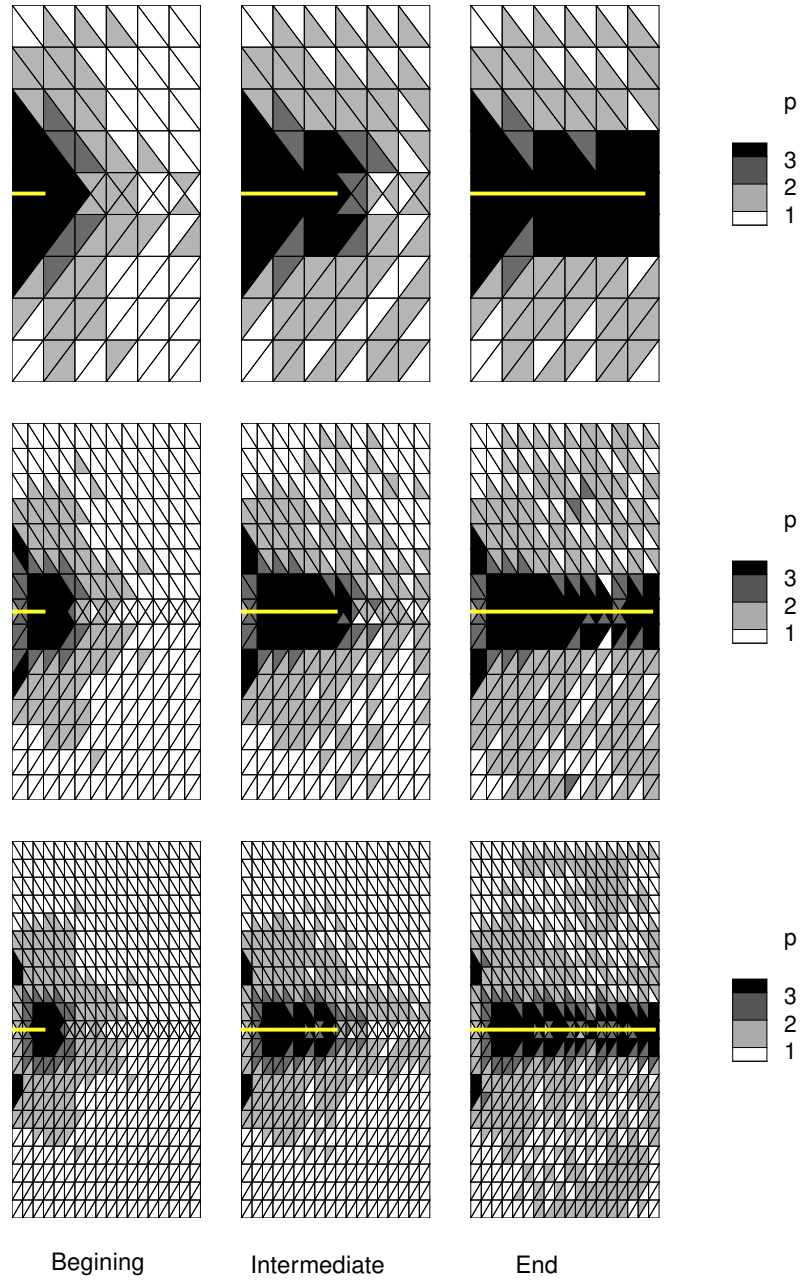
Figure 7.19 Distribution of the interpolation orders in the crossed crack test at the beginning of the computation, at the half-way crack propagation and at the end of the computation. Goal-oriented adaptivity is applied in three uniform linear meshes.
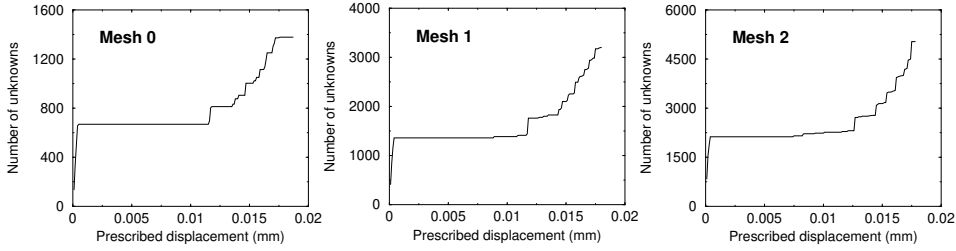
Figure 7.20 The increased number of unknowns during the p-adaptive PU-based discontinuity analysis based on goal-oriented error estimation.

measured via the goal-oriented error estimation, which is based on the solution of the dual problem (also known as the influence functions).

By solving the *dual problem* at the end of each loading increment, four influence functions, each of which corresponding to each primary unknown, are shown in Figure 7.17. These influence functions reveal how the error in each unknown field influences the average of the stress σ_{yy} in the element ahead of the crack tip. It is appealing that the residuals corresponding to the second and the fourth degrees of freedom provide major influences on the quantity of interest, as observed from the evolution of w_y^{reg} and w_y^{enh}. Compared to the evolution of the primary unknowns (cf. Figure 7.18), the evolution of the corresponding components (i.e. u_y^{reg} versus w_y^{reg}, and u_y^{enh} versus w_y^{enh}) are reversed.

(B) Analysis of the goal-oriented p-adaptivity

We employ the error in the quantity of interest via the measure $\mathcal{E}_k = \|\mathbf{e}\|_k^{\text{mod}}\|\boldsymbol{\epsilon}\|_k^{\text{mod}}$ to drive the goal-oriented *p*-adaptivity. Controlled by a threshold value ζ_{goal} (cf. Subsection 5.2.2) of 0.3%[‡], the elemental interpolation orders evolve during the crack propagation as shown in Figure 7.19.

At the beginning when the crack has not yet propagated, it is evident that the accuracy of the stress σ_{yy} in the element ahead of the crack tip (as our goal quantity) depends greatly on the discretisation in the crack tip region where the local quantity of interest resides, as well as where the supports change type (i.e at the tip of the vertical crack). This is similar to the adapted mesh shown in Figure 7.13, although it seems that more enrichment in the support-change region and more intense discretisation in the crack path are desired in the goal-oriented scheme. For the rest of the domain, the enrichment levels obtained from both adaptive schemes are comparable. Figure 7.20 shows that the number of unknowns used in this goal-oriented adaptive analysis is slightly higher than the one demanded in the earlier analysis (cf. Figure 7.14).

[‡]This is a very tight value. However, since this basic problem is for an intensive investigation, we would like to see how *p*-adaptivity works.
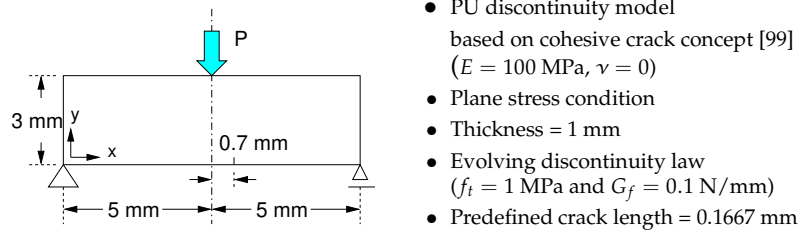
- PU discontinuity model
  based on cohesive crack concept [99]
  $\left(E = 100\,\text{MPa}, \nu = 0\right)$
- Plane stress condition
- Thickness = 1 mm
- Evolving discontinuity law
  $(f_t = 1\,\text{MPa and } G_f = 0.1\,\text{N/mm})$
- Predefined crack length = 0.1667 mm

Figure 7.21 Description of the three-point bending test [99].

Mesh	No. of Nodes	No. of Elements	p-order	No. of basic DOF
0	237	403	1	474
			2	1752
			3	3836
			4	6726
1	495	888	1	990
			2	3754
			3	8294
			4	14610
2	1520	2860	1	3040
			2	11798
			3	26276
			4	46474

Table 7.3 Information of fixed meshes used in the three-point bending test. The basic degrees of freedom includes the basic displacement modes without any jump contributions, i.e. only the first and the second degrees of freedom are counted.

7.4 Three-point bending test

The three-point bending test [99], whose detailed descriptions are given in Figure 7.21, is chosen as our third test. The pre-defined crack is located at a 0.7-mm offset to enable a curved or inclined crack. The beam, subjected to a point load at the middle, is analysed by displacement control. At the first step, the prescribed displacement of 0.1 mm is applied. Afterwards, incremental displacements of 0.001 mm are applied in each computational step in the full Newton-Raphson iterative scheme. The semi-circular interaction domain for searching the crack orientation has a constant radius for all cases equal to 0.40 mm.

To investigate the *h*-factor and the *p*-factor of the finite element discretisation, we select three uniform triangular meshes, namely Mesh 0, Mesh 1 and Mesh 2, for each of which four orders of interpolation (*p*=1 to 4) are applied. Table 7.3 gives some details of these reference meshes.
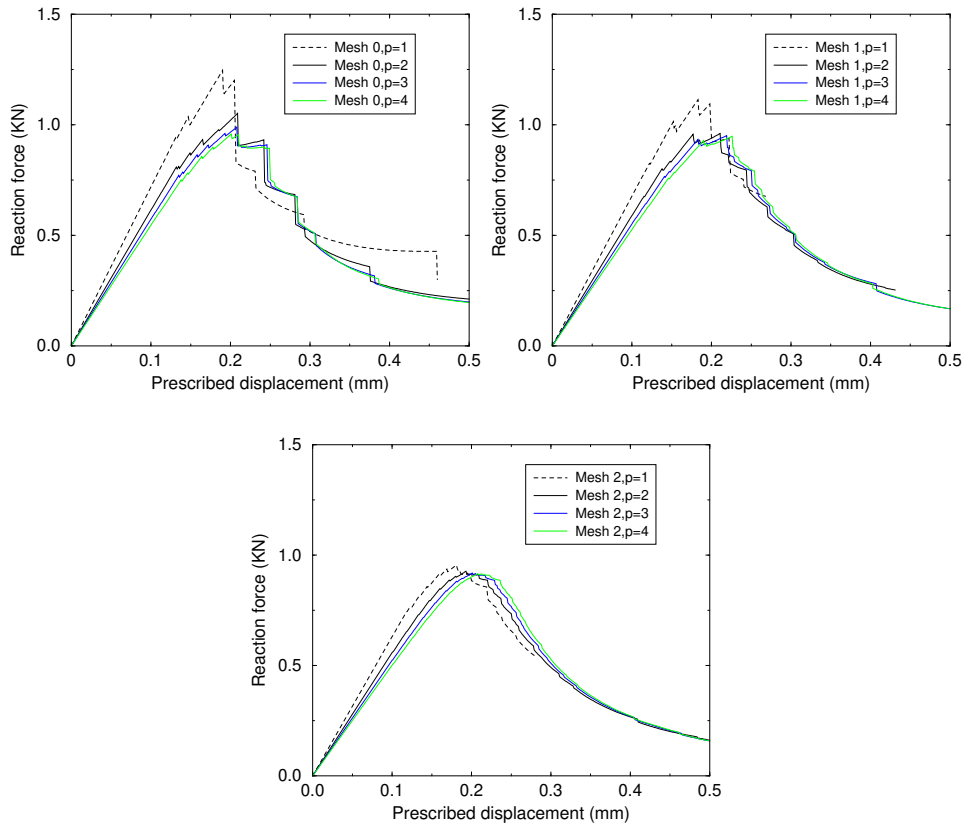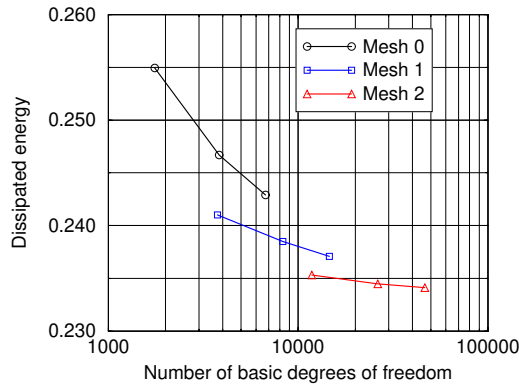
Figure 7.22 Load-displacement relations of the three-point bending test based on Mesh 0, Mesh 1 and Mesh 2. The interpolation ranges from linear order to quartic order ($p = 1$ to 4).

7.4.1 Preliminary investigation

We start our investigation by examining effects of the mesh resolution (*h*-factor) and the interpolation degree (*p*-factor) on the finite element solution. By varying the order of uniform interpolation, the load-displacement responses obtained from the three reference meshes are plotted in Figure 7.22. The most deviating results are obtained with Mesh 0 with linear interpolation. All meshes with linear interpolation fail to complete the loading process. Figure 7.25 shows the number of unknowns emerging during the computation.

The dissipated energy is plotted against the number of basic degrees of freedom in Figure 7.23. Since the linear interpolation is inadequate for this finite element computation, we select the results from all the meshes with the interpolation ranging from a quadratic order to a quartic order for this convergence test. Confirming

Figure 7.23 Dissipated energy of the three-point bending test based on Mesh 0, Mesh 1 and Mesh 2. The interpolation ranges from quadratic order to quartic order ($p = 2$ to 4). (The unit is kN · mm)
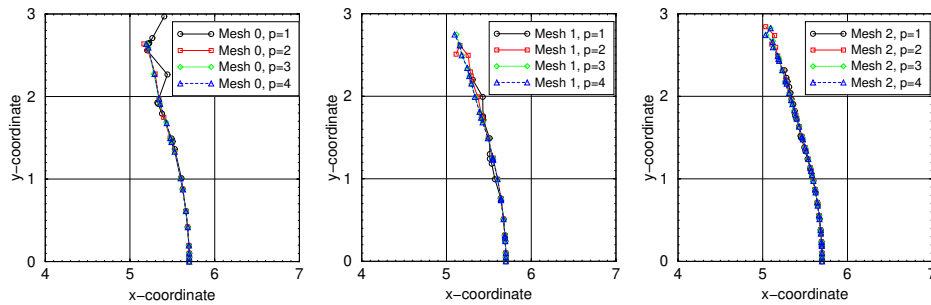
Figure 7.24 Various crack paths obtained by using Mesh 0, Mesh 1 and Mesh 2. Uniform orders of interpolation ranging from $p = 1$ to 4 are applied.

the results from Section 7.3, the mesh quality is improved upon mesh refinement (*h*-extension) and mesh enrichment (*p*-extension)[§]. The former provides better performance between the two, however. Not only it gives smoother global responses, it also leads to faster rates of convergence in terms of the dissipated energy.

In Figure 7.24, the crack paths computed from all meshes follow in the same orientation. The only obvious exception is the case of Mesh 0 with linear interpolation. According to Section 7.1, the determination of the crack direction is based on a nonlocal computation of the principal stress, and the interaction region (i.e. the region used for searching of crack orientation) is suggested to be three times the average element size in front of the crack tip [99]. Since we fix the region to be equal in all the cases, it is possible that, for a coarse mesh with a low order of interpolation, this area for nonlocal averaging is not sufficiently large, thus resulting in a wrong crack

---

[§]Note that, although the exact solution is unknown, the responses tend to converge to the same dissipated energy (cf. Figure 7.23).
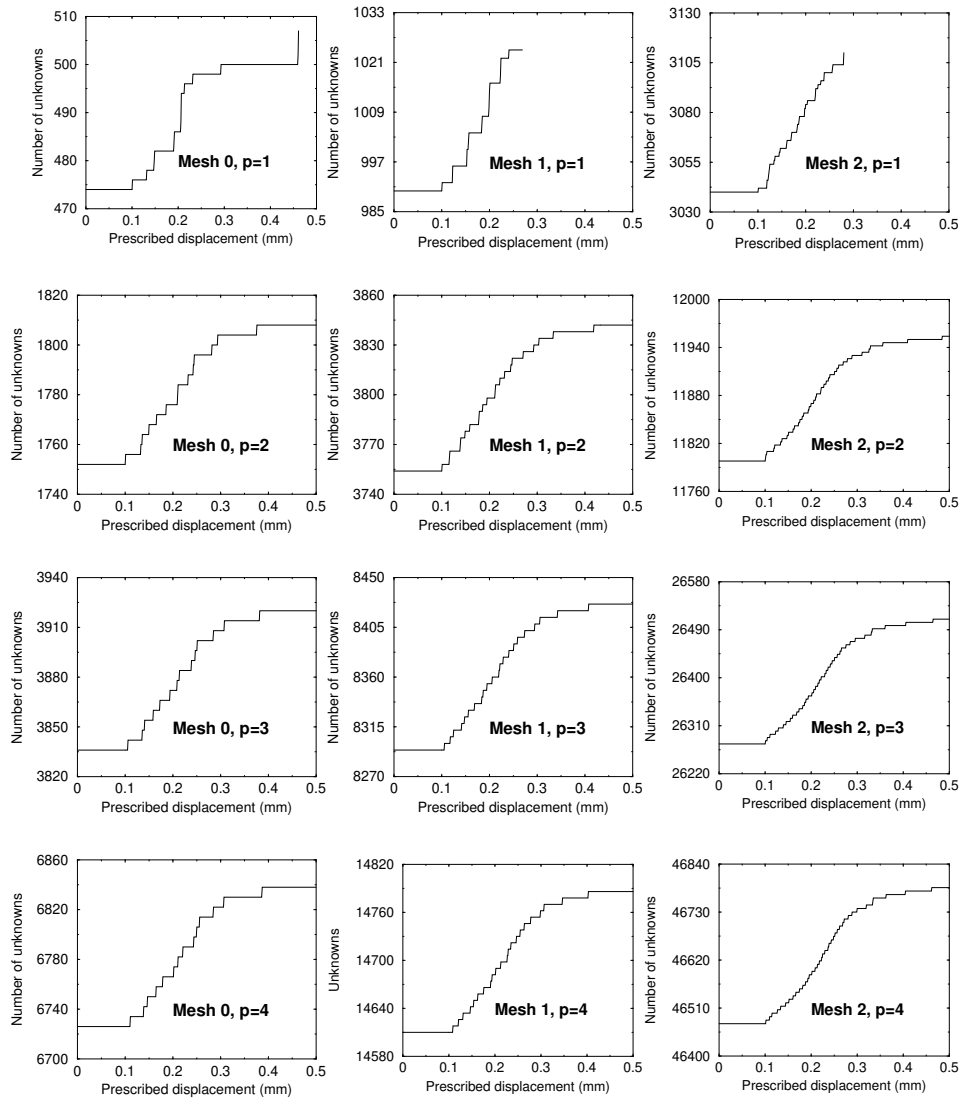
Figure 7.25 The increased number of unknowns during the PU-based discontinuity analysis in the three-point bending test. Three uniform meshes with four orders of interpolation are surveyed.

path.

7.4.2 Mesh adaptive tests

To trigger an adaptive analysis of the three-point bending test, the adaptive criteria of Section 5.2 are applied in this subsection with a prescribed global error allowance ζ_{prim} of 15%. The discretisation error is assessed at the end of some selected computational steps, by which the adaptive process may subsequently be driven. For implementational simplicity, only p-adaptivity is investigated, although the results shown in the preliminary tests suggest the competitiveness of h-adaptivity. Some aspects of the adaptive process will be investigated and discussed in this subsection.
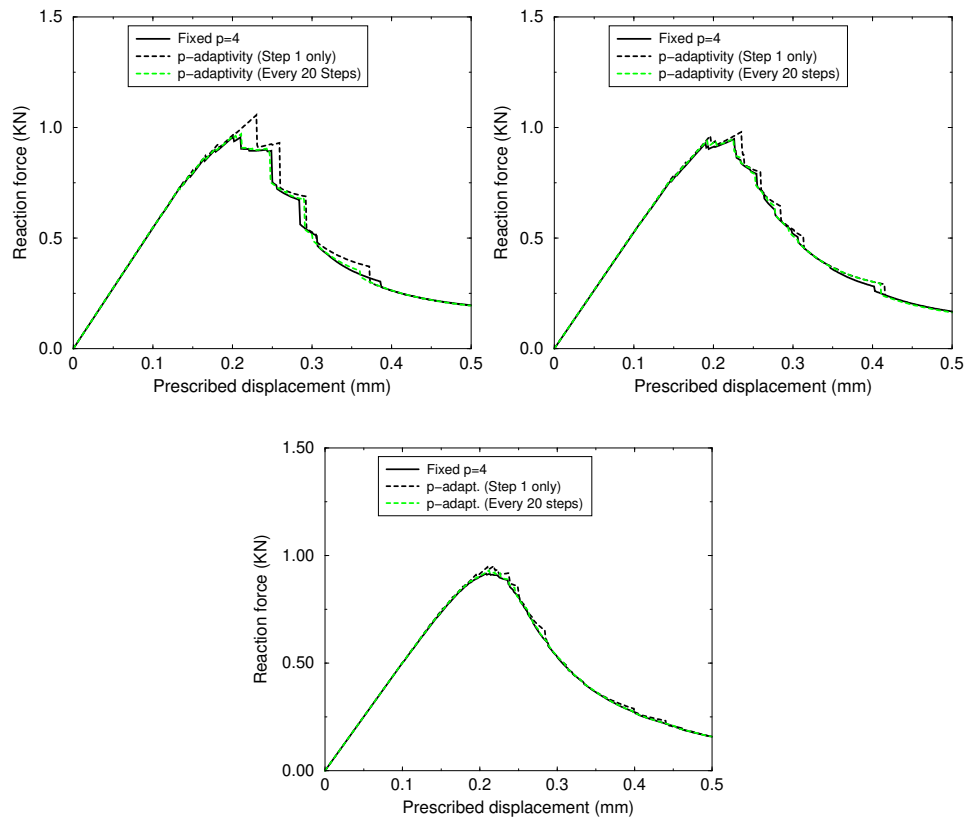
Figure 7.26 Load-displacement relations of the three-point bending test with p-adaptivity based on Mesh 0, Mesh 1 and Mesh 2.

p-adaptivity

Without change in mesh configuration, *p*-adaptivity can be regarded as the simplest adaptive technique for discontinuous modelling. In this study, the interpolation function is upgraded to one order higher at a time. Degrading is prohibited and quartic order of the interpolation is set as the maximum in our *p*-adaptive study.

There are two points to be investigated in this *p*-adaptive analysis. Firstly, we would like to make a survey about when an adaptive process should be activated. Secondly, it is essential to consider efficiency of the standard finite element analysis (i.e. with a fixed mesh), in comparison to the adaptive version. Some remarks regarding these two aspects can be made as follows.

(A) Effect of adaptive timing on FE solutions

It is widely known that the discretisation error tends to occur where a high stress (or precisely, stress gradient) exists (see also Section 7.3.1). The most critical region is at the tip of the traction-free discontinuity defined prior to the beginning of the computation, where we observe an intense amount of error. During cohesive crack propagation, the amount of error at the updated crack tip decreases. Based on this observation, a question arises whether a sufficiently good discretisation designed before the crack propagation is sufficient for use in the whole computation or not. To answer this question, we investigate two tests based on different adaptive timing schemes, i.e.

- Test 1: *p*-adaptivity is activated only once in the linear elastic regime (Step 1, $\bar{u} = 0.1$ mm), and
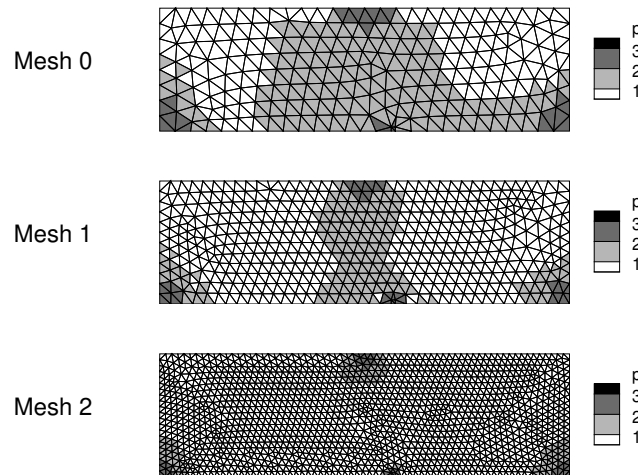
Figure 7.27 Distribution of the interpolation orders in the three-point bending test at $\bar{u} = 0.100$ mm.
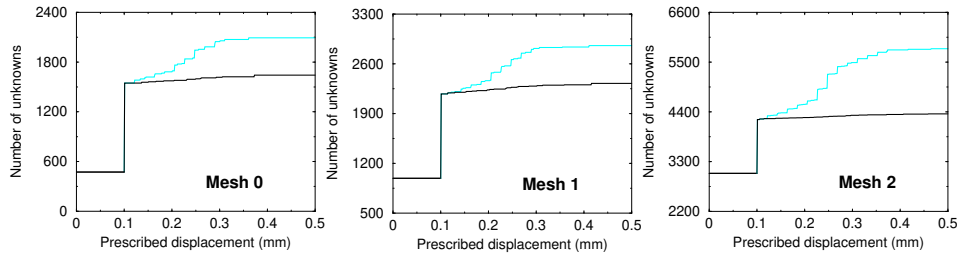
Figure 7.28 The increased number of unknowns during the *p*-adaptive PU-based discontinuity analysis. The black line refers to the adaptive process activated at Step 1 only (Test 1) and the grey line indicates the adaptive process activated at every 20 steps (Test 2).

- Test 2: *p*-adaptivity is activated every 20 steps during the computation.

The load-displacement responses obtained from both *p*-adaptive studies, in comparison to the fixed mesh of uniform quartic interpolation, are shown in Figure 7.26. It can be seen in all three subfigures, corresponding to Mesh 0, Mesh 1 and Mesh 2, that the responses obtained via Test 2 resemble the one based on the uniform quartic mesh. Although not as good, activating only one mesh adaptive process (cf. Test 1) greatly improves the original responses especially in the linear elastic regime. The corresponding mesh designs are shown in Figure 7.27, where higher-order interpolations are applied at the supports and the pre-defined crack. Nevertheless, during crack propagation, the responses based on Test 1 clearly deviate from the other two cases, especially in Mesh 0 which is the coarsest of all. The initial element size apparently plays an important role. Without any mesh adaptive process during crack propagation, a mesh of smaller elements is more capable of modelling the cracking process.

The above results clearly suggest that the adaptive process should be activated during crack propagation as well as at the beginning of the analysis. However, as the number of unknowns to be solved in Test 1 is less than the one required in Test 2 by a factor of 1.2 - 1.4 for all meshes (cf. Figure 7.28), it may be more cost-efficient in case of a fine mesh (in this context, Mesh 2) to activate *p*-adaptivity only at the beginning since it can already provide a reasonably good approximation, considering from the global responses shown in Figure 7.26.

(B) Efficiency of p-adaptivity

To investigate efficiency of the *p*-adaptive analysis, only the results from Test 2 will be discussed. The discretisation error is assessed at the end of every 20 loading increments, implying in total 20 possible adaptive processes during the entire analysis.

As mentioned earlier, the load-displacement responses obtained from *p*-adaptivity in all meshes can resemble, more or less, the ones obtained by using the same mesh with the uniform quartic interpolation (cf. Figure 7.26). Despite that, costs required during the computation for both cases are quite different. The
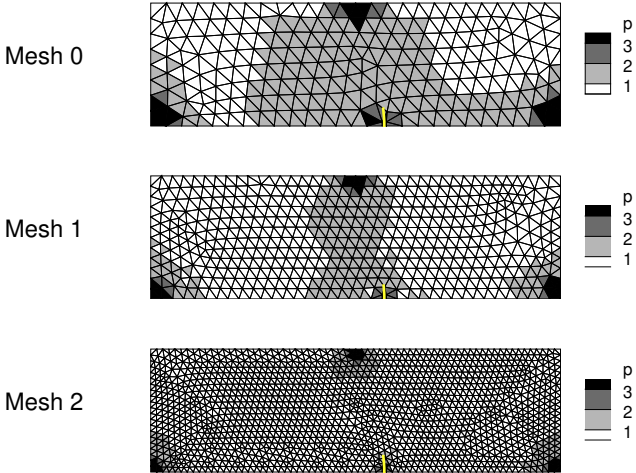
Figure 7.29 Distribution of the interpolation orders in the three-point bending test at $\bar{u} = 0.142$ mm.
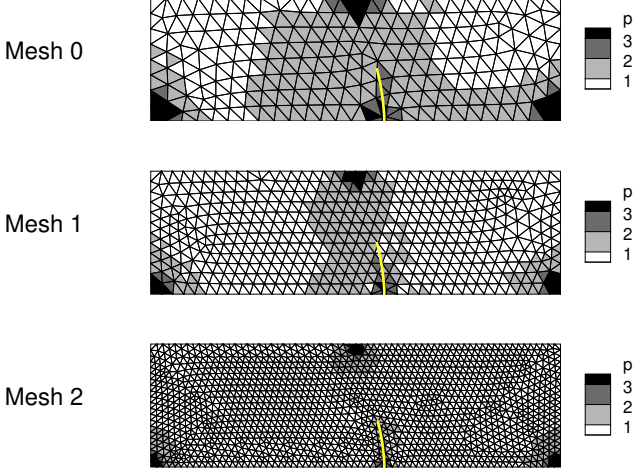

Figure 7.30 Distribution of the interpolation orders in the three-point bending test at $\bar{u} = 0.205$ mm.
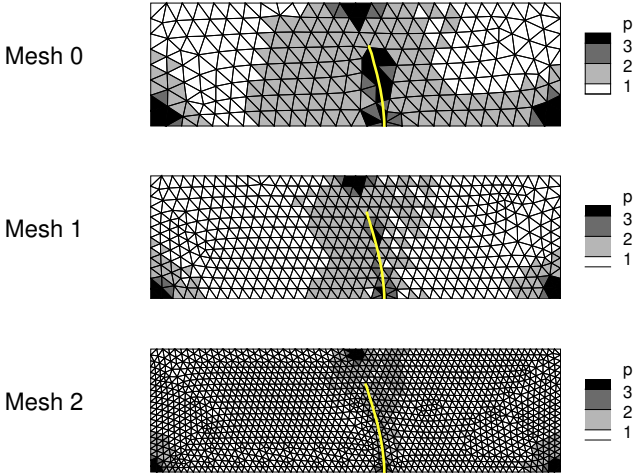
Figure 7.31 Distribution of the interpolation orders in the three-point bending test at $\bar{u} = 0.268$ mm.
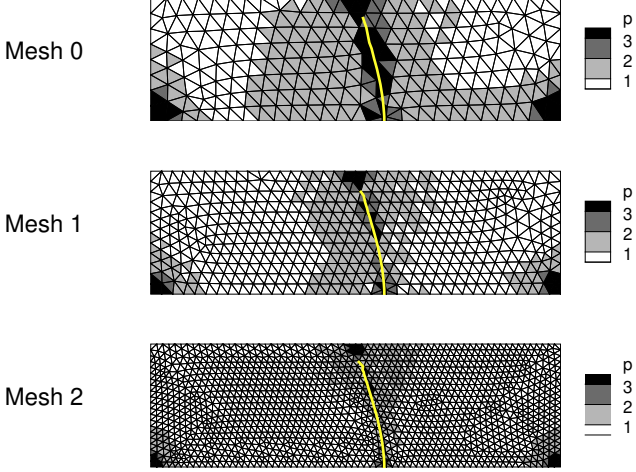

Figure 7.32 Distribution of the interpolation orders in the three-point bending test at $\bar{u} = 0.331$ mm.

number of unknowns, implying the computational cost, can be greatly reduced via p-adaptivity. In comparison to Figure 7.25, Figure 7.28 reveals that the number of unknowns at the end of the computation can be decreased via p-adaptivity by a factor of 3.25 in Mesh 0, a factor of 5.15 in Mesh 1 and even a factor of 8.00 in Mesh 2.

From Figures 7.29 to 7.32, elemental distributions of the interpolation orders are plotted during the analysis for all three base meshes. It is evident that the use of higher-order elements is concentrated at the supports and tends to follow the propagation of the discontinuity. These problem regions are smaller in size, implying more intense higher-order interpolation, for finer meshes. For the rest of the problem domain, linear interpolations are sufficient.

7.5 Remarks

In this chapter, the discontinuity modelling based on the PU-based cohesive zone concept [99] has been investigated. Despite the additional degrees of freedom, the error estimation used in this discontinuity model follows the standard concept addressed in Chapter 3. To set a proper error norm, the linear-elastic stiffness matrix is employed since it is positive-definite. The results from all numerical examples have shown that, prior to crack propagation, the discretisation error appears to concentrate at the supports and the pre-defined crack tip (where singularities exist), while intense errors follow the updated crack tip during crack propagation. Based on the average stress quantity in the element ahead of the crack tip, the goal-oriented error has shown a similar trend, although mesh improvement is clearly needed more in the crack path area.

Comparing two mesh improvement techniques, h-extension provides two superior qualities over p-extension. Firstly, for a propagating crack, a sufficiently refined discretisation is obviously needed to ensure a smooth load-displacement response. Secondly, convergence of the dissipated energy during the cracking process tends to be faster upon mesh refinement. Nevertheless, p-adaptivity has been chosen in this research as it can improve the result greatly while requiring much less implementational effort than h-adaptivity. In addition, since the PU model has been proposed with a strong intention to avoid a remeshing procedure needed in other crack models, remeshing the domain via h-adaptivity appears to be unreasonable.

Conclusions

Using finite element analysis, the approximate solution to a model problem relies largely on the spatial discretisation of the problem domain. A more refined/enriched discretisation provides a more accurate result but requires a higher computational cost. By adaptive spatial discretisation, a balance between cost and accuracy of the finite element solution can be realised. In other words, the refined/enriched discretisation needs not be applied to the whole problem domain but limited only to where needed. Thus, it is very efficient for use in problems where the critical regions are small in comparison to the rest of the problem domain. For quasi-brittle failure, which is our main focus in this research, adaptive discretisation can be beneficial as strains tend to localise in a small zone/band at the moment that the material loses its ability to carry loads.

PART I: INGREDIENTS OF ADAPTIVITY

(A) Hierarchical shape functions:

Two discretisation factors, regarding how refined the mesh is (h-factor) and how enriched the interpolation is (p-factor), have been studied in this research. We have employed hierarchical interpolation functions because their higher-order components can be added or removed without changing the original system of equations. Two hierarchical versions have been put in consideration. The node-based version (cf. Chapter 2 and Appendix A), wherein the higher-order components are added via partitions of unity, can be implemented in a more straightforward fashion (also at some slight extra cost) than the edge-based version (cf. Chapter 2). Nevertheless, the node-based version suffers from linear dependence of the system of equations, and even though the ill-conditioned system can be treated by introducing special constraints or using special solution techniques, it becomes very complicated. It

has been found also that, since the linear dependence becomes more dominant when the scheme is applied to a smaller number of elements, it is hard to assess the error based on the local patching scheme (cf. Chapter 3). For these reasons, the edge-based version has been chosen in this study.

(B) Error estimation:

The adaptive discretisation process, in this context, has been steered by objective error information, rather than heuristic assessment of the state variables. Two error estimates are presented, namely the energy-based error measure (cf. Chapter 3) and the goal-oriented error measure (also known as error measure of the local quantity of interest) (cf. Chapter 4). The estimation of error is based on solving a series of local problems, based on patches consisting of elements surrounding each node, with prescribed homogeneous boundary conditions. The order of the interpolation function is upgraded to generate a reference discretisation. With the patch-overlapping scheme, the need to equilibrate inter-element flux jumps can be avoided [29].

The performance of the error estimator has been investigated in linear elastic problems (cf. Chapter 3) and it has been shown that the method can provide a good lower-bound estimate of the actual error. A more accurate error estimate (to the actual error) has been obtained in the problem domain of more refined/enriched discretisation. Explainably, the reference discretisation is constructed by improving the original mesh quality to *another* level. For a higher-quality mesh (i.e. a more refined/enriched discretisation), a fixed degree of local improvement leads to a higher-quality reference discretisation, hence correspondingly a higher-quality error estimate to the exact error. We have also found that, in general cases, upgrading the local interpolation function only by one order can already provide a sufficiently accurate error estimate. Nevertheless, the one-order upgrading scheme may not be sufficient to provide good error estimates for all of the finite elements in some special cases, as we have also observed in one of our numerical examples (cf. Section 3.7).

Although the formulation of the error estimation is based on a measurement in a norm quantity, it has been found in our study (cf. Chapter 3) that the estimated error functions, which are obtained from solving the local error equations, can somehow resemble the actual/reference ones. However, we have found that performance of the error estimator is very sensitive to how the reference system is modelled. Without a correct reference discretisation, the estimated error can be far from the actual error.

In Chapter 4, the error estimation concept has been extended for determining the error in the quantity of interest. This goal-oriented error computation is rather straightforward at an additional price of solving a set of global equations for the influence function and also another series of local problems for its error. To facilitate the setting of adaptive criteria, two new error measures, constructed by directly summing up the elemental contributions, have been proposed and investigated in a linear elastic problem. The difference between the two measures

lies on how the error contributions from the primal and the dual problems are combined; the first measure mixes both contributions in the norm computation, while the second measure separately considers each contribution in the norm computation and then combines both norms together. The second measure may be preferred since the obvious separation of the error norms from the primal and the dual problems (cf. the second measure) gives the analyst more clue on how much the error allowance should be set. The results have shown that both measures provide similar error trends in terms of global convergence and their elementwise error distribution.

(C) Mesh adaptivity:

Once the error information is at hand, it is necessary to judge whether or not the finite element mesh should be improved and in which specific regions. The adaptive process is activated when the global error exceeds a user-specified global error allowance. Based on the uniform error distribution, the process takes place in the regions of large elemental errors. In addition to the adaptive criteria, the optimality criterion based on an *a priori* convergence assumption may be applied for an optimal design of element sizes, in case of *h*-adaptivity (mesh refinement) and *r*-adaptivity (node relocation). For *p*-adaptivity (mesh enrichment), the convergence assumption is not applicable, and, even if it were, the interpolation orders would simply be too hard to be optimally adjusted. In this study, only *integer* orders of polynomial degree have been considered. Following the same idea, the adaptive criteria and the optimality criterion for use with goal-oriented error quantity have also been introduced in Chapter 5.

In this research, a mesh gradation technique (for *r*-adaptivity) based on a weighted Laplace smoothing concept has been employed (cf. Chapter 5). This gradation proceeds in a simple node-by-node fashion and does not require solving any global system of equations. Despite its simplicity, the technique has been proven a success as nodes can be moved towards the zones of large errors, improving the local mesh resolution and successfully reducing the error in those critical regions.

PART II: APPLICATIONS

The above-mentioned concepts of error estimation and adaptive discretisation have been applied in the modelling of quasi-brittle failure. Here, we have considered the failure mechanisms described by the continuous concept, where failure is presented by means of material strength degradation (in this context: damage), as well as by the discontinuous concept, where failure is presented by means of a geometrical discontinuity.

(D) Application to continuous failure

For modelling continuous failure, the gradient-enhanced damage model [70] has been chosen (cf. Chapter 6). Since this model is regularised in the post-peak regime, the finite element solution does not suffer from pathological mesh dependence and thus converges to a physically realistic solution upon mesh refinement.

By using a softening model, it is possible that the consistent tangent stiffness matrix becomes negative-definite, and thus the resulting norm becomes meaningless. To prevent this from happening, we have selected only positive-definite parts of the stiffness matrix for computation of the error norm.

◇ *Distribution of the error:*

The error analysis has shown that, in the linear elastic regime, the discretisation error concentrates at the pre-defined crack tip where stress singularities exist. Upon damage evolution, the discretisation error in the zone where damage appears becomes dominant. This very zone is where the strains localise, and a sufficiently refined/enriched discretisation is needed most at the boundary of the localised strain profile, i.e. in the zone of high strain gradients.

◇ *Goal-oriented error:*

The goal-oriented error analysis, based on the crack mouth opening displacement, has revealed that, prior to damage initiation, the discretisation error concentrates at the pre-defined crack tip as well as the point of interest (i.e. the crack mouth). Once damage appears, large errors emerge in the region of localised strains as well. This observation has proven that the accuracy of the nonlocal equivalent strain field (and the corresponding damage quantity) greatly influences the accuracy of the quantity of interest. Our two proposed error measures have provided similar error distributions, although some small differences have been found in the softening regime.

◇ *p-, h- and r-extension:*

Based on a uniform mesh improvement, we have found that increasing the interpolation order (*p*-extension) may provide a faster rate of convergence than decreasing the element size (*h*-extension). However, for problems with highly localised strain fields, it is very important that a sufficient mesh resolution is employed in critical regions, in order to correctly capture the strain localisation phenomenon during the failure process. Without a sufficient mesh resolution, the localised strain profile and the corresponding damage profile will be dispersed, resulting also in a too ductile response, even for meshes with high degrees of interpolation.

Three adaptive techniques have been investigated in this study. With the optimally-designed element sizes, *h*-adaptivity has proven to be very efficient as the computational cost is reduced significantly. Since a high mesh

resolution is clearly needed in the region of high strain gradients, the simple mesh gradation (*r*-adaptivity) has also proven to be cost effective. This low-cost technique has provided a dramatic improvement of the result without increasing the number of unknowns. Applying *p*-adaptivity can also improve the solution greatly, although, as also confirmed by the uniform enrichment tests, the technique alone is not as efficient as the other techniques in capturing strain localisation, since it depends somewhat on the resolution of the finite element mesh. Without a good original mesh design, *h*-adaptivity or *r*-adaptivity is more preferred. However, as a recommendation for future research, it may be interesting to investigate combined techniques such as *hp*-adaptivity or *rp*-adaptivity, as they are expected to serve as ideal choices for this damage modelling.

⋄ *Selection of adaptive steps:*

The adaptive process may be activated at the end of any computational step. For the nonlinear analysis, the solution path depends greatly on the discretisation. It has been found in our numerical examples that it is equally important to activate the adaptive process before and after damage initiation. Without a good discretisation in the linear elastic regime, initiation of the failure mechanism may be wrong. A good discretisation is also needed during damage evolution to ensure a correct simulation of failure.

⋄ *Variable transfer:*

Another aspect that we have investigated in this research concerns the history transfer process. Once the discretisation is re-designed, the variables stored at the integration points as well as the nodal degrees of freedom must be transferred. Here, the variables stored at nodes (including hierarchical modes) have been transferred by a conventional interpolation technique and an auxiliary remapping scheme (named as Lagrange-equivalent transfer) for transferring of the hierarchical modes. Based on extrapolation and interpolation techniques, the so-called shape function remapping transfer is suitable for the variables stored at quadrature points, although may also create an artificial expansion of the damage/localisation zone. It has been shown that this transfer is sufficiently accurate when combined with *h*-adaptivity, especially for lower-order meshes. For *r*-adaptivity, this shape function remapping technique must be completely avoided. Techniques that possess a low diffusion property (i.e. limited expansion of localisation zone), such as the closest point transfer [69], have proven to be a good choice, which is also applicable to *h*-adaptivity and *p*-adaptivity.

(E) Application to discontinuous failure

Besides the continuous approach, quasi-brittle failure based on discontinuity modelling has also been investigated in this research (cf. Chapter 7). We have chosen the cohesive zone model based on the partition-of-unity concept [100], as dis-

continuities can pass arbitrarily through finite elements without the restriction of a particular mesh alignment.

◇ *Distribution of the error:*

Our first investigation is concerned with the performance of the error estimator for use in PU-based discontinuity modelling. For a fixed traction-free crack, stress singularities are present at the crack tip, inducing so-called *pollution* in the error estimation. The pollution in this context is the result of replacing the global error computation with a series of local computations. It has been observed in our numerical test that this pollution quantity increases for a more refined discretisation, due to decreased interaction between distant patches, and/or for a less enriched interpolation, as the finite element solution is less smooth.

The error analysis has shown that, before crack propagation (i.e. the material still exhibits linear elastic behaviour), the discretisation error concentrates in the conventional critical areas, such as at the supports and at the pre-defined crack tip. During crack propagation, intense errors tend to concentrate at the updated crack tip. It has been noticed that the error at the cohesive crack tip is less concentrated than the error at the traction-free crack tip. This can be explained as follows. Before crack propagation, stress singularities appear at the pre-defined (traction-free) crack tip. Conversely, based on the cohesive zone concept, a bounded and more uniform stress distribution is obtained at the updated (cohesive) crack tip during crack propagation.

◇ *p- and h-extension:*

Based on the PU-based discontinuity modelling, the enhanced degrees of freedom are activated to model jumps in the displacement field. Since the discontinuity is extended through the whole element [100], a more gradual increase of the number of unknowns has been observed when a more refined discretisation is employed. The smaller element size allows the crack to be extended in smaller steps, hence resulting in a smoother response. Apparently, this element size factor (h-factor) plays an important role. Although the use of higher-order elements (p-extension) can also improve the response greatly, a better performance in terms of response smoothness and faster rate of convergence has been obtained in a more refined mesh (h-extension).

As a conclusion from the uniform mesh tests, h-adaptivity appears to be the best choice in improving the finite element results of this PU-based discontinuity model. Unfortunately, the model has been originally proposed with a strong intention to avoid a remeshing procedure needed in other models, hence applying h-adaptivity seems to be unreasonable. Compromisingly, we have chosen p-adaptivity for investigation as it can improve the result greatly while requiring much less implementational effort. Based on the energy-based error measure, higher orders of interpolation are clearly needed most

in the pre-defined crack tip regions and a lower order can be applied at the propagated crack tip. The goal-oriented adaptivity, however, has given more importance to the elements in the whole crack path, as to ensure the accuracy of the average stress in the element ahead of the crack tip. The results have shown that p-adaptivity can greatly improve the result, while also saving a lot of computational effort in comparison to the amount required when employing a uniform higher-order mesh.

The application of p-adaptivity in the discontinuity modelling should be able to provide a much better result if the discontinuity extension does not need to propagate through the whole element at a time. We expect the material response to be correspondingly smooth, without the use of a smaller-size mesh. It is then very interesting to apply p-adaptivity in the model proposed in [104], as the cohesive crack can propagate anywhere inside an element. Although this point is beyond the scope of this study, it deserves consideration for future research.

◇ *Degrading the interpolation order:*

According to the error analysis, it has been found that the discretisation error tends to concentrate always at the updated crack tip. After each crack extension, we have observed a decrease of the error in the elements behind the crack tip. A simple degrading scheme has been examined and the result has shown some fluctuations of the error and correspondingly the number of unknowns during the finite element computation. Based on the proposed criterion, we do not recommend adding the degrading scheme in the p-adaptive process.

In this dissertation, we have applied p-extension and h-extension to error estimation and mesh adaptivity, with an emphasis on the p-extension. As we have demonstrated, the hierarchical element-based p-extension is a suitable technique for error estimation (either in the energy norm or of a quantity of interest) as it facilitates the construction of the reference space needed in the setting of error equations. The approach can be applied with both h-adaptivity (total mesh refinement) and p-adaptivity, and can also provide a good error estimate for use in the adaptive modelling of continuous and discontinuous crack propagation. Whereas the p-extension seems to be the optimal choice for error estimation, both applications of crack propagation suggest that further improvement may be obtained through a combination with other adaptive techniques.

Critical survey on node-based hierarchical shape functions

In Chapter 2, the concept of node-based hierarchical shape functions [94] has been presented. Although this type of shape functions has not been used in the remainder of this thesis, it is worthwhile to address, in this appendix, some critical aspects that we have found at the beginning of our study. The aspects to be presented in this section include performance analyses of the shape functions in term of convergence, enforcement of boundary conditions and, finally, the main shortcoming of the scheme that precludes a widespread use of this method, i.e. the problem of linear dependence.

A.1 Convergence

Referring to Eqs. (2.14) – (2.16), a set of higher-order shape functions can be contructed based on adding extra terms to the linear shape functions. Consider the one-dimensional example described in Section 3.7, we investigate the convergence of the finite element solutions by upgrading the linear shape functions to quadratic, cubic and quartic orders.

The convergence analyses of the enriched set of shape functions, based on a different number of primary nodes, are plotted in Figure A.1. With respect to the number of nodes (cf. Figure A.1 (left)), the results show that all higher orders of enhancement yield quadratic convergence rates. Although the error analysis shown in Figure A.1 (left) does not provide an obvious difference between the results obtained by using the cubic and the quartic shape functions, the plot of error versus number of degrees of freedom (cf. Figure A.1 (right)), implying the computational cost required in the finite element analysis to achieve the same accuracy, reveals that a *larger* number of degrees of freedom is required in the quartic interpolation than
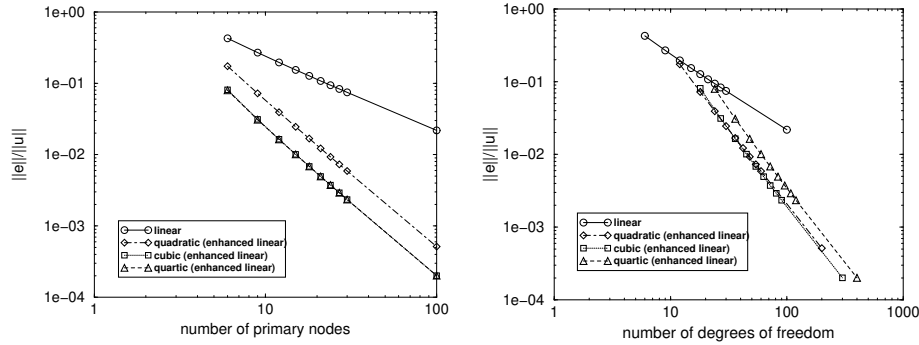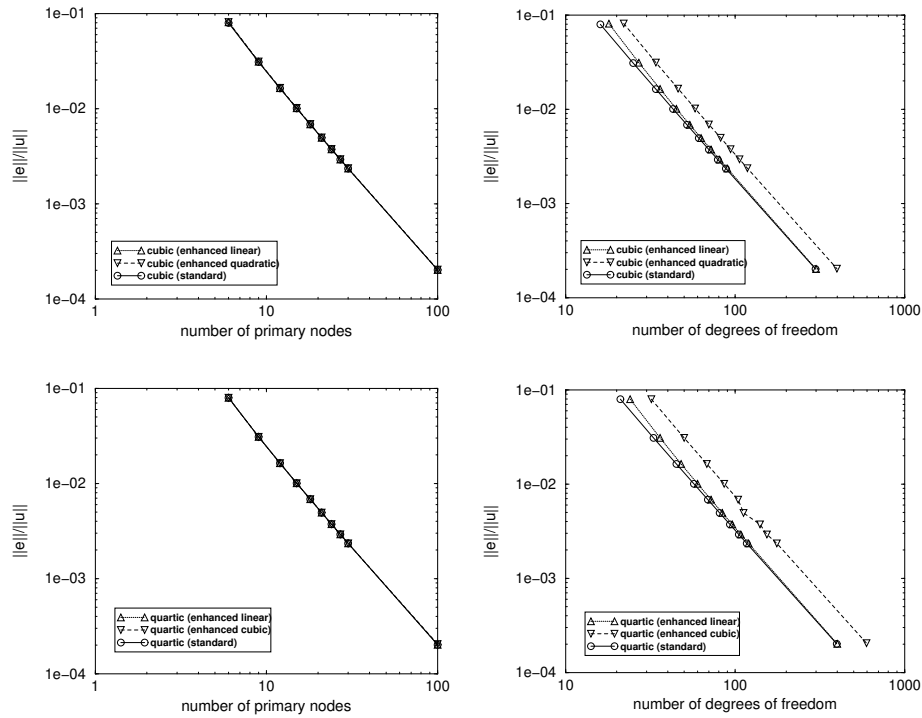
Figure A.1 Convergence analysis of the problem $-\frac{d^2u}{dx^2}(x) = 6x^2 - 3x$ in $\Omega =]0,1[$ with the prescribed boundary conditions $u(0) = u(1) = 0$ comparing different orders of enhancement.

order	required DOFs	provided DOFs
quadratic	6	12
cubic	10	24
quartic	15	39

Table A.1 Number of the equations required for achieving the higher-order interpolation and those provided in the new hierarchical method for a triangular element.

it is in the cubic interpolation. The background of this counter-intuitive finding is provided in Table A.1. An increasing unbalance between required and provided degrees of freedom is found for the higher orders of the interpolants.

Not only the new hierarchical enhancement can be applied on the linear basis function, but it can also be applied in any existing basis. For example, the cubic interpolation can be enhanced either from the linear basis using 7 extra functions per direction, as described in Eq. (2.15), or from the existing 6-node isoparametric elements (quadratic interpolation) using 4 extra functions per direction, that is

$$\gamma^{(i)}_{(2\to3)} = \{\xi_i^3, \xi_i^2\eta_i, \xi_i\eta_i^2, \eta_i^3\} \tag{A.1}$$

The convergence curves in Figure A.2 show similar accuracies obtained in the analyses applying three different cubic interpolation schemes. However, the cubic interpolation scheme based on enhancing the existing 6-node isoparametric elements is computationally more expensive as implied in Figure A.2 (right).

A.2 Enforcement of boundary conditions

As a common feature of the hierarchical finite element method, one of the serious shortcomings is how to enforce boundary conditions. Since no extra node is added

Figure A.2 Convergence analysis of the problem $-\frac{d^2u}{dx^2}(x) = 6x^2 - 3x$ in $\Omega =\,]0,1[$ with the prescribed boundary conditions $u(0) = u(1) = 0$ comparing different cubic and quartic interpolation schemes.
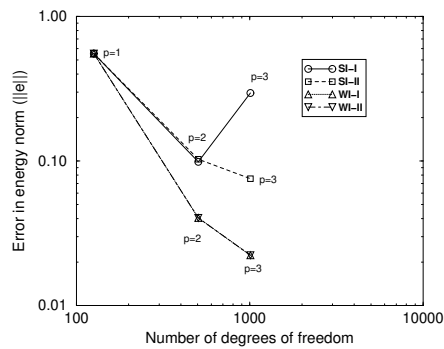
Figure A.3 Convergence analyses based on different constraint equations.

in the problem domain, constraining only the existing nodal points can be inadequate, i.e. the finite element solution may not converge to the true solution. To investigate this hypothesis, a cantilever beam (Timoshenko's beam [96]) with a set of contraints of a quantity $\overline{u}$ imposed at the left and the right ends is considered.

Four different methods of imposing boundary conditions used in this study are listed in Table A.2, namely the strong imposition schemes (SI-I and SI-II) and weak imposition schemes (WI-I and WI-II) based on the use of a penalty formulation [65]. Test SI-I imposes the Dirichlet boundary conditions by constraining the existing degrees of freedom only. This means the higher-order polynomial still exists along the boundary and the boundary conditions are interpolated using higher-order functions. Test SI-II constrains the nodal values on the Dirichlet boundary and also cancels the influence of higher-order function. This results in a linear interpolation of the constraints on the constrained edge. However, eliminating the higher-order function at the boundary nodes implies an incomplete higher-order enrichment function in the elements on the boundary. Test WI-I and WI-II are more realistic since the constraints are prescribed at the integration points on the constrained edges. The penalty parameter should be chosen to be a large value in order to impose the constraints properly*, implying an increase of the conditioning number of the stiffness matrix. Here, we have chosen the penalty parameter to be a thousand times of Young's modulus.

The finite element solutions using different constraints are presented in Figure A.3. As expected, the finite element solution obtained by constraining only the existing nodal values (cf. Test SI-I) shows an increase of the error when the enrichment reaches the cubic order. Test SI-II does not show such a divergence pattern. However, when compared with those using penalty formulations (WI-I and WI-II), the error is much larger. Note that there is no difference between the solutions obtained by Test WI-I and Test WI-II.

Method	Constraint Equation(s)
SI-I	$a_i = \overline{u}_i$
SI-II	$a_i = \overline{u}_i$ and $b_j^{(i)} = 0$
WI-I	$u(\mathbf{x}) = \overline{u}(\mathbf{x})$
WI-II	$a_i = \overline{u}_i$ and $u(\mathbf{x}) = \overline{u}(\mathbf{x})$

Table A.2 Four methods for constraint imposition on Γ_d.

---

*A good choice of the penalty parameter can be ensured by applying positive as well as negative penalty parameters to provide bounds for the finite element solution [8].
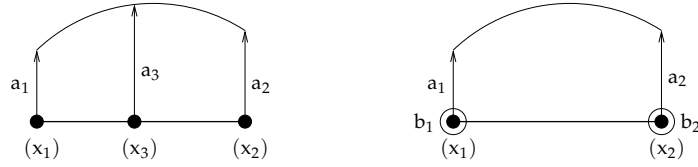
Figure A.4 One-dimensional quadratic elements: (left) standard element, (right) linear element with nodal enrichment

A.3 Linear dependence

Despite its simplicity in improving the mesh quality, the node-based hierarchical enrichment may bring about a serious problem: unsolvability of the discretised equations. It was reported in [94] that spurious zero energy modes were found in problems consisting of a small number of elements. These zero energy modes, apart from those describing three rigid body motions, bring about the singularity of the stiffness matrix.

Eigenvalue analyses of some element combinations are summarised in Table A.3. It is, however, obvious that these spurious zero energy modes occur since the resulting shape functions are linearly dependent. To explain this concept, let us consider a one-dimentional quadratic element that is constructed by the concept of nodal enrichment in comparison to the standard quadratic element shown in Figure A.4. The approximate function using the standard isoparametric element ($F(x)$) and the one using the linear element with nodal enrichment ($G(x)$) are

$$F(x) = \frac{(x_2 - x)(x_3 - x)}{(x_2 - x_1)(x_3 - x_1)}a_1 + \frac{(x_1 - x)(x_3 - x)}{(x_1 - x_2)(x_3 - x_2)}a_2 + \frac{(x_1 - x)(x_2 - x)}{(x_1 - x_3)(x_2 - x_3)}a_3 \quad \text{(A.2)}$$

and

$$G(x) = \frac{(x_2 - x)}{(x_2 - x_1)}a_1 + \frac{(x_1 - x)}{(x_1 - x_2)}a_2 + \frac{(x_2 - x)(x - x_1)^2}{(x_2 - x_1)}b_1 + \frac{(x_1 - x)(x - x_2)^2}{(x_1 - x_2)}b_2. \quad \text{(A.3)}$$

In this one-dimensional example, the next-order enrichment of the linear interpolation in $G(x)$ actually provides a cubic interpolation. Only in the special case that $b_1 = b_2$ is the interpolation given in Eq. (A.3) quadratic and equivalent to that of Eq. (A.2). However, this implies that b_1 and b_2 must be linearly dependent for the interpolation to be quadratic.

The degree of linear dependence may be measured via the so-called *nullity* (also called *rank deficiency*) in the eigenvalue analysis. The nullity shows how many zero energy modes, i.e. number of zero eigenvalues, appear in the stiffness matrix. Since the higher-order interpolation is constructed by adding the enrichment functions hierarchically, the nullity becomes larger in higher-order interpolations (cf. Table A.4). It is subsequently shown that the nullity relatively decreases with the number
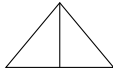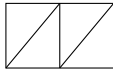
Arrangement	Order	Standard	PU-based		
		NDOFs	NDOFs	Zero Eigenvalues	Solvability
	$p = 2$	12	24	7	No
	$p = 3$	20	48	21	No
	$p = 4$	30	78	39	No
	$p = 2$	18	32	3	Yes
	$p = 3$	32	64	17	No
	$p = 4$	50	104	35	No
	$p = 2$	18	32	5	No
	$p = 3$	32	64	17	No
	$p = 4$	50	104	35	No
	$p = 2$	30	48	3	Yes
	$p = 3$	56	96	9	No
	$p = 4$	90	156	27	No
	$p = 2$	50	72	3	Yes
	$p = 3$	98	144	3	Yes
	$p = 4$	162	234	15	No
	$p = 2$	98	128	3	Yes
	$p = 3$	200	256	3	Yes
	$p = 4$	238	416	3	Yes

Table A.3 Eigenvalue analyses of some element combinations enriched by the nodal scheme. Note that the enrichment is based on enriching linear interpolation.
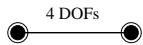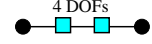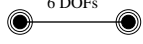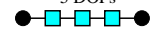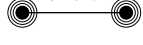
Standard	PU	Relative difference
3 DOFs	4 DOFs	1/3
4 DOFs	6 DOFs	1/2
5 DOFs	8 DOFs	3/5

Table A.4 Relative difference in number of degrees of freedom between a single standard higher-order element and a single linear element with nodal enrichment (PU).

Standard	PU	Relative difference
3 DOFs	4 DOFs	1/3
5 DOFs	6 DOFs	1/5
7 DOFs	8 DOFs	1/7

Table A.5 Relative difference in number of degrees of freedom between a set of standard quadratic elements and a set of linear elements with nodal enrichment (PU) to quadratic order.

Scheme	Linear	Quadratic	Cubic	Quartic
Standard	n	$2n - 1$	$3n - 2$	$4n - 3$
$PU_{(1\rightarrow2)}$		$2n$		
$PU_{(1\rightarrow3)}$			$3n$	
$PU_{(1\rightarrow4)}$				$4n$

Table A.6 Comparisons of global number of degrees of freedom in one-dimensional analysis using standard and PU-based hierarchical shape functions of various orders.

of degrees of freedom (cf. Table A.5 and Table A.6). This fact also implies that for problems with a few degrees of freedom, the nullity appears to be relatively large. This is precisely the situation if the PU based p-version is used for error estimation, whereby a higher-order interpolation is applied to estimate the error within an element or a patch of a few elements. Hence, the nodal hierarchical enrichment scheme is not suitable for error estimation, although it may be used for p-enrichment in the entire domain [94].

To solve the linear dependency problem, internal constraints (also known as *tyings*) between the degrees of freedom must be imposed. However, it requires a complicated mathematical derivation that is also different for each combination of elements. An obviously easier way is to choose the enrichment functions which do not belong to the span of the existing functions. For example, in upgrading from linear to quadratic interpolation, one may select

$$\boldsymbol{\gamma}^{(i)}_{(1\rightarrow2)} = \{\xi_i^2, \eta_i^2\} \tag{A.4}$$

instead of the set in Eq. (2.14). Similarly, to obtain cubic interpolation, the set

$$\boldsymbol{\gamma}^{(i)}_{(1\rightarrow3)} = \{\xi_i^2, \eta_i^2, \xi_i^3, \xi_i^2\eta_i, \xi_i\eta_i^2, \eta_i^3\} \tag{A.5}$$

may be an alternative to Eq. (2.15).

Nevertheless, two more sophisticated ways of solving the equations while still keeping all the terms are proposed in [91]. Although the linear dependencies are still in the equations, the set of equations is solved by special techniques, i.e.

- adding a small perturbation to the stiffness matrix and iteratively balance the original equations, or

- using the direct method of multifrontal sparse Gaussian elimination.

The reader is referred to [91, 95] for a more deeper investigation about the linear dependence problems in the PU-based finite element method.

Bibliography

[1] M. Ainsworth. The performance of Bank-Weiser's error estimator for quadrilateral finite elements. *Numerical Methods for Partial Differential Equations*, 10:609–623, 1994.

[2] M. Ainsworth. The influence and selection of subspaces for a posteriori error estimators. *Numerische Mathematik*, 73:399–418, 1996.

[3] M. Ainsworth and J. T. Oden. A unified approach to a posteriori error estimation based on element residual methods. *Numerische Mathematik*, 65:23–50, 1993.

[4] M. Ainsworth and J. T. Oden. *A posteriori error estimation in finite element analysis*. John Wiley and Sons, 2000.

[5] F. Armero and K. Garikipati. Recent advances in the analysis and numerical simulation of strain localization in inelastic solids. In D. R. J. Owen, E. Oñate, and E. Hinton, editors, *Computational Plasticity: Fundamentals and Applications, Proceeding of COMPLAS IV*, pages 547–561. Pineridge Press, Swansea, 1995.

[6] H. Askes. *Advanced spatial discretisation strategies for localised failure: mesh adapivity and meshless methods*. PhD thesis, Delft University of Technology, The Netherlands, 2000. (Downloadable from http://www.library.tudelft.nl/dissertations).

[7] H. Askes, E. Kuhl, and P. Steinmann. An ALE formulation based on spatial and material settings of continuum mechanics. part 2: Classification and applications. *Computer Methods in Applied Mechanics and Engineering*, 193:4223–4245, 2004.

[8] H. Askes and T. Pannachet. Letter to the editor: Negative penalty functions in the Element-Free Galerkin method. *Communications in Numerical Methods in Engineering*, 21:97–98, 2005.

[9] H. Askes and A. Rodríguez-Ferran. A combined *rh*-adaptive scheme based on domain subdivision – formulation and linear examples. *International Journal for Numerical Methods in Engineering*, 51:253–273, 2001.

[10] H. Askes, A. Rodríguez-Ferran, and A. Huerta. Adaptive analysis of yield line patterns in plates with the arbitrary Lagrangian-Eulerian method. *Computers and Structures*, 70:257–271, 1999.

[11] H. Askes and L. J. Sluys. Remeshing strategies for adaptive ALE analysis of strain localisation. *European Journal of Mechanics A/Solids*, 19:447–467, 2000.

[12] I. Babuška and C. Rheinboldt. A-posteriori error estimates for the finite element method. *International Journal for Numerical Methods in Engineering*, 12:1597–1615, 1978.

[13] I. Babuška, G. Caloz, and J. E. Osborn. Special finite element methods for a class of second order elliptic problems with rough coefficients. *SIAM Journal on Numerical Analysis*, 31(4):745–981, 1994.

[14] I. Babuška and J. M. Melenk. The partition of unity method. *International Journal for Numerical Methods in Engineering*, 40:727–758, 1997.

[15] M. J. Baines. *Moving finite elements*. Clarendon, Oxford, 1994.

[16] R. E. Bank and A. Weiser. Some a posteriori error estimators for elliptic partial differential equations. *Mathematics of Computation*, 44:283–301, 1985.

[17] G. I. Barenblatt. The mathematical theory of equilibrium cracks in brittle fracture. *Advances in Applied Mechanics*, 7:55–129, 1962.

[18] K. J. Bathe. *Finite element procedures*. Prentice-Hall, Englewood Cliffs, 1995.

[19] Z. P. Bažant and T. Belytschko aand T. P. Chang. Continuum theory for strain-softening. *Journal of Engineering Mechanics*, 110:1666–1692, 1984.

[20] Z. P. Bažant and G. Pijaudier-Cabot. Nonlocal continuum damage, localization instability and convergence. *Journal of Applied Mechanics*, 55:287–293, 1988.

[21] T. Belytschko and T. Black. Elastic crack growth in finite elements with minimal remeshing. *International Journal for Numerical Methods in Engineering*, 45:601–620, 1999.

[22] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming, and P. Krysl. Meshless methods: An overview and recent developments. *Computer Methods in Applied Mechanics and Engineering*, 139:3–47, 1996.

[23] G. Bugeda. A comparison between new adaptive remeshing strategies based on point wise stress error estimation and energy norm error estimation. *Communications in Numerical Methods in Engineering*, 18:469–482, 2002.

[24] P. Carnevali, R. B. Morris, Y. Tsuji, and G. Taylor. New basis functions and computational procedures for *p*-version finite element analysis. *International Journal for Numerical Methods in Engineering*, 36:3759–3779, 1993.

[25] C. Comi and U. Perego. Criteria for mesh refinement in nonlocal damage finite element analyses. *European Journal of Mechanics A/Solids*, 23:615–632, 2004.

[26] J. H. P. de Vree, W. A. M. Brekelmans, and M. A. J. van Gils. Comparison of nonlocal approaches in continuum damage mechanics. *Computers and Structures*, 55:581–588, 1995.

[27] L. Demkowicz. 2d *hp*-adaptive finite element package (2dhp90) version 2. Technical Report 02-06, Texas Institute for Computational and Applied Mathematics, University of Texas at Austin, USA, 2002.

[28] P. Díez, M. Arroyo, and A. Huerta. Adaptivity based on error estimation for viscoplastic softening materials. *Mechanics of Cohesive-Frictional Materials*, 5:87–112, 2000.

[29] P. Díez, J. J. Egozcue, and A. Huerta. A posteriori error estimation for standard finite element analysis. *Computer Methods in Applied Mechanics and Engineering*, 163:141–157, 1998.

[30] P. Díez and A. Huerta. A unified approach to remeshing strategies for finite element *h*-adaptivity. *Computer Methods in Applied Mechanics and Engineering*, 176:215–229, 1999.

[31] P. Díez, I. Morata, and A. Huerta. Goal-oriented adaptivity for shell structures: error estimation and remeshing criteria. *Revue Européenne des Éléments Finis*, 12:691–715, 2003.

[32] J. Dolbow, N. Moës, and T. Belytschko. Discontinuous enrichment in finite elements with a partition of unity method. *Finite Element Analysis and Design*, 36:235–260, 2000.

[33] C. A. Duarte, I. Babuška, and J. T. Oden. Generalized finite element methods for three-dimensional structural mechanics problems. *Computers and Structures*, 77:215–232, 2000.

[34] D. S. Dugdale. Yielding of steel sheets containing slits. *Journal of the Mechanics and Physics of Solids*, 8:100–108, 1960.

[35] N. B. Edgar and K. S. Surana. On the conditioning number and the selection criteria for *p*-version approximation functions. *Computers and Structures*, 60(4):521–530, 1996.

[36] W. J. Gordon. Blending function methods of bivariate and multivariate interpolation and approximation. *SIAM Journal on Numerical Analysis*, 8(1):158–177, 1971.

[37] A. A. Griffith. The theory of rupture. In *Proceedings of the First International Conference of Applied Mechanics*, pages 55–63, 1924.

[38] P. Heintz, F. Larsson, P. Hansbo, and K. Runesson. On error control and adaptivity for computing material forces in fracture mechanics. In *WCCM V - Fifth World Congress on Computational Mechanics*, Vienna, Australia, 2002. (Downloadable from http://wccm.tuwien.ac.at).

[39] P. Heintz and K. Samuelsson. On adaptive strategies and error control in fracture mechanics. *Computers and Structures*, 82:485–497, 2004.

[40] A. Hernández, J. Albizuri, M. .B. G. Ajuria, and M. V. Hormaza. An adaptive meshing automatic scheme based on the strain energy density function. *Engineering Computations*, 14:604–629, 1997.

[41] A. Hillerborg, M. Modeer, and P. E. Petersson. Analysis of crack formation and crack growth in concrete by means of fracture mechanics and finite elements. *Cement and Concrete Research*, 6:773–782, 1976.

[42] A. Huerta and P. Díez. Error estimation including pollution assessment for nonlinear finite element analysis. *Computer Methods in Applied Mechanics and Engineering*, 181:21–41, 2000.

[43] A. Huerta, A. Rodríguez-Ferran, P. Díez, and J. Sarrate. Adaptive finite element strategies based on error assessment. *International Journal for Numerical Methods in Engineering*, 46:1803–1818, 1999.

[44] T. J. R. Hughes. *The finite element method. Linear static and dynamic finite element analysis*. Dover publications, Inc., New York, 2000.

[45] G. R. Irwin. Analysis of stresses and strains near the end of a crack traversing a plate. *Journal of Applied Mechanics, Trans. ASME*, 24:361–364, 1957.

[46] Y. Krongauz and T. Belytschko. Enforcement of essential boundary conditions in meshkess approximations using finite elements. *Computer Methods in Applied Mechanics and Engineering*, 131:133–145, 1996.

[47] E. Kuhl, H. Askes, and P. Steinmann. An ALE formulation based on spatial and material settings of continuum mechanics. part 1: Generic hyperelastic formulation. *Computer Methods in Applied Mechanics and Engineering*, 193:4207–4222, 2004.

[48] P. Ladevèze and D. Leguillon. Error estimate procedure in the finite element method and applications. *SIAM Journal on Numerical Analysis*, 20:485–509, 1983.

[49] P. Ladevèze, N. Moës, and B. Douchin. Constitutive relation error estimator for (visco)plastic finite element analysis with softening. *Computer Methods in Applied Mechanics and Engineering*, 176:247–264, 1999.

[50] P. Lancaster and K. Salkauskas. Surfaces generated by moving least-squares methods. *Mathematics of Computation*, 37:141–158, 1981.

[51] F. Larsson, P. Hansbo, and K. Runesson. Strategies for goal-oriented a posteriori error measures in nonlinear elasticity. *International Journal for Numerical Methods in Engineering*, 55:879–894, 2002.

[52] L. Y. Li and P. Bettess. Notes on mesh optimal criteria in adaptive finite element computation. *Communications in Numerical Methods in Engineering*, 11:911–915, 1995.

[53] L. Y. Li, P. Bettess, J. W.Bull, T. Bond, and I. Applegarth. Theoretical formulations for adaptive finite element computations. *Communications in Numerical Methods in Engineering*, 11:857–868, 1995.

[54] X. Liu, C. K. Lee, and S. C. Fan. On using enriched cover function in the Partition-of-Unity method for singular boundary-value problem. *Computational Mechanics*, 29:212–225, 2002.

[55] J. Mazars and G. Pijaudier-Cabot. Continuum damage theory – application to concrete. *Journal of Engineering Mechanics*, 115:345–365, 1989.

[56] T. Meinders. *Developments in numerical simulations of the real-life drawing process*. PhD thesis, Universiteit Twente, The Netherlands, 2000.

[57] J. M. Melenk and I. Babuška. The partition of unity finite element method: Basic theory and applications. *Computer Methods in Applied Mechanics and Engineering*, 139:289–314, 1996.

[58] N. Moës, J. Dolbow, and T. Belytschko. A finite element method for crack growth without remeshing. *International Journal for Numerical Methods in Engineering*, 46:131–150, 1999.

[59] Y. X. Mukherjee and S. Mukherjee. On boundary conditions in the element-free Galerkin method. *Computational Mechanics*, 19:264–270, 1997.

[60] V. Nübel, A. Düster, and E. Rank. An *rp*-adaptive finite element method for the deformation theory of plasticity. *Computational Mechanics*, 2004. (Submitted).

[61] J. T. Oden and G. F. Carey. *Finite Elements: Mathematical Aspects. Volume 4*. Prentice-Hall, Englewood Cliffs, 1983.

[62] J. T. Oden, C. A. M. Duarte, and O. C. Zienkiewicz. A new cloud-based *hp* finite element method. *Computer Methods in Applied Mechanics and Engineering*, 153:117–126, 1998.

[63] J. Oliver. Modelling strong discontinuities in solid mechanics via strain softening constitutive equation. part 1: Fundamentals. part 2: Numerical simulation. *International Journal for Numerical Methods in Engineering*, 39:3575–3624, 1996.

[64] E. Oñate and G. Bugeda. A study of mesh optimality criteria in adaptive finite element analysis. *Engineering Computations*, 10:307–321, 1993.

[65] T. Pannachet and H. Askes. Some observations on the enforcement of constraint equations in the EFG method. *Communications in Numerical Methods in Engineering*, 16:819–830, 2000.

[66] T. Pannachet and H. Askes. The new hierarchical finite element method: performance analyses. In *Proceedings of the National Convention on Civil Engineering*, Khon Kaen, Thailand, 2002.

[67] T. Pannachet and H. Askes. Partition-of-unity based *p*-adaptivity and error estimation in the finite element method. In *WCCM V - Fifth World Congress on Computational Mechanics*, Vienna, Austria, 2002. (Downloadable from http://wccm.tuwien.ac.at).

[68] T. Pannachet and H. Askes. A residual-type error estimate for p-version finite element method. In *Proceedings of International Conference on Adaptive Modeling and Simulation (ADMOS 2003)*, Gothenburg, Sweden, 2003.

[69] B. Patzák and M. Jirásek. Adaptive resolution of localized damage in quasi-brittle materials. *Journal of Engineering Mechanics*, 130:720–732, 2004.

[70] R. H. J. Peerlings. *Enhanced damage modelling for fracture*. PhD thesis, Eindhoven University of Technology, The Netherlands, 1999. (Downloadable from http://www.mate.tue.nl).

[71] R. H. J. Peerlings, R. de Borst, W. A. M. Brekelmans, and J. H. P. de Vree. Gradient-enhanced damage for quasi-brittle materials. *International Journal for Numerical Methods in Engineering*, 39:3391–3403, 1996.

[72] G. Pijaudier-Cabot and Z. P. Bažant. Nonlocal damage theory. *Journal of Engineering Mechanics*, 113:1512–1533, 1987.

[73] G. Pijaudier-Cabot, L. Bodé, and A. Huerta. Arbitrary Lagrangian-Eulerian finite element analysis of strain localization in transient problem. *International Journal for Numerical Methods in Engineering*, 38:4171–4191, 1995.

[74] A. Promwungkwa. *Data structure and error estimation for an adaptive p-version finite element method in 2-D and 3-D solids*. PhD thesis, Virginia Tech, Blacksburg, USA, 1998.

[75] S. Prudhomme and J. T. Oden. On goal-oriented error estimation for elliptic problems: Application to the control of pointwise errors. *Computer Methods in Applied Mechanics and Engineering*, 176:313–331, 1999.

[76] J. R. Rice. Mathematical analysis in the mechanics of fracture. In H. Liebowitz, editor, *Fracture – An Advanced Treatise*, volume 2, pages 191–308, Academic Press, New York, 1968.

[77] J. R. Rice. A path independent integral and the approximate analysis of strain concentrations by notches and cracks. *Journal of Applied Mechanics, Trans. ASME*, 35:379–386, 1968.

[78] A. Rodríguez-Ferran and A. Huerta. Error estimation and adaptivity for non-local damage models. *International Journal of Solids and Structures*, 37:7501–7528, 2000.

[79] J. G. Rots. Smeared and discrete representations of localized fracture. *International Journal of Fracture*, 51:45–59, 1991.

[80] K. Runesson. Strategies for goal-oriented a posteriori error measures in nonlinear elasticity. In *15th Nordic Seminar on Computational Mechanics*, Aalborg, Denmark, 2002. (Downloadable from http://www.ime.auc.dk/nscm15).

[81] M. Rüter and E. Stein. Goal-oriented a posteriori error estimates in elastic fracture mechanics. In *WCCM V - Fifth World Congress on Computational Mechanics*, Vienna, Australia, 2002. (Downloadable from http://wccm.tuwien.ac.at).

[82] M. Rüter and E. Stein. Adaptive finite element analysis of crack propagation in elastic fracture mechanics based on averaging techniques. *Computational Material Science*, 31:247–257, 2004.

[83] M. Rüter, E. Stein, F. Larsson, P. Hansbo, and K. Runesson. Strategies for goal-oriented a posteriori error measures in nonlinear elasticity. In *2nd European Conference on Computational Mechanics (CDROM)*, Cracow, Poland, 2001.

[84] E. Schlangen. Fracture simulations of brittle heterogeneous materials. In *Engineering Mechanics, Proc. 10th Conf.*, pages 130–133, New York, 1995. ASCE.

[85] J. C. Simo, J. Oliver, and F. Armero. An analysis of strong discontinuities induced by strain-softening in rate-independent inelastic solids. *Computational Mechanics*, 12:277–296, 1993.

[86] A. Simone. Remarks on a gradient-enhanced damage model and its implementation. *Studi e Ricerche*, 21:243–282, 2000.

[87] A. Simone. *Continuous-Discontinuous Modelling of fracture*. PhD thesis, Delft University of Technology, The Netherlands, 2003. (Downloadable http://www.library.tudelft.nl/dissertations).

[88] A. Simone, H. Askes, R. H. J. Peerlings, and L. J. Sluys. Interpolation requirements for implicit gradient-enhanced continuum damage models. *Communications in Numerical Methods in Engineering*, 19:563–572, 2003. (See also the corrigenda: *Communications in Numerical Methods in Engineering*, 20:163–165).

[89] L. J. Sluys. *Wave propagation, localisation and dispersion in softening solids*. PhD thesis, Delft University of Technology, The Netherlands, 1992. (Downloadable from http://www.library.tudelft.nl/dissertations).

[90] L. J. Sluys, M. Cauvern, and R. de Borst. Discretization influence in strain-softening problems. *Engineering Computations*, 12:209–228, 1995.

[91] T. Strouboulis, I. Babuška, and K. Copps. The design and analysis of the generalized finite element method. *Computer Methods in Applied Mechanics and Engineering*, 181:43–69, 2000.

[92] T. Strouboulis, K. Copps, and I. Babuška. The generalized finite element method: an example of its implementation and illustration of its performance. *International Journal for Numerical Methods in Engineering*, 47:1401–1417, 2000.

[93] B. Szabó and I. Babuška. *Finite element analysis*. John Wiley & Sons, Inc., New York, 1991.

[94] R. L. Taylor, O. C. Zienkiewicz, and E. Oñate. A hierarchical finite element method based on partition of unity. *Computer Methods in Applied Mechanics and Engineering*, 152:73–84, 1998.

[95] R. Tian, G. Yagawa, and H. Terasaka. Linear dependence problems of partition of unity-based generalized FEMs. *Computer Methods in Applied Mechanics and Engineering*, 195:4768–4782, 2006.

[96] S. P. Timoshenko and J. N. Goodier. *Theory of Elasticity*. Mc Graw-Hill, New York, 3rd edition, 1970.

[97] R. Verfürth. *A review of a posteriori error estimation and adaptive mesh refinement techniques*. Wiley-Teubner, 1996.

[98] S. M. Vijayakar, H. R. Busby, and D. R. Houser. Finite element analysis of quasi-prismatic bodies using Chebychev polynomials. *International Journal for Numerical Methods in Engineering*, 24:1461–1477, 1987.

[99] G. N. Wells. *Discontinuous modelling of strain localisation and failure*. PhD thesis, Delft University of Technology, The Netherlands, 2001. (Downloadable from http://www.library.tudelft.nl/dissertations).

[100] G. N. Wells and L. J. Sluys. A new method for modelling cohesive cracks using finite elements. *International Journal for Numerical Methods in Engineering*, 50:2667–2682, 2001.

[101] G. N. Wells and L. J. Sluys. Three-dimensional embedded discontinuity model for brittle fracture. *International Journal of Solids and Structures*, 38:897–913, 2001.

[102] G. N. Wells and L. J. Sluys. A p-adaptive scheme for overcoming volumetric locking during plastic flow. *Computer Methods in Applied Mechanics and Engineering*, 191:3153–3164, 2002.

[103] J. Z. Zhu and Z. Zhang. The relationship of some a posteriori estimators. *Computer Methods in Applied Mechanics and Engineering*, 176:463–475, 1999.

[104] G. Zi and T. Belytschko. New crack-tip elements for XFEM and applications to cohesive cracks. *International Journal for Numerical Methods in Engineering*, 57:2221–2240, 2003.

[105] O. C. Zienkiewicz and R. L. Taylor. *The finite element method*, volume 1: Basic formulation and linear programs. McGraw-Hill, fourth edition, 1994.

[106] 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.

[107] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery (SPR) and adaptive finite element refinement. *Computer Methods in Applied Mechanics and Engineering*, 101:207–224, 1992.

Summary

**Error estimation and adaptive spatial discretisation
for quasi-brittle failure**

by T. Pannachet

The accuracy of a finite element solution depends greatly on the spatial discretisation of the problem domain. A more refined/enriched discretisation provides better accuracy but also requires more computational cost. Adaptive discretisation is a solution to such a dilemma. The spatial discretisation can be varied to suit the need in the problem domain. In modelling quasi-brittle failure the process zone, where failure takes place, can be very small. Thus, it is cost-effective to apply an adaptive spatial discretisation scheme. In this research, two failure models for quasi-brittle materials have been put under investigation. The adaptive scheme is driven by the mathematically formulated error estimate in the form of either the energy norm or the quantity of interest. The adaptive scheme is activated during the failure process.

The continuous model represents failure via material strength degradation. A gradient enhanced damage model has been chosen for this study. Besides the conventional critical regions such as at the supports or the area containing a stress singularity, the error is found to concentrate in the area of high strain gradients. Three adaptive techniques have been applied. The results have shown that, although p-adaptivity can greatly improve the discretisation, its performance relies on the base mesh. Obviously, h-adaptivity has shown better capabilities to capture the localisation zone. As an inexpensive alternative, an r-adaptive scheme, based on the weighted Laplace-smoothing techniques and driven by the refinement factor, has shown a good performance.

The discontinuous model represents failure via the introduction of a geometrical discontinuity. In this study, the cohesive zone concept is modelled via the use of the partition-of-unity property of the finite element interpolation. By this approach, the crack line can pass through the elements without any restriction to the underlying mesh. Despite such feature, it has been found that a sufficiently fine mesh discretisation still needs to be ensured in order to obtain a correct crack path and mechanical response. The uniform mesh tests have shown that h-adaptivity is the best choice in improving the finite element results. However, only the p-adaptive

scheme has been examined due to its implementational simplicity, in comparison to *h*-adaptivity, to be inserted in this model. The results have shown that, if considering only increasing the polynomial degree, the *p*-approach can greatly improve the results, whereas decreasing the polynomial degree was found to be unsuccessful.

Samenvatting

**Error estimation and adaptive spatial discretisation
for quasi-brittle failure**

door T. Pannachet

De nauwkeurigheid van een eindige-elementenoplossing hangt sterk af van de ruimtelijke discretisatie van het probleemdomein. Een verfijnde/verrijkte discretisatie verschaft een grotere nauwkeurigheid maar vereist ook hogere berekeningskosten. Adaptieve discretisatie is een oplossing voor dit dilemma. De ruimtelijke discretisatie kan worden gevarieerd naarmate dit nodig is in het probleemdomein. In de modellering van quasi-bros bezwijken kan de proceszone, de zone waar het bezwijken plaatsvindt, zeer klein zijn. Het is dus efficiënt een adaptieve ruimtelijke discretisatie toe te passen. In dit onderzoek zijn twee bezwijkmodellen voor quasi-brosse materialen onderzocht. Het adaptieve schema wordt gestuurd door een wiskundig geformuleerde foutschatter in de vorm van ofwel de energienorm ofwel een specifieke grootheid. Het adaptieve schema wordt geactiveerd tijdens het bezwijkproces.

Het continuümmodel simuleert bezwijken door degradatie van de materiaalsterkte. Een gradiënt-verrijkt schademodel is gekozen voor deze studie. Naast de gebruikelijke kritieke zones, zoals bij de opleggingen of het gebied met de spanningsingulariteit, zijn foutconcentraties gevonden in gebieden met grote rekgradiënten. Drie adaptieve technieken zijn toegepast. Hoewel p-adaptiviteit de discretisatie sterk kan verbeteren, laten de resultaten zien dat de prestatie afhangt van het elementennet. H-adaptiviteit laat duidelijk zien dat het de lokalisatiezone beter kan beschrijven. Als goedkoop alternatief heeft r-adaptatie, gebaseerd op gewogen Laplace-uitsmeringsmethoden en gestuurd door de verfijningsfactor, goede prestaties laten zien.

Het discontinue model beschrijft bezwijken door een geometrische discontinuïteit te introduceren. In deze studie is het cohesieve-zoneconcept gemodelleerd met de *Partition-of-Unity*-eigenschap van de eindige-elementeninterpolatie. In deze aanpak kan de scheur elementen passeren zonder beperkingen van het onderliggende elementennet. Desondanks is gevonden dat een voldoende fijn elementennet nodig is om een correct scheurpad en mechanische repons te verkrij-

gen. De voorbeelden met een uniform elementennet tonen aan dat *h*-adaptiviteit de beste keuze is om de eindige-elementenresultaten te verbeteren. De studie beperkt zich echter tot een *p*-adaptivief schema vanwege de eenvoud van implementatie in dit model, vergeleken met *h*-adaptiviteit. De resultaten laten verder zien dat de *p*-aanpak de resultaten sterk kan verbeteren als alleen verhoging van de polynoom-graad gebruikt wordt, terwijl verlaging van de polynoomgraad zonder succes is gebleken.

Propositions/Stellingen

1. As long as computer resources are limited, an error assessment should be applied in a finite element analysis to guarantee that the solution is not just 'a junk'.

   Zolang mogelijkheden van de computer beperkt zijn, zal een foutschatting moeten worden gebruikt in een eindige-elementenberekening om te garanderen dat de oplossing niet puur willekeurig is.

2. There are two main reasons why a residual-type error estimation can not give an exact discretisation error. The first relates to the introduction of the reference mesh. The second relates to the introduction of the local scheme.

   Er zijn twee hoofdredenen waarom de residu-type foutschatter geen exacte discretisatie-fout oplevert. De eerste wordt veroorzaakt door het gebruik van een referentie 'mesh'. De tweede refereert aan het gebruik van een lokaal schema.

3. Although Partition-of-Unity based discontinuity models do not suffer from mesh alignment, sufficient mesh discretisation still needs to be ensured.

   Hoewel het 'Partition-of-Unity' gebaseerde discontinuïteitenmodel niet gevoelig is voor richting-afhankelijkheid van de 'mesh', is toch een voldoende fijne discretisatie nodig.

4. The goal you try to achieve at the beginning may not be the same goal you try to achieve at the end. It depends very much on the information you collect along the way.

   Het doel dat je wilt bereiken aan het begin, is niet noodzakelijk hetzelfde doel dat je probeert te halen aan het eind. Dit is afhankelijk van de informatie die onderweg wordt verzameld.

5. The style of working of a person relates to his/her culture.

   De werkstijl van een persoon is gekoppeld aan zijn/haar cultuur.

6. The more you learn, the smarter you are, but the more stupid you feel.

   Hoe meer je leert, hoe slimmer je wordt, maar hoe dommer je jezelf voelt.

7. Good researchers listen to other people's opinions.

   Goede onderzoekers luisteren naar de mening van anderen.

8. The relation between work progress and tension may be described by a load-displacement softening curve. For the pre-peak regime holds, the more tension you feel, the higher your productivity. However, beyond your mental strength, it is reversed. To which extent your work progress is reduced depends on your mental toughness.

   De relatie tussen voortgang van werk en druk kan worden beschreven met een last-verplaatsingsdiagram met 'softening'. Voor de stijgende tak geldt dat meer druk leidt tot een hogere productiviteit. Echter, voorbij de mentale sterkte is dit juist andersom. De reductie in voortgang van het werk wordt bepaald door de mentale weerbaarheid.

9. Body and mind are important in one's life. However, the power that drives you to do things comes from your mind. With a healthy mind, you can do almost anything.

   Lichaam en geest zijn belangrijk in het leven. Echter, de kracht om dingen te doen komt van de geest. Met een gezonde geest kun je bijna alles doen.

These propositions are considered opposable and defendable and as such have been approved by the supervisors, Prof. dr. ir L. J. Sluys and Prof. dr. ir. H. Askes.

Deze stellingen worden opponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotoren, Prof. dr. ir L. J. Sluys and Prof. dr. ir. H. Askes.

Acknowledgement

<div align="right">

T. Pannachet
Delft, October 2006

</div>

[†]Department of Civil and Structural Engineering, University of Sheffield, the United Kingdom.
[‡]Departament de Matemàtica Aplicada III, Universitat Politècnica de Catalunya, Spain.

Curriculum vitae

Name:	Tanyada Pannachet
Date of birth:	May 18,1976
Place of birth:	Khon Kaen, Thailand

Jun 1993 – Jan 1997	Undergraduate student for the degree Bachelor of Engineering (BEng), Department of Civil Engineering, Khon Kaen university, Thailand.
Feb 1997 – Aug 1998	Lecturer, Department of Civil Engineering, Khon Kaen university, Thailand.
Sep 1998 – Aug 2000	Graduate student for the degree Master of Engineering (MEng), Section of Structural Engineering, School of Civil Engineering, Asian Institute of Technology, Thailand.
	Thesis title: Large deformation analyses of 2D elastic solids by the meshless local Petrov-Galerkin method.
Nov 1999 – Feb 2000	Exchange student to Section of Computational Mechanics, Faculty of Civil Engineering and Geosciences, Delft university of Technology, The Netherlands.
Sep 2000 – Oct 2006	Doctoral student for the degree Doctor of Philosophy (PhD), Section of Computational Mechanics, Faculty of Civil Engineering and Geosciences, Delft university of Technology, The Netherlands.
Mar 2002 – Apr 2002	Exchange student to Laboratori de Càlcul Numèric, Departament de Matemàtica Aplicada III, Universitat Politècnica de Catalunya, Barcelona, Spain.
Nov 2006 –	Lecturer, Department of Civil Engineering, Khon Kaen university, Thailand.