Journal of Geodynamics 70 (2013) 1–20

Contents lists available at SciVerse ScienceDirect

Journal of Geodynamics

journal homepage: http://www.elsevier.com/locate/jog

Review

Overview of adaptive finite element analysis in computational geodynamics

D.A. May[a,*], W.P. Schellart[b], L. Moresi[b,c]

[a] Institute of Geophysics, Department of Earth Sciences, ETH Zürich, Zürich, Switzerland
[b] School of Geosciences, Monash University, Victoria, Australia
[c] School of Mathematical Sciences, Monash University, Victoria, Australia

ARTICLE INFO

Article history:
Received 9 August 2012
Received in revised form 17 April 2013
Accepted 19 April 2013
Available online 7 May 2013

Keywords:
Computational geodynamics
Finite element method
Adaptivity
Numerics

ABSTRACT

The use of numerical models to develop insight and intuition into the dynamics of the Earth over geological time scales is a firmly established practice in the geodynamics community. As our depth of understanding grows, and hand-in-hand with improvements in analytical techniques and higher resolution remote sensing of the physical structure and state of the Earth, there is a continual need to develop more efficient, accurate and reliable numerical techniques. This is necessary to ensure that we can meet the challenge of generating robust conclusions, interpretations and predictions from improved observations.

In adaptive numerical methods, the desire is generally to maximise the *quality* of the numerical solution for a given amount of *computational effort*. Neither of these terms has a unique, universal definition, but typically there is a trade off between the number of unknowns we can calculate to obtain a more accurate representation of the Earth, and the resources (time and computational memory) required to compute them. In the engineering community, this topic has been extensively examined using the adaptive finite element (AFE) method. Recently, the applicability of this technique to geodynamic processes has started to be explored. In this review we report on the current status and usage of spatially adaptive finite element analysis in the field of geodynamics.

The objective of this review is to provide a brief introduction to the area of spatially adaptive finite analysis, including a summary of different techniques to define spatial adaptation and of different approaches to guide the adaptive process in order to control the discretisation error inherent within the numerical solution. An overview of the current state of the art in adaptive modelling in geodynamics is provided, together with a discussion pertaining to the issues related to using adaptive analysis techniques and perspectives for future research in this area. Additionally, we also provide a summary of available open source software relevant to AFE analysis in geodynamics with the intent to provide interested practitioners of computational geodynamics with a starting point for "first hand" geodynamic experimentation using AFE analysis.

© 2013 Elsevier Ltd. All rights reserved.

Contents

1. Introduction 2
   1.1. Computational geodynamics and the need for adaptivity 2
      1.1.1. The issue of spatial resolution in numerical models 2
2. Spatial adaptivity in finite elements 4
   2.1. Dynamics of creeping flow 4
   2.2. Overview of adaptive FE analysis 5
   2.3. Types of adaptivity 6
   2.4. Error estimators 7
      2.4.1. Residual based estimators 7
      2.4.2. Recovery based estimators 8
      2.4.3. Goal oriented estimators 9

* Corresponding author.
E-mail addresses: dave.may@erdw.ethz.ch, dave.mayhem23@gmail.com (D.A. May).

0264-3707/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jog.2013.04.002

2.5. Solvers for systems of linear equations 9
3. Adaptivity in geodynamics – the state of the art 11
3.1. Arbitrary Lagrangian–Eulerian (ALE) methods 11
3.2. Unstructured *h*-adaptivity 11
3.3. Structured *h*-adaptivity 13
4. Perspectives and future directions 16
5. Summary 17
Acknowledgments 17
Appendix A. Software availability 17
References 18

1. Introduction

1.1. Computational geodynamics and the need for adaptivity

Developing insight of the dynamics of the Earth over geological time scales via numerical models has a history in the Earth sciences dating back to the early 1970s (e.g., Minear and Toksöz, 1970; Torrance and Turcotte, 1971; Turcotte et al., 1973; Richter, 1973; Mckenzie et al., 1974; Parmentier et al., 1975; Houston and De Bremaecker, 1975).

Mathematical models to study geodynamic processes are complementary to analogue (laboratory) models, but they also have several advantages. For instance, numerical models permit tighter control to be placed upon the material properties and the rheology of the material used. They also permit access to a wider range of material properties, rheological behaviour and boundary conditions. Numerical models permit a more quantitative analysis of the model results (i.e. post processing) to be performed, enabling a quantitatively richer understanding of the physical processes to be developed. And finally, numerical models yield results which are more readily reproducible, in particular when using the same modelling code. However, we note that model reproducibility between different numerical codes, is of comparable quality to the reproducibility between laboratory experiments originating from different laboratories (e.g., Schreurs et al., 2006; Buiter et al., 2006).

It is helpful at this point to focus on one specific geodynamic application and demonstrate the evolution of numerical techniques and the manner in which the problems of interest evolve along with the sophistication of the models. Within geodynamics, the study of mantle convection has consistently been driven by advances in computation and has also produced significant innovation in numerical methods.

Early numerical models of mantle convection were two-dimensional, either planar or axisymmetric (e.g., Richter, 1973; Moore and Weiss, 1973; Mckenzie et al., 1974; Houston and De Bremaecker, 1975; Parmentier and Turcotte, 1978; Lux et al., 1979; Schubert and Zebib, 1980). Whilst two-dimensional models provide insight into mantle convection, they are obviously unable to simulate the convective planforms of global mantle convection. In Fig. 1 we provide images from a range of different mantle convection simulations illustrating the advances made in 30 years of research in this field.

In three-dimensional simulations, several practical considerations have to be overcome. These consist of addressing the usage of computational memory (RAM) and the computational time required by the numerical method. Developing numerical models of mantle convection in three-dimensions, (spherical or Cartesian geometry) pushed the frontiers of high performance computing in an attempt to seek a higher spatial resolution. Many of the earliest three-dimensional mantle convection simulations utilised spectral methods (e.g., Young, 1974; Glatzmaier, 1988; Bercovici et al., 1989; Tackley et al., 1993; Zhang and Christensen, 1993; Zhang and Yuen, 1995; Machetel et al., 1995; Harder and Christensen, 1996). Such methods are especially attractive for problems involving constant viscosity or radial viscosity profiles, however the inclusion of a laterally varying viscosity structure significantly complicates the methodology and increases the computational cost (e.g., Christensen and Harder, 1991; Čadek et al., 1992; Zhang and Christensen, 1993; Forte and Peltier, 1994; Tromp and Mitrovica, 1999).

The first three-dimensional, spherical model of mantle convection simulations was presented by Baumgardner (1985) using the code Terra. The methodology utilised in Terra consisted of a finite element formulation and used a multi-grid solver to obtain a solution to the underlying mantle dynamics equations. The methodology represents an important step in addressing the major limitations of the spectral methods as the methodology; (i) can make effective usage of parallel computing (Bunge and Baumgardner, 1995; Bunge et al., 1997); (ii) the memory requirements and computational time grow linearly with respect to increasing grid resolution (due to the usage of multi-grid); (iii) permits both radial, azimuth variations in viscosity and furthermore enables fully non-linear rheologies to be used. Following the development of Terra, finite volume formulations and staggered grid finite difference methods (Ratcliff et al., 1995; Yoshida et al., 1999; Hernlund and Tackley, 2003; Yoshida and Kageyama, 2004; Harder and Hansen, 2005; Stemmer et al., 2006; Tackley, 2008b; Hüttig and Stemmer, 2008) and the finite element method (Zhong et al., 2000; Tabata and Suzuki, 2000; Stadler et al., 2010) were applied to study convection in a spherical geometry.

1.1.1. The issue of spatial resolution in numerical models

One of the important considerations to be made when performing numerical simulations is the selection of the numerical spatial resolution used. For finite difference and linear finite element methods, this choice relates to the grid spacing, or size of the element, which we will denote via h. The quality of the numerical solution, or the discretisation error is strongly related to h. A numerical solution will accurately resolve a physical processes if "features of interest" are able to be represented over the chosen discretisation length scale. For example, a thermal field with a linear variation in space can be exactly represented via a single linear finite element, however a thermal field possessing sharp thermal gradients will require multiple linear elements to be accurately represented. In the context of mantle convection, one feature of interest to resolve is the thermal boundary layer. In a simple isoviscous, bottom heated experiment, the boundary layer thickness l scales as $l \sim R_a{}^{-1/3}$, where the Rayleigh number R_a defines the ratio between the buoyancy force driving flow and the combination of both viscous and thermal dissipation: i.e. it essentially characterizes the vigour of convection within the system. To resolve the dynamics of convection with a numerical simulation, it is thus required that $h < l$. We note that inclusion of other physical processes may introduce additional length scales, for example the introduction of a strain-rate dependent rheology, plate boundaries or a representation of melt-migration and fluid percolation.

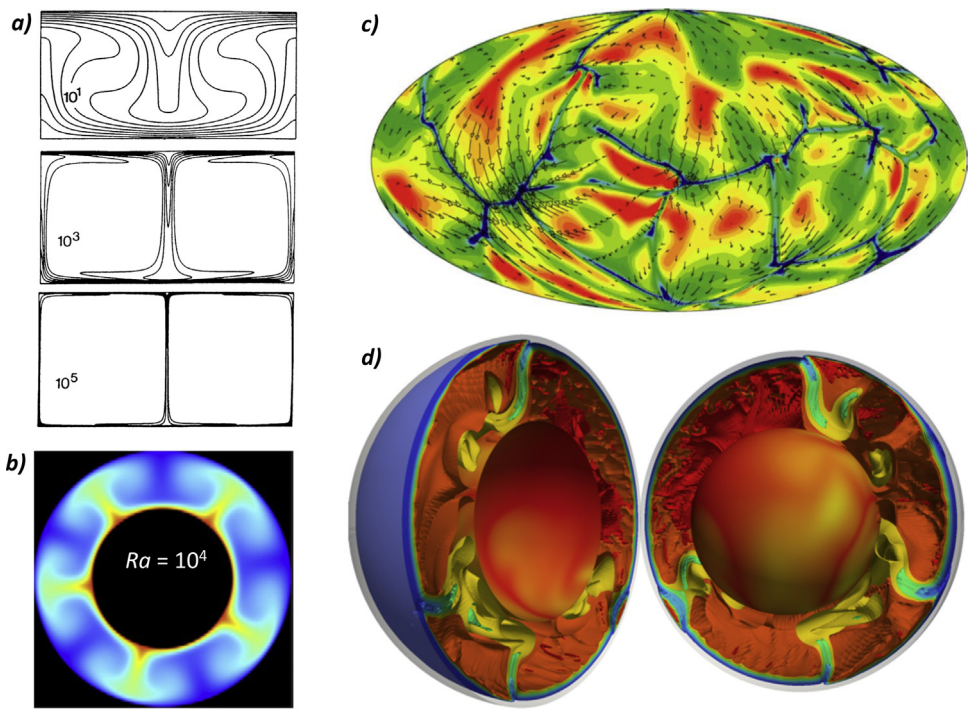

Fig. 1. Examples of numerical simulations used to study mantle convection. (a) Temperature isocontours for isoviscous 2D Cartesian models with increasing Rayleigh number (*Ra*), where inset numbers indicate the ratio between *Ra* and the critical Rayleigh number defining the onset of convection (Jarvis, 1984); (b) temperature field for isoviscous convection in the annulus (Hernlund and Tackley, 2008); (c) temperature and velocity vectors at 135 km depth obtained form a full spherical convective system (Walzer et al., 2004); (d) temperature isosurfaces for full spherical convection with a free surface (taken from the study associated with Crameri et al. (2012)).

In models of convection, a thermal boundary layer will develop along the boundaries of the model domain, and also along the surface of hot, buoyant rising plumes and cold, dense down-wellings. Consider an idealised model of the Earth which is isoviscous and possesses a Rayleigh number of 10^6 (Schubert et al., 2001), which we can argue will result in boundary layers with a thickness of the order ~10–50 km. A global model of the Earth's mantle, with a mesh employing uniform grid spacing in the range of ~50–1 km, would require $\sim O(200^3) \rightarrow O(10{,}000^3)$ grid cells. With the continued increase of the speed of computer hardware, and the widespread availability of massively parallel computational resources, practitioners of computational geodynamics naturally become more ambitious with their modelling objectives. The most common objective is to develop numerical representations at a regional scale or global (whole Earth) scale with a spatial resolution which captures all the length scales of the "relevant" physics. Twenty-five years since the first full spherical mantle convection simulations, the development of new mathematical techniques, implicit non-linear solvers and computational algorithms suitable for studying the dynamics of global mantle convection is still an active area of research (Kameyama and Sato, 2008; Tackley, 2008b; Stadler et al., 2010).

Two important observations should be emphasised in the context of thermal convection: the gradients requiring high spatial resolution are time dependent, and they are localised features restricted to a small faction of the model domain. The principal of spatially adaptive finite element analysis is to employ meshes which possess different sized elements throughout the domain, and that the size of the elements is dynamically modified to capture localised, temporally varying features of the numerical solution which require higher spatial resolution; e.g. in regions where there are sharp thermal gradients. The use of such a technique is often mandatory from a practical point of view, as employing meshes with a fixed size resolution over the entire domain yields numerical models which either cannot fit within the memory of existing computer hardware, or require too much computer time to obtain a solution due to the large number of unknowns generated from the uniformly spaced mesh.

The preceding discussion largely focused on mantle dynamics as a prototype geodynamic example to introduce and motivate the need for spatially adaptive methods. However, this class of numerical methods has wide applicability in many areas of geodynamics. To illustrate this point further, we consider the modern view of plate tectonics theory and some of the observables and conceptual ideas associated with tectonic processes. The two most striking observables which illustrate the multi-scale nature of tectonics are the presence of plate boundaries (regions of highly localised strain-rate) and topographic variations within the continents and oceans. Understanding the processes which give rise to the present day plate boundary geometry, processes associated with mountain formation are fundamentally coupled to large scale deformation within the interior of the Earth. Understanding surface expressions such as continental arc magmatism at convergent plate boundaries is a complex process which involves the coupling between the dynamics of mantle flow, the subduction of oceanic plates and the generation of partial molten mantle material on the leading edge of the subducting plate and the subsequent percolation of melt upwards through the continental crust. Such physical processes occur over a multitude of different length and time scales. In considering the evolution of chemical and thermal heterogeneities within the interior of the Earth, for instance, those associated with subduction initiation and the ultimate fate of subducted slabs over billion year time scales, is indeed another multi-scale process linking processes occurring deep in the mantle with surface observables. To summarise this short discussion, in Fig. 2 we re-emphasise the multi-scale nature of modern day tectonics with several images of surface observables and geological interpretations related to our current understanding of the Earth.

The outline of this review is as follows. In Section 2 we describe the general framework for spatially adaptive finite element (AFE)

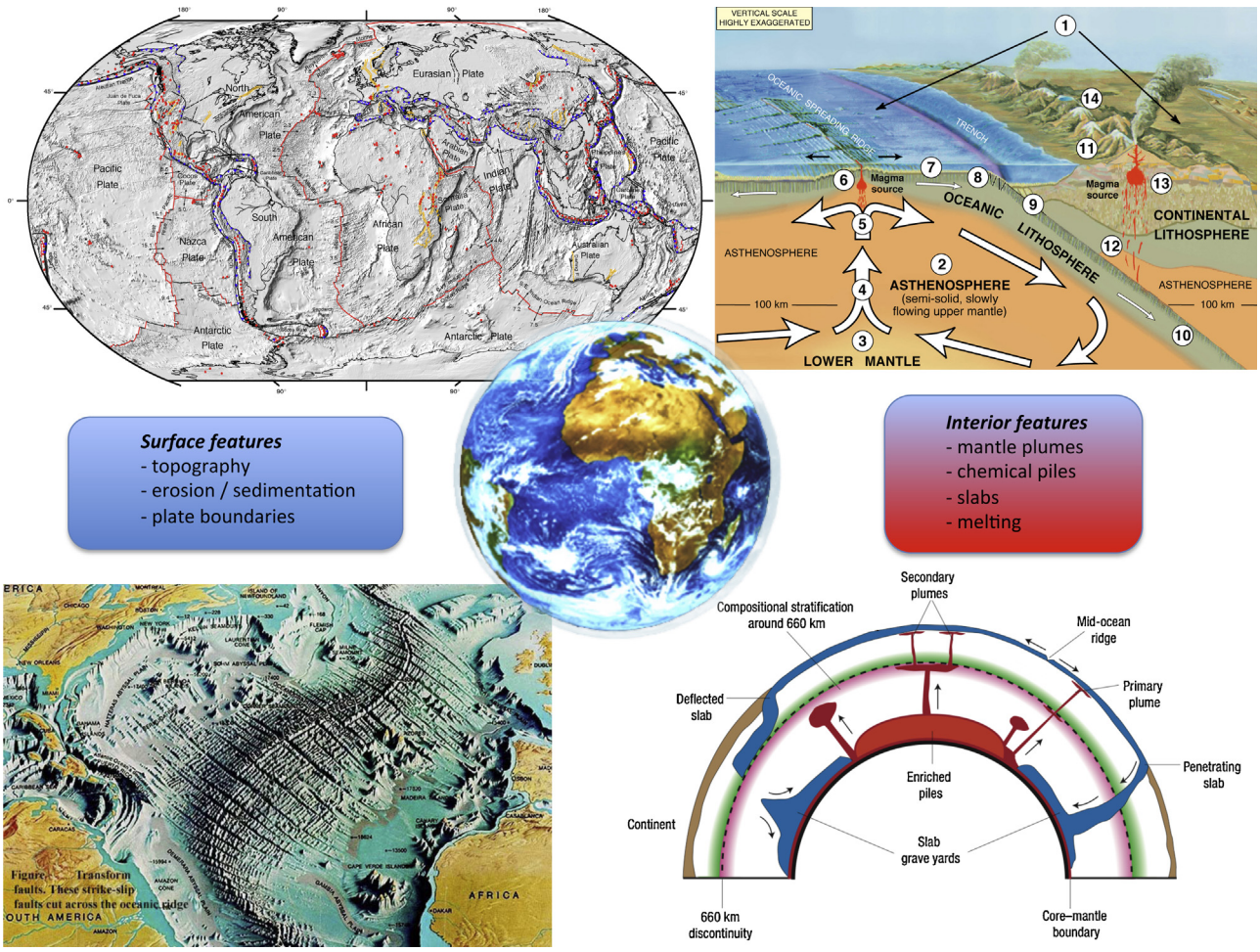

Fig. 2. Present day observations (left panels) and conceptual ideas associated with modern tectonics (right panels). Identified in these images are "features of interest" which may be relevant to consider and thus numerically resolve in computational geodynamic simulations. See text for further details. Figure modified after http://denali.gsfc.nasa.gov/dtam (upper left), http://geology.uprm.edu/Morelock (lower left), http://volcanoes.usgs.gov (upper right) and Tackley (2008a) (lower right).

analysis. Specifically, this entails a statement of the relevant governing equations for many geodynamic processes, an overview of different spatially adaptive strategies, a summary of different approaches to estimate errors and a short summary of different methods to solve systems of linear equations arising from the finite element (FE) discretisation. This summary of adaptive FE analysis is general and applicable to problems in applied mathematics and engineering as well as to geodynamics. In Section 3 we review the current state of adaptive FE which has been explored within the geodynamics community. In Section 4 we discuss some future directions which should be explored in the geodynamics community with respect to AFE analysis and in Section 5 we provide a short summary. In Appendix A, we list several open source adaptive finite element software packages and libraries with the intent to provide interested practitioners of computational geodynamics with a starting point for "first hand" geodynamic experimentation using AFE analysis.

2. Spatial adaptivity in finite elements

2.1. Dynamics of creeping flow

To describe the different components within adaptivity finite element analysis in geodynamics, we will consider the governing equations which are appropriate to describe the dynamics of rocks over million year time spans. The behaviour of such materials is akin to that of a very viscous, slowly moving (or creeping), nearly incompressible fluid. The dynamics of very viscous fluids is described by the steady Stokes equations which represent a balance of viscous stresses and pressure with buoyancy forces. The effects of elastic stresses are generally negligible over geologically relevant time and length scales, but, even where they play a role, can be incorporated with minor modifications to the Stokes equation (e.g. Gerya, 2009) which do not change the nature of the following discussion.

The momentum balance over a given domain Ω under such conditions is:

$$\nabla \cdot (\eta[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T]) - \nabla p = \rho g \boldsymbol{k}, \tag{1}$$

where $\boldsymbol{u}$, p represent the fluid velocity and pressure, η the dynamic viscosity, ρ the density, g the gravity field and $\boldsymbol{k}$ is the unit vector pointing in the direction which gravity acts.

The general evolution equation for the conservation of mass is given by

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = 0. \tag{2}$$

For studying geodynamic processes, various simplifications of Eq. (2) are typically employed. At depth, pressure p variations within the Earth are large (order of GPa's) and dominated by the hydrostatic pressure. Such high pressure will lead to significant compression of the rocks. The first-order effects of compressibility and phase changes may be introduced by assuming that the density–pressure relationship, taken at a reference temperature T_0,

varies only in the direction of gravity $\boldsymbol{k}$, and is independent of time, i.e. $\rho(\boldsymbol{x}, t) = \rho_0(\boldsymbol{k})$. Thus Eq. (2) reduces to

$$\nabla \cdot (\rho_0 \boldsymbol{u}) = 0. \tag{3}$$

Furthermore, density variations of the rock due to thermal expansion are assumed to be small. This is the *Boussinesq approximation* (Schubert et al., 2001) and assumes that the buoyancy force driving fluid motion is associated with small perturbations in fluid density ρ, relative to some reference density ρ_0 measured at a reference temperature T_0, which is induced by temperature variations T within the fluid, i.e.

$$\rho := \rho_0(\boldsymbol{k}) - \rho_0(\boldsymbol{k})\alpha(T - T_0), \tag{4}$$

where α is the coefficient of thermal expansivity. When inserted into Eq. (3), the contribution arising from the term $\rho_0(\boldsymbol{k})\alpha(T - T_0)$ is regarded as negligible and thus is not included in the continuity equation. If the depth dependence is removed from ρ and we assume that all spatial variations in density are associated with compositional differences which move with the fluid velocity $\boldsymbol{u}$, we obtain the incompressible continuity equation:

$$\nabla \cdot \boldsymbol{u} = 0. \tag{5}$$

For geodynamic processes, the dynamics of the fluid motion and the rheological behaviour of the rock is inherently related to temperature. Thus inclusion of heat transport is a fundamental component of many geodynamic processes. The conservation equation for thermal energy is given by

$$\rho C_p \left(\frac{\partial T}{\partial t} + \boldsymbol{u} \cdot \nabla T \right) = \nabla \cdot (k \nabla T) + Q, \tag{6}$$

where T is the temperature, k is the thermal conductivity, C_p the specific heat at constant pressure and Q is a source of internal heating (e.g. radiogenic, viscous dissipation, adiabatic heating/cooling).

The dynamics of mantle convection are frequently represented using Eqs. (1) and (5) together with the transport of thermal energy Eq. (6). This set of partial differential equations represents a prototype evolution equation for numerous geodynamic processes. It should be noted that time dependence does not explicitly appear in the momentum balance for the fluid (see Eq. (1)), rather it is introduced via the temperature dependence of the forcing term in the momentum equation and the temporal evolution of T, via the energy transport Eq. (6).

2.2. Overview of adaptive FE analysis

To describe the essential features of adaptive finite element analysis, we will focus our discussion around the mantle dynamics equations given by Eqs. (1), (5), and (6). In order to use the FE method, we are first required to partition the problem domain Ω into a set of non-overlapping elements. We will denote the partitioned domain via $\mathcal{T}^0$. To define the discrete representation of the mantle dynamics equations, we must specify how the discrete fields $\boldsymbol{u}_h$, p_h, T_h will vary across each element. In the FE method, polynomial basis functions (or shape functions) are used to specify how the discrete field varies within each element. For simplicity we will assume that the same set of elements in $\mathcal{T}^0$ are used to represent all fields, furthermore we will also assume that $\boldsymbol{u}_h$ and T_h will be conforming finite element spaces. This implies that these fields are continuous across element edges and faces. Readers unfamiliar with the finite element method, and in particular its application to solving Stokes equations are referred to Hughes (1987), Zienkiewicz and Taylor (2000), and Elman et al. (2005).

The selection of basis functions suitable for representing the velocity and pressure within the Stokes system requires consideration. In order that the discrete solution for $(\boldsymbol{u}_h, p_h)$ converges to the exact solution and is devoid of unphysical artifacts, such as pressure oscillations, the basis functions used for the $(\boldsymbol{u}_h, p_h)$ pair must satisfy a stability criterion known as the Ladyzanskya Babuska Brezzi (LBB) condition (also known as the inf-sup condition) (Brezzi and Fortin, 1991; Elman et al., 2005). In addition, stable and accurate treatment of the advective term $\boldsymbol{u} \cdot \nabla T$ in the energy equation requires special treatment, e.g. via a form of upwinding (Brooks and Hughes, 1982). We refer readers to Gresho and Sani (1998), Biswas et al. (1994), Ewing and Wang (2001), Hughes et al. (2007), and John and Schmeyer (2008) for discussions and comparisons of finite element based techniques to solve the advection diffusion equation.

The premise of adaptive numerical analysis is to optimise the usage of our computational resources. We will assume that the computational work W (some function of computational time and memory) required to define the discrete problem, and to obtain the finite element solution $X_h = (\boldsymbol{u}_h, p_h, T_h)$ is proportional to the length of the vector X_h. Let us further assume that we have a functional $\mathcal{E}$ which provides some measure of the error. For illustrative purposes we consider the following definition which measures the discretisation error:

$$\mathcal{E} := C_u \|\boldsymbol{u} - \boldsymbol{u}_h\|_{L_2(\Omega)} + C_p \|p - p_h\|_{L_2(\Omega)} + C_T \|T - T_h\|_{L_2(\Omega)}, \tag{7}$$

where $\boldsymbol{u}$, p, T is the exact solution to Eqs. (1), (5), and (6) and the error in each field is measured via the L_2 norm over the volume Ω, which for an arbitrary function ϕ is defined via:

$$\|\phi\|_{L_2(\Omega)} = \left[\int_\Omega \phi^2 \, dV \right]^{1/2}. \tag{8}$$

The normalisation and scaling of the individual error terms in Eq. (7) are absorbed into the constants, C_u, C_p, C_T. We note that Eq. (7) is not a practical choice as the analytic solution, will in general, not be known a priori. Optimal usage of our computational resource is expressed by either of the following statements:

C1: Given a tolerance ϵ_0, minimise the computational work W required to obtain X_h such that $\mathcal{E} < \epsilon_0$.
C2: For a fixed amount of computational work W_0, find a solution X_h which minimises $\mathcal{E}$.

The general adaptive procedure for a steady state problem can then be summarised as follows:

1 Generate an initial finite element mesh, $\mathcal{T}^0$. Specify a maximum number of refinement iterations $k_{\max}$. Initialise the adaptive cycle counter, $k = 0$.
2 **Solve:** Solve the discrete problem on the current mesh, $\mathcal{T}^k$.
3 **Estimate:** Analyze the computed solution. Estimate the error $\mathcal{E}$, and evaluate the error on each element. If either condition *C1* or *C2* are satisfied $\longrightarrow$ stop.
4 **Mark:** Based on the element-wise error, identify elements which should be modified, i.e. coarsened or refined.
5 **Adapt:** Using the elements marked for adaptation, generate a new mesh, $\mathcal{T}^{k+1}$.
6 If $k < k_{\max}$, set $k = k + 1$ and go to Step 2, else stop.

We note that for time dependent problems, any fields (i.e. temperature) associated with the previous time step, must be interpolated onto the new mesh $\mathcal{T}^{k+1}$. In Section 2.3 we discuss different means by which the finite element mesh can be modified in the "Adapt" step. In Section 2.4 we describe several different approaches to define a measure of the solution error, $\mathcal{E}$ which is used in the "Estimate" step. An overview of different methods to solve the system of algebraic equations arising from the finite element discretisation is discussed in Section 2.5.

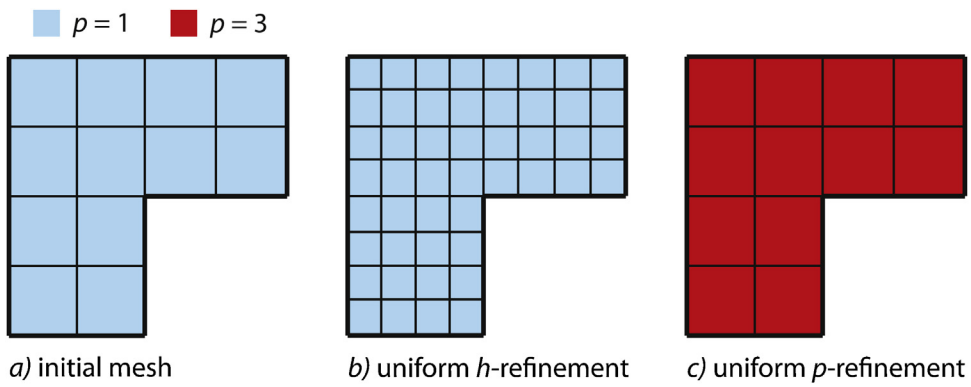

Fig. 3. Examples of different types of refinement. Colours indicate the polynomial order (p) of the basis function used over each element. Thin black lines denote element boundaries. (a) The initial mesh $\mathcal{T}^0$, prior to adaptivity. Subsequent panels illustrate refinement of $\mathcal{T}^0$ using; (b) uniform h-adaptivity and (c) uniform p-adaptivity.

2.3. Types of adaptivity

The two quantities which control the accuracy of the finite element solution are the element size, which we will denote via h, and the order of the polynomial p, which we used to approximate the unknown field over each element. The two classical types of adaptivity derive directly from modifying either the size of the elements (h-adaptivity) or the polynomial order of the interpolating function (p-adaptivity). We illustrate these types of adaptive refinement in Fig. 3b and c) respectively. The aforementioned images depict *uniform refinement*, that is, all elements in the mesh are modified in the same manner. We emphasise that the key to the adaptive procedure using either h or p adaptivity is that the modification to the discretisation only occurs locally within Ω.

In performing local h-refinement, two possibilities exist for locally decreasing the element size: element subdivision or remeshing. Both approaches are shown in Fig. 4. In the element subdivision approach, the element selected for refinement is divided into a number of child elements, whilst preserving the original element boundary. Element subdivision produces nested meshes in which nodes in the coarse (unrefined) mesh exist within the fine (refined) mesh. Direct application of element subdivision to an initially conforming mesh will create a non-conforming mesh through the introduction of "hanging nodes" (shown as open circles in Fig. 4c). To render the refined mesh conforming, one can either (i) introduce algebraic constraints at each hanging node to enforce continuity, or (ii) refine the neighbouring elements using special "blending" elements (see Fig. 4d). In the remeshing approach, a completely new mesh is generated with the desired distribution of element sizes. In general, even if the initial mesh was structured, the refined mesh obtained via the remeshing approach will be unstructured. We illustrate an example of remeshing in Fig. 4b).

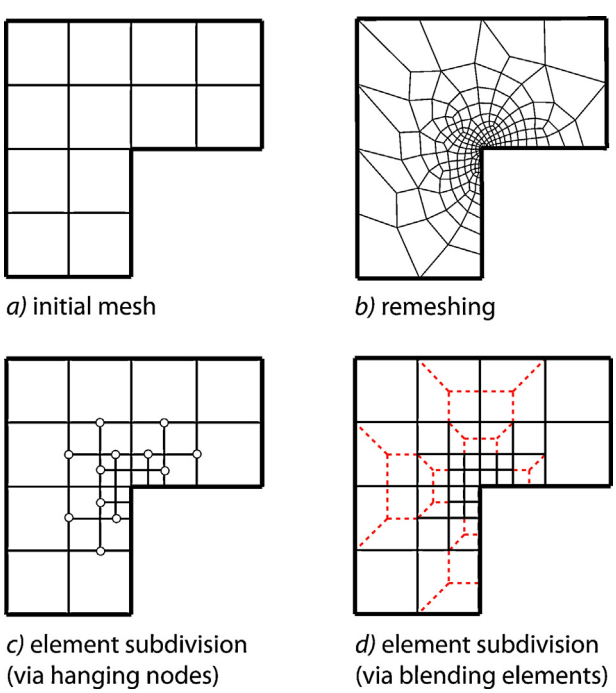

Fig. 4. Examples of local h-refinement at the re-entrant corner within an L-shaped domain. (a) The initial mesh $\mathcal{T}^0$, prior to adaptivity. Local adaptivity of $\mathcal{T}^0$ using (b) remeshing and (c and d) element subdivision. Hanging nodes are indicated via open circles and "blending" elements introduced to enforce continuity are indicated with thin, dashed red lines. All elements employ the same order basis function.

From classical interpolation theory, it is well known that high order interpolating functions (e.g. cubic, quartic) produce more accurate results than lower order functions (e.g. constant, linear) if the original function is smooth. However, in scenarios where the original function is discontinuous, the use of high order interpolating functions will introduce an artificial "wiggle" which overshoots and undershoots the function. In cases where the function is discontinuous, using low order interpolants, (piecewise constant or linear approximations) yields more accurate results. Exploiting p-adaptivity seeks to modify finite element discretisation by changing the polynomial order of the finite element space (field representation) used within each element, to more accurately approximate the numerical solution. A higher accuracy finite element solution is achieved if regions where the solution is smooth and varying rapidly are identified, and accordingly in these locations, higher order elements are employed. In regions where the solution is varying slowly (or is constant), or is discontinuous – low order polynomials are utilised. This naturally leads to the idea of combining both forms of adaptivity, i.e. modifying both the h and p value of the elements used to partition Ω, this is known as hp-adaptivity.

In the context of a simple isoviscous mantle convection simulation (see Fig. 5 – central panel), we have a nearly isothermal interior, and temperature variations occur predominantly at the boundaries within the thermal boundary layer. Considering purely h-adaptivity (see Fig. 5 – left panel), we would ideally refine the mesh only in regions where the thermal gradient is large, i.e. places where the approximation error would be largest. In the isothermal core where the temperature field is not changing rapidly, fewer elements are required. Given the temperature field is smooth, if we utilised hp-adaptivity (see Fig. 5 – right panel), it would be natural to use higher order polynomials along the boundaries to capture the sharp thermal boundary layer, and in the interior we would use a lower order polynomial as the temperature field is approximately constant.

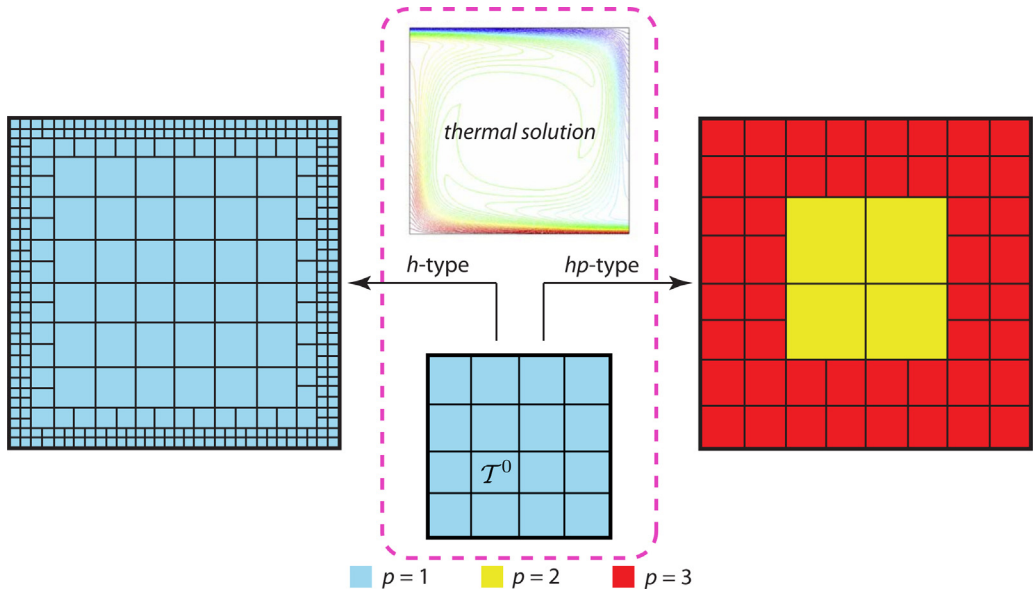

Fig. 5. Illustration of using *h*-adaptivity (left) and *hp*-adaptivity (right) in an example of mantle convection where adaptivity is designed to resolve the thermal boundary layers which develop along the domain walls. The central panels show the numerical solution of the temperature field and the mesh ($\mathcal{T}^0$) on which it was computed.

An alternative approach to modify the spatial discretisation is *r*-adaptivity. In this procedure, the mesh adaption process does not introduce any additional nodes, but rather re-distributes nodes within the original mesh, whilst preserving the mesh connectivity (node-edge-cell relationships), into a new configuration which better approximates the solution. In order to explain this further, an example of *r*-adaptivity is provided in Fig. 6, where we consider the extension of a linear elastic beam containing a small crack. The crack tip is located at the centre of the bottom edge of the domain. This simulation employed an unstructured mesh of triangles. The presence of the crack induces a large stress concentration (i.e. large gradients) in the vicinity of the tip. The movement of the nodes in this *r*-adaptive strategy utilises an energy minimization approach which produces a mesh in which the nodes are clustered around the discontinuity to better resolve the steep stress gradients in the vicinity of the crack.

2.4. Error estimators

In Section 2.3, we described several different approaches to modify the spatial discretisation of the mesh, however we did not indicate a procedure to determine *when* the discretisation should be modified. At this point, we should emphasise that "adapting" the spatial discretisation implies both the process of mesh refinement and mesh coarsening, using any combination of *h*, *p*, *r*-adaptivity.

To determine if a region within the discretised model domain should be refined or coarsened, some measure of the local error associated with the solution is required. If the local error is observed to be too large, the mesh resolution is refined. Conversely, if the error is below some threshold, the mesh resolution may be coarsened. The class of methods which provide this information are called error estimators.

Point-wise defined error estimates are in general not convenient for steering mesh adaptivity. This is evident if we consider problems with solutions which possess singularities (e.g. the stress field near a re-entrant corner). In this case, the field values may locally be infinite, yet the global solution is well behaved. Consequently, errors are frequently measured in an integral sense via some norm, for example in the L_2 norm (e.g. Eq. (8)).

A large volume of literature exists on the topic of errors estimators for a wide class of linear and non-linear problems (diffusion, convection-diffusion, Navier–Stokes, Maxwell equations). To develop an appreciation for the topic we refer readers to the pioneering work of Babuska and Rheinboldt (1978a,b) and to Verfürth (1996), Ainsworth and Oden (2000) for a comprehensive coverage of the topic. Here we briefly summarise several commonly used error indicators.

2.4.1. Residual based estimators

Residual based estimators aim to improve the quality of the numerical solution by minimising the discretisation error of the finite element approximation (u_h) using some global measure (i.e. norm). Whilst it is possible to compute the discretisation error via knowing the exact solution (u), or by solving very accurately the solution on an extremely well resolved mesh, neither option is practical. Residual based estimators approximate $e_h = u - u_h$

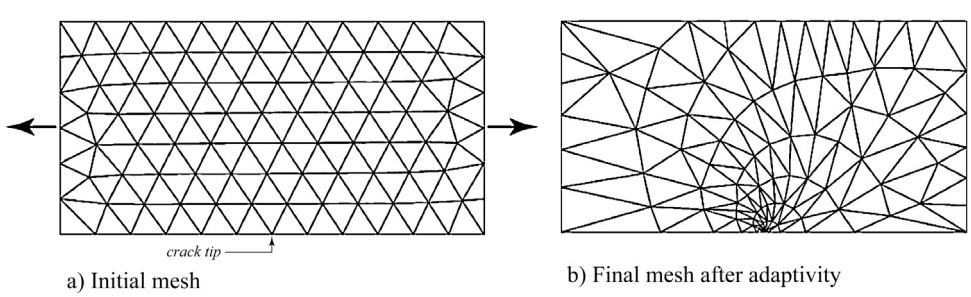

Fig. 6. Example of the *r*-refinement via a variational ALE approach (taken from Thoutireddy (2003)).

without having to resort to either of these options. The essential idea stems from the definition of the residual of the variational problem. To illustrate this concept, we consider the diffusion equation, defined in a domain Ω with boundary Γ,

$$\nabla^2 u = f \quad \text{in} \quad \Omega, \tag{9}$$

subject to the Dirichlet boundary conditions $u = u_D$ and Neumann boundary conditions $\nabla u \cdot \boldsymbol{n} = g_N$. Dirichlet and Neumann conditions are applied on the boundary segments Γ_D and Γ_N respectively and collectively these boundary segments cover all Γ (i.e. $\Gamma_D \cup \Gamma_N = \Gamma$). The variational form of the diffusion equation, for a test function v is

$$\int_\Omega \nabla v \cdot \nabla u \, dV = \int_\Omega v f \, dV + \int_{\Gamma_N} v g_N \, dS. \tag{10}$$

Residual based estimators make extensive use of the energy norm $\|\cdot\|_E$ associated with the bilinear form of the variational problem, which for the diffusion equation yields:

$$\|u\|_E = \sqrt{\left[\int_\Omega \nabla u \cdot \nabla u \, dV\right]}. \tag{11}$$

Defining the error as $e_h = u - u_h$, and substituting into Eq. (10), we obtain

$$\int_\Omega \nabla v \cdot \nabla e_h \, dV = \int_\Omega v f \, dV + \int_{\Gamma_N} v g_N \, dS - \int_\Omega \nabla v \cdot \nabla u_h \, dV = R(v), \tag{12}$$

where the right hand side of Eq. (12) is defined as the weak residual $R(\cdot)$ of the finite element approximation.

Applying integration by parts to the last term, we obtain

$$\int_\Omega \nabla v \cdot \nabla e_h \, dV = \int_\Omega v f \, dV + \int_{\Gamma_N} v g_N \, dS + \int_\Omega v \nabla^2 u_h \, dV - \int_\Gamma v (\nabla u_h \cdot \boldsymbol{n}) \, dS. \tag{13}$$

In explicit residual estimators (Babuska and Rheinboldt, 1979; Kelly et al., 1983; Babuska and Miller, 1987), the integral in Eq. (13) is split into the contributions from each element $K \in \mathcal{T}$. Considering only the last term, the boundary integral will be evaluated over each face of the element K, which is denoted via ∂K. Across any element edge (with normal $\boldsymbol{n}$) located within the interior of the domain (i.e. not contained in Γ), the quantity $\nabla u_h \cdot \boldsymbol{n}$ is discontinuous. The element-wise decomposition of Eq. (13) is stated as

$$\int_\Omega \nabla v \cdot \nabla e_h \, dV = \sum_{K \in \mathcal{T}} \int_K \mathcal{R}(u_h) v \, dV + \sum_{K \in \mathcal{T}} \int_{\partial K} \mathcal{J}(u_h) v \, dS, \tag{14}$$

where $\mathcal{R}(\cdot)$ contains all the volumetric integrals and $\mathcal{J}(\cdot)$ contains all the surface integral terms from Eq. (13).

Direct calculation of this jump in the flux can be used to construct an estimate of the discretisation error. This type of residual estimator is referred to as explicit as it can be constructed directly from the finite element solution u_h. Typically one considers an upper bound estimate for the discretisation error $\widehat{E}_h$, which can be computed via

$$\widehat{E}_h^2 = \sum_{K \in \mathcal{T}} \hat{\eta}_K^2, \tag{15}$$

where the error from each element $\hat{\eta}_K$ is

$$\hat{\eta}_K^2 = c_1 h_K^2 \|\mathcal{R}\|_{L_2(K)}^2 + c_2 h_K \|\mathcal{J}\|_{L_2(\partial K)}^2 \tag{16}$$

and h_K is a measure of the element size and c_1, c_2 are constants independent of h_K. The exact value of the constants c_1 and c_2 are generally unavailable, and in fact are dependent on the specific physical problem being solved (e.g. linear diffusion). Nevertheless, approximate values for c_1 and c_2, possibly determined via empirically studies, can lead to an error indicator which is useful in steering mesh adaptivity. Given Eq. (15), the relative error (in the energy norm) e_r of the finite element solution can be approximated via

$$e_r = \frac{\widehat{E}_h}{\|u\|_E}. \tag{17}$$

An optimal mesh is one in which each element contributes an equal amount to the total relative error e_r. Thus, adaptivity proceeds by evaluating the element wise contribution to e_r, and marking an element for refinement if the element contribution is above the expected element contribution $\bar{e}_K$ to the total error, or marking an element for coarsening if the contribution to error is below $\bar{e}_K$. Explicit residual error estimators are desirable as they are simple to implement and inexpensive to compute. However, they can be sensitive to mesh distortion and since it was derived from an upper bound, they can significantly over estimate the error, causing excessive amounts of mesh refinement.

Implicit residual estimators aim to construct solutions of Eq. (12) to directly evaluate the error. To obtain a meaningful error estimation, the function space representing e_h must be contain more information (i.e. be "richer") than those used to represent u_h – this could be achieved using higher order basis functions, or by discretising e_h on a finer mesh. Using either option to define a new problem over the entire domain Ω is more expensive than the original problem, thus implicit residual estimators solve a local version of Eq. (12) over an element (Bank and Weiser, 1985), or a patch of elements (Babuska and Rheinboldt, 1978a,b; Carstensen and Funken, 1999) to estimate e_h. Differences in implicit residual estimators arise from how the local subdomain is defined and how the boundary conditions on the local problem are defined. The solution of the auxiliary problem required is local (to the element or patch) and thus small, hence individually these are much cheaper to solve than the original problem. However, overall the cost of computing this estimator is non-trivial compared to the explicit residual approach.

2.4.2. Recovery based estimators

One of the classical techniques to define an error indicator is via the construction of higher order estimates to approximate discretisation error $e_h = u - u_h$. As stated previously, this definition is somewhat impractical (or expensive to compute if using a high resolution reference solution). Recovery based error estimators attempt to construct an approximation to the true error via

$$e_h \approx e_h^* = u^* - u_h, \tag{18}$$

where u^* is a higher order accurate estimate of u in comparison to u_h. The assumption in using higher order estimates is that since u^* converges to the exact solution faster than u_h, the error given by e_h^* is a meaningful approximation of the true discretisation error in an appropriate norm. This is mathematically justified under certain conditions (Ainsworth and Oden, 2000).

Like the explicit residual estimators, recovery methods utilise the discontinuous nature of the gradient of the finite element solution u_h across element boundaries to define an approximation of the error. Here the discretisation error is defined in terms of the L_2 norm of the gradient of the finite element solution, compared with some "recovered" or post-processed gradient,

$$E_h = \left[\int_\Omega (\mathcal{G}(u_h) - \nabla u_h)^2 \, dV\right]^{1/2}, \tag{19}$$

where $\mathcal{G}(u_h)$ is the improved estimate of the gradient of u_h which was computed via some form of post-processing procedure. The post-processing techniques employed are defined locally about each finite element node i, and utilise data from a "patch". A patch is a generic term for a set of elements. The patch may consist of a single element, or all of the elements connected to a given node, or some larger region of the domain. Since the reconstruction are applied locally, element-wise contributions to the error $\hat{\eta}_K$ can be readily defined as,

$$\hat{\eta}_K^2 = \|\mathcal{G}(u_h) - \nabla u_h\|_{L_2(K)}^2. \tag{20}$$

Using Eq. (20) together with Eq. (15), the relative error in the energy norm can be defined and adaptivity can be directed by the magnitude of the element-wise error with respect to the total relative error e_r.

Several gradient recovery techniques suitable for reconstructing higher order accurate estimates of ∇u_h exist. The classical method is the Super-convergent Patch Recovery (SPR) method (Zienkiewicz and Zhu, 1987, 1992a). In the SPR approach, a patch of elements adjacent to a node is constructed, and a least squares fit is performed of ∇u_h sampled at specific points within an element patch. The locations where the gradient is sampled are referred to as "super-convergent", implying that the order of accuracy of ∇u_h at these specific location is higher than at arbitrary locations within the element. For certain element types, the existence of super-convergent points can be proven – however the result does not extend to all elements. The application of SPR as an error indicator was initially demonstrated in Zienkiewicz and Zhu (1992b,c).

A more general recovery is the Recovery by Equilibrium of Patches (REP) method (Boroomand and Zienkiewicz, 1997a,b). In the REP method, the original PDE (or equilibrium equation) is expressed in terms of the field we seek to recover (e.g. ∇u_h), and then the PDE is solved in least squares sense over a patch of elements adjacent to a given node. Thus, the method can be regarded as a least squares fit subject to constraints imposed by the original equilibrium equation. REP has the advantage over SPR in that it is applicable for any element type, and in principal, any physical set of equations as its foundations are based on a least squares fit of a field, subject to equilibrium constraints. Additionally, REP is extremely flexible, as its permits any variable, which the PDE can be posed in terms of, to be recovered. We note that both SPR and REP apply a local least squares operation over a patch, thus the methods are not effective at capturing ∇u_h if the gradient should physically be discontinuous, i.e. as may occur if a material discontinuity (such as viscosity or thermal diffusivity) intersects the patch.

2.4.3. Goal oriented estimators

Goal oriented adaptivity seeks to modify the finite element mesh, with the objective of improving the numerical approximation of a specific quantity of interest (as opposed to minimising the discretisation error) which is computed from the FE solution, u_h (Babuska and Rheinboldt, 1978b; Ainsworth and Oden, 2000; Oden and Prudhomme, 2001; Becker and Rannacher, 2001). Under this requirement, adaptivity driven via discretisation error estimates defined in the energy norm will not generally yield the most optimal mesh. For example, in the context of the diffusion equation, we may seek to obtain the most accurate estimate of the heat flux at the surface of the domain. Accordingly, the error measured is now with respect to the specific goal (or objective) of the analysis (e.g. the surface heat flux).

In general, employing goal oriented error estimation will require the solution of the adjoint problem to our PDE. To develop the general concept of goal oriented error estimation, we will assume that $L(u)$ is a linear functional defining the quantity of interest (or goal) of our computation, e.g.

$$L(u) = \int_{\Gamma_s} \nabla u \cdot \boldsymbol{n}\, dS, \tag{21}$$

where Γ_s is the portion of the boundary on the surface of the domain Ω. From the linearity of $L(\cdot)$, we can define the error in the quantity of interest to be

$$E_L = L(u) - L(u_h) = L(u - u_h) = L(e_h). \tag{22}$$

Inspection of Eq. (12) indicates that the residual, $R(.)$, is the source term which drives the numerical error e_h. The key to goal oriented estimation is to avoid solving Eq. (12) for e_h, and then evaluating Eq. (22) to determine how the mesh should be adapted; rather the objective is to relate E_L directly to R. These quantities are related through an influence function (or adjoint solution) z defined with respect to L. The diffusion equation under consideration (the primal problem) is self adjoint, thus the influence function z relating the error E_L to the objective function L is obtained via the solution to the dual problem:

$$\int_{\Omega} \nabla v \cdot \nabla z\, dV = L(v). \tag{23}$$

Applying the FE discretisation to Eq. (23) and solving for an approximate z_h, we can follow the ideas outlined in Section 2.4.1, whereby we use an upper bound estimate of $L(e_h)$ and explicit residual based estimators to build the following error estimate in L;

$$\widehat{L}_h = \sum_{K \in \mathcal{T}} \hat{\omega}_K\, \hat{\eta}_K, \tag{24}$$

where in this instance, the error is weighted by element-wise contributions from both the solution to the primal problem $\hat{\eta}_K$, and the dual problem $\hat{\omega}_K$. The specific error contributions are given by the volume and jump contributions of the residual (cf. Eq. (16)):

$$\hat{\omega}_K^2 = c_1 h_K^2 \|\mathcal{R}(z_h)\|_{L_2(K)}^2 + c_2 h_K \|\mathcal{J}(z_h)\|_{L_2(\partial K)}^2, \tag{25}$$

$$\hat{\eta}_K^2 = c_1 h_K^2 \|\mathcal{R}(u_h)\|_{L_2(K)}^2 + c_2 h_K \|\mathcal{J}(u_h)\|_{L_2(\partial K)}^2. \tag{26}$$

In the context of goal oriented adaptivity, the relative error e_r is defined with respect to the objective L, thus we have

$$e_r = \frac{\widehat{L}_h}{|L(u_h)|} \tag{27}$$

and mesh adaptivity is steered in accordance with the size of the element error $\hat{\omega}_K \hat{\eta}_K$ contribution to e_r.

2.5. Solvers for systems of linear equations

The finite element discretisation transforms the PDE into a discrete system of algebraic equations. Within the scope of this review, we have only considered linear problems, thus the discrete system of equations is also linear and can be expressed simply as

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{28}$$

where $\boldsymbol{b}$, $\boldsymbol{x}$ are N-dimensional vectors representing the discrete forcing term and the unknown field (e.g. velocity, pressure, temperature) and $\boldsymbol{A}$ is an $N \times N$ non-singular matrix representing the discrete form of the differential operator. We note that when solving non-linear problems, linearization techniques are employed, reducing the non-linear problem to a system of the form in Eq. (28). The finite element method utilises basis functions which have compact support over adjacent elements, implying that the basis functions vanish outside of this domain. This naturally implies that

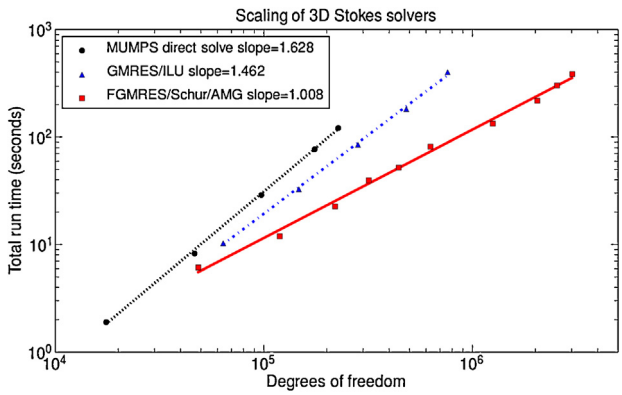

Fig. 7. CPU time required to solve a 3D Stokes problem versus degrees of freedom, using (i) a direct solution technique (dashed black line); (ii) an iterative method combined with a standard preconditioner (incomplete LU) (dashed blue line) and (iii) an iterative method with an algebraic multi-grid preconditioner (red line). All calculations are performed on a single processor. Image courtesy of Jed Brown (jed-brown@anl.mcs.gov).

$\boldsymbol{A}$ will be sparse. We now briefly discuss several approaches to obtain the solution $\boldsymbol{x}$ in the context of adaptive finite elements.

There are two distinct classes of solution algorithms; direct methods and iterative methods. We briefly distinguish the difference between the two solution methods, and refer readers to Golub and Loan (1996) and Saad (2003) where these methods are discussed in more detail.

To obtain the solution $\boldsymbol{x}$, we require the calculation of $\boldsymbol{A}^{-1}\,\boldsymbol{b}$. Naively, one could consider that this statement implies we first explicitly assemble the matrix $\boldsymbol{A}^{-1}$ and then perform the matrix vector product with $\boldsymbol{b}$. This approach is rarely ever advocated, especially when $\boldsymbol{A}$ is sparse, as the fully assembled inverse of $\boldsymbol{A}$ will be dense. The alternative statement is: we seek a method to apply the *action of the inverse of* $\boldsymbol{A}$ on a vector. Efficient direct solution methods factor the original matrix into several components, for which the action of the inverse applied to a vector is trivial. For example, LU decomposition factors $\boldsymbol{A}$ according to $\boldsymbol{A} = \boldsymbol{LU}$, where $\boldsymbol{L}$ and $\boldsymbol{U}$ are lower and upper triangular matrices. Following the factorization, the solution is obtained via first performing $\boldsymbol{y} = \boldsymbol{L}^{-1}\,\boldsymbol{b}$ (forward solve), and then computing $\boldsymbol{x} = \boldsymbol{U}^{-1}\,\boldsymbol{y}$ (backward solve). We note that key to the factorization is that the action of the inverse of upper/lower triangular matrices on a vector, is trivial. We refer to the following freely available software packages for implementations of sparse matrix factorizations which are suitable for symmetric and non-symmetric matrices – `UMFPACK` (Davis, 2004); `CHOLMOD` (Chen et al., 2008); `PARDISO` (Schenk and Gartner, 2006); `PaSTiX` (Hénon et al., 2002); `SuperLU` (Li, 2005); `MUMPS` (Amestoy et al., 2001)

Modern factorization techniques can exploit the sparsity prevalent in finite element operators, such as $\boldsymbol{A}$. To perform the factorizations, sparse direct methods require $O(N^{3/2})$ floating point operations in two-dimensions, and $O(N^2)$ operations in three-dimensions. The number of operations required to perform the forward and backward solves requires $O(N \log N)$ and $O(N^{4/3})$ operations in two and three dimensions respectively (Li and Widlund, 2007). The memory usage required by sparse direction factorizations scales like $O(N \log N)$ and $O(N^{4/3})$ in two and three dimensions respectively (Calo et al., 2011). The scaling of the memory usage and CPU time required to factor finite element matrices can become prohibitively expensive when the number of unknowns becomes large. We illustrate this point in Fig. 7 where we show the CPU time requirements when using a direct solver (`MUMPS` – black dashed line). Clearly high-resolution 3D simulations are not possible using such techniques, even with the usage of a massively parallel computer.

In contrast to direct methods which produce the solution $\boldsymbol{x}$ in a "two-step" (factorize – forward/backward solve) procedure, iterative methods obtain updates to the solution via sequence of operations. A simple iterative method (Richardson's method) is provided below:

1 Set the iteration counter: $k = 0$.
  Choose the initial guess for the solution: $\boldsymbol{x}_0 = 0$.
2 Update the solution:

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^k)$$

3 Update counter: $k = k + 1$.
  Return to step 2.

Given the solution is updated at each iteration, iterative methods require a stopping condition to terminate the sequence when $\boldsymbol{x}^k$ is sufficiently accurate. Typically the stopping condition will involve the residual $\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^k$. For instance, one strict condition would be to stop the iterative sequence if $|r_i| \le \epsilon$ for all i, $\epsilon \ll 1$. Iterative methods also have the desirable property that only matrix–vector products with $\boldsymbol{A}$ and vector–vector operations are required to obtain to the solution. Thus in terms of memory usage, such methods can be considered cheap compared to direct solution methods.

A large number of robust iterative methods exist. Typically one chooses the appropriate iterative method based on the properties of the matrix itself, e.g. symmetric, non-symmetric, positive definite, etc. See Barrett et al. (1994) for an implementation "catalog" of different iterative methods and refer to Golub and Loan (1996) and Saad (2003) for theoretical justifications of these methods.

Despite the wealth of different iterative methods, in all cases, the number of iterations required to reach a converged solution can be large, and more over, the number of iterations will generally increase as the resolution of the finite element mesh is increased. To accelerate the convergence of iterative methods, typically a preconditioner $\boldsymbol{B}$ is employed. In the context of the Richardson's method, the preconditioner is used in the following manner

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{B}^{-1}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^k).$$

If $\boldsymbol{B} = \boldsymbol{LU}$, then the sequence would obtain the exact solution in one iteration. However, as already explained, this choice is not always practical - particularly in three-dimensions. Intuitively we can infer that $\boldsymbol{B}$ should be a close approximation to $\boldsymbol{A}$ (in some norm), and the cost (memory and CPU time) to apply the action of its inverse should be cheap. In considering finite element calculations, an optimal preconditioner is one in which the number of iterations required to reach convergence is independent of the mesh resolution h. This further implies that the solution time required to obtain the solution scales like $O(N)$. The class of preconditioners which possess this property are called multi-level preconditioners. An excellent coverage of the theory and implementation of multi-level preconditioners can be found within Briggs et al. (2000), Wesseling (1992), Trottenberg et al. (2001), and Falgout (2006).

A multi-level preconditioner attempts to construct a hierarchical representation of the discrete PDE $\boldsymbol{A}$, containing a number of different levels (two or more). At the top of the hierarchy is the operator $\boldsymbol{A}$, and on each subsequent level we have a "coarser" representation of the operator from the level above. The under-lying principle of such multi-level algorithms is that we use solutions from the coarser level to accelerate the convergence on the finer levels. Information about the solution is passed between different levels in the hierarchy via interpolants (fine → coarse) and prolongation operators (coarse → fine). A more detailed definition of these inter-level transfer operators will be discussed below.

The two end member multi-level algorithms which are used to define the operators within the hierarchy are called geometric multi-grid and algebraic multi-grid. In geometric multi-grid (GMG), the hierarchy within the multi-level framework consists of a sequence of finite element meshes with an element resolution h which increases from the finest level, to the coarsest level. Given a structured grid, building a mesh hierarchy is naturally defined by simply sub-dividing each element (e.g. splitting a quadrilateral into four) and thereby producing a nested mesh with twice the resolution of the original mesh. The definition of the PDE on each level l can be defined by re-applying the finite element discretisation over the FE mesh at level l. For PDE's which contain coefficients which vary in space, a definition of these coefficients is required on each coarse grid. For example in a finite elements calculation for Stokes flow, we would require the viscosity at all quadrature points within each element on the coarse grid. The definition of the coarse grid coefficients is the essential weakness of GMG. Simply put, if the fine grid coefficient structure (for example a discontinuous viscosity) is "blurred" out when it is interpolated to the coarse grid, the coarse grid operator will not provide a good approximation – and as a result will degrade the performance of the preconditioner. If a hierarchy of coarse grids can be constructed in which element boundaries align with jumps in material properties, the preconditioner will remain optimal, however this construction is rarely possible for real geodynamic applications. The interpolation and prolongation operators in GMG are defined by geometrical interpolating values from one mesh to another using the underlying finite element basis functions (e.g. bilinear interpolation).

An algebraic multi-grid (AMG) preconditioner does not require any geometric information pertaining to either the mesh on the finest level (which was used to define $\boldsymbol{A}$), or a mesh hierarchy. Rather, AMG constructs the coarse grid operators by direct consideration of the matrix entries within $\boldsymbol{A}$. The non-zero structure of a matrix can be regarded as a edge-weighted graph in which the vertices in the graph represent the nonzero entries, and the weight placed on the edges between vertex i and j is the value of the non-zero entry A_{ij}. In AMG, the coarse operator is defined by considering only strongly coupled vertices, i.e. those in which A_{ij} is large. This approach proves to be very robust when applied to problems in which there are large jumps in coefficients as the algebraic coarsening procedure automatically preserves the coupling, whilst any contrast in coefficients in simply ignored by the geometric coarsening. The optimality of AMG is illustrated in Fig. 7 where the CPU time required to solve a 3D Stokes problem using `BoomerAMG` (red line) is shown. Here the slope of 1.008 indicates nearly optimal behaviour and implies that the solution time scales approximately linearly with respect to the number of degrees of freedom N_d in the problem – in contrast to the direct solution method which scaled like $N_d^{1.628}$.

For a didactic overview and comparison of different multi-level algorithms (geometric versus algebraic) for complex fluid simulations with large variations in material properties, we refer to MacLachlan et al. (2008). We refer readers to the following open source software packages for implementations of parallel algebraic multi-level preconditioners – `BoomerAMG` (Henson and Yang, 2002); `ML` (Gee et al., 2006).

For adaptive finite element calculations, the use of AMG can be extremely beneficial as it alleviates the need for any grid hierarchy. Whilst conceptually simple, building grid hierarchies on quad/oct-tree meshes which yield an optimal GMG preconditioner is non-trivial. This results from the fact that the entire adaptive mesh cannot be uniformly coarsened on each grid level. Instead, uniform coarsening will only occur on a subset of levels. In unstructured based adaptive FE implementations, the generation of a valid coarse grid representation is further complicated as there is no natural geometric structure (or nesting) within the mesh to exploit.

3. Adaptivity in geodynamics – the state of the art

In this section we summarise a number of recent efforts within the computational geodynamic community to incorporate various forms of spatial adaptivity within their numerical models.

3.1. Arbitrary Lagrangian–Eulerian (ALE) methods

The ability to model systems with deformable boundaries, or internal material interfaces is of great importance in geodynamics. For example, a deformable upper boundary is required when using a free surface boundary condition. This type of boundary is often required in studies examining the dynamics of subducting plates, and any regional models which are coupled to surface processes (Fullsack, 1995; Buiter et al., 2006; Popov and Sobolev, 2008; Moresi et al., 2007; Quinteros et al., 2009; Thieulot, 2011). Deformable internal boundaries are required in models of diaparism (Rayleigh Taylor instabilities) (Poliakov and Podladchikov, 1992; van Keken et al., 1997; Naimark et al., 1998; Longoni et al., 2011) and folding/buckling simulations of horizontal multi-layer sequences (Cobbold, 1977; Cobbold and Percevault, 1983; Lan and Hudleston, 1991; Schmalholz and Podladchikov, 1999; Kaus and Schmalholz, 2006). In such simulations, the mesh is required only to follow the upper boundary, or a material interface. In the applications mentioned above, following the upper boundary or material interfaces can be readily performed (up to modest strains) using structured meshes combined with a simple remeshing algorithm. The remesher is applied between interfaces and only redistributes nodes in the original mesh. Remeshing is required to ensure the elements do not become too distorted, as this will degrade the accuracy of the discrete solution. Consequently this class of geodynamic models can be classified as r-adaptive methods. However, it should be noted that the aforementioned methods, as described, do not employ any form of error estimator, thus the new configuration of the mesh obtained after advection and remeshing is not generated using any direct measures of error. Implicitly, the methods try to minimise sources of discretisation error by (i) accurately following material boundaries, thus minimising error in the definition of the boundary condition; (ii) ensuring that elements are not too distorted, thereby by minimising discretisation error in the finite element solution. We refer to Fig. 8 where examples of lithospheric deformation using this type of adaptivity are shown.

3.2. Unstructured h-adaptivity

One of the first examples utilising adaptive finite elements in geodynamics was presented in Davies et al. (2007) in which the authors exploited an automatic mesh generator to create adaptive unstructured quadrilateral meshes to study the dynamics of both thermal and thermo-chemical convection in two dimensions. See Fig. 9 (upper panels) for examples. The adaptivity employed consists of purely h-adaptivity with P_2 (quadratic) basis functions throughout the domain, with an error estimator defined via a local, interpolation-based error indicator. The interpolation error is estimated over each element by performing an L_2 projection (using a lumped mass matrix) of the first and second derivative of either temperature (thermal convection models), or both temperature and composition (thermo-chemical convection models). The usage of an unstructured mesh naturally introduces a non-nested grid during the process of refinement (or coarsening). In this work, the authors utilise a cubic interpolation to transfer time dependent nodal fields (temperature and composition) from the original mesh to the adapted mesh. One should note that the choice of transfer operator used between the original and adapted mesh will affect solution quality. For example, consistent interpolation defined using the finite element basis functions is not

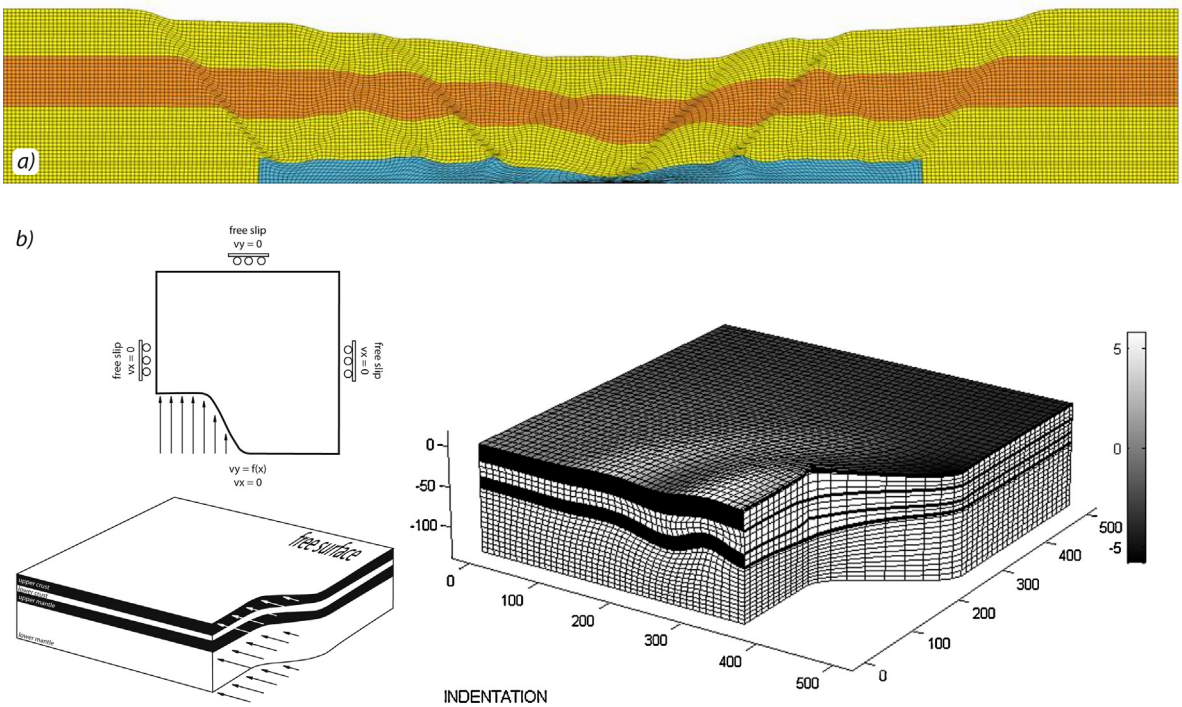

Fig. 8. Examples of lithospheric deformation that utilise full Lagrangian meshes with regular, vertical remeshing which conforms to material interfaces. (a) Numerical "sand box" experiment in which a layered visco-plastic material (yellow/orange) with a weak viscous inclusion (blue) is subjected to extension. Modified from Buiter et al. (2006). (b) Idealized model of the India–Asia collision zone which consists of a layered power-law material subjected to a spatial and temporal indentation velocity boundary condition. Modified from Lechmann et al. (2011).

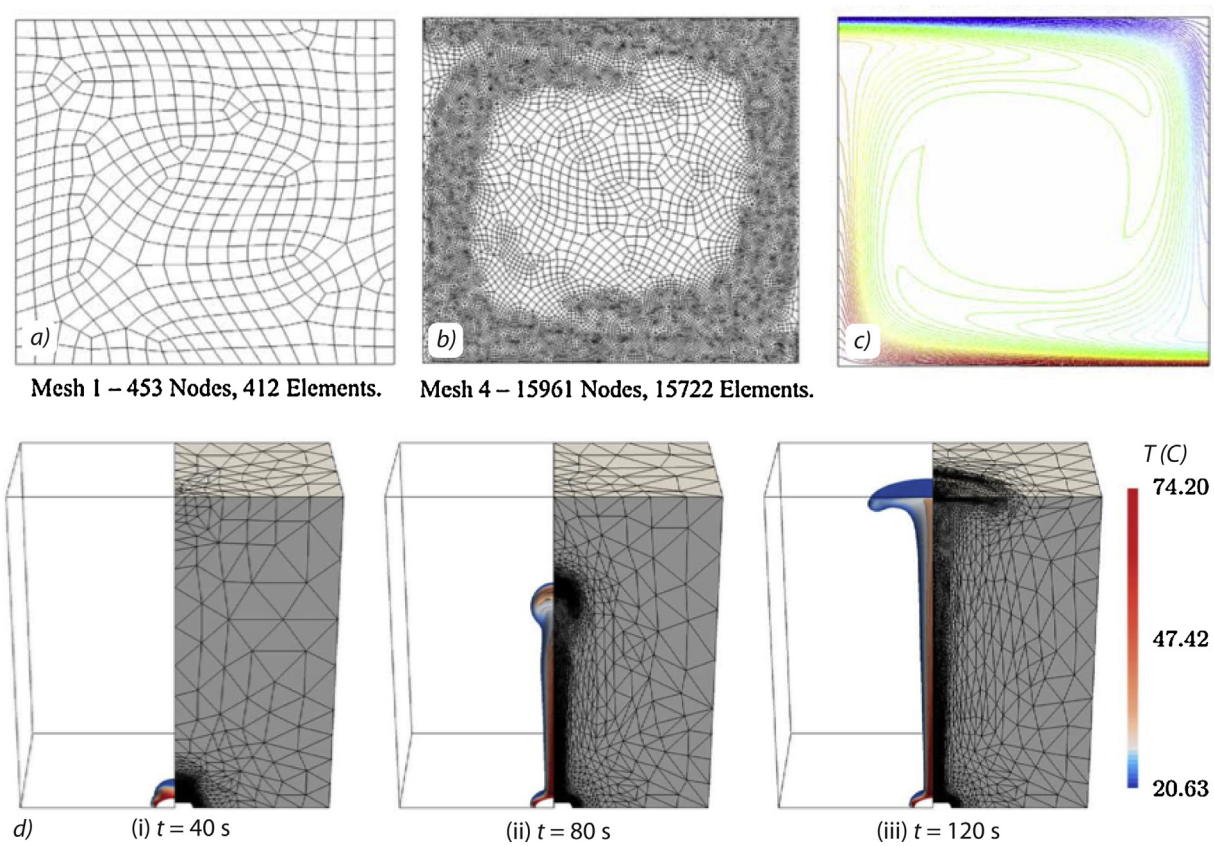

Fig. 9. Geodynamic examples using unstructured *h*-adaptivity in 2D and 3D. The upper panels (taken from Davies et al. (2007)) depict the solution of steady state thermal convection problem where (a) is the initial mesh, (b) the final mesh and (c) the resulting thermal field. The lower panels (taken from Davies et al. (2011)) depict the rise of a thermal plume over time.

conservative, i.e. the integral of the interpolated field over the domain will not match the integral of the original field. In contrast, a Galerkin projection which defines an interpolated field that minimises the L_2 norm of the interpolation error, is in fact globally conservative. Depending upon the physical application, enforcing conservation might be necessary, thus care should be taken when selecting the transfer operator used. For a discussion of the issues associated with the choice of inter-mesh transfer operator we refer to Davies et al. (2007), Farrell and Maddison (2010) and Davies et al. (2011).

The method described in Davies et al. (2007) demonstrated that spatial adaptivity, driven by an appropriate error indicator, is an effective technique to study thermal convection. The AFE method was shown to require dramatically less memory (RAM) and CPU time compared to structured grid methods to obtain the same level of numerical accuracy in measured Nusselt number, RMS velocity and heat fluxes in the corners of the upper surface. The thermo-chemical convection study indicated that the AFE method was still quite numerically diffusive, compared to tracer based or marker-chain techniques of van Keken et al. (1997) and Tackley and King (2003), which lead to erroneous estimates of entrainment rates. The diffusiveness of the results is associated with the particular grid based technique which was used to solve the pure advection equation (Brooks and Hughes, 1982) describing the evolution of the composition. Using an adaptive mesh was shown to decrease the diffusiveness of the field based advection scheme, as the AFE results were shown to be superior to those computed using a non-adaptive, uniform mesh.

In Davies et al. (2011), the unstructured, adaptive finite element framework Fluidity was introduced and the methodology was verified using previous numerical (2D and 3D) and experimental studies (3D) of thermal convection applied to mantle dynamics. Example simulations are shown in Fig. 9 (lower panels). Similar to the earlier work of Davies et al. (2007, 2008), the unstructured nature of the grid used by Fluidity permits complex domain geometries to be defined. This framework also permits the usage of highly anisotropic elements, thereby improving mesh optimization in regions where thin boundary layers may develop. Individual error estimates are defined for each physical quantity (i.e. velocity, pressure, temperature). The error estimate employed for a given field ϕ, is based on the 1D interpolation error which is approximated via

$$\epsilon = h^2 \left| \frac{\partial^2 \phi}{\partial x^2} \right| ,$$

where h is the element length. In multiple dimensions, the curvature term is generalised to consider variations of the second derivative along a direction which is oriented arbitrarily with respect to the element. This is necessary to define error estimates which are sensitive to element anisotropy. In coupled simulations where multiple fields are present, the error estimates are combined and the adaptivity rules are applied based on the composite error indicator.

In the experiments presented, both h-adaptivity and r-adaptivity were considered using a P_2-P_1 element pair to discretise the Stokes problem, together with P_2 elements to discretise the advection-diffusion equation. Stabilisation techniques for the advection-diffusion equation (e.g. SUPG Brooks and Hughes (1982)) were not employed as solutions were found to be sufficiently smooth such that no stabilisation of the Galerkin discretisation was required. Additionally, this also allowed the use of a P_2 discretisation for temperature. We note that Fluidity offers the choice of other element pairs suitable for Stokes flow and the energy equation, however different discretisations were not explored in this study. Interpolation of fields between meshes was performed using the natural basis functions associated with each field (e.g. P_2 for velocity and temperature and P_1 for pressure). The discrete Stokes problem is solved using Krylov method preconditioned with an algebraic multi-grid based on smoothed aggregation (Vanek et al., 1996; Kramer et al., 2010). The coupled system of Stokes and energy is solved using a Picard iteration.

From the comparison with previous numerical and laboratory experiments, the Fluidity framework was demonstrated to produce accurate results for both 2D and 3D, steady state, isoviscous and variable viscosity simulations. Additionally, the anisotropic adaptive unstructured meshes produced equally accurate model results with significantly fewer grid points than when uniform, structured meshes were used. Whilst not explored in Davies et al. (2011), the mesh infra-structure and algebraic multi-grid employed by Fluidity support execution on parallel computers. Demonstrating the parallel scalability of these components and the parallel performance of the entire Fluidity framework for use in geodynamic modelling will undoubtedly be the focus of upcoming studies.

Arguably fully unstructured, front tracking methods such as MILAMIN (Dabrowski et al., 2008) can also be regarded as a type of adaptive finite element method (h-type) employing an unstructured adaptive mesh technique. In MILAMIN, inclusions and internal interfaces are explicitly represented and an unstructured mesh is generated which conforms to both these internal boundaries and the external model domain boundary. During the model evolution, the interfaces are evolved in time, and the triangulated mesh is recomputed based on the new configuration of the interface. However, the MILAMIN implementation (as published) does not explicitly include any form of error estimator to guide mesh adaptivity in a manner which addresses either of the two adaptive finite element objectives defined in Section 2.2. The only consideration made in regenerating the mesh given the new configuration of the internal and external boundaries, is that it is a quality mesh, meaning that each triangle satisfies a number of geometric constraints related to its aspect ratio, interior angles and the cell volume.

3.3. Structured h-adaptivity

Another two dimensional study focusing on the usage of adaptive finite elements to examine the dynamics of thermo-chemical convection was presented in Leng and Zhong (2011). In contrast to the mesh style used in (Davies et al., 2007), here the authors employ a nested grid structure defined via a quad-tree (see Fig. 10a) in conjunction with h-adaptivity. The underlying equations for momentum and mass (Stokes) and the energy transport equation are solved using stabilised finite elements formulations with Q_1 (bilinear) basis functions for each field. The composition field is discretised using the tracer ratio method (Tackley and King, 2003) in which distinct chemical compositions are represented via a set of Lagrangian markers. In this work, the action of refinement and coarsening affects both the mesh and the set of markers, thus here we will refer to "discretisation adaptivity". The mesh is considered as the "master" and dictates how the discretisation will be modified. Under refinement or coarsening, the marker distribution within each element in the new mesh is dynamically adjusted to ensure that the number of markers within each elements is bounded within a predefined range. Removal of markers is determined randomly. Similarly, if new markers are required to be introduced within an element, their coordinates are determined randomly and their chemical composition is defined by the composition of the element. Since spatial adaptation of the mesh given by a quad tree structure requires only local operations, the marker distribution can be modified in an element-by-element fashion, thus rendering the operation of re-meshing fast and efficient in terms of both CPU time and RAM.

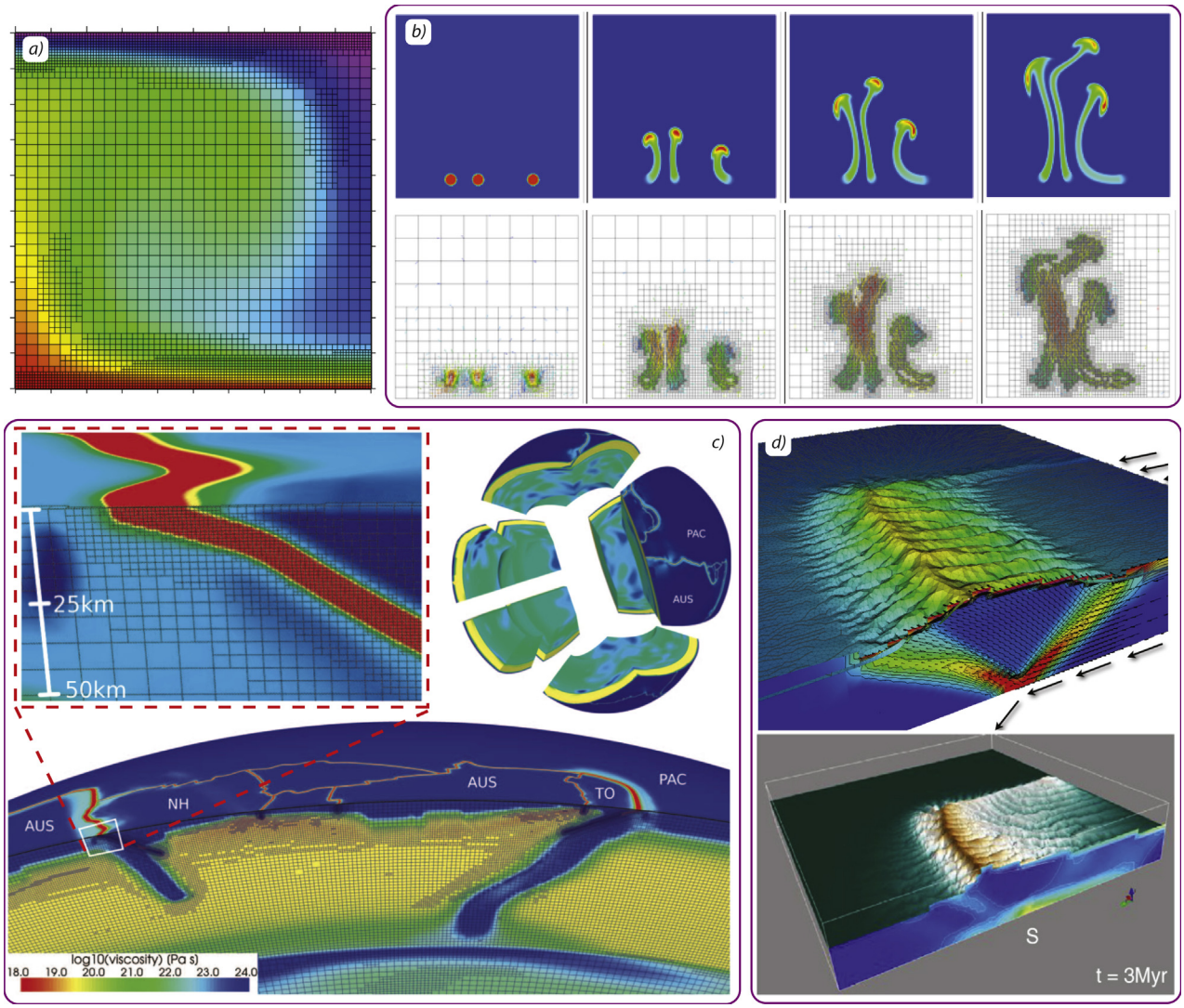

Fig. 10. Geodynamic examples utilising structured *h*-adaptivity via hanging nodes in 2D and 3D. (a) Steady state convection with $Ra = 10^5$ from Leng and Zhong (2011). (b) Development of plumes – upper panels show temperature field, lower panels show the mesh resolution (taken from http://www.dealii.org). (c) Snapshots from Stadler et al. (2010) illustrating the forest of oct-trees employed by RHEA to study convective flow in the present day Earth (Stadler et al., 2010). (d) Models of coupled tectonic and surface processes obtained using DOUAR (Braun et al., 2008).

The novelty of the method described by Leng and Zhong (2011) is that they use the composition field represented by markers as a refinement (and coarsening) indicator to drive mesh adaptivity. This indicator is used in addition to a classical gradient based error indicator applied to the temperature field. In this way, thermal boundary layers can be captured, as well as the sharp interface between the compositional boundaries. The approach alleviates much of the numerical diffusion observed in a purely grid based method (Davies et al., 2007) due to the fact that the marker based representation of compositional fields possess significantly less numerical diffusion (Tackley and King, 2003). Furthermore, the adaptivity introduced for the marker discretisation was shown to not introduce artifacts in the composition interface, or cause a significant loss of conservation in the compositional field. Furthermore, the mass conservation error was observed to decrease with increased refinement. The combination of grid based and marker based discretisation together with adaptivity applied to both discretisations was found to be suitable to study thermo-chemical convection and to measure entrainment rates.

The computational cost required to perform 3D geodynamic simulations is non-trivial in terms of both CPU time and memory (RAM) required. In 3D, the highest resolution model which can be performed is typically bounded by memory requirements, rather than CPU time. In 2D, the converse is true. Consequently, when performing 3D simulations, exploiting spatial adaptivity to resolve local, small scale features is extremely desirable. In Braun et al. (2008), an oct-tree based method combined with a level-set based surface tracking algorithm (named DOUAR – see Fig. 10d) was developed to study the coupling and feedback between tectonic processes and landscape evolution. In this approach, *h*-adaptivity was employed to resolve material interfaces within the model. This was achieved by prescribing that the oct-tree refine any elements which contained an interface represented by the level-set, down to some specified number of refinement levels. The Stokes problem was discretised using Q_1–P_0 mixed finite elements. The discrete pressure and velocity equations are decoupled via a penalty formulation (Hughes, 1987), and the resulting discrete system of equations was solved using one of the parallel direct solvers WSMP (Gupta et al., 1997) or MUMPS (Amestoy et al., 2001).

The models presented in Braun et al. (2008) clearly demonstrated the potential of using AFE techniques to study three-dimensional thermo-mechanical processes relevant to geodynamics. Among these preliminary models were the first 3D

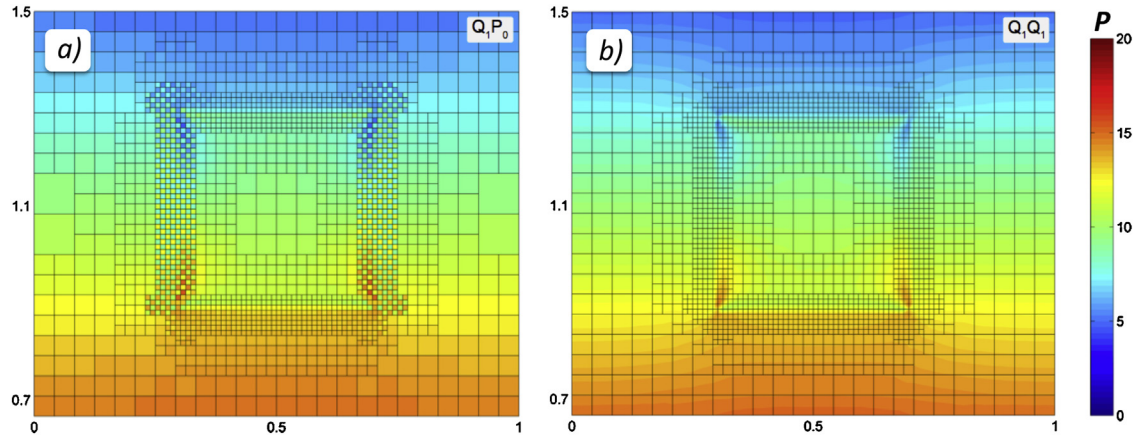

Fig. 11. Pressure distribution around a viscous sinking block obtained on an adaptive mesh using (a) an unstable element (Q_1–P_0) and (b) an LBB stable element (Q_1–Q_1). Note that the unstable element (a) produces a "checkerboard" pattern in the vicinity of the block corners where the pressure gradients are large and the mesh has been refined.

Modified from Mishin (2011).

models to couple "full erosion" surface processes with a tectonic model. In latter work (Braun and Yamato, 2010), the inclusion of visco-plastic rheology and a more complex surface process model CASCADE (Braun and Sambridge, 1997) was introduced to a study the deformation of a finite-width crustal wedge, and in Loiselet et al. (2010) the dynamics of a subducting slab were examined.

Despite the potential of the DOUAR methodology described in Braun et al. (2008), there are two factors which have hindered the approach from becoming widely used in the geodynamic community. The first issue relates to usage of the Q_1–P_0 element used to discretise Stokes equations. Fundamentally, the quality of the numerical solution for the viscous flow problem is dictated by the element used. From a practical point of view, the Q_1–P_0 element is extremely appealing to use for geodynamic applications as it permits the use of discontinuous viscosities and thus stresses (due to the piece wise constant pressure approximation) and given the low order of the velocity and pressure space, the method is fairly cheap with respect to the required memory (RAM). However, the element is mathematically proven to not satisfy the LBB stability condition (Brezzi and Fortin, 1991). In practice this means that the pressure field may exhibit spurious (unphysical) oscillations. To compound the issue, practitioners often do not see such unphysical behaviour in their results, and consequently consider that the element is "safe" to use. The appearance of the oscillation is dependent of the model setup; e.g. boundary conditions, geometry of the domain, geometry of the individual elements, viscosity structure. A priori we have no way of knowing when this element will produce spurious pressure fields, thus for general geodynamic problems the Q_1–P_0 element is not practical if reliability in the numerical result is desired. Empirically it is also known that these pressure oscillations are excited by the introduction of hanging nodes within quad-tree and oct-tree based discretisations if element subdivision occurs in a region where the pressure gradient changes. This is illustrated in Fig. 11 where we consider a dense, very viscous block of material falling through a fluid of higher density and lower viscosity. Refinement of the mesh was conducted to resolve the viscosity interface between the block and the background. It is evident in Fig. 11a) that the combination of sharp pressure variations (located at the corners of the viscous block) and the refined mesh give rise to an oscillatory pressure field.

The other main limitation of the DOUAR methodology is the usage of a penalty formulation which requires the use of a sparse direct solver. Whilst WSMP and MUMPS are both parallel algorithms, the algorithmic complexity of these factorizations is high (see Section 2.5 for more details). Consequently, increasing the resolution of the simulation, i.e. increasing the number of unknowns, will cause both the CPU time and RAM required to significantly increase. For example, a factor of four increase in the total number of unknowns will cause the solve time to increase by a factor of 16 and the memory usage to increase by a factor of 6.35.

RHEA is a scalable, massively parallel, oct-tree based adaptive FE code for solving variable viscosity Stokes flow problems in a Cartesian box (Burstedde et al., 2009) or a full spherical domain (Stadler et al., 2010) which has been specifically developed to study mantle convection processes. See Fig. 10c for examples of RHEA calculations performed in spherical geometry. RHEA employs a h-adaptivity in conjunction with a stabilised Q_1–Q_1 discretisation (Dohrmann and Bochev, 2004; Bochev et al., 2006) of the Stokes problem. Adaptivity is controlled by a collection of different error indicators including; a goal-oriented based approach which requires the solution of the adjoint equation for Stokes, a residual based indicator which utilises the complete residual of the Stokes equation, and an indicator based on the divergence of the velocity field. The discrete Stokes system is solved by exploiting the block structure associated with the partial differential equation (Elman et al., 2005) and the parallel algebraic multi-grid (AMG) preconditioner BoomerAMG (Falgout, 2006; Henson and Yang, 2002).

RHEA benefits from the fact that both the oct-tree mesh data structure and the Stokes solver can be executed on massively parallel computer systems (e.g. Ranger – 62,976 cores, located at the Texas Advanced Computing Center) and that the algorithms possess excellent parallel scalability. In Burstedde et al. (2008) it was demonstrated that their oct-tree implementation used possess excellent parallel scalability (both strong and weak) up to 62,464 cores. Furthermore, in Burstedde et al. (2009) it was established that the parallel solver employed was robust when applied to variable viscosity Stokes problems defined on a oct-tree mesh, and that the overall performance of the solver exhibited excellent parallel scalability (weak) up to 4096 cores. Furthermore, this work has established that the overhead associated with performing parallel mesh adaptation using an oct-tree description is small compared to the cost of performing a one time step of a full mantle convection simulation, which is dominated by the time required to solve the discrete (linear/non-linear) Stokes problem. In practice, RHEA was demonstrated to be able to solve full spherical models of mantle convection, with a local resolution of <1 km at the plate boundaries. These models utilised 400M – 1.2B unknowns, and using between

5000 and 7000 computational cores, required up to 9 h of CPU time to obtain a single adaptive solution to the Stokes flow problem employing a temperature and strain rate dependent (non-linear) viscosity (Stadler et al., 2010).

Many of the ideas central to the RHEA development (e.g. emphasis on variable viscosity stokes flow for convection, scalable algebraic multilevel preconditioners, parallel oct-tree data structures) have been brought together in an open source package known as ASPECT (Kronbichler et al., 2012). In Appendix A we discuss this open source project in more detail.

4. Perspectives and future directions

It is apparent that the application of adaptive finite element methods in computational geodynamics is a relatively new area of research, and thus there is much further experimentation and investigation which is required to establish the approach in the geodynamics community. The following points represent areas for further exploration in future adaptive FE developments within geodynamics.

1 The development and systematic study of error indicators for a wide range of different geodynamic applications. The overall benefits of AFE are only realised when adaptivity is employed in a manner which minimises the error of the numerical solution or physical quantity of interest. Developing a knowledge of which error indicators are appropriate for specific applications is of paramount importance to harness the full potential of AFE. First efforts to address this issue for a simulation related to a model of simplified melt migration were demonstrated in Burstedde et al. (2009) (see Figs. 6 and 7 within), in which three different error indicators were utilized and the effectiveness of adaptivity, in terms of error versus degrees of freedom and error versus total CPU time was examined. Continued efforts in this direction for problems in mantle convection, two phase flow, lithospheric dynamics, coupled models of lithospheric deformation and landscape evolution should be pursued.

2 A detailed comparison between the relative advantages of using structured versus unstructured *h*-adaptive mesh strategies has yet to be performed in the context of geodynamic applications. For example, it is not clear whether the overhead of implementing fully unstructured meshes is ameliorated by its increased flexibility. Similarly, trade-offs between computational efficiency, solution accuracy and complexity in fully parallel implementations of structured versus unstructured adaptive mesh algorithms needs to be closely examined. To some extent, this may largely hinge upon the ability to develop robust and scalable solvers for the resulting discrete problems generated on such meshes (see perspective point 5 for further details).

3 To measure the true discretisation, one of the ingredients required is an analytical solution which satisfies the PDE and boundary conditions. In Burstedde et al. (2009), an analytic solution was used in their evaluation of the error estimators. A more flexible and powerful approach to develop analytical solutions is the Method of Manufactured Solutions (MMS) (Roache, 1997, 1998). The MMS technique allows for the construction of arbitrarily complex solutions to any linear, or nonlinear PDE. This approach will prove invaluable in accessing the performance of error indicators for a wide range of geodynamic applications. Currently the MMS is almost never used in the geodynamic modelling community – in stark contrast to the applied mathematics and engineering communities. Exploiting the MMS in computational geodynamics will help to establish a higher level of robustness and reliability in both the numerical methods used, and in the verification of their implementation. Furthermore, MMS will prove useful in the development of robust error indicators for geodynamic applications.

4 The Marker-And-Cell (MAC) methodology has received wide spread acceptance as a powerful method to study large deformation processes relevant to geodynamics. This class of methods employs a mesh to discretise Stokes equations, and a set of Lagrangian particles (or markers) to represent the rock lithology. The inclusion of adaptive FE analysis mandates that the particle discretisation must also be adaptive. Further work is required to better develop the refinement and coarsening procedures which are suitable for the MAC style, particle based discretisation. Preliminary work in this direction for geodynamics was explored in Leng and Zhong (2011) in the context of the tracer-ratio method. In the context of engineering applications using the material point method, this idea is also being actively pursued (Tan and Nairn, 2002; Ma et al., 2005, 2006). Further evaluation of adaptive procedures being developed within the field of meshfree methods should be assessed.

5 The implicit solution of the linear (or non-linear) system of equations which represent the discrete Stokes operator still represents a significant fraction of the overall CPU time per time step. As reported in Burstedde et al. (2009) (see Table 2 within), even when a fully parallel oct-tree structure is utilised, the solve time consumes >95% of the total simulation time. For this reason, developing more efficient solvers will represent a continuing area of research in future AFE studies. The use of geometric multi-grid (GMG) methods on adaptive FE meshes (structured or unstructured) in geodynamics is largely unexplored. In the context of oct-tree *h*-adaptivity, a rich geometric mesh hierarchy is available which would permit the use of such an approach. Exploring the advantages of fully coupled multi-grid approaches (smoothing $\boldsymbol{u}, p$ simultaneously, e.g. Tackley, 2008b) versus block decoupled approaches (e.g. Burstedde et al., 2009) have not been systematically studied for finite element discretisations applied to geodynamic problems with large variations in viscosity.

6 Adaptive meshing techniques are of immediate importance for studying fractures and shear band formation/evolution in models of crustal and lithospheric deformation. The work of Mishin (2011) represents the first step to address the issue by performing "sandbox" inspired shear band experiments using a wavelet based adaptive FE procedure. The plasticity formulations commonly employed in geodynamics possess no intrinsic length scale, thus even in the limit of infinite mesh refinement, the numerical solution will be unresolved. In practice, the length scale of the shear band in such numerical experiments is determined by the minimum cell size. Hence, the combination of AFE techniques together with this plasticity formulation favoured by the geodynamics community raises several topics to be addressed in the future.

The lack of an inherent length scale in such plasticity formulations is well documented (Bazant and Zubelewicz, 1988; de Borst et al., 1993), however application of AFE analysis exasperated this issue due to the extremely high, local resolution permitted by mesh refinement. For one, direct application of error indicators within an adaptive procedure (see Section 2.2) will not terminate. To address this, practitioners typically specify a maximum number of refinement levels (or cycles). The required cut-off is simply to circumvent the lack of convergence in the physical system.

Real materials possesses a physical length scale determined by the underlying material microstructure (e.g. grain size), thus the thickness of the localization zone is a physical property of the material. To address this fundamental shortcoming, several alternative formulations have been developed, including the usage of a Cosserat continuum (Mühlhaus and Vardoulakis, 1987; de Borst and Sluys, 1991) and nonlocal plasticity formulations. For a comprehensive overview of nonlocal plasticity formulations we

refer to Bazant and Jirasek (2002) and Luzio and Bazant (2005) and for a comparison of different nonlocal plasticity approaches see Peerlings et al. (2001). The applicability of alternative plasticity formulations for geodynamic processes in conjunction with adaptive finite element analyses should be investigated in a systematic manner.

5. Summary

Adaptive finite element analysis represents an automated technique to locally modify the spatial discretisation used in a numerical simulation in such a manner so as to reduce the error within the numerical solution. The concept of adaptive finite element analysis has been actively developed within the engineering community since the 1980s. Within the field of computational geodynamics, this technique is relatively new, however interest in the methodology has commenced. Preliminary studies focusing on mantle convection and thermo-chemical convection all reveal that adaptive finite element methods are indeed likely to be highly beneficial in the context of numerical geodynamic simulations.

In this paper, we have provided an overview of the fundamental components of adaptive finite element analysis: namely techniques for performing spatial adaptivity and techniques for steering spatial adaptivity. The emphasis being primarily to disseminate a basic appreciation of the AFE methodology and literature from the engineering community, to the geodynamics community. In addition, we also address some of the computational issues associated with solving the discretized system of equations arising from adaptive finite element methods. We also highlighted the current state of the art research within geodynamics which has already exploited AFE technology, and furthermore, we speculate upon future research directions for AFE analysis in this field.

Lastly, students and researchers interested in exploring the applicability of the AFE method to their own geodynamic simulations are encouraged to experiment with the open source software packages listed in Appendix A.

Acknowledgments

The authors wish to thank the editor Randell Stephenson and Tim Horscroft for the invitation to write this review paper, the constructive feedback from reviewers Rhodri Davies and David Egholm. Taras V. Gerya and Paul J. Tackley are thanked for their constructive feedback and insightful comments. Alice-Agnes Gabriel is thanked for generating the unstructured quadrilateral mesh in Fig. 4b).

Appendix A. Software availability

Here we provide a brief summary of available open source software packages which are relevant to performing adaptive FE analysis in geodynamics. Whilst numerous AFE packages exist, few provided the necessary components required to solve Stokes flow. We note that due to the continued development of these packages, the facts and opinions expressed here are only relevant to the particular software release available at the time of writing.

1 Fluidity – v4.1.6: (http://amcg.ese.ic.ac.uk/Fluidity)

Fluidity is an open source, general purpose, multi-phase computational fluid dynamics code capable of numerically solving the Navier–Stokes equation and accompanying field equations on arbitrary unstructured finite element meshes in one, two and three dimensions. It is used in a number of different scientific areas including geophysical fluid dynamics, computational fluid dynamics, ocean modelling and recently extended to geodynamic processes, including mantle convection. It uses a finite element/control volume method which allows arbitrary movement of the mesh with time dependent problems, allowing mesh resolution to increase or decrease locally according to the current simulated state. It has a wide range of element choices including mixed formulations. Fluidity is parallelised using MPI and is capable of scaling to many thousands of processors. Fluidity is currently being actively developed to support a wide range of geophysical and geodynamic processes, however this work is still considered to be in a *development phase*. Nevertheless, the preliminary results presented in Davies et al. (2011) for studying thermal convection using unstructured *h* adaptivity highlight the potential of this package for studying such processes.

2 deal.II – v7.1.0: (http://www.dealii.org)

deal.II is a C++library targeted at the computational solution of partial differential equations using adaptive finite elements. It uses state-of-the-art programming techniques to offer a modern interface to the complex data structures and algorithms required for AFE analysis. The main aim of deal.II is to enable rapid development of modern finite element codes, using among other aspects adaptive meshes and a wide array of tools classes often used in finite element program. deal.II supports *hp*-adaptivity over quadrilateral and hexahedral element geometries and employs many different element types. Support for parallel linear algebra and algebraic multi-level preconditioners is provided by Trilinos (Heroux et al., 2005). Parallel support for oct-tree meshes is provided by the p4est (http://www.p4est.org) library used within Burstedde et al. (2009).

deal.II includes a huge collection of example code in the form of tutorials. Specifically, it contains a number of examples related to geodynamic applications including elasticity, two-phase flow, Stokes flow and mantle convection. Please refer to http://www.dealii.org/developer/doxygen/deal.II/group__geodynamics.html.

3 Aspect – v0.1: (http://dealii.org/aspect)

Aspect (Kronbichler et al., 2012) is a software package built on top of deal.II which has been developed to simulate thermal convection of highly viscous flow, with its primary focus towards the simulation of processes in the Earth's mantle. Aspect exploits infra-structure from deal.II and thus leverages technology such as parallel adaptive mesh refinement, linear and nonlinear solvers and stabilised discretisations for transport-dominated processes. Aspect currently supports a range of different mesh geometries (2D-box, annulus; 3D-cube, hollow sphere). The design of Aspect is intended to be extensible, so users are readily able to introduce their own material models, gravity models, post-processing routines, etc.

4 libMesh – v0.7.3.1: (http://libmesh.sourceforge.net)

The libMesh library provides a framework for the numerical simulation of partial differential equations using arbitrary unstructured discretisations on serial and parallel platforms. A major goal of the library is to provide support for adaptive mesh refinement computations in a parallel computing environment, whilst allowing the research scientist to focus on the physics they are modelling.

libMesh currently supports 1D, 2D, and 3D steady and transient simulations on a variety of popular geometric (quadrilateral, hexahedral, triangle, tetrahedra) and finite element types. *h*-adaptivity is introduced via element subdivision and *p*-adaptivity is supported. The library makes use of high-quality, existing software whenever possible. Whilst no immediately geodynamically relevant examples appear within libMesh distribution, the library offers sufficient functionality to enable users to develop mixed-finite element formulations for Stokes flow.

References

Čadek, O., Martinec, Z., Matyska, C., 1992. Spectral variational approach to the non-Newtonian stokes problem in a spherical shell. Computer Physics Communications 71, 56–70.

Ainsworth, M., Oden, J.T., 2000. A Posteriori Error Estimation in Finite Element Analysis. Wiley, New York.

Amestoy, P.R., Duff, I.S., Koster, J., L'Excellent, J.Y., 2001. A fully asynchronous multifrontal solver using distributed dynamic scheduling. SIAM Journal of Matrix Analysis and Applications 23, 15–41.

Babuska, I., Miller, A., 1987. A feedback finite element method with a posteriori error estimation: Part I. The finite element method and some basic properties of the a posteriori error estimator. Computer Methods in Applied Mechanics and Engineering 61, 1–40.

Babuska, I., Rheinboldt, W.C., 1978a. Error estimates for adaptive finite element computations. SIAM Journal of Numerical Analysis 15, 736–754.

Babuska, I., Rheinboldt, W.C., 1978b. A posteriori error estimates for the finite element method. International Journal for Numerical Methods in Engineering 12, 1597–1615.

Babuska, I., Rheinboldt, W.C., 1979. Analysis of optimal finite-element meshes in r^1. Mathematics of Computation 33, 435–463.

Bank, R.E., Weiser, A., 1985. Some a posteriori error estimators for elliptic partial differential equations. Mathematics of Computation 44, 283–301.

Barrett, R., Berry, M., Chan, T.F., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., van der Vorst, H., 1994. Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods. SIAM, Philadelphia, PA.

Baumgardner, J.R., 1985. Three-dimensional treatment of convective flow in the Earth's mantle. Journal of Statistical Physics 39, 501–511.

Bazant, Z.P., Jirasek, M., 2002. Nonlocal integral formulations of plasticity and damage: survey of progress. Journal of Engineering Mechanics 128, 1119–1149.

Bazant, Z.P., Zubelewicz, A., 1988. Strain-softening bar and beam: exact non-local solution. International Journal of Solids and Structures 7, 659–673.

Becker, R., Rannacher, R., 2001. An optimal control approach to a posteriori error estimation in finite element methods. Acta Numerica 10, 1–102.

Bercovici, D., Schubert, G., Glatzmaier, G.A., 1989. Three-dimensional spherical models of convection in the Earth's mantle. Science 244, 950–955.

Biswas, R., Devine, K.D., Flaherty, J.E., 1994. Parallel, adaptive finite element methods for conservation laws. Applied Numerical Mathematics 14, 255–283.

Bochev, P., Dohrmann, C., Gunzburger, M., 2006. Stabilization of low-order mixed finite elements for the Stokes equations. SIAM Journal on Numerical Analysis 44, 82–101.

Boroomand, B., Zienkiewicz, O.C., 1997a. An improved REP and the effectivity robustness test. International Journal for Numerical Methods in Engineering 40, 3247–3277.

Boroomand, B., Zienkiewicz, O.C., 1997b. Recovery by Equilibrium of Patches (REP). International Journal for Numerical Methods in Engineering 40, 137–164.

Braun, J., Sambridge, M., 1997. Modelling landscape evolution on geological time scales: a new method based on irregular spatial discretization. Basin Research 9, 27–52.

Braun, J., Yamato, P., 2010. Structural evolution of a three-dimensional, finite-width crustal wedge. Tectonophysics 484, 181–192.

Braun, J., Thieulot, C., Fullsack, P., DeKool, M., Beaumont, C., Huismans, R., 2008. DOUAR: a new three-dimensional creeping flow numerical model for the solution of geological problems. Physics of the Earth and Planetary Interiors 171, 76–91.

Brezzi, F., Fortin, M., 1991. Mixed and Hybrid Finite Element Methods. Springer-Verlag New York, Inc., New York, NY, USA.

Briggs, W.L., Henson, V.E., McCormick, S.F., 2000. A Multigrid Tutorial, 2nd ed. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Brooks, A.N., Hughes, T.J.R., 1982. Streamline upwind/Petrov–Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier–Stokes equations. Computer Methods in Applied Mechanics and Engineering 32, 199–259.

Buiter, S.J.H., Babeyko, A.Y., Ellis, S., Gerya, T.V., Kaus, B.J.P., Kellner, A., Schreurs, G., Yamada, Y.,2006. The numerical sandbox: comparison of model results for a shortening and an extension experiment. In: Analogue and numerical modelling of crustal-scale processes. Geological Society of London, London, pp. 29–64.

Bunge, H.P., Baumgardner, J.R., 1995. Mantle convection modeling on parallel virtual machines. Computers in Physics 9, 207–215.

Bunge, H.P., Richards, M.A., Baumgardner, J.R., 1997. A sensitivity study of three-dimensional spherical mantle convection at 10^8 Rayleigh number: effects of depth-dependent viscosity, heating mode, and an endothermic phase change. Journal of Geophysical Research 102, 11991–12007.

Burstedde, C., Ghattas, O., Gurnis, M., Stadler, G., Tan, E., Tu, T., Wilcox, L.C., Zhong, S., 2008. Scalable adaptive mantle convection simulation on petascale supercomputers. In: ACM/IEEE SC08, Austin, TX, USA, pp. 1–15.

Burstedde, C., Ghattas, O., Stadler, G., Tu, T., Wilcox, L.C., 2009. Parallel scalable adjoint-based adaptive solution of variable-viscosity Stokes flow problems. Computer Methods in Applied Mechanics and Engineering 198, 1691–1700.

Calo, V.M., Collier, N.O., Pardo, D., Paszynski, M.R., 2011. Computational complexity and memory usage for multi-frontal direct solvers used in p finite element analysis. Procedia Computer Science 4, 1854–1861.

Carstensen, C., Funken, S.A., 1999. Fully reliable localized error control in the FEM. SIAM Journal on Scientific Computing 21, 1465–1484.

Chen, Y., Davis, T.A., Hager, W.W., Rajamanickam, S., 2008. Algorithm 887: CHOLMOD, Supernodal sparse Cholesky factorization and update/downdate. ACM Transactions on Mathematical Software 35, 22:1–22:14.

Christensen, U., Harder, H., 1991. 3-D convection with variable viscosity. Geophysical Journal International 104, 213–220.

Cobbold, P.R., Percevault, M.N., 1983. Spatial integration of strains using finite elements. Journal of Structural Geology 5, 299–305 (Strain Patterns in Rocks).

Cobbold, P.R., 1977. Finite-element analysis of fold propagation – a problematic application? Tectonophysics 38, 339–353.

Crameri, F., Tackley, P.J., Meilick, I., Gerya, T.V., Kaus, B.J.P., 2012. A free plate surface and weak oceanic crust produce single-sided subduction on Earth. Geophysical Research Letters 39, L03306.

Dabrowski, M., Krotkiewski, M., Schmid, D.W., 2008. MILAMIN: MATLAB-based finite element method solver for large problems. Geochemistry, Geophysics, Geosystems 9, Q04030.

Davies, D.R., Davies, J.H., Hassan, O., Morgan, K., Nithiarasu, P., 2007. Investigations into the applicability of adaptive finite element methods to two-dimensional infinite Prandtl number thermal and thermochemical convection. Geochemistry, Geophysics and Geosystems 8, Q05010.

Davies, D.R., Davies, J.H., Hassan, O., Morgan, K., Nithiarasu, P., 2008. Adaptive finite element methods in geodynamics: convection dominated mid-ocean ridge and subduction zone simulations. International Journal of Numerical Methods for Heat & Fluid Flow 18, 1015–1035.

Davies, R., Wilson, C.R., Kramer, S.C., 2011. Fluidity: a fully unstructured anisotropic adaptive mesh computational modeling framework for geodynamics. Geochemistry, Geophysics, Geosystems 12, Q06001.

Davis, T.A., 2004. Algorithm 832: UMFPACK V4.3, An unsymmetric-pattern multifrontal method. ACM Transactions on Mathematical Software 30, 196–199.

de Borst, R., Sluys, L.J., 1991. Localisation in a Cosserat continuum under static and dynamic loading conditions. Computer Methods in Applied Mechanics and Engineering 90, 805–827.

de Borst, R., Sluys, L., Mühlhaus, H., Pamin, J., 1993. Fundamental issues in finite element analyses of localization of deformation. Engineering with Computers 10, 99–121.

Dohrmann, C., Bochev, P., 2004. A stabilized finite element method for the Stokes problem based on polynomial pressure projections. International Journal for Numerical Methods in Fluids 46, 183–201.

Elman, H., Silvester, D., Wathan, A., 2005. Finite Elements and Fast Iterative Solvers. Oxford Univ. Press, New York.

Ewing, R.E., Wang, H., 2001. A summary of numerical methods for time-dependent advection-dominated partial differential equations. Journal of Computational and Applied Mathematics 128, 423–445.

Falgout, R.D., 2006. An introduction to algebraic multigrid. Computing in Science and Engineering 8, 24–33.

Farrell, P.E., Maddison, J.R., 2010. Conservation interpolation between volume meshes by local Galerkin projection. Computer Methods in Applied Mechanics and Engineering 200, 89–100.

Forte, A.M., Peltier, W.R., 1994. The kinematics and dynamics of poloidal–toroidal coupling in mantle flow: the importance of surface plates and lateral viscosity variations. Advances in Geophysics 36, 1–119.

Fullsack, P., 1995. An arbitrary Lagrangian–Eulerian formulation for creeping flows and its application in tectonic models. Geophysical and Astrophysical Fluid Dynamics 120, 1–23.

Gee, M.W., Siefert, C.M., Hu, J.J., Tuminaro, R.S., Sala, M.G., 2006. ML 5.0 Smoothed Aggregation User's Guide. Technical Report SAND2006-2649. Sandia National Laboratories.

Gerya, T., 2009. Introduction to Numerical Geodynamic Modelling. Cambridge University Press, Cambridge.

Glatzmaier, G.A., 1988. Numerical simulations of mantle convection: time-dependent, three-dimensional, compressible, spherical shell. Geophysical and Astrophysical Fluid Dynamics 43, 223–264.

Golub, G.H., Loan, C.F.V., 1996. Matrix Computations. Johns Hopkins University Press, Baltimore, MD.

Gresho, P.M., Sani, R.L., 1998. Incompressible Flow and the Finite Element Method. Vol. 1: Advection-Diffusion and Isothermal Laminar Flow. John Wiley and Sons, Inc., New York, NY, USA.

Gupta, A., Karypis, G., Kumar, V., 1997. Highly scalable parallel algorithms for sparse matrix factorization. IEEE Transactions on Parallel and Distributed Systems 8, 502–520.

Hüttig, C., Stemmer, K., 2008. The spiral grid: A new approach to discretize the sphere and its application to mantle convection. Geochemistry, Geophysics, Geosystems 9, Q02018.

Hénon, P., Ramet, P., Roman, J., 2002. PASTIX: a high-performance parallel direct solver for sparse symmetric positive definite systems. Parallel Computing 28, 301–321.

Harder, H., Christensen, U.R., 1996. A one-plume model of Martian mantle convection. Nature 380, 507–509.

Harder, H., Hansen, U., 2005. A finite-volume solution method of thermal convection and dynamo problems in spherical shells. Geophysical and Astrophysical Fluid Dynamics 161, 522–532.

Henson, V.E., Yang, U.M., 2002. BoomerAMG: a parallel algebraic multigrid solver and preconditioner. Applied Numerical Mathematics 41, 155–177.

Hernlund, J.W., Tackley, P.J., 2003. Three-dimensional spherical shell convection at infinite Prandtl number using the 'cubed sphere' method. In: Second MIT Conference on Computational Fluid and Solid Mechanics, Elsevier, MIT.

Hernlund, J.W., Tackley, P.J., 2008. Modeling mantle convection in the spherical annulus. Physics of the Earth and Planetary Interiors 171, 48–54.

Heroux, M.A., Bartlett, R.A., Howle, V.E., Hoekstra, R.J., Hu, J.J., Kolda, T.G., Lehoucq, R.B., Long, K.R., Pawlowski, R.P., Phipps, E.T., Salinger, A.G., Thornquist, H.K., Tuminaro, R.S., Willenbring, J.M., Williams, A., Stanley, K.S., 2005. An overview of the Trilinos project. ACM Transactions on Mathematical Software 31, 397–423.

Houston, M.H., De Bremaecker, J.C., 1975. Numerical models of convection in the upper mantle. Journal of Geophysical Research 80, 742–751.

Hughes, T.J.R., Scovazzi, G., Franca, L.P., 2007. Multiscale and stabilized methods. Encyclopedia of Computational Mechanics, vol. 3. John Wiley & Sons Ltd, Baffins Lane, Chichester, UK.

Hughes, T.J.R., 1987. The Finite Element Method. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Jarvis, G.T., 1984. Time-dependent convection in the Earth's mantle. Physics of the Earth and Planetary Interiors 36, 305–327.

John, V., Schmeyer, E., 2008. Finite element methods for time-dependent convection-diffusion-reaction equations with small diffusion. Computer Methods in Applied Mechanics and Engineering 198, 475–494.

Kameyama, M., Sato, A.K.T., 2008. Multigrid-based simulation code for mantle convection in spherical shell using Yin-Yang grid. Physics of the Earth and Planetary Interiors 171, 19–32.

Kaus, B.J.P., Schmalholz, S.M., 2006. 3D finite amplitude folding: implications for stress evolution during crustal and lithospheric deformation. Geophysical Research Letters 33, L14309, http://dx.doi.org/10.1029/2006GL026341.

Kelly, D.W., Gago, J.P.D.S.R., Zienkiewicz, O.C., Babuska, I., 1983. A posteriori error analysis and adaptive processes in the finite element method: Part I—Error analysis. International Journal for Numerical Methods in Engineering 19, 1593–1619.

Kramer, S.C., Cotter, C.J., Pain, C.C., 2010. Solving the Poisson equation on small aspect ratio domains using unstructured meshes. Ocean Modelling 35, 253–263.

Kronbichler, M., Heister, T., Bangerth, W., 2012. High accuracy mantle convection simulation through modern numerical methods. Geophysical Journal International 191, 12–29.

Lan, L., Hudleston, P.J., 1991. Finite-element models of buckle folds in non-linear materials. Tectonophysics 199, 1–12.

Lechmann, S.M., May, D.A., Kaus, B.J.P., Schmalholz, S.M., 2011. Comparing thin-sheet models with 3-D multilayer models for continental collision. Geophysical Journal International 187, 10–33.

Leng, W., Zhong, S., 2011. Implementation and application of adaptive mesh refinement for thermochemical mantle convection studies. Geochemistry, Geophysics, Geosystems 12, Q04006.

Li, J., Widlund, O.B., 2007. On the use of inexact subdomain solves for BDDC algorithms. Computer Methods in Applied Mechanics and Engineering 196, 1415–1428.

Li, X.S., 2005. An overview of SuperLU: algorithms, implementation, and user interface. ACM Transactions on Mathematical Software 31, 302–325.

Loiselet, C., Braun, J., Husson, L., de Veslud, C.L.C., Thieulot, C., Yamato, P., Grujic, D., 2010. Subducting slabs: jellyfishes in the Earth's mantle. Geochemistry, Geophysics, Geosystems 11, Q08016.

Longoni, M., Malossi, A.C.I., Quarteroni, A., Villa, A., Ruffo, P., 2011. An ALE-based numerical technique for modeling sedimentary basin evolution featuring layer deformations and faults. Journal of Computational Physics 230, 3230–3248.

Lux, R.A., Davies, G.F., Thomas, J.H., 1979. Moving lithospheric plates and mantle convection. Geophysical Journal of the Royal Astronomical Society 58, 209–228.

Luzio, G.D., Bazant, Z., 2005. Spectral analysis of localization in nonlocal and over nonlocal materials with softening plasticity or damage. International Journal of Solids and Structures 42, 6071–6100.

Mühlhaus, H.B., Vardoulakis, I., 1987. The thickness of shear bands in granular materials. Geotechnique 37, 271–283.

Ma, J., Lu, H., Wang, B., Roy, S., Hornung, R., Wissink, A., Komanduri, R., 2005. Multiscale simulations using generalized interpolation material point (GIMP) method and SAMRAI parallel processing. Computer Modeling in Engineering & Sciences 8, 135–152.

Ma, J., Lu, H., Komanduri, R., 2006. Structured mesh refinement in generalized interpolation material point (GIMP) method for simulation of dynamic problems. Computer Modeling in Engineering & Sciences 12, 213–227.

Machetel, P., Thoraval, C., Brunet, D., 1995. Spectral and geophysical consequences of 3-D spherical mantle convection with an endothermic phase change at the 670 km discontinuity. Physics of the Earth and Planetary Interiors 88, 43–51.

MacLachlan, S.P., Tang, J.M., Vuik, C., 2008. Fast and robust solvers for pressure-correction in bubbly flow problems. Journal of Computational Physics 227, 9742–9761.

Mckenzie, D.P., Roberts, J.M., Weiss, N.O., 1974. Convection in the Earth's mantle: towards a numerical simulation. Journal of Fluid Mechanics 62, 465–538.

Minear, J.W., Toksöz, M.N., 1970. Thermal regime of a downgoing slab and new global tectonics. Journal of Geophysical Research 75, 1397–1419.

Mishin, Y., 2011. Adaptive multiresolution methods for problems of computational geodynamics. ETH Zürich (PhD).

Moore, D.R., Weiss, N., 1973. Two-dimensional Rayleigh–Bénard convection. Journal of Fluid Mechanics 58, 289–312.

Moresi, L., Quenette, S., Lemiale, V., Mériaux, C., Appelbe, B., Mühlhaus, H.B., 2007. Computational approaches to studying non-linear dynamics of the crust and mantle. Physics of the Earth and Planetary Interiors 163, 69–82 (Computational Challenges in the Earth Sciences).

Naimark, B.M., Ismail-Zadeh, A.T., Jacoby, W.R., 1998. Numerical approach to problems of gravitational instability of geostructures with advected material boundaries. Geophysical and Astrophysical Fluid Dynamics 134, 473–483.

Oden, J.T., Prudhomme, S., 2001. Goal-oriented error estimation and adaptivity for the finite element method. Computers and Mathematics with Applications 41, 735–756.

Parmentier, E.M., Turcotte, D.L., 1978. Two-dimensional mantle flow beneath a rigid accreting lithosphere. Physics of the Earth and Planetary Interiors 17, 281–289.

Parmentier, E.M., Turcotte, D.L., Torrance, K.E., 1975. Numerical experiments on the structure of mantle plumes. Journal of Geophysical Research 80, 4417–4424.

Peerlings, R., Geers, M., de Borst, R., Brekelmans, W., 2001. A critical comparison of nonlocal and gradient enhanced softening continua. International Journal of Solids and Structures 38, 7723–7746.

Poliakov, A., Podladchikov, Y., 1992. Diapirism and topography. Geophysical and Astrophysical Fluid Dynamics 109, 553–564.

Popov, A.A., Sobolev, S.V., 2008. Slim3d: A tool for three-dimensional thermomechanical modeling of lithospheric deformation with elasto-visco-plastic rheology. Physics of the Earth and Planetary Interiors 171, 55–75 (Recent Advances in Computational Geodynamics: Theory, Numerics and Applications).

Quinteros, J., Ramos, V.A., Jacovkis, P.M., 2009. An elasto-visco-plastic model using the finite element method for crustal and lithospheric deformation. Journal of Geodynamics 48, 83–94.

Ratcliff, J.T., Schubert, G., Zebib, A., 1995. Three-dimensional variable viscosity convection of an infinite Prandtl number Boussinesq fluid in a spherical shell. Geophysical Research Letters 22, 2227–2230.

Richter, F.M., 1973. Dynamical models for sea floor spreading. Reviews of Geophysics and Space Physics 11, 223–287.

Roache, P.J., 1997. Quantification of uncertainty in computational fluid dynamics. Annual Review of Fluid Mechanics 29, 123–160.

Roache, P.J., 1998. Verification and Validation in Computational Science and Engineering. Hermosa Publishers, Albuquerque, New Mexico.

Saad, Y., 2003. Iterative Methods for Sparse Linear Systems. SIAM, Philadelphia, PA, USA.

Schenk, O., Gartner, K., 2006. On fast factorization pivoting methods for symmetric indefinite systems. Electronic Transactions on Numerical Analysis 23, 158–179.

Schmalholz, S.M., Podladchikov, Y., 1999. Buckling versus folding: importance of viscoelasticity. Geophysical Research Letters 26, 2641–2644.

Schreurs, G., Buiter, S.J.H., Boutelier, D., Corti, G., Costa, E., Cruden, A.R., Daniel, J.M., Hoth, S., Koyi, H.A., Kukowski, N., Lohrmann, J., Ravaglia, A., Schlische, R.W., Withjack, M.O., Yamada, Y., Cavozzi, C., Delventisette, C., Brady, J.A.E., Hoffmann-Rothe, A., Mengus, J.M., Montanari, D., Nilforoushan, F., 2006. The numerical sandbox: comparison of model results for a shortening and an extension experiment. In: Analogue and numerical modelling of crustal-scale processes, Geological Society of London, London, pp. 1–27.

Schubert, G., Zebib, A., 1980. Thermal convection of an internally heated infinite Prandtl number fluid in a spherical shell. Geophysical and Astrophysical Fluid Dynamics 15, 65–90.

Schubert, G., Turcotte, D.L., Olson, P., 2001. Mantle Convection in the Earth and Planets. Cambridge University Press, Cambridge.

Stadler, G., Gurnis, M., Burstedde, C., Wilcox, L.C., Alisic, L., Ghattas, O., 2010. The dynamics of plate tectonics and mantle flow: from local to global scales. Science 329, 1033–1038.

Stemmer, K., Harder, H., Hansen, U., 2006. A new method to simulate convection with strongly temperature- and pressure-dependent viscosity in a spherical shell: applications to the Earth's mantle. Physics of the Earth and Planetary Interiors 157, 223–249.

Tabata, M., Suzuki, A., 2000. A stabilized finite element method for the Rayleigh–Bénard equations with infinite Prandtl number in a spherical shell. Computer Methods in Applied Mechanics and Engineering 190, 387–402.

Tackley, P.J., King, S.D., 2003. Testing the tracer ratio method for modeling active compositional fields in mantle convection simulations. Geochemistry, Geophysics, Geosystems 4, 8302.

Tackley, P.J., Stevenson, D.J., Glatzmaier, G.A., Schubert, G., 1993. Effects of an endothermic phase transition at 670 km depth in a spherical model of convection in the Earth's mantle. Nature 361, 699–704.

Tackley, P.J., 2008a. Geodynamics: layer cake or plum pudding? Nature Geoscience 1, 157–158.

Tackley, P.J., 2008b. Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the Yin-Yang grid. Physics of the Earth and Planetary Interiors 171, 7–18.

Tan, H., Nairn, J.A., 2002. Hierarchical, adaptive, material point method for dynamic energy release rate calculations. Computer Methods in Applied Mechanics and Engineering 191, 2123–2137.

Thieulot, C., 2011. FANTOM: Two- and three-dimensional numerical modelling of creeping flows for the solution of geological problems. Physics of the Earth and Planetary Interiors 188, 47–68.

Thoutireddy, P., 2003. Variational Arbitrary Lagrangian–Eulerian method. California Institute of Technology (PhD).

Torrance, K.E., Turcotte, D.L., 1971. Thermal convection with large viscosity variations. Journal of Fluid Mechanics 47, 113–125.

Tromp, J., Mitrovica, J.X., 1999. Surface loading of a viscoelastic earth—I. General theory. Geophysical Journal International 137, 847–855.

Trottenberg, U., Oosterlee, C.W., Schüller, A., 2001. Multigrid. Academic Press, Inc., Orlando, FL, USA.

Turcotte, D.L., Torrance, K.E., Hsui, A.T., 1973. Convection in the Earth's mantle. Methods in Computational Physics: Geophysics, vol. 13. Academic Press, Inc., pp. 431–454.
van Keken, P.E., King, S.D., Schmeling, H., Christensen, U.R., Neumeister, D., Doin, M.P., 1997. A comparison of methods for the modeling of thermochemical convection. Journal of Geophysical Research 102, 22477-–22495.
van Keken, P.E., King, S.D., Schmeling, H., Christensen, U.R., Neumeister, D., Doin, M.P., 1997. A comparison of methods for the modeling of thermochemical convection. Journal of Geophysical Research 102, 22477–22496.
Vanek, P., Mandel, J., Brezina, M., 1996. Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems. Computing 56, 179–196.
Verfürth, R., 1996. A Review of a Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques. Wiley & Teubner, Stuttgart, Germany.
Walzer, U., Hendel, R., Baumgardner, J., 2004. The effects of a variation of the radial viscosity profile on mantle evolution. Tectonophysics 384, 55–90.
Wesseling, P., 1992. An Introduction to Multigrid Methods. John Wiley & Sons, Chichester, United Kingdom.
Yoshida, M., Kageyama, A., 2004. Application of the Yin-Yang grid to a thermal convection of a Boussinesq fluid with infinite Prandtl number in a three-dimensional spherical shell. Geophysical Research Letters, 31.
Yoshida, M., Iwase, Y., Honda, S., 1999. Generation of plumes under a localized high viscosity lid in 3-D spherical shell convection. Geophysical Research Letters 26, 947–950.
Young, R.E., 1974. Finite-amplitude thermal convection in a spherical shell. Journal of Fluid Mechanics 63, 695–721.
Zhang, S., Christensen, U., 1993. Some effects of lateral viscosity variations on geoid and surface velocities induced by density anomalies in the mantle. Geophysical Journal International 114, 531–547.
Zhang, S., Yuen, D.A., 1995. The influences of lower mantle viscosity stratification on 3D spherical-shell mantle convection. Earth and Planetary Science Letters 132, 157–166.
Zhong, S., Zuber, M.T., Moresi, L., Gurnis, M., 2000. Role of temperature-dependent viscosity and surface plates in spherical shell models of mantle convection. Journal of Geophysical Research 105, 11063–11082.
Zienkiewicz, O.C., Taylor, R.L., 2000. The Finite Element Method: The Basis, vol. 1., 5th ed. Arnold, London.
Zienkiewicz, O.C., Zhu, J.Z., 1987. A simple error estimator and adaptive procedure for practical engineering analysis. International Journal for Numerical Methods in Engineering 24, 337–357.
Zienkiewicz, O.C., Zhu, J.Z., 1992a. The superconvergent patch recovery and a posteriori error estimates. Part 1: The recovery technique. International Journal for Numerical Methods in Engineering 33, 1331–1364.
Zienkiewicz, O.C., Zhu, J.Z., 1992b. The superconvergent patch recovery and a posteriori error estimates. Part 2: Error estimates and adaptivity. International Journal for Numerical Methods in Engineering 33, 1365–1382.
Zienkiewicz, O.C., Zhu, J.Z., 1992c. The superconvergent patch recovery (SPR) and adaptive finite element refinement. Computer Methods in Applied Mechanics and Engineering 101, 207–224.