Adaptive algorithms for poro-mechanics and poro-plasticity

Rita Riedlbeck

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Mathématiques et Modélisation

École doctorale Information Structures Systèmes

Unité de recherche Institut Montpelliérain Alexander Grothendieck

Algorithmes adaptatifs pour la poromécanique et la poro-plasticité

Présentée par Rita RIEDLBECK
Le 27 novembre 2017

Sous la direction de Daniele DI PIETRO et Alexandre ERN

Devant le jury composé de

Roland BECKER, Professeur des universités, Université de Pau et des Pays de l'Adour	Rapporteur
Emmanuel CREUSÉ, Professeur des universités, Université Lille 1	Rapporteur
Daniele DI PIETRO, Professeur des universités, Université de Montpellier	Directeur
Alexandre ERN, Ingénieur HDR, Ecole des Ponts ParisTech	Co-directeur
Luca FORMAGGIA, Professeur des universités, Politecnico di Milano	Examinateur
Kyrylo KAZYMYRENKO, Ingénieur, EDF R&D	Co-encadrant
Françoise KRASUCKI, Professeur des universités, Université de Montpellier	Examinatrice
Martin VOHRALÍK, Directeur de recherche, INRIA Paris	Président du jury

UNIVERSITÉ DE MONTPELLIER

Résumé

Dans cette thèse nous développons des estimations d'erreur a posteriori par équilibrage de flux pour la poro-mécanique et la poro-plasticité. En se basant sur ces estimations, nous proposons des algorithmes adaptatifs pour la résolution numérique de problèmes en mécanique des sols.

Le premier chapitre traite des problèmes en poro-élasticité linéaire. Nous obtenons une borne garantie sur l'erreur en utilisant des reconstructions équilibrées et $H(\mathrm{div})$-conformes de la vitesse de Darcy et du tenseur de contraintes mécaniques. Nous appliquons cette estimation dans un algorithme adaptif pour équilibrer les composantes de l'erreur provenant de la discrétisation en espace et en temps pour des simulations en deux dimensions. La contribution principale du chapitre porte sur la reconstruction symétrique du tenseur de contraintes.

Dans le deuxième chapitre nous proposons une deuxième technique de reconstruction du tenseur de contraintes dans le cadre de l'élasticité nonlinéaire. En imposant la symétrie faiblement, cette technique améliore les temps de calcul et facilite l'implémentation. Nous démontrons l'efficacité locale et globale des estimateurs obtenus avec cette reconstruction pour une grande classe de lois en hyperélasticité. En ajoutant un estimateur de l'erreur de linéarisation, nous introduisons des critères d'arrêt adaptatifs pour le solveur de linéarisation.

Le troisième chapitre est consacré à l'application industrielle des résultats obtenus. Nous appliquons un algorithme adaptatif à des problèmes poro-mécaniques en trois dimensions avec des lois de comportement mécanique élasto-plastiques.

Mots-clés : Algorithmes adaptatifs, estimation d'erreur a posteriori, reconstruction équilibrée de tenseur de contraintes, espace d'éléments finis de Arnold-Winther, espaces d'éléments finis de Arnold-Falk-Winther, poro-plasticité

Abstract

In this Ph.D. thesis we develop equilibrated flux a posteriori error estimates for poro-mechanical and poro-plasticity problems. Based on these estimations we propose adaptive algorithms for the numerical solution of problems in soil mechanics.

The first chapter deals with linear poro-elasticity problems. Using equilibrated $H(\mathrm{div})$-conforming flux reconstructions of the Darcy velocity and the mechanical stress tensor, we obtain a guaranteed upper bound on the error. We apply this estimate in an adaptive algorithm balancing the space and time discretisation error components in simulations in two space dimensions. The main contribution of this chapter is the symmetric reconstruction of the stress tensor.

In the second chapter we propose another reconstruction technique for the stress tensor, while considering nonlinear elasticity problems. By imposing the symmetry of the tensor only weakly, we reduce computation time and simplify the implementation. We prove that the estimate obtained using this stress reconstuction is locally and globally efficient for a wide range of hyperelasticity problems. We add a linearization error estimator, enabling us to introduce adaptive stopping criteria for the linearization solver.

The third chapter adresses the industrial application of the obtained results. We apply an adaptive algorithm to three-dimensional poro-mechanical problems involving elasto-plastic mechanical behavior laws.

Keywords: Adaptive algorithms, a posteriori error estimation, equilibrated stress tensor reconstruction, Arnold-Winther finite element space, Arnold-Falk-Winther finite element space, poroplasticity.

Remerciements

Bien evidemment, c'est à Daniele Di Pietro que vont mes premiers remerciements, aussi parce que je ne serais pas en train d'écrire ces lignes si tu m'avais pas convaincue de candidater pour cette thèse. Je te remercie pour tout ce que tu m'as appris, ta confiance, ton humour et ton soutien (et du côté scientifique et du côté humain) pendant ces dernières trois années.

Mes remerciements vont ensuite à Alexandre Ern, pour ton écoute, ta gentillesse et la facilité avec laquelle tu as répondu à mes questions, qui m'imprésionne toujours.

Un très grand merci va également à Kyryl Kazymyrenko et Sylvie Granet pour un encadrement complet, allant de la mécanique et la THM en passant par plein d'astuces pour la vie professionelle, des discussions et schémas sur l'implémentation dans Code_Aster, jusqu'au volley, la politique, le culinaire... bref, un encadrement complet que j'ai plus que apprécié.

Je tiens également à remercier Roland Becker et Emmanuel Creusé d'avoir accepté de rapporter cette thèse, ainsi que Luca Formaggia, Françoise Krasucki et Martin Vohralík pour leur présence dans le jury.

Mes remerciements vont ensuite à toutes les personnes que j'ai cotoyé et qui m'ont soutenu pendant ces trois années, en commençant par mes collègues de bureau : merci Eric pour ce bel acceuil et plein de moments inoubliables, merci Inès pour ton humour, ta sagesse et d'avoir été là dans les moments difficils, et merci Geoffroy pour ta bonne humeur et ta sérénité. Merci également à François, le meilleur partenaire sportif, pour tous les bons moments et ta sincérité. Je remercie Jacques Pellet et Mickaël Abbas pour votre aide indispensable avec Code_Aster, et Eric Lorentz pour les cours particuliers de mécanique. Et enfin je tiens à remercier tous mes autres collègues chez EDF, notamment Malika d'avoir rendu toutes les taches administratifs plus simple, et Roméo parce qu'il est un plaisir d'être doctorant(e) dans le projet Stockages.

A l'IMAG je remercie Alex, Flo, Joubine, Michele, Paul et tous les autres pour le bon acceuil qu'ils m'ont toujours fait lors de mes visites à Montpellier et les conférences passés ensembles. Bon courage à vous tous pour les thèses et/ou post-docs !

Enfin je veux remercier les filles de l'ASBAM et du PAC pour leurs amitié, les succès sportifs

et toutes les discussions autour d'un verre à l'Irish et aux Prolongations. Merci également à Caro d'avoir partagé le plus bel appartement de Paris avec moi, je me souviendrai avec plaisir de ces trois ans de coloc.

Zu guter Letzt möchte ich natürlich meinen Freunden und meiner Familie für ihre Unterstützung danken. Ich bin unglaublich froh und dankbar, dass ich mich auch nach fünfeinhalb Jahren im Ausland noch vollkommen auf euch verlassen kann und wir uns immer noch wie früher verstehen. Vielen Dank Maria, Nan, Markus, Michi, Jona, Malina, Julia und allen anderen. Und natürlich ein riesiges Danke an meine Brüder Marko und Bruno, sowie meine Eltern denen ich letztendlich alles zu verdanken habe und ohne deren Unterstützung ich nie so weit gekommen wäre.

Contents

Chapitre 1

Introduction

Contents

Le but de cette thèse est de développer des algorithmes adaptatifs pour la poro-mécanique et de les implémenter dans Code_Aster, le code éléments finis d'EDF R&D. Dans l'introduction, nous allons commencer par présenter les motivations et le contexte de la thèse, ainsi que le problème poro-mécanique. Ensuite, nous décrivons les idées et contributions principales de chaque chapitre de cette thèse. Pour conclure, nous allons donner quelques perspectives de ces travaux.

1.1 Motivations, contexte et structure du manuscrit

Nous commençons par présenter le contexte industriel de cette thèse. Ensuite nous expliquons ce que nous entendons par l'expression «algorithmes adaptatifs» et comment ceux-ci peuvent répondre aux problèmes rencontrés dans le contexte industriel. Nous résumons ensuite les résultats obtenus lors de la thèse et décrivons la structure du manuscrit.

Contexte industriel

En France, l'Agence Nationale pour la Gestion des Déchets Radioactifs (ANDRA[1]) est chargée de concevoir un centre de stockage souterrain pour des déchets dont le niveau de radioactivité et la durée de vie sont élevés. Ces déchets proviennent principalement de l'industrie éléctronucléaire. Si sa création est autorisée, ce centre industriel de stockage géologique (Cigéo[2], cf. Figure 1.1) sera construit à Bure, dans l'est de la France, où se situe depuis 2007 un laboratoire de recherche à cet effet. Il sera financé par les producteurs des déchets concernés, donc principalement par EDF, qui entretient le parc nucléaire français composé en 2017 de 58 réacteurs. EDF R&D participe donc à la recherche pour garantir la sûreté, la faisabilité et l'optimisation des installations de Cigéo.

Le contexte industriel de cette thèse est la simulation du creusement de tunnels en 3D. Il s'agit d'un problème dont la complexité vient à la fois du couplage multi-physique et des géométries complexes (voir la Figure 1.2 pour un exemple). Ces simulations sont importantes pour estimer l'endommagement du sous-sol, qui pourrait favoriser le passage de radionucléides. Le modèle poro-mécanique décrivant le comportement de la roche destinée à recevoir les déchets radioactifs prend en compte deux phénomènes physiques résultants de la structure du sol : d'une part, le comportement mécanique du squelette solide déformable ; d'autre part, le comportement hydraulique de l'eau occupant son espace poreux interstitiel. Ces deux phénomènes doivent être considérés de façon couplée, car la déformation du squelette peut influencer la pression de l'eau et, réciproquement, l'écoulement souterrain induit des déformations du squelette. On se place donc ici dans le cadre de la poro-mécanique [37].

Il est à ce jour impossible de réaliser des études poro-mécaniques avec une loi plastique sur une structure 3D complexe dans des temps de calcul raisonnables. Les deux problèmes principaux qu'EDF R&D a rencontrés pour ces simulations, et qui ont motivé cette thèse, sont les suivants :

- Le couplage des deux phénomènes entraine un problème à deux variables primales – le déplacement du squelette mécanique et la pression de l'eau – qui ont des unités et des ordres de grandeur différents. Même si dans le modèle considéré la non linéarité du système à résoudre ne provient que de la partie mécanique, la présence de la pression rend problématique la définition d'un critère de convergence purement algébrique pour le solveur de Newton.

[1] `http://www.andra.fr/`
[2] `http://www.cigeo.com/`

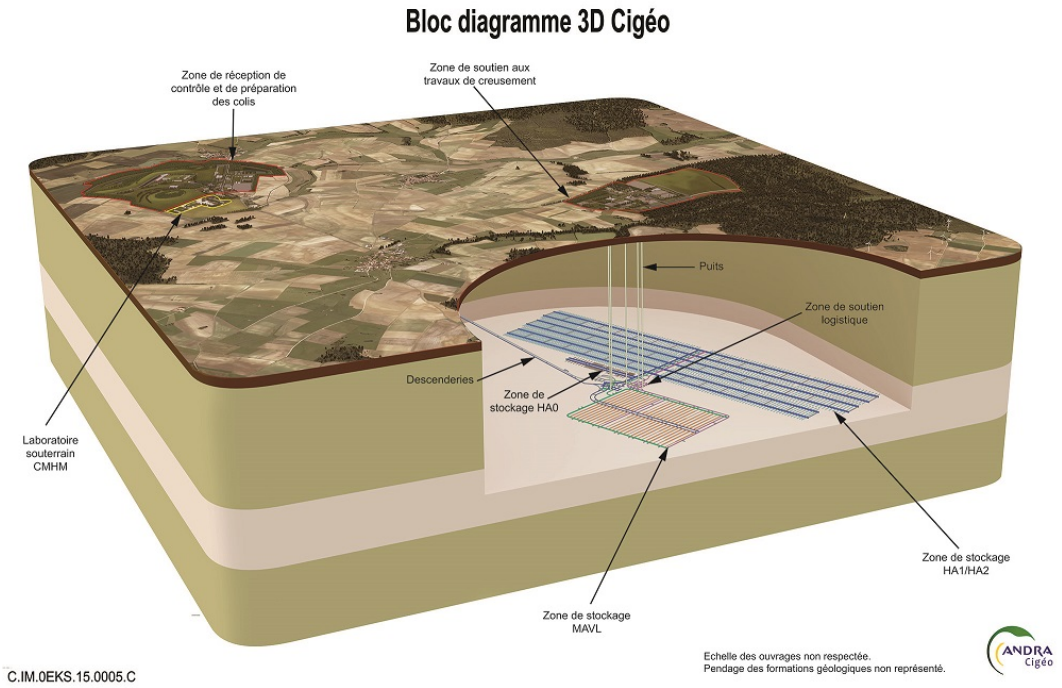


FIGURE 1.1 – Schéma du principe des installations de Cigéo, le centre industriel de stockage géologique. *Source : ANDRA*

- Dans la simulation du creusement d'un tunnel en 3D, la zone de creusement bouge. Cette zone est en général la source principale de l'erreur de discrétisation, et doit donc être maillée finement. L'utilisation d'un maillage très raffiné sur tout le chemin de creusement serait trop coûteuse : une adaptation systématique du maillage pendant le calcul est donc indispensable, non seulement pour le raffiner dans la zone où le creusement a lieu, mais également pour le déraffiner une fois que le creusement a avancé. Dans Code_Aster[3] il est possible d'utiliser l'outil de remaillage HOMARD[4], qui adapte le maillage en se basant sur un champ d'indicateurs de l'erreur qui doit être renseigné par l'utilisateur. Dans [76], le calcul d'un tel champ d'indicateurs a été développé et implémenté dans Code_Aster pour des simulations en poro-élasticité linéaire. Il existe, en effet, plusieurs lois de comportement pour décrire la relation entre le déplacement du squelette solide du sol et le tenseur de contraintes résultant, et dans certaines applications il est suffisant de considérer un comportement élastique linéaire. Toutefois, si on cherche à reproduire plus précisément le comportement de la roche, on doit avoir recours à des lois élasto-plastiques fortement non linéaires. Avant le début de cette thèse, il n'était pas possible dans Code_Aster d'obtenir un champ d'indicateurs d'erreur pour ces lois de comportement. Ces lois sont nombreuses et font l'objet de recherches continues. Dans Code_Aster elles peuvent être actualisées et enrichies à l'avenir.

[3] `http://www.code-aster.org`
[4] `http://www.code-aster.org/outils/homard/index.fr.html`

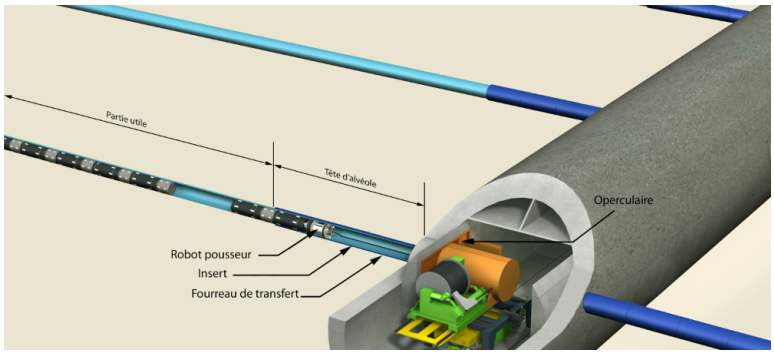


FIGURE 1.2 – Exemple de géométries complexe : tunnel et alvéoles. *Source : ANDRA*

Il est aujourd'hui possible d'effectuer des calculs 3D, soit en mécanique pure avec des lois plastiques ou visco-plastiques, soit en poro-élasticité, soit en poro-plasticité sur des géométries simples. Des études plus approfondies sont cependant nécessaires pour représenter correctement le creusement d'un tunnel et la zone fracturée qui en découle.

Algorithmes adaptatifs et estimation d'erreur *a posteriori*

L'objectif de cette thèse est de proposer une solution pour les deux problèmes ci-dessus en développant des algorithmes de résolution adaptative basés sur des estimateurs d'erreur *a posteriori* obtenus par équilibrage de flux. Le fonctionnement d'un tel algorithme est illustré dans la Figure 1.3 : l'idée derrière le terme «adaptatif» est de concentrer l'effort de calcul là où l'erreur est grande, par exemple : en utilisant des techniques de remaillage pour obtenir une distribution homogène de l'erreur de discrétisation spatiale ; en adaptant le pas de temps pour équilibrer les contributions spatiales et temporelles de l'erreur de discrétisation ; ou bien en arrêtant les solveurs itératifs linéaires et non linéaires dès que leurs contributions à l'erreur globale deviennent négligeables. Dans le contexte de la modélisation de reservoirs de pétrole, le développement et l'implémentation d'un tel algorithme a permis de reduire considérablement les temps de calcul [114].

Il est clair qu'une estimation fiable des composantes de l'erreur citées plus haut (erreur en espace, en temps, de linéarisation, etc.) est indispensable pour pouvoir les équilibrer. Contrairement aux estimations d'erreur *a priori* qui sont utilisées pour contrôler la convergence d'une méthode de discrétisation (en général en fonction de la régularité de la solution analytique), les estimations *a posteriori* donnent une borne calculable (donc notamment indépendante de la solution analytique) de l'erreur. Cette borne dépend de la solution discrète et se calcule en conséquence après (*a posteriori*) la résolution du problème discret. Il existe de nombreuses techniques d'estimation d'erreur *a posteriori*, mais, à notre connaissance, aucune n'a été appliquée à des problèmes poro-mécaniques non linéaires. Dans cette thèse nous avons choisi de nous baser sur la technique introduite dans [28, 38] pour le problème de Poisson et de l'étendre à des problèmes poro-mécaniques. Cette méthode utilise des reconstructions équili-

brées et $H(\mathrm{div})$-conformes de flux, où - en fonction du problème physique considéré - le flux satisfait une loi de type Fick et est en équilibre avec les forces extérieures. Cette technique offre plusieurs avantages qui répondent aux problèmes ci-dessus :

- L'expression de l'erreur en termes de flux permet de distinguer et, plus particulièrement, de comparer les différentes composantes de l'erreur dans l'esprit de [55]. Elle permet notamment de comparer les contributions provenant des solveurs itératifs à l'erreur de discrétisation, ce qui est indispensable pour la définition de critères d'arrêt adaptatifs pour ces solveurs. L'expression en terme de flux permet également de distinguer les deux phénomènes physiques du problème, et donc notamment de ne considérer que la partie mécanique qui induit la non linéarité quand on définit ces critères d'arrêt.

- Dans [27], Braess et al. démontrent que – dans le cadre des méthodes d'éléments finis conformes en 2D – l'estimation d'erreur est efficace et p-robuste. C'est-à-dire que l'estimation multipliée par une constante est une borne inférieure pour l'erreur, et que cette constante est indépendante du degré polynomial de la discrétisation. La p-robustesse a ensuite été démontrée pour une variété de méthodes de discrétisations dans [56], et pour différentes applications linéaires en 2D comme l'élasticité dans [48], le problème

```
1  n = 1; t = τ_n = τ_0                                  % Initialisation
   while t < t_F do                                      % Time loop
       k = 0
       while not(η_lin << η_tm + η_sp) do                % Linearization loop
           k = k + 1
6          solve linear algebraic system at iteration k
           calculate η_lin, η_tm, η_sp
       end while                                         % End of lin. loop

       if η_tm >> η_sp
11         τ_n = τ_n*0.5                                 % Space and time
       if η_tm << η_sp                                   % error balancing
           τ_n = τ_n*2
       if spatial estimator scales vary considerably
           remesh
16
       if η_lin+η_tm+η_sp <= crit_error
           n = n + 1                                     % If total estimate is too big
           τ_n = τ_n-1                                   % repeat time step
           t = t + τ_n
21
   end while                                             % End of time loop
```

FIGURE 1.3 – Schéma d'un algorithme adaptatif avec un critère d'arrêt pour le solveur de Newton (ligne 4) et l'adaptation de la discrétisation (lignes 10 – 15) en fonction des estimateurs η_{lin}, η_{sp} et η_{tm} qui dénotent respectivement l'estimateur de l'erreur de linéarisation et de la discrétisation spatiale et temporelle. t et τ_n sont l'instant et le pas de temps.

de Stokes dans [33], et des problèmes paraboliques linéaires avec une discrétisation en temps par la méthode de Galerkin discontinue dans [53]. La p-robustesse de problèmes en trois dimensions a été traitée dans [57].

- Les reconstructions des flux ne dépendent pas de la relation entre le flux et la variable primale. L'implémentation de la méthode est donc indépendante de la loi de comportement, et se prête en conséquence à l'application dans des codes offrant un grand choix de telles lois, comme c'est le cas dans Code_Aster.

Développements techniques effectués dans cette thèse

Dans les problèmes poro-mécaniques considérés dans cette thèse apparaissent deux flux : la vitesse de l'eau pour la partie hydraulique et le tenseur de contraintes pour la partie mécanique. Pour obtenir une reconstruction équilibrée du flux hydraulique nous nous basons sur [28, 38]. La contribution principale de cette thèse est de développer des méthodes de reconstruction du tenseur de contraintes en prenant en compte la symétrie de ce tenseur. Pour des calculs en 2D nous proposons deux techniques différentes, en imposant la symétrie fortement ou faiblement. Une comparaison sur un problème en élasticité linéaire montre que les deux méthodes produisent des résultats similaires. Pour les simulations en 3D, nous nous focalisons sur la deuxième variante, qui est plus facile à implémenter et moins coûteuse en termes de temps de calcul.

Nous avons progressivement intégré le calcul de ces reconstructions de flux discrets et des estimateurs d'erreur dans le logiciel industriel Code_Aster, et avons obtenu les résultats suivants en réponse aux problèmes qui ont motivé cette thèse :

- **Critère de convergence :** Comme mentionné ci-dessus, l'estimation de l'erreur en termes de flux permet de séparer la partie hydraulique et la partie mécanique du problème. Nous pouvons ainsi définir des estimateurs de l'erreur de discrétisation pour les deux parties, et de l'erreur de linéarisation pour la partie mécanique (puisque nous ne considérons que des lois hydrauliques de Darcy linéaires). En arrêtant le solveur de Newton dès que l'erreur provenant de la linéarisation est négligeable par rapport à celui de la discrétisation de la partie mécanique, nous obtenons un critère de convergence fiable qui prend en compte la physique du problème et qui, de plus, permet d'éviter des itérations inutiles.

- **Efficacité et remaillage :** Pour des problèmes en hyperélasticité, nous démontrons que l'estimation d'erreur *a posteriori* développée pendant cette thèse est efficace. De plus, nous avons vérifié sur des tests numériques avec des solutions analytiques que les estimateurs réflètent bien la distribution de l'erreur de discrétisation sur le maillage. Ces propriétés sont importantes pour fournir à HOMARD des estimateurs adaptés au problème de poro-mécanique non linéaire, et pour garantir que les maillages soient bien

adaptés au problème à chaque pas de calcul. En raffinant les maillages en fonction de la distribution des estimateurs, nous avons obtenu des ordres de convergence (en termes d'erreur en fonction du nombre de degrés de liberté) plus élevés qu'avec un raffinement uniforme.

- **Application directe à différentes lois de comportement mécanique :** Dans Code_Aster, nous avons implémenté l'outil de reconstruction locale de façon à ce qu'il puisse être appliqué à toutes les lois de comportement utilisées au sein du formalisme poro-mécanique.

Structure du manuscrit

La structure du manuscrit repose sur l'ordre chronologique des étapes de la thèse. Nous commençons au Chapitre 2 par développer une estimation d'erreur *a posteriori* pour des problèmes poro-élastiques linéaires en 2D. La reconstruction du flux hydraulique est celle présentée pour des problèmes de Darcy, que nous adaptons à la reconstruction du tenseur symétrique des contraintes en utilisant les espaces d'éléments finis mixtes d'Arnold–Winther qui imposent la symétrie du tenseur fortement. Dans le Chapitre 3, nous nous focalisons sur la partie mécanique et considérons l'élasticité non linéaire. Nous présentons également une nouvelle stratégie de reconstruction du tenseur de contraintes en utilisant les éléments d'Arnold–Falk–Winther, qui offrent des avantages d'implémentation en anticipant le cas 3D. Nous démontrons que, sous certaines hypothèses sur la loi de comportement, l'estimation multipliée par une constante est également une borne inférieure de l'erreur. Enfin, dans le Chapitre 4, nous revenons au problème poro-mécanique et considérons d'un point de vue numérique des lois de comportement mécanique élasto-plastiques en 3D.

Dans la partie restante de ce chapitre d'introduction, nous allons d'abord présenter brièvement les équations qui décrivent le comportement des sols. Dans les trois sections suivantes, nous résumons les Chapitres 2, 3, et 4. Nous finissons par donner quelques perspectives à nos travaux.

1.2 Le problème de Biot

Le problème de Biot modélise des milieux poreux constitués d'une part de grains solides formant un squelette déformable et un réseau de pores connectés entre eux, et d'autre part d'un fluide occupant ces pores. Plusieurs matériaux présentent ce type de structure, par exemple le béton, certains matériaux organiques comme le bois ou – ce qui nous intéresse dans cette thèse – les géomatériaux comme les sols et les roches. Le modèle quasi-statique est basé sur deux principes physiques : l'équilibre mécanique et la conservation de la masse du fluide. Soit Ω un domaine polyédrique de $\mathbb{R}^d$, $d \in \{2, 3\}$, occupé par le milieu poreux déformable, et soit $(0, t_F)$ avec $t_F > 0$ l'intervalle de temps considéré.

Equilibre mécanique

Nous commençons par considérer la partie mécanique. Dans toute la suite, nous allons nous restreindre au cas des petites déformations sans forcément le rappeler à chaque fois. Etant donné une force volumique $\underline{f} : \Omega \times (0, t_F) \to \mathbb{R}^d$ (souvent $\underline{f} = \rho_{\text{ref}} \underline{F}^m$, où ρ_{ref} est la masse volumique homogénéisée et $\underline{F}^m$ la force de gravité), la première équation du modèle de Biot décrit l'équilibre entre le tenseur de contraintes $\underline{\underline{\sigma}} : \Omega \times (0, t_F) \to \mathbb{R}^{d\times d}_{\text{sym}} = \{\underline{\underline{\tau}} \in \mathbb{R}^{d\times d};\ \underline{\underline{\tau}}^T = \underline{\underline{\tau}}\}$ et cette force volumique :

$$\underline{\nabla} \cdot \underline{\underline{\sigma}} + \underline{f} = \underline{0}. \tag{1.1}$$

Pour décrire le comportement mécanique d'un milieu poreux, l'hypothèse de Terzaghi [110] permet de décomposer le tenseur de contraintes comme la somme du tenseur de *contraintes effectives* $\underline{\underline{\sigma}}'$ induites par les déformations $\underline{\underline{\varepsilon}} : \Omega \times (0, t_F) \to \mathbb{R}^{d\times d}_{\text{sym}}$ de la structure solide, et du tenseur de *contraintes de pression* $\underline{\underline{\sigma}}_p$ induites par la pression $p : \Omega \times (0, t_F) \to \mathbb{R}$ du fluide :

$$\underline{\underline{\sigma}} = \underline{\underline{\sigma}}' + \underline{\underline{\sigma}}_p.$$

De ce fait, $\underline{\underline{\sigma}}$ est aussi appelé tenseur de *contraintes totales*. Les contraintes de pression dépendent linéairement et de manière isotrope de la pression. En introduisant le coefficient de Biot-Willis $b > 0$, elles s'écrivent

$$\underline{\underline{\sigma}}_p = -bp\underline{\underline{I}}_d.$$

Les contraintes effectives, quant à elles, sont exprimées en fonction de la loi de comportement mécanique. Dans le cas le plus simple de l'élasticité linéaire elles correspondent au comportement élastique formulé par la loi de Hooke

$$\underline{\underline{\sigma}}' = 2\mu\underline{\underline{\varepsilon}} + \lambda \operatorname{tr}(\underline{\underline{\varepsilon}})\underline{\underline{I}}, \qquad \underline{\underline{\varepsilon}} = \frac{1}{2}(\underline{\underline{\nabla}}\underline{u} + \underline{\underline{\nabla}}\underline{u}^T), \tag{1.2}$$

où $\underline{u} : \Omega \times (0, t_F) \to \mathbb{R}^d$ est le déplacement du squelette solide, et les paramètres de Lamé $\mu > 0$ et $\lambda \geqslant 0$ décrivent les propriétés mécaniques du matériau.

Dans le Chapitre 3 nous considérons également deux autres lois de comportement décrivant des problèmes hyperélastiques : le modèle de Hencky-Mises et un modèle d'endommagement isotrope réversible. Les deux sont des variantes non linéaires de (1.2) ; dans le modèle de Hencky-Mises, les paramètres de Lamé ne sont plus constants, mais des fonctions scalaires de la partie déviatorique des déformations $\operatorname{dev}(\underline{\underline{\varepsilon}}) = \operatorname{tr}(\underline{\underline{\varepsilon}}^2) - \frac{1}{d}\operatorname{tr}(\underline{\underline{\varepsilon}})^2$. Dans le modèle d'endommagement, on introduit une fonction d'endommagement $D : \mathbb{R}^{d\times d}_{\text{sym}} \to \mathbb{R}$ et on multiplie l'expression de $\underline{\underline{\sigma}}'$ dans (1.2) par $1 - D(\underline{\underline{\varepsilon}})$.

Enfin, nous traitons des lois élasto-plastiques dans le Chapitre 4. Ces lois combinent deux comportements différents : la réaction élastique d'un matériau tant que les contraintes n'atteignent pas un certain critère ; et une déformation plastique, donc irréversible, si le critère est atteint. Nous développons des expressions pour la loi de comportement en fonction de ce cri-

tère de charge en prenant également en compte des effets d'écrouissage, qui apparaissent si des déformations plastiques préalables ont modifié ce critère par une modification de la structure microscopique du matériau. Nous considérons deux lois de comportement mécanique utilisées dans la mécanique de sols : d'une part la loi de Drucker-Prager [49], et d'autre part la loi L&K développée au sein d'EDF [46].

Conservation de la masse du fluide

La deuxième équation du modèle de Biot exprime la conservation de la masse $m : \Omega \times (0, t_F) \to \mathbb{R}$ du fluide en postulant l'équilibre entre la variation de la masse et le flux massique $\underline{\Phi} : \Omega \times (0, t_F) \to \mathbb{R}^d$:

$$\partial_t m + \nabla \cdot \underline{\Phi} = 0. \tag{1.3}$$

La masse du fluide est le produit de la masse volumique ρ du fluide et de la porosité lagrangienne φ du squelette, qui correspond, dans le cas saturé, à la fraction du volume occupé par le fluide, donc au volume du fluide. La variation de la masse volumique du fluide peut être exprimée en fonction de sa pression par

$$\partial_t \rho = \rho \frac{\partial_t p}{K_w},$$

où K_w est la compressibilité du fluide. Le changement du volume des pores est relié à la variation de la déformation du squelette par la relation suivante :

$$\partial_t \varphi = b \partial_t \nabla \cdot \underline{u}.$$

Pour la variation de la masse, on obtient donc

$$\partial_t m = \partial_t(\rho\varphi) = \varphi \partial_t \rho + \rho \partial_t \varphi = \rho(c_0 \partial_t p + b \partial_t \nabla \cdot \underline{u}), \tag{1.4}$$

où le paramètre $c_0 = \varphi K_w^{-1}$ mesure la quantité de fluide qui peut être forcée dans l'espace poreux. Si le fluide considéré est incompressible, on a $K_w^{-1} = 0$, et donc la variation de sa masse ne dépend pas de la pression, mais uniquement de la variation du volume des pores qu'il occupe. Nous supposons ici que le flux hydraulique $\underline{\phi} : \Omega \times (0, t_F) \to \mathbb{R}^d$ suit la loi de Darcy avec une conductivité hydraulique $\kappa : \Omega \to \mathbb{R}$ donnée. On peut donc l'écrire

$$\underline{\phi}(p) = -\kappa \underline{\nabla} p.$$

La conductivité hydraulique est donnée par $\kappa = K_{\text{int}} K_{\text{visc}}^{-1}$, où $K_{\text{int}} : \Omega \to \mathbb{R}_+^*$ et $K_{\text{visc}} \in \mathbb{R}_+^*$ dénotent respectivement la perméabilité intrinsèque et la viscosité du fluide. Pour le flux massique on obtient donc

$$\underline{\Phi} = \rho\kappa(-\underline{\nabla} p + \rho \underline{F}^m). \tag{1.5}$$

Plus généralement, on peut considérer des coefficients de perméabilité qui sont des tenseurs d'ordre deux, ce qui permet de modéliser l'anisotropie du sol. En combinant (1.4) et (1.5), on peut réécrire (1.3) comme

$$\partial_t(c_0 p + b\nabla \cdot \underline{u}) + \nabla \cdot \underline{\phi}(p) = g, \tag{1.6}$$

où $g = -\nabla \cdot (\kappa\rho \underline{F}^m)$ joue le rôle d'une source volumique de fluide.

Formulation faible

Pour la résolution des équations (1.1) et (1.6) dans un domaine Ω – complétées par des conditions au bord et initiales – nous supposons que les termes source $\underline{f}$ et g sont à presque tout temps $t \in (0, t_F)$ des fonctions de carré intégrable dans Ω, et que la norme correspondante dans Ω est également de carré intégrable dans l'intervalle de temps. Pour écrire la formulation variationelle nous supposons qu'à presque tout instant de temps les variables primales $\underline{u}(\cdot, t)$ et $p(\cdot, t)$ vivent dans les espaces de Sobolev $\underline{H}^1(\Omega)$ et $H^1(\Omega)$ respectivement, et que les fonctions $t \mapsto \underline{u}(\cdot, t)$ et $t \mapsto p(\cdot, t)$ admettent une dérivée au sens faible. On cherche donc $\underline{u} \in H^1(0, t_F; \underline{H}^1(\Omega))$ et $p \in H^1(0, t_F; H^1(\Omega))$ vérifiant les conditions initiales, telles que pour presque tout $t \in (0, t_F)$ et pour toutes fonctions test $\underline{v} \in L^2(0, t_F; \underline{H}^1(\Omega))$ et $q \in L^2(0, t_F; H^1(\Omega))$,

$$\int_\Omega \underline{\underline{\sigma}}(\underline{\underline{\varepsilon}}(\underline{u}), p) : \underline{\underline{\varepsilon}}(\underline{v}) \, \mathrm{d}\underline{x} = \int_\Omega \underline{f} \cdot \underline{v} \, \mathrm{d}\underline{x}, \tag{1.7a}$$

$$\int_\Omega \partial_t(c_0 p + b\nabla \cdot \underline{u}) q \, \mathrm{d}\underline{x} - \int_\Omega \underline{\phi}(p) \cdot \underline{\nabla} q \, \mathrm{d}\underline{x} = \int_\Omega g q \, \mathrm{d}\underline{x}. \tag{1.7b}$$

En regardant la première intégrale dans (1.7b) on voit que, contrairement à $\underline{u}$ et p, il n'est pas nécessaire que les fonctions test soient dérivables en temps.

Le tenseur de contraintes $\underline{\underline{\sigma}} = \underline{\underline{\sigma}}(\underline{\underline{\varepsilon}}(\underline{u}), p)$ et la vitesse $\underline{\phi} = \underline{\phi}(p)$ résultants de (1.7) vérifient deux propriétés importantes : premièrement, ils admettent pour presque tout $t \in (0, t_F)$ une divergence au sens faible. On a donc

$$\underline{\underline{\sigma}} \in \underline{\underline{H}}_{\mathrm{s}}(\mathrm{div}, \Omega) = \{\underline{\underline{\tau}} \in \underline{\underline{L}}^2(\Omega);\ \underline{\nabla} \cdot \underline{\underline{\tau}} \in \underline{L}^2(\Omega) \text{ et } \underline{\underline{\tau}} \text{ est symétrique}\},$$
$$\underline{\phi} \in \underline{H}(\mathrm{div}, \Omega) = \{\underline{\phi} \in \underline{L}^2(\Omega);\ \nabla \cdot \underline{\phi} \in L^2(\Omega)\}.$$

Deuxièmement, ils satisfont les deux équations d'équilibre (1.1) et (1.6). Ces deux propriétés peuvent être facilement vérifiées, en rappelant qu'une fonction $w \in L^2(\Omega)$ est la divergence au sens faible de $\underline{v} \in \underline{L}^2(\Omega)$ si $\int_\Omega w\psi \mathrm{d}\underline{x} = -\int_\Omega \underline{v} \cdot \underline{\nabla}\psi \mathrm{d}\underline{x}$ pour toute $\psi \in \mathcal{D}(\Omega) \subset H_0^1(\Omega)$. Le tenseur $\underline{\underline{\sigma}}$ étant symétrique, on peut remplacer $\underline{\underline{\varepsilon}}(\underline{v})$ par $\underline{\underline{\nabla}}\underline{v}$ dans (1.7a), et on voit que $-\underline{f} \in \underline{L}^2(\Omega)$ et $g - \partial_t(c_0 p + \nabla \cdot \underline{u}) \in L^2(\Omega)$ vérifient cette définition.

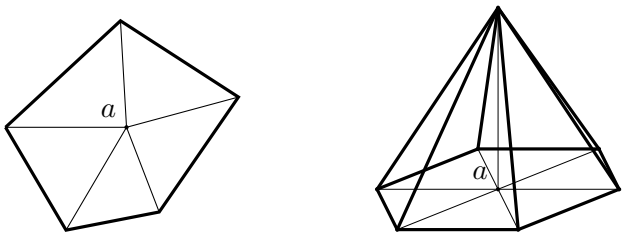


FIGURE 1.4 – Exemple d'un patch autour d'un noeud a du maillage *(gauche)*, et la fonction chapeau d'un noeud *(droite)*.

1.3 Reconstruction de flux et estimateurs d'erreur pour le problème poro-élastique de Biot en 2D

Dans cette section nous décrivons de façon plus détaillée la stratégie de reconstruction de flux. Nous expliquons ensuite comment ces développements ont été implémentés dans Code_Aster. L'évolution dans l'implémentation, qui sera de plus en plus intégrée dans la structure de Code_Aster au fur et à mesure des chapitres de la thèse, est explicitée dans chacune des sections correspondantes de l'introduction.

La stratégie de reconstruction des flux

Lorsqu'on discrétise en espace le problème (1.7) par la méthode des éléments finis de Lagrange (H^1-conformes), les variables primales discrètes vérifient de nouveau $\underline{u}_h \in \underline{H}^1(\Omega)$ et $p_h \in H^1(\Omega)$ à chaque étape de calcul dans l'algorithme de la Figure 1.3, tandis que les flux $\underline{\underline{\sigma}}(\underline{\underline{\varepsilon}}(\underline{u}_h), p_h)$ et $\underline{\phi}(p_h)$ calculés en fonction de $\underline{u}_h$ et p_h ont en général des composantes normales discontinues à travers les faces du maillage utilisé pour la discrétisation, et ne vérifient pas les équations (1.1) et (1.6). Cette représentation non-physique des contraintes et de la vitesse du fluide sont la première raison pour laquelle nous souhaitons calculer des reconstructions de flux, le but étant d'obtenir des flux plus physiques (donc $H(\mathrm{div})$-conformes et équilibrés) et proches des flux discrets. La deuxième (et principale) motivation est l'utilisation de telles reconstructions de flux pour calculer des estimations d'erreur avec les avantages décrits en Section 1.1.

Pour la partie hydraulique, nous nous basons sur la reconstruction de flux introduite dans [28,38], en utilisant la formulation de [56]. L'idée intuitive constiste à chercher dans un espace discret $H(\mathrm{div})$-conforme approprié la fonction la plus proche du flux discret $\underline{\phi}(p_h)$ qui vérifie l'équilibre. Il s'agit donc de résoudre un problème de minimisation sous contrainte. Effectuer à chaque pas de temps ce calcul global sur tout Ω serait trop coûteux ; on se base donc sur des patchs autour des sommets du maillage (on rappelle qu'un patch est l'ensemble des mailles partageant un sommet donné) pour y résoudre le même type de problème de minimisation, et on additionne ensuite les solutions afin d'obtenir un champ global. Pour que cette somme approche le flux discret, on multiplie $\underline{\phi}(p_h)$ par la fonction chapeau du sommet, illustrée dans la Figure 1.4. Sur le bord des patchs on impose des conditions de Neumann homogènes pour pouvoir prolonger la solution par zéro sur le reste du domaine sans introduire des discontinuités

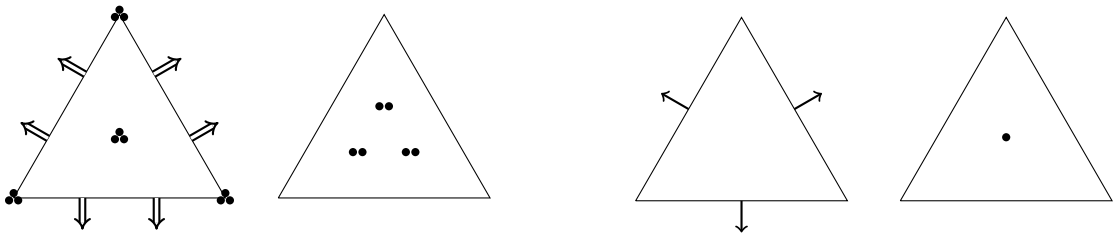


FIGURE 1.5 – Diagrammes des éléments finis d'Arnold–Winther et de Raviart–Thomas de degré le plus bas

à travers les bords du patch. De telle manière, la somme des solutions locales est de nouveau $H(\mathrm{div})$-conforme. L'espace discret utilisé est l'espace de Raviart-Thomas de plus bas degré, dont les degrés de liberté sont représentés dans la partie droite de la Figure 1.5 avec ceux de l'espace utilisé pour imposer la divergence du flux.

La contribution principale de ce chapitre est d'appliquer le même principe à la partie mécanique pour obtenir une reconstruction équilibrée, $H(\mathrm{div})$-conforme et symétrique du tenseur de contraintes totales. Les espaces discrets de tenseurs symétriques et $H(\mathrm{div})$-conformes sont les espaces d'Arnold-Winther introduits dans [10] pour les maillages triangulaires et dans [8] pour des maillages de tétraèdres en 3D. Dans l'élément triangulaire, il existe trois groupes de degrés de liberté (voir la partie gauche de la Figure 1.5). Le premier contient les moments jusqu'à un certain degré polynomial des composantes normales à travers les arêtes pour garantir leur continuité. Dans le deuxième groupe se trouvent les valeurs des composantes aux sommets du triangle pour garantir la symétrie du tenseur. Le troisième groupe de degrés de liberté contient les moments à l'intérieur du triangle de chaque composante.

En utilisant les propriétés des reconstructions de flux ainsi obtenues, nous pouvons établir une borne supérieure garantie (donc sans constantes inconnues) sur la norme duale du résidu de la formulation faible du problème de Biot. Cette mesure d'erreur se prête naturellement au développement d'estimateurs d'erreur de type flux équilibrés. Pour pouvoir bien définir le résidu et pour comparer les composantes de l'estimation provenant des parties hydraulique et mécanique du problème, nous avons introduit des paramètres d'adimensionnement. Ces paramètres n'interviennent que dans la phase de calcul d'estimateurs : le calcul initial de la solution discrète reste inchangé.

Mise en œuvre dans Code_Aster

Dans un premier temps, les reconstructions et le calcul des estimateurs d'erreur ont été implémentés de façon provisoire dans Code_Aster. Ce code développé par EDF R&D est principalement écrit en Fortran90 (initialement Fortran77), completé par des fonctions C pour réaliser quelques tâches impossibles en Fortran77 (comme l'allocation dynamique), et par des catalogues Python qui gèrent d'une part la communication entre l'utilisateur et le code et d'autre part les différentes options de calculs élémentaires. Les paramètres d'entrée pour réa-

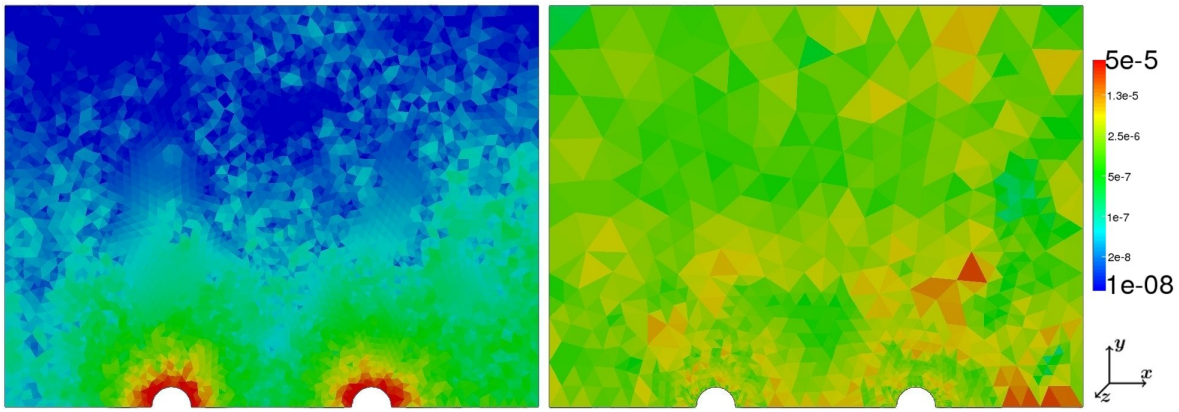


FIGURE 1.6 – Comparaison de la distribution des estimateurs de discrétisation spatiale sans et avec remaillage adaptatif (cf. Figure 2.10)

liser une étude avec Code_Aster sont un maillage et un fichier de commandes s'appuyant sur le langage Python, qui enchaîne des commandes pour la lecture du maillage, la définition du problème physique étudié, sa résolution numérique et le post-traitement. Pendant la résolution du problème exécutée par la commande `STAT_NON_LINE`, le flux hydraulique et le tenseur de contraintes sont sauvegardés à chaque étape du calcul dans une structure de données. Nous avons créé une commande de post-traitement réutilisant cette structure de données. A chaque pas de temps, elle calcule les estimateurs d'erreur de discrétisation spatiale et temporelle en fonction des flux discrets et les sauvegarde dans la même structure de données. Les termes sources $\underline{f}$ et g, ainsi que les conditions aux bords, doivent être renseignés dans les sources Fortran.

Nous avons validé les estimateurs sur deux cas tests avec une solution analytique, dont un test purement mécanique. Ensuite, nous avons effectué deux tests d'application industrielle : le quart de five-spot simulant l'injection et la production d'eau dans deux points distincts, et la simulation d'excavations de tunnels dans le contexte du stockage des déchets radioactifs (cf. Figure 1.6). Pour ces deux tests industriels nous avons écrit des fichiers de commandes mettant en œuvre un algorithme adaptatif. Puisque le calcul des estimateurs est effectué dans une commande séparée de la commande `STAT_NON_LINE`, nous utilisons une boucle qui, pour chaque pas de temps, résout le problème en prenant le résultat de l'instant de calcul précédent comme donnée initiale, estime ensuite l'erreur, et utilise cette estimation pour adapter le pas de temps et le maillage. Dans la Figure 1.6, nous comparons la distribution des estimateurs de discrétisation spatiale de deux calculs – l'un avec un maillage spatial et un pas de temps fixé durant le calcul (à gauche), et l'autre en adaptant la discrétisation en fonction des estimateurs de l'erreur spatial et temporel (à droite). On voit que la distribution de l'erreur est significativement plus homogène si on utilise l'algorithme adaptatif, et nous observons pour les deux tests industriels que, à erreur donné, le nombre de degrés de liberté à erreur donnée peut être réduit considérablement en utilisant un algorithme adaptatif.

Cependant, la reconstruction des flux – et notamment du tenseur de contraintes – coûte cher en termes de temps de calcul. Ceci est d'une part dû au fait qu'on compare le temps de reconstruction avec celui de la résolution du problème initial par Code_Aster qui est, par le caractère industriel du logiciel, améliorée et optimisée depuis plusieurs années. Mais ces longs temps de calcul s'expliquent surtout par la taille et la nature des problèmes locaux à résoudre : les éléments d'Arnold–Winther sont très riches et, contrairement à d'autres familles d'éléments finis mixtes, l'hybridation des problèmes locaux n'est pas possible, à cause des degrés de liberté sur les sommets des éléments. Cette technique est utilisée pour «transformer» le problème de type point selle résultant de l'utilisation de méthodes mixtes en un problème défini positif, ce qui rend la résolution du problème plus facile et rapide en pratique. Une possibilité afin d'améliorer les temps de calcul est donc de relâcher les contraintes de symétrie, ce qui sera discuté dans la section suivante.

Bibliographie

Nous terminons cette section avec un récapitulatif bibliographique. Le problème de Biot a été proposé par Biot [19] et von Terzaghi [110]. Ženíšek [111] a démontré qu'il est bien posé (voir aussi [98]). La discrétisation H^1-conforme utilisée dans Code_Aster utilise les éléments de Taylor-Hood [102], qui ont initialement été introduits pour le problème de Navier-Stokes. Son analyse d'erreur *a priori* est discutée dans [77–79]. Une analyse d'erreur *a posteriori* pour le problème poro-élastique linéaire a été effectué dans [52, 76], en utilisant une approche d'estimation d'erreur par résidus [12, 13, 104], donc en mesurant les sauts des flux discrets.

A notre connaissance, il n'existe pas d'estimation par équilibrage de flux pour le problème de Biot antérieure à celle proposée dans ce manuscrit. Ces méthodes ont été développées à partir de travaux de Ladevèze [69, 70], qui s'appuyent sur le théorème de Prager et Synge [89]. L'idée est de décomposer l'erreur en utilisant une reconstruction de flux qui est plus régulière que le flux discret et qui permet donc de remplacer la solution analytique inconnue dans l'erreur par des termes connus. Plusieurs méthodes pour obtenir une telle reconstruction de flux en résolvant des problèmes de Neumann sur chaque élément ont été développées dans [2, 4, 30, 71, 73, 84] et les références qui y sont contenues. Comme mentionné ci-dessus, nous allons utiliser des problèmes locaux de Neumann sur des patchs, ce qui a été proposé dans [28, 38, 48, 56, 75].

Il existe de nombreuses autres méthodes pour obtenir une estimation de l'erreur, une vue d'ensemble peut être trouvée dans [2, 3, 31, 32]. Quelques exemples hormis les estimateurs par résidus et équilibrage de flux sont la méthode de résidus équilibrés [1, 3, 71], les estimateurs fonctionels [81, 93], les techniques de lissage [116–118], ou les méthodes hiérarchiques [3, 14]. Il existe également des méthodes pour estimer l'erreur dans une certaine fonctionelle en utilisant des techniques de dualité, par exemple dans [16, 51, 83].

Dans le cas des problèmes non stationnaires, il est possible de distinguer les erreurs de dis-

crétisation en espace et en temps, et d'adapter le pas de temps et le maillage (cf. Figure 1.3) pour équilibrer ces deux sources d'erreur, comme proposé dans [18, 54, 72, 106].

1.4 Reconstruction de contraintes et estimateurs d'erreur pour l'élasticité non linéaire

Comme nous l'avons vu, l'originalité de cette thèse est le développement de reconstructions équilibrées du tenseur de contraintes mécaniques pour obtenir des estimateurs d'erreur *a posteriori* pour le problème de Biot. Pour simplifier l'analyse, nous nous focalisons dans le Chapitre 3 sur des problèmes purement mécaniques. Plus précisement, nous nous plaçons dans un cadre stationnaire de problèmes d'élasticité non linéaires, dans le but d'étendre les résultats à des problèmes élasto-plastiques. Du fait que l'état d'un matériau avec un comportement plastique à un instant ne dépend pas seulement des forces auxquelles il est sujet à cet instant, mais également de celles qui ont été appliquées avant, ces problèmes sont forcément non stationnaires. Nous les traitons dans le Chapitre 4.

Dans le Chapitre 3 nous utilisons une nouvelle reconstruction du tenseur de contraintes, que nous avons développé dans l'Annexe A. Cette reconstruction est obtenue en imposant la contrainte de symétrie faiblement, et non plus fortement comme au chapitre précedent, ce qui facilite l'implémentation et nous permet d'améliorer les temps de calcul. Les deux contributions du Chapitre 3 sont les suivantes :

- Nous étendons les développements obtenus pour des problèmes linéaires dans le Chapitre 2 à des problèmes avec un comportement mécanique non linéaire. Dans l'estimation d'erreur, nous ajoutons un estimateur de l'erreur de linéarisation dans l'esprit de [55], ce qui nous permet d'introduire des critères d'arrêt adaptatifs pour le solveur de linéarisation.

- Nous démontrons l'efficacité locale et globale de l'estimation d'erreur pour des problèmes hyperélastiques.

Symétrie faible du tenseur de contraintes reconstruit

Dans le Chapitre 3 nous utilisons les espaces d'éléments finis mixtes d'Arnold–Falk–Winther introduits dans [9] (voir Figure 1.7) pour la reconstruction du tenseur de contraintes. Contrairement aux éléments d'Arnold–Winther, les éléments d'Arnold–Falk–Winther ne fournissent pas des tenseurs symétriques mais imposent la symétrie faiblement via un multiplicateur de Lagrange. Par conséquent, l'espace $\underline{\underline{\Sigma}}$ pour les contraintes (à gauche dans la Figure 1.7) n'a pas de degrés de liberté sur les sommets. Il est l'extension aux tenseurs de l'espace d'éléments finis de Brezzi-Douglas-Marini [29], dans le sens où chaque ligne du tenseur $\underline{\underline{\Sigma}}$ est dans l'espace de Brezzi-Douglas-Marini. Il est donc facile de voir qu'un tenseur dans $\underline{\underline{\Sigma}}$ est dans $\underline{\underline{H}}(\mathrm{div}, \Omega)$,

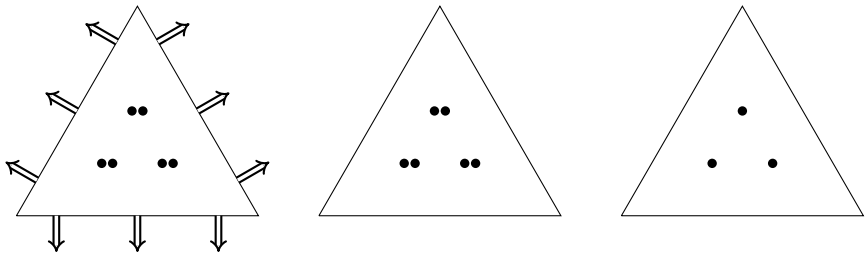


FIGURE 1.7 – Diagrammes des éléments d'Arnold–Falk–Winther utilisés pour la reconstruction du tenseur de contraintes dans Code_Aster. De gauche à droite : l'espace pour le tenseur de contrainte et les espaces des multiplicateurs de Lagrange pour la divergence et la symétrie du tenseur

mais en général pas dans $\underline{\underline{H}}_{\mathrm{s}}(\mathrm{div}, \Omega)$. La symétrie du tenseur de contraintes est ensuite imposée en exigeant que sa projection L^2 dans l'espace des tenseurs antisymétriques jusqu'à un certain degré polynomial (à droite dans la Figure 1.7) soit égale au tenseur nul.

Le relâchement de la contrainte de symétrie et l'absence de degrés de liberté sur les sommets ont deux avantages : premièrement, la formulation des fonctions de forme pour les tenseurs est plus facile, il existe notamment des expressions utilisant les coordonnés barycentriques [22]. Ce point est d'autant plus important en trois dimensions, où l'élément tétraédrique d'Arnold–Winther de degré polynomial le plus bas a 162 degrés de liberté. Deuxièmement, il est possible d'effectuer l'hybridation des problèmes locaux. Comme mentionné dans la section précédente, cette technique permet de transformer des problèmes de type point selle en problèmes coercifs permettant l'utilisation de solveurs algébriques plus rapides. Sur des triangles, la reconstruction du tenseur de contraintes est implémentée sans hybridation, ce qui réduit déjà le temps de calcul par rapport à la méthode précédente. Nous discutons l'hybridation dans la section suivante, où nous l'avons appliquée aux reconstructions des flux sur des tétraèdres. Une comparaison des deux techniques de reconstructions du tenseur de contraintes en 2D est effectuée dans l'Annexe A.

Calcul de l'erreur de linéarisation et critère adaptatif de convergence

Pour obtenir des estimations séparées des erreurs de discrétisation et de linéarisation à une itération de Newton k donnée, nous reconstruisons deux tenseurs de contraintes : le premier est la reconstruction équilibrée des contraintes discrètes correspondant au déplacement $\underline{u}_h^k$ qui résout le problème linéarisé ; le deuxième mesure la différence entre ces contraintes discrètes et les contraintes linéarisées. Ce deuxième tenseur tend donc vers le tenseur nul quand le solveur de Newton converge, et nous définissons l'estimateur de l'erreur de linéarisation comme sa norme L^2. L'estimateur de l'erreur de discrétisation se calcule comme la norme L^2 de la différence entre la première reconstruction et les contraintes discrètes, complétée par un terme d'erreur d'oscillation. Ces deux estimateurs expriment la partie de l'erreur correspondante en termes de normes L^2 de tenseurs de contraintes, ce qui les rend comparables. Nous pouvons alors arrêter le solveur itératif de linéarisation dès que l'erreur de linéarisation devient négligeable par rapport à celle de discrétisation.

Efficacité

En faisant quelques hypothèses sur la relation entre les contraintes et la déformation, nous pouvons démontrer l'équivalence entre la norme duale du résidu de la formulation faible (la mesure d'erreur utilisée pour obtenir l'estimation) et la norme d'énergie. Plus précisement, nous supposons que la fonction qui exprime le tenseur de contraintes en termes du tenseur de déformations est lipschitzienne et strictement monotone. Cette équivalence nous permet d'obtenir une borne garantie sur l'erreur d'énergie.

De plus, cette équivalence de normes est la base de la troisième contribution du Chapitre 3, où on considère le problème mécanique et la norme d'énergie : la démonstration de l'efficacité locale et globale de l'estimation d'erreur pour le problème hyperélastique en utilisant les techniques de [50, 55, 62]. Cette borne inférieure est importante pour être sûr de ne pas surestimer l'erreur. L'idée est de borner l'estimation de type flux équilibrés localement par l'estimation d'erreur de type résidus, qui est quant à elle bornée par l'erreur d'énergie [104]. La première borne exprime la propriété d'approximation locale pour la reconstruction du tenseur de contraintes discrètes. Plus précisément, la différence entre cette reconstruction et les contraintes discrètes mesurée en norme L^2 (ce qui correspond à l'estimateur d'erreur de discrétisation) sur chaque élément est bornée par la somme des estimateurs par résidus sur tous les éléments voisins d'un élément donné. La technique utilisée dans la preuve a initialement été proposée pour établir des estimations d'erreurs *a priori* pour des méthodes d'élément finis mixtes (cf [109]). Les estimateurs par résidu mesurent l'erreur (comme indiqué par le nom) par la différence entre la divergence des contraintes discrètes et le terme source qui est égal à la divergence des contraintes sur chaque élément et les sauts des contraintes discrètes à travers les interfaces. Pour pouvoir comparer les contraintes à leurs divergence on introduit pour chaque patch un déplacement dans l'espace discret obtenu en hybridisant le problème local sur le patch. En imposant les bonnes conditions sur ce déplacement, l'estimateur de discrétisation est borné par sa norme duale, qui quant à elle peut être bornée par les estimateurs par résidus.

Mise en œuvre dans Code_Aster

Pour pouvoir estimer les deux contributions d'erreur à chaque itération de Newton dans Code_Aster, nous avons intégré le calcul des deux reconstructions de contraintes (et également celle du flux hydraulique pour les problèmes poro-mécaniques) dans la commande `STAT_NON_LINE`. Cette intégration nous permet de comparer les deux contributions après chaque itération, et d'arrêter le solveur non linéaire dès que l'estimation de l'erreur de linéarisation devient négligeable par rapport à celle de discrétisation. Pour des problèmes non stationnaires, nous avons également rajouté la possibilité de diminuer le pas de temps, si l'estimation de l'erreur de discrétisation temporelle est plus petite que celle provenant de la discrétisation spatiale multipliée par un paramètre renseigné par l'utilisateur. Les autres adaptations de la discrétisation dans le pseudo-code de la Figure 1.3 (donc l'augmentation du pas

de temps et le remaillage) se font toujours à l'extérieur de la commande `STAT_NON_LINE`.

Nous avons testé la nouvelle reconstruction et l'arrêt du solveur de Newton sur deux cas tests, en considérant au total trois relations différentes entre contraintes et déformations : le comportement élastique linéaire, le modèle non linéaire de Hencky-Mises, et un modèle d'endommagement isotrope réversible.

Bibliographie

Après avoir donné un aperçu de différentes techniques d'estimation d'erreur *a posteriori* dans la section précédente, nous nous focalisons dans cette partie bibliographique sur les problèmes en mécanique non linéaire. Des estimateurs par résidu pour des problèmes en hyperélasticité sont proposés dans [104]. Dans [90,99–101] des estimateurs en quantité d'intérêt ont été introduits pour l'erreur de modélisation (donc pour la vérification des modèles) et pour la discrétisation pour des problèmes en élasto-plasticité. La méthode de lissage de Zienkiewicz et Zhu a été étendue aux problèmes de visco-plasticité et élasto-plasticité dans [24,26,115]

Une distinction des composantes de discrétisation et de linéarisation de l'erreur pour des problèmes non linéaires en général dans le but de définir des critères d'arrêt adaptatif pour le solveur de linéarisation a été proposée dans [35] et dans [50,55] avec une estimation d'erreur par équilibrage de flux.

1.5 Algorithmes adaptatifs pour des problèmes poro-mécaniques en 3D

Dans le Chapitre 4 nous appliquons les techniques d'estimation d'erreur développées dans les Chapitres 2 et 3 à des problèmes poro-mécaniques en trois dimensions d'espace avec des lois de comportement mécaniques élasto-plastiques. Grâce aux estimations obtenues nous effectuons une simulation de creusement de tunnel en équilibrant les sources d'erreur comme proposé dans l'algorithme de la Figure 1.3.

Lois de comportement mécanique pour la modélisation des roches

Les modèles élasto-plastiques décrivent des matériaux qui montrent un comportement élastique sous faible chargement, et des déformations plastiques - donc irréversibles - si les contraintes résultantes de ce chargement dépassent un certain seuil. Ce seuil est en général exprimé comme surface convexe, appelée surface de charge, dans l'espace des contraintes : si le tenseur de contraintes pour un point du matériau est à l'intérieur de cette surface, le comportement de ce point sera élastique ; s'il est sur la surface et va vers son extérieur, le comportement sera plastique et les déformations irréversibles. Cette surface de charge est le premier ingrédient pour formuler une loi de comportement élasto-plastique en définissant *si* un point du matériau

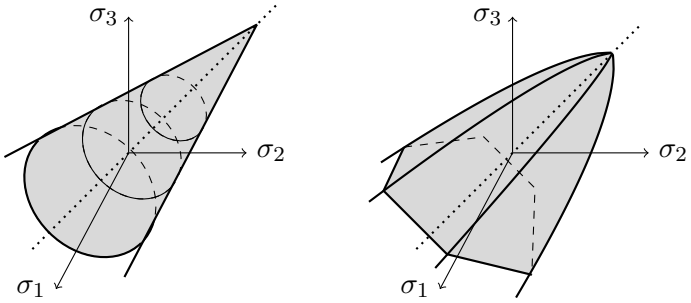


FIGURE 1.8 – Surfaces de charge de Drucker–Prager (à gauche) et de Hoek–Brown (à droite) dans l'espace des contraintes principales

se comporte de façon élastique ou plastique. Le deuxième ingrédient est la loi d'écoulement, définissant *comment* il se comporte si son comportement est plastique. Dans le Chapitre 4 nous utilisons la loi d'écoulement de Prandtl–Reuss, qui, entre autre, fait l'hypothèse que, pendant l'écoulement plastique, l'incrément de déformation plastique est proportionnel à la normale de la surface de charge. Finalement, le troisième ingrédient est la loi d'écrouissage, qui exprime comment des déformations antérieures modifient la surface de charge. Pour faire cela sans devoir considérer toute l'histoire du matériau, on fait l'hypothèse que l'état actuel du matériau peut s'exprimer par une variable dite interne, qui est mise à jour à chaque déformation plastique. La loi d'écrouissage est donc une fonction scalaire de cette variable interne, qui se rajoute à la fonction définissant la surface de charge.

Une des propriétés des géomatériaux est qu'ils deviennent «plus élastiques» sous compression hydrostatique (donc si on appuie uniformément dans toutes les directions). Sous traction hydrostatique, en revanche, ils ont tendance à plastifier. Cette propriété est prise en compte dans la surface de charge. La Figure 1.8 montre deux exemples de surfaces de charge utilisées dans la modélisation de géomatériaux dans l'espace des contraintes principales, où la ligne pointillée correspond aux contraintes hydrostatiques. Une deuxième propriété est que les géomatériaux sont adoucissants, c'est à dire des déformations plastiques endommagent la structure du matériau de façon à ce qu'il resiste moins aux futures sollicitations. La loi d'écrouissage prend en compte cette propriété en diminuant le domaine des contraintes élastiques à l'intérieur de la surface de charge.

Mise en œuvre dans Code_Aster

Nous avons implémenté les reconstructions équilibrées de la vitesse du fluide et du tenseur de contraintes mécaniques en 3D dans la commande de résolution `STAT_NON_LINE` dans Code_Aster. Nous utilisons de nouveau les éléments finis de Raviart–Thomas (à gauche dans la Figure 1.9) pour reconstruire la vitesse, et les éléments finis d'Arnold–Falk–Winther (à droite dans la Figure 1.9) pour le tenseur de contraintes. Les évolutions par rapport aux développements en 2D sont l'hybridation des problèmes locaux et le calcul des matrices dans une phase

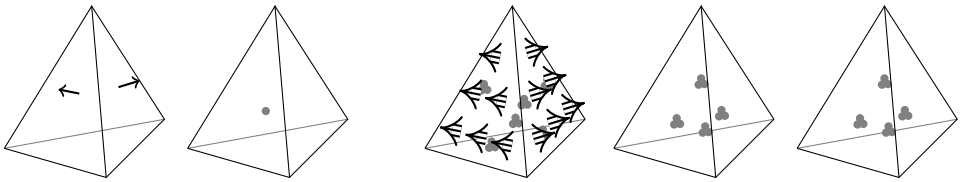


FIGURE 1.9 – Diagrammes des éléments de Raviart–Thomas et d'Arnold–Falk–Winther en 3D

de preprocessing, au lieu de les construire à chaque itération du calcul. L'hybridation s'applique aux méthodes mixtes et permet d'écrire un problème coercif, à condition que les seuls degrés de liberté de l'élément pour la variable duale qui imposent une continuité entre deux éléments voisins soient les composantes normales aux interfaces. On peut alors relâcher cette condition de continuité et l'imposer par multiplicateurs de Lagrange. Les inconnues duales et primales peuvent ensuite être éliminées élément par élément en utilisant des compléments de Schur.

La structure du code nous permet d'appliquer l'estimation d'erreur directement à toutes les lois de comportement mécanique utilisées dans des simulations poro-mécaniques, mais également aux problèmes quasi-statiques de mécanique pure.

Résultats numériques

Dans le Chapitre 4 nous estimons l'erreur d'une simulation de creusement de tunnel en 3D pour trois lois de comportement mécanique différentes : l'élasticité linéaire, la loi de Drucker–Prager et la loi L&K développée au sein d'EDF. Nous utilisons des algorithmes adaptatifs pour équilibrer les estimateurs d'erreur de linéarisation et de discrétisation spatiale et temporelle. En comparant les estimations d'erreur obtenues en équilibrant les sources d'erreurs aux estimations obtenues avec des maillages et pas de temps fixés et le critère de convergence pour le solveur itératif de linéarisation actuellement utilisé dans Code_Aster, nous obtenons les résultats suivants :

- Pour les trois lois de comportement nous réduisons le nombre d'inconnues dans le calcul pour une estimation de l'erreur comparable
- En arrêtant le solveur de linéarisation dès que l'erreur de linéarisation est négligeable par rapport à l'erreur de discrétisation, nous réduisons le nombre d'itérations jusqu'à 80% par rapport aux calculs avec un critère de convergence standard
- En comparant la distribution des estimateurs d'erreur de discrétisation spatiale sur des maillages utilisés par des ingénieurs pour ces simulations et sur les maillages obtenu par remaillage adaptatif, nous observons que la distribution des estimateurs est plus équilibrée sur les maillages issus de remaillage adaptatif. La Figure 1.10 montre les estimateurs d'erreur de discrétisation spatiale à la fin d'un calcul pour la loi de comportement mécanique L&K sur un maillage fixé à gauche et un maillage adapté à droite

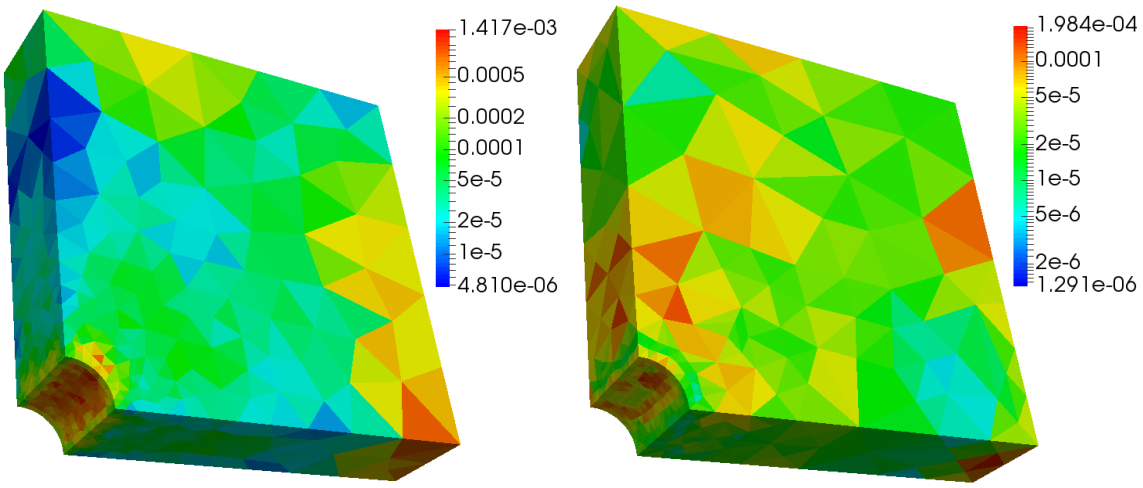


FIGURE 1.10 – Comparaison des estimateurs de discrétisation spatiale sans et avec remaillage adaptatif

1.6 Perspectives

La perspective principale du travail présenté dans cette thèse est de rendre l'utilisation des estimateurs d'erreur plus attractive pour les ingénieurs utilisant le code. Pour ce faire, nous voyons deux points clefs : d'une part la réduction du temps de calcul et de la mémoire utilisée en preprocessing, et d'autre part l'application à plus de problèmes et discrétisations.

La difficulté du premier point résulte essentiellement du fait que Code_Aster est un code spécialisé aux éléments finis de Lagrange. Ses routines d'assemblage et de calculs élémentaires sont optimisées pour des degrés de liberté aux nœuds, mais il n'existe pas d'outils d'assemblage pour des degrés de liberté aux faces, comme c'est le cas pour les éléments $H(\mathrm{div})$-conformes. Nous utilisons les outil de calcul élémentaires pour pouvoir profiter du calcul rapide des fonctions de formes et des intégrales, mais la récupération de ces matrices hors de l'outil d'assemblage se fait valeur par valeur (en passant à chaque fois par d'autres routines pour récuperer l'adresse de la valeur). Cela coûte donc très cher en termes de temps de calcul. En conséquence nous calculons toutes les matrices utilisées dans les calculs sur les patchs en preprocessing. Cela inclut également des matrices utilisées dans l'hybridation pour calculer le vecteur de droite (qui dépend des flux discrets) et les degrés de liberté des flux reconstruits en fonction de la solution du problème hybridé, ce qui consomme beaucoup de mémoire. Il est - dans le cadre d'une autre thèse - prévu d'introduire des degrés de liberté sur les faces dans Code_Aster. L'implémentation de la reconstruction des flux pourra alors à ce moment être revue et améliorée.

Pour pouvoir appliquer les algorithmes adaptatifs à un plus grand nombre de simulations, il serait utile de les implémenter également sur des maillages plus généraux, donc pour des quadrangles en 2D et des hexaèdres et prismes en 3D. Sur des quadrangles, l'équivalent des éléments d'Arnold–Falk–Winther d'ordre plus bas a été introduit dans [11], tandis que des éléments de Raviart–Thomas pour chaque ligne du tenseur en imposant la symétrie de ce

dernier faiblement sont considérés dans [7]. A notre connaissance, il n'existe pas d'extension de ces éléments en trois dimensions d'espaces.

Enfin, on pourrait envisager d'étendre la méthode à des problèmes de poro-mécanique plus généraux, en considérant par exemple non seulement un fluide, mais deux (eau et air typiquement), qui peuvent être présents en phase fluide ou gazeuse, ce qui introduit de nouvelles non linéarités dans le problème. Il est également possible de prendre en compte des phénomènes thermiques en ajoutant une équation parabolique, résultant dans un couplage thermo-hydro-mécanique.

Chapter 2

Stress and flux reconstruction in Biot's poro-elasticity problem with application to a posteriori error analysis

This chapter consists of an article published in Computers and Mathematics with Applications, written with Daniele Di Pietro, Alexandre Ern, Sylvie Granet and Kyrylo Kazymyrenko.

Contents

Abstract

We derive equilibrated reconstructions of the Darcy velocity and of the total stress tensor for Biot's poro-elasticity problem. Both reconstructions are obtained from mixed finite element solutions of local Neumann problems posed over patches of elements around mesh vertices. The Darcy velocity is reconstructed using Raviart–Thomas finite elements and the stress tensor using Arnold–Winther finite elements so that the reconstructed stress tensor is symmetric. Both reconstructions have continuous normal component across mesh interfaces. Using these reconstructions, we derive a posteriori error estimators for Biot's poro-elasticity problem, and we devise an adaptive space-time algorithm driven by these estimators. The algorithm is illustrated on test cases with analytical solution, on the quarter five-spot problem, and on an industrial test case simulating the excavation of two galleries.

2.1 Introduction

Biot's poro-elasticity problem was originally proposed by von Terzaghi [110] and Biot [19] to describe the hydro-mechanical coupling between the displacement field $\underline{u}$ of a linearly elastic, porous material and the pressure p of an incompressible, viscous fluid saturating its pores. Let $\Omega \subset \mathbb{R}^2$ be the simply connected polygonal region occupied by the porous material, and let $t_F > 0$ denote the simulation time. For the sake of simplicity, we assume that the material is clamped at its impermeable boundary, and fix the Biot–Willis coefficient equal to 1. We also assume that the deformation of the material is much slower than the flow rate, so that the problem can be considered in quasi-static form. Then, the displacement field $\underline{u} : \Omega \times (0, t_F) \to \mathbb{R}^2$ and the pore pressure $p : \Omega \times (0, t_F) \to \mathbb{R}$ are determined by

$$-\underline{\nabla} \cdot \underline{\sigma}(\underline{u}) + \underline{\nabla} p = \underline{f} \qquad \text{in } \Omega \times (0, t_F), \tag{2.1a}$$

$$\partial_t(\nabla \cdot \underline{u} + c_0 p) - \nabla \cdot (\kappa \underline{\nabla} p) = g \qquad \text{in } \Omega \times (0, t_F), \tag{2.1b}$$

$$\underline{u} = \underline{0} \qquad \text{on } \partial\Omega \times (0, t_F), \tag{2.1c}$$

$$\kappa \underline{\nabla} p \cdot \underline{n}_\Omega = 0 \qquad \text{on } \partial\Omega \times (0, t_F), \tag{2.1d}$$

$$\underline{u}(\cdot, 0) = \underline{u}_0 \qquad \text{in } \Omega, \tag{2.1e}$$

where $\underline{f}$ denotes the volumetric body force acting on the material, g a volumetric fluid source (which, if $c_0 = 0$, is assumed to verify the compatibility condition $\int_\Omega g(\underline{x}, t) \mathrm{d}\underline{x} = 0$ for each

$t \in (0, t_F)$), and the effective stress tensor $\underline{\underline{\sigma}}$ is linked to the strain tensor $\underline{\underline{\epsilon}}$ through Hooke's law

$$\underline{\underline{\sigma}}(\underline{u}) = 2\mu\underline{\underline{\epsilon}}(\underline{u}) + \lambda\,\mathrm{tr}(\underline{\underline{\epsilon}}(\underline{u}))\underline{\underline{I}}_2, \qquad \underline{\underline{\epsilon}}(\underline{u}) = \frac{1}{2}(\underline{\underline{\nabla}}\underline{u} + \underline{\underline{\nabla}}\underline{u}^{\mathrm{T}}), \tag{2.2}$$

where $\underline{\underline{I}}_2$ is the two-dimensional identity matrix. The Lamé parameters λ and μ, describing the mechanical properties of the material, are assumed such that $\mu > 0$ and $\lambda + \frac{2}{3}\mu > 0$ uniformly in Ω. The scalar field $\kappa : \Omega \to \mathbb{R}$ describes the mobility of the fluid and we assume that there exist positive real numbers $\kappa_\flat$ and $\kappa_\sharp$ such that $\kappa_\flat \leqslant \kappa \leqslant \kappa_\sharp$ a.e. in Ω. When the specific storage coefficient c_0 is zero, we enforce uniqueness of the pore-pressure in (2.1) by further requiring that

$$\int_\Omega p(\cdot, t)\mathrm{d}\underline{x} = 0 \qquad \text{in } (0, t_F). \tag{2.3}$$

On the other hand, for $c_0 > 0$, we complement (2.1) by the following initial condition on the pressure:

$$p(\cdot, 0) = p_0 \qquad \text{in } \Omega. \tag{2.4}$$

In practice, even when $c_0 = 0$, the initial velocity field $\underline{u}_0$ in (2.1e) is usually obtained by first setting p_0 equal to the solution of a hydrostatic computation and then calculating $\underline{u}_0$ by solving (2.1a) with $p = p_0$ (cf. Remark 2.1 in [85]). The well-posedness of Biot's consolidation problem has been analyzed in [98,111]. A suitable approximation method consists of using Taylor–Hood H^1-conforming finite elements in space (using piecewise polynomials of order $k \geqslant 1$ for the pressure and of order $(k+1)$ for the displacement) and a backward Euler scheme in time. The corresponding a priori error analysis can be found in [77–79]. This discretization strategy is adopted in the `Code_Aster`[1] software, which is used for the numerical examples presented in this work. Several other discretization methods have been studied in the literature, among which we cite, in particular, the fully coupled algorithm of [20], where the Hybrid High-Order method of [39] is used for the elasticity operator, while the weighted discontinuous Galerkin method of [40] is used for the Darcy operator.

The two governing equations (2.1a) and (2.1b) express, respectively, the conservation of mechanical momentum and fluid mass. In particular, the Darcy velocity $\underline{\phi}(p) := -\kappa\underline{\nabla}p$ and the total stress tensor $\underline{\underline{\theta}}(\underline{u}, p) := \underline{\underline{\sigma}}(\underline{u}) - p\underline{\underline{I}}_2$ have continuous normal component across any interface in the domain Ω, and the divergence of these fields is locally in equilibrium with the sources (and the accumulation terms) in any control volume. It is well known that the use of H^1-conforming finite elements does not lead to discrete fluxes $\underline{\phi}(p_h^n)$ and $\underline{\underline{\theta}}(\underline{u}_h^n, p_h^n)$ (where $(\underline{u}_h^n, p_h^n)$ denotes the discrete solution at a given discrete time t^n) that satisfy the discrete counterpart of the above properties across mesh interfaces and in mesh cells. The first contribution of this work is to fill this gap by reconstructing equilibrated fluxes from local mixed finite element solves on cell patches around mesh vertices. The Darcy velocity reconstruction uses, as in [28,38,56], Raviart–Thomas mixed finite elements [92] on cell patches around vertices of the original mesh. The construction we propose for the total stress tensor is, to our knowledge,

[1] `http://web-code-aster.org`

novel and is based on the use of the Arnold–Winther mixed finite element [10], again on the same vertex-based cell patches. This construction provides, in particular, a symmetric total stress tensor. The Darcy velocity and the total stress tensor are reconstructed at each discrete time, they have continuous normal component across any mesh interface, and their divergence is locally in equilibrium with the sources (averaged over the time interval) in any mesh cell. In steady-state linear elasticity, element-wise (as opposed to patch-wise) reconstructions of equilibrated tractions from local Neumann problems can be found in [4, 36, 73, 84], whereas direct prescription of the degrees of freedom in the Arnold–Winther finite element space is considered in [82].

The second contribution of this work is to perform an a posteriori error analysis of Biot's poro-elasticity problem using the above reconstructed fluxes to compute the error indicators. Equilibrated-flux a posteriori error estimates for poro-elasticity appear to be a novel topic (residual-based error estimates can be found, e.g., in [52, 76]). Equilibrated-flux a posteriori error estimates offer several advantages. On the one hand, error upper bounds are obtained with fully computable constants. The idea can be traced back to [89] and was advanced amongst others by [3, 28, 38, 54, 56, 69, 71, 75, 93]. Another interesting property is the polynomial-degree robustness proved recently for the Poisson problem in [27, 56]. A third attractive feature introduced in [55] is to distinguish among various error components, e.g., discretization, linearization, and algebraic solver error components, and to equilibrate adaptively these components in the iterative solution of nonlinear problems. This idea was applied to multi-phase, multi-components (possibly non isothermal) Darcy flows in [41–43]. For simplicity, we consider in the present work a global error measure which lends itself naturally to the development of equilibrated-flux error estimators, and defined as the dual energy-norm of the residual of the weak formulation.

This paper is organized as follows. In Section 2.2, we introduce the weak and discrete formulations of Biot's poro-elasticity problem (2.1), along with some useful notation and preliminary results. In Section 2.3, we present the equilibrated reconstruction for the Darcy velocity and the total stress tensor. In Section 2.4, we derive a fully computable upper bound on the residual dual norm. We then distinguish two different error sources in the upper bound, namely the spatial and the temporal discretization, and we propose an algorithm adapting the mesh and the time step so as to equilibrate these error sources. Finally, we show numerical results in Section 2.5.

2.2 Setting

In this section we introduce some notation, the weak formulation, and the discrete solution of problem (2.1).

2.2.1 Weak formulation

We denote by $L^2(\Omega)$, $\underline{L}^2(\Omega)$ and $\underline{\underline{L}}^2(\Omega)$ the spaces composed of square-integrable functions taking values in $\mathbb{R}$, $\mathbb{R}^2$ and $\mathbb{R}^{2\times 2}$ respectively, and by $(\cdot,\cdot)$ and $\|\cdot\|$ the corresponding inner product and norm. We also let $L^2_0(\Omega) := \{q \in L^2(\Omega) \mid (q,1) = 0\}$. $\underline{H}^1(\Omega)$ stands for the Sobolev space composed of $\underline{L}^2(\Omega)$ functions with weak gradients in $\underline{\underline{L}}^2(\Omega)$ and $\underline{H}^1_0(\Omega)$ for its zero-trace subspace. $\underline{\underline{H}}(\mathrm{div},\Omega)$ and $\underline{H}(\mathrm{div},\Omega)$ denote the spaces composed of $\underline{\underline{L}}^2(\Omega)$ and $\underline{L}^2(\Omega)$ functions with weak divergence in $\underline{L}^2(\Omega)$ and $L^2(\Omega)$, respectively, $\underline{\underline{H}}_{\mathrm{s}}(\mathrm{div},\Omega)$ the subspace of $\underline{\underline{H}}(\mathrm{div},\Omega)$ composed of symmetric-valued tensors, and $\underline{H}_0(\mathrm{div},\Omega) := \{\underline{\varphi} \in \underline{H}(\mathrm{div},\Omega) \mid \underline{\varphi}\cdot\underline{n}_\Omega = 0 \text{ on } \partial\Omega\}$.

We assume henceforth, for the sake of simplicity, that the volumetric body force $\underline{f}$ and the fluid source g lie in $L^2(0,t_F;\underline{L}^2(\Omega))$ and $L^2(0,t_F;L^2_0(\Omega))$, respectively. In order to write a weak formulation of this poro-elastic problem, we define

$$\underline{U} := \underline{H}^1_0(\Omega), \qquad P := H^1(\Omega), \tag{2.5}$$

where in the case $c_0 = 0$ we require additionally that $P = H^1(\Omega) \cap L^2_0(\Omega)$, and introduce the following Bochner spaces:

$$X := L^2(0,t_F;\underline{U}) \times L^2(0,t_F;P), \tag{2.6a}$$
$$Y := H^1(0,t_F;\underline{U}) \times H^1(0,t_F;P). \tag{2.6b}$$

Let $\underline{u},\underline{v} \in \underline{U}$ and $p,q \in P$. We define the bilinear forms

$$a(\underline{u},\underline{v}) := (\underline{\underline{\sigma}}(\underline{u}),\underline{\underline{\epsilon}}(\underline{v})), \tag{2.7a}$$
$$b(\underline{v},q) := -(q,\nabla\cdot\underline{v}), \tag{2.7b}$$
$$c(p,q) := (c_0 p,q), \tag{2.7c}$$
$$d(p,q) := (\kappa\underline{\nabla}p,\underline{\nabla}q). \tag{2.7d}$$

Then, we consider the following weak formulation: find $(\underline{u},p) \in Y$, verifying the initial condition (2.1e) with $\underline{u}_0 \in \underline{H}^1_0(\Omega)$ and (2.4) with $p_0 \in H^1(\Omega)$ if $c_0 > 0$, and such that, for a.e. $t \in (0,t_F)$,

$$a(\underline{u}(t),\underline{v}) + b(\underline{v},p(t)) = (\underline{f}(t),\underline{v}) \qquad \forall \underline{v} \in \underline{U}, \tag{2.8a}$$
$$-b(\partial_t\underline{u}(t),q) + c(\partial_t p(t),q) + d(p(t),q) = (g(t),q) \qquad \forall q \in P. \tag{2.8b}$$

The well-posedness of Biot's consolidation problem in slightly different weak formulations is shown in [98,111]. The uniqueness of the solution to (2.8) can be shown by energy arguments. Assuming the existence of the solution in Y, we denote by $\underline{\underline{\sigma}}(\underline{u})$ the resulting effective stress tensor, by $\underline{\underline{\theta}}(\underline{u},p) = \underline{\underline{\sigma}}(\underline{u}) - p\underline{\underline{I}}_2$ the total stress tensor and by $\underline{\phi}(p) = -\kappa\underline{\nabla}p$ the Darcy velocity.

They verify the following properties:

$$\underline{\underline{\theta}}(\underline{u},p) \in L^2(0,t_F;\underline{\underline{H}}_{\rm s}(\mathrm{div},\Omega)), \qquad -\underline{\nabla}\cdot\underline{\underline{\theta}}(\underline{u},p) = \underline{f}, \tag{2.9a}$$

$$\underline{\phi}(p) \in L^2(0,t_F;\underline{H}_0(\mathrm{div},\Omega)), \qquad \nabla\cdot\underline{\phi}(p) = g - \partial_t(\nabla\cdot\underline{u} + c_0 p). \tag{2.9b}$$

2.2.2 Discrete setting

For the time discretization, we consider a sequence of discrete times $(t^n)_{0\leqslant n\leqslant N}$ such that $t^i < t^j$ whenever $i < j$, $t^0 = 0$, and $t^N = t_F$. For each $1 \leqslant n \leqslant N$, let $I_n := (t^{n-1}, t^n)$ and $\tau_n := t^n - t^{n-1}$. For a space-time function v, we denote $v^n := v(\cdot, t^n)$ and define the backward differencing operator $\partial_t^n v = \tau_n^{-1}(v^n - v^{n-1})$.

At each time step $1 \leqslant n \leqslant N$, the space discretization is based on a conforming triangulation $\mathcal{T}_h^n$ of Ω, i.e. a set of closed triangles with union equal to $\overline{\Omega}$ and such that, for any distinct $T_1, T_2 \in \mathcal{T}_h^n$, the set $T_1 \cap T_2$ is either a common edge, a vertex or the empty set. We assume that $\mathcal{T}_h^n$ verifies the minimum angle condition, i.e., there exists $\alpha_{\min} > 0$ uniform with respect to all considered meshes such that the minimum angle α_T of each triangle $T \in \mathcal{T}_h^n$ satisfies $\alpha_T \geqslant \alpha_{\min}$. The set of vertices of the mesh is denoted by $\mathcal{V}_h^n$; it is decomposed into interior vertices $\mathcal{V}_h^{n,\mathrm{int}}$ and boundary vertices $\mathcal{V}_h^{n,\mathrm{ext}}$. For any subdomain $\omega \subset \Omega$ we denote $\mathcal{V}_\omega^n$ the set of vertices in ω. For all $a \in \mathcal{V}_h^n$, $\mathcal{T}_a^n$ is the patch of elements sharing the vertex a, and ω_a the corresponding open subset of Ω. For all $T \in \mathcal{T}_h^n$, $\mathcal{V}_T^n$ denotes the set of vertices of T, h_T its diameter and $\underline{n}_T$ its unit outward normal vector.

For all $n \in \mathbb{N}$ and all $k \in \mathbb{N}$, we denote by $\mathbb{P}_k(T)$ the space of bivariate polynomials in $T \in \mathcal{T}_h^n$ of total degree at most k and by $\mathbb{P}_k(\mathcal{T}_h^n) = \{\varphi \in L^2(\Omega) \mid \varphi|_T \in \mathbb{P}_k(T)\ \forall T \in \mathcal{T}_h^n\}$ the corresponding broken space over $\mathcal{T}_h^n$.

The following Poincaré's inequality holds for all $T \in \mathcal{T}_h^n$:

$$\|v - \Pi_{0,T} v\|_T \leqslant C_{P,T} h_T \|\underline{\nabla} v\|_T \qquad \forall v \in H^1(T), \tag{2.10}$$

where $\Pi_{0,T} : L^1(T) \to \mathbb{P}_0(T)$ is such that $\int_T (v - \Pi_{0,T} v)\mathrm{d}\underline{x} = 0$ and $C_{P,T} = 1/\pi$ owing to the convexity of the mesh elements (see e.g. [15]). Let

$$\underline{RM} := \{\underline{b} + c(x_2, -x_1)^{\mathrm{T}} \mid \underline{b} \in \mathbb{R}^2, c \in \mathbb{R}\} \tag{2.11}$$

denote the space of rigid body motions. We have the following Korn's inequality, again valid for all $T \in \mathcal{T}_h^n$:

$$\|\underline{\underline{\nabla}}(\underline{v} - \Pi_{RM,T}\underline{v})\|_T \leqslant C_{K,T} \|\underline{\underline{\epsilon}}(\underline{v})\|_T \qquad \forall \underline{v} \in \underline{H}^1(T), \tag{2.12}$$

where $\Pi_{RM,T} : \underline{H}^1(T) \to \underline{RM}$ is such that $\int_T (\underline{v} - \Pi_{RM,T}\underline{v})\mathrm{d}\underline{x} = \underline{0}$ and $\int_T \mathrm{rot}(\underline{v} - \Pi_{RM,T}\underline{v})\mathrm{d}\underline{x} = 0$ (with $\mathrm{rot}(\underline{v}) := \partial_{x_1} v_2 - \partial_{x_2} v_1$), and the constant $C_{K,T}$ is bounded by $\sqrt{2}(\sin(\alpha_T/4))^{-1}$

(cf [66]). Combining (2.10) and (2.12), and accounting for the bounds on the corresponding constants, we infer that, for all $T \in \mathcal{T}_h^n$,

$$\|\underline{v} - \Pi_{RM,T}\underline{v}\|_T \leqslant \frac{h_T}{\pi}\frac{\sqrt{2}}{\sin(\alpha_T/4)}\|\underline{\underline{\epsilon}}(\underline{v})\|_T \qquad \forall \underline{v} \in \underline{H}^1(T). \tag{2.13}$$

2.2.3 Discrete problem

We will focus on the conforming Taylor–Hood finite element method using for each time step $1 \leqslant n \leqslant N$ the spaces

$$\underline{U}_h^n := \underline{\mathbb{P}}_{k+1}(\mathcal{T}_h^n) \cap \underline{U}, \quad P_h^n := \mathbb{P}_k(\mathcal{T}_h^n) \cap P, \tag{2.14}$$

with $k \geqslant 1$. This method was first proposed in [102] for incompressible flows and is known to provide stable pore pressure approximations (cf. [79] and [96,113]) and is a classical choice for the discretization of poro-mechanical problems by conforming finite elements.

Assumption 2.1 (Piecewise-constant-in-time source terms)**.** *For simplicity of exposition, we assume henceforth that the functions $\underline{f}$ and g are constant-in-time on each time interval I_n and denote $\underline{f}^n := \underline{f}|_{I_n}$ and $g^n := g|_{I_n}$.*

Using the Taylor–Hood finite element spaces and a backward Euler scheme to march in time, the discrete problem reads: given $\underline{u}_h^0$ and, if $c_0 > 0$, p_h^0, find $(\underline{u}_h^n, p_h^n) \in \underline{U}_h^n \times P_h^n$, for all $1 \leqslant n \leqslant N$, such that

$$a(\underline{u}_h^n, \underline{v}_h) + b(\underline{v}_h, p_h^n) = (\underline{f}^n, \underline{v}_h) \quad \forall \underline{v}_h \in \underline{U}_h^n, \tag{2.15a}$$

$$-b(\partial_t \underline{u}_{h\tau}^n, q_h) + c(\partial_t p_{h\tau}^n, q_h) + d(p_h^n, q_h) = (g^n, q_h) \quad \forall q_h \in P_h^n, \tag{2.15b}$$

where we denote by $\underline{u}_{h\tau}, p_{h\tau}$ the discrete space-time functions which are continuous and piecewise affine in time, and such that, for each $0 \leqslant n \leqslant N$, $(\underline{u}_{h\tau}, p_{h\tau})(\cdot, t^n) = (\underline{u}_h^n, p_h^n)$, so that $\partial_t \underline{u}_{h\tau}^n := \partial_t \underline{u}_{h\tau}|_{I_n} = \tau_n^{-1}(\underline{u}_h^n - \underline{u}_h^{n-1})$ and $\partial_t p_{h\tau}^n := \partial_t p_{h\tau}|_{I_n} = \tau_n^{-1}(p_h^n - p_h^{n-1})$.

2.3 Quasi-static flux reconstructions

In contrast to (2.9), we have in general $\underline{\underline{\theta}}(\underline{u}_h^n, p_h^n) \notin \underline{\underline{H}}_{\mathrm{s}}(\mathrm{div}, \Omega)$ and $\underline{\phi}(p_h^n) \notin \underline{H}_0(\mathrm{div}, \Omega)$. In this section, we restore these properties by reconstructing $H(\mathrm{div})$-conforming discrete fluxes. These reconstructions are devised locally on patches of elements around mesh vertices. We first present the reconstructions in an abstract setting; then we apply these reconstructions to Biot's poro-elasticity problem. Since the time variable is irrelevant in devising the reconstructions, we drop the index n in the abstract presentation.

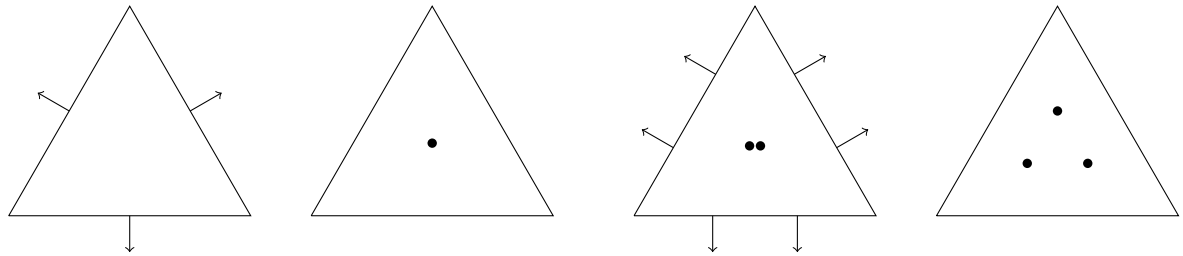


Figure 2.1 – The mixed Raviart–Thomas finite element for $l = 0$ (left) and $l = 1$ (right)

2.3.1 Darcy velocity

We reconstruct the Darcy velocity using mixed Raviart–Thomas finite elements of order $l \geqslant 0$. For each element $T \in \mathcal{T}_h$, the local Raviart–Thomas polynomial spaces are defined by

$$\underline{W}_T := \underline{\mathbb{P}}_l(T) + \underline{x}\mathbb{P}_l(T), \tag{2.16a}$$

$$Q_T = \mathbb{P}_l(T). \tag{2.16b}$$

Figure 2.1 shows the corresponding degrees of freedom for $l = 0$ and $l = 1$. For each vertex $a \in \mathcal{V}_h$, the mixed Raviart–Thomas finite element spaces on the patch domain ω_a are then defined as

$$\tilde{\underline{W}}_h^a := \{\underline{v}_h \in \underline{H}(\mathrm{div}, \omega_a) \mid \underline{v}_h|_T \in \underline{W}_T \ \forall T \in \mathcal{T}_a\}, \tag{2.17a}$$

$$\tilde{Q}_h^a := \{q_h \in L^2(\omega_a) \mid q_h|_T \in Q_T \ \forall T \in \mathcal{T}_a\}. \tag{2.17b}$$

We need to consider the following subspaces associated with the setting where a zero normal component is enforced on the velocity:

$$\underline{W}_h^a := \{\underline{v}_h \in \tilde{\underline{W}}_h^a \mid \underline{v}_h \cdot \underline{n}_{\omega_a} = 0 \text{ on } \partial\omega_a\}, \tag{2.18a}$$

$$Q_h^a := \{q_h \in \tilde{Q}_h^a \mid (q_h, 1)_{\omega_a} = 0\}. \tag{2.18b}$$

The distribution of the degrees of freedom of functions in $\underline{W}_h^a$ is presented in Figure 2.2. Note that we enforce the zero normal condition also on patches associated with boundary vertices since a zero normal Darcy velocity is prescribed in the exact problem; see Remark 2.4 for other types of boundary conditions.

Construction 2.2 (Darcy velocity $\underline{\phi}_h$)**.** *For each $a \in \mathcal{V}_h$, let $\gamma_a \in L^2(\omega_a)$ be such that $(\gamma_a, 1)_{\omega_a} = 0$, and let $\underline{\Gamma}_a \in \underline{L}^2(\omega_a)$. Consider the following constrained minimization problem:*

$$\underline{\varphi}_h^a = \underset{\underline{w}_h \in \underline{W}_h^a,\ \nabla\cdot\underline{w}_h = \Pi_{Q_h^a}\gamma_a}{\mathrm{argmin}} \|\underline{w}_h - \underline{\Gamma}_a\|_{\omega_a}, \tag{2.19}$$

where $\Pi_{Q_h^a}$ denotes the L^2–orthogonal projection on Q_h^a. Then, extending $\underline{\varphi}_h^a$ by zero outside ω_a, set

$$\underline{\phi}_h := \sum_{a \in \mathcal{V}_h} \underline{\varphi}_h^a. \tag{2.20}$$

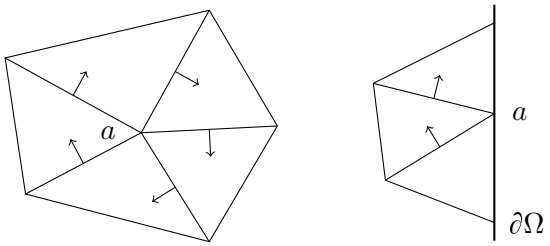


Figure 2.2 – The degrees of freedom of the space $\underline{W}_h^a$ in the case $l = 0$ on a patch for $a \in \mathcal{V}_h^{\mathrm{int}}$ (left) and $a \in \mathcal{V}_h^{\mathrm{ext}}$ (right)

Since functions in $\underline{W}_h^a$ have zero normal component on $\partial\omega_a$, condition $(\gamma_a, 1)_{\omega_a} = 0$ is crucial for the well-posedness of the constrained minimization problem (2.19). This problem is classically solved by finding $\underline{\varphi}_h^a \in \underline{W}_h^a$ and $s_h^a \in Q_h^a$ such that

$$\left(\underline{\varphi}_h^a, \underline{w}_h\right)_{\omega_a} - \left(s_h^a, \nabla\cdot\underline{w}_h\right)_{\omega_a} = \left(\underline{\Gamma}_a, \underline{w}_h\right)_{\omega_a} \quad \forall \underline{w}_h \in \underline{W}_h^a, \tag{2.21a}$$

$$\left(\nabla\cdot\underline{\varphi}_h^a, q_h\right)_{\omega_a} = \left(\gamma_a, q_h\right)_{\omega_a} \quad \forall q_h \in Q_h^a. \tag{2.21b}$$

This problem is well-posed owing to the properties of mixed Raviart–Thomas finite elements, and we obtain the following result (cf. [28,38,56]):

Lemma 2.3 (Properties of $\underline{\phi}_h$)**.** *Let $\underline{\phi}_h$ be prescribed by Construction 2.2. Then, $\underline{\phi}_h \in \underline{H}_0(\mathrm{div},\Omega)$, and letting $\gamma_h \in L^2(\Omega)$ be defined such that $\gamma_h|_T = \sum_{a\in\mathcal{V}_T}\gamma_a$ for all $T \in \mathcal{T}_h$, we have*

$$(\gamma_h - \nabla\cdot\underline{\phi}_h, q)_T = 0 \qquad \forall q \in Q_T \quad \forall T \in \mathcal{T}_h. \tag{2.22}$$

Remark 2.4 (Other boundary conditions)**.** *Suppose that we are given a partition of the boundary as $\partial\Omega = \partial\Omega_{N,P} \cup \partial\Omega_{D,P}$ (the subsets $\partial\Omega_{N,P}$ and $\partial\Omega_{D,P}$ are conventionally closed in $\partial\Omega$, i.e., $\partial\Omega_{N,P} \cap \partial\Omega_{D,P}$ is the common boundary of the two subsets) and that an inhomogeneous Neumann condition is enforced on the flux $\phi(p)$ on $\partial\Omega_{N,P}$ (and a Dirichlet condition is enforced on p in $\partial\Omega_{D,P}$). Assume that the mesh is fitted to the boundary partition, so that any mesh edge on the boundary belongs to either $\partial\Omega_{N,P}$ or $\partial\Omega_{D,P}$. As detailed in [47], this situation can be accommodated in Construction 2.2 up to minor modifications for all $a \in \mathcal{V}_h^{\mathrm{ext}}$ (the construction is unmodified for all $a \in \mathcal{V}_h^{\mathrm{int}}$). For the flux, we consider for the trial and test spaces, respectively,*

$$\underline{W}_{h,N}^a := \{\underline{w}_h \in \underline{\tilde{W}}_h^a \mid \underline{w}_h\cdot\underline{n}_{\omega_a}|_{\partial\omega_a\setminus\partial\Omega} = 0,\ \underline{w}_h\cdot\underline{n}_{\omega_a}|_{\partial\omega_a\cap\partial\Omega_{N,P}} = \Phi_{a,N}\},$$
$$\underline{W}_{h,0}^a := \{\underline{w}_h \in \underline{\tilde{W}}_h^a \mid \underline{w}_h\cdot\underline{n}_{\omega_a}|_{\partial\omega_a\setminus\partial\Omega} = 0,\ \underline{w}_h\cdot\underline{n}_{\omega_a}|_{\partial\omega_a\cap\partial\Omega_{N,P}} = 0\},$$

with $\Phi_{a,N}$ related to the Neumann condition, whereas we set $Q_h^a := \tilde{Q}_h^a$ if a lies on some edge in $\partial\Omega_{D,P}$ and Q_h^a as in (2.18b) otherwise.

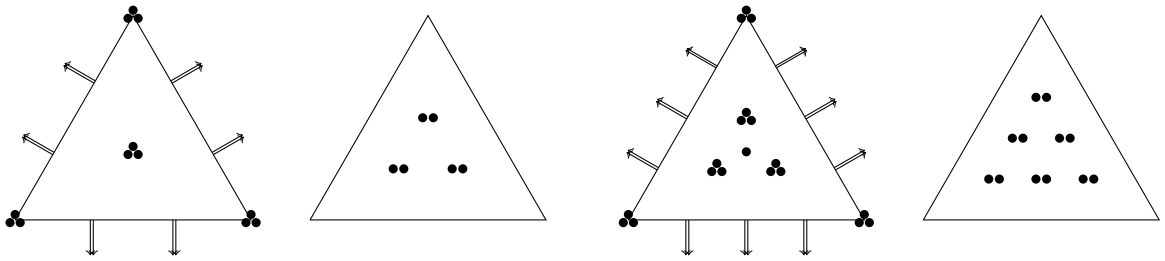


Figure 2.3 – Element diagrams for the pair $(\underline{\underline{\Sigma}}_T, \underline{V}_T)$ in the cases $m = 1$ (left) and $m = 2$ (right)

2.3.2 Total stress tensor

We reconstruct the total stress tensor using mixed Arnold–Winther finite elements of order $m \geqslant 1$. One advantage of using these elements is that the reconstructed stress tensor is symmetric. For each element $T \in \mathcal{T}_h$, the local Arnold–Winther polynomial spaces are defined by

$$\begin{aligned} \underline{\underline{\Sigma}}_T &:= \underline{\underline{\mathbb{P}}}_{s,m+1}(T) + \{\underline{\underline{\tau}} \in \underline{\underline{\mathbb{P}}}_{s,m+2}(T) \mid \underline{\nabla} \cdot \underline{\underline{\tau}} = \underline{0}\} && \text{(2.23a)} \\ &= \{\underline{\underline{\tau}} \in \underline{\underline{\mathbb{P}}}_{s,m+2}(T) \mid \underline{\nabla} \cdot \underline{\underline{\tau}} \in \underline{\underline{\mathbb{P}}}_m(T)\}, \\ \underline{V}_T &:= \underline{\mathbb{P}}_m(T), && \text{(2.23b)} \end{aligned}$$

where $\underline{\underline{\mathbb{P}}}_{s,m}(T)$ denotes the subspace of $\underline{\underline{\mathbb{P}}}_m(T)$ composed of symmetric-valued tensors.

For each vertex $a \in \mathcal{V}_h$, the mixed Arnold–Winther finite element spaces on the patch domain ω_a are defined as

$$\begin{aligned} \tilde{\underline{\underline{\Sigma}}}_h^a &:= \{\underline{\underline{\tau}}_h \in \underline{\underline{H}}_{\mathrm{s}}(\mathrm{div}, \omega_a) \mid \underline{\underline{\tau}}_h|_T \in \underline{\underline{\Sigma}}_T \ \forall T \in \mathcal{T}_a\}, && \text{(2.24a)} \\ \tilde{\underline{V}}_h^a &:= \{\underline{v}_h \in \underline{L}^2(\omega_a) \mid \underline{v}_h|_T \in \underline{V}_T \ \forall T \in \mathcal{T}_a\}. && \text{(2.24b)} \end{aligned}$$

Figure 2.3 shows the corresponding degrees of freedom in the cases $m = 1$ and $m = 2$. The dimension of $\underline{V}_T$ is $(m+1)(m+2)$, and it is shown in [10] that $\dim(\underline{\underline{\Sigma}}_T) = (3m^2+17m+28)/2$. For the lowest-order case $m = 1$, the 24 degrees of freedom in $\underline{\underline{\Sigma}}_T$ are

- The values of the three components of the (symmetric) stress tensor at each vertex of the triangle (9 dofs);
- The values of the moments of degree 0 and 1 of the normal components of the stress tensor on each edge (12 dofs);
- The value of the moment of degree 0 of each component of the stress tensor on the triangle (3 dofs).

We need to consider subspaces where a zero normal component is enforced on the stress tensor. Since the boundary condition in the exact problem prescribes the displacement and not the normal stress, we distinguish the case whether a is an interior vertex or a boundary vertex

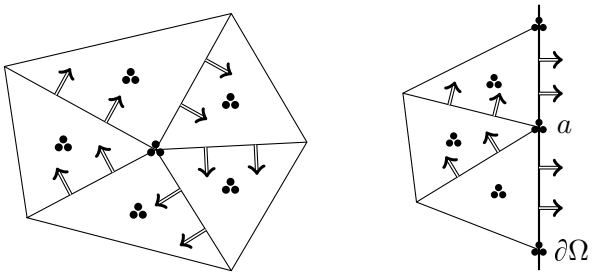


Figure 2.4 – The degrees of freedom of the space $\underline{\underline{\Sigma}}_h^a$ in the case $m = 1$ on a patch for $a \in \mathcal{V}_h^{\mathrm{int}}$ (left) and $a \in \mathcal{V}_h^{\mathrm{ext}}$ (right)

(see Remark 2.7 for other types of boundary conditions). For $a \in \mathcal{V}_h^{\mathrm{int}}$, we set

$$\underline{\underline{\Sigma}}_h^a := \{\underline{\underline{\tau}}_h \in \tilde{\underline{\underline{\Sigma}}}_h^a \mid \underline{\underline{\tau}}_h \underline{n}_{\omega_a} = \underline{0} \text{ on } \partial\omega_a,\ \underline{\underline{\tau}}_h(b) = \underline{\underline{0}}\ \forall b \in \mathcal{V}_{\omega_a} \cap \partial\omega_a\}, \tag{2.25a}$$

$$\underline{V}_h^a := \{\underline{v}_h \in \tilde{\underline{V}}_h^a \mid (\underline{v}_h, \underline{z})_{\omega_a} = 0\ \forall \underline{z} \in \underline{RM}\}, \tag{2.25b}$$

and for $a \in \mathcal{V}_h^{\mathrm{ext}}$, we set

$$\underline{\underline{\Sigma}}_h^a := \{\underline{\underline{\tau}}_h \in \tilde{\underline{\underline{\Sigma}}}_h^a \mid \underline{\underline{\tau}}_h \underline{n}_{\omega_a} = \underline{0} \text{ on } \partial\omega_a \backslash \partial\Omega,\ \underline{\underline{\tau}}_h(b) = \underline{\underline{0}}\ \forall b \in \mathcal{V}_{\omega_a} \cap (\partial\omega_a \backslash \partial\Omega)\}, \tag{2.26a}$$

$$\underline{V}_h^a := \tilde{\underline{V}}_h^a. \tag{2.26b}$$

Note that, as argued in [10], the nodal degrees of freedom on $\partial\omega_a$ are set to zero if the vertex separates two edges where the normal stress is enforced to be zero. The distribution of the degrees of freedom in $\underline{\underline{\Sigma}}_h^a$ is presented in Figure 2.4.

Construction 2.5 (Total stress tensor $\underline{\underline{\theta}}_h$)**.** *For each $a \in \mathcal{V}_h$, let $\underline{\lambda}_a \in \underline{L}^2(\omega_a)$ be such that $(\underline{\lambda}_a, \underline{z})_{\omega_a} = 0$ for all $\underline{z} \in \underline{RM}$ and all $a \in \mathcal{V}_h^{\mathrm{int}}$, and let $\underline{\underline{\Lambda}}_a \in \underline{\underline{L}}^2(\omega_a)$. Consider the following constrained minimization problem:*

$$\underline{\underline{\vartheta}}_h^a = \underset{\underline{\underline{\tau}}_h \in \underline{\underline{\Sigma}}_h^a,\ \underline{\nabla}\cdot\underline{\underline{\tau}}_h = \Pi_{\underline{V}_h^a}\underline{\lambda}_a}{\operatorname{argmin}} \|\underline{\underline{\tau}}_h - \underline{\underline{\Lambda}}_a\|_{\omega_a}, \tag{2.27}$$

where $\Pi_{\underline{V}_h^a}$ denotes the L^2-orthogonal projection onto $\underline{V}_h^a$. Then, extending $\underline{\underline{\vartheta}}_h^a$ by zero outside ω_a, set

$$\underline{\underline{\theta}}_h := \sum_{a \in \mathcal{V}_h} \underline{\underline{\vartheta}}_h^a. \tag{2.28}$$

The condition on $\underline{\lambda}_a$ for all $a \in \mathcal{V}_h^{\mathrm{int}}$ ensures that the constrained minimization problem (2.27) is well-posed. This problem is classically solved by finding $\underline{\underline{\vartheta}}_h^a \in \underline{\underline{\Sigma}}_h^a$ and $\underline{s}_h^a \in \underline{V}_h^a$ such that

$$\left(\underline{\underline{\vartheta}}_h^a, \underline{\underline{\tau}}_h\right)_{\omega_a} - \left(\underline{s}_h^a, \underline{\nabla}\cdot\underline{\underline{\tau}}_h\right)_{\omega_a} = \left(\underline{\underline{\Lambda}}_a, \underline{\underline{\tau}}_h\right)_{\omega_a} \qquad \forall \underline{\underline{\tau}}_h \in \underline{\underline{\Sigma}}_h^a, \tag{2.29a}$$

$$\left(\underline{\nabla}\cdot\underline{\underline{\vartheta}}_h^a, \underline{v}_h\right)_{\omega_a} = (\underline{\lambda}_a, \underline{v}_h)_{\omega_a} \qquad \forall \underline{v}_h \in \underline{V}_h^a. \tag{2.29b}$$

This problem is well-posed (see [10]), and we obtain the following result:

Lemma 2.6 (Properties of $\underline{\underline{\theta}}_h$). *Let $\underline{\underline{\theta}}_h$ be prescribed by Construction 2.5. Then, $\underline{\underline{\theta}}_h \in \underline{\underline{H}}_{\mathrm{s}}(\mathrm{div},\Omega)$, and letting $\underline{\lambda}_h \in \underline{L}^2(\Omega)$ be defined such that $\underline{\lambda}_h|_T = \sum_{a\in\mathcal{V}_T} \underline{\lambda}_a$ for all $T \in \mathcal{T}_h$, we have*

$$(\underline{\lambda}_h - \underline{\nabla}\cdot\underline{\underline{\theta}}_h, \underline{v})_T = 0 \qquad \forall \underline{v} \in \underline{V}_T \quad \forall T \in \mathcal{T}_h. \tag{2.30}$$

Proof. All the fields $\underline{\underline{\vartheta}}_h^a$ are in $\underline{\underline{H}}_{\mathrm{s}}(\mathrm{div},\omega_a)$ and satisfy appropriate zero normal conditions so that their zero-extension to Ω is in $\underline{\underline{H}}_{\mathrm{s}}(\mathrm{div},\Omega)$. Hence, $\underline{\underline{\theta}}_h \in \underline{\underline{H}}_{\mathrm{s}}(\mathrm{div},\Omega)$. Let us prove (2.30). Let $a \in \mathcal{V}_h^{\mathrm{int}}$. Since $(\underline{\lambda}_a, \underline{z})_{\omega_a} = 0$ for all $\underline{z} \in \underline{RM}$, we infer that (2.29b) actually holds for all $\underline{v}_h \in \underline{\tilde{V}}_h^a$. The same holds true if $a \in \mathcal{V}_h^{\mathrm{ext}}$ by definition of $\underline{V}_h^a$. Hence, $\big(\underline{\nabla}\cdot\underline{\underline{\vartheta}}_h^a, \underline{v}_h\big)_{\omega_a} = (\underline{\lambda}_a, \underline{v}_h)_{\omega_a}$ for all $\underline{v}_h \in \underline{\tilde{V}}_h^a$ and all $a \in \mathcal{V}_h$. Since $\underline{\tilde{V}}_h^a$ is composed of piecewise polynomials that can be chosen independently in each cell $T \in \mathcal{T}_a$, we conclude that (2.30) holds. □

Remark 2.7 (Other boundary conditions). *In the spirit of Remark 2.4 with the boundary partition $\partial\Omega = \partial\Omega_{N,U} \cup \partial\Omega_{D,U}$, the minor modifications of Construction 2.5 are as follows for all $a \in \mathcal{V}_h^{\mathrm{ext}}$ (the construction is unmodified for all $a \in \mathcal{V}_h^{\mathrm{int}}$): For the stress tensor, we consider for the trial and test spaces, respectively,*

$$\begin{aligned}
\underline{\underline{\Sigma}}_{h,N}^a := \{\underline{\underline{\tau}}_h \in \underline{\underline{\tilde{\Sigma}}}_h^a \mid\ & \underline{\underline{\tau}}_h\underline{n}_{\omega_a}|_{\partial\omega_a\setminus\partial\Omega} = \underline{0},\ \underline{\underline{\tau}}_h\underline{n}_{\omega_a}|_{\partial\omega_a\cap\partial\Omega_{N,U}} = \underline{\Theta}_{a,N},\\
& \underline{\underline{\tau}}_h(b) = \underline{\underline{0}}\ \forall b \in \mathcal{V}_{\omega_a} \cap (\partial\omega_a\setminus\partial\Omega),\ \underline{\underline{\tau}}_h(b) = \underline{\underline{\theta}}_{a,N}\ \forall b \in \mathcal{V}_{\omega_a} \cap (\partial\Omega\setminus\partial\Omega_{D,U})\},\\
\underline{\underline{\Sigma}}_{h,0}^a := \{\underline{\underline{\tau}}_h \in \underline{\underline{\tilde{\Sigma}}}_h^a \mid\ & \underline{\underline{\tau}}_h\underline{n}_{\omega_a}|_{\partial\omega_a\setminus\partial\Omega} = \underline{0},\ \underline{\underline{\tau}}_h\underline{n}_{\omega_a}|_{\partial\omega_a\cap\partial\Omega_{N,U}} = \underline{0},\\
& \underline{\underline{\tau}}_h(b) = \underline{\underline{0}}\ \forall b \in \mathcal{V}_{\omega_a} \cap (\partial\omega_a\setminus\partial\Omega),\ \underline{\underline{\tau}}_h(b) = \underline{\underline{0}}\ \forall b \in \mathcal{V}_{\omega_a} \cap (\partial\Omega\setminus\partial\Omega_{D,U})\},
\end{aligned}$$

with $\underline{\Theta}_{a,N}$ and $\underline{\underline{\theta}}_{a,N}$ related to the Neumann condition, whereas we set $\underline{V}_h^a$ as in (2.26b) *if a lies on some edge in $\partial\Omega_{D,U}$ and $\underline{V}_h^a$ as in* (2.25b) *otherwise.*

Remark 2.8 (Extension to 3D). *The extension of Construction 2.5 to three dimensions hinges on the existence of mixed finite element spaces producing three dimensional, $\underline{\underline{H}}(\mathrm{div})$-conforming, symmetric tensors. These were introduced in [8], but are complex to implement and require significant computational effort, due to the high number of degrees of freedom per element (162 for the stress tensor).*

2.3.3 Application to Biot's poro-elasticity problem

In this section, we apply the above constructions to the discrete Biot poro-elasticity problem (2.15). The reconstructed Darcy velocity and total stress tensor are space-time functions that are piecewise constant in time, i.e. these functions are calculated at every time step. We use Constructions 2.2 and 2.5 where we now specify the data γ_a, $\underline{\Gamma}_a$, $\underline{\lambda}_a$, and $\underline{\underline{\Lambda}}_a$ for each $1 \leqslant n \leqslant N$. For this purpose, we consider for any mesh vertex $a \in \mathcal{V}_h^n$, the piecewise affine "hat" function $\psi_a \in \mathbb{P}_1(\mathcal{T}_h) \cap H^1(\Omega)$ supported in ω_a, which takes the value 1 at the vertex a and zero at the vertices lying on the boundary of ω_a; cf. Figure 2.5. An important property

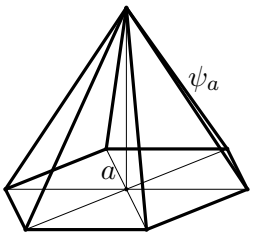


Figure 2.5 – Hat function

for our constructions is the following partition of unity:

$$\sum_{a\in\mathcal{V}_T^n} \psi_a|_T = 1 \qquad \forall T \in \mathcal{T}_h^n. \tag{2.31}$$

Construction 2.9 (Darcy velocity and total stress reconstructions)**.** *Let $1 \leqslant n \leqslant N$. Define for all $a \in \mathcal{V}_h^n$,*

$$\gamma_a = \psi_a g^n - \psi_a \partial_t(\nabla\cdot\underline{u}_{h\tau} + c_0 p_{h\tau})^n + \underline{\nabla}\psi_a \cdot \underline{\phi}(p_h^n), \qquad \underline{\Gamma}_a = \psi_a \underline{\phi}(p_h^n), \tag{2.32a}$$

$$\underline{\lambda}_a = -\psi_a \underline{f}^n + \underline{\underline{\theta}}(\underline{u}_h^n, p_h^n)\underline{\nabla}\psi_a, \qquad \underline{\underline{\Lambda}}_a = \psi_a \underline{\underline{\theta}}(\underline{u}_h^n, p_h^n), \tag{2.32b}$$

where we recall that $\underline{\phi}(p_h^n) = -\kappa\nabla p_h^n$ and $\underline{\underline{\theta}}(\underline{u}_h^n, p_h^n) = \underline{\underline{\sigma}}(\underline{u}_h^n) - p_h^n \underline{\underline{I}}_2$. Then define $\underline{\phi}_h^n \in \underline{H}_0(\mathrm{div},\Omega)$ and $\underline{\underline{\theta}}_h^n \in \underline{\underline{H}}_{\mathrm{s}}(\mathrm{div},\Omega)$ using Constructions 2.2 and 2.5, respectively, with $l \in \{k-1, k\}$ and $m = k$, where k is the degree of the used Taylor–Hood element in (2.15)*.*

Remark 2.10 (Choice of l and m in Construction 2.9)**.** *The polynomial degree k in the Taylor–Hood finite element method* (2.14) *corresponds to degree k for the pressure and degree $k+1$ for the displacement, implying polynomial degree $k-1$ for $\underline{\phi}(p_h)$ and k for $\underline{\underline{\theta}}(\underline{u}_h, p_h)$. Thus, it seems reasonable that the polynomial degree for the reconstruction of the velocity is one lower than for the reconstruction of the total stress tensor. It has been shown in [27, 56] that $k-1$ or k are suitable choices for the velocity reconstruction, and for $k = 1$ we can observe the expected convergence rates for $l = 0$ and $m = 1$ in the numerical test of Section 2.5.2.*

Lemma 2.11 (Darcy velocity and total stress reconstructions)**.** *Construction 2.9 is well defined, and the following holds:*

$$(-\underline{\nabla}\cdot\underline{\underline{\theta}}_h^n, \underline{z})_T = (\underline{f}^n, \underline{z})_T \qquad \forall T \in \mathcal{T}_h^n, \quad \forall \underline{z} \in \underline{RM}, \tag{2.33a}$$

$$(\nabla\cdot\underline{\phi}_h^n, 1)_T = (g^n - \partial_t(\nabla\cdot\underline{u}_{h\tau} + c_0 p_{h\tau})^n, 1)_T \qquad \forall T \in \mathcal{T}_h^n. \tag{2.33b}$$

Proof. Construction 2.2 is well defined provided $(\gamma_a, 1)_{\omega_a} = 0$ holds for all $a \in \mathcal{V}_h$, and this follows by taking $q_h = \psi_a$ in (2.15b) (this is possible since hat functions are contained in the discrete space for the pressure). Construction 2.5 is well defined provided $(\underline{\lambda}_a, \underline{z})_{\omega_a} = 0$ for all $\underline{z} \in \underline{RM}$ and all $a \in \mathcal{V}_h^{n,\mathrm{int}}$, and this follows by taking $\underline{v}_h = \psi_a \underline{z}$ in (2.15a) (recall that $k \geqslant 1$ in

(2.14), so that this choice is legitimate) and using that

$$\begin{aligned}(\underline{\underline{\theta}}(\underline{u}_h^n, p_h^n), \underline{\epsilon}(\psi_a \underline{z})) &= (\underline{\underline{\theta}}(\underline{u}_h^n, p_h^n), \underline{\nabla}(\psi_a \underline{z})^{\mathrm{T}}) \\ &= (\underline{\underline{\theta}}(\underline{u}_h^n, p_h^n), \underline{z}(\underline{\nabla}\psi_a)^{\mathrm{T}}) + (\underline{\underline{\theta}}(\underline{u}_h^n, p_h^n), \psi_a(\underline{\underline{\nabla}}\underline{z})^{\mathrm{T}}) \\ &= (\underline{\underline{\theta}}(\underline{u}_h^n, p_h^n), \underline{z}(\underline{\nabla}\psi_a)^{\mathrm{T}}).\end{aligned}$$

Finally, the properties on the divergence follow from Lemmas 2.6 and 2.3 and the partition of unity (2.31) which implies that $\sum_{a \in \mathcal{V}_T^n} \nabla \psi_a|_T = \underline{0}$. □

Remark 2.12 (Other boundary conditions)**.** *If inhomogeneous Neumann boundary conditions* $\underline{\phi}(p) \cdot \underline{n}_\Omega = \Phi_N$ *or* $\underline{\underline{\theta}}(\underline{u}, p)\underline{n}_\Omega = \underline{\Theta}_N$ *are imposed on* $\partial\Omega_{N,P}$ *and* $\partial\Omega_{N,U}$*, respectively, we modify the constructions using Remarks 2.4 and 2.7. In particular, in Remark 2.4,* $\Phi_{a,N}$ *is the* L^2*-projection of* $\psi_a \Phi_N$ *onto* $\underline{\tilde{W}}_h^a \cdot \underline{n}_\Omega$*, whereas in Remark 2.7,* $\underline{\Theta}_{a,N}$ *is the* L^2*-projection of* $\psi_a \underline{\Theta}_N$ *onto* $\underline{\underline{\tilde{\Sigma}}}_h^a \underline{n}_\Omega$*.*

2.4 A posteriori error analysis and space-time adaptivity

In this section, we derive an a posteriori error estimate at every time step n based on the quasi-equilibrated flux reconstructions of Section 2.3.3 for Biot's poro-elasticity problem. Using these estimators, we devise an adaptive algorithm including the adaptive choice of the mesh size and of the time step.

2.4.1 A posteriori error estimate

To derive the a posteriori error estimate, we consider the residual of the weak formulation (2.8). To combine the two equations in (2.8) into a single residual, the two equations must be written using the same physical units. Therefore, we introduce a reference time scale $t^\star$ and a reference length scale $l^\star$ which we will use as scaling parameters together with the Young modulus $E = \mu(3\lambda + 2\mu)(\lambda + \mu)^{-1}$. Defining the bilinear map $\mathcal{B} : Y \times X \to L^2(0, t_F; \mathbb{R})$ such that

$$\mathcal{B}((\underline{u}, p), (\underline{v}, q))(t) := a(\underline{u}, \underline{v})(t) + b(\underline{v}, p)(t) + t^\star \left(-b(\partial_t \underline{u}, q)(t) + c(\partial_t p, q)(t) + d(p, q)(t)\right), \tag{2.34}$$

we can restate (2.8) as follows: find $(\underline{u}, p) \in Y$, verifying the initial condition (2.1e), and (2.4) if $c_0 > 0$, and such that for a.e. $t \in (0, t_F)$,

$$\mathcal{B}((\underline{u}, p), (\underline{v}, q))(t) = (\underline{f}, \underline{v})(t) + t^\star (g, q)(t) \qquad \forall (\underline{v}, q) \in X. \tag{2.35}$$

For any pair $(\underline{u}_{h\tau}, p_{h\tau}) \in Y$, we define the residual $\mathcal{R}(\underline{u}_{h\tau}, p_{h\tau}) \in X'$ of (2.35) as

$$\langle \mathcal{R}(\underline{u}_{h\tau}, p_{h\tau}), (\underline{v}, q) \rangle_{X',X} := \int_0^{t_F} \mathcal{B}((\underline{u} - \underline{u}_{h\tau}, p - p_{h\tau}), (\underline{v}, q))(t) \mathrm{d}t.$$

Its dual norm is defined as

$$\|\mathcal{R}(\underline{u}_{h\tau},p_{h\tau})\|_{X'} := \sup_{(\underline{v},q)\in X,\|(\underline{v},q)\|_X=1} \langle \mathcal{R}(\underline{u}_{h\tau},p_{h\tau}),(\underline{v},q)\rangle_{X',X},$$

with

$$\|(\underline{v},q)\|_X^2 := \int_0^{t_F} (E\|\underline{\underline{\epsilon}}(\underline{v})\|)^2 + (l^\star\|\underline{\nabla} q\|)^2 \mathrm{d}t.$$

We first derive a local-in-time a posteriori error estimate. Let the time step $1 \leqslant n \leqslant N$ be fixed. Let us set

$$X_n := L^2(I_n;\underline{U}) \times L^2(I_n;P) \quad \text{and} \quad Y_n := H^1(I_n;\underline{U}) \times H^1(I_n;P),$$

and define the norm $\|\cdot\|_{X_n}$ on X_n in the same way as $\|\cdot\|_X$ on X. The local error measure for the time step n is then defined as

$$\begin{aligned} e^n &:= \sup_{(\underline{v},q)\in X_n,\|(\underline{v},q)\|_{X_n}=1} \int_{I_n} \mathcal{B}((\underline{u}-\underline{u}_{h\tau},p-p_{h\tau}),(\underline{v},q))\mathrm{d}t \\ &= \sup_{(\underline{v},q)\in X_n,\|(\underline{v},q)\|_{X_n}=1} \int_{I_n} e_U^n(\underline{v}) + e_P^n(q)\mathrm{d}t, \end{aligned} \tag{2.36}$$

with

$$e_U^n(\underline{v}) := \int_{I_n} (\underline{\underline{\theta}}(\underline{u},p) - \underline{\underline{\theta}}(\underline{u}_{h\tau},p_{h\tau}),\underline{\underline{\epsilon}}(\underline{v}))(t)\mathrm{d}t, \tag{2.37a}$$

$$e_P^n(q) := t^\star \int_{I_n} \big(\partial_t(\nabla\cdot\underline{u}+c_0 p) - \partial_t(\nabla\cdot\underline{u}_{h\tau}+c_0 p_{h\tau}),q\big)^n(t) - \big(\underline{\phi}(p)-\underline{\phi}(p_{h\tau}),\underline{\nabla} q\big)(t)\mathrm{d}t, \tag{2.37b}$$

where we recall that $I_n = (t^{n-1},t^n)$ and that both $\underline{u}_{h\tau}$ and $p_{h\tau}$ are continuous, piecewise affine functions in time.

For all $1 \leqslant n \leqslant N$, let $\underline{\phi}_h^n$ and $\underline{\underline{\theta}}_h^n$ be the constant in time fields over I_n defined by Construction 2.9. For all $T \in \mathcal{T}_h^n$, we define the *residual estimators* $\eta_{R,T,U}^n$, $\eta_{R,T,P}^n$ by

$$\eta_{R,T,U}^n := \frac{h_T}{\pi}\frac{\sqrt{2}}{\sin(\alpha_T/4)} E^{-1}\|\underline{f}^n + \underline{\nabla}\cdot\underline{\underline{\theta}}_h^n\|_T, \tag{2.38a}$$

$$\eta_{R,T,P}^n := \frac{h_T}{\pi}\frac{t^\star}{l^\star}\|g^n - \partial_t^n(\nabla\cdot\underline{u}_{h\tau}+c_0 p_{h\tau}) - \nabla\cdot\underline{\phi}_h^n\|_T, \tag{2.38b}$$

where α_T denotes the minimum angle of the triangle T, and the *flux estimators* $\eta_{F,T,U}^n(t)$, $\eta_{F,T,P}^n(t)$, $t \in I_n$, by

$$\eta_{F,T,U}^n(t) := E^{-1}\|\underline{\underline{\theta}}_h^n - \underline{\underline{\theta}}(\underline{u}_{h\tau},p_{h\tau})(t)\|_T, \tag{2.39a}$$

$$\eta_{F,T,P}^n(t) := \frac{t^\star}{l^\star}\|\underline{\phi}_h^n - \underline{\phi}(p_{h\tau})(t)\|_T. \tag{2.39b}$$

Theorem 2.13 (Local-in-time a posteriori error estimate)**.** *Let $(\underline{u},p) \in Y$ be the weak solution of (2.8) and let $(\underline{u}_{h\tau},p_{h\tau}) \in Y$ be the discrete solution of (2.15). Let $1 \leqslant n \leqslant N$. Let e^n be defined by (2.36) with estimators defined by (2.38) and (2.39). Then the following holds:*

$$e^n \leqslant \left(\int_{I_n} \sum_{T \in \mathcal{T}_h^n} \{ (\eta_{R,T,U}^n + \eta_{F,T,U}^n(t))^2 + (\eta_{R,T,P}^n + \eta_{F,T,P}^n(t))^2 \} \mathrm{d}t \right)^{1/2}. \tag{2.40}$$

Proof. Let $(\underline{v}, q) \in X_n$. Recalling (2.37a), we have

$$\begin{aligned} e_U^n(\underline{v}) &= \int_{I_n} \big(\underline{\underline{\theta}}(\underline{u}, p) - \underline{\underline{\theta}}(\underline{u}_{h\tau}, p_{h\tau}), \underline{\underline{\epsilon}}(\underline{v}) \big)(t) \mathrm{d}t \\ &= \int_{I_n} \big(\underline{f}, \underline{v} \big)(t) - \big(\underline{\underline{\theta}}(\underline{u}_{h\tau}, p_{h\tau}), \underline{\underline{\epsilon}}(\underline{v}) \big)(t) \mathrm{d}t \\ &= \int_{I_n} \Big(\underbrace{\big(\underline{f}^n + \underline{\nabla} \cdot \underline{\underline{\theta}}_h^n, \underline{v}(t) \big)}_{\mathfrak{T}_1(t)} + \underbrace{\big(\underline{\underline{\theta}}_h^n - \underline{\underline{\theta}}(\underline{u}_{h\tau}, p_{h\tau}), \underline{\underline{\epsilon}}(\underline{v}) \big)(t)}_{\mathfrak{T}_2(t)} \Big) \mathrm{d}t, \end{aligned} \tag{2.41}$$

where we have used (2.8a) to pass to the second line and we have inserted $(\underline{\nabla} \cdot \underline{\underline{\theta}}_{h\tau}^n, \underline{v}) + (\underline{\underline{\theta}}_{h\tau}^n, \underline{\underline{\epsilon}}(\underline{v})) = 0$ inside the integral to conclude. For the first term we have, for a.e. $t \in (0, t_F)$,

$$|\mathfrak{T}_1(t)| = \left| \sum_{T \in \mathcal{T}_h^n} (\underline{f}^n + \underline{\nabla} \cdot \underline{\underline{\theta}}_h^n, (\underline{v} - \Pi_{K,T} \underline{v})(t))_T \right| \leqslant \sum_{T \in \mathcal{T}_h^n} \eta_{R,T,U}^n E \| \underline{\underline{\epsilon}}(\underline{v})(t) \|_T,$$

where we have used (2.33a) to insert $\Pi_{K,T} \underline{v}$ inside the integral and (2.13) to conclude. For the second term, using the Cauchy–Schwarz inequality readily yields

$$|\mathfrak{T}_2(t)| \leqslant \sum_{T \in \mathcal{T}_h^n} \| \underline{\underline{\theta}}_h^n - \underline{\underline{\theta}}(\underline{u}_{h\tau}, p_{h\tau})(t) \|_T \| \underline{\underline{\epsilon}}(\underline{v})(t) \|_T = \sum_{T \in \mathcal{T}_h^n} \eta_{F,T,U}^n(t) E \| \underline{\underline{\epsilon}}(\underline{v})(t) \|_T.$$

Inserting these results into (2.41) and applying the Cauchy-Schwarz inequality yields

$$|e_U(\underline{v})| \leqslant \left(\int_{I_n} \sum_{T \in \mathcal{T}_h^n} (\eta_{R,T,U}^n + \eta_{F,T,U}^n(t))^2 \mathrm{d}t \right)^{1/2} \times \left(\int_{I_n} (E \| \underline{\underline{\epsilon}}(\underline{v})(t) \|)^2 \mathrm{d}t \right)^{1/2}.$$

Proceeding in a similar way for e_P^n using (2.33b) and Poincaré's inequality (2.10) in place of (2.33a) and (2.13), respectively, we obtain

$$|e_P(q)| \leqslant \left(\int_{I_n} \sum_{T \in \mathcal{T}_h^n} (\eta_{R,T,P}^n + \eta_{F,T,P}^n(t))^2 \mathrm{d}t \right)^{1/2} \times \left(\int_{I_n} (l^\star \| \underline{\nabla} q(t) \|)^2 \mathrm{d}t \right)^{1/2}.$$

Combining these results and using again the Cauchy-Schwarz inequality yields

$$|e_U(\underline{v}) + e_P(q)| \leqslant \left(\int_{I_n} \sum_{T \in \mathcal{T}_h^n} (\eta_{R,T,U}^n + \eta_{F,T,U}^n(t))^2 + (\eta_{R,T,P}^n(t) + \eta_{F,T,P}^n(t))^2 \mathrm{d}t \right)^{1/2} \times \| (\underline{v}, q) \|_{X_n},$$

and passing to the supremum concludes the proof. □

Remark 2.14 (Data oscillation). *Lemmas 2.3 and 2.6, and the mixed finite element space property* $\underline{\nabla} \cdot \underline{\underline{\Sigma}}_h^n(T) = \underline{V}_h^n(T)$ *and* $\nabla \cdot \underline{W}_h^n(T) = Q_h^n(T)$ *for any* $T \in \mathcal{T}_h^n$ *imply*

$$\eta_{R,T,U}^n = \frac{h_T}{\pi} \frac{\sqrt{2}}{\sin(\alpha_T/4)} E^{-1} \|\underline{f}^n - \Pi_{\underline{V}_h^n(T)} \underline{f}\|_T,$$
$$\eta_{R,T,P}^n = \frac{h_T}{\pi} \frac{t^\star}{l^\star} \|g^n - \partial_t^n(\nabla \cdot \underline{u}_{h\tau} + c_0 p_{h\tau}) - \Pi_{Q_h^n(T)}(g^n - \partial_t^n(\nabla \cdot \underline{u}_{h\tau} + c_0 p_{h\tau}))\|_T.$$

For the sake of convenience, we assumed the source terms $\underline{f}$ *and* g *to be piecewise constant in time. When this is not the case, an additional data time-oscillation term appears in the right-hand side of the bound* (2.40).

Remark 2.15 (Other types of boundary conditions). *If we consider inhomogeneous Neumann boundary conditions* $\underline{\underline{\theta}}(\underline{u}, p)\underline{n}_\Omega = \underline{\theta}_N$ *on* $\partial\Omega_{N,U} \subseteq \partial\Omega$ *and* $\underline{\phi}(p) \cdot \underline{n}_\Omega = \phi_N$ *on* $\partial\Omega_{N,P} \subseteq \partial\Omega$*, two more error estimators appear in* (2.40). *The details of how they are obtained for the hydraulic part are shown in [47] and can be directly applied to linear elasticity. For each* $T \in \mathcal{T}_h^n$*, let* $\mathcal{E}_T^{N,U}$ *and* $\mathcal{E}_T^{N,P}$ *be the set of edges lying on* $\partial\Omega_{N,U}$ *and* $\partial\Omega_{N,P}$ *respectively, and let* $\underline{\underline{\theta}}_h^n$ *and* $\underline{\phi}_h^n$ *be the flux reconstructions of Remarks 2.4 and 2.7. Then we set*

$$\eta_{N,T,U}^n = \sum_{e \in \mathcal{E}_T^{N,U}} \frac{h_T (2h_e C_t)^{1/2}}{E \sin(\alpha_T/4) |T|^{1/2}} \|\underline{\underline{\theta}}_h^n \underline{n}_\Omega - \underline{\theta}_N\|_e,$$
$$\eta_{N,T,P}^n = \sum_{e \in \mathcal{E}_T^{N,P}} \frac{t^\star h_T (h_e C_t)^{1/2}}{l^\star |T|^{1/2}} \|\underline{\phi}_h^n \cdot \underline{n}_\Omega - \phi_N\|_e,$$

where $C_t \approx 0.77708$*, and* (2.40) *now reads*

$$e^n \leqslant \left(\int_{I_n} \sum_{T \in \mathcal{T}_h^n} \{(\eta_{R,T,U}^n + \eta_{F,T,U}^n(t) + \eta_{N,T,U}^n)^2 + (\eta_{R,T,P}^n + \eta_{F,T,P}^n(t) + \eta_{N,T,P}^n)^2\} \mathrm{d}t \right)^{1/2}.$$

To define a global-in-time a posteriori error estimate, we additionally define the *initial condition errors* $\eta_{\mathrm{IC},T,U}$ and $\eta_{\mathrm{IC},T,P}$ by setting

$$\eta_{\mathrm{IC},T,U} := \left(\frac{1}{2} E^{-1} t^\star \left(\underline{\underline{\sigma}}(\underline{u}_0 - \underline{u}_{h\tau}(\cdot, 0)), \underline{\underline{\epsilon}}(\underline{u}_0 - \underline{u}_{h\tau}(\cdot, 0)) \right)_T \right)^{1/2}, \tag{2.42a}$$
$$\eta_{\mathrm{IC},T,P} := \left(\frac{1}{2} E^{-1} (t^\star)^2 c_0 \left((p_0 - p_{h\tau}(\cdot, 0)), p_0 - p_{h\tau}(\cdot, 0) \right)_T \right)^{1/2}, \tag{2.42b}$$

and we set

$$e_{\mathrm{IC}} = \eta_{\mathrm{IC}} = \left(\sum_{T \in \mathcal{T}_h^0} \eta_{\mathrm{IC},T,U}^2 + \eta_{\mathrm{IC},T,P}^2 \right)^{1/2}. \tag{2.43}$$

We define the global error as

$$e := \|\mathcal{R}(\underline{u}_{h\tau}, p_{h\tau})\|_{X'} + e_{\mathrm{IC}}. \tag{2.44}$$

Corollary 2.16 (Global-in-time a posteriori error estimate)**.** *The following holds:*

$$e \leqslant \left(\sum_{n=1}^{N} \int_{I_n} \sum_{T \in \mathcal{T}_h^n} \{(\eta_{R,T,U}^n + \eta_{F,T,U}^n(t))^2 + (\eta_{R,T,P}^n + \eta_{F,T,P}^n(t))^2\} \mathrm{d}t \right)^{1/2} + \eta_{\mathrm{IC}}. \tag{2.45}$$

Proof. For each $1 \leqslant n \leqslant N$, let $\xi_n \in X$ be the Riesz-representative of $J_n : X_n \to \mathbb{R}$ with $J_n(\underline{v}, q) = \int_{I_n} \mathcal{B}((\underline{u} - \underline{u}_{h\tau}, p - p_{h\tau}), (\underline{v}, q)) \mathrm{d}t$. Then the function $\xi \in X$ defined by $\xi_{|I_n} := \xi_n$ will be the Riesz-representative of $J : X \to \mathbb{R}$ with $(\underline{v}, q) \mapsto \int_0^{t_F} \mathcal{B}((\underline{u} - \underline{u}_{h\tau}, p - p_{h\tau}), (\underline{v}, q)) \mathrm{d}t$, so that

$$\|\mathcal{R}(\underline{u}_{h\tau}, p_{h\tau})\|_{X'}^2 = \|J\|_{X'}^2 = \|\xi\|_X^2 = \sum_{n=1}^{N} \|\xi_n\|_{X_n}^2 = \sum_{n=1}^{N} \|J_n\|_{X_n'}^2 = \sum_{n=1}^{N} (e^n)^2. \tag{2.46}$$

Inserting this result into (2.44) and applying Theorem 2.13 concludes the proof. □

2.4.2 Distinguishing the space and time error components

The goal of this section is to elaborate the error estimate (2.40) so as to distinguish the error components resulting from the spatial and the temporal discretization. This is essential for the development of Algorithm 2.18 below, where the space mesh and the time step are chosen adaptively. Therefore, we add and subtract the discrete fluxes in the flux estimators (2.39) and apply the triangle inequality. We obtain, for all $T \in \mathcal{T}_h^n$, the following local *spatial and temporal discretization error estimators*:

$$\eta_{\mathrm{sp},T,U}^n := \eta_{R,T,U}^n + E^{-1} \|\underline{\underline{\theta}}_h^n - \underline{\underline{\theta}}(\underline{u}_h^n, p_h^n)\|_T, \tag{2.47a}$$

$$\eta_{\mathrm{sp},T,P}^n := \eta_{R,T,P}^n + \frac{t^\star}{l^\star} \|\underline{\phi}_h^n - \underline{\phi}(p_h^n)\|_T, \tag{2.47b}$$

$$\eta_{\mathrm{tm},T,U}^n(t) := E^{-1} \|\underline{\underline{\theta}}(\underline{u}_h^n, p_h^n) - \underline{\underline{\theta}}(\underline{u}_{h\tau}, p_{h\tau})(t)\|_T, \tag{2.47c}$$

$$\eta_{\mathrm{tm},T,P}^n(t) := \frac{t^\star}{l^\star} \|\underline{\phi}(p_h^n) - \underline{\phi}(p_{h\tau})(t)\|_T. \tag{2.47d}$$

For each of these local estimators we can define a global version by setting

$$\eta_{\bullet,\{U,P\}}^n := \left(2 \int_{I_n} \sum_{T \in \mathcal{T}_h^n} \left(\eta_{\bullet,T,\{U,P\}}^n(t) \right)^2 \mathrm{d}t \right)^{1/2}. \tag{2.48}$$

Inserting them into (2.40) and applying the triangle inequality yields the following result.

Theorem 2.17 (A posteriori error estimate distinguishing the error components)**.** *Let $1 \leqslant n \leqslant N$. Let $(\underline{u}, p)$ be the weak solution of (2.8) and let $(\underline{u}_{h\tau}, p_{h\tau}) \in Y^n$ be the discrete solution of (2.15). Let $\underline{\underline{\theta}}_h^n$, $\underline{\phi}_h^n$ be the equilibrated fluxes of Construction 2.9. Then the following holds for the error e^n defined by (2.36) with estimators defined by (2.47) and (2.48):*

$$e^n \leqslant \eta_{\mathrm{sp},U}^n + \eta_{\mathrm{sp},P}^n + \eta_{\mathrm{tm},U}^n + \eta_{\mathrm{tm},P}^n. \tag{2.49}$$

2.4.3 Adaptive algorithm

Based on the error estimate of Theorem 2.17, we propose an adaptive algorithm where the mesh size and time step are locally adapted. The idea is to compare the estimators for the two error sources with each other in order to concentrate the computational effort on reducing the dominant one. Thus, both the spatial mesh and the time step are adjusted until space and time discretization contribute nearly equally to the overall error. For this purpose, let $\Gamma_{\mathrm{tm}} > 1 > \gamma_{\mathrm{tm}} > 0$ be user-given weights and crit^n, for all $0 \leqslant n \leqslant N$, a chosen threshold that the error on the time interval I_n should not exceed. For each of the considered error sources we define the corresponding estimator as $\eta_\bullet := \eta_{\bullet,U} + \eta_{\bullet,P}$, so that (2.49) becomes

$$e^n \leqslant \eta_{\mathrm{sp}}^n + \eta_{\mathrm{tm}}^n. \tag{2.50}$$

Algorithm 2.18 (Adaptive algorithm)**.**

1. **Initialisation**
   (a) Choose an initial triangulation $\mathcal{T}_h^0$, an initial time step τ_0, and set $t^0 := 0$
   (b) **Initial mesh adaptation loop**
   - Calculate $\eta_{\mathrm{IC},U}$ and $\eta_{\mathrm{IC,P}}$
   - Refine or coarsen the mesh $\mathcal{T}_h^0$ such that the local initial condition error estimators $\eta_{\mathrm{IC},T,\bullet}$ are distributed equally

   End of the loop if $\eta_{\mathrm{IC}} \leqslant \mathrm{crit}^0$
2. **Time loop**
   (a) Set $n := n + 1$, $\mathcal{T}_h^n := \mathcal{T}_h^{n-1}$, and $\tau_n := \tau_{n-1}$
   (b) Calculate $(\underline{u}_h^n, p_h^n)$ and the estimators η_{sp}^n and η_{tm}^n
   (c) **Space refinement loop**
      i. **Space and time error balancing loop**
         A. if $\gamma_{\mathrm{tm}}\eta_{\mathrm{sp}}^n > \eta_{\mathrm{tm}}^n$: Set $\tau_n := 2\tau_n$
         B. if $\Gamma_{\mathrm{tm}}\eta_{\mathrm{sp}}^n < \eta_{\mathrm{tm}}^n$: Set $\tau_n := \frac{1}{2}\tau_n$

         End of the space-time error balancing loop if

$$\gamma_{\mathrm{tm}}\eta_{\mathrm{sp}}^n \leqslant \eta_{\mathrm{tm}}^n \leqslant \Gamma_{\mathrm{tm}}\eta_{\mathrm{sp}}^n \qquad \text{or} \qquad \tau_n \leqslant \tau_{\min}$$

      ii. Refine or coarsen the mesh $\mathcal{T}_h^n$ such that the local spatial error estimators $\eta_{\mathrm{sp},T}^n$ are distributed equally

   End of the space refinement loop if

$$\eta_{\mathrm{sp}}^n + \eta_{\mathrm{tm}}^n \leqslant \mathrm{crit}^n \tag{2.51}$$

   End of the time loop if $t^n \geqslant t_F$

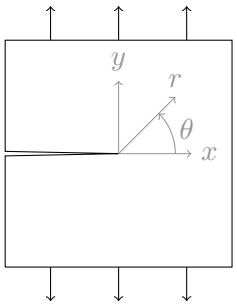


Figure 2.6 – Loading of a cracked plate

Owing to (2.44), (2.46), (2.50) and (2.51) the obtained discrete solution satisfies

$$e \leqslant \left(\sum_{n=1}^{N} (\mathrm{crit}^{\mathrm{n}})^2 \right)^{1/2} + \mathrm{crit}^0. \tag{2.52}$$

In order to keep computational costs in the algorithm low, the initial mesh and time step should be chosen in a way that they match the criteria crit^0 and crit^1. This can be achieved by performing only one time step before running the whole computation, and by modifying the initial discretization if they do not.

2.5 Numerical results

In this section we illustrate numerically our theoretical results on four test cases. For all tests we use the Taylor–Hood finite elements (2.14) with $k = 1$ and Construction 2.9 with $l = 0$ and $m = 1$. In the first two test cases, analytical solutions are known; the first one is a purely elastic, stationary problem and the second one a Biot's poro-elasticity problem. We analyze the convergence rates of the error estimators and compare them to those of an energy-type norm of the analytical error. The third test is the quarter five-spot problem, where we compare the results of the adaptive algorithm to a "standard" solution with fixed mesh and time steps. In the fourth test, the excavation of two parallel tunnels is simulated. It shows an industrial application of the error estimators used for remeshing and again compares the performance of Algorithm 2.18 to a standard resolution.

2.5.1 Purely mechanical analytical test

For this stationary, purely mechanical test we consider the mode I loading of a cracked plate, corresponding to pure tension at the top and the bottom applied at the infinity. Following [112], an analytical solution around the crack tip is given by

$$\underline{u}(r,\theta) = \frac{1+\nu}{E\sqrt{2\pi}}\sqrt{r} \begin{pmatrix} -\cos(\frac{\theta}{2})(3-4\nu-\cos(\theta)) \\ \sin(\frac{\theta}{2})(3-4\nu-\cos(\theta)) \end{pmatrix}, \tag{2.53}$$

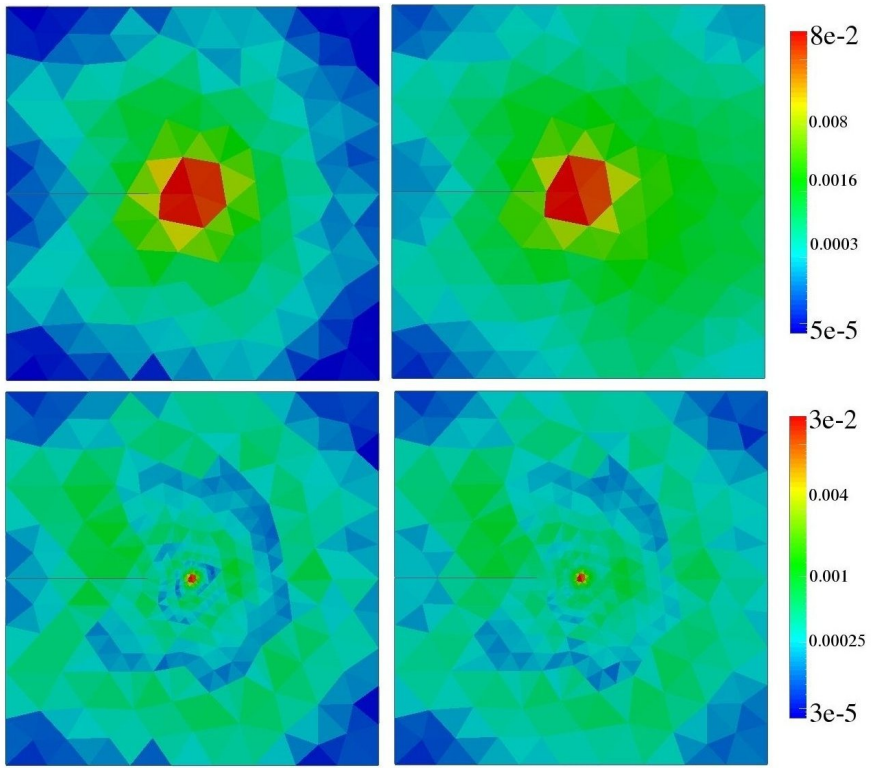


Figure 2.7 – Error estimation (left) and analytical error (right) on an initial mesh and after three mesh refinements

leading to a singularity of the stress tensor at the crack tip. For our test, we restrain ourselves to the domain $\Omega = (-\frac{1}{2}, \frac{1}{2}) \times (-\frac{1}{2}, \frac{1}{2})$ with a straight crack from $(-\frac{1}{2}, 0)$ (cf. Figure 2.6), and impose the analytical solution (2.53) as Dirichlet boundary condition on $\partial\Omega$ and the crack faces to obtain the discrete solution $\underline{u}_h$. The Young modulus and the Poisson ratio are set to $E = 1$ and $\nu = 0$, leading to the Lamé parameters $\mu = 0.5$ and $\lambda = 0$. Since in this purely mechanical test case there is no need for nondimensionalization, we omit the scaling factor E^{-1} in the error estimators. Figure 2.7 compares the distribution of the error estimators and the analytical error measured in the energy norm $\|\underline{u} - \underline{u}_h\|_{\text{en}} = a(\underline{u} - \underline{u}_h, \underline{u} - \underline{u}_h)^{1/2}$. Besides detecting the dominating error at the crack tip due to the singularity of $\underline{\underline{\sigma}}(\underline{u})$, the error estimators reflect the distribution of the analytical error in the whole domain, as can be seen in the lower panel for the finer mesh.

2.5.2 Poro-elastic analytical test

Let $\Omega = (0, 1) \times (0, 1)$. Following [20, 52], we consider the analytical solution of Biot's consolidation problem (2.1)

$$\underline{u}(t, x, y) = \cos(-\pi t) \begin{pmatrix} \cos(\pi x) \sin(\pi y) \\ \sin(\pi x) \cos(\pi y) \end{pmatrix}, \quad p(t, x, y) = \sin(-\pi t) \sin(\pi x) \sin(\pi y),$$

h^{-1}	$\eta_{\mathrm{sp},U}$		$\eta_{\mathrm{sp},P}$		$\|\underline{u}-\underline{u}_{h\tau}\|_U$		$\|p-p_{h\tau}\|_P$		e_{en}	I_{eff}
4	3.45e-2	—	1.58	—	3.44e-2	—	4.67e-1	—	5.12e-1	3.15
8	8.13e-3	2.09	7.62e-1	1.05	8.11e-3	2.08	2.33e-1	1.07	2.46e-1	3.14
16	1.96e-3	2.05	3.76e-1	1.02	2.00e-3	2.02	1.10e-1	1.02	1.21e-1	4.04
32	4.85e-4	2.01	1.87e-1	1.01	9.03e-4	1.15	5.46e-2	1.01	6.03e-2	3.16

Table 2.1 – Error estimators and analytical errors under space refinement with $t_F = 0.5$, τ = 5e-5

τ^{-1}	$\eta_{\mathrm{tm},U}$		$\eta_{\mathrm{tm},P}$		$\|\underline{u}-\underline{u}_{h\tau}\|_U$		$\|p-p_{h\tau}\|_P$		e_{en}	I_{eff}
4	4.73e-1	—	2.54e-1	—	1.96e-1	—	2.09e-1	—	2.32e-1	3.34
8	2.40e-1	0.78	1.40e-1	0.86	9.88e-2	1.00	1.14e-1	0.87	1.27e-1	3.35
16	1.20e-1	1.00	7.31e-2	0.94	4.94e-2	1.00	6.03e-2	0.92	6.94e-2	3.46
32	6.00e-2	1.00	3.74e-2	0.97	2.47e-2	1.00	3.17e-2	0.93	3.85e-2	3.76

Table 2.2 – Error estimators and analytical errors under time refinement with $t_F = 0.5$, $h = 1/128$

with $\kappa = 1$, $c_0 = 0$, and the Lamé coefficients $\mu = \lambda = 0.4$, yielding a Young modulus $E = 1$ and a Poisson ratio $\nu = 0.25$. The resulting source terms are given by

$$\underline{f}(t,x,y) = (2.4\pi^2\cos(-\pi t) + \pi\sin(-\pi t))\begin{pmatrix}\cos(\pi x)\sin(\pi y)\\ \sin(\pi x)\cos(\pi y)\end{pmatrix},$$

and $g = 0$.

To evaluate convergence rates under space or time uniform refinement, we measure the analytical error in the energy norm

$$|||(v,q)|||_{\mathrm{en}}^2 = \int_0^{t_F}\mathcal{B}((\underline{v},q),(t^\star\partial_t\underline{v},q))\mathrm{d}t = \frac{1}{2}t^\star\left(a(\underline{v},\underline{v})(t_F) - a(\underline{v},\underline{v})(t_0)\right) + t^\star\int_0^{t_F} d(q,q)\mathrm{d}t, \tag{2.54}$$

and the mechanical and hydraulic parts separately in the following norms:

$$\|\underline{v}\|_U^2 = \int_0^{t_F} a(\underline{v},\underline{v})\mathrm{d}t \quad \text{and} \quad \|q\|_P^2 = \int_0^{t_F} d(q,q)\mathrm{d}t. \tag{2.55}$$

where in this dimensionless test, the nondimensionalization parameters $t^\star$ and $l^\star$ are both equal to one, and we also omit the factor E.

Tables 2.1 and 2.2 compare the convergence rates under space and time refinement of the corresponding error estimators to the analytical error in the norms defined by (2.54) and (2.55). The last column shows the effectivity index defined by

$$I_{\mathrm{eff}} := \frac{\eta_{\mathrm{sp},U} + \eta_{\mathrm{sp},P} + \eta_{\mathrm{tm},U} + \eta_{\mathrm{tm},P}}{|||(\underline{u}-\underline{u}_{h\tau}, p-p_{h\tau})|||_{\mathrm{en}}}. \tag{2.56}$$

For both the spatial and the temporal refinement, we obtain the expected convergence rates of the Taylor–Hood finite element method (2.14) with $k = 1$, and a backward Euler scheme in

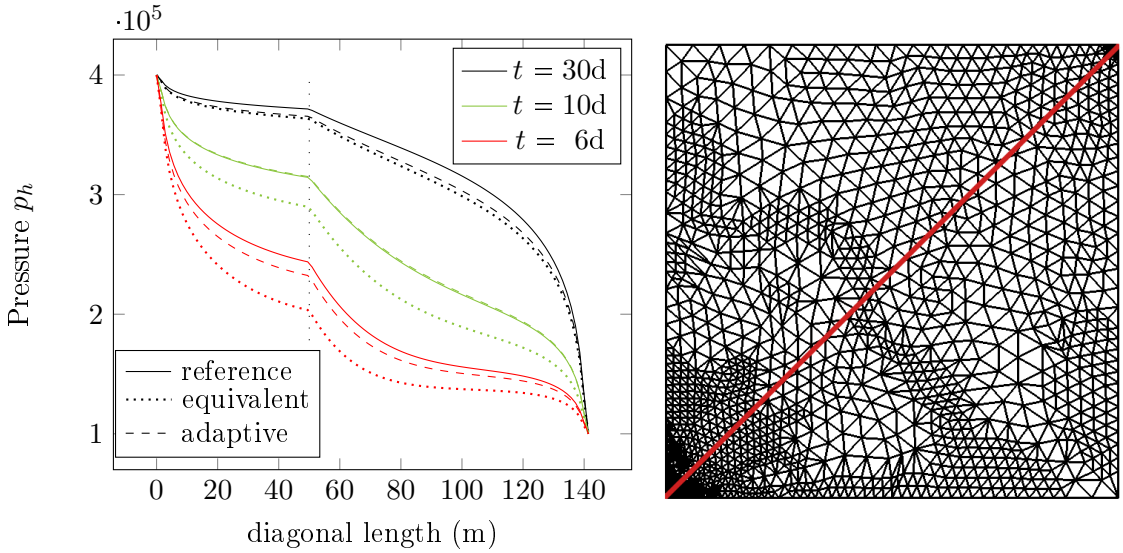


Figure 2.8 – Left: Comparison of the pressure in the quarter five-spot problem along the diagonal between the three algorithms in Table 2.3. Right: An adapted mesh at time $t = 10$ days

time. The last value of the analytical mechanical error under space refinement is due to the error in time discretization which starts playing a role for the finest mesh. We observe that the orders of magnitude of the mechanical and hydraulic part are comparable in this test, and that the effectivity index is dominated by the hydraulic part under space refinement, and by the mechanical part under time refinement.

2.5.3 Quarter five-spot problem

In this standard configuration considered in petroleum engineering, the injection of water at the center of a square domain and the production at the four corners is simulated on a quarter of the domain. In our test, this quarter is a square of 100m side length, divided into two parts with different mobilities; a circle around the injection point of radius 50m with $\kappa = 8 \cdot 10^{-9}\mathrm{m}^2\mathrm{Pa}^{-1}\mathrm{s}^{-1}$, and $\kappa = 10^{-9}\mathrm{m}^2\mathrm{Pa}^{-1}\mathrm{s}^{-1}$ in the rest of the domain. The Young modulus and the Poisson ratio are given by $E = 10^9\mathrm{Pa}$, $\nu = 0.3$, and we set $c_0 = 0$. The initial state is given by $\underline{\underline{\theta}}_0 = \underline{\underline{0}}$, $\underline{\phi}_0 = \underline{0}$ and $p_0 = 10^5\mathrm{Pa}$. During the computation time of 30 days, we set $p = p_0$ in the top right corner and $p = 4 \cdot 10^5\mathrm{Pa}$ in the bottom left corner, simulating the production and the injection respectively. The nondimensionalization parameters are $l^\star = 140\mathrm{m}$ and $t^\star = 1\mathrm{h}$. The problem is dominated by hydraulic processes.

	# space-time unknowns	# iterations	η_{sp}	η_{tm}	$\eta_{\mathrm{sp}}+\eta_{\mathrm{tm}}$
reference	13,754,520	120	0.204	0.315	0.519
equivalent	973,620	45	1.14	1.54	2.68
adaptive	846,174	71	0.462	0.507	0.969

Table 2.3 – The three computations in our test for the quarter five-spot problem

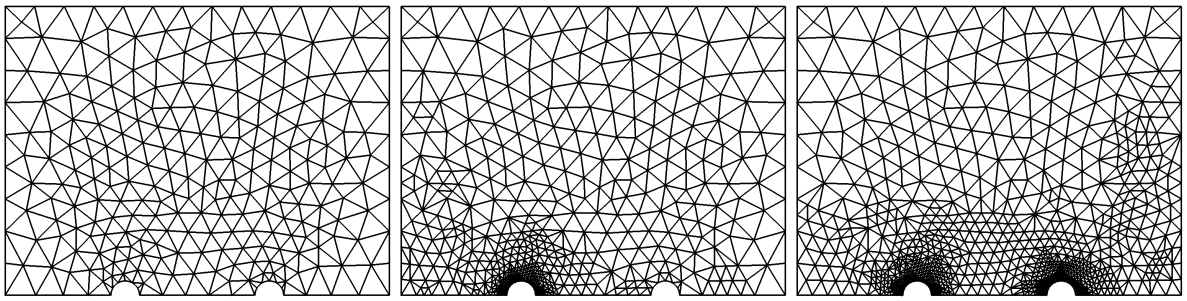


Figure 2.9 – Initial mesh (left) and meshes at the end of the first (center) and the second (right) excavation in the adaptive algorithm

We use Algorithm 2.18 to perform space-time adaptivity. We start with an initial mesh of 10,638 vertices and with an initial time step of $\tau_0 = 12\text{h}$. For the space-time error balancing, we set $\gamma_{tm} = 0.8$ and $\Gamma_{tm} = 1.3$ and fix the error limit for each time step to $\text{crit}^n = 0.005\tau_n$. We compare the performance of the adaptive algorithm to two static computations (i.e. with fixed meshes and time steps), one, called equivalent, where the discretization is chosen in a way to have approximately the same number of space-time unknowns as in the adaptive algorithm, and one where the discretization is very fine, so its solution can be taken as a reference solution. Table 2.3 compares the number of space-time unknowns and performed iterations (i.e. the number of time steps, counting repetitions in the adaptive algorithm), and the values of the error estimators of the three computations.

The left graphic in Figure 2.8 shows the discrete pressure along the diagonal going from the bottom left to the top right of the domain (as indicated in the right graphic) at three different times obtained by the static computations (solid and dotted lines) and the adaptive algorithm (dashed lines). The loosely dotted vertical line marks the edge between the two parts of Ω with different permeabilities. At each of these times, the discrete solution of the adaptive algorithm is closer to the reference solution than the equivalent computation using a fixed mesh and time step. At the last time step, all the results get closer as the solution converges in time to a constant state.

2.5.4 Excavation damage test

In the context of the conception of a radioactive waste repository site, the excavation of tunnels destined to contain waste packages is numerically simulated. The domain Ω is a 80m $\times$ 60m quadrilateral, vertical cutout of the rock, in which two galleries are digged time-delayed in the z-direction, first left, then right. Both excavations take 17.4 days ($1.5{\cdot}10^6$s) and the second one starts 11.6 days (10^6s) after the end of the first one. For both excavations we first calculate the initial total equilibrium of the hole-free geometry. Then the digging is simulated by linearly decreasing boundary conditions on the tunnel (convergence confinement method). These are of Neumann type for the mechanical part and of Dirichlet type for the hydraulic part and start with the total stress measured at the equilibrium state and the pressure $p_0 = 4.7\text{MPa}$. Homo-

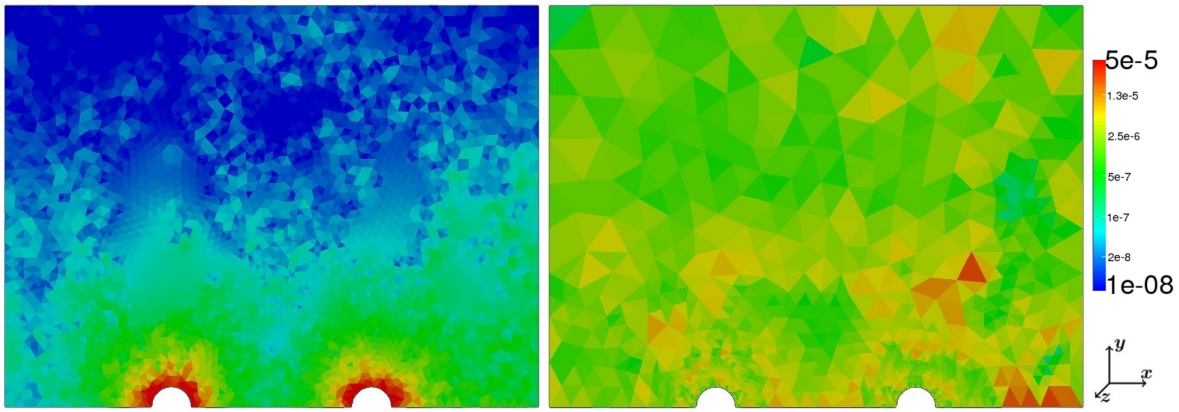


Figure 2.10 – Spatial discretization error estimators at t_F on a fixed mesh (left, 29,275 dofs) and on the last mesh of an adaptive algorithm (right, 15,064 dofs)

geneous Dirichlet boundary conditions for the y-component of the displacement and $p = p_0$ are imposed on the bottom of Ω (except for the tunnel parts), while on the top, the left and the right sides of Ω, we set $\underline{\underline{\theta}}\underline{n} = \underline{\underline{\theta}}_{\mathrm{ref}}\underline{n}$ with $(\underline{\underline{\theta}}_{\mathrm{ref},xx}, \underline{\underline{\theta}}_{\mathrm{ref},yy}, \underline{\underline{\theta}}_{\mathrm{ref},xy}) := (-11\mathrm{MPa}, -15.4\mathrm{MPa}, 0)$ and $p = p_0$. These boundary conditions have to be taken into account for the stress reconstruction (cf. Remark 2.7) and the a posteriori error estimate (cf. Remark 2.15). The initial fluxes are given by $\underline{\underline{\theta}}_0 = \underline{\underline{\theta}}_{\mathrm{ref}}$ and $\underline{\phi}_0 = \underline{0}$, while the initial pressure is p_0. The parameters describing the rock are the Young modulus $E = 5800\mathrm{MPa}$, the Poisson ratio $\nu = 0.3$, the specific storage coefficient $c_0 = 0$, and the hydraulic mobility $\kappa = 10^{-13}\mathrm{m}^2\mathrm{Pa}^{-1}\mathrm{s}^{-1}$. For the nondimensionalization of the problem, we used, along with E, the parameters $t^\star = 1\mathrm{h}$ and $l^\star = 100\mathrm{m}$.

The performance of Algorithm 2.18 is tested on four different initial meshes with $\mathrm{crit}^n = 7 \cdot 10^{-3}\tau_n$ for the coarsest one and, with the mesh getting finer, $\mathrm{crit}^n = 4{\cdot}10^{-3}\tau_n$, $\mathrm{crit}^n = 2{\cdot}10^{-3}\tau_n$ and $\mathrm{crit}^n = 1 \cdot 10^{-3}\tau_n$. In all the calculations we fix $\gamma_{\mathrm{tm}} = 0.8$, $\Gamma_{\mathrm{tm}} = 1.5$ and $\tau_0 = 3.9\mathrm{d}$. Figure 2.9 illustrates the evolution of the second coarsest mesh with $\mathrm{crit}^n = 4 \cdot 10^{-3}\tau_n$. During the first excavation, the refinement takes only place around the left tunnel, whereas the area around the right tunnel is only refined after the beginning of the second excavation. The calculations resulting from the adaptive algorithm are compared to calculations with fixed meshes and time steps. Each of these meshes is slightly finer around the tunnels than in the rest of Ω, and the time steps are chosen in a way that $\eta_{\mathrm{sp}} \approx \eta_{\mathrm{tm}}$. Figure 2.10 compares the spatial discretization error estimators at the final time t_F of the static algorithm to those of the adaptive algorithm, which are much more evenly distributed over the domain. Furthermore, the left graphic in Figure 2.11 shows that in our test, the use of the adaptive algorithm reduces the number of space-time-unknowns for a similar value of the error estimator.

In the right graphic of Figure 2.11, we plot the evolution of the error estimators in the two computations circled in the left graphic. Each mark stands for an iteration and shows the error estimate e^n of the current time interval divided by τ_n. For the plain algorithm, an iteration is

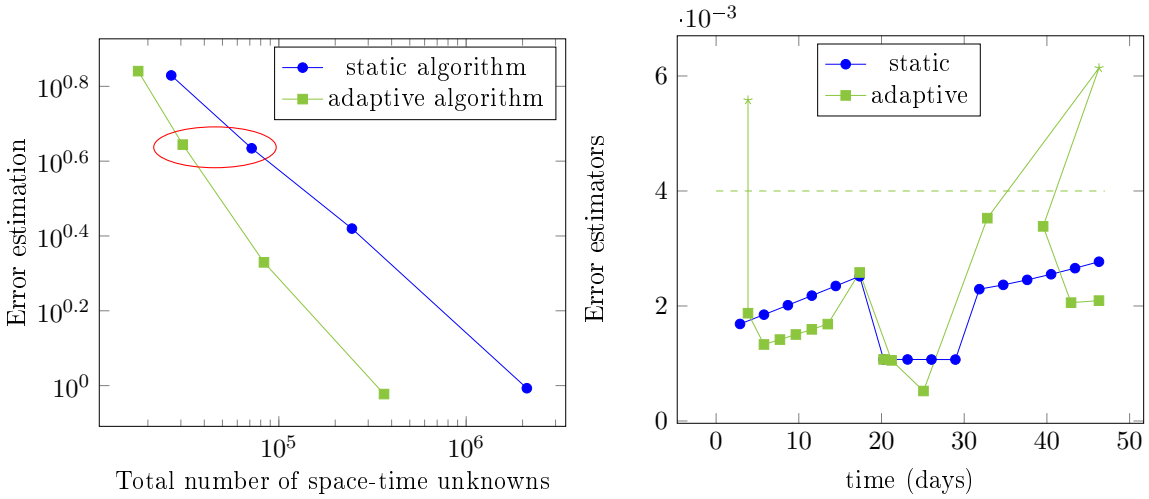


Figure 2.11 – Comparison between a static algorithm with fixed mesh and time step and the adaptive algorithm 2.18 for the excavation damage test

equal to a time step. The adaptive algorithm recalculates the solution at a time step whenever the error estimate lies over crit^n (illustrated by the dashed line) by refining τ_n or the mesh (or both). Thus, only the square shaped points in the graphic contribute to the overall error estimate. In the consolidation phase between the two excavations (from $t = 17.4$ days to $t = 29$ days), the mesh is slightly coarsened and the time step considerably increased, since the dominating error source in this phase is the spatial discretization.

2.5.5 Conclusion

The analytical test cases show that the distribution and convergence rates of our error estimators reflect those of the analytical error. The efficiency of Algorithm 2.18 has been illustrated in industrial tests, where the number of space-time unknowns is considerably decreased for a comparable overall error estimate. We also observe that the price for computing the flux reconstructions can be substantially reduced by pre-processing, a task that is fully parallelizable. As shown in the first test, the stress reconstruction and a posteriori estimate presented in this work are directly applicable to pure linear elasticity problems. The second test shows that the presented error estimate also delivers sharp bounds (as reflected by moderate effectivity indices) of more accessible error measures computed using energy-type norms. In the third and fourth tests, comparing the proportions of the estimators for the hydraulic and the mechanical parts reflects the physical properties of the problem: in the quarter five-spot test, the dominating estimators are those for the hydraulic part; for the excavation damage test, they are approximately of the same order of magnitude, with the mechanical estimator dominating in regions of stress concentration.

Chapter 3

Equilibrated stress tensor reconstruction and a posteriori error estimation for nonlinear elasticity

This chapter consists of an article in preparation, written with Michele Botti.

Contents

Abstract

We consider hyperelastic problems and their numerical solution using a conforming finite element discretization and some iterative linearization algorithm. For these problems, we present equilibrated, weakly symmetric, $H(\mathrm{div})$-conforming stress tensor reconstructions, obtained from local problems on patches around vertices using the Arnold–Falk–Winther finite element spaces. We distinguish two stress reconstructions, one for the discrete stress and one representing the linearization error. The reconstructions are independent of the mechanical behavior law. Based on these stress tensor reconstructions we derive an a posteriori error estimate distinguishing the discretization, linearization, and quadrature error estimates, and propose an adaptive algorithm balancing these different error sources. We prove the efficiency of the estimate, and confirm it on a numerical test with analytical solution for the linear elasticity problem. We then apply the adaptive algorithm to a more application-oriented test, considering the Hencky-Mises and an isotropic damage model.

3.1 Introduction

In this work we develop equilibrated $H(\mathrm{div})$-conforming stress tensor reconstructions for a class of (linear and) nonlinear elasticity problems in the small deformation regime. Based on these reconstructions, we can derive an a posteriori error estimate distinguishing the discretization and linearization errors for conforming discretizations of the problem.

Let $\Omega \in \mathbb{R}^d$, $d \in \{2,3\}$, be a bounded, simply connected polyhedron, which is occupied by a body subjected to a volumetric force field $\underline{f} \in \underline{L}^2(\Omega)$. For the sake of simplicity, we assume that the body is fixed on its boundary $\partial\Omega$. The nonlinear elasticity problem consists in finding a vector-valued displacement field $\underline{u} : \Omega \to \mathbb{R}^d$ solving

$$-\underline{\nabla} \cdot \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}) = \underline{f} \qquad \text{in } \Omega, \tag{3.1a}$$

$$\underline{u} = \underline{0} \qquad \text{on } \partial\Omega, \tag{3.1b}$$

where $\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u} = \frac{1}{2}((\underline{\underline{\nabla}} \underline{u})^T + \underline{\underline{\nabla}} \underline{u})$ denotes the symmetric gradient and expresses the strain tensor associated to $\underline{u}$. The stress-strain law $\underline{\underline{\sigma}} : \Omega \times \mathbb{R}^{d\times d}_{\mathrm{sym}} \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$ is assumed to satisfy regularity requirements inspired by [25, 86, 87]. Problem (3.1) describes the mechanical behavior of soft materials [103] and metal alloys [88]. Examples of stress-strain relations of common use in the engineering practice are given in Section 3.2. In these applications, the solution is often approximated using H^1-conforming finite elements. For nonlinear mechanical behavior laws, the resulting discrete nonlinear equation can then be solved using an iterative linearization algorithm yielding at each iteration a linear algebraic system to be solved, until the residual of the nonlinear equation lies under a predefined threshold.

In this paper we develop an a posteriori error estimate allowing to distinguish between the

error stemming from the linearization of the problem and the one due to its discretization, as proposed in [55] for nonlinear diffusion problems. Thanks to this distinction we can, at each iteration, compare these two error contributions and stop the linearization algorithm once its contribution is negligible compared to the discretization error.

The a posteriori error estimate is based on equilibrated stress reconstructions. It is well known that, in contrast to the analytical solution, the discrete stress tensor resulting from the conforming finite element method does not have continuous normal components across mesh interfaces, and that its divergence is not locally in equilibrium with the source term $\underline{f}$ on mesh elements. In this paper we consider the stress tensor reconstruction proposed in [94] for linear elasticity to restore these two properties. This reconstruction uses the Arnold–Falk–Winther mixed finite element spaces [9], leading to weakly symmetric tensors . In [94] this reconstruction is compared to a similar reconstruction introduced in [95] using the Arnold–Winther finite element spaces [10], yielding a symmetric tensor, and very good agreement was observed while saving substantial computational effort. In Section 3.3 we apply this reconstruction to the nonlinear case by constructing two stress tensors: one playing the role of the discrete stress and one expressing the linearization error. They are obtained by summing up the solutions of constrained minimization problems on cell patches around each mesh vertex, so that they are $H(\mathrm{div})$-conforming and the sum of the two reconstructions verifies locally the mechanical equilibrium (3.1a). The patch-wise equilibration technique was introduced in [28,38] for the Poisson problem using the Raviart–Thomas finite element spaces. In [48] it is extended to linear elasticity without any symmetry constraint by using linewise Raviart–Thomas reconstructions. Elementwise reconstructions from local Neumann problems requiring some pre-computations to determine the normal fluxes to obtain an equilibrated stress tensor can be found in [4,36,73,84], whereas in [82] the direct prescription of the degrees of freedom in the Arnold–Winther finite element space is considered.

Based on the equilibrated stress reconstructions, we develop the a posteriori error estimate in Section 3.4 and prove that this error estimate is efficient, meaning that, up to a generic constant, it is also a local lower bound for the error. The idea goes back to [89] and was advanced amongst others by [3, 69, 71, 93] for the upper bound. Local lower error bounds are derived in [28, 38, 54, 56, 75]. Using equilibrated fluxes for a posteriori error estimation offers several advantages. The first one is, as mentioned above, the possible distinction and comparison of error components by expressing them in terms of fluxes. Secondly, the error upper bound is obtained with fully computable constants. In our case these constants depend only on the parameters of the stress-strain relation. Thirdly, since the estimate is based on the discrete stress (and not the displacement), it does not depend on the mechanical behavior law (except for the constant in the upper bound). Therefore, its implementation is independent and directly applicable to these laws, which makes the method convenient for FEM softwares in solid mechanics, which often provide a large choice of behavior laws. In addition, equilibrated error estimates were proven to be polynomial-degree robust for several linear problems in 2D,

as the Poisson problem in [27,56], linear elasticity in [48] and the related Stokes problem in [33] and recently in 3D in [57].

This paper is organized as follows. In Section 3.2 we first formulate the assumptions on the stress-strain function $\underline{\underline{\sigma}}$ and provide three examples of models used in the engineering practice. We then introduce the weak and the discrete formulations of problem (3.1) and its linearization, along with some useful notation. In Section 3.3 we present the equilibrated stress tensor reconstructions, first assuming that we solve the nonlinear discrete problem exactly, and then, based on this first reconstruction, distinguish its discrete and its linearization error part at each iteration of a linearization solver. In Section 3.4 we derive the a posteriori error estimate, again first assuming the exact solution of the discrete problem and then distinguishing the different error components. We then propose an algorithm equilibrating the error sources using adaptive stopping criteria for the linearization and adaptive remeshing. We finally show the efficiency of the error estimate. In Section 3.5 we evaluate the performance of the estimates for the three behavior laws given as examples on numerical test cases.

3.2 Setting

In this section we will give three examples of hyperelastic behavior laws, before writing the weak and the considered discrete formulation of problem (3.1).

3.2.1 Continuous setting

Assumption 3.1 (Stress-strain relation)**.** *We assume that the symmetric stress tensor* $\underline{\underline{\sigma}} : \mathbb{R}^{d\times d}_{\mathrm{sym}} \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$ *is continuous on* $\mathbb{R}^{d\times d}_{\mathrm{sym}}$ *and that* $\underline{\underline{\sigma}}(\underline{\underline{0}}) = \underline{\underline{0}}$*. Moreover, we assume that there exist real numbers* $C_{\mathrm{gro}}, C_{\mathrm{mon}} \in (0, +\infty)$ *such that, for all* $\underline{\underline{\tau}}, \underline{\underline{\eta}} \in \mathbb{R}^{d\times d}_{\mathrm{sym}}$*,*

$$|\underline{\underline{\sigma}}(\underline{\underline{\tau}})|_{d\times d} \leqslant C_{\mathrm{gro}} |\underline{\underline{\tau}}|_{d\times d}, \qquad \textit{(growth)} \tag{3.2a}$$

$$\Big(\underline{\underline{\sigma}}(\underline{\underline{\tau}}) - \underline{\underline{\sigma}}(\underline{\underline{\eta}})\Big) : \Big(\underline{\underline{\tau}} - \underline{\underline{\eta}}\Big) \geqslant C^2_{\mathrm{mon}} |\underline{\underline{\tau}} - \underline{\underline{\eta}}|^2_{d\times d}, \quad \textit{(strong monotonicity)} \tag{3.2b}$$

where $\underline{\underline{\tau}} : \underline{\underline{\eta}} := \mathrm{tr}(\underline{\underline{\tau}}^T \underline{\underline{\eta}})$ *with* $\mathrm{tr}(\underline{\underline{\tau}}) := \sum_{i=1}^d \tau_{ii}$*, and* $|\underline{\underline{\tau}}|^2_{d\times d} = \underline{\underline{\tau}} : \underline{\underline{\tau}}$*.*

We next discuss a number of meaningful stress-strain relations for hyperelastic materials that satisfy the above assumptions. Hyperelasticity is a type of constitutive model for ideally elastic materials in which the stress is determined by the current state of deformation by deriving a stored energy density function $\Psi : \mathbb{R}^{d\times d}_{\mathrm{sym}} \to \mathbb{R}$, namely

$$\underline{\underline{\sigma}}(\underline{\underline{\tau}}) := \frac{\partial \Psi(\underline{\underline{\tau}})}{\partial \underline{\underline{\tau}}}.$$

Example 3.2 (Linear elasticity). *The stored energy density function leading to the linear elasticity model is*

$$\Psi_{\mathrm{lin}}(\underline{\underline{\tau}}) := \frac{\lambda}{2}\,\mathrm{tr}(\underline{\underline{\tau}})^2 + \mu\,\mathrm{tr}(\underline{\underline{\tau}}^2), \tag{3.3}$$

where $\mu > 0$ and $\lambda \geqslant 0$ are the Lamé parameters. Deriving (3.3) *yields the usual Cauchy stress tensor*

$$\underline{\underline{\sigma}}(\underline{\underline{\tau}}) = \lambda\,\mathrm{tr}(\underline{\underline{\tau}})\underline{\underline{I}}_d + 2\mu\underline{\underline{\tau}}. \tag{3.4}$$

Being linear, the previous stress-strain relation clearly satisfies Assumption 3.1.

Example 3.3 (Hencky–Mises model). *The nonlinear Hencky–Mises model of [61, 80] corresponds to the stored energy density function*

$$\Psi_{\mathrm{hm}}(\underline{\underline{\tau}}) := \frac{\alpha}{2}\,\mathrm{tr}(\underline{\underline{\tau}})^2 + \Phi(\mathrm{dev}(\underline{\underline{\tau}})), \tag{3.5}$$

where $\mathrm{dev} : \mathbb{R}^{d\times d}_{\mathrm{sym}} \to \mathbb{R}$ *defined by* $\mathrm{dev}(\underline{\underline{\tau}}) = \mathrm{tr}(\underline{\underline{\tau}}^2) - \frac{1}{d}\,\mathrm{tr}(\underline{\underline{\tau}})^2$ *is the deviatoric operator. Here, $\alpha \in (0, +\infty)$ and $\Phi : [0, +\infty) \to \mathbb{R}$ is a function of class C^2 satisfying, for some positive constants C_1, C_2, and C_3,*

$$C_1 \leqslant \Phi'(\rho) < \alpha, \quad |\rho\Phi''(\rho)| \leqslant C_2 \quad \text{and} \quad \Phi'(\rho) + 2\rho\Phi''(\rho) \geqslant C_3 \quad \forall \rho \in [0, +\infty). \tag{3.6}$$

We observe that taking $\alpha = \lambda + \frac{2}{d}\mu$ and $\Phi(\rho) = \mu\rho$ in (3.5) *leads to the linear case* (3.3). *Deriving the energy density function* (3.5) *yields*

$$\underline{\underline{\sigma}}(\underline{\underline{\tau}}) = \tilde{\lambda}(\mathrm{dev}(\underline{\underline{\tau}}))\,\mathrm{tr}(\underline{\underline{\tau}})\underline{\underline{I}}_d + 2\tilde{\mu}(\mathrm{dev}(\underline{\underline{\tau}}))\underline{\underline{\tau}}, \tag{3.7}$$

with nonlinear Lamé functions $\tilde{\mu}(\rho) := \Phi'(\rho)$ and $\tilde{\lambda}(\rho) := \alpha - \Phi'(\rho)$. Under conditions (3.6) *it can be proven that the previous stress-strain relation satisfies Assumption 3.1.*

In the previous example the nonlinearity of the model only depends on the deviatoric part of the strain. In the following model it depends on the term $\underline{\underline{\tau}} : \underset{\approx}{\underline{\underline{C}}}\,\underline{\underline{\tau}}$.

Example 3.4 (An isotropic reversible damage model). *The isotropic reversible damage model of [34] can also be interpreted in the framework of hyperelasticity by setting up the energy density function as*

$$\Psi_{\mathrm{dam}}(\underline{\underline{\tau}}) := \frac{(1 - D(\underline{\underline{\tau}}))}{2}\underline{\underline{\tau}} : \underset{\approx}{\underline{\underline{C}}}\,\underline{\underline{\tau}} + \Phi(D(\underline{\underline{\tau}})), \tag{3.8}$$

where $D : \mathbb{R}^{d\times d}_{\mathrm{sym}} \to [0, 1]$ is the scalar damage function and $\underset{\approx}{\underline{\underline{C}}}$ is a fourth-order symmetric and uniformly elliptic tensor, namely, for some positive constants C_* and C^*, it holds

$$C_*|\underline{\underline{\tau}}|^2_{d\times d} \leqslant \underset{\approx}{\underline{\underline{C}}}\,\underline{\underline{\tau}} : \underline{\underline{\tau}} \leqslant C^*|\underline{\underline{\tau}}|^2_{d\times d}, \quad \forall \underline{\underline{\tau}} \in \mathbb{R}^{d\times d}. \tag{3.9}$$

The function $\Phi : [0,1] \to \mathbb{R}$ *defines the relation between* $\underline{\underline{\tau}}$ *and* D *by* $\frac{\partial \phi}{\partial D} = \frac{1}{2}\underline{\underline{\tau}} : \underline{\underline{\underline{\underline{C}}}}\,\underline{\underline{\tau}}$. *The resulting stress-strain relation reads*

$$\underline{\underline{\sigma}}(\underline{\underline{\tau}}) = (1 - D(\underline{\underline{\tau}}))\underline{\underline{\underline{\underline{C}}}}\,\underline{\underline{\tau}}. \tag{3.10}$$

If there exists a continuous function $f : [0, +\infty) \to [a, b]$ *for some* $0 < a \leqslant b \leqslant 1$, *such that* $s \in [0, +\infty) \to sf(s)$ *is strictly increasing and, for all* $\underline{\underline{\tau}} \in \mathbb{R}^{d\times d}_{\mathrm{sym}}$, $D(\underline{\underline{\tau}}) = 1 - f(\underline{\underline{\tau}} : \underline{\underline{\underline{\underline{C}}}}\,\underline{\underline{\tau}})$, *the damage model constitutive relation satisfies Assumption 3.1.*

Before presenting the variational formulation of problem (3.1), some useful notations are introduced. For $X \subset \overline{\Omega}$, we respectively denote by $(\cdot,\cdot)_X$ and $\|\cdot\|_X$ the standard inner product and norm in $L^2(X)$, with the convention that the subscript is omitted whenever $X = \Omega$. The same notation is used in the vector- and tensor-valued cases. $\underline{H}^1(\Omega)$ and $\underline{\underline{H}}(\mathrm{div}, \Omega)$ stand for the Sobolev spaces composed of vector-valued $\underline{L}^2(\Omega)$ functions with weak gradient in $\underline{\underline{L}}^2(\Omega)$, and tensor-valued $\underline{\underline{L}}^2(\Omega)$ functions with weak divergence in $\underline{L}^2(\Omega)$, respectively. Multiplying equation (3.1a) by a test function $\underline{v} \in \underline{H}^1_0(\Omega)$ and integrating by parts one has

$$(\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}), \underline{\underline{\nabla}}_{\mathrm{s}}\underline{v}) = (\underline{f}, \underline{v}). \tag{3.11}$$

Owing to the growth assumption (3.2a), for all $\underline{v}, \underline{w} \in \underline{H}^1(\Omega)$, the form

$$a(\underline{v}, \underline{w}) := (\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{v}), \underline{\underline{\nabla}}_{\mathrm{s}}\underline{w}) \tag{3.12a}$$

is well defined and, from equation (3.11), we can derive the following weak formulation of (3.1):

$$\text{Given } \underline{f} \in \underline{L}^2(\Omega), \text{ find } \underline{u} \in \underline{H}^1_0(\Omega) \text{ s.t., } \forall \underline{v} \in \underline{H}^1_0(\Omega),\ a(\underline{u}, \underline{v}) = (\underline{f}, \underline{v}). \tag{3.13}$$

From (3.13) it is clear that the analytical stress tensor $\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u})$ lies in the space $\underline{\underline{H}}_{\mathrm{s}}(\mathrm{div}, \Omega) := \{\underline{\underline{\tau}} \in \underline{\underline{L}}^2(\Omega) \mid \underline{\nabla} \cdot \underline{\underline{\tau}} \in \underline{L}^2(\Omega) \text{ and } \underline{\underline{\tau}} \text{ is symmetric}\}$.

3.2.2 Discrete setting

The discretization (3.13) is based on a conforming triangulation $\mathcal{T}_h$ of Ω, i.e. a set of closed triangles or tetrahedra with union $\overline{\Omega}$ and such that, for any distinct $T_1, T_2 \in \mathcal{T}_h$, the set $T_1 \cap T_2$ is either a common edge, a vertex, the empty set or, if $d = 3$, a common face. We assume that $\mathcal{T}_h$ verifies the minimum angle condition, i.e., there exists $\alpha_{\min} > 0$ uniform with respect to all considered meshes such that the minimum angle α_T of each $T \in \mathcal{T}_h$ satisfies $\alpha_T \geqslant \alpha_{\min}$. The set of vertices of the mesh is denoted by $\mathcal{V}_h$; it is decomposed into interior vertices $\mathcal{V}_h^{\mathrm{int}}$ and boundary vertices $\mathcal{V}_h^{\mathrm{ext}}$. For all $a \in \mathcal{V}_h$, $\mathcal{T}_a$ is the patch of elements sharing the vertex a, ω_a the corresponding open subdomain in Ω and $\mathcal{V}_a$ the set of vertices in ω_a. For all $T \in \mathcal{T}_h$, $\mathcal{V}_T$ denotes the set of vertices of T, h_T its diameter and $\underline{n}_T$ its unit outward normal vector.

For all $p \in \mathbb{N}$ and all $T \in \mathcal{T}_h$, we denote by $\mathbb{P}^p(T)$ the space of d-variate polynomials in T of total degree at most p and by $\mathbb{P}^p(\mathcal{T}_h) = \{\varphi \in L^2(\Omega) \mid \varphi_{|T} \in \mathbb{P}^p(T)\ \forall T \in \mathcal{T}_h\}$ the corresponding broken space over $\mathcal{T}_h$. In the same way we denote by $\underline{\mathbb{P}}^p(T)$ and $\underline{\underline{\mathbb{P}}}^p(T)$, respectively, the space of vector-valued and tensor-valued polynomials of total degree p over T, and by $\underline{\mathbb{P}}^p(\mathcal{T}_h)$ and $\underline{\underline{\mathbb{P}}}^p(\mathcal{T}_h)$ the corresponding broken spaces over $\mathcal{T}_h$.

In this work we will focus on conforming discretizations of problem (3.11) of polynomial degree $p \geqslant 2$ to avoid numerical locking, cf [107]. The discrete formulation reads: find $\underline{u}_h \in \underline{H}^1_0(\Omega) \cap \underline{\mathbb{P}}^p(\mathcal{T}_h)$ such that

$$\forall \underline{v}_h \in \underline{H}^1_0(\Omega) \cap \underline{\mathbb{P}}^p(\mathcal{T}_h), \quad a(\underline{u}_h, \underline{v}_h) = (\underline{f}, \underline{v}_h). \tag{3.14}$$

This problem is usually solved using some iterative linearization algorithm defining at each iteration $k \geqslant 1$ a linear approximation $\underline{\underline{\sigma}}^{k-1}$ of $\underline{\underline{\sigma}}$. Then the linearized formulation reads: find $\underline{u}_h^k \in \underline{H}^1_0(\Omega) \cap \underline{\mathbb{P}}^p(\mathcal{T}_h)$ such that

$$\forall \underline{v}_h \in \underline{H}^1_0(\Omega) \cap \underline{\mathbb{P}}^p(\mathcal{T}_h), \quad (\underline{\underline{\sigma}}^{k-1}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k), \underline{\underline{\nabla}}_{\mathrm{s}} \underline{v}_h) = (\underline{f}, \underline{v}_h). \tag{3.15}$$

For the Newton algorithm the linearized stress tensor is defined as

$$\underline{\underline{\sigma}}^{k-1}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k) := \frac{\partial \underline{\underline{\sigma}}(\underline{\underline{\tau}})}{\partial \underline{\underline{\tau}}}\Big|_{\underline{\underline{\tau}} = \underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^{k-1}} \underline{\underline{\nabla}}_{\mathrm{s}} (\underline{u}_h^k - \underline{u}_h^{k-1}) + \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^{k-1}). \tag{3.16}$$

3.3 Equilibrated stress reconstruction

In general, the discrete stress tensor $\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h)$ resulting from (3.14) does not lie in $\underline{\underline{H}}(\mathrm{div}, \Omega)$ and thus cannot verify the equilibrium equation (3.1a). In this section we will reconstruct from $\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h)$ a discrete stress tensor $\underline{\underline{\sigma}}_h$ satisfying these properties. Based on this reconstruction, we then devise two equilibrated stress tensors representing the discrete stress and the linearization error respectively, which will be useful for the distinction of error components in the a posteriori error estimate of Section 3.4.2.

3.3.1 Patchwise construction in the Arnold–Falk–Winther mixed finite element spaces

Let us for now suppose that $\underline{u}_h$ solves (3.14) exactly, before considering iterative linearization methods such as (3.15) in Section 3.3.2. For the stress reconstruction we will use mixed finite element formulations on patches around mesh vertices in the spirit of [94, 95]. The mixed finite elements based on the dual formulation of (3.1a) will provide a stress tensor lying in $\underline{\underline{H}}(\mathrm{div}, \Omega)$. A global computation is too expensive for this post-processing reconstruction, so we solve local problems on patches of elements around mesh vertices. The goal is to obtain

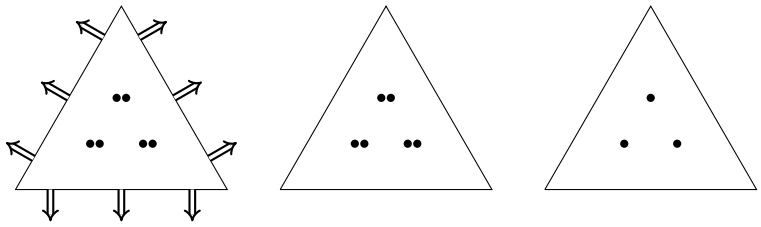


Figure 3.1 – Element diagrams for $(\underline{\underline{\Sigma}}_T, \underline{V}_T, \underline{\underline{\Lambda}}_T)$ in the case $d = q = 2$

a stress tensor $\underline{\underline{\sigma}}_h$ in a suitable (i.e. $H(\mathrm{div})$-conforming) finite element space by summing up these local solutions. The local problems are posed such that this global stress tensor is close to the discrete stress tensor $\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h)$ obtained from (3.14), and that it satisfies the mechanical equilibrium on each element.

In [95] the stress tensor is reconstructed in the Arnold–Winther finite element space [10], directly providing symmetric tensors, but requiring high computational effort. In this work, as in [94], we weaken the symmetry constraint and impose it weakly, as proposed in [9]: for each element $T \in \mathcal{T}_h$, the local Arnold–Falk–Winther mixed finite element spaces of degree $q \geqslant 1$ hinge on the Brezzi–Douglas–Marini mixed finite element spaces [29] for each line of the stress tensor and are defined by

$$\begin{aligned}
\underline{\underline{\Sigma}}_T &:= \underline{\underline{\mathbb{P}}}^q(T), \\
\underline{V}_T &:= \underline{\mathbb{P}}^{q-1}(T), \\
\underline{\underline{\Lambda}}_T &:= \{\underline{\underline{\mu}} \in \underline{\underline{\mathbb{P}}}^{q-1}(T) \mid \underline{\underline{\mu}} = -\underline{\underline{\mu}}^T\}.
\end{aligned}$$

For $q = 2$, the degrees of freedom are displayed in Figure 3.1. On a patch ω_a the global space $\underline{\underline{\Sigma}}_h(\omega_a)$ is the subspace of $\underline{\underline{H}}(\mathrm{div}, \omega_a)$ composed of functions belonging piecewise to $\underline{\underline{\Sigma}}_T$. The spaces $\underline{V}_h(\omega_a)$ and $\underline{\underline{\Lambda}}_h(\omega_a)$ consist of functions lying piecewise in $\underline{V}_T$ and $\underline{\underline{\Lambda}}_T$ respectively, with no continuity conditions between two elements.

Let now $q := p$. On each patch we need to consider subspaces where a zero normal component is enforced on the stress tensor on the boundary of the patch, so that the sum of the local solutions will have continous normal component across any mesh face inside Ω. Since the boundary condition in the exact problem prescribes the displacement and not the normal stress, we distinguish the case whether a is an interior vertex or a boundary vertex. If $a \in \mathcal{V}_h^{\mathrm{int}}$ we set

$$\underline{\underline{\Sigma}}_h^a := \{\underline{\underline{\tau}}_h \in \underline{\underline{\Sigma}}_h(\omega_a) \mid \underline{\underline{\tau}}_h \underline{n}_{\omega_a} = \underline{0} \text{ on } \partial\omega_a\}, \tag{3.18a}$$

$$\underline{V}_h^a := \{\underline{v}_h \in \underline{V}_h(\omega_a) \mid (\underline{v}_h, \underline{z})_{\omega_a} = 0 \ \forall \underline{z} \in \underline{RM}^d\}, \tag{3.18b}$$

$$\underline{\underline{\Lambda}}_h^a := \underline{\underline{\Lambda}}_h(\omega_a), \tag{3.18c}$$

where $\underline{RM}^2 := \{\underline{b} + c(x_2, -x_1)^T \mid \underline{b} \in \mathbb{R}^2, c \in \mathbb{R}\}$ and $\underline{RM}^3 := \{\underline{b} + \underline{a} \times \underline{x} \mid \underline{b} \in \mathbb{R}^3, \underline{a} \in \mathbb{R}^3\}$ are

the spaces of rigid-body motions respectively for $d = 2$ and $d = 3$. If $a \in \mathcal{V}_h^{\rm ext}$ we set

$$\underline{\underline{\Sigma}}_h^a := \{\underline{\underline{\tau}}_h \in \underline{\underline{\Sigma}}_h(\omega_a) \mid \underline{\underline{\tau}}_h \underline{n}_{\omega_a} = \underline{0} \text{ on } \partial\omega_a \backslash \partial\Omega\}, \tag{3.18d}$$

$$\underline{V}_h^a := \underline{V}_h(\omega_a), \tag{3.18e}$$

$$\underline{\underline{\Lambda}}_h^a := \underline{\underline{\Lambda}}_h(\omega_a). \tag{3.18f}$$

For each vertex $a \in \mathcal{V}_h$ we define its hat function $\psi_a \in \mathbb{P}^1(\mathcal{T}_h)$ as the piecewise linear function taking value one at the vertex a and zero on all other mesh vertices.

Construction 3.5 (Stress tensor reconstruction)**.** *Let $\underline{u}_h$ solve* (3.14)*. For each $a \in \mathcal{V}_h$ find $(\underline{\underline{\sigma}}_h^a, \underline{r}_h^a, \underline{\underline{\lambda}}_h^a) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a$ such that for all $(\underline{\underline{\tau}}_h, \underline{v}_h, \underline{\underline{\mu}}_h) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a$,*

$$(\underline{\underline{\sigma}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} + (\underline{r}_h^a, \underline{\nabla} \cdot \underline{\underline{\tau}}_h)_{\omega_a} + (\underline{\underline{\lambda}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} = (\psi_a \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\rm s} \underline{u}_h), \underline{\underline{\tau}}_h)_{\omega_a}, \tag{3.19a}$$

$$(\underline{\nabla} \cdot \underline{\underline{\sigma}}_h^a, \underline{v}_h)_{\omega_a} = (-\psi_a \underline{f} + \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\rm s} \underline{u}_h) \underline{\nabla} \psi_a, \underline{v}_h)_{\omega_a}, \tag{3.19b}$$

$$(\underline{\underline{\sigma}}_h^a, \underline{\underline{\mu}}_h)_{\omega_a} = 0. \tag{3.19c}$$

Then, extending $\underline{\underline{\sigma}}_h^a$ by zero outside ω_a, set $\underline{\underline{\sigma}}_h := \sum_{a \in \mathcal{V}_h} \underline{\underline{\sigma}}_h^a$.

For interior vertices, the source term in (3.19b) has to verify the Neumann compatibility condition

$$(-\psi_a \underline{f} + \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\rm s} \underline{u}_h) \underline{\nabla} \psi_a, \underline{z})_{\omega_a} = 0 \quad \forall \underline{z} \in \underline{RM}^d. \tag{3.20}$$

Taking $\psi_a \underline{z}$ as a test function in (3.14), we see that (3.20) holds and we obtain the following result.

Lemma 3.6 (Properties of $\underline{\underline{\sigma}}_h$)**.** *Let $\underline{\underline{\sigma}}_h$ be prescribed by Construction 3.5. Then $\underline{\underline{\sigma}}_h \in \underline{\underline{H}}(\mathrm{div}, \Omega)$, and for all $T \in \mathcal{T}_h$, the following holds:*

$$(\underline{f} + \underline{\nabla} \cdot \underline{\underline{\sigma}}_h, \underline{v})_T = 0 \quad \forall \underline{v} \in \underline{V}_T \quad \forall T \in \mathcal{T}_h. \tag{3.21}$$

Proof. All the fields $\underline{\underline{\sigma}}_h^a$ are in $\underline{\underline{H}}(\mathrm{div}, \omega_a)$ and satisfy appropriate zero normal conditions so that their zero-extension to Ω is in $\underline{\underline{H}}(\mathrm{div}, \Omega)$. Hence, $\underline{\underline{\sigma}}_h \in \underline{\underline{H}}(\mathrm{div}, \Omega)$. Let us prove (3.21). Since (3.20) holds for all $a \in \mathcal{V}_h^{\rm int}$, we infer that (3.19b) is actually true for all $\underline{v}_h \in \underline{V}_h(\omega_a)$. The same holds if $a \in \mathcal{V}_h^{\rm ext}$ by definition of $\underline{V}_h^a$. Since $\underline{V}_h(\omega_a)$ is composed of piecewise polynomials that can be chosen independently in each cell $T \in \mathcal{T}_a$, and using $\underline{\underline{\sigma}}_h|_T = \sum_{a \in \mathcal{V}_T} \underline{\underline{\sigma}}_h^a|_T$ and the partition of unity $\sum_{a \in \mathcal{V}_T} \psi_a = 1$, we infer that $(\underline{f} + \underline{\nabla} \cdot \underline{\underline{\sigma}}_h, \underline{v})_T = 0$ for all $\underline{v} \in \underline{V}_T$ and all $T \in \mathcal{T}_h$. □

3.3.2 Discretization and linearization error stress reconstructions

Let now, for $k \geqslant 1$, $\underline{u}_h^k$ solve (3.15). We will construct two different equilibrated $H(\mathrm{div})$-conforming stress tensors. The first one, $\underline{\underline{\sigma}}_{h,\rm disc}^k$, represents as above the discrete stress tensor

$\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)$, for which we will have to modify Construction 3.5, because the Neumann compatibility condition (3.20) is not satisfied anymore. The second stress tensor $\underline{\underline{\sigma}}_{h,\mathrm{lin}}^k$ will be a measure for the linearization error and approximate $\underline{\underline{\sigma}}^{k-1}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k) - \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)$. The matrix resulting from the left side of (3.19) will stay unchanged and we will only modify the source terms.

We denote by $\overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)$ the L^2-orthogonal projection of $\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)$ onto $\underline{\underline{\mathbb{P}}}^{p-1}(\mathcal{T}_h)$ such that $(\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k) - \overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k), \underline{\tau}_h) = 0$ for any $\underline{\tau}_h \in \underline{\underline{\mathbb{P}}}^{p-1}(\mathcal{T}_h)$.

Construction 3.7 (Discrete stress reconstruction)**.** *For each $a \in \mathcal{V}_h$ solve* (3.19) *with $\underline{u}_h^k$ instead of $\underline{u}_h$, $\overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)$ instead of $\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)$ and the source term in* (3.19b) *replaced by*

$$-\psi_a \underline{f} + \overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)\underline{\nabla}\psi_a - \underline{y}_{\mathrm{disc}}^k,$$

where $\underline{y}_{\mathrm{disc}}^k \in \underline{RM}^d$ is the unique solution of

$$(\underline{y}_{\mathrm{disc}}^k, \underline{z})_{\omega_a} = -(\underline{f}, \psi_a \underline{z})_{\omega_a} + (\overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k), \underline{\underline{\nabla}}_{\mathrm{s}}(\psi_a \underline{z}))_{\omega_a} \quad \forall \underline{z} \in \underline{RM}^d. \tag{3.22}$$

The so obtained problem reads: find $(\underline{\underline{\sigma}}_h^a, \underline{r}_h^a, \underline{\underline{\lambda}}_h^a) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a$ such that for all $(\underline{\tau}_h, \underline{v}_h, \underline{\underline{\mu}}_h) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a$,

$$\begin{aligned}
(\underline{\underline{\sigma}}_h^a, \underline{\tau}_h)_{\omega_a} + (\underline{r}_h^a, \underline{\nabla} \cdot \underline{\tau}_h)_{\omega_a} + (\underline{\underline{\lambda}}_h^a, \underline{\tau}_h)_{\omega_a} &= (\psi_a \overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k), \underline{\tau}_h)_{\omega_a}, \\
(\underline{\nabla} \cdot \underline{\underline{\sigma}}_h^a, \underline{v}_h)_{\omega_a} &= (-\psi_a \underline{f} + \overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)\underline{\nabla}\psi_a - \underline{y}_{\mathrm{disc}}^k, \underline{v}_h)_{\omega_a}, \\
(\underline{\underline{\sigma}}_h^a, \underline{\underline{\mu}}_h)_{\omega_a} &= 0.
\end{aligned}$$

Then set $\underline{\underline{\sigma}}_{h,\mathrm{disc}}^k := \sum_{a \in \mathcal{V}_h} \underline{\underline{\sigma}}_h^a$.

Construction 3.8 (Linearization error stress reconstruction)**.** *For each $a \in \mathcal{V}_h$ solve* (3.19) *with $\underline{u}_h^k$ instead of $\underline{u}_h$, the source term in* (3.19a) *replaced by*

$$\psi_a(\underline{\underline{\sigma}}^{k-1}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k) - \overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)),$$

and the source term in (3.19b) *replaced by*

$$(\underline{\underline{\sigma}}^{k-1}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k) - \overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k))\underline{\nabla}\psi_a + \underline{y}_{\mathrm{disc}}^k,$$

where $\underline{y}_{\mathrm{disc}}^k \in \underline{RM}^d$ is defined by (3.22)*. The corresponing local problem is to find $(\underline{\underline{\sigma}}_h^a, \underline{r}_h^a, \underline{\underline{\lambda}}_h^a) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a$ such that for all $(\underline{\tau}_h, \underline{v}_h, \underline{\underline{\mu}}_h) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a$,*

$$\begin{aligned}
(\underline{\underline{\sigma}}_h^a, \underline{\tau}_h)_{\omega_a} + (\underline{r}_h^a, \underline{\nabla} \cdot \underline{\tau}_h)_{\omega_a} + (\underline{\underline{\lambda}}_h^a, \underline{\tau}_h)_{\omega_a} &= (\psi_a(\underline{\underline{\sigma}}^{k-1}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k) - \overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)), \underline{\tau}_h)_{\omega_a}, \\
(\underline{\nabla} \cdot \underline{\underline{\sigma}}_h^a, \underline{v}_h)_{\omega_a} &= ((\underline{\underline{\sigma}}^{k-1}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k) - \overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k))\underline{\nabla}\psi_a + \underline{y}_{\mathrm{disc}}^k, \underline{v}_h)_{\omega_a}, \\
(\underline{\underline{\sigma}}_h^a, \underline{\underline{\mu}}_h)_{\omega_a} &= 0.
\end{aligned}$$

Then set $\underline{\underline{\sigma}}_{h,\mathrm{lin}}^k := \sum_{a \in \mathcal{V}_h} \underline{\underline{\sigma}}_h^a$.

Notice that the role of $\underline{y}^k_{\mathrm{disc}}$ is to guarantee that, for interior vertices, the source terms in Constructions 3.7 and 3.8 satisfy the Neumann compatibility conditions

$$\begin{aligned}(-\psi_a \underline{f} + \overline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}^k_h)\underline{\nabla}\psi_a - \underline{y}^k_{\mathrm{disc}}, \underline{z})_{\omega_a} &= 0 \quad \forall \underline{z} \in \underline{RM}^d,\\ ((\underline{\underline{\sigma}}^{k-1}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}^k_h) - \overline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}^k_h))\underline{\nabla}\psi_a + \underline{y}^k_{\mathrm{disc}}, \underline{z})_{\omega_a} &= 0 \quad \forall \underline{z} \in \underline{RM}^d.\end{aligned}$$

Lemma 3.9 (Properties of the discretization and linearization error stress reconstructions)**.** *Let $\underline{\underline{\sigma}}^k_{h,\mathrm{disc}}$ and $\underline{\underline{\sigma}}^k_{h,\mathrm{lin}}$ be prescribed by Constructions 3.7 and 3.8. Then it holds*

1. $\underline{\underline{\sigma}}^k_{h,\mathrm{disc}}, \underline{\underline{\sigma}}^k_{h,\mathrm{lin}} \in \underline{\underline{H}}(\mathrm{div}, \Omega)$,
2. $(\underline{f} + \underline{\nabla} \cdot (\underline{\underline{\sigma}}^k_{h,\mathrm{disc}} + \underline{\underline{\sigma}}^k_{h,\mathrm{lin}}), \underline{v})_T = 0 \quad \forall \underline{v} \in \underline{V}_T \quad \forall T \in \mathcal{T}_h$,
3. *As the Newton solver converges,* $\underline{\underline{\sigma}}^k_{h,\mathrm{lin}} \to \underline{\underline{0}}$.

Proof. The proof is similar to the proof of Lemma 3.6. The first property is again satisfied due to the definition of $\underline{\underline{\Sigma}}^a_h$. In order to show that the second property holds, we add the two equations (3.19b) obtained for each of the constructions. The right hand side of this sum then reads $(-\psi_a \underline{f} + \underline{\underline{\sigma}}^{k-1}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}^k_h)\underline{\nabla}\psi_a, \underline{v}_h)_{\omega_a}$. Once again we can, for any $\underline{z} \in \underline{RM}^d$, take $\psi_a \underline{z}$ as a test function in (3.15) to show that this term is zero if $\underline{v}_h \in \underline{RM}^d$, and so the equation holds for all $\underline{v}_h \in \underline{V}_h(\omega_a)$. Then we proceed as in the proof of Lemma 3.6. □

3.4 A posteriori error estimate and adaptive algorithm

In this section we first derive an upper bound on the error between the analytical solution of (3.13) and the solution $\underline{u}_h$ of (3.14), in which we then identify and distinguish the discretization and linearization error components at each Newton iteration for the solution $\underline{u}^k_h$ of (3.15). Based on this distinction, we present an adaptive algorithm stopping the Newton iterations once the linearization error estimate is dominated by the estimate of the discretization error. Finally, in a more theoretical part, we show the effectivity of the error estimate.

3.4.1 Guaranteed upper bound

We measure the error in the energy norm

$$\|\underline{v}\|^2_{\mathrm{en}} := a(\underline{v}, \underline{v}) = (\underline{\sigma}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{v}), \underline{\underline{\nabla}}_{\mathrm{s}} \underline{v}), \tag{3.25}$$

for which we obtain the properties

$$C^2_{\mathrm{mon}} C_K^{-2} \|\underline{\nabla} \underline{v}\|^2 \leqslant \|\underline{v}\|^2_{\mathrm{en}} \leqslant C_{\mathrm{gro}} \|\underline{\underline{\nabla}}_{\mathrm{s}} \underline{v}\|^2, \tag{3.26}$$

by applying (3.2b) and the Korn inequality for the left inequality, and the Cauchy–Schwarz inequality and (3.2a) for the right one. In our case it holds $C_K = \sqrt{2}$, owing to (3.1b).

Theorem 3.10 (Basic a posteriori error estimate)**.** *Let $\underline{u}$ be the analytical solution of* (3.13) *and $\underline{u}_h$ the discrete solution of* (3.14)*. Let $\underline{\underline{\sigma}}_h$ be the stress tensor defined in Construction 3.5. Then,*

$$\|\underline{u}-\underline{u}_h\|_{\mathrm{en}} \leqslant \sqrt{2} C_{\mathrm{gro}} C_{\mathrm{mon}}^{-3} \left(\sum_{T\in\mathcal{T}_h} \left(\frac{h_T}{\pi} \|\underline{f} + \underline{\nabla}\cdot\underline{\underline{\sigma}}_h\|_T + \|\underline{\underline{\sigma}}_h - \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h)\|_T \right)^2 \right)^{1/2}. \tag{3.27}$$

Remark 3.11 (Constants C_{gro} and C_{mon})**.** *For the estimate to be computable, the constants C_{gro} and C_{mon} have to be specified. For the linear elasticity model* (3.4) *we set $C_{\mathrm{gro}} := 2\mu + d\lambda$ and $C_{\mathrm{mon}} := \sqrt{2\mu}$, whereas for the Hencky–Mises model* (3.7) *we set $C_{\mathrm{gro}} := 2\tilde{\mu}(0) + d\tilde{\lambda}(0)$ and $C_{\mathrm{mon}} := \sqrt{2\tilde{\mu}(0)}$. For the damage model* (3.10) *we take $C_{\mathrm{gro}} := C^*$ and $C_{\mathrm{mon}} := \sqrt{C_*}$, where C_* and C^* are the constants appearing in* (3.9)*. Following [94], we obtain a sharper bound in the case of linear elasticity, with $\mu^{-1/2}$ instead of $\sqrt{2}C_{\mathrm{gro}}C_{\mathrm{mon}}^{-3}$ in* (3.27)*.*

Proof of Theorem 3.10. We start by bounding the energy norm of the error by the dual norm of the residual of the weak formulation (3.13). Using (3.26), (3.2b), the linearity of a in its second argument, and (3.13) we obtain

$$\begin{aligned}
\|\underline{u}-\underline{u}_h\|_{\mathrm{en}}^2 &\leqslant C_{\mathrm{gro}} \|\underline{\underline{\nabla}}_{\mathrm{s}}(\underline{u}-\underline{u}_h)\|^2 \leqslant C_{\mathrm{gro}} C_{\mathrm{mon}}^{-2} |a(\underline{u},\underline{u}-\underline{u}_h) - a(\underline{u}_h,\underline{u}-\underline{u}_h)| \\
&= C_{\mathrm{gro}} C_{\mathrm{mon}}^{-2} \|\underline{\underline{\nabla}}(\underline{u}-\underline{u}_h)\| \left| a\left(\underline{u}, \frac{\underline{u}-\underline{u}_h}{\|\underline{\underline{\nabla}}(\underline{u}-\underline{u}_h)\|}\right) - a\left(\underline{u}_h, \frac{\underline{u}-\underline{u}_h}{\|\underline{\underline{\nabla}}(\underline{u}-\underline{u}_h)\|}\right)\right| \\
&\leqslant C_{\mathrm{gro}} C_{\mathrm{mon}}^{-3} C_K \|\underline{u}-\underline{u}_h\|_{\mathrm{en}} \sup_{\underline{v}\in\underline{H}_0^1(\Omega), \|\underline{\underline{\nabla}}\underline{v}\|=1} \{a(\underline{u},\underline{v}) - a(\underline{u}_h,\underline{v})\} \\
&= C_{\mathrm{gro}} C_{\mathrm{mon}}^{-3} C_K \|\underline{u}-\underline{u}_h\|_{\mathrm{en}} \sup_{\underline{v}\in\underline{H}_0^1(\Omega), \|\underline{\underline{\nabla}}\underline{v}\|=1} \{(\underline{f},\underline{v}) - (\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h), \underline{\underline{\nabla}}_{\mathrm{s}}\underline{v})\}.
\end{aligned}$$

and thus

$$\|\underline{u}-\underline{u}_h\|_{\mathrm{en}} \leqslant C_{\mathrm{gro}} C_{\mathrm{mon}}^{-3} C_K \sup_{\underline{v}\in\underline{H}_0^1(\Omega), \|\underline{\underline{\nabla}}\underline{v}\|=1} \{(\underline{f},\underline{v}) - (\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h), \underline{\underline{\nabla}}_{\mathrm{s}}\underline{v})\}. \tag{3.28}$$

Note that, due to the symmetry of $\underline{\underline{\sigma}}$ we can replace $\underline{\underline{\nabla}}_{\mathrm{s}}\underline{v}$ by $\underline{\underline{\nabla}}\underline{v}$ in the second term inside the supremum. Now fix $\underline{v} \in \underline{H}_0^1(\Omega)$, such that $\|\underline{\underline{\nabla}}\underline{v}\| = 1$. Since $\underline{\underline{\sigma}}_h \in \underline{\underline{H}}(\mathrm{div},\Omega)$, we can insert $(\underline{\nabla}\cdot\underline{\underline{\sigma}}_h,\underline{v}) + (\underline{\underline{\sigma}}_h, \underline{\underline{\nabla}}\underline{v}) = 0$ into the term inside the supremum and obtain

$$(\underline{f},\underline{v}) - (\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h), \underline{\underline{\nabla}}\underline{v}) = (\underline{f} + \underline{\nabla}\cdot\underline{\underline{\sigma}}_h, \underline{v}) + (\underline{\underline{\sigma}}_h - \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h), \underline{\underline{\nabla}}\underline{v}). \tag{3.29}$$

For the first term of the right hand side of (3.29) we obtain, using (3.21) on each $T \in \mathcal{T}_h$ to insert $\underline{\Pi}_T^0\underline{v}$, which denotes the $\underline{L}^2$-projection of $\underline{v}$ onto $\underline{\mathbb{P}}^0(T)$, the Cauchy–Schwarz inequality and the Poincaré inequality on simplices,

$$\left|(\underline{f} + \underline{\nabla}\cdot\underline{\underline{\sigma}}_h, \underline{v})\right| \leqslant \left| \sum_{T\in\mathcal{T}_h} (\underline{f} + \underline{\nabla}\cdot\underline{\underline{\sigma}}_h, \underline{v} - \underline{\Pi}_T^0\underline{v})_T \right| \leqslant \sum_{T\in\mathcal{T}_h} \frac{h_T}{\pi} \|\underline{f} + \underline{\nabla}\cdot\underline{\underline{\sigma}}_h\|_T \|\underline{\underline{\nabla}}\underline{v}\|_T, \tag{3.30}$$

whereas the Cauchy–Schwarz inequality applied to the second term directly yields

$$\left|(\underline{\underline{\sigma}}_h - \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_s \underline{u}_h), \underline{\underline{\nabla}} \underline{v})\right| \leqslant \sum_{T \in \mathcal{T}_h} \|\underline{\underline{\sigma}}_h - \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_s \underline{u}_h)\|_T \|\underline{\underline{\nabla}} \underline{v}\|_T.$$

Inserting these results in (3.28) and again applying the Cauchy–Schwarz inequality yields the result. □

3.4.2 Distinguishing the different error components

The goal of this section is to elaborate the error estimate (3.27) so as to distinguish different error components using the equilibrated stress tensors of Constructions 3.7 and 3.8. This distinction is essential for the development of Algorithm 3.14, where the mesh and the stopping criteria for the iterative solver are chosen adaptively.

Notice that in Theorem 3.10 we don't necessarily need $\underline{\underline{\sigma}}_h$ to be the stress tensor obtained in Construction 3.5. We only need it to satisfy two properties: First, equation (3.29) requires $\underline{\underline{\sigma}}_h$ to lie in $\underline{\underline{H}}(\mathrm{div}, \Omega)$. Second, in order to be able to apply the Poincaré inequality in (3.30), $\underline{\underline{\sigma}}_h$ has to satisfy the local equilibrium relation

$$(\underline{f} - \underline{\nabla} \cdot \underline{\underline{\sigma}}_h, \underline{v})_T = 0 \quad \forall \underline{v} \in \underline{\mathbb{P}}^0(T) \ \forall T \in \mathcal{T}_h. \tag{3.31}$$

Thus, the theorem also holds for $\underline{\underline{\sigma}}_h := \underline{\underline{\sigma}}^k_{h,\mathrm{disc}} + \underline{\underline{\sigma}}^k_{h,\mathrm{lin}}$, where $\underline{\underline{\sigma}}^k_{h,\mathrm{disc}}$ and $\underline{\underline{\sigma}}^k_{h,\mathrm{lin}}$ are defined in Constructions 3.7 and 3.8 and we obtain the following result.

Theorem 3.12 (A posteriori error estimate distinguishing different error sources)**.** *Let $\underline{u}$ be the analytical solution of* (3.13)*, $\underline{u}^k_h$ the discrete solution of* (3.15)*, and $\underline{\underline{\sigma}}_h := \underline{\underline{\sigma}}^k_{h,\mathrm{disc}} + \underline{\underline{\sigma}}^k_{h,\mathrm{lin}}$. Then,*

$$\|\underline{u} - \underline{u}^k_h\|_{\mathrm{en}} \leqslant \sqrt{2} C_{\mathrm{gro}} C^{-3}_{\mathrm{mon}} \big(\eta^k_{\mathrm{disc}} + \eta^k_{\mathrm{lin}} + \eta^k_{\mathrm{quad}} + \eta^k_{\mathrm{osc}}\big), \tag{3.32}$$

where the local discretization, linearization, quadrature and oscillation error estimators on each $T \in \mathcal{T}_h$ are defined as

$$\eta^k_{\mathrm{disc},T} := \|\underline{\underline{\sigma}}^k_{h,\mathrm{disc}} - \overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_s \underline{u}^k_h)\|_T, \tag{3.33a}$$

$$\eta^k_{\mathrm{lin},T} := \|\underline{\underline{\sigma}}^k_{h,\mathrm{lin}}\|_T, \tag{3.33b}$$

$$\eta^k_{\mathrm{quad},T} := \|\overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_s \underline{u}^k_h) - \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_s \underline{u}^k_h)\|_T, \tag{3.33c}$$

$$\eta^k_{\mathrm{osc},T} := \frac{h_T}{\pi} \|\underline{f} - \underline{\Pi}^{p-1}_T \underline{f}\|_T, \tag{3.33d}$$

with $\underline{\Pi}^{p-1}_T$ denoting the $\underline{L}^2$-projection onto $\underline{\mathbb{P}}^{p-1}(T)$, and for each error source the global estimator is given by

$$\eta^k_{\cdot} := \Big(4 \sum_{T \in \mathcal{T}_h} (\eta^k_{\cdot,T})^2\Big)^{1/2}. \tag{3.34}$$

Proof. Using $\underline{\underline{\sigma}}_h := \underline{\underline{\sigma}}^k_{h,\mathrm{disc}} + \underline{\underline{\sigma}}^k_{h,\mathrm{lin}}$ in Theorem 3.10, we obtain

$$\|\underline{u}-\underline{u}^k_h\|_{\mathrm{en}} \leqslant \sqrt{2}C_{\mathrm{gro}}C_{\mathrm{mon}}^{-3}\left(\sum_{T\in\mathcal{T}_h}\big(\frac{h_T}{\pi}\|\underline{f}+\underline{\nabla}\cdot(\underline{\underline{\sigma}}^k_{h,\mathrm{disc}}+\underline{\underline{\sigma}}^k_{h,\mathrm{lin}})\|_T+\|\underline{\underline{\sigma}}^k_{h,\mathrm{lin}}+\underline{\underline{\sigma}}^k_{h,\mathrm{disc}}-\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}^k_h)\|_T\big)^2\right)^{1/2}.$$

Applying the second property of Lemma 3.9 in the first term yields the oscillation error estimator. In the second term we add and substract $\overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}^k_h)$ and apply the triangle inequality to obtain

$$\|\underline{u}-\underline{u}^k_h\|_{\mathrm{en}} \leqslant \sqrt{2}C_{\mathrm{gro}}C_{\mathrm{mon}}^{-3}\left(\sum_{T\in\mathcal{T}_h}\big(\eta^k_{\mathrm{disc},T}+\eta^k_{\mathrm{lin},T}+\eta^k_{\mathrm{quad},T}+\eta^k_{\mathrm{osc},T}\big)^2\right)^{1/2}.$$

Owing to (3.34), the previous bound yields the conclusion. □

Remark 3.13 (Quadrature error)**.** *In practice, the projection* $\overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}^k_h)$ *of* $\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}^k_h)$ *onto* $\underline{\underline{\mathbb{P}}}^{p-1}(\mathcal{T}_h)$ *for a general nonlinear stress-strain relation cannot be computed exactly. The quadrature error estimator* $\eta^k_{\mathrm{quad},T}$ *measures the quality of this projection.*

3.4.3 Adaptive algorithm

Based on the error estimate of Theorem 3.12, we propose an adaptive algorithm where the mesh size is locally adapted, and a dynamic stopping criterion is used for the linearization iterations. The idea is to compare the estimators for the different error sources with each other in order to concentrate the computational effort on reducing the dominant one. For this purpose, let $\gamma_{\mathrm{lin}}, \gamma_{\mathrm{quad}} \in (0,1)$, be user-given weights and $\Gamma > 0$ a chosen threshold that the error should not exceed.

Algorithm 3.14 (Adaptive algorithm)**.**

Mesh adaptation loop

1. Choose an initial function $\underline{u}^0_h \in \underline{H}^1_0(\Omega) \cap \underline{\mathbb{P}}^p(\mathcal{T}_h)$ and set $k := 1$
2. Set the initial quadrature precision $\nu := 2p$ (exactness for polynomials up to degree ν)
3. **Linearization iterations**
   (a) Calculate $\underline{\underline{\sigma}}^{k-1}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}^k_h)$, $\underline{u}^k_h$, $\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}^k_h)$ and $\overline{\underline{\underline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}^k_h)$
   (b) Calculate the stress reconstructions $\underline{\underline{\sigma}}^k_{h,\mathrm{disc}}$ and $\underline{\underline{\sigma}}^k_{h,\mathrm{lin}}$ and the error estimators η^k_{disc}, η^k_{lin}, η^k_{osc} and η^k_{quad}
   (c) Improve the quadrature rule (setting $\nu := \nu + 1$) and go back to step 3(a) until

$$\eta^k_{\mathrm{quad}} \leqslant \gamma_{\mathrm{quad}}(\eta^k_{\mathrm{disc}}+\eta^k_{\mathrm{lin}}+\eta^k_{\mathrm{osc}}) \tag{3.35}$$

(d) **End of the linearization loop** if

$$\eta_{\text{lin}}^k \leqslant \gamma_{\text{lin}}(\eta_{\text{disc}}^k + \eta_{\text{osc}}^k) \tag{3.36}$$

4. Refine or coarsen the mesh $\mathcal{T}_h$ such that the local discretization error estimators $\eta_{\text{disc},T}^k$ are distributed evenly

End of the mesh adaptation loop if $\eta_{\text{disc}}^k + \eta_{\text{osc}}^k \leqslant \Gamma$

Instead of using the global stopping criteria (3.35) and (3.36), which are evaluated over all mesh elements, we can also define the local criteria

$$\eta_{\text{quad},T}^k \leqslant \gamma_{\text{quad}}(\eta_{\text{disc},T}^k + \eta_{\text{lin},T}^k + \eta_{\text{osc},T}^k) \qquad \forall T \in \mathcal{T}_h, \tag{3.37a}$$

$$\eta_{\text{lin},T}^k \leqslant \gamma_{\text{lin}}(\eta_{\text{disc},T}^k + \eta_{\text{osc},T}^k) \qquad \forall T \in \mathcal{T}_h, \tag{3.37b}$$

where it is also possible to define local weights $\gamma_{\text{lin},T}$ and $\gamma_{\text{quad},T}$ for each element. These local stopping criteria are necessary to establish the local efficiency of the error estimators in the following section, whereas the global criteria are only sufficient to prove global efficiency.

3.4.4 Local and global efficiency

Let us start by introducing some additional notation used in this section. For a given element $T \in \mathcal{T}_h$, the set $\mathcal{T}_T$ collects the elements sharing at least a vertex with T. The set $\mathcal{F}_h$ contains all faces (if $d = 2$ we will, for simplicity, refer to the edges as faces) of the mesh and is decomposed into boundary faces $\mathcal{F}_h^{\text{ext}}$ and interfaces $\mathcal{F}_h^{\text{int}}$. For a vertex $a \in \mathcal{V}_h$, we denote by $\mathcal{F}_a$ the faces containing a, and by $\mathcal{F}_{\omega_a}$ the faces of the patch ω_a. For any $T \in \mathcal{T}_h$ the set $\mathcal{F}_T$ contains the faces of T, whereas $\mathcal{F}_{\mathcal{T}_T}$ collects all faces sharing at least a vertex with T and we denote $\mathcal{F}_{\mathcal{T}_T}^{\text{int}} = \mathcal{F}_{\mathcal{T}_T} \cap \mathcal{F}_h^{\text{int}}$. In what follows we let $a \lesssim b$ stand for $a \leqslant Cb$ with a generic constant C, which is independent of the mesh size, the domain Ω and the stress-strain relation, but that can depend on the shape regularity of the mesh family $\{\mathcal{T}_h\}_h$ and on the polynomial degree p.

To prove efficiency, we will use a posteriori error estimators of residual type. Following [104,105] we define for $X \subseteq \Omega$ the functional $\mathcal{R}_X : \underline{H}^1(X) \to \underline{H}^{-1}(X)$ such that, for all $\underline{v} \in \underline{H}^1(X), \underline{w} \in \underline{H}_0^1(X)$,

$$\langle \mathcal{R}_X(\underline{v}), \underline{w} \rangle_X := (\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_s \underline{v}), \underline{\underline{\nabla}}_s \underline{w})_X - (\underline{f}, \underline{w})_X.$$

Define, for each $T \in \mathcal{T}_h$,

$$(\eta_{\sharp,T}^k)^2 := \sum_{T' \in \mathcal{T}_T} h_{T'}^2 \|\underline{\nabla} \cdot \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k) + \underline{\Pi}_T^p \underline{f}\|_{T'}^2 + \sum_{F \in \mathcal{F}_{\mathcal{T}_T}^{\mathrm{int}}} h_F \|[\![\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k) \underline{n}_F]\!]\|_F^2,$$
$$(\eta_{\flat,T}^k)^2 := \sum_{T' \in \mathcal{T}_T} h_{T'}^2 \|\underline{\nabla} \cdot (\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k) - \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k))\|_{T'}^2 + \sum_{F \in \mathcal{F}_{\mathcal{T}_T}^{\mathrm{int}}} h_F \|[\![(\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k) - \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k)) \underline{n}_F]\!]\|_F^2. \tag{3.38}$$

The quantity $\eta_{\flat,T}^k$ obviously measures the quality of the approximation of $\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k)$ by $\underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k)$ and can be estimated explicitly. The following result is shown in [104, Section 3.3]. Denoting by $\eta_{\mathrm{osc},\mathcal{T}_T}^k := \{2 \sum_{T' \in \mathcal{T}_T} (\eta_{\mathrm{osc},T'}^k)^2\}^{1/2}$, it holds

$$\eta_{\sharp,T}^k \lesssim \|\mathcal{R}_{\mathcal{T}_T}(\underline{u}_h^k)\|_{\underline{H}^{-1}(\mathcal{T}_T)} + \eta_{\flat,T}^k + \eta_{\mathrm{osc},\mathcal{T}_T}^k. \tag{3.39}$$

In order to bound the dual norm of the residual, we need an additional assumption on the stress-strain relation which, in particular, implies the growth assumption (3.2a).

Assumption 3.15 (Stress-strain relation II). *There exists a real number $C_{\mathrm{Lip}} \in (0, +\infty)$ such that, for all $\underline{\underline{\tau}}, \underline{\underline{\eta}} \in \mathbb{R}_{\mathrm{sym}}^{d \times d}$,*

$$|\underline{\underline{\sigma}}(\underline{\underline{\tau}}) - \underline{\underline{\sigma}}(\underline{\underline{\eta}})|_{d \times d} \leqslant C_{\mathrm{Lip}} |\underline{\underline{\tau}} - \underline{\underline{\eta}}|_{d \times d}. \quad \textit{(Lipschitz continuity)} \tag{3.40}$$

Notice that the three stress-strain relations presented in Examples 3.2, 3.3, and 3.4 satisfy the previous Lipschitz continuity assumptions. Owing to the definition of the functional $\mathcal{R}_{\mathcal{T}_T}$ and to the fact that $-\underline{\nabla} \cdot \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}) = \underline{f} \in \underline{L}^2(\mathcal{T}_T)$, using the Cauchy–Schwarz inequality and the Lipschitz continuity (3.40) of $\underline{\underline{\sigma}}$, it is inferred that

$$\begin{aligned}
\|\mathcal{R}_{\mathcal{T}_T}(\underline{u}_h^k)\|_{\underline{H}^{-1}(\mathcal{T}_T)} &:= \sup_{\underline{w} \in \underline{H}_0^1(\mathcal{T}_T),\, \|\underline{w}\|_{\underline{H}_0^1(\mathcal{T}_T)} \leqslant 1} (\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k), \underline{\underline{\nabla}}_{\mathrm{s}} \underline{w})_{\mathcal{T}_T} - (\underline{f}, \underline{w})_{\mathcal{T}_T} \\
&= \sup_{\underline{w} \in \underline{H}_0^1(\mathcal{T}_T),\, \|\underline{w}\|_{\underline{H}_0^1(\mathcal{T}_T)} \leqslant 1} (\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k) - \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}), \underline{\underline{\nabla}}_{\mathrm{s}} \underline{w})_{\mathcal{T}_T} \\
&\leqslant \sup_{\underline{w} \in \underline{H}_0^1(\mathcal{T}_T),\, \|\underline{w}\|_{\underline{H}_0^1(\mathcal{T}_T)} \leqslant 1} \|\underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k) - \underline{\underline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u})\|_{\mathcal{T}_T} \|\underline{\underline{\nabla}}_{\mathrm{s}} \underline{w}\|_{\mathcal{T}_T} \\
&\leqslant C_{\mathrm{Lip}} \|\underline{\underline{\nabla}}_{\mathrm{s}} (\underline{u} - \underline{u}_h^k)\|_{\mathcal{T}_T}.
\end{aligned}$$

Thus, by (3.39), the previous bound, and the strong monotonicity (3.2b) it holds

$$\eta_{\sharp,T}^k \lesssim C_{\mathrm{Lip}} C_{\mathrm{mon}}^{-1} \|\underline{u} - \underline{u}_h^k\|_{\mathrm{en},\mathcal{T}_T} + \eta_{\flat,T}^k + \eta_{\mathrm{osc},\mathcal{T}_T}^k. \tag{3.41}$$

Theorem 3.16 (Local efficiency). *Let $\underline{u} \in \underline{H}_0^1(\Omega)$ be the solution of (3.13), $\underline{u}_h^k \in \underline{H}_0^1(\Omega) \cap \underline{\mathbb{P}}^p(\mathcal{T}_h)$ be arbitrary and $\underline{\underline{\sigma}}_{h,\mathrm{disc}}^k$ and $\underline{\underline{\sigma}}_{h,\mathrm{lin}}^k$ defined by Constructions 3.7 and 3.8. Let the local*

stopping criteria (3.37) *be verified. Then it holds for all* $T \in \mathcal{T}_h$,

$$\eta^k_{\mathrm{disc},T} + \eta^k_{\mathrm{lin},T} + \eta^k_{\mathrm{quad},T} + \eta^k_{\mathrm{osc},T} \lesssim C_{\mathrm{Lip}} C^{-1}_{\mathrm{mon}} \|\underline{u} - \underline{u}^k_h\|_{\mathrm{en},\mathcal{T}_\mathrm{T}} + \eta^k_{\flat,T} + \eta^k_{\mathrm{osc},\mathcal{T}_T}. \tag{3.42}$$

It is well known that there exist nonconforming finite element methods which are equivalent to mixed finite element methods using the Brezzi–Douglas–Marini spaces (see e.g. [6]). Following the ideas of [50,55,62] and references therein, we use these spaces to prove Theorem 3.16. We will denote by $\underline{M}_h(\omega_a)$ the extension to vector valued functions of the nonconforming space introduced in [6] on a patch ω_a. Recall that $\underline{\underline{\Sigma}}_h(\omega_a)$ is the subspace of $\underline{\underline{H}}(\mathrm{div},\Omega)$ containing tensor-valued piecewise polynomials of degree at most p.

For $d = 2$, the space $\underline{M}_T$ on a triangle $T \in \mathcal{T}_h$ is given by

$$\underline{M}_T := \begin{cases} \{\underline{v} \in \underline{\mathbb{P}}^{p+2}(T) \quad | \ \underline{v}_{|F} \in [\underline{\mathbb{P}}^{p+1}(F) \quad \forall F \in \mathcal{F}_T\} & \text{if } p \text{ is even,} \\ \{\underline{v} \in \underline{\mathbb{P}}^{p+2}(T) \quad | \ \underline{v}_{|F} \in \underline{\mathbb{P}}^{p}(F) \oplus \tilde{\underline{\mathbb{P}}}^{p+2}(F) \quad \forall F \in \mathcal{F}_T\} & \text{if } p \text{ is odd,} \end{cases} \tag{3.43}$$

where $\tilde{\underline{\mathbb{P}}}^{p+2}(F)$ is the $\underline{L}^2(F)$–orthogonal complement of $\underline{\mathbb{P}}^{p+2}(F)$ in $\underline{\mathbb{P}}^{p+1}(F)$. The degrees of freedom are given by the moments up to degree $(p-1)$ inside each $T \in \mathcal{T}_h$ and up to degree p on each edge $F \in \mathcal{F}_h$. On a patch ω_a this means that $\underline{M}_h(\omega_a)$ contains vector-valued functions lying piecewise in $\underline{M}_T$ such that

$$([\![\underline{m}_h]\!], \underline{v}_h)_F = 0 \quad \forall F \in \mathcal{F}_a \backslash \mathcal{F}^{\mathrm{ext}}_h \quad \forall \underline{v}_h \in \underline{\mathbb{P}}^p(F). \tag{3.44}$$

We will denote by $\underline{M}^a_h$ the subspace of $\underline{M}_h(\omega_a)$ with functions $\underline{m}_h$ verifying

$$(\underline{m}_h, \underline{z})_{\omega_a} = 0 \quad \forall \underline{z} \in \underline{RM}^d, \tag{3.45}$$

if $a \in \mathcal{V}^{\mathrm{int}}_h$, and

$$(\underline{m}_h, \underline{v}_h)_F = 0 \quad \forall F \in \mathcal{F}_a \cap \mathcal{F}^{\mathrm{ext}}_h \quad \forall \underline{v}_h \in \underline{\mathbb{P}}^p(F), \tag{3.46}$$

if $a \in \mathcal{V}^{\mathrm{ext}}_h$.

We will use the space $\underline{M}^a_h$ together with Proposition 3.17 to prove Theorem 3.16. For Proposition 3.17 we introduce two equivalent formulations of Construction 3.7 based on the following spaces

$$\tilde{\underline{\underline{\Sigma}}}_T := \{\underline{\underline{\tau}} \in \underline{\underline{\Sigma}}_T \mid (\underline{\underline{\tau}}, \underline{\underline{\mu}})_T = 0 \ \forall \underline{\underline{\mu}} \in \underline{\underline{\Lambda}}_T\}, \tag{3.47}$$

$$\tilde{\underline{\underline{\Sigma}}}_h(\omega_a) := \{\underline{\underline{\tau}}_h \in \underline{\underline{L}}^2(\Omega) \mid \underline{\underline{\tau}}_h \in \tilde{\underline{\underline{\Sigma}}}_T \ \forall T \in \mathcal{T}_a\}, \tag{3.48}$$

$$\tilde{\underline{\underline{\Sigma}}}^a_h := \underline{\underline{\Sigma}}^a_h \cap \tilde{\underline{\underline{\Sigma}}}_h(\omega_a) = \{\underline{\underline{\tau}}_h \in \underline{\underline{\Sigma}}^a_h \mid (\underline{\underline{\tau}}_h, \underline{\underline{\mu}}_h)_{\omega_a} = 0 \ \forall \underline{\underline{\mu}}_h \in \underline{\underline{\Lambda}}^a_h\}, \tag{3.49}$$

$$\underline{L}^a_h := \{\underline{l}_h \in \underline{\mathbb{P}}^p(\mathcal{F}_{\omega_a}) \mid \ \underline{l}_h = \underline{0} \text{ on } \partial\omega_a \text{ if } a \in \mathcal{V}^{\mathrm{int}}_h, \ \underline{l}_h = \underline{0} \text{ on } \partial\omega_a \backslash \partial\Omega \text{ if } a \in \mathcal{V}^{\mathrm{ext}}_h\}, \tag{3.50}$$

where $\mathcal{F}_{\omega_a}$ collects the faces of the patch. The first equivalent formulation of Construction 3.7

consists in finding $\underline{\underline{\sigma}}_h^a \in \underline{\underline{\tilde{\Sigma}}}_h^a$ and $\underline{r}_h^a \in \underline{V}_h^a$ such that for all $(\underline{\underline{\tau}}_h, \underline{v}_h) \in \underline{\underline{\tilde{\Sigma}}}_h^a \times \underline{V}_h^a$

$$(\underline{\underline{\sigma}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} + (\underline{r}_h^a, \underline{\nabla} \cdot \underline{\underline{\tau}}_h)_{\omega_a} = (\psi_a \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_s \underline{u}_h^k), \underline{\underline{\tau}}_h)_{\omega_a}, \tag{3.51a}$$

$$(\underline{\nabla} \cdot \underline{\underline{\sigma}}_h^a, \underline{v}_h)_{\omega_a} = (-\psi_a \underline{f} + \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_s \underline{u}_h^k) \underline{\nabla} \psi_a - \underline{y}_{\text{disc}}^k, \underline{v}_h)_{\omega_a}. \tag{3.51b}$$

The second formulation is the first step when hybridizing the mixed problem (3.51). Following [6] it consists in using the broken space $\underline{\underline{\tilde{\Sigma}}}_h(\omega_a)$ instead of $\underline{\underline{\tilde{\Sigma}}}_h^a$ and imposing the continuity of the normal stress components by Lagrange multipliers. Its solution is $(\underline{\underline{\sigma}}_h^a, \underline{r}_h^a, \underline{l}_h^a) \in \underline{\underline{\tilde{\Sigma}}}_h(\omega_a) \times \underline{V}_h^a \times \underline{L}_h^a$ such that for all $(\underline{\underline{\tau}}_h, \underline{v}_h, \underline{l}_h) \in \underline{\underline{\tilde{\Sigma}}}_h(\omega_a) \times \underline{V}_h^a \times \underline{L}_h^a$

$$(\underline{\underline{\sigma}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} + \sum_{T \in \mathcal{T}_a} (\underline{r}_h^a, \underline{\nabla} \cdot \underline{\underline{\tau}}_h)_T - \sum_{F \in \mathcal{F}_{\omega_a}} (\underline{l}_h^a, [\![\underline{\underline{\tau}}_h \underline{n}_F]\!]_F)_F = (\psi_a \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_s \underline{u}_h^k), \underline{\underline{\tau}}_h)_{\omega_a}, \tag{3.52a}$$

$$\sum_{T \in \mathcal{T}_a} (\underline{\nabla} \cdot \underline{\underline{\sigma}}_h^a, \underline{v}_h)_T = (-\psi_a \underline{f} + \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_s \underline{u}_h^k) \underline{\nabla} \psi_a - \underline{y}_{\text{disc}}^k, \underline{v}_h)_{\omega_a}, \tag{3.52b}$$

$$- \sum_{F \in \mathcal{F}_{\omega_a}} ([\![\underline{\underline{\sigma}}_h^a \underline{n}_F]\!]_F, \underline{l}_h)_F = 0, \tag{3.52c}$$

where we denote by $\underline{n}_{TF}$ the outward normal vector of T on F and by $\underline{n}_F$ the normal vector of F with an arbitrary, but fixed direction. In particular, (3.52a) can be reformulated as

$$(\underline{\underline{\sigma}}_h^a - \psi_a \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_s \underline{u}_h^k), \underline{\underline{\tau}}_T)_T + (\underline{r}_h^a, \underline{\nabla} \cdot \underline{\underline{\tau}}_T)_T = \sum_{F \in \mathcal{F}_T} (\underline{l}_h^a, \underline{\underline{\tau}}_T \underline{n}_{TF})_F \quad \forall \underline{\underline{\tau}}_T \in \underline{\underline{\tilde{\Sigma}}}_T \ \forall T \in \mathcal{T}_a. \tag{3.53}$$

Proposition 3.17. *Let $a \in \mathcal{V}_h$ and let $(\underline{\underline{\sigma}}_h^a, \underline{r}_h^a, \underline{l}_h^a) \in \underline{\underline{\tilde{\Sigma}}}_h(\omega_a) \times \underline{V}_h^a \times \underline{L}_h^a$ be defined by (3.52). Let $\underline{\tilde{r}}_h^a$ be a vector-valued function verifying for all $T \in \mathcal{T}_a$ and for all $F \in \mathcal{F}_T$,*

$$\underline{\tilde{r}}_{h|T}^a \in \underline{M}_T, \tag{3.54a}$$

$$\underline{\Pi}_{\underline{L}_F} \underline{\tilde{r}}_{h|F}^a = \underline{l}_{h|F}^a, \tag{3.54b}$$

$$\underline{\Pi}_{\underline{V}_T} \underline{\tilde{r}}_{h|T}^a = \underline{r}_{h|T}^a, \tag{3.54c}$$

where $\underline{\Pi}_{\underline{L}_F}$ and $\underline{\Pi}_{\underline{V}_T}$ denote, respectively, the L^2-projections on $\underline{L}_F = \underline{\mathbb{P}}^p(F)$ and $\underline{V}_T = \underline{\mathbb{P}}^{p-1}(T)$. Then $\underline{\tilde{r}}_h^a \in \underline{M}_h^a$.

Proof. From $\dim(\underline{V}_T) + 3\dim(\underline{L}_F) = p^2 + p + 3(2p+2) = p^2 + 7p + 6 = \dim(\underline{M}_T)$ we infer that problem (3.54) is well-posed. Plugging (3.54b) and (3.54c) into (3.53) yields

$$\underline{\underline{\Pi}}_{\underline{\underline{\tilde{\Sigma}}}_T} (\underline{\underline{\nabla}} \underline{\tilde{r}}_h^a)_{|T} = (\underline{\underline{\sigma}}_h^a - \psi_a \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_s \underline{u}_h^k))_{|T}. \tag{3.55}$$

Since the formulations (3.51) and (3.52) are equivalent, we can insert (3.55) and (3.54c) into (3.51a) and obtain

$$(\underline{\underline{\nabla}} \underline{\tilde{r}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} + (\underline{\tilde{r}}_h^a, \underline{\nabla} \cdot \underline{\underline{\tau}}_h)_{\omega_a} = 0 \quad \forall \underline{\underline{\tau}}_h \in \underline{\underline{\tilde{\Sigma}}}_h^a.$$

Choosing a basis function of $\underline{\underline{\tilde{\Sigma}}}_h^a$ having zero normal trace across all edges except one edge F

and applying the Green theorem, we see that $\tilde{\underline{r}}_h^a$ satisfies (3.44) for faces $F \in \mathcal{F}_a \backslash \mathcal{F}_h^{\mathrm{ext}}$, since the normal components across F of a basis of $\underline{\underline{\tilde{\Sigma}}}_T$ span $\underline{\mathbb{P}}^p(F)$. If $a \in \mathcal{V}_h^{\mathrm{ext}}$ we can proceed in the same way for $F \in \mathcal{F}_a \cap \mathcal{F}_h^{\mathrm{ext}}$ to obtain (3.46). Finally, for $a \in \mathcal{V}_h^{\mathrm{int}}$ it holds $(\underline{r}_h^a, \underline{z})_{\omega_a} = 0$ for any $\underline{z} \in \underline{RM}^d$ by the definition (3.18b) of $\underline{V}_h^a$, and by (3.54c) it follows that $\tilde{\underline{r}}_h^a$ satisfies (3.45). We conclude that $\tilde{\underline{r}}_h^a$ lies in $\underline{M}_h^a$. □

Proof of Theorem 3.16. We start by proving the local approximation property of the discrete stress reconstruction for any $T \in \mathcal{T}_h$

$$\eta_{\mathrm{disc},T}^k = \|\underline{\underline{\sigma}}_{h,\mathrm{disc}}^k - \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k)\|_T \lesssim \eta_{\sharp,T}^k + \eta_{\mathrm{osc},\mathcal{T}_T}^k. \tag{3.56}$$

We define $\tilde{\underline{r}}_h^a$ by (3.54). Then using the fact that $\tilde{\underline{r}}_h^a \in \underline{M}_h^a$ by Proposition 3.17 and [109, Lemma 5.4], stating that the dual norm on $\underline{M}_h$ is an upper bound for the $\underline{H}^1$-seminorm, we obtain

$$\|\underline{\underline{\sigma}}_h^a - \psi_a \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k)\|_{\omega_a} \leqslant \|\underline{\underline{\nabla}} \tilde{\underline{r}}_h^a\|_{\omega_a} \lesssim \sup_{\underline{m}_h \in \underline{M}_h^a, \|\underline{\underline{\nabla}} \underline{m}_h\| = 1} (\underline{\underline{\sigma}}_h^a - \psi_a \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k), \underline{\underline{\nabla}} \underline{m}_h)_{\omega_a}. \tag{3.57}$$

Now fix $\underline{m}_h \in \underline{M}_h^a$ such that $\|\underline{\underline{\nabla}} \underline{m}_h\|_{\omega_a} = 1$. Then, by (3.44), it follows

$$\begin{aligned}
&(\underline{\underline{\sigma}}_h^a - \psi_a \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k), \underline{\underline{\nabla}} \underline{m}_h)_{\omega_a} \\
&\qquad = \sum_{T \in \mathcal{T}_a} (\underline{\underline{\sigma}}_h^a - \psi_a \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k), \underline{\underline{\nabla}} \underline{m}_h)_T \\
&\qquad = \underbrace{- \sum_{T \in \mathcal{T}_a} (\underline{\nabla} \cdot \underline{\underline{\sigma}}_h^a - \underline{\nabla} \cdot (\psi_a \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k), \underline{m}_h)_T}_{=: \mathfrak{T}_1} + \underbrace{\sum_{F \in \mathcal{F}_a} ([\![\psi_a \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k) \underline{n}_F]\!], \underline{m}_h)_F}_{=: \mathfrak{T}_2}.
\end{aligned}$$

Using (3.51b) (which, as in the proof of Lemma 3.9, is valid for all $\underline{v}_h \in \underline{V}_h(\omega_a)$) and the fact that for all $T \in \mathcal{T}_a$ and $\underline{\underline{\tau}} \in \underline{\underline{\Sigma}}_T$ it holds $(\underline{\nabla} \cdot \underline{\underline{\tau}}, \underline{m}_h)_T = (\underline{\nabla} \cdot \underline{\underline{\tau}}, \underline{\Pi}_{\underline{V}_T} \underline{m}_h)_T$, due to the property $\underline{\nabla} \cdot \underline{\underline{\Sigma}}_T = \underline{V}_T$, we can write for the first term

$$\begin{aligned}
\mathfrak{T}_1 &= - \sum_{T \in \mathcal{T}_a} (-\psi_a \underline{f} + \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k) \underline{\nabla} \psi_a - \underline{\nabla} \cdot (\psi_a \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k)), \underline{\Pi}_{\underline{V}_T} \underline{m}_h)_T \\
&= - \sum_{T \in \mathcal{T}_a} (\psi_a (\underline{f} + \underline{\nabla} \cdot \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k)), \underline{\Pi}_{\underline{V}_T} \underline{m}_h)_T \\
&= - \sum_{T \in \mathcal{T}_a} (\underline{\Pi}_T^p \underline{f} + \underline{\nabla} \cdot \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k), \psi_a \underline{\Pi}_{\underline{V}_T} \underline{m}_h)_T \\
&\leqslant \left(\sum_{T \in \mathcal{T}_a} h_T^2 \|\psi_a (\underline{\Pi}_T^p \underline{f} + \underline{\nabla} \cdot (\underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k)))\|_T^2 \right)^{1/2} \left(\sum_{T \in \mathcal{T}_a} h_T^{-2} \|\underline{m}_h\|_T^2 \right)^{1/2} \\
&\lesssim \left(\sum_{T \in \mathcal{T}_a} h_T^2 \|\underline{\Pi}_T^p \underline{f} + \underline{\nabla} \cdot \underline{\underline{\overline{\sigma}}}(\underline{\underline{\nabla}}_{\mathrm{s}} \underline{u}_h^k)\|_T^2 \|\psi_a\|_{L^\infty(T)}^2 \right)^{1/2} \|\underline{\underline{\nabla}} \underline{m}_h\|_{\omega_a},
\end{aligned}$$

where we used the Cauchy-Schwarz, the discrete Poincaré inequality of [108, Theorem 8.1]

together with (3.45) if $a \in \mathcal{V}_h^{\mathrm{int}}$ and the discrete Friedrichs inequality of [108, Theorem 5.4] together with (3.46) if $a \in \mathcal{V}_h^{\mathrm{ext}}$, and $\|\psi_a\|_{L^\infty(T)} = 1$. For the second term we proceed in a similar way, using the discrete trace inequality $\|\underline{m}_h\|_F \lesssim h_F^{-1/2}\|\underline{m}_h\|_T$, and obtain

$$\begin{aligned}
\mathfrak{T}_2 &= \sum_{F\in\mathcal{F}_a^{\mathrm{int}}} (\psi_a [\![\underline{\overline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)\underline{n}_F]\!], \underline{m}_h)_F \\
&\leqslant \left(\sum_{F\in\mathcal{F}_a^{\mathrm{int}}} h_F \|\psi_a [\![\underline{\overline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)\underline{n}_F]\!]\|_F^2\right)^{1/2} \left(\sum_{F\in\mathcal{F}_a^{\mathrm{int}}} h_F^{-1}\|\underline{m}_h\|_F^2\right)^{1/2} \\
&\lesssim \left(\sum_{F\in\mathcal{F}_a^{\mathrm{int}}} h_F \|[\![\underline{\overline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)\underline{n}_F]\!]\|_F^2\right)^{1/2} \|\underline{\underline{\nabla}}\underline{m}_h\|_{\omega_a}.
\end{aligned}$$

Inserting these results into (3.57) yields (3.56).

From the local stopping criteria (3.37), the definition of the discretization error estimator (3.33a) and the local approximation property (3.56) it follows that

$$\eta_{\mathrm{disc},T}^k + \eta_{\mathrm{lin},T}^k + \eta_{\mathrm{quad},T}^k \lesssim \eta_{\mathrm{disc},T}^k = \|\underline{\underline{\sigma}}_{h,\mathrm{disc}}^k - \underline{\overline{\sigma}}(\underline{\underline{\nabla}}_{\mathrm{s}}\underline{u}_h^k)\|_T \lesssim \eta_{\sharp,T}^k + \eta_{\mathrm{osc},\mathcal{T}_T}^k.$$

Then (3.41) yields the result. □

Theorem 3.18 (Global efficiency). *Let $\underline{u} \in \underline{H}_0^1(\Omega)$ be the solution of (3.13), $\underline{u}_h^k \in \underline{H}_0^1(\Omega) \cap \underline{\mathbb{P}}^p(\mathcal{T}_h)$ be arbitrary and $\underline{\underline{\sigma}}_{h,\mathrm{disc}}^k$ and $\underline{\underline{\sigma}}_{h,\mathrm{lin}}^k$ defined by Constructions 3.7 and 3.8. Let the stopping criteria (3.35) and (3.36) be verified. Then it holds*

$$\eta_{\mathrm{disc}}^k + \eta_{\mathrm{lin}}^k + \eta_{\mathrm{quad}}^k + \eta_{\mathrm{osc}}^k \lesssim C_{\mathrm{Lip}} C_{\mathrm{mon}}^{-1} \|\underline{u} - \underline{u}_h^k\|_{\mathrm{en}} + \eta_\flat^k + \eta_{\mathrm{osc}}^k. \tag{3.58}$$

Proof. Proceeding as above, using the global stopping criteria (3.35) and (3.36), and owing to (3.56) we obtain

$$\eta_{\mathrm{disc}}^k + \eta_{\mathrm{lin}}^k + \eta_{\mathrm{quad}}^k + \eta_{\mathrm{osc}}^k \lesssim \eta_{\mathrm{disc}}^k + \eta_{\mathrm{osc}}^k \lesssim \left(2\sum_{T\in\mathcal{T}_h} (\eta_{\sharp,\mathcal{T}_T}^k + \eta_{\mathrm{osc},\mathcal{T}_T}^k)^2\right)^{1/2} + \eta_{\mathrm{osc}}^k \lesssim \eta_\sharp^k + \eta_{\mathrm{osc}}^k. \tag{3.59}$$

Then, using again (3.41) we obtain the result. □

3.5 Numerical results

In this section we illustrate numerically our results on two test cases, both performed with the `Code_Aster`[1] software, which uses conforming finite elements of degree $p = 2$. Our intention is, first, to show the relevance of the discretization error estimators used as mesh refinement

[1] `http://web-code-aster.org`

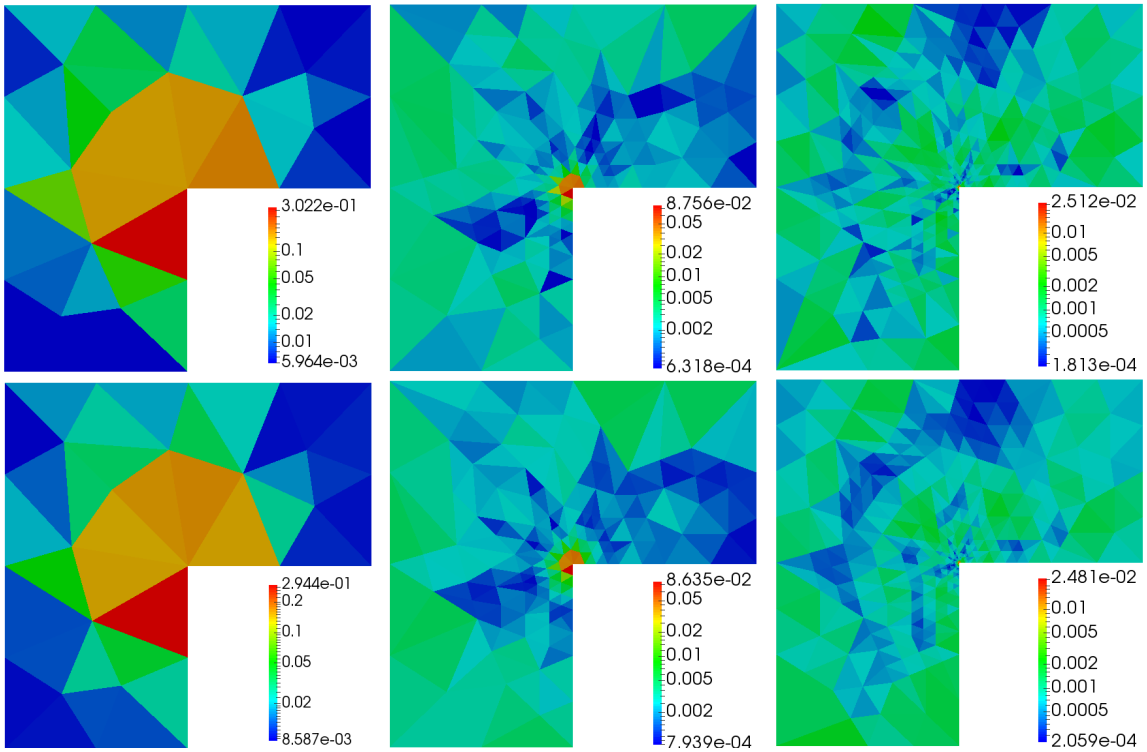


Figure 3.2 – L-shaped domain with linear elasticity model. Distribution of the error estimators (top) and the analytical error (bottom) for the initial mesh (left) and after three (middle) and six (right) adaptive mesh refinements.

indicators, and second, to propose a stopping criterion for the Newton iterations based on the linearization error estimator. All the triangulations are conforming, since in the remeshing progress hanging nodes are removed by bisecting the neighboring element.

3.5.1 L-shaped domain

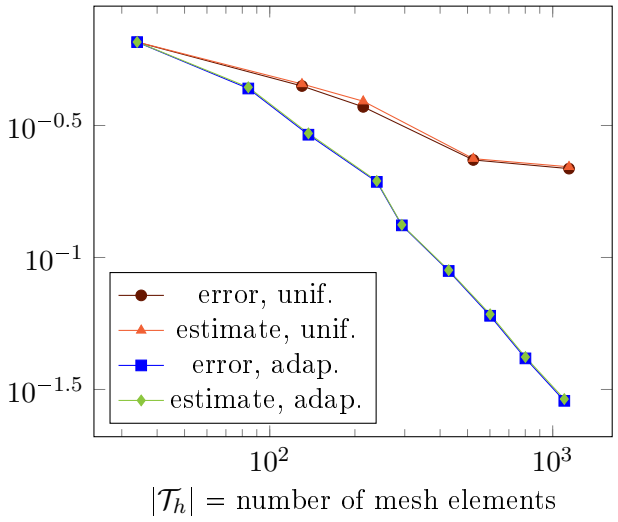
Following [4, 66, 82], we consider the L-shaped domain $\Omega = (-1,1)^2 \backslash ([0,1] \times [-1,0])$, where for the linear elasticity case an analytical solution is given by

$$\underline{u}(r,\theta) = \frac{1}{2\mu} r^{\alpha} \begin{pmatrix} \cos(\alpha\theta) - \cos((\alpha-2)\theta) \\ A\sin(\alpha\theta) + \sin((\alpha-2)\theta) \end{pmatrix},$$

with the parameters

$$\mu = 1.0, \quad \lambda = 5.0, \quad \alpha = 0.6, \quad A = 31/9.$$

This solution is imposed as Dirichlet boundary condition on $\partial\Omega$, together with $\underline{f} = \underline{0}$ in Ω. We perform this test for two different stress-strain relations. First on the linear elasticity model (3.4), where we can compare the error estimate (3.27) to the analytical error $\|\underline{u} - \underline{u}_h\|_{\mathrm{en}}$. The second relation is the nonlinear Hencky–Mises model (3.7), for which we distinguish the discretization and linearization error components and use the adaptive algorithm from Section 3.4.3.



$\lvert\mathcal{T}_h\rvert$	I_{eff}
34	1.00
84	1.01
130	1.02
137	1.01
214	1.05
239	1.01
293	1.00
429	1.01
524	1.01
601	1.01
801	1.01
1099	1.02
1142	1.02

Figure 3.3 – L-shaped domain with linear elasticity model. *Left:* Comparison of the error estimate (3.27) and $\|\underline{u} - \underline{u}_h\|_{\text{en}}$ on two series of meshes, obtained by uniform and adaptive remeshing. *Right:* Effectivity indices of the estimate for each mesh, with the meshes stemming from uniform refinement highlighted in gray.

Linear elasticity model

We compute the analytical error and its estimate on two series of unstructered meshes. Starting with the same initial mesh, we use uniform mesh refinement for the first one and adaptive refinement based on the error estimate for the second series.

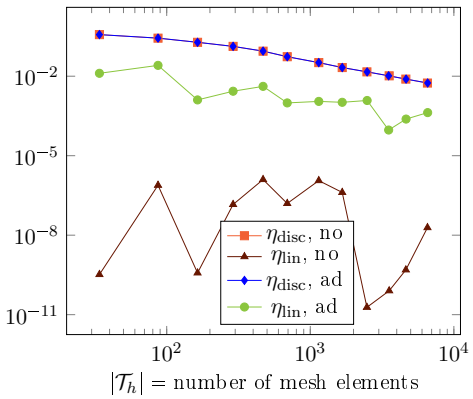
Figure 3.2 compares the distribution of the error and the estimators on the initial and two adaptively refined meshes. The error estimators reflect the distribution of the analytical error, which makes them a good indicator for adaptive remeshing. Figure 3.3 shows the global estimates and errors for each mesh, as well as their effectivity index corresponding to the ratio of the estimate to the error. We obtain effectivity indices close to one, showing that the estimated error value lies close to the actual one, what we can also observe in the graphics on the left. As expected, the adaptively refined mesh series has a higher convergence rate, with corresponding error an order of magnitude lower for 10^3 elements.

Hencky–Mises model

For the Hencky–Mises model we choose the Lamé functions

$$\tilde{\mu}(\rho) := a + b(1+\rho^2)^{-1/2}, \quad \tilde{\lambda}(\rho) := \kappa - \frac{3}{2}\tilde{\mu}(\rho),$$

corresponding to the Carreau law for elastoplastic materials (see, e.g. [60, 74, 97]), and we set $a = 1/20$, $b = 1/2$, and $\kappa = 17/3$ so that the shear modulus reduces progressively to approximately 10% of its initial value. This model allows us to soften the singularity observed



	norm.	adap.
0	4	2
1	4	2
2	5	3
3	5	3
4	5	3
5	6	4
6	6	4
7	6	4
8	7	4
9	7	5
10	8	5
11	9	5

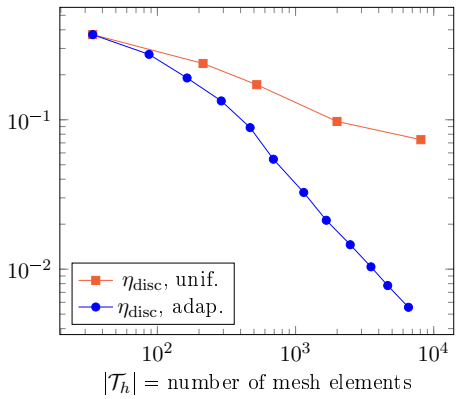


Figure 3.4 – L-shaped domain with Hencky–Mises model. *Left:* Comparison of the global discretization and linearization error estimators on a series of meshes, without and with adaptive stopping criterion for the Newton algorithm. *Middle:* Number of Newton iterations without and with adaptive stopping criterion for each mesh. *Right:* Discretization error estimate for uniform and adaptive remeshing.

in the linear case and to validate our results on more homogeneous error distributions. We apply Algorithm 3.14 with $\gamma_{\text{lin}} = 0.1$ and compare the obtained results to those without the adaptive stopping criterion for the Newton solver. In both cases, we use adaptive remeshing based on the spatial error estimators.

The results are shown in Figure 3.4. In the left graphic we observe that the linearization error estimate in the adaptive case is much higher than in the one without adaptive stopping criterion. We see that this does not affect the discretization error estimator. The table shows the number of performed Newton iterations for both cases. The algorithm using the adaptive stopping criterion is more efficient. In the right graphic we compare the global distretization error estimate on two series of meshes, one refined uniformly and the other one adaptively, based on the local discretization error estimators. Again the convergence rate is higher for the adaptively refined mesh series.

3.5.2 Notched specimen plate

In our second test we use the two nonlinear models of Examples 3.3 and 3.4 on a more application-oriented test. The idea is to set a special sample geometry yielding to a model discrimination test, namely different physical results for different models. We simulate the uniform traction of a notched specimen under plain strain assumption (cf. Figure 3.5). The notch is meant to favor strain localization phenomenon. We consider a domain $\Omega = (0, 10\text{m}) \times (-10\text{m}, 10\text{m}) \backslash \{\underline{x} \in \mathbb{R}^2 \mid \|\underline{x}\text{m} - (0, 11\text{m})^T\| \leqslant 2\text{m}\}$, we take $\underline{f} = \underline{0}$, and we prescribe a displacement on the boundary leading to the following Dirichlet conditions:

$$u_x = 0\text{m} \quad \text{if } x = 0\text{m}, \qquad u_y = -1.1 \cdot 10^{-3}\text{m} \quad \text{if } y = -10\text{m}, \qquad u_y = 1.1 \cdot 10^{-3}\text{m} \quad \text{if } y = 10\text{m}.$$

In many applications, the information about the material properties are obtained in uniaxial

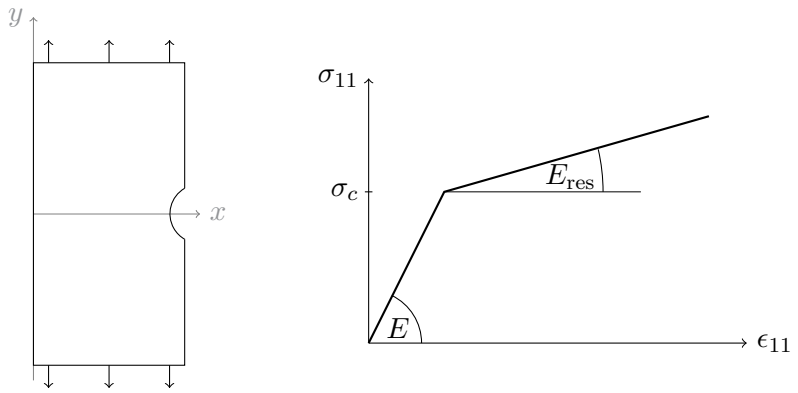


Figure 3.5 – *Left:* Notched specimen plate. *Right:* Uniaxial traction curve

experiments, yielding a relation between σ_{ii} and ϵ_{ii} for a space direction x_i. Since we only consider isotropic materials, we can choose $i = 1$. From this curve one can compute the nonlinear Lamé functions of (3.7) and the damage function in (3.10). Although the uniaxial relation is the same, the resulting stress-strain relations will be different. In our test, we use the $\sigma_{11} - \epsilon_{11}$ – relation indicated in the right of Figure 3.5 with

$$\sigma_c = 3 \cdot 10^4 \text{Pa}, \quad \text{E} = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu} = 3 \cdot 10^8 \text{Pa}, \quad \text{E}_{\text{res}} = 3 \cdot 10^7 \text{Pa},$$

corresponding to the Lamé parameters $\mu = \frac{3}{26} \cdot 10^9 \text{Pa}$ and $\lambda = \frac{9}{52} \cdot 10^9 \text{Pa}$. For both stress-strain relations we apply Algorithm 3.14 with $\gamma_{\text{lin}} = 0.1$. We first compare the results to a computation on a very fine mesh to evaluate the remeshing based on the discretization error estimators. Secondly, we perform adaptive remeshing based on these estimators but without applying the adaptive stopping of the Newton iterations and compare the two series of meshes. As in Section 3.5.1, we verify if the reduced number of iterations impacts the discretization error.

Figure 3.6 shows the result of the first part of the test. In each of the four images the left specimen corresponds to the Hencky–Mises and the right to the isotropic damage model. To illustrate the difference of the two models, the top left picture shows the trace of the strain tensor. This scalar value is a good indicator for both models, representing locally the relative volume increase which could correspond to either a damage or shear band localization zone. In the top right picture we see the distribution of the discretization error estimators in the reference computation (209,375 elements), whereas the distribution of the estimators on the sixth adaptively refined mesh is shown in the bottom right picture (60,618 elements for Hencky–Mises, 55,718 elements for the damage model). The corresponding meshes and the initial mesh for the adaptive algorithm are displayed in the bottom left of the figure. To ensure a good discretization of the notch after repeated mesh refinement, the initial mesh cannot be too corse in this curved area. We observe that the adaptively refined meshes match the distribution of the discretization error estimators on the uniform mesh, and that the estimators are more

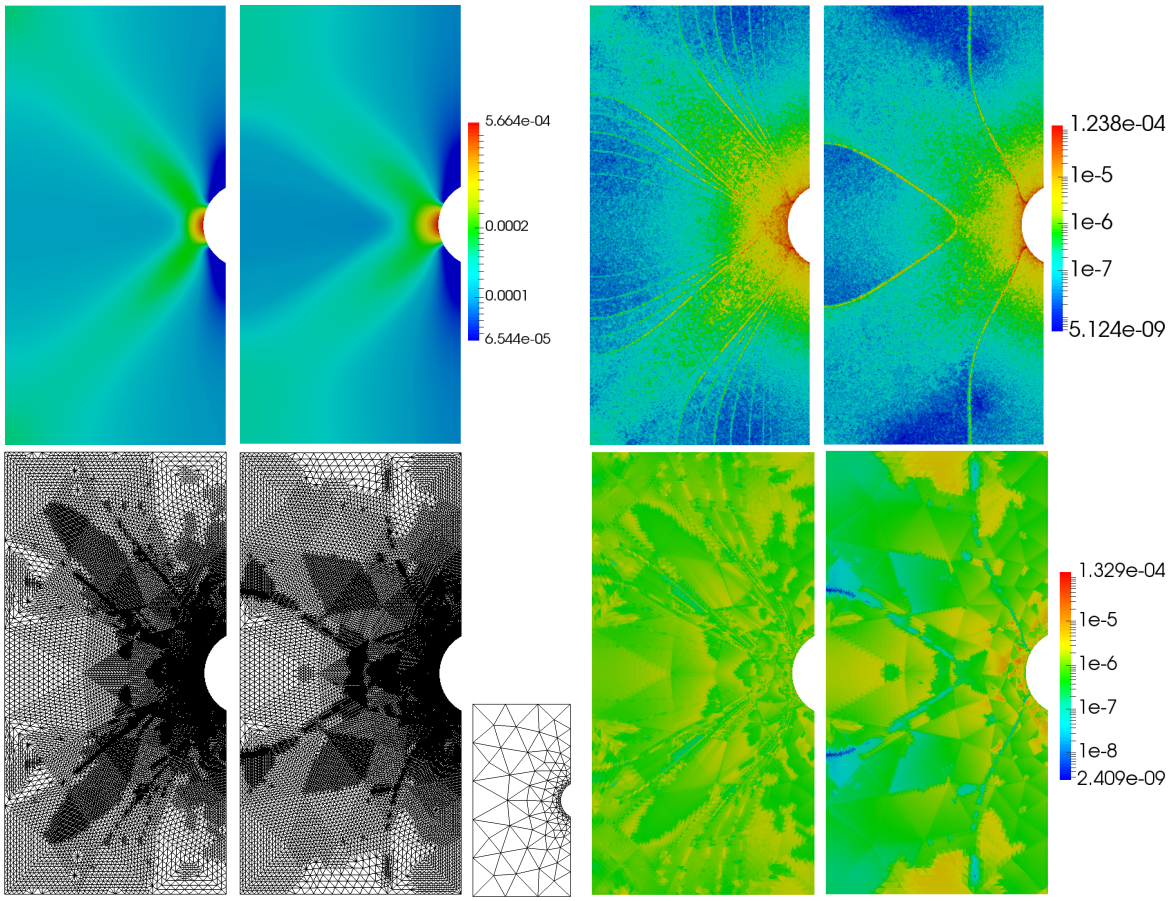


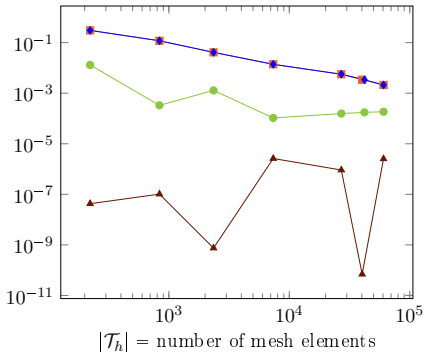
Figure 3.6 – Notched specimen plate, comparison between Hencky–Mises (left in each picture) and damage model (right). *Top left:* $\mathrm{tr}(\underline{\underline{\nabla}}_s \underline{u}_h)$. *Top right:* η_{disc} on a fine mesh (no adaptive refinement). *Bottom left:* meshes after six adaptive refinements. *Bottom middle:* initial mesh. *Bottom right:* η_{disc} on the adaptively refined meshes.

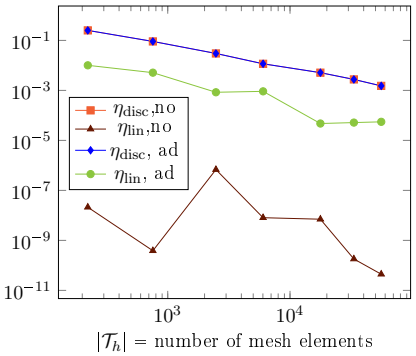
evenly distributed on these meshes.

The results of the second part of the test are illustrated in Figure 3.7. As for the L-shape test, we observe that the reduced number of Newton iterations does not affect the discretization error estimate, nor the overall error estimate which is dominated by the discretization error estimate if the Newton algorithm is stopped.

3.6 Conclusions

In this work we have developed an a posteriori error estimate for a wide class of hyperelastic problems. The estimate is based on stress tensor reconstructions and thus independent of the stress-strain relation, except for two constants. In a finite element software providing different mechanical behavior laws it can be directly applied to any of these laws. The assumptions we make on the stress-strain relation are only used to obtain the equivalence of the energy norm and the dual norm of the residual of the weak formulation. Using the latter as error





	Hencky–Mises		Damage	
	nor.	adap.	nor.	adap.
0	5	3	6	4
1	5	4	6	4
2	6	4	6	5
3	6	5	7	5
4	6	5	7	6
5	7	5	8	6
6	6	5	8	6

Figure 3.7 – Notched specimen plate. Comparison of the global discretization and linearization error estimators without and with adaptive stopping criterion for the Hencky–Mises model (left) and the damage model (middle), and comparison of the number of perfomed Newton iterations (right).

measure, the method can be applied to a wider range of behavior laws. Exploring both numerical tests, we have promising results for general plasticity and damage models. These results come at the price of solving local mixed finite element problems at each iteration of the linearization solver. In practice, the corresponding saddle point problems can be transformed into symmetric positive definite systems using the spaces of Section 3.4.4. Furthermore, these matrices (or their decomposition) can be computed once in a preprocessing stage, and only need to be recomputed if one or more elements in the patch have changed due to remeshing.

Chapter 4

Adaptive algorithms for poro-mechanical problems in 3D

Contents

It is a well established concept that a solid body resists to applied forces by deforming itself. The main goal of engineering and research in solid mechanics is to predict this deformation, which depends on several factors, such as the geometry and the physical properties of the material, and of course the way the forces are applied. The first step for this prediction is the definition of variables describing the intrinsic rigidity of the material and the observation of their reaction during tests simulating different loadings on material specimen of simple geometries. Let $\Omega \subset \mathbb{R}^3$ be the considered body and $(0, t_F)$ with $t_F > 0$ the time interval. By defining in each point of the structure a local mechanical state, it is possible to distinguish the contribution of the geometry to the deformation from the one due to material properties. Under the hypothesis of small deformations, this local state is typically characterized by the stress $\underline{\underline{\sigma}} : \Omega \times (0, t_F) \to \mathbb{R}^{3\times3}_{\text{sym}}$ and the strain $\underline{\underline{\varepsilon}} : \Omega \times (0, t_F) \to \mathbb{R}^{3\times3}_{\text{sym}}$. The evaluation of the test results then yields a model describing the material properties as the relation between these quantities or their time derivatives $\dot{\underline{\underline{\sigma}}}$ and $\dot{\underline{\underline{\varepsilon}}}$. The second step is to apply this model to a specific case of interest, involving in general more complicated geometries and loading terms than in the tests. In general, this can not be done analytically, so the challenge of the second step is the numerical resolution of the resulting problem, taking into account the material behaviour described by the model. Nowadays, some well-established models for groups of materials exist, involving material-specific parameters. Step one then often consists in identifying the parameters corresponding to the considered material through a series of experiments.

In this chapter we start by reviewing a general elasto-plastic stress-strain relation in Section 4.1, which we then, in Section 4.2, consider in the context of poro-mechanical problems. Section 4.3 is dedicated to the numerical solution of the obtained problem. In Section 4.4 we discuss an error measure and its estimation using the equilibrated flux reconstruction techniques elaborated in the previous chapters. In Section 4.5 we present an adaptive algorithm based on the error estimate. We then discuss some mechanical models used in the modeling of soils and rocks in Section 4.6. Finally, in Section 4.7 we assess the error estimate on an analytical test case and apply the adaptive algorithm to the simulation of a tunnel excavation in three space dimensions.

4.1 Elasto-Plasticity

Let us recall that elasticity is based on the assumption that every deformation is reversible, meaning that a body always returns to its original size and shape when loadings are removed. A typical example for elastic behavior is a spring. On the other hand, we speak of plastic behaviour, whenever the deformation in response to an applied loading is non reversible and the shape of the material stays changed after the loading is removed, as for example with a piece of modeling clay. Some materials only behave elastically until the stress reaches a certain threshold and show plastic behaviour once the threshold is exceeded - like a spring that always returns in its originial position, unless it is pulled too hard. These materials are called *elasto-plastic* and the threshold expressing the transition between elastic and plastic behaviour is called *yield* and plays a fundamental role in models describing these materials. An important property for the modelization of elasto-plastic materials is that the strain tensor $\underline{\underline{\varepsilon}}$ is composed of an elastic and a plastic part, i.e.

$$\underline{\underline{\varepsilon}} = \underline{\underline{\varepsilon}}^e + \underline{\underline{\varepsilon}}^p, \tag{4.1}$$

such that the stress $\underline{\underline{\sigma}}$ and the elastic part $\underline{\underline{\varepsilon}}^e$ satisfy Hooke's law. There are different ways of classifying materials. In this work we only consider classical plasticity phenomena, meaning that the time and the speed of the deformation play an ancillary role; the deformation depends primarily on the current loading and possibly the different states the material went through. Counterexamples are phenomena of creep and fatigue or dynamic plasticity, involving additional strain parts in (4.1).

4.1.1 The yield function

To modelize the material behavior, we need a criterion telling us if the reaction at a point $\underline{x}$ in the material to a stress $\underline{\underline{\sigma}}$ will be elastic or plastic deformation. This criterion is expressed as a convex function $F : \mathbb{R}^{3\times 3}_{\text{sym}} \to \mathbb{R}$, called the yield function, defining the *yield surface* $\{\underline{\underline{\sigma}};\ F(\underline{\underline{\sigma}}) = 0\}$ in the space spanned by the components of $\underline{\underline{\sigma}}$. The reaction of a material to the stress $\underline{\underline{\sigma}}$ is defined by the position of $\underline{\underline{\sigma}}$ with regard to the yield surface of the material :

- if $\underline{\underline{\sigma}}$ lies inside the yield surface, i.e. $F(\underline{\underline{\sigma}}) < 0$: the material is **elastic** at $\underline{\underline{\sigma}}$. In particular this means that $\underline{\underline{\varepsilon}}^p = 0$ and thus that stress and strain satisfy Hooke's law.

- if the point $\underline{\underline{\sigma}}$ lies on the yield surface, i.e. $F(\underline{\underline{\sigma}}) = 0$, and the direction of $\dot{\underline{\underline{\sigma}}}$ points inwards the yield surface: the material reaction is still **elastic**. Again, stress and strain satisfy Hooke's law.

- if the point $\underline{\underline{\sigma}}$ lies on the yield surface and the direction of $\dot{\underline{\underline{\sigma}}}$ points outwards the yield surface: the material is **plastic** at $\underline{\underline{\sigma}}$.

4.1.2 Hardening and softening

It is quite intuitive to understand that once a body has been irreversibly deformed, its behavior might change. This is the case if the energy absorbed by the material during the deformation process modifies the internal structure of the material. The resulting effect is called *hardening* or *softening*, depending on whether the deformed material will support more or less stress than the original one before reaching the yield criterion. This means that the value of F for a given $\underline{\underline{\sigma}}$ increases or decreases with ongoing plastic deformation and consequently that the convex elasticity zone $\{\underline{\underline{\sigma}};\ F(\underline{\underline{\sigma}}) < 0\}$ decreases or increases respectively. To account for this change of the yield surface, we introduce the parameter $\alpha : \Omega \times (0, t_F) \to \mathbb{R}$ in the function F. This parameter, called *accumulated plastic strain*, describes the length of the deformation path until the time t, and thus is a way of tracking the strain history in the time interval $(0, t_F)$.

In general, geomaterials, which are the topic of this chapter, show pure softening behavior, meaning that after every plastic deformation the resulting yield surface $\{(\underline{\underline{\sigma}}, \alpha);\ F(\underline{\underline{\sigma}}, \alpha) = 0\}$ lies entirely inside the previous one. Equivalently, we speek of pure hardening behavior when the yield surface lies entirely outside the previous surfaces after any plastic deformation. At this point, it is already worth noticing that only elasto-plasticity problems describing purely hardening materials have a unique solution. The reason can be explained in a simplified way using Figure 4.1, showing two basic stress-strain relations for a one-dimensional problem; left for a hardening material, and right for a softening one. Both materials show an elastic behavior until σ first reaches σ_{crit} and plastic behavior if ε increases from that point on. At every point of the domain, the couple (ε, σ) lies on the curve indicated in Figure 4.1. Let us assume that the material satisfies the equilibrium $\nabla \cdot \sigma = 0$, which - in one dimension - means that σ is constant over the whole domain, and that the domain is under traction with an increasing displacement u_D. Then, since σ is constant, we see that in the case of the left picture, ε will be constant over the domain too. In the right picture on the other hand, once σ reaches the value σ_{crit}, each point can either take the left or the right branch of the curve, as long as the integral over ε is equal to u'.

Figure 4.2 shows, again on a one dimensional example, how softening changes the properties of a material. The left picture corresponds to the reaction of a material point under increasing

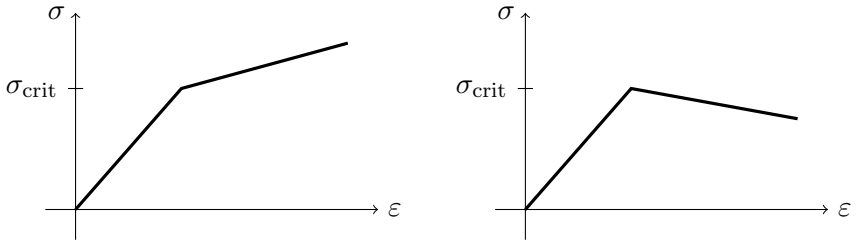


Figure 4.1 – Hardening and softening

traction. The stress reaches the value $\sigma_{\text{crit},0}$, and then decreases with further increasing strain. The middle picture shows what happens when the traction then decreases to zero: only the elastic deformation ε^e is reversible and obeys Hooke's law $\sigma = E\varepsilon^e$, where $E > 0$ is Young's modulus, whereas the plastic deformation remains. When - as illustrated in the right picture - the traction increases again, the material resists to less stress than the undeformed one.

4.1.3 The elasto-plastic behavior law

We recall that the mechanical behavior describing the relation between stress and strain is essential for the resolution of a mechanical problem (where external forces are known and the resulting displacement is sought), as the stress satisfies the equilibrium with the external forces, and the strain is the symmetric gradient of the displacement. To derive this stress-strain relation taking into account the yield criterion, we assume that the yield function $F : \mathbb{R}^{3\times3}_{\text{sym}} \times \mathbb{R} \to \mathbb{R}$ satisfies the following properties:

- F is a convex and piecewise analytical function of $(\underline{\underline{\sigma}}, \alpha)$
- The point $(\underline{\underline{0}}, 0)$ does not lie on the yield surface, i.e. $F(\underline{\underline{0}}, 0) < 0$
- F is differentiable in all points $(\underline{\underline{\sigma}}, \alpha)$ satisfying $F(\underline{\underline{\sigma}}, \alpha) = 0$

In plasticity, the behavior law is usually expressed as the relation between the time derivatives of the stress and the strain. For a given time $t \in (0, t_F)$, we distinguish between the elastic behavior, which takes place if $(\underline{\underline{\sigma}}, \alpha)$ is such that $F(\underline{\underline{\sigma}}, \alpha) < 0$ or if $F(\underline{\underline{\sigma}}, \alpha) = 0$ and $(\partial F/\partial \underline{\underline{\sigma}}) : \dot{\underline{\underline{\sigma}}} \leqslant 0$ (i.e. $(\underline{\underline{\sigma}}, \alpha)$ lies inside or will move towards the inside of the yield surface), and the plastic behavior occuring if $F(\underline{\underline{\sigma}}, \alpha) = 0$ and $(\partial F/\partial \underline{\underline{\sigma}}) : \dot{\underline{\underline{\sigma}}} > 0$. In the elastic case we know that $\underline{\underline{\sigma}}$ and $\underline{\underline{\varepsilon}}$ satisfy Hooke's law, and that the plastic increment is zero:

$$\dot{\underline{\underline{\sigma}}} = \underline{\underline{\underline{\underline{D}}}} : \dot{\underline{\underline{\varepsilon}}}, \tag{4.2a}$$

$$\dot{p} = 0, \tag{4.2b}$$

where $\underline{\underline{\underline{\underline{D}}}}$ is a symmetric, positive definite fourth-order tensor, and $\underline{\underline{\underline{\underline{a}}}} : \underline{\underline{b}}$ produces a second order tensor by $(\underline{\underline{\underline{\underline{a}}}} : \underline{\underline{b}})_{ij} := \sum_{k,l=1}^{3} a_{ijkl} b_{kl}$. Similarly we define $(\underline{\underline{b}} : \underline{\underline{\underline{\underline{a}}}})_{kl} := \sum_{i,j=1}^{3} b_{ij} a_{ijkl}$.

As mentioned above, in the plastic case we split the strain into two parts. On the one hand

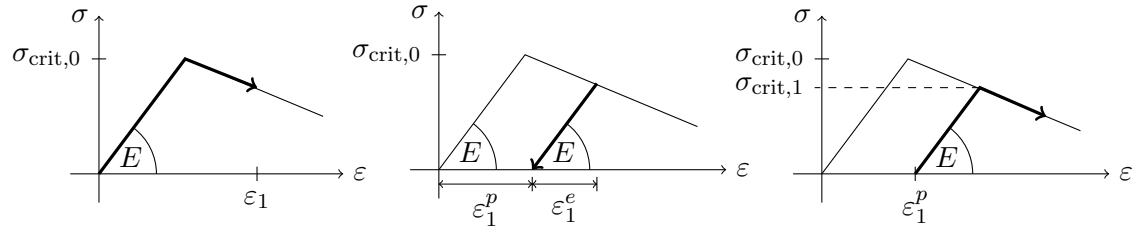


Figure 4.2 – Plastic deformation with softening

the elastic strain $\underline{\underline{\varepsilon}}^e$ still satisfying Hooke's law (4.2a), and on the other hand the plastic strain $\underline{\underline{\varepsilon}}^p = \underline{\underline{\varepsilon}} - \underline{\underline{\varepsilon}}^e$. In order to write the mechanical behavior law, we eliminate this plastic strain, using the normality assumption, also called Prandtl-Reuss flow rule. It states that, for every $t \in (0, t_F)$, there exists $\lambda \geqslant 0$, such that

$$\dot{\underline{\underline{\varepsilon}}}^p = \lambda \frac{\partial F}{\partial \underline{\underline{\sigma}}}(\underline{\underline{\sigma}}, \alpha) \quad \text{and} \quad \dot{\alpha} = -\lambda \frac{\partial F}{\partial \alpha}(\underline{\underline{\sigma}}, \alpha), \tag{4.3}$$

meaning that, on the one hand, the direction of the plastic flow is equal to the outward normal to the yield surface in the stress space, and that, on the other hand, the direction of $(\underline{\underline{\varepsilon}}^p, \alpha)$ in the $\underline{\underline{\sigma}} - \alpha-$ space is tangential to the yield surface. We also know that the point $(\underline{\underline{\sigma}}, p)$ moves along the yield surface, i.e.

$$\frac{\partial F}{\partial \underline{\underline{\sigma}}}(\underline{\underline{\sigma}}, \alpha) : \dot{\underline{\underline{\sigma}}} + \frac{\partial F}{\partial \alpha}(\underline{\underline{\sigma}}, \alpha)\dot{\alpha} = 0. \tag{4.4}$$

In the following we denote $\underline{\underline{\partial}}_\sigma F := \frac{\partial F}{\partial \underline{\underline{\sigma}}}(\underline{\underline{\sigma}}, \alpha)$ and $\partial_\alpha F := \frac{\partial F}{\partial \alpha}(\underline{\underline{\sigma}}, \alpha)$. Since $\dot{\underline{\underline{\varepsilon}}}^e$ satisfies Hooke's law, we can write

$$\underline{\underline{0}} = \underline{\underline{\underline{\underline{D}}}} : (\dot{\underline{\underline{\varepsilon}}} - \dot{\underline{\underline{\varepsilon}}}^p) - \dot{\underline{\underline{\sigma}}}. \tag{4.5}$$

Multiplying this equation by $\underline{\underline{\partial}}_\sigma F$, and using (4.3) and (4.4) we obtain

$$\begin{aligned} \underline{\underline{0}} &= \underline{\underline{\partial}}_\sigma F : \underline{\underline{\underline{\underline{D}}}} : \dot{\underline{\underline{\varepsilon}}} - \underline{\underline{\partial}}_\sigma F : \underline{\underline{\underline{\underline{D}}}} : \dot{\underline{\underline{\varepsilon}}}^p - \underline{\underline{\partial}}_\sigma F : \dot{\underline{\underline{\sigma}}} \\ &= \underline{\underline{\partial}}_\sigma F : \underline{\underline{\underline{\underline{D}}}} : \dot{\underline{\underline{\varepsilon}}} - (\underline{\underline{\partial}}_\sigma F : \underline{\underline{\underline{\underline{D}}}} : \underline{\underline{\partial}}_\sigma F + (\partial_\alpha F)^2)\lambda, \end{aligned}$$

leading to

$$\lambda = \frac{\underline{\underline{\partial}}_\sigma F : \underline{\underline{\underline{\underline{D}}}} : \dot{\underline{\underline{\varepsilon}}}}{\underline{\underline{\partial}}_\sigma F : \underline{\underline{\underline{\underline{D}}}} : \underline{\underline{\partial}}_\sigma F + (\partial_\alpha F)^2}.$$

Plugging this expression for λ into (4.3) and the so obtained $\underline{\underline{\varepsilon}}^p$ into (4.5) we get the stress-strain relation

$$\dot{\underline{\underline{\sigma}}} = \left(\underline{\underline{\underline{\underline{D}}}} - \frac{\underline{\underline{\underline{\underline{D}}}} : \underline{\underline{\partial}}_\sigma F \otimes \underline{\underline{\partial}}_\sigma F : \underline{\underline{\underline{\underline{D}}}}}{\underline{\underline{\partial}}_\sigma F : \underline{\underline{\underline{\underline{D}}}} : \underline{\underline{\partial}}_\sigma F + (\partial_\alpha F)^2} \right) \dot{\underline{\underline{\varepsilon}}},$$

where for two second-order tensors $\underline{\underline{a}}, \underline{\underline{b}}$ the tensor product yielding a fourth-order tensor is defined as $(\underline{\underline{a}} \otimes \underline{\underline{b}})_{ijkl} := a_{ij} b_{kl}$. The variation of the accumulated plastic strain as a function of $\dot{\underline{\underline{\sigma}}}$ is directly obtained from (4.4).

Summing up, we can state the general elasto-plastic behavior law as follows:

- if $(\underline{\underline{\sigma}}, \alpha)$ are such that $F(\underline{\underline{\sigma}}, \alpha) < 0$ or $F(\underline{\underline{\sigma}}, \alpha) = 0$ and $\underline{\underline{\partial}}_\sigma F : \dot{\underline{\underline{\sigma}}} \leqslant 0$, then

$$\dot{\underline{\underline{\sigma}}} = \underline{\underline{\underline{\underline{D}}}} : \dot{\underline{\underline{\varepsilon}}}, \tag{4.6a}$$

$$\dot{\alpha} = 0. \tag{4.6b}$$

- if $(\underline{\underline{\sigma}}, \alpha)$ are such that $F(\underline{\underline{\sigma}}, \alpha) = 0$ and $\underline{\underline{\partial}}_\sigma F : \dot{\underline{\underline{\sigma}}} > 0$, then

$$\dot{\underline{\underline{\sigma}}} = \left(\underline{\underline{\underline{\underline{D}}}} - \frac{\underline{\underline{\underline{\underline{D}}}} : \underline{\underline{\partial}}_\sigma F \otimes \underline{\underline{\partial}}_\sigma F : \underline{\underline{\underline{\underline{D}}}}}{\underline{\underline{\partial}}_\sigma F : \underline{\underline{\underline{\underline{D}}}} : \underline{\underline{\partial}}_\sigma F + (\partial_\alpha F)^2} \right) \dot{\underline{\underline{\varepsilon}}}, \tag{4.6c}$$

$$\dot{\alpha} = -(\partial_\alpha F)^{-1} \underline{\underline{\partial}}_\sigma F : \dot{\underline{\underline{\sigma}}}. \tag{4.6d}$$

4.2 Poro-mechanical coupling

In this section we integrate the stress-strain relation (4.6) into the mechanical constitutive relation of the coupled hydro-mechanical problem introduced in Section 1.2. We express the effective stress $\underline{\underline{\sigma}}'$, which together with the stress $\underline{\underline{\sigma}}_p$ induced by the pressure yields the total stress $\underline{\underline{\sigma}}$. We recall that $\underline{f} : \Omega \times (0, t_F) \to \mathbb{R}^3$ and $g : \Omega \times (0, t_F) \to \mathbb{R}$ denote respectively the volumetric body force and a fluid source, $b > 0$ is the Biot-Willis coefficient, c_0 the specific storage coefficient, and $\kappa : \Omega \to [\kappa_\flat, \kappa_\sharp]$ with $0 < \kappa_\flat < \kappa_\sharp < \infty$ the hydraulic conductivity. The goal is to determine the displacement field $\underline{u} : \Omega \times (0, t_F) \to \mathbb{R}^3$ and the pressure $p : \Omega \times (0, t_F) \to \mathbb{R}$ verifying the equilibrium equations

$$-\underline{\nabla} \cdot \left(\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}), \alpha) + bp\underline{\underline{I}} \right) = \underline{f} \qquad \text{in } \Omega \times (0, t_F), \tag{4.7a}$$

$$\partial_t (c_0 p + b \underline{\nabla} \cdot \underline{u}) + \nabla \cdot \underline{\phi}(p) = g \qquad \text{in } \Omega \times (0, t_F), \tag{4.7b}$$

where the Darcy velocity is defined as $\underline{\phi}(p) = -\kappa \underline{\nabla} p$ and $\underline{\underline{\sigma}}'$ and α are derived from (4.6), with $\underline{\underline{\sigma}}'$ instead of $\underline{\underline{\sigma}}$, taking into account the initial conditions

$$\underline{u}(\cdot, 0) = \underline{u}_0, \quad \alpha(\cdot, 0) = 0 \quad \text{and} \quad p(\cdot, 0) = p_0 \qquad \text{in } \Omega, \tag{4.7c}$$

with functions $\underline{u}_0 : \Omega \to \mathbb{R}^3$ and $p_0 : \Omega \to \mathbb{R}$. For simplicity, we consider homogenous Dirichlet boundary conditions

$$\underline{u} = \underline{0} \quad \text{and} \quad p = 0 \qquad \text{on } \partial\Omega \times (0, t_F). \tag{4.7d}$$

As in Chapter 2, we next write the weak formulation. We define the following sets and spaces

$$\begin{aligned}
\mathcal{E} &= \{(\underline{\underline{\sigma}}, \alpha) \in \mathbb{R}^{3\times 3}_{\text{sym}} \times \mathbb{R};\ F(\underline{\underline{\sigma}}, \alpha) < 0 \text{ or } F(\underline{\underline{\sigma}}, \alpha) = 0 \text{ and } \underline{\underline{\partial}}_\sigma F : \dot{\underline{\underline{\sigma}}} \leqslant 0\}, \\
\mathcal{P} &= \{(\underline{\underline{\sigma}}, \alpha) \in \mathbb{R}^{3\times 3}_{\text{sym}} \times \mathbb{R};\ F(\underline{\underline{\sigma}}, \alpha) = 0 \text{ and } \underline{\underline{\partial}}_\sigma F : \dot{\underline{\underline{\sigma}}} > 0\} \\
\underline{U} &= H^1(0, t_F; \underline{H}^1_0(\Omega)), \qquad P = H^1(0, t_F; H^1_0(\Omega)).
\end{aligned}$$

Then the problem reads: find $(\underline{u}, p, \alpha) \in \underline{U} \times P \times L^2(0, t_F; L^2(\Omega))$ verifying (4.7c), such that

for a. e. $t \in (0, t_F)$,

$$(\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}), \alpha), \underline{\underline{\varepsilon}}(\underline{v})) - b(p, \nabla \cdot \underline{v}) = (\underline{f}, \underline{v}) \qquad \forall \underline{v} \in \underline{H}^1_0(\Omega), \tag{4.9a}$$

$$c_0(\dot{p}, q) + b(\nabla \cdot \dot{\underline{u}}, q) - (\phi(p), \underline{\nabla} q) = (g, q) \qquad \forall q \in H^1_0(\Omega), \tag{4.9b}$$

$$\dot{\underline{\underline{\sigma}}}'(\underline{\underline{\varepsilon}}(\underline{u}), \alpha) = \begin{cases} \underline{\underline{\underline{\underline{D}}}} : \dot{\underline{\underline{\varepsilon}}}(\underline{u}) & \text{if } (\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}), \alpha), \alpha) \in \mathcal{E} \\ (\underline{\underline{\underline{\underline{D}}}} - \underline{\underline{\underline{\underline{D}}}}') : \dot{\underline{\underline{\varepsilon}}}(\underline{u}) & \text{if } (\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}), \alpha), \alpha) \in \mathcal{P} \end{cases} \quad \text{a.e. in } \Omega \times (0, t_F), \tag{4.9c}$$

$$\dot{\alpha} = \begin{cases} 0 & \text{if } (\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}), \alpha), \alpha) \in \mathcal{E} \\ -(\partial_\alpha F)^{-1} \underline{\underline{\partial}}_\sigma F : \dot{\underline{\underline{\sigma}}}'. & \text{if } (\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}), \alpha), \alpha) \in \mathcal{P} \end{cases} \quad \text{a.e. in } \Omega \times (0, t_F), \tag{4.9d}$$

where

$$\underline{\underline{\underline{\underline{D}}}}' = \frac{\underline{\underline{\underline{\underline{D}}}} : \underline{\underline{\partial}}_\sigma F \otimes \underline{\underline{\partial}}_\sigma F : \underline{\underline{\underline{\underline{D}}}}}{\underline{\underline{\partial}}_\sigma F : \underline{\underline{\underline{\underline{D}}}} : \underline{\underline{\partial}}_\sigma F + (\partial_\alpha F)^2}.$$

For hardening materials, existence and uniqueness of a solution for linear poroplasticity (i.e. if there exists a real forth-order tensor $\underline{\underline{\underline{\underline{D}}}}^*$ such that $\dot{\underline{\underline{\sigma}}}' = \underline{\underline{\underline{\underline{D}}}}^* : \dot{\underline{\underline{\varepsilon}}}^p + \underline{\underline{\underline{\underline{D}}}} : \underline{\underline{\varepsilon}}^e$) is proven for example in [37]. The well-posedness of the pure mechanical problem for a class of hardening elasto-plastic materials has been shown in [23]. As discussed in Section 4.1.2, for softening materials the problem does in general not have a unique solution. In this case, it is possible to add regularization parameters in the mechanical behavior law, as for example in [58].

4.3 Numerical solution

In this section we explain how (4.9) is numerically solved in Code_Aster. We assume that the problem is well posed, using for example the regularization techniques of [58], see also [45,46].

We start by introducing some notation before we identify the space and time discretization, which is the same as in Chapter 2 in the two-dimensional case: an H^1-conforming space discretization using the Taylor–Hood finite element spaces and an implicit Euler time stepping.

4.3.1 Notation

For the discretization of the time interval $(0, t_F)$, we consider a sequence of discrete times $(t^n)_{0 \leqslant n \leqslant N}$ with $t^0 = 0$ and $t^N = t_F$, and such that $t^i < t^j$ if $i < j$. For each $1 \leqslant n \leqslant N$ we define the time step $\tau_n := t^n - t^{n-1}$, the time interval $I_n := (t^{n-1}, t^n)$, and the discrete backward differencing operator $\partial_t^n \varphi := \tau_n^{-1}(\varphi^n - \varphi^{n-1})$, where for a space-time function φ we denote $\varphi^n := \varphi(\cdot, t^n)$.

At each time step n, let $\mathcal{T}_h^n$ be a conforming triangulation of Ω (i.e. a set of closed tetrahedra with union $\overline{\Omega}$ and such that, for any distinct $T_1, T_2 \in \mathcal{T}_h^n$, their intersection $T_1 \cap T_2$ is either a common face, a common edge, a vertex or the empty set) verifying the minimum angle condition (i.e. there exists $\alpha_{\min} > 0$ such that the minimum angle α_T of a tetrahedron $T \in \mathcal{T}_h^n$

satisfies $\alpha_T \geqslant \alpha_{\min}$). $\mathcal{V}_h^n = \mathcal{V}_h^{n,\mathrm{int}} \cup \mathcal{V}_h^{n,\mathrm{ext}}$ denotes the set of vertices, divided into interior vertices and exterior vertices lying on the boundary $\partial\Omega$. For any subdomain $\omega \subset \Omega$ the set $\mathcal{V}_\omega^n$ contains the vertices in ω, and for all vertices $a \in \mathcal{V}_h^n$ the set $\mathcal{T}_a$ collects the elements sharing the vertex a. The corresponding open subset of Ω is ω_a and is called a patch. For any element $T \in \mathcal{T}_h^n$, h_T denotes its diameter and $\underline{n}_T$ its unit outward normal vector.

4.3.2 Discrete formulation

We next introduce the used discretization spaces. Let $\mathbb{P}^k(T)$ be the space of trivariate polynomials in $T \in \mathcal{T}_h^n$ of total degree at most $k \geqslant 1$, and let $\mathbb{P}^k(\mathcal{T}_h^n) = \{\varphi \in L^2(\Omega);\ \varphi_{|T} \in \mathbb{P}^k(T)\ \forall T \in \mathcal{T}_h^n\}$ denote the corresponding broken space over $\mathcal{T}_h^n$. The Taylor–Hood finite element spaces of degree $k \geqslant 1$ are defined as

$$\underline{U}_h^n := \underline{\mathbb{P}}^{k+1}(\mathcal{T}_h^n) \cap \underline{H}_0^1(\Omega) \quad \text{and} \quad P_h^n := \mathbb{P}^k(\mathcal{T}_h^n) \cap H_0^1(\Omega). \tag{4.10}$$

We indicate discrete space-time functions by the subscript "$h\tau$", whereas the discrete space function obtained when fixing one time step t^n is denoted as before by $\varphi_h^n := \varphi_{h\tau}(\cdot, t^n)$.

Assumption 4.1. *For simplicity, we assume that the source functions $\underline{f}$ and g are piecewise constant in time and in space, i.e. that the following holds:*

- *$\underline{f}$ and g are constant in each time interval. We denote $\underline{f}^n := \underline{f}_{|I_n}$ and $g^n := g_{|I_n}$.*
- *For each $1 \leqslant n \leqslant N$ it holds $\underline{f}^n \in \underline{\mathbb{P}}^0(\mathcal{T}_h^n)$ and $g^n \in \mathbb{P}^0(\mathcal{T}_h^n)$.*

We denote by $(\cdot,\cdot)_h$ the discrete inner L^2-product calculated using Gauss integration. We omit the subscript h, whenever the integral can be calculated exactly. The discrete formulation of the problem then reads: fixing $\underline{u}^0 = \underline{u}_0$, $p^0 = p_0$ and $\alpha^0 = 0$, find for each $1 \leqslant n \leqslant N$ with given $\underline{u}_h^{n-1}$, p_h^{n-1} and α^{n-1} the functions $(\underline{u}_h^n, p_h^n) \in \underline{U}_h^n \times P_h^n$ and α^n, such that

$$(\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^n), \alpha^n, \underline{\underline{\varepsilon}}(\underline{v}_h))_h + b(p_h^n, \nabla \cdot \underline{v}_h) = (\underline{f}^n, \underline{v}_h) \quad \forall \underline{v}_h \in \underline{U}_h^n, \tag{4.11a}$$

$$(\rho^{-1} m(\partial_t^n \underline{u}_{h\tau}, \partial_t^n p_{h\tau}), q_h) - (\underline{\phi}(p_h^n), \underline{\nabla} q_h) = (g^n, q_h) \quad \forall q_h \in P_h^n, \tag{4.11b}$$

with the effective stress tensor and accumulated plastic strain increment at each Gauss point given by

$$\partial_t^n \underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_{h\tau}), \alpha^n) = \begin{cases} \underline{\underline{\underline{\underline{D}}}} : \partial_t^n \underline{\underline{\varepsilon}}(\underline{u}_{h\tau}) & \text{if } (\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^n), \alpha^n), \alpha^n) \in \mathcal{E}, \\ (\underline{\underline{\underline{\underline{D}}}} - \underline{\underline{\underline{\underline{D}}}}') : \partial_t^n \underline{\underline{\varepsilon}}(\underline{u}_{h\tau}) & \text{if } (\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^n), \alpha^n), \alpha^n) \in \mathcal{P}, \end{cases} \tag{4.11c}$$

$$\partial_t^n \alpha^n = \begin{cases} 0 & \text{if } (\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^n), \alpha^n) \in \mathcal{E}, \\ -(\partial_\alpha F)^{-1} \underline{\underline{\partial}}_\sigma F : \partial_t^n \underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_{h\tau}), \alpha^n) & \text{if } (\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^n), \alpha^n) \in \mathcal{P}, \end{cases} \tag{4.11d}$$

where we denote by $\underline{u}_{h\tau}$ and $p_{h\tau}$ the discrete space-time functions such that, at each time step $0 \leqslant n \leqslant N$, it holds $(\underline{u}_{h\tau}, p_{h\tau})(\cdot, t^n) = (\underline{u}_h^n, p_h^n)$ and $\partial_t \underline{u}_{h\tau}^n := \partial_t \underline{u}_{h\tau|I_n} = \tau_n^{-1}(\underline{u}_h^n - \underline{u}_h^{n-1})$ and $\partial_t p_{h\tau}^n := \partial_t p_{h\tau|I_n} = \tau_n^{-1}(p_h^n - p_h^{n-1})$ if $n \geqslant 1$.

4.3.3 Linearization

In order to solve the nonlinear problem (4.11) numerically, we apply a linearization algorithm, leading at each $1 \leqslant n \leqslant N$ to a series of linear algebraic systems, whose solutions converge to the solution of (4.11). In Code_Aster, this is done using the Newton–Raphson method. It is applied only to the nonlinear term, expressing the dependency of the effective stress tensor on the primal variable $\underline{u}$ through the strain tensor $\underline{\underline{\varepsilon}}(\underline{u})$. Note that, since $\underline{\underline{\varepsilon}}$ depends linearly on $\underline{u}$, it holds $\frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{u}}(\underline{u}) \cdot \underline{v} = \frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u})) : \underline{\underline{\varepsilon}}(\underline{v})$ for any $\underline{v} \in \underline{U}$. Let $(\underline{u}_h^{n,0}, p_h^{n,0}) \in \underline{U}_h^n \times P_h^n$ be an initial guess and $\alpha^{n,0} = \alpha^{n-1}$. Then the linearized discrete problem at time step n and iteration $i \geqslant 1$ is to find, for given $\underline{u}_h^{n-1}$, p_h^{n-1}, α^{n-1}, $\underline{u}_h^{n,i-1}$, $p_h^{n,i-1}$ and $\alpha^{n,i-1}$, the functions $(\underline{u}_h^{n,i}, p_h^{n,i}) \in \underline{U}_h^n \times P_h^n$ such that, for all $(\underline{v}_h, q_h) \in \underline{U}_h^n \times P_h^n$,

$$\begin{aligned}
&\Big(\frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1})) : \underline{\underline{\varepsilon}}(\underline{u}_h^{n,i}), \underline{\underline{\varepsilon}}(\underline{v}_h)\Big) - b(p_h^{n,i}, \nabla \cdot \underline{v}_h) \\
&\qquad = \left(\frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1})) : \underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}) - \underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}), \alpha^{n,i-1}), \underline{\underline{\varepsilon}}(\underline{v}_h)\right) + (\underline{f}^n, \underline{v}_h), \qquad (4.12a) \\
&(c_0(\partial_t^n p_{h\tau}^{n,i}, q) + b(\nabla \cdot \partial_t^n \underline{u}_{h\tau}^{n,i}, q_h) - (\phi(p_h^{n,i}), \underline{\nabla} q_h) = (g^n, q_h). \qquad (4.12b)
\end{aligned}$$

This problem corresponds to a linear equation system $A^{n,i-1}U^{n,i} = G^{n,i-1}$. We stop the iterations, when

$$\|A^{n,i-1}U^{n,i} - G^{n,i-1}\|_\infty \leqslant \Gamma_{\text{res}} \|G^{n,i-1}\|_\infty, \qquad (4.13)$$

where Γ_{res} is typically in the order of 10^{-6}. Let j denote the corresponding iteration of the Newton solver. We then set $(\underline{u}_h^n, p_h^n) = (\underline{u}_h^{n,j}, p_h^{n,j})$ and move forward to the next time step.

4.3.4 Initial guess

It is well-known that the convergence of the Newton method depends on the choice of the initial guess $(\underline{u}_h^{n,0}, p_h^{n,0})$, especially if the yield function F is such that the convergence is not guaranteed. In any case, a good initial guess reduces in practice the number of iterations necessary to approximate the solution of the nonlinear problem (4.11) to a sufficient accuracy. In Code_Aster the initial guess at each time step n is obtained by linearizing equation (4.11) at $(\underline{u}_h^{n-1}, p_h^{n-1})$ with respect to the time [44]. Since the only dependency on t is through $(\underline{u}, p)$, we obtain, similarly to (4.12),

$$\begin{aligned}
&\Big(\frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u}_h^{n-1})) : \underline{\underline{\varepsilon}}(\underline{u}_h^{n,0}), \underline{\underline{\varepsilon}}(\underline{v}_h)\Big) - b(p_h^{n,0}, \nabla \cdot \underline{v}_h) \\
&\qquad = \left(\frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u}_h^{n-1})) : \underline{\underline{\varepsilon}}(\underline{u}_h^{n-1}) - \underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n-1}), \alpha^{n-1}), \underline{\underline{\varepsilon}}(\underline{v}_h)\right) + (\underline{f}^{n-1}, \underline{v}_h), \\
&\qquad\qquad (c_0(\partial_t^n p_{h\tau}^{n,0}, q) + b(\nabla \cdot \partial_t^n \underline{u}_{h\tau}^{n,0}, q_h) - (\phi(p_h^{n,0}), \underline{\nabla} q_h) = (g^{n-1}, q_h),
\end{aligned}$$

for all $(\underline{v}_h, q_h) \in \underline{U}_h^n \times P_h^n$.

4.3.5 Integration of the mechanical behavior law

To solve (4.12), we have to calculate $\frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}))$ and $\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}), \alpha^{n,i-1})$ before each iteration. We start with $\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}), \alpha^{n,i-1})$, using (4.11c,4.11d). Therefore, we first have to find out if the displacement $\underline{u}_h^{n,i-1}$ causes elastic or plastic deformation in the material, taking into account the deformation history expressed by α^{n-1}. This is done by testing if an elastic deformation leads to an admissible pair $(\underline{\underline{\sigma}}', \alpha)$, i.e. if $F(\underline{\underline{\sigma}}', \alpha) \leqslant 0$. According to (4.11c,4.11d), the elastic deformation leads to

$$\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}), \alpha^{n,i-1}) = \underline{\underline{D}} : \underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}) \quad \text{and} \quad \alpha^{n,i-1} = \alpha^{n-1}.$$

We then distinguish the following two cases:

(1) If this pair satisfies $F(\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}), \alpha^{n,i-1}), \alpha^{n,i-1}) \leqslant 0$, then $\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}), \alpha^{n,i-1})$ is given by the above expression and for any $\underline{v}_h \in \underline{U}_h^n$ it holds

$$\frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1})) : \underline{\underline{\varepsilon}}(\underline{v}_h) = \underline{\underline{D}} : \underline{\underline{\varepsilon}}(\underline{v}_h).$$

(2) If $F(\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}), \alpha^{n,i-1}), \alpha^{n,i-1}) > 0$, the deformation $\underline{u}_h^{n,i-1}$ cannot lead to an elastic deformation, we thus have to apply the plastic behavior law and obtain

$$\begin{aligned}\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}), \alpha^{n,i-1}) &= \underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n-1}), \alpha^{n-1}) + (\underline{\underline{D}} - \underline{\underline{D}}') : \underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1} - \underline{u}_h^{n-1}) \\ \alpha^{n,i-1} &= \alpha^{n-1} - (\partial_\alpha F)^{-1} \underline{\underline{\partial}}_\sigma F : (\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}), \alpha^{n,i-1}) - \underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n-1}), \alpha^{n-1})),\end{aligned}$$

and, for any $\underline{v}_h \in \underline{U}_h^n$, it holds

$$\frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1})) : \underline{\underline{\varepsilon}}(\underline{v}_h) = \underline{\underline{D}} : \underline{\underline{\varepsilon}}(\underline{v}_h) - \frac{\partial (\underline{\underline{D}}' : \underline{\underline{\varepsilon}})}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1})) : \underline{\underline{\varepsilon}}(\underline{v}_h).$$

4.4 Equilibrated flux a posteriori error estimate

In this section we apply the results of Chapters 2 and 3 to the poro-plastic problem introduced in the previous section. We start by presenting the equilibrated fluxes. The velocity reconstruction is the extension to three space dimensions of the reconstruction in Section 2.3.1. The stress tensor reconstructions are the same as presented in Section 3.3.2 for nonlinear elasticity problems. We then present the error measure, which, as in Chapter 2, is the dual norm of the residual of the weak formulation (4.9), and derive an upper bound on this error, calculated at each time step and each Newton iteration using the equilibrated flux reconstructions. In the estimate, we distinguish between the linearization error for the mechanical part and the space and time discretization errors for both the mechanical and the hydraulic parts. Finally, we explain the hybridization technique used in the implementation of the flux reconstruction.

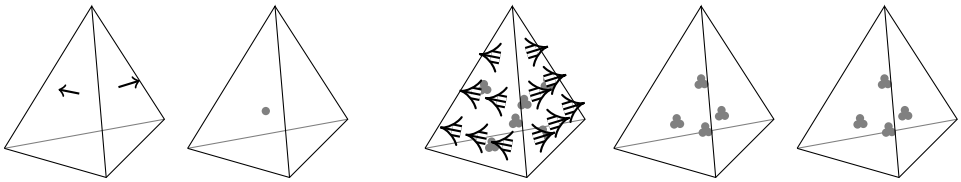


Figure 4.3 – The Raviart–Thomas and Arnold–Falk–Winther finite elements for $k = 1$

4.4.1 Quasi-static flux reconstruction

Comparing the properties of the continous and the discrete fluxes we observe the same differences as in the linear case: Recalling that $\underline{\underline{\sigma}}(\underline{u}, p, \alpha) = \underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}), \alpha) + \underline{\underline{\sigma}}_p(p)$, where $\underline{\underline{\sigma}}_p(p) = bp\underline{\underline{I}}$, the fluxes from (4.9) satisfy, for a. e. $t \in (0, t_F)$,

$$\begin{aligned} \underline{\underline{\sigma}}(\underline{u}, p, \alpha)(t) &\in \underline{\underline{H}}_{\mathrm{s}}(\mathrm{div}, \Omega), & -\underline{\nabla} \cdot \underline{\underline{\sigma}}(\underline{u}, p, \alpha)(t) &= \underline{f}(t), \\ \underline{\phi}(p)(t) &\in \underline{H}(\mathrm{div}, \Omega), & \nabla \cdot \underline{\phi}(p)(t) &= g(t) - \partial_t(\nabla \cdot \underline{u} + c_0 p)(t), \end{aligned}$$

wheras the discrete fluxes from (4.12) at time step $1 \leqslant n \leqslant N$ and Newton iteration $i \geqslant 0$ are in general such that

$$\begin{aligned} \underline{\underline{\sigma}}(\underline{u}_{h\tau}^{n,i}, p_{h\tau}^{n,i}, \alpha^{n,i}) &\notin \underline{\underline{H}}_{\mathrm{s}}(\mathrm{div}, \Omega), & -\underline{\nabla} \cdot \underline{\underline{\sigma}}(\underline{u}, p, \alpha) &\neq \underline{f}^{n,i}, \\ \underline{\phi}(p_{h\tau}^{n,i}) &\notin \underline{H}(\mathrm{div}, \Omega), & \nabla \cdot \underline{\phi}(p_{h\tau}^{n,i}) &\neq g^{n,i} - \partial_t^n(\nabla \cdot \underline{u}_{h\tau}^{n,i} + c_0 p_{h\tau}^{n,i}). \end{aligned}$$

For each vertex $a \in \mathcal{V}_h$, we define the corresponding hat function $\psi_a \in \mathbb{P}^1(\mathcal{T}_h)$ as the piecewise linear function taking value one at the vertex a and zero on all other mesh vertices. In this section, we recall briefly the two discrete reconstructions restoring these properties, starting with the Darcy velocity.

Equilibrated Darcy velocity reconstruction

Following the reconstruction in Section 2.3.1 we use the Raviart–Thomas finite elements of order $k-1$, illustrated in the left part of Figure 4.3 for $k = 1$. On one tetrahedron $T \in \mathcal{T}_h^n$ they are given by

$$\underline{W}_T = \underline{\mathbb{P}}^{k-1}(T) + \underline{x}\mathbb{P}^{k-1}(T).$$

For each vertex $a \in \mathcal{V}_h^n$ we then define the spaces on the element patches around a. If $a \in \mathcal{V}_h^{n,\mathrm{int}}$ we set

$$\begin{aligned} \underline{W}_h^a &:= \{\underline{v}_h \in \underline{H}(\mathrm{div}, \omega_a);\ \underline{v}_{h|T} \in \underline{W}_h(T)\ \forall T \in \mathcal{T}_a^n \text{ and } \underline{v}_h \cdot \underline{n}_{\omega_a} = 0 \text{ on } \partial\omega_a\}, \\ Q_h^a &:= \{q_h \in L^2(\omega_a);\ q_{h|T} \in \mathbb{P}^{k-1}(T)\ \forall T \in \mathcal{T}_a^n \text{ and } (q_h, 1)_{\omega_a} = 0\}, \end{aligned}$$

and if $a \in \mathcal{V}_h^{n,\mathrm{ext}}$, they are defined as

$$\begin{aligned} \underline{W}_h^a &:= \{\underline{v}_h \in \underline{H}(\mathrm{div}, \omega_a);\ \underline{v}_{h|T} \in \underline{W}_h(T)\ \forall T \in \mathcal{T}_a^n \text{ and } \underline{v}_h \cdot \underline{n}_{\omega_a} = 0 \text{ on } \partial\omega_a \backslash \partial\Omega\}, \\ Q_h^a &:= \{q_h \in L^2(\omega_a);\ q_{h|T} \in \mathbb{P}^{k-1}(T)\ \forall T \in \mathcal{T}_a^n\}. \end{aligned}$$

Construction 4.2 (Darcy velocity). *For each $a \in \mathcal{V}_h^n$ find $(\underline{\phi}_h^a, r_h^a) \in \underline{W}_h^a \times Q_h^a$ such that for all $(\underline{v}_h, q_h) \in \underline{W}_h^a \times Q_h^a$*

$$\left(\underline{\phi}_h^a, \underline{v}_h\right)_{\omega_a} - \left(r_h^a, \nabla \cdot \underline{v}_h\right)_{\omega_a} = \left(\psi_a \underline{\phi}(p_h^{n,i}), \underline{v}_h\right)_{\omega_a}, \tag{4.16a}$$

$$\left(\nabla \cdot \underline{\phi}_h^a, q_h\right)_{\omega_a} = \left(\psi_a g^n - \psi_a \partial_t^n (\nabla \cdot \underline{u}_{h\tau}^{n,i} + c_0 p_{h\tau}^{n,i}) + \underline{\nabla} \psi_a \cdot \underline{\phi}(p_h^n), q_h\right)_{\omega_a}. \tag{4.16b}$$

Then set $\underline{\phi}_h^{n,i} := \sum_{a \in \mathcal{V}_h^n} \underline{\sigma}_h^a$.

For interior vertices it holds $\left(\psi_a g^n - \psi_a \partial_t (\nabla \cdot \underline{u}_{h\tau} + c_0 p_{h\tau})^n + \underline{\nabla} \psi_a \cdot \underline{\phi}(p_h^n), 1\right)_{\omega_a} = 0$, guaranteeing that the pure Neumann problems (4.16) are well-posed for all $a \in \mathcal{V}_h^{n,\mathrm{int}}$. The resulting flux $\underline{\phi}_h^n$ lies in $\underline{H}(\mathrm{div}, \Omega)$ and satisfies the equilibrium

$$\nabla \cdot \underline{\phi}_h^n = g^n - \partial_t^n (\nabla \cdot \underline{u}_{h\tau} + c_0 p_{h\tau}). \tag{4.17}$$

The proof is the same as in Lemma 2.11, using also Assumption 4.1.

Equilibrated stress tensor reconstructions

For the three-dimensional computations we chose to rely on the stress reconstructions of introduced in Sections A.4 and 3.3, using the Arnold–Falk–Winther finite elements illustrated in the right part of Figure 4.3. As in Section 3.3, we compute two stress tensor reconstructions: one corresponding to the discrete stress and one expressing the linearization error in terms of a stress tensor. This distinction will be useful in Section 4.4.3, where we estimate the different error sources separately. We recall the corresponding mixed finite element spaces and the reconstructions for the sake of completeness. The Arnold–Falk–Winther spaces on one tetrahedron are the extension to tensors of the Brezzi–Douglas–Marini space for the stress, and Lagrange elements for the Lagrange multipliers imposing the equilibrium and the weak symmetry of the stress tensor. They are defined as

$$\begin{aligned}
\underline{\underline{\Sigma}}_T &:= \underline{\underline{\mathbb{P}}}^{k+1}(T), \\
\underline{V}_T &:= \underline{\mathbb{P}}^k(T), \\
\underline{\underline{\Lambda}}_T &:= \{\underline{\underline{\mu}} \in \underline{\underline{\mathbb{P}}}^k(T);\ \underline{\underline{\mu}} = -\underline{\underline{\mu}}^T\}.
\end{aligned}$$

Like before, we define for each vertex $a \in \mathcal{V}_h^n$ the mixed spaces on the patches around a, distinguishing between interior and boundary vertices. In the former case, we set

$$\underline{\underline{\Sigma}}_h^a := \{\underline{\underline{\tau}}_h \in \underline{\underline{H}}(\mathrm{div}, \omega_a);\ \underline{\underline{\tau}}_{h|T} \in \underline{\underline{\Sigma}}_T\ \forall T \in \mathcal{T}_h^n \text{ and } \underline{\underline{\tau}}_h \underline{n}_{\omega_a} = \underline{0} \text{ on } \partial\omega_a\}, \tag{4.19a}$$

$$\underline{V}_h^a := \{\underline{v}_h \in \underline{L}^2(\omega_a);\ \underline{v}_{h|T} \in \underline{V}_T\ \forall T \in \mathcal{T}_h^n \text{ and } (\underline{v}_h, \underline{z})_{\omega_a} = 0\ \forall \underline{z} \in \underline{RM}\}, \tag{4.19b}$$

$$\underline{\underline{\Lambda}}_h^a := \{\underline{\underline{\mu}}_h \in \underline{\underline{L}}^2(\omega_a);\ \underline{\underline{\mu}}_{h|T} \in \underline{\underline{\Lambda}}_T\ \forall T \in \mathcal{T}_h^n\}, \tag{4.19c}$$

where $\underline{RM} := \{\underline{b} + \underline{a} \times \underline{x};\ \underline{a}, \underline{b} \in \mathbb{R}^3\}$ is the space of rigid-body motions. If $a \in \mathcal{V}_h^{n,\mathrm{ext}}$ we set

$$\underline{\underline{\Sigma}}_h^a := \{\underline{\underline{\tau}}_h \in \underline{\underline{H}}(\mathrm{div}, \omega_a);\ \text{ and } \underline{\underline{\tau}}_{h|T} \in \underline{\underline{\Sigma}}_T\ \forall T \in \mathcal{T}_h^n \text{ and } \underline{\underline{\tau}}_h \underline{n}_{\omega_a} = \underline{0} \text{ on } \partial\omega_a \backslash \partial\Omega\}, \tag{4.19d}$$

$$\underline{V}_h^a := \{\underline{v}_h \in \underline{L}^2(\omega_a);\ \underline{v}_{h|T} \in \underline{V}_T\ \forall T \in \mathcal{T}_h^n\}, \tag{4.19e}$$

$$\underline{\underline{\Lambda}}_h^a := \{\underline{\underline{\mu}}_h \in \underline{\underline{L}}^2(\omega_a);\ \underline{\underline{\mu}}_{h|T} \in \underline{\underline{\Lambda}}_T\ \forall T \in \mathcal{T}_h^n\}. \tag{4.19f}$$

Then the two stress tensor reconstructions are defined as in Section 3.3.2:

Construction 4.3 (Discrete stress reconstruction). *For each $a \in \mathcal{V}_h^n$ find $(\underline{\underline{\sigma}}_h^a, \underline{r}_h^a, \underline{\underline{\lambda}}_h^a) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a$ such that for all $(\underline{\underline{\tau}}_h, \underline{v}_h, \underline{\underline{\mu}}_h) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a$,*

$$\begin{aligned}
(\underline{\underline{\sigma}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} + (\underline{r}_h^a, \underline{\nabla} \cdot \underline{\underline{\tau}}_h)_{\omega_a} + (\underline{\underline{\lambda}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} &= (\psi_a \underline{\underline{\sigma}}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i}), \underline{\underline{\tau}}_h)_{\omega_a}, \\
(\underline{\nabla} \cdot \underline{\underline{\sigma}}_h^a, \underline{v}_h)_{\omega_a} &= (-\psi_a \underline{f} + \underline{\underline{\sigma}}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i}) \underline{\nabla} \psi_a - \underline{y}_{\mathrm{disc}}^{n,i}, \underline{v}_h)_{\omega_a}, \\
(\underline{\underline{\sigma}}_h^a, \underline{\underline{\mu}}_h)_{\omega_a} &= 0,
\end{aligned}$$

where $\underline{y}_{\mathrm{disc}}^{n,i} \in \underline{RM}$ is the unique solution of

$$(\underline{y}_{\mathrm{disc}}^{n,i}, \underline{z})_{\omega_a} = -(\underline{f}, \psi_a \underline{z})_{\omega_a} + (\underline{\underline{\sigma}}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i}), \underline{\underline{\varepsilon}}(\psi_a \underline{z}))_{\omega_a} \quad \forall \underline{z} \in \underline{RM}^d.$$

Then set $\underline{\underline{\sigma}}_{h,\mathrm{disc}}^{n,i} := \sum_{a \in \mathcal{V}_h} \underline{\underline{\sigma}}_h^a$.

To express the linearization error, we observe that we can rewrite equation (4.12a) as

$$(\underline{\underline{\sigma}}^{i-1}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i}), \underline{\underline{\varepsilon}}(\underline{v}_h)) = (\underline{f}^n, \underline{v}_h),$$

where the linear function $\underline{\underline{\sigma}}^{i-1}$ maps $(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i})$ onto

$$\underline{\underline{\sigma}}^{i-1}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i}) := \frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1})) : \underline{\underline{\varepsilon}}(\underline{u}_h^{n,i} - \underline{u}_h^{n,i-1}) + \underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}), \alpha^{n,i-1}) + \underline{\underline{\sigma}}_p(p_h^{n,i}).$$

Construction 4.4 (Linearization error stress reconstruction). *For each $a \in \mathcal{V}_h$ find $(\underline{\underline{\sigma}}_h^a, \underline{r}_h^a, \underline{\underline{\lambda}}_h^a) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a$ such that for all $(\underline{\underline{\tau}}_h, \underline{v}_h, \underline{\underline{\mu}}_h) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a$,*

$$\begin{aligned}
(\underline{\underline{\sigma}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} + (\underline{r}_h^a, \underline{\nabla} \cdot \underline{\underline{\tau}}_h)_{\omega_a} + (\underline{\underline{\lambda}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} &= (\psi_a(\underline{\underline{\sigma}}^{i-1}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i}) - \underline{\underline{\sigma}}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i})), \underline{\underline{\tau}}_h)_{\omega_a}, \\
(\underline{\nabla} \cdot \underline{\underline{\sigma}}_h^a, \underline{v}_h)_{\omega_a} &= ((\underline{\underline{\sigma}}^{i-1}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i}) - \underline{\underline{\sigma}}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i})) \underline{\nabla} \psi_a + \underline{y}_{\mathrm{disc}}^{n,i}, \underline{v}_h)_{\omega_a}, \\
(\underline{\underline{\sigma}}_h^a, \underline{\underline{\mu}}_h)_{\omega_a} &= 0.
\end{aligned}$$

Then set $\underline{\underline{\sigma}}_{h,\mathrm{lin}}^{n,i} := \sum_{a \in \mathcal{V}_h} \underline{\underline{\sigma}}_h^a$.

As shown in Lemma 3.9 together with Assumption 4.1, these two stress tensors lie in $\underline{\underline{H}}(\mathrm{div}, \Omega)$ and verify for all $1 \leqslant n \leqslant N$

$$-\underline{\nabla} \cdot (\underline{\underline{\sigma}}_{h,\mathrm{disc}}^k + \underline{\underline{\sigma}}_{h,\mathrm{lin}}^k) = \underline{f}. \tag{4.22}$$

4.4.2 Error measure

We proceed as in Chapter 2 and introduce reference parameters allowing us to define the residual of (4.9a,4.9b), which are not written in the same physical units. We use the Young modulus E, a reference time scale $t^\star$ and a reference length scale $l^\star$.

We denote $X := L^2(0,t_F;\underline{H}^1_0(\Omega)) \times L^2(0,t_F;H^1_0(\Omega))$, $Y = (\underline{U} \times P \times L^2(0,t_F;L^2(\Omega)))$, and their restrictions to a time interval I_n by X_n and Y_n. Then we can define the map $\mathcal{B} : Y \times X \to L^2(0,t_F;\mathbb{R})$ by

$$\mathcal{B}((\underline{u},p,\alpha),(\underline{v},q)) := (\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}),\alpha),\underline{\underline{\varepsilon}}(\underline{v})) - b(p,\nabla\cdot\underline{v}) + t^\star(c_0(\dot{p},q)+b(\nabla\cdot\underline{\dot{u}},q) - (\phi(p),\underline{\nabla}q)), \tag{4.23}$$

and thus the weak formulation (4.9) is equivalent to finding $(\underline{u},p,\alpha) \in Y$ verifying the initial conditions (4.7c), such that for a. e. $t \in (0,t_F)$,

$$\mathcal{B}((\underline{u},p,\alpha),(\underline{v},q)) = (\underline{f},\underline{v}) + t^\star(g,q) \qquad \forall(\underline{v},q) \in X, \tag{4.24}$$

with (4.9c,4.9d). We can then, similarly to Section 2.4.1, define on every time interval I_n the residual of (4.24) for any $(\underline{u}_{h\tau},p_{h\tau},\alpha) \in Y_n$ and $(\underline{v},q) \in X_n$

$$\langle\mathcal{R}(\underline{u}_{h\tau},p_{h\tau},\alpha),(\underline{v},q)\rangle_{X'_n,X_n} := \int_{I_n} \mathcal{B}((\underline{u}_{h\tau},p_{h\tau},\alpha),(\underline{v},q)) - (\underline{f}^n,\underline{v}) - t^\star(g^n,q)\mathrm{d}t \tag{4.25}$$

On the space X_n we define the norm

$$\|(\underline{v},q)\|^2_{X_n} := \int_{I_n} (E\|\underline{\underline{\varepsilon}}(\underline{v})\|)^2 + (l^\star\|\underline{\nabla}q\|)^2\mathrm{d}t.$$

Then the error measure $e^{n,i}$ for the error between the weak solution of (4.9) and the discrete solution $(\underline{u}^{n,i}_{h\tau},p^{n,i}_{h\tau},\alpha^{n,i})$ of (4.12) at the time step $1 \leqslant n \leqslant N$ and iteration $i \geqslant 0$ is defined by the dual norm of the residual

$$e^{n,i} := \|\mathcal{R}(\underline{u}^{n,i}_{h\tau},p^{n,i}_{h\tau},\alpha^{n,i})\|_{X'_n} = \sup_{(\underline{v},q)\in X_n,\|(\underline{v},q)\|_{X_n}=1} \langle\mathcal{R}(\underline{u}^{n,i}_{h\tau},p^{n,i}_{h\tau},\alpha^{n,i}),(\underline{v},q)\rangle_{X'_n,X_n}. \tag{4.26}$$

The global error is defined as

$$\begin{aligned} e &:= \sum_{n=1}^{N} \|\mathcal{R}(\underline{u}_{h\tau},p_{h\tau},\alpha)\|_{X'_n} = \|\mathcal{R}(\underline{u}_{h\tau},p_{h\tau},\alpha)\|_{X'} \\ &= \sup_{(\underline{v},q)\in X,\|(\underline{v},q)\|_X=1} \int_0^{t_F} \mathcal{B}((\underline{u}_{h\tau},p_{h\tau},\alpha),(\underline{v},q)) - (\underline{f},\underline{v}) - (g,q)\mathrm{d}t, \end{aligned}$$

where the equation in the first line is shown in Corollary 2.16.

4.4.3 A posteriori error estimate

In this chapter, we directly present the a posteriori error estimate distinguishing different error sources. For each time step $1 \leqslant n \leqslant N$ and each iteration $i \geqslant 0$, let $\underline{\phi}_h^n$, $\underline{\underline{\sigma}}_{h,\mathrm{disc}}^{n,i}$ and $\underline{\underline{\sigma}}_{h,\mathrm{lin}}^{n,i}$ be respectively the velocity and stress reconstructions from Constructions 4.2, 4.3 and 4.4. For all $T \in \mathcal{T}_h^n$ we define the following local estimators of the space and time discretization error, as well as of the linearization error for the mechanical part,

$$\eta_{\mathrm{sp,mec},T}^{n,i} := E^{-1} \|\underline{\underline{\sigma}}_{h,\mathrm{disc}}^{n,i} - \underline{\underline{\sigma}}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i})\|_T, \tag{4.27a}$$

$$\eta_{\mathrm{tm,mec},T}^{n,i}(t) := E^{-1} \|\underline{\underline{\sigma}}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i}) - \underline{\underline{\sigma}}(\underline{u}_{h\tau}^{n,i}, p_{h\tau}^{n,i}, \alpha^{n,i})(t)\|_T, \tag{4.27b}$$

$$\eta_{\mathrm{lin,mec},T}^{n,i} := E^{-1} \|\underline{\underline{\sigma}}_{h,\mathrm{lin}}^{n,i}\|_T, \tag{4.27c}$$

and the local spatial and temporal discretization error estimators for the hydraulic part:

$$\eta_{\mathrm{sp,hyd},T}^{n,i} := \frac{t^\star}{l^\star} \|\underline{\phi}_h^{n,i} - \underline{\phi}(p_h^{n,i})\|_T, \tag{4.27d}$$

$$\eta_{\mathrm{tm,hyd},T}^{n,i}(t) := \frac{t^\star}{l^\star} \|\underline{\phi}(p_h^{n,i}) - \underline{\phi}(p_{h\tau}^{n,i})(t)\|_T. \tag{4.27e}$$

For each of these error sources, the global estimator is given by

$$\eta_\bullet := \left(3 \int_{I_n} \sum_{T \in \mathcal{T}_h^n} \left(\eta_\bullet^{n,i}(t) \right)^2 \mathrm{d}t \right)^{1/2} \tag{4.28}$$

We can then present the following local and global in time error estimates:

Theorem 4.5 (Local in time a posteriori error estimate)**.** *Let $(\underline{u}, p) \in Y$ be the weak solution of (4.9), let Assumption 4.1 hold and let, at a time step $1 \leqslant n \leqslant N$ and a Newton iteration $i \geqslant 0$, $(\underline{u}_h^{n,i}, p_h^{n,i}) \in \underline{U}_h^n \times P_h^n$ be the discrete solution of the linearized system (4.12). Let $e^{n,i}$ be defined by (4.26) and the local error estimators given by (4.27). Then the following holds:*

$$e^{n,i} \leqslant \eta_{\mathrm{sp,mec}}^{n,i} + \eta_{\mathrm{tm,mec}}^{n,i} + \eta_{\mathrm{lin,mec}}^{n,i} + \eta_{\mathrm{sp,hyd}}^{n,i} + \eta_{\mathrm{tm,hyd}}^{n,i}. \tag{4.29}$$

Proof. Proceeding as in the proof of Theorem 2.13, using $\underline{\underline{\sigma}}_{h,\mathrm{disc}}^{n,i} + \underline{\underline{\sigma}}_{h,\mathrm{lin}}^{n,i}$ instead of $\underline{\underline{\theta}}_h^n$ and taking into account (4.22) and (4.17), we obtain

$$e^{n,i} \leqslant \left(\sum_{n=1}^{N} \int_{I_n} \sum_{T \in \mathcal{T}_h^n} \left(E^{-1} \|\underline{\underline{\sigma}}_{h,\mathrm{disc}}^{n,i} + \underline{\underline{\sigma}}_{h,\mathrm{lin}}^{n,i} - \underline{\underline{\sigma}}(\underline{u}_{h\tau}^{n,i}, p_{h\tau}^{n,i}, \alpha^{n,i})(t)\|_T \right)^2 + \left(\|\underline{\phi}_h^{n,i} - \underline{\phi}(p_{h\tau}^{n,i})(t)\|_T \right)^2 \right)^{1/2}.$$

Applying the triangle inequality to distinguish the error sources, together with (4.28), yields

$$\begin{aligned} e^{n,i} &\leqslant \left(\sum_{n=1}^{N} \int_{I_n} \sum_{T \in \mathcal{T}_h^n} \left(\eta_{\mathrm{sp,mec},T}^{n,i} + \eta_{\mathrm{tm,mec},T}^{n,i}(t) + \eta_{\mathrm{lin,mec},T}^{n,i} \right)^2 + \left(\eta_{\mathrm{sp,hyd},T}^{n,i} + \eta_{\mathrm{tm,hyd},T}^{n,i}(t) \right)^2 \right)^{1/2} \\ &\leqslant \eta_{\mathrm{sp,mec}}^{n,i} + \eta_{\mathrm{tm,mec}}^{n,i} + \eta_{\mathrm{lin,mec}}^{n,i} + \eta_{\mathrm{sp,hyd}}^{n,i} + \eta_{\mathrm{tm,hyd}}^{n,i}. \end{aligned}$$

□

Corollary 4.6 (Global in time a posteriori error estimate)**.** *Let $(\underline{u}, p) \in Y$ be the weak solution of* (4.9)*, let Assumption 4.1 hold and let $(\underline{u}_{h\tau}, p_{h\tau})$ be the discrete solution of* (4.11)*. Let e be defined by* (4.27a) *and the global error estimators given by* (4.28) *with* (4.27)*. Then the following holds:*

$$e \leqslant \eta_{\mathrm{sp,mec}} + \eta_{\mathrm{tm,mec}} + \eta_{\mathrm{lin,mec}} + \eta_{\mathrm{sp,hyd}} + \eta_{\mathrm{tm,hyd}}. \tag{4.30}$$

The proof of Corollary 4.6 is similar to the proof of Corollary 2.16.

4.4.4 Hybridization of the local problems

The local problems (4.16), (4.20) and (4.21) result in saddle point problems, whose algebraic resolution is in general more costly than for symmetric positive definite systems of similar size. Following the idas of [59], the systems can be tranformed into symmetric positive definite systems by removing the inter-element continuities of the flux variables and imposing them using Lagrange multipliers. The system can then be reduced using static condensation, since flux and potential variables are discontinous from one element to the other.

We present the approach for the stress reconstructions. The same procedure (ignoring the space of skew-symmetric matrices used in the stress reconstruction to impose the weak symmetry) can be applied to the velocity reconstruction and is presented for example in [22, Chapter 7.2]

We start by replacing the space $\underline{\underline{\Sigma}}_h^a$ in (4.19) by the broken Brezzi–Douglas–Marini spaces, i.e. for each $a \in \mathcal{V}_h^n$ we set

$$\underline{\underline{\Sigma}}_h^a := \{ \underline{\underline{\tau}}_h \in \underline{\underline{L}}^2(\omega_a);\ \underline{\underline{\tau}}_{h|T} \in \underline{\underline{\Sigma}}_T\ \forall T \in \mathcal{T}_h^n \}.$$

The spaces $\underline{V}_h^a$ and $\underline{\underline{\Lambda}}_h^a$ are still defined as in (4.19). The continuity of the normal components of the stress functions is thus no longer imposed by the choice of the discretization space. Instead, we use Lagrange multipliers living in the space

$$\underline{\underline{\Xi}}_h^a := \underline{\underline{\mathbb{P}}}^{k+1}(\mathcal{F}_a),$$

where $\mathcal{F}_a$ denotes the set of all faces in ω_a if $a \in \mathcal{V}_h^{n,\mathrm{int}}$, and the set of all faces in ω_a not lying

on the boundary $\partial\Omega$ if $a \in \mathcal{V}_h^{n,\mathrm{ext}}$. Using these spaces we can reformulate the local problems of Constructions (4.20) and (4.21). Note that the left hand sides of these two systems are equivalent. We introduce the functions Γ^a and γ^a, which for the discrete stress reconstructions are

$$\begin{aligned}\underline{\underline{\Gamma}}^a &= \psi_a \underline{\underline{\sigma}}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i}),\\ \underline{\gamma}^a &= -\psi_a \underline{f} + \underline{\underline{\sigma}}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i})\underline{\underline{\nabla}}\psi_a - \underline{y}_{\mathrm{disc}}^{n,i},\end{aligned}$$

whereas for the linearization error stress reconstruction they read

$$\begin{aligned}\underline{\underline{\Gamma}}^a &= (\psi_a(\underline{\underline{\sigma}}^{i-1}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i}) - \underline{\underline{\sigma}}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i})),\\ \underline{\gamma}^a &= (\underline{\underline{\sigma}}^{i-1}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i}) - \underline{\underline{\sigma}}(\underline{u}_h^{n,i}, p_h^{n,i}, \alpha^{n,i}))\underline{\underline{\nabla}}\psi_a + \underline{y}_{\mathrm{disc}}^{n,i}.\end{aligned}$$

Then, the equivalent problems consist in seeking $(\underline{\underline{\sigma}}_h^a, \underline{r}_h^a, \underline{\underline{\lambda}}_h^a, \underline{\chi}_h^a) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a \times \underline{\underline{\Xi}}_h^a$ such that for all $(\underline{\underline{\tau}}_h, \underline{v}_h, \underline{\underline{\mu}}_h, \underline{\xi}_h) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a \times \underline{\underline{\Xi}}_h^a$,

$$(\underline{\underline{\sigma}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} + \sum_{T\in\mathcal{T}_{\omega_a}} (r_h^a, \underline{\nabla}\cdot\underline{\underline{\tau}}_h)_T + (\underline{\underline{\lambda}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} - \sum_{F\in\mathcal{F}_a} (\underline{\chi}_h^a, [\![\underline{\underline{\tau}}_h \underline{n}_F]\!])_F = (\underline{\underline{\Gamma}}^a, \underline{\underline{\tau}}_h)_{\omega_a} \tag{4.33a}$$

$$\sum_{T\in\mathcal{T}_{\omega_a}} (\underline{\nabla}\cdot\underline{\underline{\sigma}}_h^a, \underline{v}_h)_T = (\underline{\gamma}^a, \underline{v}_h)_{\omega_a}, \tag{4.33b}$$

$$(\underline{\underline{\sigma}}_h^a, \underline{\underline{\mu}}_h)_{\omega_a} = 0, \tag{4.33c}$$

$$-\sum_{F\in\mathcal{F}_a} ([\![\underline{\underline{\sigma}}_h^a \underline{n}_F]\!]_F, \underline{\xi}_h)_F = 0. \tag{4.33d}$$

Note that, in this formulation, we impose the homogenous Neumann condition of (4.19a) and (4.19d) weakly using the space $\underline{\underline{\Xi}}_h^a$ and the definition of $\mathcal{F}_a$, which has the practical advantage that for any element $T \in \mathcal{T}_h^n$ the spaces $\underline{\underline{\Sigma}}_{h|T}^a$ are the same, independently of the vertex a.

The problem (4.33a) corresponds to a matrix system of the form

$$\begin{pmatrix} A & B_1^t & B_2^t & C^t \\ B_1 & & & \\ B_2 & & & \\ C & & & \end{pmatrix} \begin{pmatrix} \Sigma \\ R \\ \Lambda \\ X \end{pmatrix} = \begin{pmatrix} F \\ G' \\ 0 \\ 0 \end{pmatrix}. \tag{4.34}$$

Choosing basis functions equal to zero on all but exactly one element for the spaces $\underline{\underline{\Sigma}}_h^a$, $\underline{V}_h^a$ and $\underline{\underline{\Lambda}}_h^a$, the matrices A, B_1 and B_2 are block diagonal. Defining $B = (B_1, B_2)^T$, $Y = (R, \Lambda)^T$ and $G = (G', 0)^T$, we can eliminate the variable Σ by

$$\Sigma = A^{-1}(-B^t Y - C^t X + F), \tag{4.35}$$

leading to the system

$$\begin{pmatrix} -BA^{-1}B^t & -BA^{-1}C^t \\ -CA^{-1}B^t & -CA^{-1}C^t \end{pmatrix} \begin{pmatrix} Y \\ X \end{pmatrix} = \begin{pmatrix} -BA^{-1}F + G \\ -CA^{-1}F \end{pmatrix}.$$

Our next step is to eliminate Y, writing

$$Y = (BA^{-1}B^t)^{-1}[-BA^{-1}C^tX + BA^{-1}F - G].$$

Defining now

$$D = CA^{-1}B^t(BA^{-1}B^t)^{-1}BA^{-1}C^t - CA^{-1}C^t, \tag{4.36a}$$

$$H = CA^{-1}B^t(BA^{-1}B^t)^{-1}[BA^{-1}F - G] - CA^{-1}F, \tag{4.36b}$$

the system becomes

$$DX = H.$$

The matrix D is symmetric and positive definite, owing to the fact that the sum of the two bilinear forms corresponding to the matrix B verifies the discrete inf-sup inequality (c.f. [9,21]).

4.5 Adaptive Algorithm

We propose, similarly to Section 2.4.3 and 3.4.3, an adaptive algorithm balancing the error sources considered in the error estimate (4.29), namely the space and time discretization and the linearization error estimate. The algorithm is based on Algorithm 2.18 for the adaptation of the discretization, without considering an initial error estimate, since we assume that the initial conditions can be imposed exactly. Let therefore $\Gamma_{\text{tm}} > 1 > \gamma_{\text{tm}} > 0$ be user-given weights and crit^n, for all $1 \leqslant n \leqslant N$, a chosen threshold that the error on the time interval I_n should not exceed. We add an adaptive stopping criterion for the linearization loop. To this purpose, we introduce the user-given weight $\gamma_{\text{lin}} > 0$, which typically takes values around 0.1. In order to avoid an influence of the reference parameters $t^\star$ and $l^\star$ on the adaptive stopping of the linearization iterations, we compare the linearization error estimate only to the discretization error estimate of the mechanical part and stop the iterations when it becomes negligible in this comparison.

Algorithm 4.7 (Adaptive algorithm)**.**

1. **Initialisation**: choose an initial mesh $\mathcal{T}_h^0$, an initial time step τ_0, and set $t^0 := 0$
2. **Time loop**
   (a) Set $n := n + 1$, $\mathcal{T}_h^n := \mathcal{T}_h^{n-1}$, and $\tau_n := \tau_{n-1}$
   (b) Calculate the initial guess $(\underline{u}_h^{n,0}, p_h^{n,0})$

(c) Calculate $\frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,0}))$ and $\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,0}), \alpha^{n,0})$

(d) **Linearization loop**

i. Calculate $(\underline{u}_h^{n,i}, p_h^{n,i})$ and the terms $\alpha^{n,i}$, $\frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i}))$ and $\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i}), \alpha^{n,i})$

ii. Calculate the estimators $\eta_{\mathrm{sp,mec}}^{n,i}$, $\eta_{\mathrm{tm,mec}}^{n,i}$ and $\eta_{\mathrm{lin,mec}}^{n,i}$

End of the linearization loop if

$$\eta_{\mathrm{lin,mec}}^{n,i} \leqslant \gamma_{\mathrm{lin}}(\eta_{\mathrm{sp,mec}}^{n,i} + \eta_{\mathrm{tm,mec}}^{n,i}) \tag{4.37}$$

(e) Set $\underline{u}_h^n := \underline{u}_h^{n,i}$, $p_h^n := p_h^{n,i}$, $\eta_{\mathrm{sp,mec}}^n := \eta_{\mathrm{sp,mec}}^{n,i}$ and $\eta_{\mathrm{tm,mec}}^n := \eta_{\mathrm{tm,mec}}^{n,i}$

(f) Calculate the estimators $\eta_{\mathrm{sp,hyd}}^n$, $\eta_{\mathrm{tm,hyd}}^n$ and set $\eta_{\mathrm{sp}}^n := \eta_{\mathrm{sp,mec}}^n + \eta_{\mathrm{sp,hyd}}^n$ and $\eta_{\mathrm{tm}}^n := \eta_{\mathrm{tm,mec}}^n + \eta_{\mathrm{tm,hyd}}^n$

(g) **Space refinement loop**

i. **Space and time error balancing loop**

A. if $\gamma_{\mathrm{tm}}\eta_{\mathrm{sp}}^n > \eta_{\mathrm{tm}}^n$: Set $\tau_n := 2\tau_n$

B. if $\Gamma_{\mathrm{tm}}\eta_{\mathrm{sp}}^n < \eta_{\mathrm{tm}}^n$: Set $\tau_n := \frac{1}{2}\tau_n$

End of the space-time error balancing loop if

$$\gamma_{\mathrm{tm}}\eta_{\mathrm{sp}}^n \leqslant \eta_{\mathrm{tm}}^n \leqslant \Gamma_{\mathrm{tm}}\eta_{\mathrm{sp}}^n \qquad \text{or} \qquad \tau_n \leqslant \tau_{\min} \tag{4.38}$$

ii. Refine or coarsen the mesh $\mathcal{T}_h^n$ such that the local spatial error estimators $\eta_{\mathrm{sp},T}^n$ are distributed equally

End of the space refinement loop if

$$\eta_{\mathrm{sp}}^n + \eta_{\mathrm{tm}}^n \leqslant \mathrm{crit}^n \tag{4.39}$$

End of the time loop if $t^n \geqslant t_F$

Owing to (4.30), (2.51) and (4.37) the obtained discrete solution satisfies

$$e \leqslant \left(\sum_{n=1}^{N} (\mathrm{crit}^{\mathrm{n}})^2 \right)^{1/2}. \tag{4.40}$$

4.6 Examples of elasto-plastic laws used in geomechanics

For isotropic materials, it is useful to describe the material independently of the used basis for the space variables. It is thus common not to express the yield as a function of $\underline{\underline{\sigma}}$ directly, but to use a set of functions of $\underline{\underline{\sigma}}'$ which are invariant under rotations. Some often used variables (cf. e.g. [17]) are the following:

- The *principal stresses* :

  $\sigma_1, \sigma_2, \sigma_3 \in \mathbb{R}$, such that $\sigma_1 \geqslant \sigma_2 \geqslant \sigma_3$ and $\begin{pmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3 \end{pmatrix}$ is the diagonalization of $\underline{\underline{\sigma}}'$.

  This diagonalization always exists, since the image of $\underline{\underline{\sigma}}$ is the space of real symmetric matrices.

- The *invariants of the stress tensor*:

$$I_1(\underline{\underline{\sigma}}') = \operatorname{tr} \underline{\underline{\sigma}}', \; I_2(\underline{\underline{\sigma}}') = \frac{1}{2}\underline{\underline{\sigma}}' : \underline{\underline{\sigma}}', \; I_3(\underline{\underline{\sigma}}') = \det(\underline{\underline{\sigma}}').$$

- The *equivalent von Mises stress* depending on the *deviator* $\underline{\underline{s}} := \underline{\underline{\sigma}}' - \frac{1}{3}\operatorname{tr} \underline{\underline{\sigma}}' \underline{\underline{I}}$:

$$\sigma_{\text{eq}}(\underline{\underline{\sigma}}') = \sqrt{\tfrac{3}{2}\underline{\underline{s}} : \underline{\underline{s}}} = \sqrt{\tfrac{1}{2}\left((\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2\right)}.$$

Depending on the properties of the material, some of these variables can be more convenient than others for expressing the model. The first invariant of the stress tensor expresses for example the volumetric part of the stress. Some materials, like metals, do not show any plastic deformation under hydrostatic stress (which corresponds to $\sigma_1 = \sigma_2 = \sigma_3$, for example applying a uniform pressure on a sphere). Then, it is practical to use the invariant σ_{eq} of the deviator, where the volumetric part of the stress is eliminated, since $\operatorname{tr} \underline{\underline{s}} = 0$. It is clear, that $\sigma_{\text{eq}} = 0$ corresponds to a hydrostatic stress. In the $\sigma_1 - \sigma_2 - \sigma_3-$system these points correspond to the line $\{\lambda(1,1,1)^t;\ \lambda \in \mathbb{R}\}$. The use of the invariant σ_{eq} is convenient to describe materials like metals, which show an elastic behaviour for any hydrostatic stress but react by plastifying whenever the variation between σ_1, σ_2 and σ_3 (called *shear stress*) exceeds a certain level. The set $\{\underline{\underline{\tau}};\ \sigma_{\text{eq}}(\underline{\underline{\tau}}) = k\}$ with $k > 0$ describes an open cylinder with radius k around the line $\{\underline{\underline{\tau}};\ \sigma_{\text{eq}}(\underline{\underline{\tau}}) = 0\}$. The following example is the yield function of metals, called von Mises criterion, which will help understand the slightly more complicated criteria used to modelize soils and rocks.

4.6.1 The von Mises criterion

Let $\sigma_{\text{y}} > 0$ be the tensile strength of the material. This is the limit from which on the material will undergo plastic deformations under uniaxial tension, and thus easy to determine in experiments. The von Mises yield function is defined as

$$F(\underline{\underline{\sigma}}') = \sigma_{\text{eq}}(\underline{\underline{\sigma}}') - \sigma_{\text{y}}.$$

There are different ways of visualizing the corresponding yield surface. In the left picture in Figure 4.4 it is displayed in the space of the principal stress. The middle figure shows it in the deviatoric plane, which is the projection of space of principal stress onto the plane $\{(\sigma_1, \sigma_2, \sigma_3);\ \sigma_1 + \sigma_2 + \sigma_3 = 0\}$. In the right picture the criterion is visualized in the $\sigma_1 - \sigma_2-$plane.

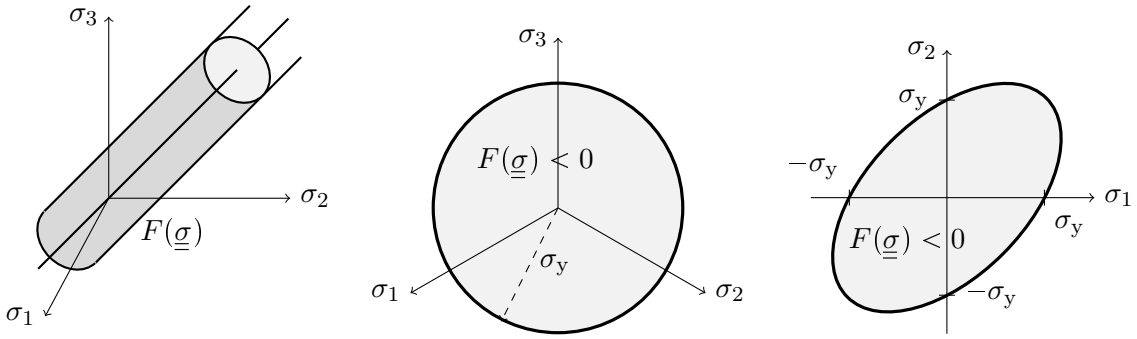


Figure 4.4 – The von Mises yield criterion

4.6.2 The Drucker–Prager criterion

Unlike metals, geomaterials deform plastically under too high tension, whereas hydrostatic compression makes them more elastic under shear stress. Therefore, the Drucker–Prager yield function is usually written in terms of σ_{eq} and I_1. In the $I_1 - \sigma_{\text{eq}}-$ plane the set $\{\underline{\underline{\tau}};\ F(\underline{\underline{\tau}}) = 0\}$ is a half-line, as shown in the right picture of Figure 4.5. Introducing the parameter $A > 0$ denoting its gradient, the yield function reads

$$F(\underline{\underline{\sigma}}') = \sigma_{\text{eq}}(\underline{\underline{\sigma}}') + AI_1(\underline{\underline{\sigma}}') - \sigma_{\text{y}}. \tag{4.41}$$

In the principal stress space, the yield surface is an open cone, as can be seen in the left picture of Figure 4.5. As mentioned earlier, we only consider softening behavior in this work. We recall that softening corresponds to a decrease of the yield function F for constant $\underline{\underline{\sigma}}'$ and increasing α. The way the yield surface moves depends once again on the material, and is expressed as a function $R : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ of α. For the Drucker–Prager yield criterion we can for example consider linear softening

$$R(\alpha) = \begin{cases} \sigma_{\text{y}} + a\alpha & \text{if } \alpha \leqslant \alpha_{\text{ult}}, \\ \sigma_{\text{y}} + a\alpha_{\text{ult}} & \text{if } \alpha > \alpha_{\text{ult}}, \end{cases}$$

or parabolic softening expressed by

$$R(\alpha) = \begin{cases} \sigma_{\text{y}} \left(1 - \left(1 - \sqrt{\frac{\sigma_{\text{y,ult}}}{\sigma_{\text{y}}}}\right) \frac{\alpha}{\alpha_{\text{ult}}}\right)^2 & \text{if } \alpha \leqslant \alpha_{\text{ult}}, \\ \sigma_{\text{y,ult}} & \text{if } \alpha > \alpha_{\text{ult}}, \end{cases} \tag{4.42}$$

where $a, \sigma_{\text{y,ult}}, \alpha_{\text{ult}} > 0$ with $\sigma_{\text{y,ult}} < \sigma_{\text{y}}$ are experimentally determined parameters. The yield criterion is then written as

$$f(\underline{\underline{\sigma}}', \alpha) = \sigma_{\text{eq}}(\underline{\underline{\sigma}}') + AI_1(\underline{\underline{\sigma}}') - R(\alpha),$$

which, in the case $\alpha = 0$, is equal to (4.41).

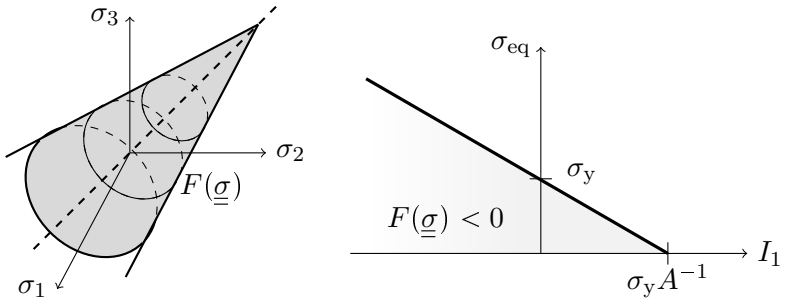


Figure 4.5 – The Drucker–Prager yield criterion

At the cone tip, F is not differentiable, implying that in the case of a plastic deformation, the direction of the latter is not defined. There are different possibilities of handling this singularity. One way consists in smoothening the cone tip, such that it becomes differentiable. The drawback of this approach is that it introduces a (non physical) regularization parameter the solution will depend on. In Code_Aster, a supplementary step is added to the steps defined in Section 4.3.5: after verifying if the pair $(\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}), \alpha^{n,i-1}), \alpha^{n,i-1})$ is elastic in step (1), we verify if it lies on the cone tip in step (1b):

(1b) Since at the cone tip it holds $\sigma_{\text{eq}}(\underline{\underline{\sigma}}') = 0$ and $I_1(\underline{\underline{\sigma}}') = R(\alpha)A^{-1}$, the stress is given by

$$\underline{\underline{\sigma}}' = \frac{1}{3} R(\alpha^{n,i-1}) A^{-1} \underline{\underline{I}},$$

where $\alpha^{n,i-1}$ is approximated by solving

$$R(\alpha^{n,i-1} - \alpha^{n-1}) = A I_1(\underline{\underline{D}} : \underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1} - \underline{u}_h^{n-1})),$$

i.e. by projecting the stress obtained using a linearly elastic stress increment onto the cone tip. We then verify if the corresponding plastic strain

$$\underline{\underline{\varepsilon}}^p = \underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}) - \underline{\underline{D}}^{-1} : \underline{\underline{\sigma}}' = \underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}) - \frac{1}{3} R(\alpha^{n,i-1}) A^{-1} \underline{\underline{D}}^{-1} : \underline{\underline{I}}$$

lies inside the normal cone of the yield surface, i.e. if for any $\underline{\underline{\tau}}$ verifying $F(\underline{\underline{\tau}}) \leqslant 0$ it holds $\underline{\underline{\tau}} : \underline{\underline{\varepsilon}}^p \leqslant 0$. If this is the case, we already have the values for $\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1}), \alpha^{n,i-1})$ and $\alpha^{n,i-1}$, and, in order to apply Newton's method (4.12), we have to calculate $\frac{\partial \underline{\underline{\sigma}}'}{\partial \underline{\underline{\varepsilon}}}(\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i-1})) : (\underline{\underline{\varepsilon}}(\underline{u}_h^{n,i} - \underline{u}_h^{n,i-1})$. In the case of the Drucker–Prager law, it is possible to express this term only using the increment $\alpha^{n,i-1} - \alpha^{n-1}$, for details we refer to [45, Section 2.2.3].

If $\underline{\underline{\varepsilon}}^p$ does not lie inside the normal cone the stress tensor violates the nomal condition and is not admissible. We then proceed as in step (2) in Section 4.3.5, assuming that the strain is plastic and that $\underline{\underline{\sigma}}'$ does not lie on the cone tip, meaning that the plastic behavior law depending on $\underline{\underline{\partial}}_\sigma F$ is well-defined.

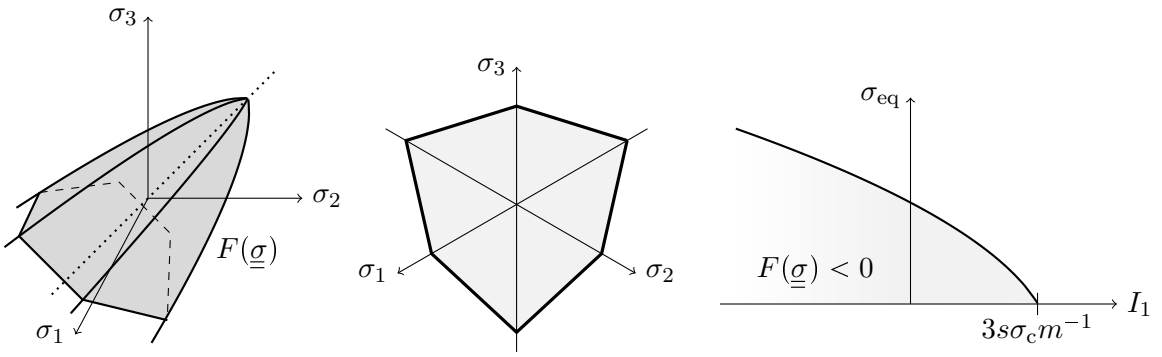


Figure 4.6 – The Hoek–Brown yield criterion displayed from left to right in the principal stress space, on the plane $\{\underline{\underline{\sigma}};\ \sigma_{\text{eq}}(\underline{\underline{\sigma}}) = 0\}$ and in the $I_1 - \sigma_{\text{eq}}-$ plane

4.6.3 The Hoek–Brown criterion

The Hoek–Brown criterion [63–65] is one of the most used plasticity criteria in mining engineering. Its yield function, diplayed in Figure 4.6 is defined as

$$F(\underline{\underline{\sigma}}') = \sigma_3(\underline{\underline{\sigma}}') - \sigma_1(\underline{\underline{\sigma}}') - \sqrt{-m\sigma_{\text{c}}\sigma_3(\underline{\underline{\sigma}}') + s\sigma_c^2}, \tag{4.43}$$

where σ_c is the uniaxial compressive strength of the intact rock, linked to the uniaxial tensile strength by $\sigma_{\text{c}} = 2\sigma_{\text{y}}(m - (m^2 + 4s)^{1/2})^{-1}$, $m > 0$ is a material parameter and $0 \leqslant s \leqslant 1$ describes the damage and fracturation state, with $s = 1$ corresponding to the intact rock. The major difference between the Drucker–Prager and the Hoek–Brown criterion is that, in the latter, the curve $\{\underline{\underline{\sigma}};\ F(\underline{\underline{\sigma}}) = 0\}$ in the $I_1 - \sigma_{\text{eq}}$–space is not a half-line, but part of half a parabola, as can be seen in the right parts of Figures 4.5 and 4.6. For the Hoek–Brown criterion, the damage leading to softening is taken into account in the variables s and m.

4.7 Numerical results

In this section we present our numerical results. We start with a poro-elastic test with an analytical solution in order to compare the convergence of the space discretization error estimators to the convergence of the error. We then apply Algorithm 4.7 to an industrial test simulating the excavation of a tunnel, using different mechanical behavior laws.

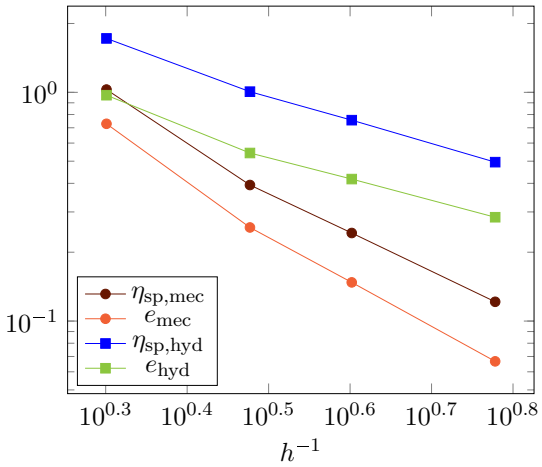


Figure 4.7 – Comparison of the space discretization error estimate and the error in the analytical test case

4.7.1 Analytical test

We consider the extension to three dimensions of the analytical solution of the linear elastic Biot problem of Section 2.5.2. Let $\Omega = (0,1) \times (0,1) \times (0,1)$ and $t \in (0, \frac{1}{2})$ and

$$\underline{u}(t,x,y,z) = \sin(-\pi t) \begin{pmatrix} \cos(\pi x)\sin(\pi y)\sin(\pi z) \\ \sin(\pi x)\cos(\pi y)\sin(\pi z) \\ \sin(\pi x)\sin(\pi y)\cos(\pi z) \end{pmatrix}$$
$$p(t,x,y) = -\cos(-\pi t)\sin(\pi x)\sin(\pi y)\sin(\pi z)$$

with $\kappa = 1$, $c_0 = 0$, and the stress-strain relation

$$\underline{\underline{\sigma}}'(\underline{\underline{\varepsilon}}) = 2\mu\underline{\underline{\varepsilon}} + \lambda \operatorname{tr}(\underline{\underline{\varepsilon}})\underline{\underline{I}},$$

with the Lamé parameters $\mu = \lambda = 0.4$. The resulting source terms $\underline{f}$ and g in (4.7) are given by

$$\underline{f} = (3.6\pi^2 \sin(-\pi t) - \pi\cos(-\pi t)) \begin{pmatrix} \cos(\pi x)\sin(\pi y)\sin(\pi z) \\ \sin(\pi x)\cos(\pi y)\sin(\pi z) \\ \sin(\pi x)\sin(\pi y)\cos(\pi z) \end{pmatrix},$$

and $g = 0$. The exact solution is imposed as Dirichlet condition on the boundary $\partial\Omega$. To evaluate the convergence of the space discretization error estimators we compare the mechanical estimators to the error

$$e_{\text{mec}} := \left(\sum_{n=1}^{N} \int_{I_n} \|\underline{\underline{\sigma}}'(\underline{u} - \underline{u}_h^n)\|^2 \mathrm{d}t \right)^{1/2},$$

and the hydraulic estimators to

$$e_{\text{hyd}} := \left(\sum_{n=1}^{N} \int_{I_n} \|\underline{\phi}(p - p_h^n)\|^2 \mathrm{d}t \right)^{1/2}.$$

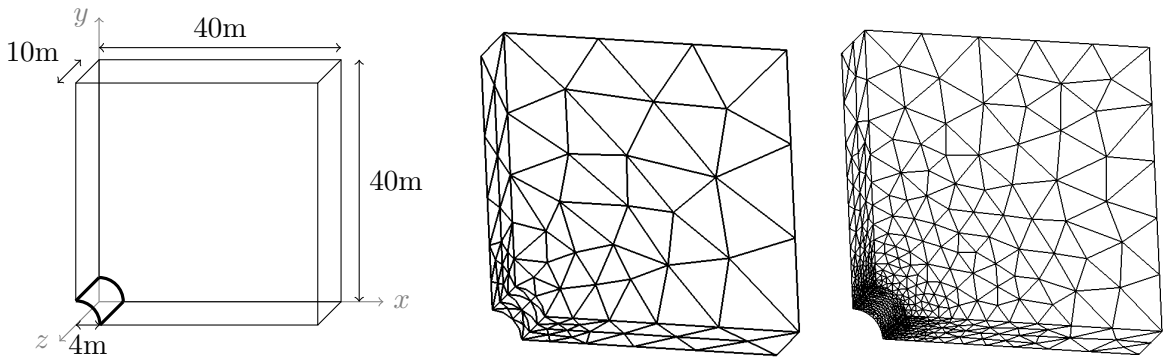


Figure 4.8 – The geometry of the rock with the hole for the tunnel and the coarsest and finest mesh for the static computations

Figure 4.7 shows the comparison of the spatial discretization error estimators to the mechanical and hydraulic errors obtained on a series of uniformly refined meshes with the fixed time step $\tau = 0.5 \cdot 10^{-4}$. We observe that the error estimators reflect the convergence of these errors.

4.7.2 Tunnel excavation

In this test we apply the adaptive algorithm of Section 4.5 to a simulation of a tunnel excavation. The simulation is a three dimensional extension of the numerical test presented in Section 2.5.4 with only one gallery. We perform the same simulation using different mechanical behavior laws: one in linear elasticity, one using the Drucker–Prager model and finally one using the viscoplastic L&K model [68] based on the Hoek–Brown criterion.

The domain Ω, illustrated in the left of Figure 4.8, is a 10m $\times$ 40m $\times$ 40m cutout of the rock, in which a tunnel is digged in the (horizontal) z-direction. The excavation time is $t_F = 1.728 \cdot 10^6$s, corresponding to 20 days. Like in the two-dimensional case, we simulate the excavation by first calculating the initial total equilibrium of the hole-free geometry, and then letting the boundary condition decrease linearly from the so obtained values to zero on the tunnel wall. At the top and the right of $\partial\Omega$ we set $p = p_0 = 4.5$MPa and $\underline{\underline{\sigma}}\,\underline{n}_\Omega = \underline{\underline{\sigma}}_0 \underline{n}_\Omega$, where $(\underline{\underline{\sigma}}_{0,xx}, \underline{\underline{\sigma}}_{0,yy}, \underline{\underline{\sigma}}_{0,zz}, \underline{\underline{\sigma}}_{0,xy}, \underline{\underline{\sigma}}_{0,xz}, \underline{\underline{\sigma}}_{0,yz}) = (-16.4\text{MPa}, -12.7\text{MPa}, -12.4\text{MPa}, 0, 0, 0)$. On the rest of $\partial\Omega$ we apply symmetry conditions, i.e. $u_x = 0$ on the left, $u_y = 0$ at the bottom and $u_z = 0$ at the front and back of Ω. The initial conditions are $p(\cdot, 0) = p_0$ and $\underline{\underline{\sigma}}(\cdot, 0) = \underline{\underline{\sigma}}_0$, and the source terms $\underline{f}$ and g are equal to zero.

In all simulations we use the realistic parameter set

$$E = 3.62 \cdot 10^9\text{Pa}, \quad \nu = 0.3, \quad b = 0.6, \quad c_0 = 9 \cdot 10^{-11}\text{Pa}^{-1}, \quad \kappa = 10^{-17}\text{m}^2\text{Pa}^{-1}s^{-1},$$

where the Young modulus E and the Poisson ratio ν lead to the Lamé parameters $\mu \approx 1.4 \cdot 10^9$Pa and $\lambda \approx 2.1 \cdot 10^9$Pa. The reference parameters are chosen as $l^\star = 40$m and $t^\star = 86400$s, corresponding to one day.

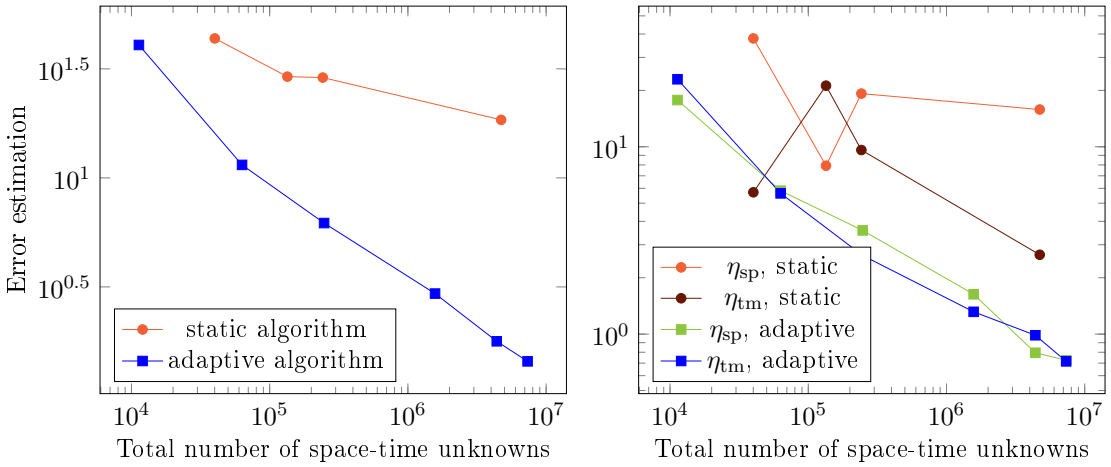


Figure 4.9 – Comparison of the global error estimates obtained using a static and an adaptive algorithm for the poro-elastic problem

For each of the three considered mechanical behavior laws we perform four static tests, meaning that we use a fixed mesh and time step. Table 4.1 recapitulates for each test the number of tetrahedra and time steps. For test 2 we chose a finer mesh and a larger time step than for test 3, such that the number of space-time unknowns for both tests is similar. Our goal is to see if the spatial and temporal discretization error estimators will capture this difference. For the two nonlinear mechanical behavior laws, the linearization is done using the Newton method of Section 4.3.3 with the convergence criterium $\Gamma_{\text{res}} = 10^{-6}$ in (4.13). If after ten iterations this convergence criterium is not verified, the current time step is divided by four.

Linear elastic mechanical behavior

We start with the linear elastic behavior law, where we compare the error estimates of the static algorithms to the estimates obtained applying Agorithm 4.7 which, in this case, only equilibrates the discretization errors by adapting the time step and the mesh, since we consider a linear problem. The adaptation parameters are chosen as $\gamma_{\text{tm}} = 0.8$ and $\Gamma_{\text{tm}} = 1.3$. Figure 4.9 shows that in our tests the use of the adaptive algorithm reduces significantly the number of space-time unknowns and equilibrates the two discretization error sources. In the left part of Figure 4.10 we compare the distribution of the spatial discretization error estimators at $t = t_F$ on the finest static mesh, which is the right mesh in Figure 4.8, to the ones of the last adaptive test (i.e. the one with the lowest error estimate). The corresponding

	test 1	test 2	test 3	test 4
Number of elements	511	1948	1612	3454
Number of time steps	10	10	16	20

Table 4.1 – Number of elements and time steps in the four static algorithms

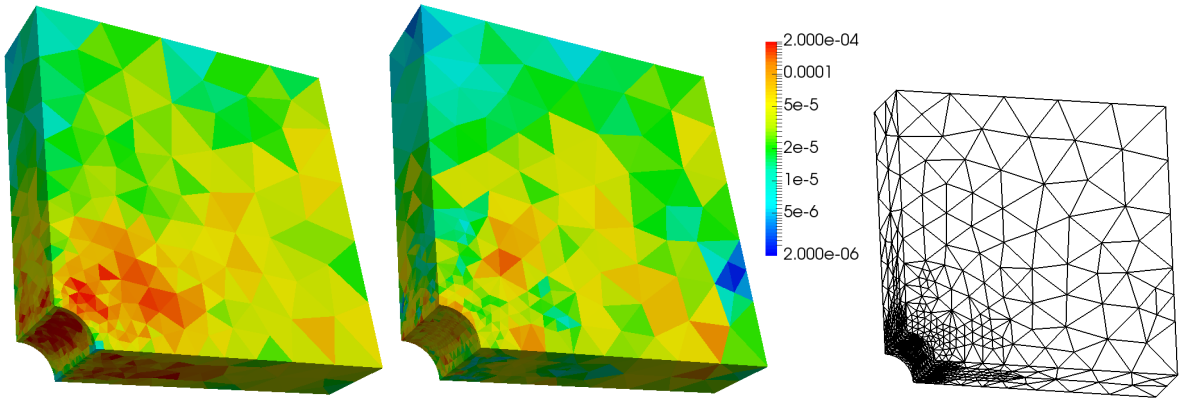


Figure 4.10 – Comparison of spatial discretization error estimators after a static and an adaptive algorithm using a linear elatic behavior law (left) and the adaptively refined mesh (right)

mesh has 4356 elements and is displayed in the right of Figure 4.10. The error estimators are distributed more evenly on the adaptively refined mesh, especially around the tunnel wall the values of the estimators are smaller than on the fixed mesh.

Drucker–Prager behavior law

In the Drucker–Prager yield function (4.41) we set $A = 0.33$. We use parabolic softening with the parameters $\alpha_{\text{ult}} = 0.01$, $\sigma_{\text{y}} = 2.57\text{MPa}$, and $\sigma_{\text{y,ult}} = 0.57\text{MPa}$ in (4.42).

We can only present the results for the three static computations 2, 3 and 4, since in the first test the algorithm did not converge. We start by comparing each of the static computations with a computation using the adaptive stopping criterion (4.37) with $\gamma_{\text{lin}} = 0.1$ for the Newton method, but without adapting the discretization. Table 4.2 shows the number of iterations, computation time and obtained error estimates for each of the three considered mesh and time step configurations for both the static and the adaptive case. We observe that the number of iterations is considerably reduced using the adaptive stopping criterion. The computation time of the static algorithms are measured without estimating the error. Since we performed each simulation only once, the given computation times should only be considered as an orientation, since they can vary for identic simulations. Nevertheless, we observe that the estimation of

test	static iterations	static comp. time	static η	adaptive iterations	adaptive comp. time	adaptive η
2	122	1min 9s	61.86	66	15min 38s	64.76
3	170	1min 17s	55.30	31	4min 20s	60.03
4	61	1min 6s	22.97	39	15min 38s	23.14

Table 4.2 – Comparison of the number of iterations, the computation time and the global estimate of the static algorithms and algorithms with the same mesh and time step using the adaptive stopping criterion (4.37) with $\gamma_{\text{lin}} = 0.1$ for the Drucker–Prager model

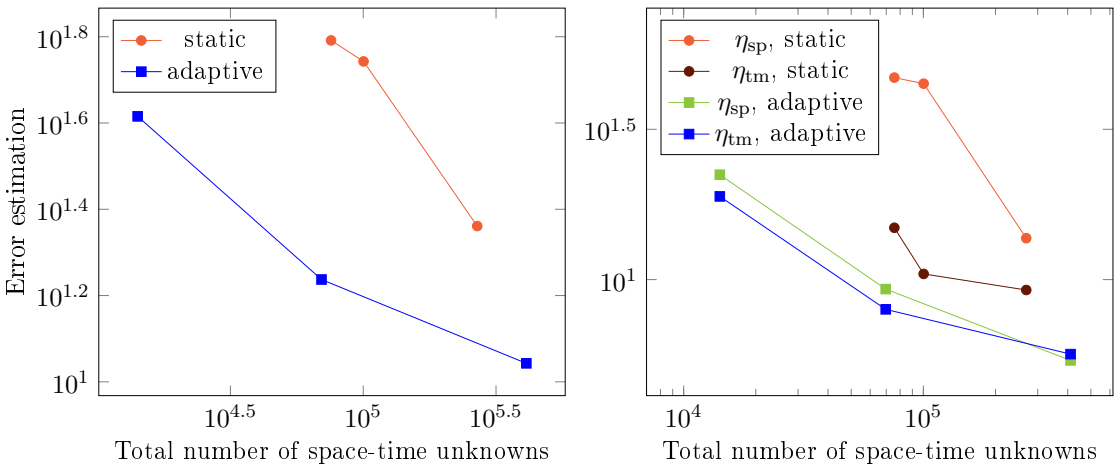


Figure 4.11 – Comparison of the global error estimates obtained for the Drucker–Prager behavior law using a static and an adaptive algorithm equilibrating the space and time discretization error

the error is quite expensive. This is, on the one hand, due to the size of the local problems for the equilibrated stress reconstructions. In the hybridized problems, there are 18 unknowns on each face of the patch, whereas the fluid velocity reconstruction only requires one unknown per face. On the other hand, we integrated the computation into an industrial software, in which the resolution of the global problem is optimized, and whose architecture is not adapted for the use of mixed finite elements. We will detail this point in Section 4.7.3

We then perform a series of simulations, where we adapt the mesh and the time step such that the corresponding discretization error estimators satisfy (4.38) and (4.39). Again we compare the global error estimates obtained in the static computations to the ones from the adaptive algorithm. Figure 4.11 shows that the total number of space-time unknowns is reduced for comparable error estimate, and that the discretization error estimators are balanced using the adaptive algorithm. We observe that, without adaptation, the space discretization error estimators dominates the estimate, which was not the case for the linear elastic behavior law, and that the space discretization error estimate in test 2 is higher than in test 3, althouh the second mesh is finer than the third (cf. Table 4.1).

L&K behavior law

The L&K behavior law [46, 68, 91] is a viscoplastic law developped by EDF to modelize the behavior of rocks, and, in particular, the rock formation destined for the nuclear waste storage project Cigéo in the East of France (cf. Section 1.1). The plasticity criterion in this model is based on the Hoek–Brown yield function (4.43). We will not detail the model and its parameters here and refer to [46].

As in the Drucker–Prager test, we start by comparing the number of performed Newton

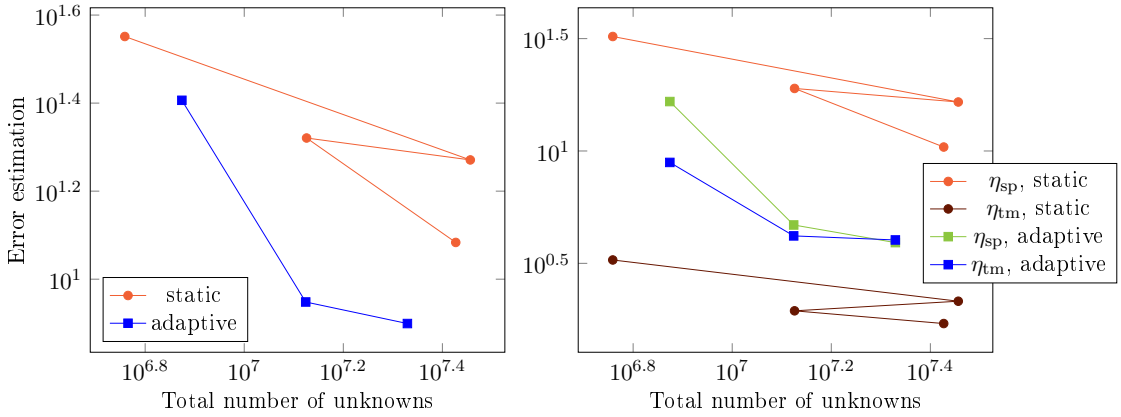


Figure 4.12 – Comparison of the global error estimates obtained for the L&K behavior law using a static and an adaptive algorithm equilibrating the linearization and the space and time discretization error

iterations with the original convergence criterion (4.13) with $\Gamma_{\text{res}} = 10^{-6}$ and the adaptive criterion (4.37) with $\gamma_{\text{lin}} = 0.1$ in algorithms with fixed time steps and meshes. In Table 4.3 we see that we perform less iterations with the adaptive stopping criterion. Compared to the Drucker–Prager test, we save less iterations but the error estimates are closer to the ones without the adaptive stopping of the Newton iterations. Again, we see that the computation time is much longer due to the calculation of the error estimate.

We then compare the error estimates obtained in the static computations to the error estimates obtained when using Algorithm 4.7 combining the discretization adaptation and the adaptive stopping criterion. In the left of Figure 4.12 we show the estimates as a function of the total number of unknowns. This number is calculated by summing up the number of unknowns of each performed iteration. Since the second static test requires a lot of iterations compared to the other tests (cf. Table 4.3), this number is higher than for test 4, which uses a finer mesh and more timsteps. Again, the estimated error for a comparable number of unknowns is reduced by the use of the adaptive algorithm.

Figure 4.13 shows the distribution of the space discretization error estimators at $t = t_F$ on the finest mesh of the static algorithms (in the right of Figure 4.8) and on an adaptively refined mesh with 2352 elements, which is displayed in the right of the figure. Again we observe that the distribution is more evenly on the adaptively refined mesh, and that, although the mesh

test	static iterations	static comp. time	static η	adaptive iterations	adaptive comp. time	adaptive η
1	2865	12min 53s	35.59	2167	1h 15min 35s	35.60
2	4562	1h 30min 13s	18.66	3625	8h 51min 7s	18.65
3	1765	30min 54s	20.92	1376	4h 13min 38s	20.93
4	1989	1h 46min 1s	12.12	1386	8h 29min 40s	12.12

Table 4.3 – Comparison of the number of iterations, the computation time and the global estimate of the static algorithm and an algorithm using the adaptive stopping criterion (4.37) with $\gamma_{\text{lin}} = 0.1$ for the L&K model

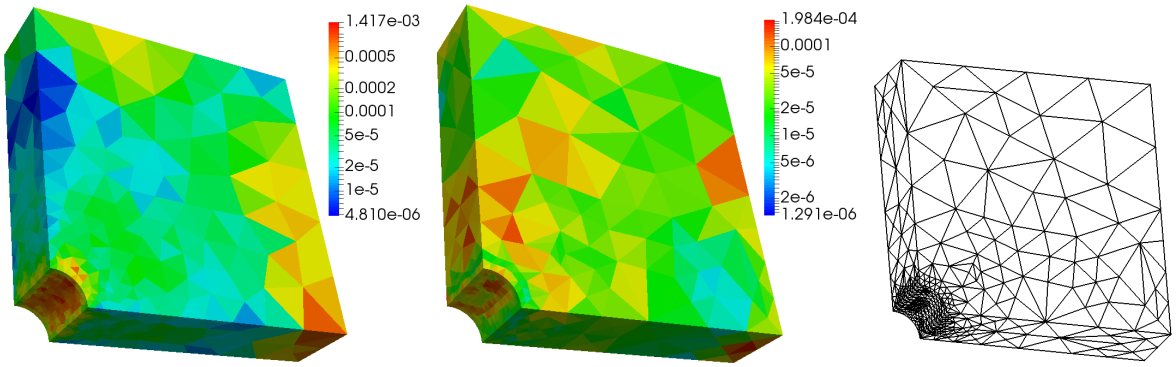


Figure 4.13 – Comparison of spatial discretization error estimators after a static and an adaptive algorithm using the L&K behavior law (left) and the adaptively refined mesh (right)

has less elements, the maximum value is smaller than on the static mesh.

4.7.3 Comment on the implementation

The computation of the error estimators is implemented in the finite element software Code_Aster, but is not yet mature for industrial use. Code_Aster is a finite element code based on Lagrange finite elements, meaning that degrees of freedom can only exist on vertices. The computation of integrals on elements and their following assembling in the global matrix are optimized. However, if degrees of freedom do not lie on vertices, like it is the case for the finite element spaces we use to reconstruct the equilibrated fluxes, the assembly tools provided in the code can not be used. Owing to the hybridization of the local problems, the matrices A and B in (4.34) are block-diagonal and we can use the tools for element computations to calculate them. We only have to assembly the matrix C, whose element contributions are also computed using these tools. This choice is essential for the integration of the error estimators into the logic of the code, and also convenient since the elementary computation routines provide a fast calculation of the barycentric coordinates, which facilitates the computation of the Raviart–Thomas and Brezzi–Douglas–Marini basis functions. The drawback of the use of these elementwise tools is that the resulting elementary matrices are – anticipating their assembly – not stored as a table one can access with one pointer. Values in an elementary matrix have to be accessed one by one by calling a routine providing their address. Especially for the stress tensor reconstruction, the recuperation of these matrices is extemely expensive. In order to save computation time, we construct and save for each vertex the matrix D in (4.36a), but also the necessary matrices to construct H and Σ in (4.36b) and (4.35) respectively, which requires a lot of memory. We then only have to compute the vectors F and G in (4.34), using again the elementary computation tools.

The introduction of degrees of freedom on mesh faces in Code_Aster is planned in the context of another PhD thesis at EDF R&D. The computation of the equilibrated flux reconstructions can then be reimplemented and industrialized in order to reduce the computation time. An-

other possible improvement is to use parallelization, since the local problems are independent of each other.

4.7.4 Conclusion

The results we obtain balancing the different error sources in the excavation test are promising. In our tests the adaptive stopping of the linearization iterations has an insignificant effect on the discretization error, while the number of total iterations is reduced by up to 80%. In each of the three test series (corresponding to the three considered mechanical behavior laws) the number of unknowns in the computation for a comparable error estimate was reduced by using adaptive algorithms.

Appendix A

Comparison of the two stress reconstruction techniques

This chapter consists of a conference proceeding published as "Equilibrated stress reconstructions for linear elasticity problems with application to a posteriori error analysis" in Finite Volumes for Complex Applications VIII – Methods and Theoretical Aspects with Application, written with Daniele Di Pietro and Alexandre Ern.

Contents

Abstract

We present an a posteriori error estimate for the linear elasticity problem. The estimate is based on an equilibrated reconstruction of the Cauchy stress tensor, which is obtained from mixed finite element solutions of local Neumann problems. We propose two different reconstructions: one using Arnold–Winther mixed finite element spaces providing a symmetric stress tensor, and one using Arnold–Falk–Winther mixed finite element spaces with a weak symmetry constraint. The performance of the estimate is illustrated on a numerical test with analytical solution.

A.1 Introduction

We consider the linear elasticity problem on a simply connected polygon $\Omega \subset \mathbb{R}^2$:

$$-\underline{\nabla} \cdot \underline{\underline{\sigma}}(\underline{u}) = \underline{f} \qquad \text{in } \Omega, \tag{A.1a}$$

$$\underline{u} = \underline{0} \qquad \text{on } \partial\Omega, \tag{A.1b}$$

where $\underline{u} : \Omega \to \mathbb{R}^2$ the displacement, and $\underline{f} : \Omega \to \mathbb{R}^2$ the volumetric body force. The Cauchy stress tensor $\underline{\underline{\sigma}}$ is given by Hooke's law $\underline{\underline{\sigma}}(\underline{u}) = \lambda \operatorname{tr}(\underline{\underline{\varepsilon}}(\underline{u}))\underline{\underline{I}}_2 + 2\mu\underline{\underline{\varepsilon}}(\underline{u})$, where λ and μ are the Lamé parameters, and the symmetric gradient $\underline{\underline{\varepsilon}}(\underline{u}) = \frac{1}{2}((\underline{\underline{\nabla}}\underline{u})^T + \underline{\underline{\nabla}}\underline{u})$ describes the infinitesimal strain.

In many applications, this problem is approximated using H^1-conforming finite elements. It is well known that, in contrast to the analytical solution, the resulting discrete stress tensor does not have continous normal components across mesh interfaces, and its divergence is not locally in equilibrium with the source term $\underline{f}$ on mesh cells. In this paper we propose an a posteriori error estimate based on stress tensor functions which are reconstructed from the discrete stress tensor such that they verify both of the above properties. Such equilibrated-flux a posteriori error estimates offer serveral advantages. First, error upper bounds are obtained with fully computable constants. Second, polynomial-degree robustness can be achieved for the Poisson problem in [27,56], for linear elasticity in [48], and for the related Stokes problem in [33]. Third, they allow one to distinguish among various error components, e.g., discretization, linearization, and algebraic solver error components, and to equilibrate adaptively these components in the iterative solution of nonlinear problems [55]. An advantage for more general problems in solid mechanics is that the stress reconstruction is based on the discrete stress (not the displacement) and thus the estimate does not depend on the mechanical behaviour law.

We present two stress reconstructions. Both use mixed finite elements on cell patches around mesh vertices, as proposed for the Poisson problem in [28,38]. The first one was introduced in [95] and uses the Arnold–Winther (AW) mixed finite element spaces [10] providing a symmetric stress tensor. The second one follows the same approach, but imposing the symmetry only weakly and using the Arnold–Falk–Winther (AFW) mixed finite element spaces [9]. Elementwise reconstructions of equilibrated stress tensors from local Neumann problems can be found in [5,67,71], whereas direct prescription of the degrees of freedom in the AW finite element space is considered in [82].

A.2 Setting

We denote by $L^2(\Omega)$ the space of square-integrable functions taking values in $\mathbb{R}$, and by $(\cdot,\cdot)$ and $\|\cdot\|$ the corresponding inner product and norm. $H^1(\Omega)$ stands for the Sobolev space com-

posed of $L^2(\Omega)$ functions with weak gradients in $\underline{L}^2(\Omega)$ and $H^1_0(\Omega)$ for its zero-trace subspace. The weak formulation of problem (A.1) reads: find $\underline{u} \in \underline{H}^1_0(\Omega)$ such that

$$(\underline{\underline{\sigma}}(\underline{u}), \underline{\underline{\varepsilon}}(\underline{v})) = (\underline{f}, \underline{v}) \qquad \forall \underline{v} \in \underline{H}^1_0(\Omega). \tag{A.2}$$

The discretization of (A.2) is based on a conforming triangulation $\mathcal{T}_h$ of Ω, verifying the minimum angle condition. We will use a conforming finite element method of order $p \geqslant 2$. Let $\mathbb{P}^p(\mathcal{T}_h) := \{v \in L^2(\Omega) \mid \forall T \in \mathcal{T}_h \ v|_T \in \mathbb{P}^p(T)\}$, where $\mathbb{P}^p(T)$ is the space of polynomials on T of degree less than or equal to p. For the sake of simplicity we assume that $\underline{f}$ lies in $\underline{\mathbb{P}}^{p-1}(\mathcal{T}_h)$. Then the discrete problem reads: find $\underline{u}_h \in \underline{H}^1_0(\Omega) \cap \underline{\mathbb{P}}^p(\mathcal{T}_h)$ such that

$$(\underline{\underline{\sigma}}(\underline{u}_h), \underline{\underline{\varepsilon}}(\underline{v}_h)) = (\underline{f}, \underline{v}_h) \qquad \forall \underline{v}_h \in \underline{H}^1_0(\Omega) \cap \underline{\mathbb{P}}^p(\mathcal{T}_h). \tag{A.3}$$

A.3 A Posteriori Error Estimate

In this section, we derive an upper bound on the error between the analytical solution of (A.2) and an arbitrary function $\underline{u}_h \in \underline{H}^1_0(\Omega) \cap \underline{\mathbb{P}}^p(\mathcal{T}_h)$. We will measure this error in the energy norm

$$\|\underline{v}\|^2_{\mathrm{en}} := (\underline{\underline{\sigma}}(\underline{v}), \underline{\underline{\varepsilon}}(\underline{v})) = 2\mu \|\underline{\underline{\varepsilon}}(\underline{v})\|^2 + \lambda \|\underline{\nabla} \cdot \underline{v}\|^2 \geqslant 2\mu C_{\mathrm{K}} \|\underline{\underline{\nabla}} \underline{v}\|^2, \tag{A.4}$$

where the last bound follows from $\lambda \geqslant 0$ and Korn's inequality. Owing to (A.1b), we have $C_{\mathrm{K}} = \frac{1}{2}$ (this value would have been different if we had chosen mixed boundary conditions). We start by introducing reconstructed stress tensors that are more "physical" than $\underline{\underline{\sigma}}(\underline{u}_h)$, which in general does not lie in $\underline{\underline{H}}(\mathrm{div}, \Omega) = \{\underline{\underline{\tau}} \in \underline{\underline{L}}^2(\Omega) \mid \underline{\nabla} \cdot \underline{\underline{\tau}} \in \underline{L}^2(\Omega)\}$ and thus cannot verify the equilibrium equation (A.1a). Unlike $\underline{\underline{\sigma}}(\underline{u}_h)$, however, these reconstructed tensors may not be symmetric.

Definition A.1 (Equilibrated stress reconstruction)**.** *We call* equilibrated stress reconstruction *any function* $\underline{\underline{\sigma}}_h \in \underline{\underline{H}}(\mathrm{div}, \Omega)$ *constructed from* $\underline{\underline{\sigma}}(\underline{u}_h)$ *such that*

$$(-\underline{\nabla} \cdot \underline{\underline{\sigma}}_h, \underline{z})_T = (\underline{f}, \underline{z})_T \quad \forall \underline{z} \in \underline{RM} \ \forall T \in \mathcal{T}_h, \tag{A.5}$$

where $\underline{RM} := \{\underline{b} + c(x_2, -x_1)^T \mid \underline{b} \in \mathbb{R}^2, c \in \mathbb{R}\}$ *is the space of rigid body motions.*

Theorem A.2 (A posteriori error estimate)**.** *Let* $\underline{u} \in \underline{H}^1_0(\Omega)$ *solve* (A.2) *and* $\underline{u}_h \in \underline{H}^1_0(\Omega)$ *be arbitrary. Let* $\underline{\underline{\sigma}}_h$ *be a stress reconstruction verifying Definition A.1. Then*

$$\|\underline{u} - \underline{u}_h\|_{\mathrm{en}} \leqslant \mu^{-1/2} \left(\sum_{T \in \mathcal{T}_h} \left(\frac{h_T}{\pi} \|\underline{f} + \underline{\nabla} \cdot \underline{\underline{\sigma}}_h\|_T + \|\underline{\underline{\sigma}}_h - \underline{\underline{\sigma}}(\underline{u}_h)\|_T \right)^2 \right)^{1/2}. \tag{A.6}$$

Proof. From (A.4) and the symmetry of $\underline{\underline{\sigma}}(\underline{u}-\underline{u}_h)$, we infer that

$$\begin{aligned}\|\underline{u}-\underline{u}_h\|_{\mathrm{en}} &= \left(\underline{\underline{\sigma}}(\underline{u}-\underline{u}_h), \frac{\underline{\underline{\varepsilon}}(\underline{u}-\underline{u}_h)}{\|\underline{u}-\underline{u}_h\|_{\mathrm{en}}}\right) \leqslant \mu^{-1/2}\left(\underline{\underline{\sigma}}(\underline{u}-\underline{u}_h), \frac{\underline{\underline{\varepsilon}}(\underline{u}-\underline{u}_h)}{\|\underline{\underline{\nabla}}(\underline{u}-\underline{u}_h)\|}\right)\\ &\leqslant \mu^{-1/2}\sup_{\underline{v}\in\underline{H}^1_0(\Omega);\ \|\underline{\underline{\nabla}}\underline{v}\|=1}(\underline{\underline{\sigma}}(\underline{u}-\underline{u}_h),\underline{\underline{\nabla}}\underline{v}). \end{aligned} \tag{A.7}$$

Fix $\underline{v}\in\underline{H}^1_0(\Omega)$, such that $\|\underline{\underline{\nabla}}\underline{v}\|=1$. Using the fact that $\underline{u}$ verifies (A.2), and inserting $(\underline{\nabla}\cdot\underline{\underline{\sigma}}_h,\underline{v})+(\underline{\underline{\sigma}}_h,\underline{\underline{\nabla}}\underline{v})=0$ into the term inside the supremum yields

$$(\underline{\underline{\sigma}}(\underline{u}-\underline{u}_h),\underline{\underline{\nabla}}\underline{v}) = (\underline{f},\underline{v})-(\underline{\underline{\sigma}}(\underline{u}_h),\underline{\underline{\nabla}}\underline{v}) = (\underline{f}+\underline{\nabla}\cdot\underline{\underline{\sigma}}_h,\underline{v})+(\underline{\underline{\sigma}}_h-\underline{\underline{\sigma}}(\underline{u}_h),\underline{\underline{\nabla}}\underline{v}). \tag{A.8}$$

For the first term in the right hand side of (A.8) we use (A.5) to insert the mean value $\Pi^0_T\underline{v}$ of $\underline{v}$ on T, the Cauchy–Schwarz inequality, and the Poincaré inequality $\|\underline{v}-\Pi^0_T\underline{v}\|_T\leqslant\frac{h_T}{\pi}\|\underline{\underline{\nabla}}\underline{v}\|_T$ on simplexes $T\in\mathcal{T}_h$, and obtain

$$\left|(\underline{f}+\underline{\nabla}\cdot\underline{\underline{\sigma}}_h,\underline{v})\right| \leqslant \left|\sum_{T\in\mathcal{T}_h}(\underline{f}+\underline{\nabla}\cdot\underline{\underline{\sigma}}_h,\underline{v}-\Pi^0_T\underline{v})_T\right| \leqslant \sum_{T\in\mathcal{T}_h}\frac{h_T}{\pi}\|\underline{f}+\underline{\nabla}\cdot\underline{\underline{\sigma}}_h\|_T\|\underline{\underline{\nabla}}\underline{v}\|_T,$$

whereas the Cauchy–Schwarz inequality applied to the second term directly yields

$$\left|(\underline{\underline{\sigma}}_h-\underline{\underline{\sigma}}(\underline{u}_h),\underline{\underline{\nabla}}\underline{v})\right| \leqslant \sum_{T\in\mathcal{T}_h}\|\underline{\underline{\sigma}}_h-\underline{\underline{\sigma}}(\underline{u}_h)\|_T\|\underline{\underline{\nabla}}\underline{v}\|_T.$$

Inserting these results in (A.7) and again applying the Cauchy-Schwarz inequality yields the result. □

A.4 Stress Tensor Reconstructions

The set of vertices of the mesh $\mathcal{T}_h$ is denoted by $\mathcal{V}_h$; it is decomposed into interior vertices $\mathcal{V}_h^{\mathrm{int}}$ and boundary vertices $\mathcal{V}_h^{\mathrm{ext}}$. For all $a\in\mathcal{V}_h$, $\mathcal{T}_a$ is the patch of elements sharing the vertex a, ω_a the corresponding open subdomain in Ω, $\underline{n}_{\omega_a}$ its unit outward normal vector, and ψ_a the piecewise affine "hat" function which takes the value 1 at the vertex a and zero at all the other vertices. For all $T\in\mathcal{T}_h$, $\mathcal{V}_T$ denotes the set of vertices of T and h_T its diameter.

From now on, $\underline{u}_h$ denotes the solution of (A.3). The goal is to minimize the error estimate (A.6) avoiding global computations. As a result, both of the proposed reconstructions are based on local minimization problems on the patches ω_a :

$$\underline{\underline{\sigma}}_h^a := \arg\min_{\underline{\underline{\tau}}_h\in\underline{\underline{\Sigma}}_h^a;\ \underline{\nabla}\cdot\underline{\underline{\tau}}_h=\psi_a\underline{f}}\|\underline{\underline{\tau}}_h-\psi_a\underline{\underline{\sigma}}(\underline{u}_h)\|_{\omega_a}, \tag{A.9}$$

where we define $\underline{\underline{\Sigma}}_h^a$ separately for each construction and add a weak symmetry constraint in the second (AFW) construction. The global reconstructed stress tensor $\underline{\underline{\sigma}}_h$ is then obtained assembling the local solutions $\underline{\underline{\sigma}}_h^a$.

A.4.1 Arnold–Winther Stress Reconstruction

For each element $T \in \mathcal{T}_h$, the local AW spaces of degree $k \geqslant 1$ are defined by [10]

$$\underline{\underline{S}}_T^{\mathrm{AW}} := \{\underline{\underline{\tau}} \in \underline{\underline{\mathbb{P}}}_{\mathrm{sym}}^{k+2}(T) \mid \underline{\nabla} \cdot \underline{\underline{\tau}} \in [\underline{\mathbb{P}}^k(T)\}, \qquad \underline{V}_T^{\mathrm{AW}} := \underline{\mathbb{P}}^k(T),$$

where $\underline{\underline{\mathbb{P}}}_{\mathrm{sym}}^k(T)$ denotes the subspace of $\underline{\underline{\mathbb{P}}}^k(T)$ composed of symmetric-valued tensors. Figure A.1 shows the corresponding 24 degrees of freedom for the symmetric stress tensor in the lowest-order case $k = 1$: the values of the three components at each vertex of the triangle, the values of the moments of degree zero and 1 of the normal components each edge, and the value of the moment of degree zero of each component on the triangle. On a patch ω_a, the AW mixed finite element spaces are defined as

$$\begin{aligned}
\underline{\underline{S}}_h^{\mathrm{AW}}(\omega_a) &:= \{\underline{\underline{\tau}}_h \in \underline{\underline{H}}(\mathrm{div}, \omega_a) \cap \underline{\underline{\mathbb{P}}}_{\mathrm{sym}}^{k+2}(T) \mid \underline{\underline{\tau}}_h|_T \in \underline{\underline{S}}_T^{\mathrm{AW}} \ \forall T \in \mathcal{T}_a\}, \\
\underline{V}_h^{\mathrm{AW}}(\omega_a) &:= \{\underline{v}_h \in [L^2(\omega_a)]^2 \mid \underline{v}_h|_T \in \underline{V}_T^{\mathrm{AW}} \ \forall T \in \mathcal{T}_a\}.
\end{aligned}$$

Let now $k := p - 1$. We need to consider subspaces where a zero normal component is enforced on the stress tensor. Since the boundary condition in the exact problem prescribes the displacement and not the normal stress, we distinguish the case whether a is an interior vertex or a boundary vertex. For $a \in \mathcal{V}_h^{\mathrm{int}}$, we set

$$\underline{\underline{\Sigma}}_h^a := \{\underline{\underline{\tau}}_h \in \underline{\underline{S}}_h^{\mathrm{AW}}(\omega_a) \mid \underline{\underline{\tau}}_h \underline{n}_{\omega_a} = \underline{0} \text{ on } \partial\omega_a,\ \underline{\underline{\tau}}_h(b) = \underline{\underline{0}}\ \forall b \in \mathcal{V}_{\omega_a}^{\mathrm{ext}}\}, \tag{A.12a}$$

$$\underline{V}_h^a := \{\underline{v}_h \in \underline{V}_h^{\mathrm{AW}}(\omega_a) \mid (\underline{v}_h, \underline{z})_{\omega_a} = 0\ \forall \underline{z} \in \underline{RM}\}, \tag{A.12b}$$

with $\mathcal{V}_{\omega_a}^{\mathrm{ext}} = \mathcal{V}_h \cap \partial\omega_a$, and for $a \in \mathcal{V}_h^{\mathrm{ext}}$, we set

$$\underline{\underline{\Sigma}}_h^a := \{\underline{\underline{\tau}}_h \in \underline{\underline{S}}_h^{\mathrm{AW}}(\omega_a) \mid \underline{\underline{\tau}}_h \underline{n}_{\omega_a} = \underline{0} \text{ on } \partial\omega_a \backslash \partial\Omega,\ \underline{\underline{\tau}}_h(b) = \underline{\underline{0}}\ \forall b \in \mathcal{V}_{\omega_a}^{\mathrm{ext}}\}, \tag{A.13a}$$

$$\underline{V}_h^a := \underline{V}_h^{\mathrm{AW}}(\omega_a), \tag{A.13b}$$

with $\mathcal{V}_{\omega_a}^{\mathrm{ext}} = \mathcal{V}_h \cap (\partial\omega_a \backslash \partial\Omega)$. As argued in [10], the nodal degrees of freedom on $\partial\omega_a$ are set to zero if the vertex separates two edges where the normal stress is zero.

Construction A.3 (AW stress reconstruction)**.** *Find $\underline{\underline{\sigma}}_h^a \in \underline{\underline{\Sigma}}_h^a$ and $\underline{r}_h^a \in \underline{V}_h^a$ such that for all $(\underline{\underline{\tau}}_h, \underline{v}_h) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a$,*

$$(\underline{\underline{\sigma}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} + (\underline{r}_h^a, \underline{\nabla} \cdot \underline{\underline{\tau}}_h)_{\omega_a} = (\psi_a \underline{\underline{\sigma}}(\underline{u}_h), \underline{\underline{\tau}}_h)_{\omega_a}, \tag{A.14a}$$

$$(\underline{\nabla} \cdot \underline{\underline{\sigma}}_h^a, \underline{v}_h)_{\omega_a} = (-\psi_a \underline{f} + \underline{\underline{\sigma}}(\underline{u}_h) \underline{\nabla} \psi_a, \underline{v}_h)_{\omega_a}. \tag{A.14b}$$

Then, extending $\underline{\underline{\sigma}}_h^a$ by zero outside ω_a, set $\underline{\underline{\sigma}}_h := \sum_{a \in \mathcal{V}_h} \underline{\underline{\sigma}}_h^a$.

Using the definitions (A.12) and (A.13), the formulation (A.14) is equivalent to (A.9). For interior vertices, the source term in (A.14a) has to verify the Neumann compatibility condition

$$(-\psi_a \underline{f} + \underline{\sigma}(\underline{u}_h)\underline{\nabla}\psi_a, \underline{z})_{\omega_a} = 0 \quad \forall \underline{z} \in \underline{RM}. \tag{A.15}$$

Taking $\psi_a \underline{z}$ as a test function in (A.3), we see that (A.15) holds.

A.4.2 Arnold–Falk–Winther Stress Reconstruction

For each element $T \in \mathcal{T}_h$, the local AFW mixed finite element spaces [9] of degree $k \geqslant 1$ hinge on the Brezzi–Douglas–Marini mixed finite element spaces [29] for each line of the stress tensor and are defined by

$$\underline{\underline{S}}_T^{\mathrm{AFW}} := \underline{\underline{\mathbb{P}}}^k(T), \quad \underline{V}_T^{\mathrm{AFW}} := \underline{\mathbb{P}}^{k-1}(T), \quad \underline{\underline{\Lambda}}_T := \{\underline{\underline{\mu}} \in \underline{\underline{\mathbb{P}}}^{k-1}(T) \mid \underline{\underline{\mu}} = -\underline{\underline{\mu}}^T\}.$$

On a patch ω_a the global space $\underline{\underline{S}}_h^{\mathrm{AFW}}(\omega_a)$ is the subspace of $\underline{\underline{H}}(\mathrm{div}, \omega_a)$ composed of functions belonging piecewise to $\underline{\underline{S}}_T^{\mathrm{AFW}}$. The spaces $\underline{V}_h^{\mathrm{AFW}}(\omega_a)$ and $\Lambda_h(\omega_a)$ consist of functions lying piecewise in $\underline{V}_T$ and Λ_T respectively, with no continuity conditions between two elements.

As for the previous construction, we define subspaces with zero normal components enforced on the stress tensor, and distinguish between interior and boundary vertices. Let $k := p$ and set

$$\begin{aligned}
\underline{\underline{\Sigma}}_h^a &:= \{\underline{\underline{\tau}}_h \in \underline{\underline{S}}_h^{\mathrm{AFW}}(\omega_a) \mid \underline{\underline{\tau}}_h \underline{n}_{\omega_a} = \underline{0} \text{ on } \partial\omega_a \quad \text{if } a \in \mathcal{V}_h^{\mathrm{int}}, \\
&\qquad\qquad \underline{\underline{\tau}}_h \underline{n}_{\omega_a} = \underline{0} \text{ on } \partial\omega_a \backslash \partial\Omega \quad \text{if } a \in \mathcal{V}_h^{\mathrm{ext}}\}, && \text{(A.17a)} \\
\underline{V}_h^a &:= \{\underline{v}_h \in \underline{V}_h^{\mathrm{AFW}}(\omega_a) \mid (\underline{v}_h, \underline{z})_{\omega_a} = 0 \ \forall \underline{z} \in \underline{RM} \text{ if } a \in \mathcal{V}_h^{\mathrm{int}}\}, && \text{(A.17b)} \\
\Lambda_h^a &:= \Lambda_h(\omega_a). && \text{(A.17c)}
\end{aligned}$$

Construction A.4 (AFW stress reconstruction)**.** *Find* $\underline{\underline{\sigma}}_h^a \in \underline{\underline{\Sigma}}_h^a$, $\underline{r}_h^a \in \underline{V}_h^a$ *and* $\lambda_h^a \in \underline{\underline{\Lambda}}_h^a$ *such that for all* $(\underline{\underline{\tau}}_h, \underline{v}_h, \underline{\underline{\mu}}_h) \in \underline{\underline{\Sigma}}_h^a \times \underline{V}_h^a \times \underline{\underline{\Lambda}}_h^a$,

$$\begin{aligned}
(\underline{\underline{\sigma}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} + (\underline{r}_h^a, \underline{\nabla} \cdot \underline{\underline{\tau}}_h)_{\omega_a} + (\underline{\underline{\lambda}}_h^a, \underline{\underline{\tau}}_h)_{\omega_a} &= (\psi_a \underline{\underline{\sigma}}(\underline{u}_h), \underline{\underline{\tau}}_h)_{\omega_a}, && \text{(A.18a)} \\
(\underline{\nabla} \cdot \underline{\underline{\sigma}}_h^a, \underline{v}_h)_{\omega_a} &= (-\psi_a \underline{f} + \underline{\sigma}(\underline{u}_h)\underline{\nabla}\psi_a, \underline{v}_h)_{\omega_a}, && \text{(A.18b)} \\
(\underline{\underline{\sigma}}_h^a, \underline{\underline{\mu}}_h)_{\omega_a} &= 0. && \text{(A.18c)}
\end{aligned}$$

Then, extending $\underline{\underline{\sigma}}_h^a$ *by zero outside* ω_a, *set* $\underline{\underline{\sigma}}_h := \sum_{a \in \mathcal{V}_h} \underline{\underline{\sigma}}_h^a$.

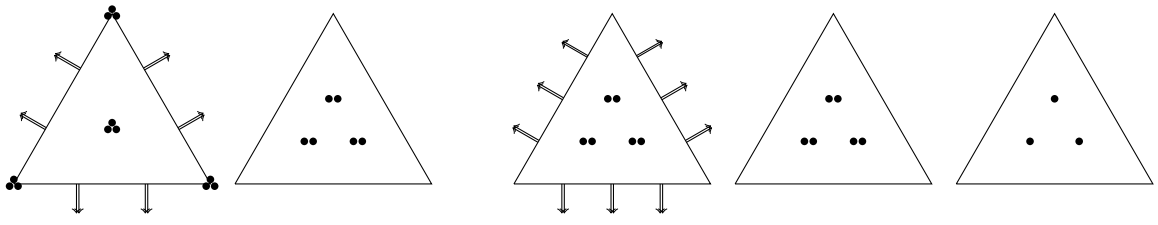


Figure A.1 – Element diagrams for $(\underline{\underline{S}}_T^{\mathrm{AW}}, \underline{V}_T^{\mathrm{AW}})$ with $k = 1$ (left) and $(\underline{\underline{S}}_T^{\mathrm{AFW}}, \underline{V}_T^{\mathrm{AFW}}, \Lambda_T)$ with $k = 2$ (right)

Using the definitions (A.17), the formulation (A.18) is equivalent to a modified version of (A.9), adding the weak symmetry constraint (A.18c). The condition (A.15) for all $a \in \mathcal{V}_h^{\text{int}}$ ensures that the constrained minimization problem (A.18) is well-posed.

A.4.3 Properties of the Stress Reconstructions

For both stress reconstructions we obtain the following result, recalling that we assume $\underline{f}$ to be piecewise polynomial of degree $p-1$.

Lemma A.5 (Properties of $\underline{\underline{\sigma}}_h$)**.** *Let $\underline{\underline{\sigma}}_h$ be prescribed by Construction A.3 or Construction A.4. Then $\underline{\underline{\sigma}}_h \in \underline{\underline{H}}(\mathrm{div}, \Omega)$, and for all $T \in \mathcal{T}_h$, the following holds:*

$$\underline{f} + \underline{\nabla} \cdot \underline{\underline{\sigma}}_h = \mathbf{0}. \tag{A.19}$$

Proof. All the fields $\underline{\underline{\sigma}}_h^a$ are in $\underline{\underline{H}}(\mathrm{div}, \omega_a)$ and satisfy appropriate zero normal conditions so that their zero-extension to Ω is in $\underline{\underline{H}}(\mathrm{div}, \Omega)$. Hence, $\underline{\underline{\sigma}}_h \in \underline{\underline{H}}(\mathrm{div}, \Omega)$. Let us prove (A.19). Since (A.15) holds for all $a \in \mathcal{V}_h^{\text{int}}$, we infer that (A.14b) or (A.18b) is actually true for all $\underline{v}_h \in \underline{V}_h(\omega_a)$. The same holds if $a \in \mathcal{V}_h^{\text{ext}}$ by definition of $\underline{V}_h^a$. Hence, $(\psi_a \underline{f} + \underline{\nabla} \cdot \underline{\underline{\sigma}}_h^a, \underline{v}_h)_{\omega_a} = 0$ for all $\underline{v}_h \in \underline{V}_h(\omega_a)$ and all $a \in \mathcal{V}_h$. Since $\underline{V}_h(\omega_a)$ is composed of piecewise polynomials that can be chosen independently in each cell $T \in \mathcal{T}_a$, and using $\underline{\underline{\sigma}}_h|_T = \sum_{a \in \mathcal{V}_T} \underline{\underline{\sigma}}_h^a|_T$ and the partition of unity $\sum_{a \in \mathcal{V}_T} \psi_a = 1$, we infer that $(\underline{f} + \underline{\nabla} \cdot \underline{\underline{\sigma}}_h, \underline{v}) = 0$ for all $\underline{v} \in \underline{V}_T$ and all $T \in \mathcal{T}_h$. The fact that $(\underline{f} + \underline{\nabla} \cdot \underline{\underline{\sigma}}_h)|_T \in \underline{V}_T$ for any $T \in \mathcal{T}_h$, concludes the proof. □

A.5 Numerical Results

We illustrate numerically our theoretical results on a test case with a known analytical solution. We analyze the convergence rates of the error estimates and compare them to those of the analytical error. The computations were performed using the `Code_Aster`[1] software. The exact solution $\underline{u} = (u_x, u_y)$ on the unit square $\Omega = (0,1)^2$ is given by

$$u_x = \frac{1}{\pi} \sin(\pi x) \cos(\pi y), \qquad u_y = -\frac{1}{\pi} \sin(\pi x) \cos(\pi y),$$

with the Lamé parameters $\mu = \lambda = 1$, and the corresponding body force $\underline{f}$. The exact solution is imposed as Dirichlet condition on the whole boundary $\partial\Omega$. The discretization is done on a series of unstructured grids with the polynomial degree $p = 2$ in the conforming finite element method (A.3). For each computation, two error estimates are calculated, one for each stress reconstruction. The AFW reconstruction offers some advantages over the AW one: it is cheaper (since by hybridization techniques we can avoid the resolution of saddle point problems), and the implementation for three-dimensional problems is easier (the lowest-order AW element in 3D has 162 degrees of freedom per element).

[1] `http://web-code-aster.org`

h^{-1}	estimate AFW		estimate AW		$\|\underline{u} - \underline{u}_h\|_{\text{en}}$		$I_{\text{eff,AFW}}$	$I_{\text{eff,AW}}$
4	1.707e-2	—	1.707e-2	—	1.704e-2	—	1.00	1.00
8	4.141e-3	2.05	4.124e-3	2.05	4.026e-3	2.08	1.03	1.02
16	1.175e-3	1.82	1.120e-3	1.88	1.116e-3	1.85	1.05	1.00
32	2.835e-4	2.05	2.736e-4	2.03	2.707e-4	2.04	1.05	1.01
64	7.384e-5	1.94	7.244e-5	1.92	7.021e-5	1.95	1.05	1.03

Table A.1 – Error estimators, analytical error, and effectivity indices under space refinement

Table A.1 shows the error estimates calculated using the stress reconstruction in the AFW (Const. A.4) and in the AW spaces (Const. A.3), the analytical error in the energy norm, as well as their convergence rates. The two columns on the right indicate the effectivity indices (overestimation factors) for both reconstruction methods, calculated as the ratio of the estimate to the analytical error. Since we chose $p = 2$, the convergence rates are close to 2, with the rates of the estimates reproducing very closely the ones of the actual error. Furthermore, the effectivity indices close to 1 indicate the reliability of the estimates.

Bibliography

[1] M. Ainsworth. A synthesis of a posteriori error estimation techniques for conforming, non-conforming and discontinuous galerkin finite element methods. *Contemp. Math.*, 383:23–50, 2005.

[2] M. Ainsworth and J. T. Oden. A unified approach to a posteriori error estimation using element residual methods. *Numer. Math.*, 65:23–50, 1993.

[3] M. Ainsworth and J. T. Oden. A posteriori error estimation in finite element analysis. *Pure and Applied Mathematics (New York), Wiley-Interscience [John Wiley & Sons], New York*, 2000.

[4] M. Ainsworth and R. Rankin. Guaranteed computable error bounds for conforming and nonconforming finite element analysis in planar elasticity. *Internat. J. Numer. Methods Engrg*, 82:1114–1157, 2010.

[5] M. Ainsworth and R. Rankin. Realistic computable error bounds for three dimensional finite element analyses in linear elasticity. *Comput. Methods Appl. Mech. Engrg.*, 200(21–22):1909–1926, 2011.

[6] T. Arbogast and Z. Chen. On the implementation of mixed methods as nonconforming methods for second-order elliptic problems. *Math. Comp.*, 64(211):943–972, 1995.

[7] D. N. Arnold, G. Awanou, and W. Qiu. Mixed finite elements for elasticity on quadrilateral meshes. *Adv. Comput. Math.*, 41:553–572, 2015.

[8] D. N. Arnold, G. Awanou, and R. Winther. Finite elements for symmetric tensors in three dimensions. *Math. Comp.*, 77:1229–1251, 2008.

[9] D. N. Arnold, R. S. Falk, and R. Winther. Mixed finite element methods for linear elasticity with weakly imposed symmetry. *Math. Comput.*, 76:1699–1723, 2007.

[10] D. N. Arnold and R. Winther. Mixed finite elements for elasticity. *Numer. Math.*, 92:401–419, 2002.

[11] G. Awanou. Rectangular mixed elements for elasticity with weakly imposed symmetry condition. *Adv. Comput. Math.*, 38:351–367, 2013.

[12] I. Babuška and A. Miller. A feedback finite element method with a posteriori error estimation, I: The finite element method and some basic properties of the a posteriori error estimator. *Comput. Methods Appl. Mech. Eng.*, 61:1–40, 1987.

[13] I. Babuška and W. C. Rheinboldt. A posteriori error estimates for the finite element method. *SIAM J. Numer. Anal*, 15:736–754, 1978.

[14] R. E. Bank and A. Weiser. Some a posteriori error estimators for elliptic partial differential equations. *Math. Comp.*, 44:283–301, 1985.

[15] M. Bebendorf. A note on the Poincaré inequality for convex domains. *Z. Anal. Anwendungen*, 22:751–756, 2003.

[16] R. Becker and R. Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numer.*, 10:1–102, 2001.

[17] J. Besson, G. Cailletaud, J. L. Chaboche, and S. Forest. *Mécanique non linéaire des matériaux*. HERMES Science Europe Ltd, 2001.

[18] M. Bieterman and I. Babuška. The finite element method for parabolic equations. *Numer. Math.*, 40(3):373–406, 1982.

[19] M. A. Biot. General theory of three-dimensional consolidation. *J. Appl. Phys.*, 12:155–169, 1941.

[20] D. Boffi, M. Botti, and D. A. Di Pietro. A nonconforming high-order method for the Biot problem on general meshes, 2016.

[21] D. Boffi, F. Brezzi, and M. Fortin. Reduced symmetry elements in linear elasticity. *Commun. Pur. Appl. Anal.*, 8:95–121, 2009.

[22] D. Boffi, F. Brezzi, and M. Fortin. *Mixed Finite Element Methods and Applications*, volume 44 of *Computational Mathematics*. Springer, 2013.

[23] E. Bonnetier. *Mathematical Treatment of the Uncertainties Appearing in the Formulation of Some Models of Plasticity*. PhD thesis, The University of Maryland at College Park, 1988.

[24] B. Boroomand and O. C. Zienkiewicz. Recovery procedures in error estimation and adaptivity. II. Adaptivity in nonlinear problems of elasto-plasticity behaviour. *Comput. Methods. Appl. Mech. Engrg.*, 176:127–146, 1999.

[25] M. Botti, D. A. Di Pietro, and P. Sochala. A hybrid high-order method for nonlinear elasticity. arXiv:1707.02154, submitted for publication, 2017.

[26] R. Boussetta and L. Fourment. A posteriori error estimation and three-dimensional adaptive remeshing: application to error control of non-steady metal forming simulations. In *International conference on numerical methods in industrial forming processes*. Ohio State, NUMIFORM 04, 1999.

[27] D. Braess, V. Pillwein, and J. Schöberl. Equilibrated residual error estimates are p-robust. *Comput. Methods Appl. Mech. Engrg.*, 198:1189–1197, 2009.

[28] D. Braess and J. Schöberl. Equilibrated residual error estimator for edge elements. *Math. Comp.*, 77(262):651–672, 2008.

[29] F. Brezzi, J. Douglas, and L. D. Marini. Recent results on mixed finite element methods for second order elliptic problems. In Dorodnitsyn Balakrishnan and Lions Eds., editors, *Vistas in applied mathematics. Numerical analysis, atmospheric sciences, immunology*, pages 25–43. Optimization Software Inc., Publications Division, New York, 1986.

[30] U. Brink and E. Stein. A posteriori error estimaion in large-strain elasticity using equilibrated local Neumann problems. *Comput. Methods Appl. Mech. Eng.*, 161:77–101, 1998.

[31] C. Carstensen. A unifying theory of a posteriori finite element error control. *Numer. Math.*, 100(5):163–175, 2005.

[32] C. Carstensen, M. Eigel, R. H. W. Hoppe, and C. Löbhard. A review of unified a posteriori finite element error control. *Numer. Math. Theory Methods Appl.*, 5(4):509–558, 2012.

[33] M. Cermak, F. Hecht, Z. Tang, and M. Vohralik. Adaptive inexact iterative algorithms based on polynomial-degree-robust a posteriori estimates for the Stokes problem. HAL Preprint 01097662, submitted for publication, 2017.

[34] M. Cervera, M. Chiumenti, and R. Codina. Mixed stabilized finite element methods in nonlinear solid mechanics: Part II: Strain localization. *Comput. Methods in Appl. Mech. and Engrg.*, 199(37–40):2571–2589, 2010.

[35] A. L. Chaillou and M. Suri. Computable error estimators for the approximation of nonlinear problems by linearized models. *Comput. Meth. Appl. Mech. Engrg.*, 196:210–224, 2006.

[36] L. Chamoin, P. Ladevèze, and F. Pled. An enhanced method with local energy minimization for the robust a posteriori construction of equilibrated stress field in finite element analysis. *Comput. Mech.*, 49:357–378, 2012.

[37] O. Coussy. *Poromechanics*. John Wiley and Sons, 2004.

[38] P. Destuynder and B. Métivet. Explicit error bounds in a conforming finite element method. *Math. Comput.*, 68(228):1379–1396, 1999.

[39] D. A. Di Pietro and A. Ern. A hybrid high-order locking-free method for linear elasticity on general meshes. *Comput. Meth. Appl. Mech. Engrg*, 283:1–21, 2015.

[40] D. A. Di Pietro, A. Ern, and J.-L. Guermond. Discontinuous Galerkin methods for anisotropic semi-definite diffusion with advection. *SIAM J. Numer. Anal.*, 46(2):805–831, 2008.

[41] D. A. Di Pietro, E. Flauraud, M. Vohralík, and S. Yousef. A posteriori error estimates, stopping criteria, and adaptivity for multiphase compositional Darcy flows in porous media. *J. Comput. Phys.*, 276:163–187, 2014.

[42] D. A. Di Pietro, M. Vohralík, and S. Yousef. An posteriori-based, fully adaptive algorithm for thermal multiphase compositional flows in porous media with adaptive mesh refinement. *Comput. and Math. with Appl.*, 68(12):2331–2347, 2014.

[43] D. A. Di Pietro, M. Vohralík, and S. Yousef. Adaptive regularization, linearization, and discretization and a posteriori error control for the two-phase Stefan problem. *Math. Comp*, 84(291):153–186, 2015.

[44] Code_Aster document [R5.03.01]. Algorithme non linéaire quasi-statique. `www.code-aster.org/doc/default/en/man_r/r5/r5.03.01.pdf`.

[45] Code_Aster document [R7.01.16]. Intégration des comportements mécaniques élasto-plastiques de Drucker–Prager, associé(druck_prager) et non-associé (druck_prag_n_a) et post-traitements. `www.code-aster.org/doc/default/en/man_r/r7/r7.01.16.pdf`.

[46] Code_Aster document [R7.01.24]. Loi de comportement viscoplastique LETK. `www.code-aster.org/doc/default/en/man_r/r7/r7.01.24.pdf`.

[47] V. Dolejší, A. Ern, and M. Vohralík. *hp*–adaption driven by polynomial-degree-robust a posteriori error estimates for elliptic problems. *SIAM J. Sci. Comput.*, 38(5):A3220–A3246, 2016.

[48] P. Dörsek and J. Melenk. Symmetry-free, p-robust equilibrated error indication for the *hp*-version of the FEM in nearly incompressible linear elasticity. *Comput. Methods Appl. Math.*, 13:291–304, 2013.

[49] D.C. Drucker and W. Prager. Soil mechanics and plastic analysis or limit design. *Quart. Appl. Math.*, 10(2):157–165, 1952.

[50] L. El Alaoui, A. Ern, and M. Vohralík. Guaranteed and robust a posteriori error estimates and balancing discretization and linearization errors for monotone nonlinear problems. *Comput. Methods Appl. Mech. Engrg.*, 200:2782–2795, 2011.

[51] K. Eriksson and C. Johnson. An adaptive finite element method for linear elliptic problems. *Math. Comput.*, 50:361–383, 1988.

[52] A. Ern and S. Meunier. A posteriori error analysis of Euler-Galerkin approximations to coupled elliptic-parabolic problems. *ESAIM Math. Mod. Numer. Anal.*., 43:353–375, 2009.

[53] A. Ern, I. Smears, and M. Vohralík. Equilibrated flux a posteriori error estimates in $L^2(H^1)$-norms for high-order discretizations of parabolic problems. preprint hal-01489721, 2017.

[54] A. Ern and M. Vohralík. A posteriori error estimation based on potential and flux reconstruction for the heat equation. *SIAM J. Numer. Anal.*, 48(1):198–223, 2010.

[55] A. Ern and M. Vohralík. Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs. *SIAM J. Sci. Comput.*, 35(4):A1761–A1791, 2013.

[56] A. Ern and M. Vohralík. Polynomial-degree-robust a posteriori estimates in a unified setting for conforming, nonconforming, discontinuous Galerkin, and mixed discretizations. *SIAM J. Numer. Anal.*, 53(2):1058–1081, 2015.

[57] A. Ern and M. Vohralík. Broken stable H^1 and $H(\mathrm{div})$ polynomial extensions for polynomial-degree-robust potential and flux reconstruction in three space dimensions. preprint hal-01422204, 2017.

[58] R. Fernandes. *Modélisation numérique objective des problèmes couplés hydromécaniques dans le cas des géomatériaux*. PhD thesis, Université Joseph Fourier - Grenoble I, 2009.

[59] B. Fraeijs de Veubeke. *Displacement and equilibrium models in the finite element method*. John Wiley and Sons, New York, 1965.

[60] G. N. Gatica, A. Marquez, and W. Rudolph. A priori and a posteriori error analyses of augmented twofold saddle point formulations for nonlinear elasticity problems. *Comput. Meth. Appl. Mech. Engrg.*, 264(1):23–48, 2013.

[61] G. N. Gatica and E. P. Stephan. A mixed-FEM formulation for nonlinear incompressible elasticity in the plane. *Numer. Methods Partial Differ. Equ.*, 18(1):105–128, 2002.

[62] A. Hannukainen, R. Stenberg, and M. Vohralík. A unified framework for a posteriori error estimation for the Stokes problem. *Numer. Math.*, 122:725–769, 2012.

[63] E. Hoek and E.T. Brown. Empirical strength criterion for rock masses. *J. Geotech. Engng. Div.*, 106:1013–1035, 1980.

[64] E. Hoek and E.T. Brown. *Underground excavations in rock*. London: Instn. Min. Metall., 1980.

[65] E. Hoek and E.T. Brown. The Hoek–Brown failure criterion - a 1988 update. *Proc. 15th Canadian Rock Mech. Symp.*, pages 31–38, 1988.

[66] K.-Y. Kim. Guaranteed a posteriori error estimator for mixed finite element methods of linear elasticity with weak stress symmetry. *SIAM J. Numer. Anal.*, 48:2364–2385, 2011.

[67] K.-Y. Kim. A posteriori error estimator for linear elasticity based on nonsymmetric stress tensor approximation. *J. Korean Soc. Ind. Appl. Math.*, 16(1):1–13, 2012.

[68] A. Kleine. *Modélisation numérique du comportement des ouvrages souterrains par une approche viscoplastique*. PhD thesis, Université de Lorraine, 2007.

[69] P. Ladevèze. *Comparaison de modèles de milieux continus*. PhD thesis, Université Pierre et Marie Curie (Paris 6), 1975.

[70] P. Ladevèze. Nouvelles procédures d'estimation d'erreur relative à la méthode des éléments finis et applications. *Journées éléments finis*, 1977.

[71] P. Ladevèze and D. Leguillon. Error estimate procedure in the finite element method and applications. *SIAM J. Numer. Anal.*, 20:485–509, 1983.

[72] P. Ladevèze and N. Moës. A new a posteriori error estimation for nonlinear time-dependent finite element analysis. *Comput. Meth. Appl. Mech. Engrg.*, 157(1–2):45–68, 1998.

[73] P. Ladevèze, J. P. Pelle, and P. Rougeot. Error estimation and mesh optimization for classical finite elements. *Engrg. Comp.*, 8(1):69–80, 1991.

[74] A. F. D. Loula and J. N. C. Guerreiro. Finite element analysis of nonlinear creeping flows. *Comput. Meth. Appl. Mech. Engrg.*, 79(1):87–109, 1990.

[75] R. Luce and B. I. Wohlmuth. A local a posteriori error estimator based on equilibrated fluxes. *SIAM J. Numer. Anal.*, 42:1394–1414, 2004.

[76] S. Meunier. *Analyse d'erreur a posteriori pour les couplages hydro-Mécaniques et mise en œuvre dans Code_Aster*. PhD thesis, École des Ponts ParisTech, 2007.

[77] M. A. Murad and A. F. D. Loula. Improved accuracy in finite element analysis of Biot's consolidation problem. *Comput. Meth. Appl. Mech. Engrg.*, 95:359–382, 1992.

[78] M. A. Murad and A. F. D. Loula. On stability and convergence of finite element analysis of Biot's consolidation problem. *Internat. J. Numer. Methods Engrg.*, 37:645–667, 1994.

[79] M. A. Murad, V. Thomée, and A. F. D. Loula. Asymptotic behaviour of semidiscrete finite-element approximations of Biot's consolidation problem. *SIAM J. Numer. Anal.*, 33(3):1065–1083, 1996.

[80] J. Nečas. *Introduction to the theory of nonlinear elliptic equations.* A Wiley-Interscience Publication. John Wiley & Sons Ltd., Chichester, 1986. Reprint of the 1983 edition.

[81] P. Neittaanmaäki and S. Repin. *Reliable methods for computer simulation: error control and a posteriori error estimates*, volume 33 of *Studies in Mathematics and Its Applications.* Elsevier, 2004.

[82] S. Nicaise, K. Witowski, and B. Wohlmuth. An a posteriori error estimator for the Lamé equation based on H(div)-conforming stress approximations. *IMA J. Numer. Anal.*, 28:331–353, 2008.

[83] J. T. Oden and S. Prudhomme. Goal-oriented error estimation and adaptivity for the finite element method. *Comput. Math. Appl.*, 41(5-6):735–756, 2001.

[84] S. Ohnimus, E. Stein, and E. Walhorn. Local error estimates of FEM for displacements and stresses in linear elasticity by solving local Neumann problems. *Int. J. Numer. Meth. Engng.*, 52:727–746, 2001.

[85] P. J. Phillips and M. J. Wheeler. A coupling of mixed and continous Galerkin finite element methods for poroelasticity II: the discrete-in-time case. *Comput Geosci*, 11:145–158, 2007.

[86] D. A. Di Pietro and J. Droniou. A Hybrid High-Order method for Leray–Lions elliptic equations on general meshes. *Math. Comp.*, 86(307):2159–2191, 2017.

[87] D. A. Di Pietro and J. Droniou. $W^{s,p}$-approximation properties of elliptic projectors on polynomial spaces, with application to the error analysis of a Hybrid High-Order discretisation of Leray–Lions problems. *Math. Models Methods Appl. Sci.*, 27(5):879–908, 2017.

[88] M. Pitteri and G. Zanotto. *Continuum Models for Phase Tansitions and Twinning in Crystals.* Chapman & Hall/CRC, 2002.

[89] W. Prager and J. L. Synge. Approximations in elasticity based on the concept of function space. *Quart. Appl. Math.*, 5:241–269, 1947.

[90] R. Rannacher and F. T. Suttmeier. A posteriori error estimation and mesh adaptation for finite element models in elasto-plasticity. *Comput. Methods Appl. Mech. Engrg.*, 176:333–361, 1999.

[91] S. Raude. *Prise en compte des sollicitations thermiques sur les comportements instantané et différé des géomatériaux.* PhD thesis, Université de Lorraine, 2015.

[92] P. A. Raviart and J. M. Thomas. *A mixed finite element method for second order elliptic problems*, volume 606 of *Lecture Notes in Math.* Springer, 1975.

[93] S. I. Repin. *A posteriori estimates for partial differential equations*, volume 4 of *Radon Series on Computational and Applied Mathematics.* Walter de Gruyter GmbH & Co. KG, Berlin, 2008.

[94] R. Riedlbeck, D. A. Di Pietro, and A. Ern. Equilibrated stress reconstructions for linear elasticity problems with application to a posteriori error analysis. In *Finite Volumes for Complex Applications VIII – Methods and Theoretical Aspects*, pages 293–301, 2017.

[95] R. Riedlbeck, D. A. Di Pietro, A. Ern, S. Granet, and K. Kazymyrenko. Stress and flux reconstruction in Biot's poro-elasticity problem with application to a posteriori analysis. *Comput. and Math. with Appl.*, 73(7):1593–1610, 2017.

[96] R. S. Sandhu and E. L. Wilson. Finite element analysis of seepage in elastic media. *J. Engrg. Mech. Div. Amer. Soc. Civil. Engrg.*, 95:641–652, 1969.

[97] D. Sandri. Sur l'approximation numérique des écoulements quasi-Newtoniens dont la viscosité suit la loi puissance ou la loi de Carreau. *Math. Modelling and Num. Anal.*, 27(2):131–155, 1993.

[98] R. E. Showalter. Diffusion in poro-elastic media. *J. Math. Anal. Appl.*, 251:310–340, 2000.

[99] E. Stein and S. Ohnimus. Anisotropic discretization- and model-error estimation in solid mechanics by local Neumann problems. *Comput. Meth. Appl. Mech. Engrg.*, 176:363–385, 1999.

[100] E. Stein, M. Rüter, and S. Ohnimus. Adaptive finite element analysis and modelling of solids and structures. Findings, problems and trends. *Int. J. Numer. Methods Engrg.*, 60:103–138, 2004.

[101] E. Stein, M. Rüter, and S. Ohnimus. Error-controlled adaptive goal-oriented modeling and finite element approximations in elasticity. *Comput. Meth. Appl. Mech. Engrg.*, 196:3598–3613, 2007.

[102] C. Taylor and P. Hood. A numerical solution of the Navier-Stokes equations using the finite element technique. *Comput. & Fluids*, 1:73–100, 1973.

[103] L. R. G. Treloar. *The Physics of Rubber Elasticity.* Oxford University Press, USA, 1975.

[104] R. Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques.* 1996.

[105] R. Verfürth. A review of a posteriori error estimation techniques for elasticity problems. *Comput. Meth. Appl. Mech. Engrg.*, 176:419–440, 1999.

[106] R. Verfürth. A posteriori error estimates for finite element discretizations of the heat equation. *Calcolo*, 40(3):195–212, 2003.

[107] M. Vogelius. An analysis of the p-version of the finite element method for nearly incompressible materials. uniformly valid, optimal error estimates. *Numer. Math.*, 41:39–53, 1983.

[108] M. Vohralík. On the discrete Poincaré–Friedrichs inequalities for nonconforming approximations of the Sobolev space H^1. *Numer. Funct. Anal. Opim.*, 26(7–8):925–952, 2005.

[109] M. Vohralík. Unified primal formulation-based a priori and a posteriori error analysis of mixed finite element methods. *Math. Comp.*, 79:2001–2032, 2010.

[110] K. von Terzaghi. *Theoretical soil mechanics.* Wiley, New York, 1943.

[111] A. Ženíšek. The existence and uniqueness theorem in Biot's consolidation theory. *Aplikace Matematiky*, 29:194–211, 1984.

[112] M. Williams. On the stress distribution at the base of a stationary crack. *J. Appl. Mech.*, 24:109–114, 1957.

[113] Y. Yokoo, K. Yamagata, and H. Nagaoka. Finite element method applied to Biot's consolidation theory. *Soils and Foundations*, 11:29–46, 1971.

[114] S. Yousef. *A posteriori error estimates and adaptivity based on stopping criteria and adaptive mesh refinement for multiphase and thermal flows. Application to steam-assisted gravity drainage.* PhD thesis, Université Paris 6, 2013.

[115] O. C. Zienkiewicz, Y. C. Liu, and G. C. Huang. Error estimation and adaptivity in flow formulation for forming problems. *Internat. J. Numer. Methods Engrg.*, 25(1):23–42, 1988.

[116] O. C. Zienkiewicz and J. Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis. *Internat. J. Numer. Methods Engrg.*, 33(2):337–357, 1987.

[117] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and a posteriori error estimates. part 1: The recovery technique. *Internat. J. Numer. Methods Engrg.*, 33:1331–1364, 1992.

[118] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and a posteriori error estimates. part 2: Error estimates and adaptivity. *Internat. J. Numer. Methods Engrg.*, 33(7):1365–1382, 1992.