Interface problems for dam modeling

Ilaria Fontana

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

En Mathématique et Modélisation

École doctorale : Information, Structures, Systèmes

Unité de recherche : Institut Montpelliérain Alexander Grothendieck

Interface problems for dam modeling

Présentée par Ilaria FONTANA
Le 31 mars 2022

Sous la direction de Daniele DI PIETRO

Devant le jury composé de

Franz CHOULY, Professeur des universités, Université de Bourgogne	Rapporteur
Daniele DI PIETRO, Professeur des universités, Université de Montpellier	Directeur
Luca FORMAGGIA, Professeur des universités, Politecnico di Milano	Rapporteur
Kyrylo KAZYMYRENKO, Ingénieur, EDF R&D	Co-encadrant
Djimédo KONDO, Professeur des universités, Sorbonne Université	Président du jury
Stella KRELL, Maître de conférences, Université de Nice	Examinatrice
Martin VOHRALÍK, Directeur de recherche, INRIA Paris	Examinateur

UNIVERSITÉ DE MONTPELLIER

Remerciements

Au cours de ces trois dernières années de thèse, j'ai collaboré et eu le soutien de plusieurs personnes. Je voudrais dédier ce court espace à tous ceux sans qui je n'aurais pas atteint ce résultat.

I miei primi ringraziamenti vanno sicuramente al mio direttore di tesi Daniele Di Pietro. La tua guida, la tua fiducia, i tuoi consigli e il tuo supporto sono stati indispensabili durante il mio percorso di tesi. Nonostante la distanza e le restrizioni degli ultimi anni non ci abbiamo permesso di incontrarci spesso, ti sei sempre dimostrato disponibile nei miei confronti e sei riuscito a trasmettermi la tua passione per la ricerca.

Secondly, I am really thankful to my industrial supervisor Cyril Kazymyrenko for convincing me to start this experience, for your kindness, your listening and your support, and for the numerous fruitful discussions we had during these years. I have learned a lot about Computational Mechanics and your advice was invaluable.

Je souhaite remercier Franz Chouly et Luca Formaggia d'avoir accepté d'être rapporteurs de ma thèse. Merci également à Djimédo Kondo, Stella Krell et Martin Vohralík d'avoir fait partie de mon jury de thèse.

Je tiens ensuite à remercier le groupe du CIH, en particulier Romain Tajetti, Patrick Divoux, Emmanuel Robbe, Charles Bodel, de m'avoir accueilli à Chambery et pour les échanges qu'on a fait pendant ces parcours. Merci également à toute l'équipe T66 (maintenant T6D) dans laquelle j'ai passé la majeure partie de ma thèse, merci à tous les collègues d'EDF, Thibaut A., Francesco B., Stéphan C., Jérémy D., Robin D., Nicolas D., Florian E., Astrid F., Laila F., Samuel G., Dominique G., Gilles G., Eilin G., David H., Samuel J., Matthieu L. C., Eric L., Tanguy M., Fannie M., pour m'avoir accompagné au cours de ces années. J'ai beaucoup apprécié la qualité du cadre scientifique et l'environnement agréable rencontrés. Je voudrais aussi à remercier Jean-Jacques Marigo, Nunziante Valoroso et Claude Stolz pour leur conseils sur mon travail.

À l'IMAG, je voudrais remercier en particulier André et Michele, toujours disponibles pour répondre à mes nombreuses questions, et, pour la partie administrative, Nathalie, Carmela et Brigitte, en particulier pour m'avoir aidé à organiser ma soutenance. J'ai aussi un remerciement particulier pour les autres doctorants d'EDF que j'ai croisé et qui m'ont donné de précieux conseils, Silun, Youbin, Goustan, Leysmir, Nicolas, Amine, Maxime.

Vorrei ringraziare le persone dell'Università di Udine che fin dall'inizio mi hanno sostenuto e guidato sul piano accademico. In particolare, i membri del CDLab, Rossana Vermiglio,

Dimitri Breda, Francesca Scarabel. Un ricordo va anche al dipartimento di matematica e alla Scuola Superiore dell'università.

Infine, voglio rivolgere un pensiero speciale ad Ariel, per essermi sempre stato accanto (anche da lontano) e per non aver mai smesso di incoraggiarmi e di consigliarmi. Voglio inoltre ringraziare i miei genitori, per il vostro costante supporto e affetto che mi avete dimostrato durante questa avventura che mi ha portato lontano da casa. Un pensiero va anche alla mia famiglia, a Germano, Elisa, Telemaco, ai miei amici, Silvia, Martina, Francesca, Emanuela, Klest, Lisa, Luca, per il loro continuo supporto.

Résumé

Les équipes d'ingénierie ont souvent recours aux simulations numériques par éléments finis pour étudier et analyser le comportement des ouvrages hydrauliques de grande dimension. Pour les ouvrages en béton, les modèles doivent être en mesure de prendre en compte le comportement non-linéaire des discontinuités aux diverses zones d'interfaces localisées en fondation, dans le corps du barrage ou à l'interface entre la structure et la fondation. Il faut non seulement être capable de représenter le comportement mécanique non-linéaire de ces interfaces (rupture, glissement, contact), mais également de prendre en compte l'écoulement hydraulique à travers ces ouvertures.

Dans le cadre de cette thèse, nous nous focalisons d'abord sur la question du comportement des interfaces, que nous abordons à travers le modèle des zones cohésives (CZM). Ce dernier, introduit dans divers codes de calcul par éléments finis (avec des *éléments finis de joint*), est une approche pertinente pour décrire la physique des problèmes de fissuration et de frottement au niveau de discontinuités géométriques. Bien que le CZM a été initialement introduit pour prendre en compte que le phénomène de rupture, nous montrons dans cette thèse que son utilisation peut être étendue aux problèmes de glissement en s'appuyant sur le formalisme élasto-plastique éventuellement couplé à l'endommagement. En outre, des lois de comportement hydromécaniques non-linéaires peuvent être introduites pour modéliser la notion d'ouverture de fissure et le couplage avec les lois d'écoulement fluide. Au niveau mécanique, nous travaillons dans le cadre des matériaux standard généralisés (SGM) [66], qui fournit une classe de modèles qui satisfont d'une manière automatique des principes de la thermodynamique tout en possédant des bonnes propriétés mathématiques utiles pour la modélisation numérique robuste. Nous adaptons le formalisme SGM volumique à la description des zones d'interface. Dans cette première partie de la thèse, nous présentons nos développements faites dans l'hypothèse de SGM adaptée aux CZM, capable de reproduire les phénomènes physiques observés expérimentalement : rupture, frottement, adhésion.

En pratique, les non-linéarités du comportement des zones d'interface sont dominées par la présence de contact, ce qui engendre des difficultés numériques importantes pour la convergence des calculs par élément fini. Le développement de méthodes numériques efficaces pour le problème de contact est donc une étape clé pour atteindre l'objectif de simulateurs numériques industriels robustes. Récemment, l'utilisation de techniques d'imposition faible des conditions de contact à la Nitsche a été proposée comme moyen pour réduire la complexité numérique [28]. Cette technique présente plusieurs avantages, dont les plus importants pour nos travaux sont : 1) possibilité de gérer une vaste gamme de conditions (glissement avec ou

sans frottement, non interpénétration, etc) ; 2) la technique se prête à une analyse d'erreur a posteriori rigoureuse. Ce schéma basé sur les conditions d'interface faibles répresente le point de départ pour l'estimation d'erreur a posteriori par reconstruction équilibrée de la contrainte. Cette analyse est utilisée pour estimer les différentes composantes d'erreur (p.e., spatiale, non-linéaire), et pour mettre en place un algorithme de résolution adaptatif, ainsi que des critères d'arrêt pour les solveurs itératifs et le réglage automatique d'éventuels paramètres numériques.

L'objectif principal de la thèse est donc de rendre robuste la simulation numérique par éléments finis des ouvrages présentant des discontinuités géométriques. On aborde cette question sous angle double : d'un côté on revisite les méthodes existantes de représentation de fissuration en travaillant sur la loi de comportement mécanique pour les joints ; de l'autre on introduit une nouvelle méthode a posteriori pour traiter le problème de contact et propose son adaptation pour les modèles d'interfaces génériques.

Mots clefs : Lois de comportement pour la modélisation des joint des barrages, couplage endommagement-plasticité, hyper-élasticité, méthode à la Nitsche pour les problèmes de contact, estimation d'erreur a posteriori par reconstruction de la contrainte, algorithmes adaptatifs

Abstract

Engineering teams often use finite element numerical simulations for the design, study and analysis of the behavior of large hydraulic structures. For concrete structures, models of increasing complexity must be able to take into account the nonlinear behavior of discontinuities at the various interfaces located in the foundation, in the body of the dam or at the interface between structure and foundation. Besides representing the nonlinear mechanical behavior of these interfaces (rupture, sliding, contact), one should also be able to take into account the hydraulic flow through these openings.

In this thesis, we first focus on the topic of interface behavior modeling, which we address through the Cohesive Zone Model (CZM). This model was introduced in various finite element codes (with the *joint elements*), and it is a relevant approach to describe the physics of cracking and friction problems at the geometrical discontinuities level. Although initially the CZM was introduced to take into account the phenomenon of rupture, we show in this thesis that it can be extended to sliding problems by possibly relying on the elasto-plastic formalism coupled to the damage. In addition, nonlinear hydro-mechanical constitutive relations can be introduced to model the notion of crack opening and the coupling with the laws of fluid flow. At the mechanical level, we work in the Standard Generalized Materials (SGM) framework [66], which provides a class of models automatically satisfying some thermodynamical principles, while having good mathematical and numerical properties that are useful for robust numerical modeling. We adapt the formalism of volumetric SGM to the interface zones description. In this first part of the thesis, we present our developments under the hypothesis of SGM adapted to CZM, capable of reproducing the physical phenomena observed experimentally: rupture, friction, adhesion.

In practice, nonlinearities of behavior of interface zones are dominated by the presence of contact, which generates significant numerical difficulties for the convergence of finite element computations. The development of efficient numerical methods for the contact problem is thus a key stage for achieving the goal of robust industrial numerical simulators. Recently, the weak enforcement of contact conditions à la Nitsche has been proposed as a mean to reduce numerical complexity [28]. This technique displays several advantages, among which the most important for our work are: 1) it can handle a wide range of conditions (slip with or without friction, no interpenetration, etc.); 2) it lends itself for a rigorous a posteriori error analysis. This scheme based on the weak contact conditions represents in this work the starting point for the a posteriori error estimation via equilibrated stress reconstruction. This analysis is then used to estimate the different error components (e.g., spatial, nonlinear), and

to develop an adaptive resolution algorithm, as well as stopping criteria for iterative solvers and the automatic tuning of possible numerical parameters.

The main goal of this thesis is thus to make the finite element numerical simulation of structures with geometrical discontinuities robust. We address this question from two angles: on one side, we revisit the existing methods for the crack representation working on the mechanical constitutive relation for joints; on the other, we introduce a new a posteriori method for the contact problem and we propose its adaptation for the generic interface models.

Key words: Constitutive relations for joint modeling in dams, damage-plasticity coupling, hyperelasticity, Nitsche's method for contact problems, a posteriori error estimate via stress reconstruction, adaptive algorithms

Contents

1 Introduction

Dans ce chapitre introductif, on présente le contexte industriel et académique de la thèse, et on décrit le problème mécanique considéré. Ensuite, on présente le plan du manuscrit en soulignant la contribution de la thèse.

Contents

1.1 Contexte industriel et motivations du travail

En France, le groupe Électricité de France (EDF) est le premier producteur et fournisseur d'électricité, et il doit assurer la consolidation du parc de production avec le maintien en état de ses installations. En particulier, depuis une douzaine d'année, le département ERMES (ElectRotechnique et Mécanique des Structures) d'EDF R&D travaille en étroite collaboration avec le Centre d'Ingénierie Hydraulique (CIH EDF) sur la modélisation mécanique et numérique des ouvrages hydrauliques en béton. L'évaluation de la sûreté de ces structures de grande dimension est un enjeu majeur pour EDF. Dans ce contexte, les ingénieurs d'EDF ont recours aux simulations numériques éléments finis avec le logiciel open-source `code_aster` (https://code-aster.org), actuellement développé par le département ERMES. `code_aster` est intégré dans la plateforme `salome_meca`, qui fournit une interface graphique ainsi que des outils pour la construction de géométries et maillages, et pour les post-traitements.

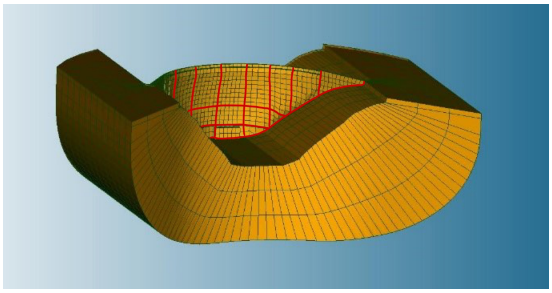


FIGURE 1.1 – Quelques zones d'interface (*en rouge*) dans un maillage d'un barrage voûte.

FIGURE 1.2 – Photos des barrages du Gleno (*gauche*) et de Malpasset (*droite*) après les incidents.

Cette thèse s'inscrit aussi dans le contexte du Groupe de Travail sur la caractérisation du contact, qui a été créé en 2010. Ce groupe réunit plusieurs divisions d'EDF : Division Production et Ingénierie Hydraulique, CIH, département Technique d'Essais en Géologie, Géotechnique et de la Génie civil, R&D PRISME, et R&D ERMES.

La complexité de la modélisation des ouvrages hydrauliques est le résultat de la combinaison des effets mécaniques, hydrauliques et thermiques sur leur comportement. Cependant, les non-linéarités de ces structures sont concentrées dans des zones bien localisées, ce qui rend les études industrielles accessibles, bien que toujours complexes. Ces régions coïncident avec les interfaces, qui sont les zones les plus faibles de la structure, comme le montrent les incidents et les auscultations sur les barrages. Il s'agit principalement :

- des interfaces entre le béton et la roche dans la fondation de la structure,
- des joints entre les blocs de béton,
- des reprises de bétonnage.

On peut voir certaines de ces zones d'interfaces en rouge dans la Figure 1.1, qui représente le maillage d'un barrage voûte. La Figure 1.2 montre les résultats de la rupture de deux barrages [119] : le barrage du Gleno (*gauche*), incident survenu en 1923 en Italie, et le barrage de Malpasset (*droite*), incident survenu en 1959 en France. Dans les images on peut bien observer les zones d'interface dans la région de rupture.

Actuellement, les ingénieurs d'EDF peuvent utiliser deux lois de comportement pour la modélisation numérique des zones d'interfaces avec le code éléments finis `code_aster` :

- `JOINT_MECA_RUPT` [49, Section 2], qui décrit la phase d'ouverture de la fissure et est basée sur une formulation cohésive de la rupture ;
- `JOINT_MECA_FROT` [49, Section 3], qui décrit la phase de glissement en cisaillement avec une version élastoplastique de la loi de frottement de Mohr–Coulomb.

Ces lois sont assez simples et ont été développées pour répondre aux besoins initiaux du CIH. En particulier, elles introduisent chacune un seul phénomène mécanique (endommagement et plasticité, respectivement) couplé avec le flux hydraulique à travers les ouvertures dans les zones d'interface. Toutefois, il a été constaté par le CIH qu'en considérant séparément les effets de la plasticité et de l'endommagement on ne peut pas retrouver dans son intégralité le comportement qu'on observe expérimentalement. En outre, ces lois ne garantissent pas les principes thermodynamiques (conservation de l'énergie élastique, dissipation liée à l'endommagement) et, pour chaque zone d'interface, l'utilisateur doit décider quel phénomène on veut modéliser, c'est-à-dire, si utiliser `JOINT_MECA_RUPT` ou `JOINT_MECA_FROT` dans le code, ce qui est un point bloquant pendant la modélisation numérique. Un premier volet de la thèse est donc dédié au développement de lois de comportement pour les zones d'interface à couplage hydromécanique, obtenues à partir de principes thermodynamiques et capables de reproduire les phénomènes physiques principaux observés expérimentalement (frottement, rupture, adhésion).

Un autre difficulté de la modélisation des barrages consiste à définir des critères d'arrêt efficaces pour la convergence des solveurs. En effet, pendant un calcul mécanique sur des grandes structures, souvent la convergence des solveurs est lente et des astuces numériques sont nécessaires pour aider cette convergence. Dans ce contexte, l'application d'une méthode d'estimation d'erreur a posteriori fournit une borne supérieure calculable de l'erreur locale et permet de distinguer les différentes composantes de l'erreur. En comparant ces composantes, on peut développer des algorithmes adaptatifs, avec adaptation local du maillage, critères d'arrêt pour les solveurs itératifs (par exemple, algébrique et non-linéaire), et réglage automatique de certains paramètres numériques. La deuxième contribution principale de la thèse est représentée par le développement d'une analyse d'erreur a posteriori par reconstruction équilibrée de la contrainte pour le problème de contact unilatéral, qui peut être considéré comme un premier pas vers la modélisation des structures avec zones d'interface. La partie finale, plus exploratoire, de ce travail de thèse établit les étapes principales pour l'extension de cette estimation a posteriori à des problèmes plus complexes avec forces cohésives.

1.2 Équations pour la modélisation des barrages

Dans la modélisation des barrages, on peut considérer plusieurs types de phénomènes : mécaniques, hydrauliques, thermiques. De plus, ces structures ont diverses zones d'interfaces présentant de fortes non-linéarités. Ces zones peuvent être représentées par une discontinuité rugueuse, remplie par des matériaux tels que, par exemple, de l'argile, des coulis, ou des roches, et montrent différents phénomènes importants : frottement, comportement élastique pour de très petits déplacements, perte de la force de traction, disparition progressive du pic de la contrainte de cisaillement avec un essai cyclique. Le but de cette section est de décrire certains modèles mathématiques pertinents pour ces types de structures, avec une attention particulière aux formulations discrètes qui permettent de résoudre en pratique le problème

d'équilibre mécanique.

1.2.1 Équilibre mécanique

On utilise la notation classique en mécanique : $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, est un domaine ouvert et connexe qui représente la région d'espace occupé par un milieu continu, $\boldsymbol{u} : \Omega \to \mathbb{R}^d$ est le déplacement du domaine, $\boldsymbol{\varepsilon}(\boldsymbol{u}) \coloneqq \frac{1}{2}(\boldsymbol{\nabla u} + (\boldsymbol{\nabla u})^T)$ est le tenseur des déformations, et $\boldsymbol{\sigma}$ est le tenseur des contraintes.

Dans le cadre des petites perturbations, les équations d'équilibre linéarisées s'écrivent

$$\mathbf{div}\,\boldsymbol{\sigma} + \boldsymbol{f} = \mathbf{0} \qquad \text{dans } \Omega, \qquad \text{(équilibre volumique)} \tag{1.1}$$

où **div** est l'opérateur de divergence et $\boldsymbol{f}$ représente le champ de forces volumiques. Le comportement du milieu continu est caractérisé par sa *loi de comportement*, qui établit une relation entre l'histoire de chargement et le tenseur de contrainte. Dans le cas particulier d'élasticité de Cauchy, on suppose que la contrainte ne dépend que du tenseur des déformations et d'autres éventuelles variables dissipatives (par exemple, la composante plastique des déformations pour les modèles élasto-plastiques). Nous adoptons ici une écriture formelle simplifiée (en règle générale, les autres variables traçant l'historique d'évolution doivent apparaître également) :

$$\boldsymbol{\sigma} = \boldsymbol{F}(\boldsymbol{\varepsilon}) \qquad \text{dans } \Omega. \qquad \text{(loi de comportement)} \tag{1.2}$$

Par exemple, dans le cas de comportement élastique linéaire isotrope homogène, on a

$$\boldsymbol{\sigma} = \lambda \mathrm{Tr}\,\boldsymbol{\varepsilon}\,\mathbf{I}_d + 2\mu\boldsymbol{\varepsilon} \tag{1.3}$$

où $\mathbf{I}_d$ est le tenseur identité d'ordre 2, λ et μ sont les coefficients de Lamé du matériau. Enfin, les conditions aux limites complètent le problème. Par exemple,

$$\boldsymbol{u} = \boldsymbol{u}_D \qquad \text{sur } \Gamma_D, \qquad \text{(conditions aux limites de Dirichlet)} \tag{1.4}$$

$$\boldsymbol{\sigma}\,\boldsymbol{n} = \boldsymbol{g}_N \qquad \text{sur } \Gamma_N, \qquad \text{(conditions aux limites de Neumann)} \tag{1.5}$$

où Γ_D et Γ_N sont le parties du bord où les déplacements et les forces sont imposés, respectivement, $\boldsymbol{u}_D : \Gamma_D \to \mathbb{R}^d$ est le déplacement imposé, $\boldsymbol{g}_N : \Gamma_N \to \mathbb{R}^d$ est la force surfacique imposée, et $\boldsymbol{n}$ et le vecteur normal unitaire sortant de Ω.

En supposant que les forces volumiques et surfaciques sont de carré intégrable sur leur domaine de définition, c'est-à-dire, $\boldsymbol{f} \in \boldsymbol{L}^2(\Omega) \coloneqq [L^2(\Omega)]^d$ et $\boldsymbol{g}_N \in \boldsymbol{L}^2(\Gamma_N) \coloneqq [L^2(\Gamma_N)]^d$, la formulation faible du problème d'équilibre mécanique (1.1)–(1.5) s'écrit : Trouver le déplacement $\boldsymbol{u} \in \boldsymbol{V}_D$ tel que

$$(\boldsymbol{\sigma}(\boldsymbol{\varepsilon}(\boldsymbol{u})), \boldsymbol{\varepsilon}(\boldsymbol{v})) = (\boldsymbol{f}, \boldsymbol{v}) + (\boldsymbol{g}_N, \boldsymbol{v})_{\Gamma_N} \qquad \forall \boldsymbol{v} \in \boldsymbol{V}_0, \tag{1.6}$$

où $(\,\cdot\,,\,\cdot\,)$ et $(\,\cdot\,,\,\cdot\,)_{\Gamma_N}$ sont les produits scalaires des espace $\boldsymbol{L}^2(\Omega)$ et $\boldsymbol{L}^2(\Gamma_N)$, respectivement, et les espaces $\boldsymbol{V}_D$ et $\boldsymbol{V}_0$ sont constitués par les fonctions de l'espace de Sobolev $\boldsymbol{H}^1(\Omega) \coloneqq [H^1(\Omega)]^d$ qui satisfont des conditions appropriées sur Γ_D :

$$\boldsymbol{V}_D \coloneqq \left\{\boldsymbol{v} \in \boldsymbol{H}^1(\Omega) \;:\; \boldsymbol{v} = \boldsymbol{u}_D \text{ sur } \Gamma_D\right\},$$
$$\boldsymbol{V}_0 \coloneqq \left\{\boldsymbol{v} \in \boldsymbol{H}^1(\Omega) \;:\; \boldsymbol{v} = \mathbf{0} \text{ sur } \Gamma_D\right\}.$$

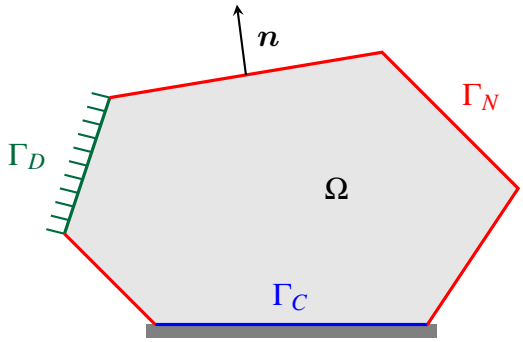


FIGURE 1.3 – Exemple de configuration où on impose des conditions aux limites de Dirichlet sur Γ_D (en *vert*), de Neumann sur Γ_N (en *rouge*), et de contact sur Γ_C (en *bleu*).

En particulier, dans le cas où $\boldsymbol{u}_D = \mathbf{0}$, on a simplement $\boldsymbol{V}_D = \boldsymbol{V}_0 =: \boldsymbol{V}$. Pour simplifier, tout au long de ce manuscrit on suppose $\boldsymbol{u}_D = \mathbf{0}$.

Pour résoudre numériquement le problème (1.6), on considère un maillage $\mathcal{T}_h$ qui discrétise le domaine Ω et un espace de dimension finie $\boldsymbol{V}_h$ (construit à partir de $\mathcal{T}_h$) dans lequel on cherche une approximation $\boldsymbol{u}_h$ de $\boldsymbol{u}$. La définition et les propriétés de cet espace dépendent de la méthode numérique utilisée. Vu que `code_aster` est un code de calcul basé sur les éléments finis de Lagrange, dans cette thèse on s'intéresse particulièrement à ce type de méthodes [110, 30, 18]. L'espace de dimension finie correspondant est

$$\boldsymbol{V}_h := \{\boldsymbol{v}_h \in \boldsymbol{V} \ : \ \boldsymbol{v}_h|_T \in \boldsymbol{\mathcal{P}}^p(T) \text{ pour tout élément } T \in \mathcal{T}_h\} \subset \boldsymbol{V},$$

où $\boldsymbol{\mathcal{P}}^p(T) := [\mathcal{P}^p(T)]^d$ et $\mathcal{P}^p(T)$ dénote la restriction à l'élément T de l'espace de polynômes de d variables de degré total $\leq p$. Le problème approchant (1.6) s'écrit alors : Trouver $\boldsymbol{u}_h \in \boldsymbol{V}_h$ tel que

$$(\boldsymbol{\sigma}(\boldsymbol{\varepsilon}(\boldsymbol{u}_h)), \boldsymbol{\varepsilon}(\boldsymbol{v}_h)) = (\boldsymbol{f}, \boldsymbol{v}_h) + (\boldsymbol{g}_N, \boldsymbol{v}_h)_{\Gamma_N} \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_h. \tag{1.7}$$

D'autres exemples de méthodes couramment utilisées en mécanique incluent : les méthodes de volumes finis [58, 84], les méthodes d'éléments finis mixtes [19, 14], les méthodes de Galerkine discontinues [40], et, plus récemment, les méthodes Hybrid High–Order [15, 39].

1.2.2 Problème de contact unilatéral

Le modèle de contact unilatéral est le plus simple permettant de représenter le contact entre deux objets. Tout d'abord, on considère le contact entre un objet avec comportement élastique et une surface rigide, puis on passe au contact entre deux objets élastiques. Dans ces modèles, on suppose qu'il n'y a pas de forces cohésives ou de friction entre les deux corps.

Pour le problème de contact unilatéral entre un objet élastique et une surface rigide, on utilise la notation de la Figure 1.3. En particulier, la zone de contact dans la configuration de référence est notée Γ_C, voir Figure 1.3. Sur le bord du domaine on peut écrire la décomposition (unique) en composante normale et tangente

$$\boldsymbol{v} = v_n \boldsymbol{n} + \boldsymbol{v}_t \qquad \text{et} \qquad \boldsymbol{\sigma}(\boldsymbol{v})\boldsymbol{n} = \sigma_n(\boldsymbol{v})\boldsymbol{n} + \boldsymbol{\sigma}_t(\boldsymbol{v}) \qquad \text{sur } \partial\Omega, \tag{1.8}$$

pour tout champ de déplacement $\boldsymbol{v}$ et pour toute densité de force surfaciques $\boldsymbol{\sigma}(\boldsymbol{v})\boldsymbol{n}$ définie sur $\partial\Omega$. On remarque que la composante tangente peut être identifiée avec une fonction définie

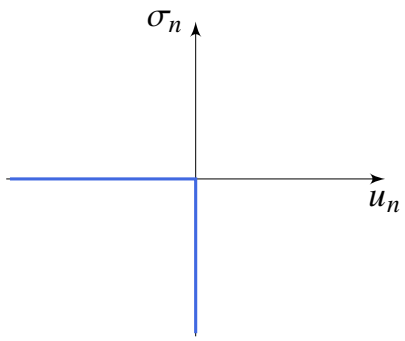


FIGURE 1.4 – Illustration de la condition de contact (1.9a).

sur $\partial\Omega$ à valeurs dans $\mathbb{R}^{d-1}$. Alors, les conditions de contact sans frottement, introduites par Signorini [118], s'écrivent :

$$u_n \leq 0,\ \sigma_n(\boldsymbol{u}) \leq 0,\ \sigma_n(\boldsymbol{u})u_n = 0 \qquad \text{sur } \Gamma_C, \tag{1.9a}$$

$$\boldsymbol{\sigma_t}(\boldsymbol{u}) = \boldsymbol{0} \qquad \text{sur } \Gamma_C. \tag{1.9b}$$

La première condition (1.9a), représenté graphiquement par la Figure 1.4, est une condition de complémentarité qui représente deux caractéristiques de la zone de contact Γ_C : l'interdiction d'interpénétration ($u_n \leq 0$) et l'absence de forces cohésives ($u_n < 0$ implique $\sigma_n(\boldsymbol{u}) = 0$). Le deuxième condition (1.9b) représente simplement l'absence de frottement.

Si on définit le cône des déplacements admissibles

$$\boldsymbol{K} \coloneqq \{\boldsymbol{v} \in \boldsymbol{V} \ :\ v_n \leq 0 \text{ on } \Gamma_C\} \subset \boldsymbol{V},$$

la formulation faible du problème (1.1) avec loi de comportement élastique (1.3) et avec conditions aux limites de Dirichlet (1.4) sur Γ_D (avec $\boldsymbol{u}_D = \boldsymbol{0}$), de Neumann (1.5) sur Γ_N, et de contact (1.9) sur Γ_C correspond à l'inégalité variationnelle suivante (voir [67, 26]) : Trouver le déplacement $\boldsymbol{u} \in \boldsymbol{K}$ tel que

$$(\boldsymbol{\sigma}(\boldsymbol{u}), \varepsilon(\boldsymbol{v} - \boldsymbol{u})) \geq (\boldsymbol{f}, \boldsymbol{v} - \boldsymbol{u}) + (\boldsymbol{g}_N, \boldsymbol{v} - \boldsymbol{u})_{\Gamma_N} \qquad \forall \boldsymbol{v} \in \boldsymbol{K}.$$

Discrétisation avec impositions faible à la Nitsche

Les méthodes numériques proposées dans la littérature pour approcher les problèmes de contact unilatéral comprennent les formulations de pénalisation, les formulations mixtes, et l'imposition faible des conditions à la Nitsche. Dans ce travail, on considère cette dernière approche, qui conduit à une formulation discrète consistante et coercive facile à implémenter, et qui ne demande pas l'introduction de multiplicateurs de Lagrange. Pour plus d'information sur ces trois méthodes, on se réfère aux articles de revue [127, 26].

L'imposition faible des conditions de contact à la Nitsche est basée sur la réécriture de la condition (1.9a) en une seule équation à travers un opérateur de projection. Cela permet d'obtenir une formulation discrète qui ne diffère de (1.7) que pour un terme défini sur la partie du bord Γ_C. Étant donné une fonction positive $\gamma\colon \Gamma_C \to \mathbb{R}^+$, la condition (1.9a) peut s'écrire (voir [28]) :

$$\sigma_n(\boldsymbol{u}) = [\sigma_n(\boldsymbol{u}) - \gamma u_n]_{\mathbb{R}^-} \qquad \text{sur } \Gamma_C,$$

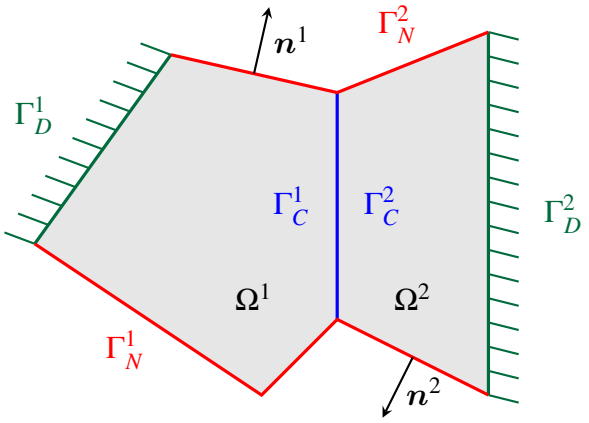


FIGURE 1.5 – Exemple de deux domaines Ω^1 et Ω^2 en contact sur la partie du bord de contact Γ_C^i, $i \in \{1,2\}$.

où $[\,\cdot\,]_{\mathbb{R}^-} : \mathbb{R} \to \mathbb{R}^-$ est la projection sur la demi-droite des réels négatifs. On considère donc la méthode suivante, originellement introduite dans [28], pour approcher le problème de contact : Trouver $\boldsymbol{u}_h \in \boldsymbol{V}_h$ tel que

$$(\boldsymbol{\sigma}(\boldsymbol{u}_h), \boldsymbol{\varepsilon}(\boldsymbol{v}_h)) - \left(\left[\sigma_n(\boldsymbol{u}_h) - \gamma u_{h,n} \right]_{\mathbb{R}^-}, v_{h,n} \right)_{\Gamma_C} = (\boldsymbol{f}, \boldsymbol{v}_h) + (\boldsymbol{g}_N, \boldsymbol{v}_h)_{\Gamma_N} \tag{1.10}$$

pour tout $\boldsymbol{v}_h \in \boldsymbol{V}_h$. On rappelle qu'ici $\sigma_n(\boldsymbol{u}_h)$ et $u_{h,n}$ indiquent la composant normal su Γ_C des champs $\boldsymbol{\sigma}(\boldsymbol{u}_h)$ et $\boldsymbol{u}_h$, respectivement, selon la décomposition (1.8). Cette méthode est consistante et bien posée (voir [26, Section 3]).

Le problème de contact entre deux corps élastiques

La méthode ci-dessus s'étend de façon naturelle aux problèmes de contact unilatéral entre deux corps élastiques. Dans ce contexte, on utilise la notation de Figure 1.5. En particulier, on note Ω^1 et Ω^2 les domaines qui représentent les deux corps élastiques. Par simplicité, dorénavant l'indice i indiquera une quantité définie sur chaque domaine indifféremment, c'est-à-dire $i \in \{1,2\}$. On suppose que, dans la configuration de référence, les domaines sont en contact sur Γ_C^i, avec une correspondance parfaite entre Γ_C^1 et Γ_C^2.

Pour la modélisation des conditions de contact, on utilise la stratégie maître/esclave, où les conditions unilatérales de contact sont imposées sur la surface esclave, Γ_C^1 dans notre cas. Le problème de contact entre deux corps élastiques (sans frottement) s'écrit alors :

$$\begin{aligned}
\mathbf{div}\, \boldsymbol{\sigma}^i(\boldsymbol{u}^i) + \boldsymbol{f}^i &= \mathbf{0} && \text{dans } \Omega^i, && (1.11a)\\
\boldsymbol{\sigma}^i(\boldsymbol{u}^i) &= \lambda^i \mathrm{Tr}\, \boldsymbol{\varepsilon}(\boldsymbol{u}^i)\, \mathbf{I}_d + 2\mu^i \boldsymbol{\varepsilon}(\boldsymbol{u}^i) && \text{dans } \Omega^i, && (1.11b)\\
\boldsymbol{u}^i &= \mathbf{0} && \text{sur } \Gamma_D^i, && (1.11c)\\
\boldsymbol{\sigma}^i(\boldsymbol{u}^i)\boldsymbol{n}^i &= \boldsymbol{g}_N^i && \text{sur } \Gamma_N^i, && (1.11d)\\
[\![u]\!]_n^1 \le 0,\ \sigma_n^1(\boldsymbol{u}^1) \le 0,\ \sigma_n^1(\boldsymbol{u}^1)[\![u]\!]_n^1 &= 0 && \text{sur } \Gamma_C^1, && (1.11e)\\
\boldsymbol{\sigma}_t^1(\boldsymbol{u}^1) &= \mathbf{0} && \text{sur } \Gamma_C^1, && (1.11f)
\end{aligned}$$

où

$$[\![u]\!]_n^1 := (\boldsymbol{u}^1 - \boldsymbol{u}^2)\boldsymbol{n}^1 = u_n^1 + u_n^2. \tag{1.12}$$

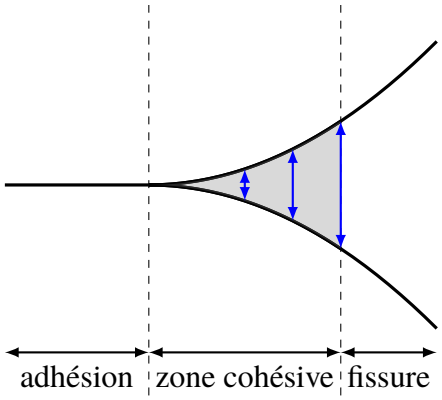


FIGURE 1.6 – Représentation d'une fissure avec le modèle de zone cohésive.

Comme pour le problème de contact entre un corps élastique et une surface rigide, les conditions de contact (1.11e) sont équivalentes à

$$\sigma_n^1(\boldsymbol{u}^1) = \left[\sigma_n^1(\boldsymbol{u}^1) - \gamma[\![u]\!]_n^1\right]_{\mathbb{R}^-}.$$

Étant donné l'espace $\boldsymbol{V}^i$ et un maillage $\mathcal{T}_h^i$ pour chaque domaine Ω^i, on définit l'espace de fonctions

$$\boldsymbol{V}_h := \boldsymbol{V}_h^1 \times \boldsymbol{V}_h^2, \qquad \text{où} \qquad \boldsymbol{V}_h^i := \left\{\boldsymbol{v}_h^i \in \boldsymbol{V}^i \ : \ \boldsymbol{v}_h^i|_T \in \boldsymbol{\mathcal{P}}^p(T) \text{ pour tout } T \in \mathcal{T}_h^i\right\},$$

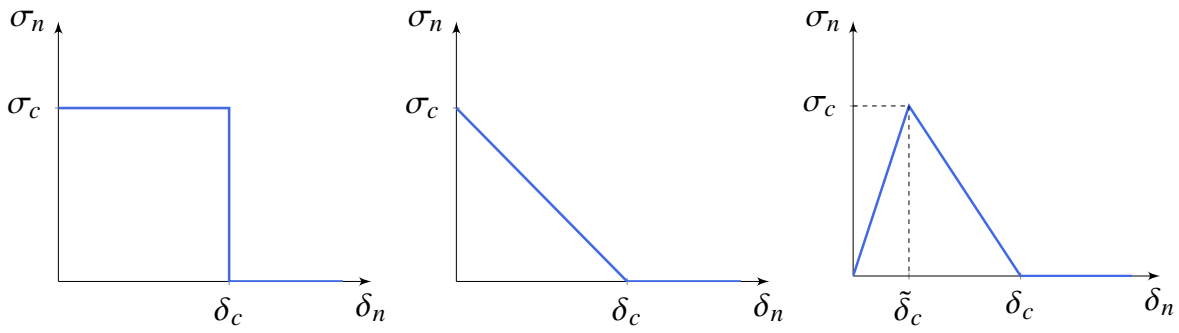
et la méthode à la Nitsche pour le problème (1.11) consiste à : Trouver $\boldsymbol{u}_h \in \boldsymbol{V}_h$ tel que

$$\begin{aligned}\sum_{i=1}^{2} \left(\boldsymbol{\sigma}^i(\boldsymbol{u}_h^i), \boldsymbol{\varepsilon}(\boldsymbol{v}_h^i)\right)_{\Omega^i} - \left(\left[\sigma_n^1(\boldsymbol{u}_h^1) - \gamma[\![u_h]\!]_n^1\right]_{\mathbb{R}^-}, [\![v_h]\!]_n^1\right)_{\Gamma_C^1} \\ = \sum_{i=1}^{2} \left(\left(\boldsymbol{f}^i, \boldsymbol{v}_h^i\right)_{\Omega^i} + \left(\boldsymbol{g}_N^i, \boldsymbol{v}_h^i\right)_{\Gamma_N^i}\right)\end{aligned} \tag{1.13}$$

pour tout $\boldsymbol{v}_h \in \boldsymbol{V}_h$.

1.2.3 Modèles avec zones cohésives

Les conditions de contact introduites dans la sous-section précédente sont assez simples à implémenter grâce aux formulations à la Nitsche. Pour modéliser et implémenter des comportements plus complexes, des modèles de joint peuvent être utilisés. Ils permettent de modéliser le comportement de discontinuités dont la localisation est connue, comme les zones d'interfaces des barrages, à travers d'une *loi de comportement* qui relie le saut du déplacement, noté $\boldsymbol{\delta}$, et la force entre les deux côtés de l'interface. En particulier, les lois de comportement présentées par la suite sont basées sur la notion de *zones cohésives*, voir Figure 1.6. Contrairement aux problèmes de contact précédents, qui ne présentent qu'une région avec adhésion ($u_n = 0$ et $\sigma_n \leq 0$) et une zone complètement fissuré ($u_n < 0$ et $\sigma_n = 0$), les modèles des zones cohésives introduisent une zone de transition entre celles-ci avec des forces non-nulles entre les lèvre de l'interface, dites *forces cohésives*. Ces types de modèles ont été



(a) Modèle de Dugdale (b) Modèle linéaire adoucissant (c) Modèle bilinéaire

FIGURE 1.7 – Trois simples exemples des modèles avec zones cohésives.

introduits dans les années '60 par Dugdale [51] et Barenblatt [11], et de simples exemples de lois cohésives sont montrés dans la Figure 1.7 : le modèle de Dugdale (Figure 1.7a), un modèle linéaire avec comportement adoucissant (Figure 1.7b), et un modèle bilinéaire avec comportement élastique linéaire puis adoucissant (Figure 1.7c). Initialement, ce formalisme a été utilisé pour décrire le phénomène de rupture en traction, mais il peut être étendu pour représenter aussi le comportement en compression en obtenant lois de comportement plus riches. On décrit ci-dessous l'extension avec force cohésives des deux problèmes de contact déjà introduits.

Remarque 1.1 (Saut du déplacement $\boldsymbol{\delta}$)**.** *Dans le contexte des modèles des zones cohésives, le comportement d'une zone d'interface est exprimé à travers le saut du déplacement $\boldsymbol{\delta}$ entre les deux côtés du joint. En général, en considérant à nouveau Γ_C^1 comme surface esclave, il est défini par*

$$\boldsymbol{\delta} := -[\![\boldsymbol{u}]\!]^1 := -(\boldsymbol{u}^1 - \boldsymbol{u}^2). \tag{1.14}$$

Par conséquent, $\delta_n = \boldsymbol{\delta}\,\boldsymbol{n}^1 = -[\![u]\!]_n^1$, voir (1.12). En particulier, dans le cas plus simple de contact entre un corps élastique et un surface rigide immobile, $\delta_n = -\boldsymbol{u}\,\boldsymbol{n} = -u_n$, et ceci est la raison pour laquelle en ouverture $u_n \leq 0$ (voir Figure 1.4), alors que $\delta_n \geq 0$ (voir les modèles de Figure 1.7).

Par simplicité, dans la suite, on continue à utiliser la notation des sous-sections précédentes, sans faire intervenir le saut de déplacement δ_n. La notation avec le saut de déplacement sera utilisé dans le chapitre suivantes.

Corps élastique et surface rigide

On considère à nouveau les cas de contact entre un corps élastique, occupant le domaine Ω, et une surface rigide, Figure 1.3. Cette fois, on suppose l'existence des forces cohésives à l'interface. On peut imaginer, par exemple, la présence d'un matériau collant entre l'objet élastique et la surface rigide. Alors, au lieu des conditions de contact (1.9), on a la loi de comportement d'interface sur la partie du bord de contact :

$$\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n} = \boldsymbol{F}(\boldsymbol{u}) \qquad \text{sur } \Gamma_C.$$

La fonction $\boldsymbol{F}$ qui intervient dans cette relation peut dépendre aussi d'autres variables internes éventuellement dissipatives, par exemple la composante plastique du déplacement dans un

modèle élasto-plastique ou une variable d'endommagement dans un modèle de mécanique de la rupture. En utilisant une approche classique (multiplication par une fonction test, application de la formule de Green et des conditions aux limites), le problème discrétisé devient : Trouver $\boldsymbol{u}_h \in \boldsymbol{V}_h$ tel que

$$(\boldsymbol{\sigma}(\boldsymbol{u}_h), \boldsymbol{\varepsilon}(\boldsymbol{v}_h)) - \left(\boldsymbol{F}(\boldsymbol{u}_h), \boldsymbol{v}_h\right)_{\Gamma_C} = (\boldsymbol{f}, \boldsymbol{v}_h) + (\boldsymbol{g}_N, \boldsymbol{v}_h)_{\Gamma_N} \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_h.$$

On note qu'il est possible de retrouver la formulation à la Nitsche pour le problème de contact (1.10) en imposant

$$\begin{cases} F_n(\boldsymbol{u}) = [\sigma_n(\boldsymbol{u}) - \gamma u_n]_{\mathbb{R}^-}, \\ \boldsymbol{F_t}(\boldsymbol{u}) = \boldsymbol{0}. \end{cases}$$

Deux corps élastiques

Dans le cas de contact entre deux domaines Ω^1 et Ω^2, Figure 1.5, en rajoutant des forces cohésives, les conditions (1.11e)-(1.11f) sur la surface esclave Γ_C^1 deviennent

$$\boldsymbol{\sigma}^1(\llbracket \boldsymbol{u} \rrbracket^1)\boldsymbol{n}^1 = \boldsymbol{F}(\llbracket \boldsymbol{u} \rrbracket^1) \qquad \text{sur } \Gamma_C^1, \tag{1.15}$$

où $\llbracket \boldsymbol{u} \rrbracket^1$ est toujours défini par (1.14). Alors, le problème discret est : Trouver $\boldsymbol{u}_h \in \boldsymbol{V}_h$ tel que

$$\sum_{i=1}^{2} \left(\boldsymbol{\sigma}^i(\boldsymbol{u}_h^i), \boldsymbol{\varepsilon}(\boldsymbol{v}_h^i)\right)_{\Omega^i} - \left(\boldsymbol{F}(\llbracket \boldsymbol{u}_h \rrbracket^1), \llbracket \boldsymbol{v}_h \rrbracket^1\right)_{\Gamma_C^1} = \sum_{i=1}^{2} \left(\left(\boldsymbol{f}^i, \boldsymbol{v}_h^i\right)_{\Omega^i} + \left(\boldsymbol{g}_N^i, \boldsymbol{v}_h^i\right)_{\Gamma_N^i} \right) \tag{1.16}$$

pour tout $\boldsymbol{v}_h \in \boldsymbol{V}_h$. À nouveau, on peut obtenir la formulation à la Nitsche correspondante (1.13) en imposant

$$\begin{cases} F_n(\llbracket \boldsymbol{u} \rrbracket^1) = \left[\sigma_n^1(\boldsymbol{u}^1) - \gamma \llbracket u \rrbracket_n^1\right]_{\mathbb{R}^-}, \\ \boldsymbol{F_t}(\llbracket \boldsymbol{u} \rrbracket^1) = \boldsymbol{0}. \end{cases}$$

Couplage hydromécanique

Pour la modélisation des barrages, on peut introduire le couplage hydromécanique à travers une fonction scalaire p définie sur l'interface. Cette fonction représente la pression due à la présence d'un fluide. En utilisant le principe de Terzaghi [120] adapté aux modèles de joint, la composante normale de la contrainte de la loi de comportement (1.15) devient donc :

$$F_n(\llbracket \boldsymbol{u} \rrbracket^1) = \sigma_n(\llbracket \boldsymbol{u} \rrbracket^1, p) = \sigma_n^{\text{meca}}(\llbracket \boldsymbol{u} \rrbracket^1) - p. \tag{1.17}$$

Une première possibilité consiste à imposer cette pression p a priori dans la zone d'interface (par exemple, lorsqu'on connaît le profil de sous-pression). Ce choix permet de formuler toujours le problème de manière purement mécanique, en utilisant la loi de comportement (1.17) dans (1.16).

La deuxième possibilité consiste à calculer explicitement la pression p pendant le calcul en introduisant une équation qui représente la conservation du flux hydraulique $\boldsymbol{w}(\llbracket \boldsymbol{u} \rrbracket^1, p)$ sur l'interface Γ_C^1 :

$$\boldsymbol{\nabla} \cdot \boldsymbol{w}(\llbracket \boldsymbol{u} \rrbracket^1, p) = 0 \qquad \text{sur } \Gamma_C^1.$$

Sur le bord de Γ_C^1, des conditions homogenées de Neumann sont considérées pour le gradient de pression $\boldsymbol{\nabla} p$. Dans ce cas, le problème discret (1.16) devient : Trouver $(\boldsymbol{u}_h, p_h) \in \boldsymbol{V}_h \times Q_h$ tels que

$$\sum_{i=1}^{2} \left(\boldsymbol{\sigma}^i(\boldsymbol{u}_h^i), \boldsymbol{\varepsilon}(\boldsymbol{v}_h^i)\right)_{\Omega^i} - \left(\boldsymbol{F}(\llbracket \boldsymbol{u}_h \rrbracket^1), \llbracket \boldsymbol{v}_h \rrbracket^1\right)_{\Gamma_C^1} = \sum_{i=1}^{2} \left(\left(\boldsymbol{f}^i, \boldsymbol{v}_h^i\right)_{\Omega^i} + \left(\boldsymbol{g}_N^i, \boldsymbol{v}_h^i\right)_{\Gamma_N^i} \right)$$
$$\left(\boldsymbol{w}(\llbracket \boldsymbol{u}_h \rrbracket^1, p_h), \boldsymbol{\nabla} q_h\right)_{\Gamma_C^1} = 0,$$

pour tout $(\boldsymbol{v}_h, q_h) \in \boldsymbol{V}_h \times Q_h$, où Q_h est l'espace des polynômes de degré $\leq p$ par morceaux continus sur Γ_C^1. En particulier, dans `code_aster`, le couplage est caractérisé par les relations suivantes :

$$F_n(\llbracket \boldsymbol{u} \rrbracket^1) = \sigma_n(\llbracket \boldsymbol{u} \rrbracket^1, p) = \sigma_n^{\text{meca}}(\llbracket \boldsymbol{u} \rrbracket^1) - p,$$
$$\boldsymbol{w}(\llbracket \boldsymbol{u} \rrbracket^1, p) = \frac{\rho}{12\bar{\mu}} \left(\max\{\epsilon_{\min}, \epsilon_{\min} - u_n\}\right)^3 \boldsymbol{\nabla} p,$$

où ρ et $\bar{\mu}$ sont, respectivement, la masse volumique et la viscosité dynamique du fluide, et $\epsilon_{\min}$ détermine la valeur minimale du flux hydraulique. Ces relations prennent en compte la contrainte normale supplémentaire liée à la présence de la pression hydraulique p (influence de l'hydraulique sur la mécanique), et le fait que la propagation du fluide est plus élevée quand le joint est ouvert (influence de la mécanique sur l'hydraulique).

Implémentation des interfaces : éléments finis de joint

Sur le plan numérique, les zones d'interface avec forces cohésives peuvent être modélisées à l'aide d'éléments finis particuliers qui se distinguent par leur forme aplatie : les élément finis de joint (voir [45, Partie II], [82, Section 3.I] ou la documentation de `code_aster` [46]). Ils permettent de représenter l'évolution d'une fissure avec trajet de discontinuité connu a priori en restant dans le cadre classique des éléments finis, et ont pour support géométrique des éléments finis dégénérés. De plus, l'ordre des élément de joint est conforme avec les éléments de Lagrange du maillage qui discrétise le domaine Ω, et, dans `code_aster` sont disponibles les éléments linéaires et quadratiques pour $d \in \{2, 3\}$. On donne plus de détails sur ces éléments dans le Chapitre 2.

1.3 Contribution de la thèse et organisation du manuscrit

Le principal défi de la modélisation des barrages consiste à caractériser et reproduire le comportement très complexe de leurs zones d'interface, avec un modèle robuste numériquement. En général, il existe deux approches pour introduire une loi de comportement : en partant d'observations physiques (voir par exemple [45]), ou en partant de concepts et principes physiques (voir par exemple [94]). La première façon est plutôt empirique, alors que la deuxième se base sur les principes de la thermodynamique, notamment la conservation de l'énergie élastique. Les lois actuelles dans `code_aster` sont "empiriques" et trop pauvres pour capturer l'ensemble des phénomènes. La première contribution de la thèse est l'introduction et l'analyse de nouvelles lois de comportement pour les zones d'interface dans une cadre approprié à la fois du point de vue physique et de la robustesse numérique. Les deux

prochaines chapitre traitent de ce travail : on commence dans le Chapitre 2 en proposant une loi de comportement qui couple plasticité et endommagement, dans l'esprit de [94] ; ensuite, le Chapitre 3 montre l'influence de l'introduction de dépendances hyper-élastiques dans les coefficients de rigidité du joint. Enfin, ces deux lois sont couplées dans la deuxième partie du Chapitre 5.

La deuxième originalité de la thèse consiste à introduire une technique d'analyse d'erreur a posteriori pour le problème de contact qui permet d'améliorer la robustesse du modèle numérique avec l'introduction de critères efficaces. Dans le Chapitre 4 on considère le problème de contact unilatéral sans frottement entre un objet élastique et une surface rigide et on présente une analyse d'erreur a posteriori basée sur la reconstruction équilibrée de la contrainte. Cette dernière est le point de départ pour l'introduction d'un algorithme adaptatif avec des critères d'arrêt pour l'algorithme de Newton et le réglage automatique d'un paramètre de régularisation. Ensuite, dans le chapitre final, de possibles extensions de l'estimation d'erreur a posteriori sont proposées.

Ce manuscrit inclut deux articles qui ont été soumis pour publication :

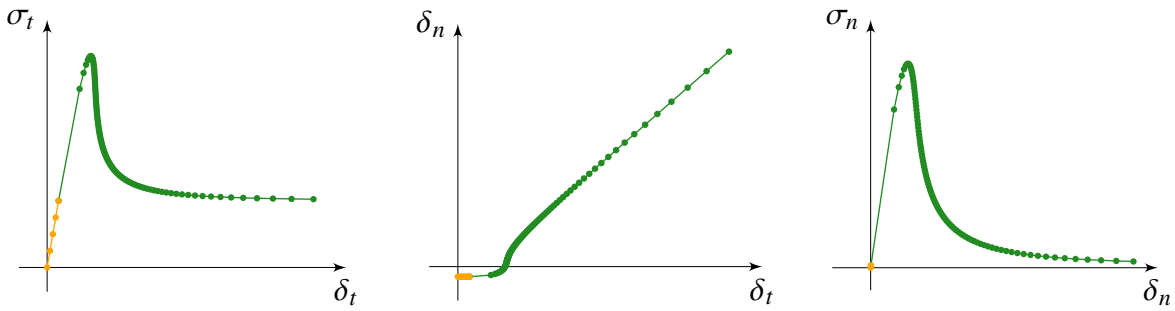
- "A posteriori error estimates via equilibrated stress reconstructions for contact problems approximated by Nitsche's method" [43] représente la majorité du Chapitre 4 ; le preprint est accessible dans ArXiv : https://arxiv.org/abs/2109.11944 ;
- "Hyperelastic nature of Hoek–Brown criterion" [61] représente la première partie du Chapitre 3 ; le preprint est accessible dans Hal : https://hal.archives-ouvertes.fr/hal-03501788.

1.3.1 Développement de lois de comportement

Les lois de comportement pour les joints qu'on analyse dans ce travail sont développées dans le cadre des *matériaux standard généralisés* adaptés aux zones d'interface. Ce formalisme théorique permet d'obtenir des modèles avec de bonnes propriétés mathématiques et compatibles avec les principes de la thermodynamique.

Dans le contexte de la modélisation des matériaux, dans les années '70 Halphen et Nguyen [66] ont introduit la notion de matériaux standard généralisés, c'est-à-dire des matériaux qui satisfont l'hypothèse de dissipativité normale [96, 97]. Cette théorie présente plusieurs avantages, notamment elle établit une classe de matériaux élasto-visco-plastiques et élasto-plastiques qui respectent automatiquement l'inégalité de Clausius–Duhem, et offre une formulation énergétique pour construire une loi de comportement mécanique, ce qui nous amène à la résolution d'un problème de minimisation. De plus, dans le cas de matériaux élasto-plastiques, on peut retrouver le principe du travail plastique maximal de Hill, qui impose la convexité du domaine de réversibilité et la normalité de l'écoulement plastique. Pour plus détails voir [63, 92, 93] et, pour l'extension aux modèles avec endommagement, [90, 91, 94].

L'adaptation aux modèles avec zones d'interface est possible grâce une approche variationnelle, basée sur le principe de minimisation de l'énergie. Ce dernier a été proposé par Francfort et Marigo [62] pour le problème de rupture fragile, en introduisant une énergie de surface au niveau de la fissure. En partant de cet article et en analogie avec le travail de thèse [82], l'idée est ici d'introduire une énergie de surface définie sur l'interface qui dépend



(a) Contrainte de frottement (b) Dilantance (c) Contrainte de traction

FIGURE 1.8 – Illustration de l'évolution pendant un essai de cisaillement à compression fixé (*gauche* et *milieu*) et un essai de traction (*droite*) ; cf. Figures 2.12 et 2.14, Chapitre 2.

du saut du déplacement et ensuite résoudre un problème de minimisation. En particulier, le modèle est complètement identifié par :

- les *variables d'états*, c'est-à-dire le saut du déplacement et les éventuelles variables internes dissipatives,
- la fonction de densité d'énergie de surface,
- le *domaine de réversibilité* (ou *d'élasticité*), caractérisé par une fonction seuil convexe.

L'intégration de ces lois de comportement dans `code_aster` a été faite avec une approche incrémentale. Dans chaque chapitre on montre en détail cette technique, et on reporte le calcul de la matrice tangent dans l'Annexe. En particulier, la stratégie qu'on présente dans ce travail, contrairement aux lois actuelles, permet de traiter directement la singularité du domaine de réversibilité.

Résumé du Chapitre 2 - Physique des interfaces : couplage mécanique entre plasticité et endommagement

Après avoir illustré les hypothèses et propriétés des matériaux standard généralisés et de l'approche variationnelle, on présente brièvement les lois d'interfaces actuelles dans `code_aster`. Ensuite, on décrit le premier modèle original pour les zones d'interfaces. L'idée clé est d'introduire dans la fonction d'énergie de surface un terme d'écrouissage cinématique qui couple plasticité et endommagement, dans l'esprit de [94], pour pouvoir reproduire les deux phénomènes majeures des joints : glissement avec frottement et rupture en traction. De plus, le domaine de réversibilité est déterminé par un critère linéaire de type Drucker–Prager [50] et reste fixe dans l'espace des forces généralisées. La loi de comportement "couplée" qu'on dérive permet de représenter le comportement élastique à faibles chargements et le pic et de la contrainte de frottement et de la contrainte de traction, avec une phase adoucissante et un palier asymptotique, voir les Figures 1.8a et 1.8c. En particulier, la présence de ces pics est liée au fait que, pendant l'évolution de plasticité et endommagement, le domaine de réversibilité est mobile dans l'espace des contraintes. De plus, on voit apparaître de la dilatance pour des essais de cisaillement à compression fixée, Figure 1.8b, ce qui est possible de constater expérimentalement à faible cisaillement. Malheureusement, pour des cisaillements importants, on ne peut pas avoir la saturation de la dilatance, ce qui

semble être lié à règle d'écoulement normale de la plasticité (qui est fixée par l'hypothèse des matériaux standard généralisé). Ce problème est plus évident pour des essais de cisaillement cycliques. Enfin, on présente aussi une analyse de l'influence des neuf paramètres sur les courbes d'un essai de cisaillement à compression fixée.

Résumé du Chapitre 3 - Hyper-élasticité

Dans le but d'améliorer le modèle proposé dans le chapitre précédent sans sortir du cadre des matériaux standard généralisés, différents changements ont été considérés. En particulier, dans cette thèse on présente un modèle avec élasticité hyperbolique non-linéaire obtenu en modifiant les rigidités normale et tangentielle du joint. Pour mieux étudier l'influence de ce type d'hyper-élasticité, dans le Chapitre 3 on présente l'analyse d'un modèle simplifié où on néglige l'endommagement. Cette analyse inclut les étude des réponses sous des chargement typiques (essais de cisaillement, compression et traction), des notes sur l'implémentation dans `code_aster`, et des résultats sur un modèle d'un barrage. Les principaux nouveaux résultats obtenus sont :

- la création d'un lien entre une classe de critères de plasticité linéaires et une classe de critères quadratiques : en supposant que le domaine de réversibilité dans l'espace des forces généralisés est toujours fixe et défini via un critère linéaire de type Drucker–Prager, le domaine correspondant dans l'espace des contraintes est établi par un critère quadratique de type Hoek–Brown ;
- la réduction de la pente de la dilatance en compression et l'obtention de la saturation de la dilatance pour des essais de cisaillement cycliques.

En outre, on propose aussi l'adaptation de ce modèle au contexte des géomatériaux, pour modéliser les non-linéarités observées pour certains matériaux et la saturation de la dilatance pour des essais uni-axiales cycliques. Dans ce cas, on a réalisé des simulations numériques avec le générateur de code `mfront` ([70], `http://tfel.sourceforge.net`), pour obtenir la réponse d'un point matériel sous différents types de chargement. Cet outil permet la mise en œuvre de lois de comportement complexes, qui peuvent être utilisées pour faire des calculs avec `code_aster`.

1.3.2 Analyse d'erreur a posteriori

Au cours d'une simulation numérique, la qualité de la solution numérique peut être mesurée à travers des estimateurs d'erreur a posteriori, qui fournissent une borne supérieure calculable de l'erreur locale [123, 37, 2, 17, 23, 126]. Une propriété importante pour un estimateur a posteriori est l'efficacité, exprimant le fait que l'estimateur local représente aussi une borne inférieure de l'erreur dans le voisinage de l'élément considéré. Une conséquence immédiate de l'efficacité est qu'on peut prédire la localisation de l'erreur.

Dans la littérature, on peut distinguer différentes méthodes d'estimation d'erreur a posteriori : les méthodes par résidus [2, 22, 114, 124], les méthodes par reconstruction de flux équilibrés [37, 87, 17, 57], les méthodes de résidus équilibrés [81, 1, 2], les méthodes hiérarchiques [10, 54, 4], les méthodes par la moyenne [60, 89, 32], les méthodes par estimateurs fonctionnels [101, 114, 113].

Dans cette thèse on se concentrera sur les méthodes d'estimation par flux équilibrés, en suivant l'approche proposée par [37, 17, 57] pour le problème de Poisson. Dans ce contexte, on cite aussi les travaux de thèse [128, 115, 34]. Les avantages de cette approche sont multiples. Plus en détail, elle permet de :

- avoir une borne supérieure de l'erreur sans constantes inconnues du type

$$|||\boldsymbol{u} - \boldsymbol{u}_h||| \leq \left(\sum_{T \in \mathcal{T}_h} (\eta_T(\boldsymbol{u}_h))^2 \right)^{1/2},$$

  où $\eta_T(\boldsymbol{u}_h)$ est l'estimateur local total, c'est-à-dire, une quantité défini sur l'élément $T \in \mathcal{T}_h$ et qui ne dépend que de la solution approchée ;
- obtenir des estimateurs qui représentent des propriétés physiques non satisfaites par $\boldsymbol{\sigma}(\boldsymbol{u}_h)$ (e.g., en général, elle n'est pas continue à travers les interfaces des éléments du maillage) en introduisant la notion de reconstruction équilibrée de la contrainte ;
- distinguer et comparer les différentes composantes de l'erreur (par exemple, discrétisation spatiale et temporelle, linéarisation, régularisation, etc.) ;
- introduire des critères d'arrêt pour les différents solveurs et des algorithmes adaptatifs pour le raffinement du maillage.

Ici, on présente brièvement cette méthode pour un problème mécanique (en statique, quasi-statique ou dynamique) défini sur un domaine Ω. En discrétisant ce type de problème avec une méthode d'élément finis $\boldsymbol{H}^1$-conformes, on obtient par définition une solution approchée $\boldsymbol{u}_h$ qui appartient à l'espace de Sobolev $\boldsymbol{H}^1(\Omega)$. Par contre, le tenseur des contraintes correspondant $\boldsymbol{\sigma}(\boldsymbol{u}_h)$ est, en général, non–physique, au sens où ses composantes normales à travers les interfaces du maillage sont discontinues et il ne vérifie pas localement ni la condition d'équilibre (1.1), ni les conditions aux limites de Neumann (1.5) ou les éventuelles conditions d'interfaces. Par conséquent, $\boldsymbol{\sigma}(\boldsymbol{u}_h) \notin \mathbb{H}(\mathbf{div}, \Omega)$, où $\mathbb{H}(\mathbf{div}, \Omega)$ est l'espace décrit par les fonctions de $\mathbb{L}^2(\Omega) := [L^2(\Omega)]^{d \times d}$ avec divergence faible dans $\boldsymbol{L}^2(\Omega)$. L'idée clé de l'introduction d'une reconstruction équilibrée est de fournir une "correction" de la contrainte en construisant un tenseur $\mathbb{H}(\mathbf{div})$-conforme tel que sa divergence et sa composante normale sont localement en équilibre avec les forces volumiques et surfaciques, respectivement. Afin de ne pas rendre le calcul trop coûteux, on résout des problèmes locaux de minimisation sous contrainte définis sur les patchs (c'est-à-dire, union des simplexes) autour de chaque sommet du maillage, Figure 1.9. Les solutions de ces problèmes sont ensuite assemblées pour obtenir la reconstruction des contraintes globale $\boldsymbol{\sigma}_h$.

Résumé du Chapitre 4 - Estimation a posteriori pour le problème de contact unilatéral

Le but de ce chapitre est d'introduire une méthode d'estimation d'erreur a posteriori basée sur une reconstruction du tenseur des contraintes pour le problème de contact unilatéral sans frottement entre un objet élastique et une surface rigide, c'est-à-dire, le problème (1.1) avec conditions aux limites (1.4), (1.5) et (1.9). L'enjeu principal a consisté à étendre l'approche adoptée dans [16] au problème avec conditions de contact. La première nouveauté est dans le choix de la mesure de l'erreur : pour notre analyse on a utilisé la norme duale d'un opérateur résiduel défini à partir de la formulation discrète éléments finis. Cet opérateur est défini de

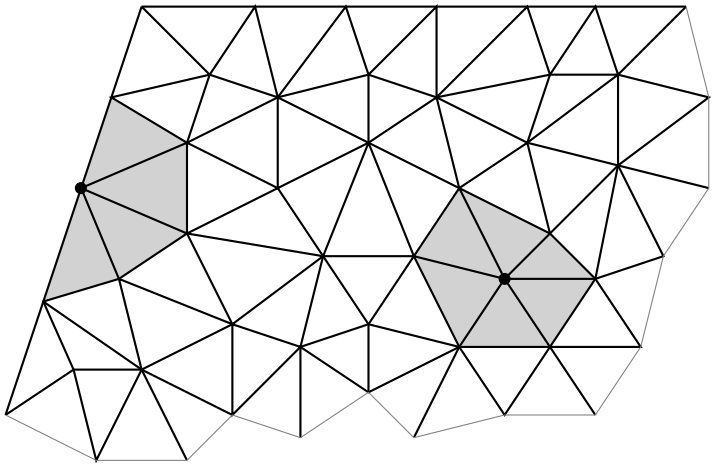


FIGURE 1.9 – Deux exemples de patch en considérant un nœud sur le bord et un nœud dans l'intérieur du domaine Ω.

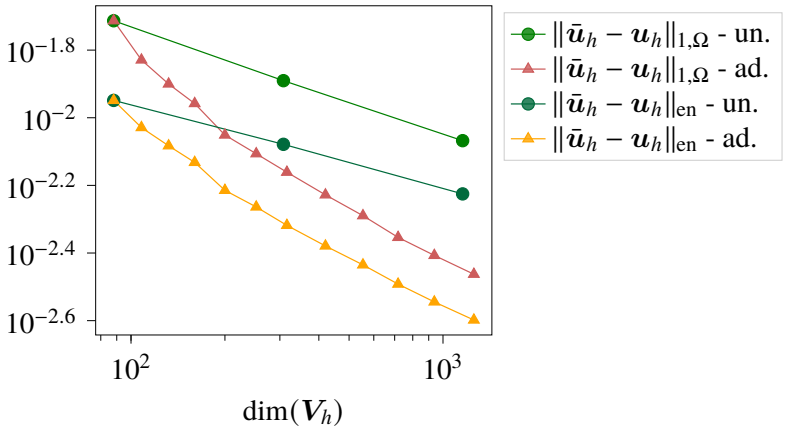


FIGURE 1.10 – Comparaison entre raffinement uniforme et adaptatif pour la norme $\boldsymbol{H}^1$ et la norma d'énergie ; cf. Figure 4.7a, Chapitre 4.

façon naturelle à partir de la formulation à la Nitsche (1.10) et inclut aussi des termes qui tiennent compte des conditions aux limites.

Une deuxième nouveauté réside dans la définition des problèmes locaux sur les patches autour des sommets qui déterminent la reconstruction du tenseur des contraintes $\boldsymbol{\sigma}_h$. Motivés par les résultats de [116, 16] pour les problèmes d'élasticité linéaire et non-linéaire, ces problèmes locaux sont définis sur l'espace des éléments finis mixtes de Arnold–Falk–Winther [6], qui imposent la symétrie de $\boldsymbol{\sigma}_h$ de manière faible à travers des multiplicateurs de Lagrange. La reconstruction ainsi obtenue appartient automatiquement à l'espace $\mathbb{H}(\mathbf{div}, \Omega)$ et est en équilibre locale avec les forces volumiques et surfaciques.

À partir de ces deux éléments (mesure de l'erreur via la norme duale de l'opérateur résidu et reconstruction équilibrée par résolution de problèmes locales) on obtient une analyse complète, au sens où : on montre une estimation a posteriori sans constantes inconnues qui distingue les composantes de l'erreur ; on compare le norme duale du résidu avec une norme plus standard ; on prouve un résultat d'efficacité des estimateurs. Grâce à la définition des estimateurs locaux, on définit un critère adaptatif pour le raffinement du maillage qui équilibre

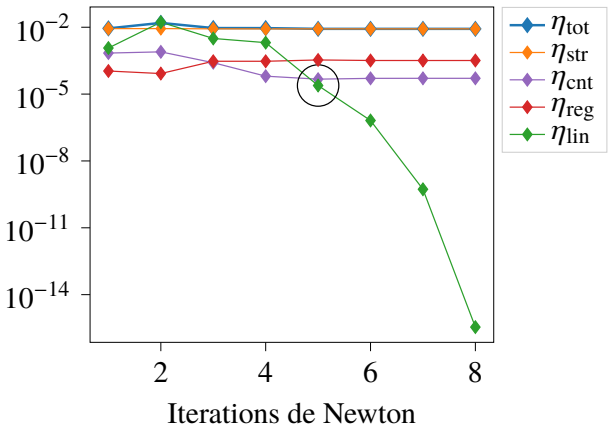


FIGURE 1.11 – Exemple d'un critère d'arrêt pour le solveur non-linéaire (algorithme de Newton) ; cf. Figure 4.10, Chapitre 4.

leur distribution : seuls les éléments ayant la valeur de l'estimateur la plus élevée sont raffinés. On vérifie avec un exemple numérique que cette technique adaptative permet d'obtenir un meilleur taux de convergence comparée à une technique de raffinement uniforme, Figure 1.10. En outre, dans l'estimation de l'erreur, on sépare un estimateur de régularisation et un estimateur de linéarisation qui sont liés à la méthode choisie pour résoudre numériquement le problème discret. En comparant ces dernières avec les autres estimateurs, on introduit des critères d'arrêt pour la valeur du paramètre de régularisation et pour le nombre d'itérations de Newton, ce qui permettent de réduire le nombre d'itérations du solveur. Une illustration d'un critère d'arrêt pour le solveur non-linéaire est montré dans la Figure 1.11 : les itérations de Newton sont arrêtées après seulement 5 pas, lorsque la valeur de l'estimateur de linéarisation η_{lin} est suffisamment petite par rapport à la somme des autres estimateurs. De plus, on note que, à cette itération, l'estimateur de linéarisation n'influence plus le comportement de l'estimateur total η_{tot} et que les autres estimateurs se sont déjà stabilisés. Les critères d'arrêt et la technique adaptative pour le raffinement de maillage sont les points clé pour l'introduction d'un algorithme adaptatif efficace.

Résumé du Chapitre 5 - Extension de l'analyse a posteriori et modèle de joint unifié

Ce chapitre conclusif est consacré à explorer les extensions des études des chapitres précédents. Dans une premier partie, on considère la possibilité d'étendre l'analyse a posteriori pour des problèmes de contact plus complexes avec forces cohésives, en partant des résultats du Chapitre 4. Pour les lois de comportement qui peuvent être reconduite à des relations hyper-élastique (c'est à dire, avec comportement réversible), l'analyse s'étend de façon naturelle aux problèmes où on considère une zone d'interface avec force cohésive entre un domaine élastique et une surface rigide ou entre deux domaines élastiques. Ensuite, on termine le manuscrit en présentant un modèle pour les zones d'interface unifié qui combine les lois de comportement des Chapitres 2 e 3, c'est à dire, un modèle dans le cadre des matériaux standard généralisés adaptés aux zone d'interface avec un terme d'écrouissage cinématique qui couple plasticité et endommagement, et une dépendance hyperbolique du saut du déplacement dans les coefficients de rigidité.

2 Physics of interfaces: mechanical coupling of plasticity and damage

In this chapter, after recalling the context and the assumptions under which the mechanical constitutive relations for interface zones are developed (cohesive zone model, generalized standard materials, energy formalism), we analyze the relations already implemented in the software `code_aster` *(*`JOINT_MECA_RUPT` *and* `JOINT_MECA_FROT`*). Then, a first new model that couples plasticity and damage is described and its main properties are analyzed. We conclude the chapter with some notes about the implementation of the joint constitutive relation using an incremental approach.*

Contents

2.1 Introduction

Engineer teams often use finite element simulations to study the behavior of large hydraulic structures like concrete dams. It is well known that the interface zones are the weakest regions of these structures and that the stability of the structure largely depends on their hydro-mechanical behavior. These zones are mainly the contact between the concrete and rock in the foundation, the joints between different blocks of the structure, and the joints in concrete. These interfaces generally can be represented as a rough discontinuity, filled with some material (e.g. clay, grout or rock elements), and they display various important phenomena: friction, loss of tensile strength, elastic behavior for very small displacements, progressive disappearance of the peak of the shear stress with cyclic loading.

In the context of joint modeling, the Cohesive Zone Model (CZM) introduced in the sixties by Barenblatt [11] and Dugdale [51], can be used to describe the phenomena of rupture and friction at the interface level. This model introduces the notion of cohesive force between the two sides of the interface. The behavior of the interface is then characterized by a constitutive relation between the displacement jump and the cohesive stress vector.

In the seventies, the framework of Standard Generalized Materials (SGM) for geomaterials, which is based on the hypothesis of normal dissipativity of Moreau [96, 97], was introduced by Nguyen and Halphen [102]. This theory establishes a class of elasto-viscoplastic and elasto-plastic materials that satisfy the Clausius–Duhem inequality, and offers an energetic formulation for constructing a constitutive relation, obtaining a model with good mathematical and numerical properties. This framework can be extended to the joint modeling using the variational formulation of [62].

Before presenting the problem setting, the current constitutive relations and the new model we propose, we describe these two frameworks (CZM and SGM).

2.1.1 Cohesive Zone Model for joints

Interface models can be used to represent discontinuities of the structure with known localization, like the joints in dams. They make it possible to consider complex behaviors, and they are characterized by a *constitutive relation* that links the displacement jump and the force density between the two sides of the interface. The constitutive relations we will present in this and in the following chapters are based on the notion of cohesive zones and forces, see Figure 2.1. In particular, we can distinguish three zones in the interface region:

- an adhesion zone where the two sides of the joint are in contact,
- a completely fissured zone,
- a transition zone between the latter with nonzero forces.

These models have been introduced in the sixties by Dugdale [51] and Barenblatt [11]. Three simple examples of cohesive laws are shown in Figure 2.2: the Dugdale model (Figure 2.2a), a linear model with softening behavior (Figure 2.2b), and a bilinear model with linear elastic and then softening phase (Figure 2.2c).

Let $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, be a domain representing an elastic body, and let Γ_C be an interface that divides Ω into two subdomains denoted by Ω^1 and Ω^2, i.e., $\Omega \setminus \Gamma_C = \Omega^1 \cup \Omega^2$, Figure 2.3a. Given the displacement $\boldsymbol{u}\colon \Omega \to \mathbb{R}^d$ and fixing the unit normal vector $\boldsymbol{n}$ on

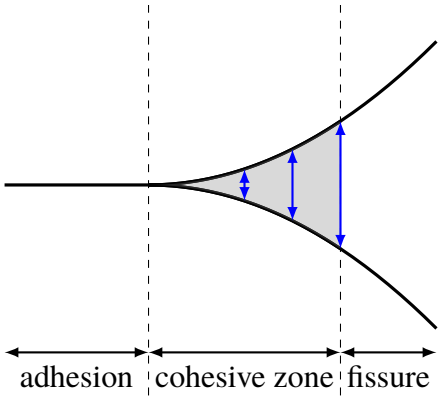
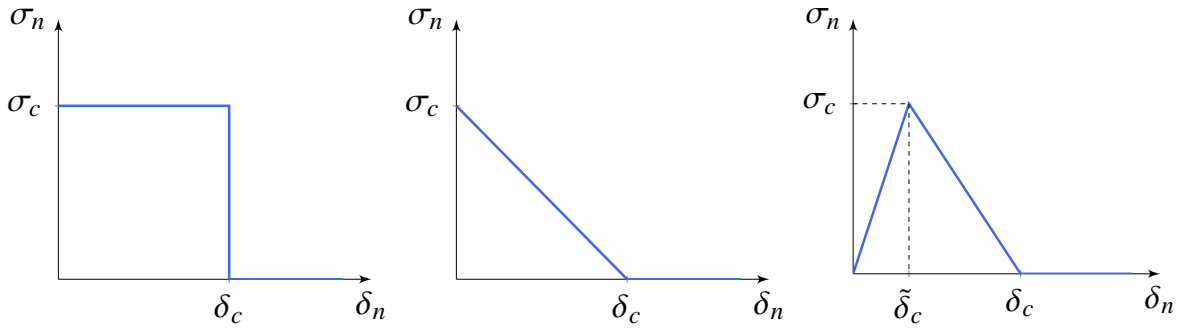


Figure 2.1 – Representation of a fissure in the Cohesive Zone Model framework, with separation into three regions.



(a) Dugdale model (b) Softening linear model (c) Bilinear model

Figure 2.2 – Three simple examples of models with cohesive forces.

Γ_C pointing from Ω^1 towards Ω^2 as shown by Figure 2.3a, the displacement jump on the interface Γ is defined by

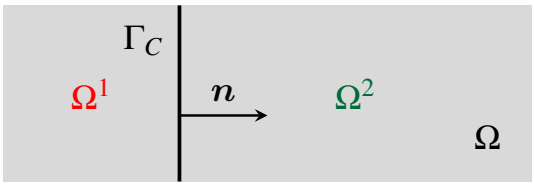
$$\boldsymbol{\delta} := -[\![\boldsymbol{u}]\!] = -(\boldsymbol{u}^1 - \boldsymbol{u}^2), \tag{2.1}$$

where $\boldsymbol{u}^1 := \boldsymbol{u}|_{\Omega^1}$ and $\boldsymbol{u}^2 := \boldsymbol{u}|_{\Omega^2}$, see Figure 2.3b and 2.3c. Then, the behavior of the interface region is determined by its mechanical constitutive relation, that is a relation of the type

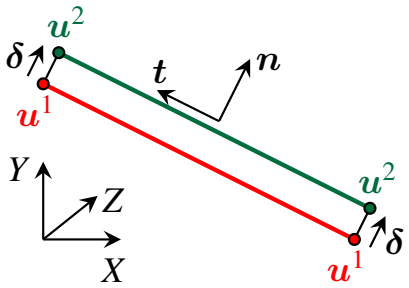
$$\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\delta}) \qquad \text{on } \Gamma_C, \tag{2.2}$$

where $\boldsymbol{\sigma}$ represents the cohesive (density) force on the interface. Notice that this is a simplified notation since, in general, also some other state variables $\boldsymbol{\alpha}$ has to be included in this relation. The vectors quantities defined on the joint can be decomposed into normal and tangential components, choosing a suitable local coordinate system: $(\boldsymbol{n}, \boldsymbol{t})$ if $d = 2$, and $(\boldsymbol{n}, \boldsymbol{t}_1, \boldsymbol{t}_2)$ if $d = 3$, where $\boldsymbol{t}$ and $(\boldsymbol{t}_1, \boldsymbol{t}_2)$ are unit tangent vectors to the joint, and we recall that $\boldsymbol{n}$ is the unit normal vector that points from Ω^1 towards Ω^2, Figure 2.3. For example, the displacement jump and the cohesive force can be written as

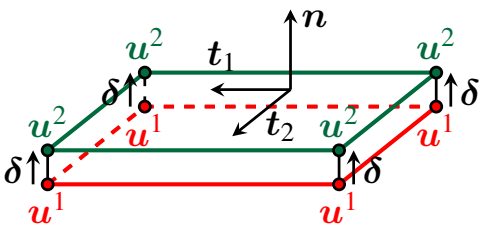
$$\begin{aligned}
\boldsymbol{\delta} = (\delta_n, \delta_t)^\top \quad &\text{and} \quad \boldsymbol{\sigma} = (\sigma_n, \sigma_t)^\top && \text{if } d = 2, \\
\boldsymbol{\delta} = (\delta_n, \boldsymbol{\delta_t})^\top = (\delta_n, \delta_{t_1}, \delta_{t_2})^\top \quad &\text{and} \quad \boldsymbol{\sigma} = (\sigma_n, \boldsymbol{\sigma_t})^\top = (\sigma_n, \sigma_{t_1}, \sigma_{t_2})^\top && \text{if } d = 3.
\end{aligned}$$



(a) Domain Ω with an interface Γ_C



(b) Bi-dimensional joint



(c) Tri-dimensional joint

Figure 2.3 – Example of domain Ω divided into Ω^1 and Ω^2 by an interface Γ_C (*top*); representation of the displacement jump $\boldsymbol{\delta}$ with local reference system, for $d = 2$ (*bottom-left*) and $d = 3$ (*bottom-right*).

From now on, we will use the notation of the case $d = 3$, whereas, for simplicity, the figures will be presented only for the case $d = 2$, which can be recovered with $\delta_{t_1} = \delta_t$ and $\delta_{t_2} = 0$.

2.1.2 Standard Generalized Materials

In the context of continuous materials modeling, the elasto-plastic models rely on four ingredients: a relation between the stress tensor $\boldsymbol{\sigma}$ and the deformation tensor $\boldsymbol{\varepsilon}$ (i.e., the *constitutive relation*); a yield criterion fixing the loading surface for plasticity and the region where the behavior is reversible and elastic (i.e., the *reversibility domain* or *elastic domain*); a *flow rule* that establishes the evolution of plasticity; a possible *cinematic* or *isotropic hardening* that describes the changes of the yield criterion during the evolution. As a consequence, different choices have to be done for fixing a model. Some elements can be identified straight through normalized experimental setup, like the yield criterion, others have more complex influence and are usually fitted via full mechanical response analysis. The most common yield criteria for geomaterials modeling only depend on two parameter. They include the Mohr–Coulomb criterion [80], the Drucker–Prager criterion [50, 3], and the Hoek–Brown criterion [72, 52]. On the other hand, the identification of the flow rule is a more delicate task. In practice, physical principles can be considered for reducing the possible choices.

In the seventies, Halphen et Nguyen [66] introduced the notion of Standard Generalized Materials (SGM), i.e., materials that satisfy the hypothesis of normal dissipativity [96, 97]. This theory has multiple advantages, in particular it identifies a class of elasto-viscoplastic and elasto-plastic materials that automatically respect the Clausius–Duhem inequality and offers an energetic formulation to construct a mechanical constitutive relation, leading us to the resolution of a minimization problem. Furthermore, in the case of elasto-plastic materials,

one can recover the Hill maximum plastic work principle, which imposes the convexity of the reversibility domain and the normality of the flow rule. For more details we refer to [63, 92, 93] and, for the extension to damage models, to [90, 91, 94].

The principal ingredients of this theory are:

1. the state variables,
2. a convex potential ϕ which represent the free energy,
3. a convex dissipation potential φ.

For example, in the elasto-plastic case without hardening, the state variables are the deformation tensor ε and its plastic component $\boldsymbol{p}$, the free energy ϕ corresponds to the elastic energy, and the dissipation potential φ is the indicator function of the reversibility domain. The constitutive relation, along with the dissipative thermodynamical forces, is obtained by derivation of the function ϕ, whereas the evolution rule of the dissipative state variables is related to the subgradient of φ.

In this manuscript, motivated by the good mathematical properties, we propose some models remaining in the framework of SGM adapted to the interface zones with cohesive forces modeling. This adaptation is possible thanks to a variational approach, based on the energy minimization principle. The latter was proposed by Francfort and Marigo [62] for the problem of fragile rupture, by introducing a surface energy at the crack level.

2.2 Problem setting and state of art in code_aster

After describing the mechanical problem we are considering, we present the standard mechanical tests performed in laboratory or in situ on joints, and we describe the current constitutive relations available in code_aster (https://code-aster.org).

2.2.1 Joint problem with cohesive forces

In this work, we only consider static or quasi-static problems for dam modeling, neglecting the viscous and thermal effects. Beside adopting the notations and hypothesis of Subsection 2.1.1, we assume that the boundary $\partial\Omega$ of the domain Ω is partitioned into two non-overlapping parts Γ_D and Γ_N in which we impose some Dirichlet and Neumann boundary conditions. The domain is subjected to a volume force $\boldsymbol{f} \in \boldsymbol{L}^2(\Omega) := [L^2(\Omega)]^d$ and a surface load $\boldsymbol{g}_N \in \boldsymbol{L}^2(\Gamma_N) := [L^2(\Gamma_N)]^d$. Furthermore, we denote with $\varepsilon(\boldsymbol{v}) := \frac{1}{2}(\boldsymbol{\nabla v} + \boldsymbol{\nabla v}^\top)$ the strain tensor field, and with $\boldsymbol{\sigma}(\boldsymbol{v}) \in \mathbb{R}^{d\times d}_{\text{sym}}$ the Cauchy stress tensor. Then, the static problem with an interface zone Γ_C is expressed by: Find the displacement field $\boldsymbol{u}\colon \Omega \to \mathbb{R}^d$ such that

$$\begin{aligned}
\mathbf{div}\,\boldsymbol{\sigma}(\boldsymbol{u}) + \boldsymbol{f} &= 0 && \text{in } \Omega \setminus \Gamma_C, && (2.3a)\\
\boldsymbol{\sigma}(\boldsymbol{u}) &= \mathbb{E}\,\varepsilon(\boldsymbol{u}) && \text{in } \Omega \setminus \Gamma_C, && (2.3b)\\
\boldsymbol{u} &= 0 && \text{on } \Gamma_D, && (2.3c)\\
\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n} &= \boldsymbol{g}_N && \text{on } \Gamma_N, && (2.3d)\\
\boldsymbol{\sigma} &= \boldsymbol{\sigma}(\boldsymbol{\delta}) && \text{on } \Gamma_C, && (2.3e)
\end{aligned}$$

where **div** is the divergence operator acting row-wise on tensor valued functions, and $\mathbb{E}$ is the fourth order symmetric elasticity tensor such that, for all second-order tensor $\boldsymbol{\tau}$,

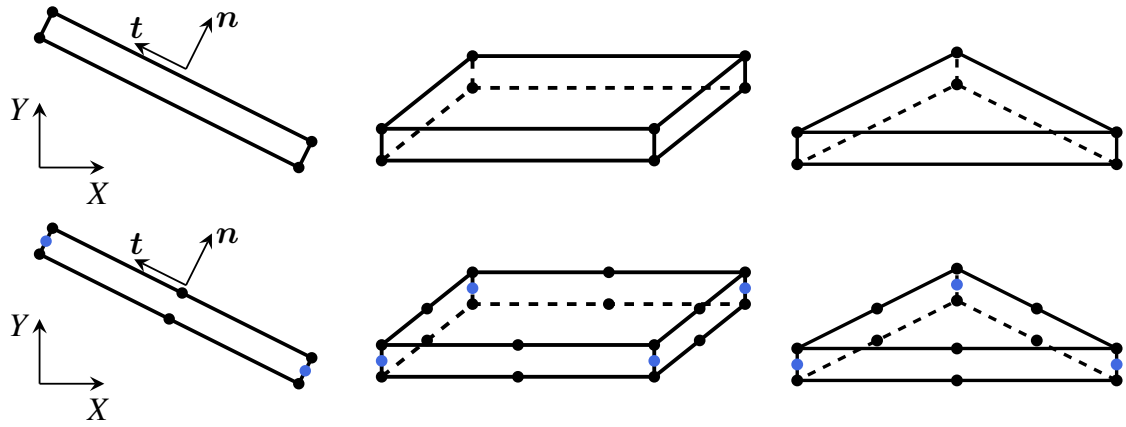


Figure 2.4 – Linear (1^{st} *line*) and quadratic (2^{nd} *line*) joint elements in `code_aster` [46]. In particular, they are: a 4-node and 8-node quadrangle degenerated in a segment (*left*), a 8-node and 20-node hexahedron degenerated in a rectangle (*middle*), and a 6-node and 15-node pentrahedron degenerated in a triangle (*right*).

$\mathbb{E}\boldsymbol{\tau} = \lambda \mathrm{Tr}(\boldsymbol{\tau})\boldsymbol{I}_d + 2\mu\boldsymbol{\tau}$, with λ and μ denoting the Lamé parameters. Notice that in the condition on the interface (2.3e) we omit the normal vector $\boldsymbol{n}$, using the notation $\boldsymbol{\sigma}$ for both stress tensor and stress vector, i.e., $\boldsymbol{\sigma} \equiv \boldsymbol{\sigma}\boldsymbol{n}$ on Γ_C.

Denoting by $\boldsymbol{H}^1_D(\Omega^i)$ the subspace of $\boldsymbol{H}^1(\Omega^i)$ incorporating the Dirichlet boundary condition on Γ_D, $i \in \{1,2\}$, defining $\boldsymbol{V} := \boldsymbol{H}^1_D(\Omega^1) \times \boldsymbol{H}^1_D(\Omega^2)$, and using a standard technique (i.e., multiplying by a test function $\boldsymbol{v} \in \boldsymbol{V}$, applying the Green formula and the conditions (2.3d)-(2.3e)) we obtain the following problem: Find $\boldsymbol{u} \in \boldsymbol{V}$ such that

$$(\boldsymbol{\sigma}(\boldsymbol{u}), \varepsilon(\boldsymbol{v}))_{\Omega\setminus\Gamma_C} + (\boldsymbol{\sigma}(\boldsymbol{\delta}), \boldsymbol{\delta}^v)_{\Gamma_C} = (\boldsymbol{f}, \boldsymbol{v})_{\Omega\setminus\Gamma_C} + (\boldsymbol{g}_N, \boldsymbol{v})_{\Gamma_N} \qquad \forall \boldsymbol{v} \in \boldsymbol{V}, \tag{2.4}$$

where $\boldsymbol{\delta}^v := -[\![\boldsymbol{v}]\!] = -(\boldsymbol{v}^1 - \boldsymbol{v}^2)$.

Discretization

Let $\mathcal{T}_h^1$ and $\mathcal{T}_h^2$ two meshes that discretize the subdomains Ω^1 and Ω^2, respectively. The interface zone Γ_C can be numerically modeled by means of particular finite elements which are characterized by their flat shape: the *joint elements* (see [45, Part II], [82, Section 3.I] or the documentation of `code_aster` [46]). They enable easily the implementation of complex behaviors remaining in the classical framework of finite elements, and they have degenerate finite elements as geometrical support. Figure 2.4 show the linear and quadratic joint elements that are available in `code_aster`. The order of the joint elements is consistent with the order of the meshes $\mathcal{T}_h^1$ and $\mathcal{T}_h^2$.

Remark 2.1 (Joint and interface models)**.** *In* `code_aster`*, another type of finite elements used for the modeling of interfaces are the* interface element *(see [86] and the documentation of* `code_aster` *[47]). They also have degenerate finite elements as geometrically support but the implementation is more delicate since they leads to a mixed problem based on a augmented Lagrangian formulation. Although in* `code_aster`*, there is a distinction between joint models and interface models, in this manuscript we continue to use the words* joint *and* interface *as synonyms to denote the discontinuities of the domain.*

Remark 2.2 (Creation of joint elements with salome_meca)**.** *With salome_meca, i.e., the platform that includes the mechanical solver code_aster, joint elements can be introduced between the meshes $\mathcal{T}_h^1$ and $\mathcal{T}_h^2$ with the option* Duplicate Node or/and Elements*, which duplicates the nodes on the joint and creates elements with zero thickness.*

Let

$$\boldsymbol{V}_h := \boldsymbol{V}_h^1 \times \boldsymbol{V}_h^2, \qquad \text{where} \qquad \boldsymbol{V}_h^i := \left\{ \boldsymbol{v}_h^i \in \boldsymbol{H}_D^1(\Omega^i) \ : \ \boldsymbol{v}_h^i|_T \in \boldsymbol{\mathcal{P}}^p(T) \text{ for any } T \in \mathcal{T}_h^i \right\},$$

$i \in \{1, 2\}$. Then, the discrete formulation reads: Find $\boldsymbol{u}_h \in \boldsymbol{V}_h$ such that

$$(\boldsymbol{\sigma}(\boldsymbol{u}_h), \boldsymbol{\varepsilon}(\boldsymbol{v}_h))_{\Omega\setminus\Gamma_C} + (\boldsymbol{\sigma}(\boldsymbol{\delta}_h), \boldsymbol{\delta}_h^v)_{\Gamma_C} = (\boldsymbol{f}, \boldsymbol{v}_h)_{\Omega\setminus\Gamma_C} + (\boldsymbol{g}_N, \boldsymbol{v}_h)_{\Gamma_N} \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_h, \tag{2.5}$$

where $\boldsymbol{\delta}_h := -[\![\boldsymbol{u}_h]\!]$ and $\boldsymbol{\delta}_h^v := -[\![\boldsymbol{v}_h]\!]$. In general, an iterative method is used to solve this nonlinear problem numerically. The most common one, which is also available in code_aster [48], is the Newton method. At each iteration $k \geq 1$, we have to solve a linearized discrete problem in which the nonlinear term $\boldsymbol{\sigma}(\boldsymbol{\delta}_h)$ is replaced by a linear approximation $\boldsymbol{\sigma}_{\text{lin}}^{k-1}(\boldsymbol{\delta}_h)$: Find $\boldsymbol{u}_h^k \in \boldsymbol{V}_h$ such that

$$(\boldsymbol{\sigma}(\boldsymbol{u}_h^k), \boldsymbol{\varepsilon}(\boldsymbol{v}_h))_{\Omega\setminus\Gamma_C} + (\boldsymbol{\sigma}_{\text{lin}}^{k-1}(\boldsymbol{\delta}_h^k), \boldsymbol{\delta}_h^v)_{\Gamma_C} = (\boldsymbol{f}, \boldsymbol{v}_h)_{\Omega\setminus\Gamma_C} + (\boldsymbol{g}_N, \boldsymbol{v}_h)_{\Gamma_N} \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_h,$$

where $\boldsymbol{\delta}_h^k := -[\![\boldsymbol{u}_h^k]\!]$. In particular, the linear approximation is obtained by:

$$\boldsymbol{\sigma}_{\text{lin}}^{k-1}(\boldsymbol{\delta}_h^k) := \boldsymbol{\sigma}(\boldsymbol{\delta}_h^{k-1}) + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\delta}_h^{k-1})}{\partial \boldsymbol{\delta}_h^{k-1}} (\boldsymbol{\delta}_h^k - \boldsymbol{\delta}_h^{k-1}).$$

The term $\dfrac{\partial \boldsymbol{\sigma}(\boldsymbol{\delta}_h^{k-1})}{\partial \boldsymbol{\delta}_h^{k-1}}$ is usually call *tangential matrix*. We recall that in general the constitutive relation between the stress vector and the displacement jump can also depend on other state variables $\boldsymbol{\alpha}$ that take into account the history of the joint (e.g, plasticity or damage variables). It is thus essential to implement the constitutive relation in a way that make it possible to consider the evolution of these variables while computing $\boldsymbol{\sigma}(\boldsymbol{\delta}_h^{k-1})$ and the tangent matrix. The joint elements enable this kind of implementation.

Energetic formulation

Starting from the article of Francfort and Marigo [62] and following the thesis work [82], the idea is to introduce a surface energy defined on the interface Γ that depends on the displacement jump $\boldsymbol{\delta}$, and then to solve a minimization problem. In particular, the displacement field $\boldsymbol{u}$ of the domain Ω is the one that minimize the total energy of the system, which is given by

$$\begin{aligned} E_{\text{tot}}(\boldsymbol{u}) = E_{\text{tot}}(\boldsymbol{u}, \boldsymbol{\delta}) &= \Phi(\boldsymbol{u}) + \Psi(\boldsymbol{\delta}) - W^{\text{ext}}(\boldsymbol{u}) \\ &= \int_{\Omega\setminus\Gamma_C} \phi(\boldsymbol{u}) \, \mathrm{d}\Omega + \int_{\Gamma_C} \psi(\boldsymbol{\delta}) \, \mathrm{d}\Gamma - \int_{\Omega\setminus\Gamma_C} \boldsymbol{f}\boldsymbol{u} \, \mathrm{d}\Omega - \int_{\Gamma_N} \boldsymbol{g}_N \boldsymbol{u} \, \mathrm{d}\Gamma, \end{aligned} \tag{2.6}$$

where $\phi(\boldsymbol{u}) \coloneqq \frac{1}{2}\,\boldsymbol{\varepsilon}(\boldsymbol{u}) : \mathbb{E} : \boldsymbol{\varepsilon}(\boldsymbol{u})$ represents the elastic energy density on the domain Ω, and $\psi(\boldsymbol{\delta})$ the surface energy density on the interface Γ_C. This approach allows us to adapt the SGM formalism to the interface modeling. In particular, some vector state variables replace the tensor ones (e.g. the displacement jump $\boldsymbol{\delta}$ replaces the strain vector $\boldsymbol{\varepsilon}$), the surface energy density function ψ the free energy ϕ, and the dissipation potential φ is the indicator function of the convex reversibility domain. Then, the constitutive relation (2.2) can be obtained by derivation of the surface energy density function:

$$\boldsymbol{\sigma} = \frac{\partial \psi}{\partial \boldsymbol{\delta}},$$

and the evolution of the dissipative state variables (e.g. plasticity or damage) is determined by the normal flow rule.

Starting from the problem of finding the displacement $\boldsymbol{u}$ that minimizes the total energy $E_{\text{tot}}(\boldsymbol{u})$ and using the procedure described in [82, Section 3.I], one can recover the formulations (2.4) and (2.5).

2.2.2 Standard mechanical tests

The real behavior of joints can be determined through some mechanical in situ or laboratory tests. The latter are more common since they are generally small dimension tests performed on samples which are obtained from dams through core drilling or manufactured in laboratory. In this subsection we briefly describe the main standard tests: shear and traction (or tensile) test.

Shear tests

Shear tests can be performed to analyze the shear behavior of joints under constant compression during monotonic or cycling loading. After enforcing a normal stress σ_n on the joint, the shear test consists in applying a tangential displacement $\boldsymbol{\delta}_t$ and measuring the shear stress $\boldsymbol{\sigma}_t$ and the normal displacement δ_n. Figure 2.5 shows the typical curves of monotonic shear tests with fixed compression. Initially, the shear stress $\boldsymbol{\sigma}_t$ increases up to a maximum value, and then there is a softening evolution reaching asymptotically a residual value. These maximum and residual values depend on the compression σ_n (see Figure 2.6) and on the level of asperities of the joint (see e.g. the plots on the top of Figure 2.5). Notice that, performing shear tests with different levels of compression and identifying these two values, one can have an indication of the shape of the reversibility elastic domain, Figure 2.6. Regarding the evolution of the normal displacement, we can observe the presence of dilatancy with asymptotic stabilization. For further discussion and analysis of monotonic and cycling shear tests on joints we refer to [45, 99, 98, 56, 77] and reference therein.

Traction tests

Traction tests enable the determination of the tensile strength of joints [117, 38, 98]. During a traction test, the normal displacement δ_n is controlled while the tangential component $\boldsymbol{\delta}_t$ is maintained constant, and we measure the normal displacement. In general, during a traction test we can distinguish three phases: at first, the tensile stress increases up to reach

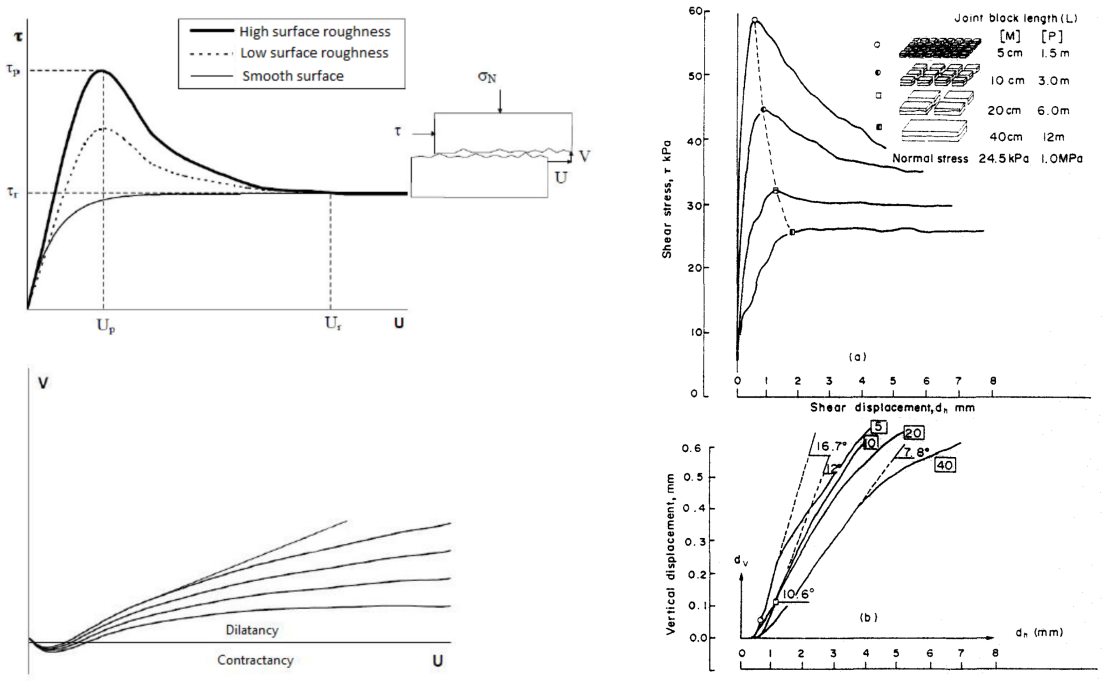


Figure 2.5 – Typical curves of shear tests for joints: evolution of the shear stress (*top*) and of the normal displacement (*bottom*); cf. [56] (*left*) and [9] (*right*).

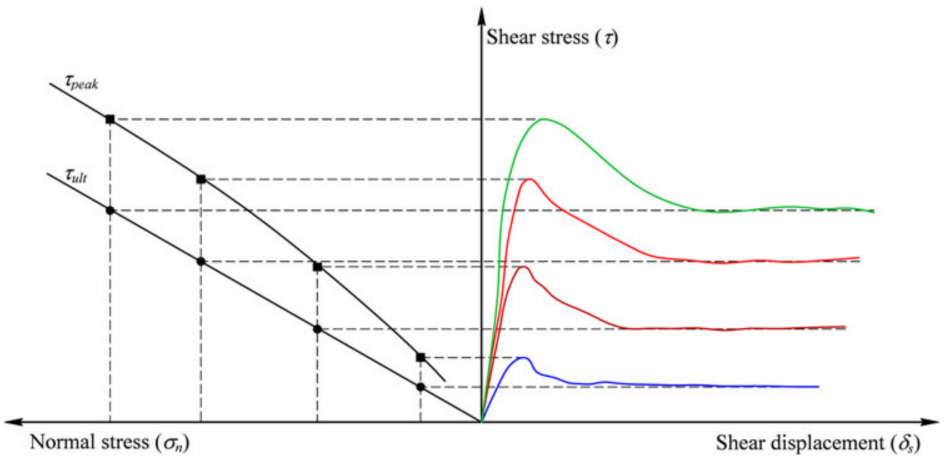


Figure 2.6 – Maximum and residual value of the shear test for different values of compression; cf. [99].

its maximum value which is called *ultimate tensile strength*; then, σ_n decreases and finally reaches 0. Schematically, this behavior can be observed in bilinear CZM of Figure 2.2c.

2.2.3 Existing constitutive relations in `code_aster`

At the moment, two constitute relations are available in the simulation tool `code_aster`. They can be used to perform static and quasi–static studies and describe two different phenomena: `JOINT_MECA_RUPT` for rupture in traction, and `JOINT_MECA_FROT` for friction in shear [49]. Both of them leave the framework of SGM, as in the first one the tangential stress is not obtained from a surface energy density, and in the second the flow rule for plasticity does not follow the normality rule. These relations are the starting point for the introduction of a new unified model for joints.

`JOINT_MECA_RUPT`

This first constitutive relation introduces a nonlinear dependence between $\boldsymbol{\delta}$ and $\boldsymbol{\sigma}$, but it leaves the framework of SGM and of variational formulation since the stress is not the derivative of a surface energy density function. Indeed, while the normal part of the stress σ_n is obtained from a energy function ψ_n by derivation, the tangential one $\boldsymbol{\sigma_t}$ is given explicitly and it depends on the normal opening δ_n. The normal surface energy density function ψ_n distinguishes three possible situations depending on the value of δ_n: contact in compression, linear evolution for small positive displacements, and dissipative evolution for displacements bigger than a threshold value. The main mechanical parameters of this relation are:

- the normal and tangential rigidity coefficient K_n and K_t,
- the maximum value of the normal stress $\sigma_{\max}$,
- a penalization parameter for compressions P_{cont},
- a penalization parameter for tractions P_{rupt}
- a parameter $\alpha \in [0, 2]$ related to the roughness of the joint.

The irreversibility of evolution is governed by a threshold internal variable κ, that indicates the maximum value of normal opening reached in the dissipative domain. It is initialized by $\kappa_0 := \sigma_{\max}/K_n$ (i.e., value of normal displacement when $\sigma_n = \sigma_{\max}$), and, during an incremental evolution, its value is updated by $\kappa^+ = \max\{\kappa^-, \delta_n^+\}$.

In this model, the normal surface energy density function is

$$\psi_n(\delta_n) = H(-\delta_n)\psi_n^{\text{con}}(\delta_n) + H(\delta_n)H(\kappa - \delta_n)\psi_n^{\text{lin}}(\delta_n) + H(\delta_n - \kappa)\psi_n^{\text{dis}}(\delta_n), \tag{2.7}$$

where $H(x) := \mathbb{1}_{x>0}$ is the Heaviside step function, and

$$\psi_n^{\text{con}}(\delta_n) = \sigma_{\max}\left(1 + \frac{1}{P_{\text{rupt}}}\right)\frac{\kappa - \kappa_0}{2} + P_{\text{cont}}K_n\frac{\delta_n^2}{2} \tag{2.8}$$

$$\psi_n^{\text{lin}}(\delta_n) = \sigma_{\max}\left(1 + \frac{1}{P_{\text{rupt}}}\right)\frac{\kappa - \kappa_0}{2} + \left[\frac{\sigma_{\max}}{\kappa}\left(1 + \frac{1}{P_{\text{rupt}}}\right) - \frac{K_n}{P_{\text{rupt}}}\right]\frac{\delta_n^2}{2} \tag{2.9}$$

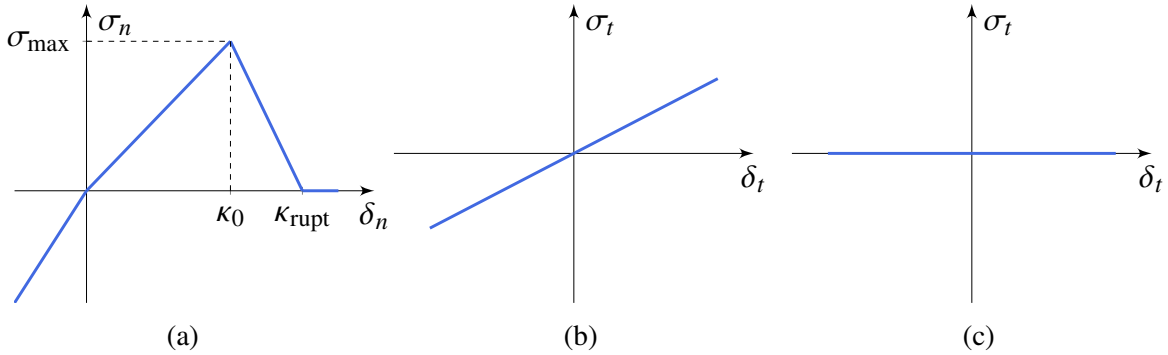


Figure 2.7 – Illustration of the evolution of normal stress in compression and traction (*left*), and of tangential stress in compression or for a partially open joint (*middle*) and for a completely open joint (*right*).

$$\psi_n^{\text{dis}}(\delta_n) = \begin{cases} \sigma_{\max}\left(1+\dfrac{1}{P_{\text{rupt}}}\right)\left(\delta_n - \dfrac{\kappa_0}{2}\right) - \dfrac{K_n}{2P_{\text{rupt}}}\delta_n^2 & \text{if } \delta_n < \kappa_{\text{rupt}}, \\ \sigma_{\max}^2 \dfrac{1+P_{\text{rupt}}}{K_n} & \delta_n \geq \kappa_{\text{rupt}}. \end{cases} \tag{2.10}$$

In the last expression, we have introduced a threshold value κ_{rupt} that identifies the complete rupture of the joint. Its values is $\kappa_{\text{rupt}} := \sigma_{\max}(1+P_{\text{rupt}})/K_n$. Thanks to (2.7)–(2.10) we can easily obtain the expression of the normal stress which is piecewise linear, Figure 2.7a:

$$\sigma_n(\delta_n) = \begin{cases} P_{\text{cont}} K_n \delta_n & \text{if } \delta_n < 0, \\ \left[\dfrac{\sigma_{\max}}{\kappa}\left(1+\dfrac{1}{P_{\text{rupt}}}\right) - \dfrac{K_n}{P_{\text{rupt}}}\right]\delta_n & \text{if } 0 \leq \delta_n < \kappa, \\ \sigma_{\max}\left(1+\dfrac{1}{P_{\text{rupt}}}\right) - \dfrac{K_n}{P_{\text{rupt}}}\delta_n & \text{if } \kappa \leq \delta_n < \kappa_{\text{rupt}}, \\ 0 & \text{if } \delta_n \geq \kappa_{\text{rupt}}. \end{cases}$$

Finally, the tangential stress is defined by

$$\boldsymbol{\sigma_t}(\boldsymbol{\delta_t}, \delta_n) = \begin{cases} K_t(\boldsymbol{\delta_t} - \boldsymbol{\delta}_{\text{shift}}) & \text{if } \delta_n < 0, \\ \left(1 - \dfrac{\delta_n}{\kappa_{\text{rupt}}^{\tan}}\right) K_t(\boldsymbol{\delta_t} - \boldsymbol{\delta}_{\text{shift}}) & \text{if } 0 \leq \delta_n < \kappa_{\text{rupt}}^{\tan}, \\ 0 & \text{if } \delta_n \geq \kappa_{\text{rupt}}^{\tan}, \end{cases}$$

where $\kappa_{\text{rupt}}^{\tan} := \kappa_{\text{rupt}} \tan\left(\dfrac{\alpha\pi}{4}\right)$, and $\boldsymbol{\delta}_{\text{shift}}$ is an history variable that corresponds to the last value of $\boldsymbol{\delta_t}$ for which the joint was completely open, i.e., $\delta_n \geq \kappa_{\text{rupt}}^{\tan}$.

JOINT_MECA_FROT

This constitutive relation consists in the Coulomb standard law described with an non-associative elastoplastic model, and makes it possible to represent the phase of friction in shear. The mechanical parameters are:

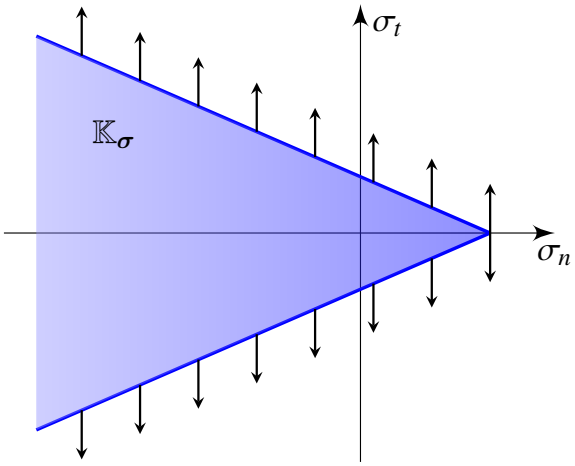


Figure 2.8 – Reversibility domain in the stress space with the illustration of the flow rule for the evolution of the plasticity for $d = 2$.

- the normal and tangential rigidity coefficient K_n and K_t,
- the friction coefficient μ,
- the adhesion c,
- the isotropic hardening coefficient K.

There are two state variables: the displacement jump $\boldsymbol{\delta} \in \mathbb{R}^d$ and its plastic part $\boldsymbol{p} \in \mathbb{R}^d$. An important ingredient of the elastoplastic models is the reversibility domain in the stress space $\mathbb{K}_{\boldsymbol{\sigma}}$. In this constitutive relation a criterion of Drucker–Prager type [50] is used to define a conic loading surface. In addition, the evolution of plasticity is characterized by a "vertical" flow rule, in the sense that only the tangential component of plasticity can evolve. Figure 2.8 shows the reversibility domain and the plasticity flow rule for $d = 2$, whereas the case $d = 3$ can be obtained by rotation around the normal axis. Their expressions are, respectively,

$$f(\boldsymbol{\sigma}, \lambda) \coloneqq \|\boldsymbol{\sigma_t}\| + \mu\sigma_n - c - K\lambda \leq 0,$$

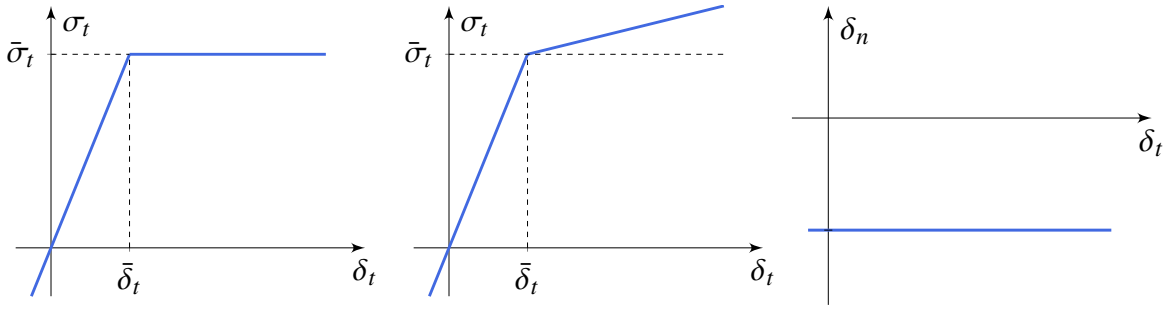
and

$$\begin{cases} \dot{\delta}_n = 0 \\ \dot{\boldsymbol{\delta}}_{\boldsymbol{t}} = \dot{\lambda}\dfrac{\boldsymbol{\sigma_t}}{\|\boldsymbol{\sigma_t}\|} \end{cases} \qquad \text{with} \qquad \begin{cases} f(\boldsymbol{\sigma}, \lambda) \cdot \dot{\lambda} = 0 \\ \dot{\lambda} \geq 0 \end{cases}$$

where $\|\cdot\|$ is the Euclidean norm of the space $\mathbb{R}^{d-1}$, λ is a plastic multiplier, and the dot superscript denotes the time derivative. Moreover, the value of the maximum tensile strength, i.e, maximum possible value of the normal stress σ_n, is automatically fixed by the values of cohesion and friction coefficient: $R_{\text{ten}} = c/\mu$.

The constitutive relation `JOINT_MECA_FROT` can be obtained by derivation from the following surface energy density function:

$$\psi(\delta_n, \boldsymbol{\delta_t}, \boldsymbol{p_t}) = \psi_n(\delta_n) + K_t\frac{\|\boldsymbol{\delta_t} - \boldsymbol{p_t}\|^2}{2},$$



(a) (b) (c)

Figure 2.9 – Illustration of the evolution during a shear test with fixed compression: tangential stress with $K = 0$ (*left*) and $K > 0$ (*middle*), and normal displacement (*right*).

where

$$\psi_n(\delta_n) := \begin{cases} K_n \dfrac{\delta_n^2}{2} & \text{if } \delta_n \leq \dfrac{R_{\text{ten}}}{K_n}, \\ R_{\text{ten}}\delta_n - \dfrac{R_{\text{ten}}^2}{2K_n} & \text{otherwise.} \end{cases}$$

As a consequence, we achieve

$$\begin{cases} \sigma_n = \min\{K_n\delta_n, R_{\text{ten}}\} & \text{(2.11a)} \\ \boldsymbol{\sigma_t} = K_t(\boldsymbol{\delta_t} - \boldsymbol{p}_t) & \text{(2.11b)} \end{cases}$$

Figure 2.9 shows the curves of a shear test with fixed compression.

2.3 Coupling plasticity and damage

Our goal is to develop a new constitutive law in a well-defined framework, that provides good numerical properties, represents the phenomena observed experimentally, possibly uses the parameters of the existing laws, and introduces as few new parameters as possible. This section is devoted to the description of a model for joints with cohesive forces which couples friction and rupture, via state variables representing plasticity and damage. The main idea is to adapt the model proposed by [94] for geomaterials to the joint context.

2.3.1 Notation and assumptions

We recall that the main building blocks to construct a GSM in the context of elastoplastic CZM for joints are:

1. choice of the state variables, including some dissipative variables;
2. choice of a surface energy density function ψ, which gives by derivation the constitutive relation and the other thermodynamical forces;
3. choice of a convex dissipation potential φ, which is the indicator function of a reversibility domain and which leads to a normal flow rule for the dissipative state variables.

Notation	Definition
$\boldsymbol{\delta} = (\delta_n, \delta_{t_1}, \delta_{t_2})^\top$	Displacement jump
$\boldsymbol{p} = (p_n, p_{t_1}, p_{t_2})^\top$	Plasticity vector
$\alpha \in [0, 1]$	Damage scalar variable
$\psi = \psi(\boldsymbol{\delta}, \boldsymbol{p}, \alpha)$	Surface energy density function
$\boldsymbol{\sigma} = (\sigma_n, \sigma_{t_1}, \sigma_{t_2})^\top$	Cohesive stress vector
$\boldsymbol{X} = (X_n, X_{t_1}, X_{t_2})^\top$	Plasticity generalized force
Y	Damage generalized force
$\|\cdot\|$	Euclidean norm of the space $\mathbb{R}^{d-1}$

Table 2.1 – Notation for the model coupling plasticity and damage

Table 2.1 summarizes the main notation of our model. In addition to the two state variable of `JOINT_MECA_FROT` already introduced in the previous section (i.e., the displacement jump $\boldsymbol{\delta}$ and its plastic component $\boldsymbol{p}$), we add a scalar state variable α which represents the evolution of damage. In particular, $\alpha = 0$ corresponds to the joint in its sound state, whereas $\alpha = 1$ to the completely fractured joint. The surface energy density function is $\psi(\boldsymbol{\delta}, \boldsymbol{p}, \alpha)$, and it is supposed to be at least differentiable and to be convex with respect to the three state variables $\boldsymbol{\delta}$, $\boldsymbol{p}$ and α separately. The three thermodynamical forces associated to the state variables are then obtained by differentiation:

$$\boldsymbol{\sigma} = \frac{\partial \psi}{\partial \boldsymbol{\delta}}, \qquad \boldsymbol{X} = -\frac{\partial \psi}{\partial \boldsymbol{p}} \qquad \text{and} \qquad Y = -\frac{\partial \psi}{\partial \alpha}.$$

In the following we will call $\boldsymbol{X}$ and Y also *generalized forces*.

We describe separately the evolution of plasticity and of damage. In the space of thermodynamical forces related to plasticity, the reversibility elastic domain is a fixed non-empty convex set. This domain is denoted by $\mathbb{K}_{\boldsymbol{X}}$ and is determined by a convex function $f_{\boldsymbol{X}} : \mathbb{R}^d \to \mathbb{R}$:

$$\mathbb{K}_{\boldsymbol{X}} := \left\{ \boldsymbol{X}^* \in \mathbb{R}^d \ : \ f_{\boldsymbol{X}}(\boldsymbol{X}^*) \leq 0 \right\}.$$

The plasticity criterion is simply $f_{\boldsymbol{X}}(\boldsymbol{X}) \leq 0$, i.e., $\boldsymbol{X} \in \mathbb{K}_{\boldsymbol{X}}$, and the plastic vector $\boldsymbol{p}$ can only evolve when $f_{\boldsymbol{X}}(\boldsymbol{X}) = 0$ is satisfied, i.e, when $\boldsymbol{X} \in \partial\mathbb{K}_{\boldsymbol{X}}$. Moreover, the normal flow rule establishes that the velocity of plastic displacements $\dot{\boldsymbol{p}}$ belongs to the cone of outer normals to $\partial\mathbb{K}_{\boldsymbol{X}}$ which is defined by

$$N(\boldsymbol{X}) := \left\{ \boldsymbol{g} \in \mathbb{R}^d \ : \ (\boldsymbol{X} - \boldsymbol{X}^*)\, \boldsymbol{g} \geq 0 \ \ \forall \boldsymbol{X}^* \in \mathbb{K}_{\boldsymbol{X}} \right\} \qquad \forall \boldsymbol{X} \in \partial\mathbb{K}_{\boldsymbol{X}}.$$

In particular, if the convex set $\mathbb{K}_{\boldsymbol{X}}$ has a smooth boundary, we obtain

$$\dot{\boldsymbol{p}} = \dot{\lambda}\, \frac{\partial f_{\boldsymbol{X}}}{\partial \boldsymbol{X}} \qquad \text{with} \quad \begin{cases} f_{\boldsymbol{X}}(\boldsymbol{X})\, \dot{\lambda} = 0 \\ \dot{\lambda} \geq 0 \end{cases} \tag{2.12}$$

where λ is a plastic multiplier.

Remark 2.3 (Normal flow rule and maximum plastic work principle of Hill)**.** *The maximum plastic work principle of Hill is automatically satisfied assuming the normality of the flow rule. Indeed, given* $\boldsymbol{X} \in \partial\mathbb{K}_{\boldsymbol{X}}$ *we have* $\dot{\boldsymbol{p}} \in N(\boldsymbol{X})$*, i.e.,*

$$(\boldsymbol{X} - \boldsymbol{X}^*)\dot{\boldsymbol{p}} \geq 0 \qquad \forall \boldsymbol{X}^* \in \mathbb{K}_{\boldsymbol{X}},$$

which is the maximum plastic work principle of Hill.

Now, we consider the damage evolution. The damage criterion reads

$$Y \leq D'(\alpha) \tag{2.13}$$

where the superscript $'$ denotes the derivative with respect to α, and $D(\alpha)$ is a function that represents the surface damage dissipated energy. By defining a reversibility domain in the space of thermodynamical forces related to damage

$$\mathbb{K}_Y := \left\{Y^* \ : \ Y^* \leq D'(\alpha)\right\}, \tag{2.14}$$

we can describe the evolution of damage with the same formalism as plasticity: the damage criterion is $Y \in \mathbb{K}_Y$, and the damage can only evolve when $Y \in \partial\mathbb{K}_Y$, i.e., when $Y = D'(\alpha)$. This is summarized by the following consistency relation

$$(Y - D'(\alpha))\,\dot{\alpha} = 0. \tag{2.15}$$

Finally, the normal flow rule leads here to a damage irreversibility condition: $\dot{\alpha} \geq 0$.

2.3.2 Choice of surface energy density function

In the model we present in this chapter, the energy density ψ is the sum of a term representing the elastic energy, and a cinematic hardening term which couples plasticity and damage:

$$\begin{aligned}\psi(\boldsymbol{\delta}, \boldsymbol{p}, \alpha) &= \psi_{\text{ela}}(\boldsymbol{\delta}, \boldsymbol{p}) + \psi_{\text{har}}(\boldsymbol{p}, \alpha) \\ &= \frac{1}{2}\,(\boldsymbol{\delta} - \boldsymbol{p})^\top \mathbb{E}\,(\boldsymbol{\delta} - \boldsymbol{p}) + \frac{1}{2}\,\boldsymbol{p}^\top \mathbb{H}(\alpha)\,\boldsymbol{p},\end{aligned} \tag{2.16}$$

where

$$\mathbb{E} = \begin{pmatrix} K_n & 0 & 0 \\ 0 & K_t & 0 \\ 0 & 0 & K_t \end{pmatrix} \qquad \text{and} \qquad \mathbb{H}(\alpha) = \begin{pmatrix} A_n(\alpha) & 0 & 0 \\ 0 & A_t(\alpha) & 0 \\ 0 & 0 & A_t(\alpha) \end{pmatrix}, \tag{2.17}$$

K_n and K_t are respectively the normal and tangential rigidity coefficients, $A_n(\alpha)$ and $A_t(\alpha)$ are some positive functions on $\alpha \in [0, 1]$. In particular, using the decomposition into normal and tangential components, we have

$$\psi(\boldsymbol{\delta}, \boldsymbol{p}, \alpha) = K_n \frac{(\delta_n - p_n)^2}{2} + K_t \frac{\|\boldsymbol{\delta_t} - \boldsymbol{p_t}\|^2}{2} + A_n(\alpha)\frac{p_n^2}{2} + A_t(\alpha)\frac{\|\boldsymbol{p_t}\|^2}{2}. \tag{2.18}$$

Motivated by the paper [94], we consider the following family of functions

$$A_n(\alpha) = B_n \frac{(1-\alpha)^{m_1}}{\alpha^{m_2}} \qquad \text{and} \qquad A_t(\alpha) = B_t \frac{(1-\alpha)^{m_1}}{\alpha^{m_2}}, \tag{2.19}$$

where $B_n, B_t > 0$, and $0 < m_2 < 1 < m_1$. B_n and B_t control the influence of the hardening term to the surface energy density ψ, m_1 control the behavior for $\alpha \sim 1$, and m_2 the behavior for $\alpha \sim 0$. Notice that

$$A_s'(\alpha) = B_s\left((m_2 - m_1)\alpha - m_2\right)\frac{(1-\alpha)^{m_1-1}}{\alpha^{m_2+1}} \tag{2.20}$$

and

$$A_s''(\alpha) = B_s\big((m_1 - m_2)(m_1 - m_2 - 1)\alpha^2 + 2m_2(m_1 - m_2 - 1)\alpha + \\ + m_2(m_2+1)\big)\frac{(1-\alpha)^{m_1-2}}{\alpha^{m_2+2}}, \tag{2.21}$$

for $s \in \{n, t\}$. It is easy to check that

$$A_s(\alpha) \geq 0, \qquad A_s'(\alpha) \leq 0, \qquad A_s''(\alpha) \geq 0 \qquad\qquad \forall \alpha \in [0,1] \tag{2.22}$$

for $s \in \{n, t\}$.

For the following proposition, we introduce the notation

$$S(\alpha) := \frac{(1-\alpha)^{m_1}}{\alpha^{m_2}} \qquad \text{and} \qquad R(\alpha) := \frac{2(S'(\alpha))^2}{S''(\alpha)} - S(\alpha). \tag{2.23}$$

Proposition 2.4 (Convexity properties of ψ)**.** *Let $\psi(\boldsymbol{\delta}, \boldsymbol{p}, \alpha)$ be the surface energy density function defined by (2.18). Then,*

1) $\psi(\boldsymbol{\delta}, \boldsymbol{p}, \alpha)$ is convex with respect to $\boldsymbol{\delta}$, $\boldsymbol{p}$ and α separately.

2) $\psi(\boldsymbol{\delta}, \boldsymbol{p}, \alpha)$ is convex with respect to $(\boldsymbol{p}, \alpha)$ if

$$K_n \geq B_n \max_{\alpha\in[0,1]} R(\alpha) \qquad \text{and} \qquad K_t \geq B_t \max_{\alpha\in[0,1]} R(\alpha). \tag{2.24}$$

Proof. 1) The convexity with respect to $\boldsymbol{\delta}$ and $\boldsymbol{p}$ is straightforward and comes from the positivity of the rigidity coefficients K_s and of the damage functions $A_s(\alpha)$, $s \in \{n, t\}$. Indeed, it is sufficient to verify that $\mathbb{E}$ and $\mathbb{E} + \mathbb{H}(\alpha)$ defined by (2.17) are positive definite for all $\alpha \in [0, 1]$. In order to prove also the convexity with respect to α it is sufficient to check the sign of the second derivative of ψ with respect to the damage variable α:

$$\frac{\partial^2\psi}{\partial\alpha^2} = A_n''(\alpha)\frac{p_n^2}{2} + A_t''(\alpha)\frac{\|\boldsymbol{p_t}\|^2}{2} \geq 0 \qquad\qquad \forall \boldsymbol{p} \in \mathbb{R}^d, \forall \alpha \in [0,1],$$

since $A_s(\alpha)$ is convex on the interval $[0, 1]$, $s \in \{n, t\}$, by (2.22).

2) The Hessian matrix of $\psi(\boldsymbol{\delta}, \boldsymbol{p}, \alpha)$ with respect to $(\boldsymbol{p}, \alpha)$ is

$$\mathbb{H}_{(\boldsymbol{p},\alpha)} := \begin{pmatrix} K_n + A_n(\alpha) & 0 & A_n'(\alpha)p_n \\ 0 & (K_t + A_t(\alpha))\mathbf{I}_2 & A_t'(\alpha)\boldsymbol{p_t} \\ A_n'(\alpha)p_n & A_t'(\alpha)\boldsymbol{p_t} & A_n''(\alpha)\dfrac{p_n^2}{2} + A_t''(\alpha)\dfrac{\|\boldsymbol{p_t}\|^2}{2} \end{pmatrix}.$$

Here, $\mathbf{I}_2$ is the identity matrix of the space $\mathbb{R}^{2\times 2}$. By applying an extension of the Sylvester's criterion [100], $\mathbb{H}_{(\boldsymbol{p},\alpha)}$ is semidefinite if and only if all its principal minors are nonnegative. In particular, all the diagonal elements are nonnegatives (since $K_n > 0$ and $K_t > 0$, and $A_n(\alpha) \leq 0$ and $A_t(\alpha)$ for any $\alpha \in [0,1]$ by (2.22)), and the determinant of the Hessian matrix $\mathbb{H}_{(\boldsymbol{p},\alpha)}$ is

$$\det\left(\mathbb{H}_{(\boldsymbol{p},\alpha)}\right) = (K_t + A_t(\alpha))\left[(K_t + A_t(\alpha))\left((K_n + A_n(\alpha))\,A_n''(\alpha) - 2\left(A_n'(\alpha)\right)^2\right)\frac{p_n^2}{2}\right.$$
$$\left. + (K_n + A_n(\alpha))\left((K_t + A_t(\alpha))\,A_t''(\alpha) - 2\left(A_t'(\alpha)\right)^2\right)\frac{\|\boldsymbol{p_t}\|^2}{2}\right],$$

which is nonnegative if for every $\alpha \in [0,1]$

$$(K_s + A_s(\alpha))\,A_s''(\alpha) - 2\left(A_s'(\alpha)\right)^2 \geq 0 \qquad s \in \{n,t\},$$

or equivalently

$$K_s \geq \frac{2(A_s'(\alpha))^2}{A_s''(\alpha)} - A_s(\alpha) = B_s R(\alpha) \qquad s \in \{n,t\}.$$

Notice that $R(\alpha)$ is a continuous function in $(0,1)$ that depends only on the parameters m_1 and m_2. Moreover, one can easily show that $R(0) = -\infty$, $R(1) = 0$, $R(\alpha)$ is positive near 1, and it has exactly one root in $(0,1)$. As a consequence,

$$0 < \max_{\alpha\in[0,1]} R(\alpha) < +\infty, \tag{2.25}$$

and, assuming (2.24), we get $\det(\mathbb{H}_\psi) \geq 0$. In addition, on can easily check that the assumptions (2.24) ensure also the nonnegativity of all the others principal minors. □

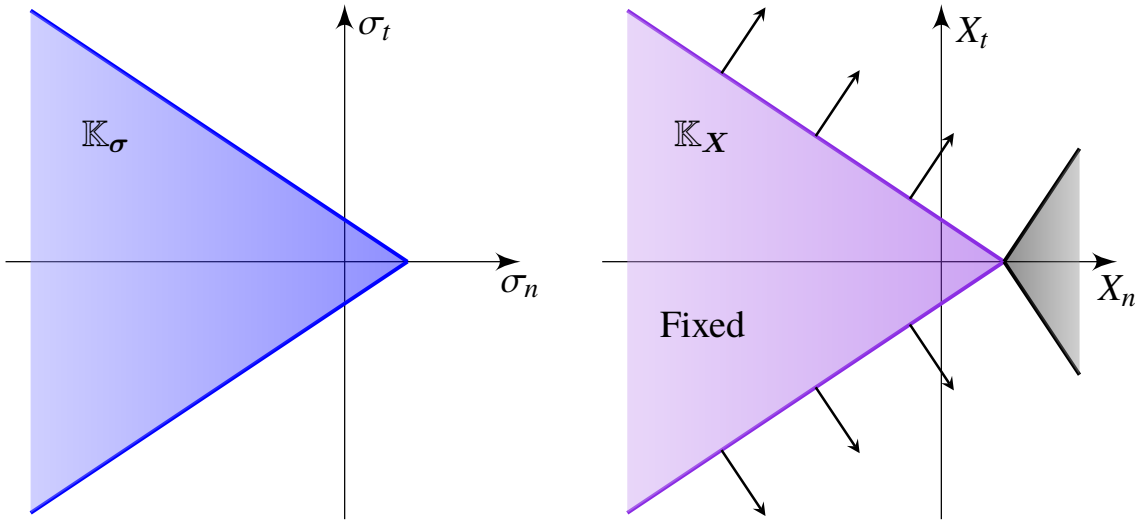
Remark 2.5 (Convexity properties of ψ_{ela} and ψ_{har})**.** *While the elastic part of the surface energy density ψ (2.16), i.e., ψ_{ela}, is convex for any value of $(\boldsymbol{\delta}, \boldsymbol{p}, \alpha)$, the hardening part ψ_{har} is convex with respect to $(\boldsymbol{p},\alpha)$ only for $0 \leq \alpha \leq \bar{\alpha}$, where*

$$\bar{\alpha} = \frac{1}{m_1 - m_2}\left(-m_2 + \sqrt{\frac{m_1 m_2}{m_1 - m_2 + 1}}\right). \tag{2.26}$$

Indeed, the determinant of the Hessian matrix of ψ_{har} with respect to $(\boldsymbol{p},\alpha)$ is

$$-\frac{1}{2}B_n B_t^2\,(S(\alpha))^2\,S''(\alpha)R(\alpha)\left(B_n p_n + B_t\|\boldsymbol{p_t}\|^2\right)$$

which is nonnegative if and only if $R(\alpha) \leq 0$. In the interval $[0,1]$, this holds if and only if $\alpha \in [0,\bar{\alpha}]$. This is the reason why, for the second statement of Proposition 2.4, we cannot simply consider ψ_{ela} and ψ_{har} separately. As we will se later (see Remark 2.9), $\bar{\alpha}$ represents the value of damage that identifies the peak of stress in a shear or in a traction evolution. Finally, it is possible to show that ψ is fully convex with respect to $(\boldsymbol{\delta},\boldsymbol{p},\alpha)$ only inside the interval $\alpha \in [0,\bar{\alpha}]$, i.e., we loose full convexity during the softening phase, but we do preserve $(\boldsymbol{p},\alpha)$ convexity under the conditions (2.24).



(a) Initial $\mathbb{K}_{\boldsymbol{\sigma}}$ (b) $\mathbb{K}_{\boldsymbol{X}}$ and plasticity flow rule

Figure 2.10 – Initial reversibility domain $\mathbb{K}_{\boldsymbol{\sigma}}$ determined by (*left*), and reversibility domain $\mathbb{K}_{\boldsymbol{X}}$ determined by (2.29) with the representation of the flow rule for plasticity (*right*).

By derivation, we obtain the expression for the three thermodynamical forces:

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_n \\ \boldsymbol{\sigma_t} \end{pmatrix} = \begin{pmatrix} K_n(\delta_n - p_n) \\ K_t(\boldsymbol{\delta_t} - \boldsymbol{p_t}) \end{pmatrix}, \tag{2.27}$$

$$\boldsymbol{X} = \begin{pmatrix} X_n \\ \boldsymbol{X_t} \end{pmatrix} = \begin{pmatrix} K_n(\delta_n - p_n) - A_n(\alpha)p_n \\ K_t(\boldsymbol{\delta_t} - \boldsymbol{p_t}) - A_t(\alpha)\boldsymbol{p_t} \end{pmatrix},$$

and

$$Y = -A_n'(\alpha)\frac{p_n^2}{2} - A_t'(\alpha)\frac{\|\boldsymbol{p_t}\|^2}{2}.$$

In particular, we notice that the cohesive stress and the thermodynamical force associated with plasticity are connected through the following relation:

$$\boldsymbol{\sigma} = \boldsymbol{X} + \mathbb{H}(\alpha)\boldsymbol{p}. \tag{2.28}$$

2.3.3 Choice of plasticity and damage criteria

In literature, different criteria have been proposed for describing the behavior of materials, notably the Mohr–Coulomb criterion, the Drucker–Prager criterion, or the Hoek–Brown criterion. In the spirit of [94], we assume that the reversibility domain $\mathbb{K}_{\boldsymbol{X}}$ is determined by the Drucker–Prager criterion:

$$f_{\boldsymbol{X}}(\boldsymbol{X}) = \|\boldsymbol{X_t}\| + \mu X_n - \bar{c} \leq 0, \tag{2.29}$$

where $\mu > 0$ is the friction coefficient and $\bar{c} \geq 0$ represents the residual cohesion. As a consequence, $\mathbb{K}_{\boldsymbol{X}}$ is an unbounded cone with axis of symmetry $\boldsymbol{X_t} = \boldsymbol{0}$ and vertex in

$(\bar{c}/\mu, \mathbf{0})$, Figure 2.10b, and we assume that it remains fixed during all types of evolution. Outside of this singular point the flow rule (2.12) becomes

$$\begin{cases} \dot{p}_n = \mu\dot{\lambda}, \\ \dot{\boldsymbol{p}}_{\boldsymbol{t}} = \dot{\lambda}\dfrac{\boldsymbol{X}_{\boldsymbol{t}}}{\|\boldsymbol{X}_{\boldsymbol{t}}\|}, \end{cases} \qquad \text{with} \qquad \begin{cases} f_{\boldsymbol{X}}(\boldsymbol{X})\,\dot{\lambda} = 0, \\ \dot{\lambda} \geq 0, \end{cases} \tag{2.30}$$

whereas in the vertex $(c/\mu, \mathbf{0})$ the rate of the plastic displacement $\dot{\boldsymbol{p}}$ belongs to the cone of the outer normal vectors:

$$\begin{cases} \dot{p}_n = \mu\dot{\lambda}, \\ \dot{\boldsymbol{p}}_{\boldsymbol{t}} = \dot{\lambda}\boldsymbol{\beta}, \end{cases} \qquad \text{with} \qquad \begin{cases} f_{\boldsymbol{X}}(\boldsymbol{X})\,\dot{\lambda} = 0, \\ \dot{\lambda} \geq 0, \\ \|\boldsymbol{\beta}\| \leq 1, \end{cases} \tag{2.31}$$

Figure 2.10b. Furthermore, using the relation (2.28), we obtain the expression of the criterion in the stress space, which defines the reversibility domain $\mathbb{K}_{\boldsymbol{\sigma}}$, Figure 2.10a:

$$f_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}_{\boldsymbol{t}} + A_t(\alpha)\boldsymbol{p}_{\boldsymbol{t}}\| + \mu(\sigma_n + A_n(\alpha)p_n) - c \leq 0. \tag{2.32}$$

In particular, at the beginning of the evolution, i.e., when $\boldsymbol{p} = \mathbf{0}$ and $\alpha = 0$, $\mathbb{K}_{\boldsymbol{\sigma}}$ coincides with $\mathbb{K}_{\boldsymbol{X}}$, and, since the latter is fixed by hypothesis, during the evolution of plasticity and damage it shifts along the direction of the vector $\mathbb{H}(\alpha)\boldsymbol{p}$.

Finally, by a possible change of variable of damage, one can assume that the surface damage dissipated energy is linear in the damage variable, i.e., $D(\alpha) = D_1\alpha$ for all $\alpha[0, 1]$, where $D_1 > 0$. Then, the damage criterion (2.13) simply becomes

$$Y \leq D_1, \tag{2.33}$$

and we can rewrite (2.14) and (2.15) as follows:

$$\mathbb{K}_Y = \left\{Y^* \ : \ Y^* \leq D_1\right\} \qquad \text{and} \qquad (Y - D_1)\dot{\alpha} = 0 \tag{2.34}$$

The following proposition establishes some properties regarding the evolution of plasticity and damage, and, in particular, a condition under which they evolve simultaneously.

Proposition 2.6 (Simultaneous evolution of plasticity and damage)**.** *Assume that the surface energy density ψ is defined by (2.18), that the damage functions $A_n(\alpha)$ and $A_t(\alpha)$ are defined by (2.19), that $f_{\boldsymbol{X}}$ is defined by (2.29), and that $D(\alpha) = D_1\alpha$.*

1) Plasticity and damage start at the same time.

2) If $\dot{\alpha} > 0$, then $\dot{\boldsymbol{p}} \neq \mathbf{0}$.

3) If $\mu^2 B_n \geq B_t$, then plasticity and damage evolve simultaneously.

Proof. 1) Initially, let us assume that damage can start before of plasticity. By the consistency criterion for damage (2.34), we have $Y = D_1$, and since $\boldsymbol{p} = \mathbf{0}$

$$0 = -A_n'(\alpha)\frac{p_n^2}{2} - A_t'(\alpha)\frac{\|\boldsymbol{p}_{\boldsymbol{t}}\|^2}{2} = D_1,$$

which is a contradiction because $D_1 > 0$. Conversely, if plasticity can start before damage, then

$$D_1 \geq -A_n'(0)\frac{p_n^2}{2} - A_t'(0)\frac{\|\boldsymbol{p_t}\|^2}{2} = +\infty,$$

since, by definition, $A_s'(0) = -\infty$, $s \in \{n, t\}$. We conclude that plasticity and damage start at the same time.

2) Again, (2.34) implies that damage evolves only when $Y = D_1$, i.e., when, by derivation with respect of time

$$\left[A_n''(\alpha)\frac{p_n^2}{2} + A_t''(\alpha)\frac{\|\boldsymbol{p_t}\|^2}{2}\right]\dot{\alpha} = -A_n'(\alpha)p_n\dot{p}_n - A_t'(\alpha)\boldsymbol{p_t}\dot{\boldsymbol{p}}_{\boldsymbol{t}}.$$

Then, $\dot{\boldsymbol{p}} = \boldsymbol{0}$ implies $\dot{\alpha} = 0$, and we reach the thesis.

3) In order to conclude the proof, we show an incremental approach and we proceed by induction on the number of increments. Since we have already shown that plasticity and damage start at the same time, we assume that plasticity and damage have evolved simultaneously until the k-th step, and we denote with $\boldsymbol{\delta}^k$, $\boldsymbol{p}^k$ and $\alpha^k < 1$ the values of the state variables at this step. By (2.29) and (2.34), we have

$$f_{\boldsymbol{X}}(\boldsymbol{X}(\boldsymbol{\delta}^k, \boldsymbol{p}^k, \alpha^k)) \leq 0$$

and

$$-A_n'(\alpha^k)\frac{(p_n^k)^2}{2} - A_t'(\alpha^k)\frac{\|\boldsymbol{p_t^k}\|^2}{2} = D_1. \tag{2.35}$$

Now, we suppose that the displacement jump is increased by the increment $\Delta\boldsymbol{\delta}^k$, with $f(\boldsymbol{\delta}^k + \Delta\boldsymbol{\delta}^k, \boldsymbol{p}^k, \alpha^k) > 0$, i.e., either plasticity or damage has to evolve. If damage evolves without plasticity, i.e., $\Delta\alpha^k > 0$ and $\Delta\boldsymbol{p}^k = \boldsymbol{0}$, then by (2.35)

$$\begin{aligned} D_1 = &-\frac{1}{2}A_n'(\alpha^k + \Delta\alpha^k)(p_n^k)^2 - \frac{1}{2}A_t'(\alpha^k + \Delta\alpha^k)(p_t^k)^2 < \\ &< -\frac{1}{2}A_n'(\alpha^k)(p_n^k)^2 - \frac{1}{2}A_t'(\alpha^k)(p_t^k)^2 = D_1. \end{aligned}$$

Here, we use the fact that $A_n'(\alpha)$ and $A_t'(\alpha)$ are increasing functions. Conversely, assume that $\Delta\boldsymbol{p}^k \neq 0$ and $\Delta\alpha^k = 0$. Then, using again (2.35),

$$\begin{aligned} D_1 \geq &- A_n'(\alpha^k)\frac{(p_n^k + \Delta p_n^k)^2}{2} - A_t'(\alpha^k)\frac{\|\boldsymbol{p_t^k} + \Delta\boldsymbol{p_t^k}\|^2}{2} \\ &= D_1 - A_n'(\alpha^k)\frac{(\Delta p_n^k)^2}{2} - A_n'(\alpha^k)p_n^k\Delta p_n^k - A_t'(\alpha^k)\frac{\|\Delta\boldsymbol{p_t^k}\|^2}{2} - A_t'(\alpha^k)\boldsymbol{p_t^k}\Delta\boldsymbol{p_t^k}. \end{aligned}$$

Notice that the second and the forth terms are (strickly) positive. Let us focus on the remaining terms. By the normal flow rule for plasticity (2.30) and (2.31), we get

$$\begin{cases} \Delta p_n^k = \mu\Delta\lambda^k \\ p_n^k = \mu\lambda^k \end{cases} \quad \text{and} \quad \begin{cases} \Delta\boldsymbol{p_t^k} = \beta\Delta\lambda^k \text{ with } \|\beta\| \leq 1 \\ \|\boldsymbol{p_t^k}\| \leq \lambda^k \end{cases}$$

As a consequence,

$$\begin{aligned}-A_n'(\alpha^k)p_n^k\Delta p_n^k - A_t'(\alpha^k)\boldsymbol{p}_{\boldsymbol{t}}^k\Delta\boldsymbol{p}_{\boldsymbol{t}}^k &= -S'(\alpha^k)\Delta\lambda^k\left(B_n\mu p_n^k + B_t\boldsymbol{\beta}\boldsymbol{p}_{\boldsymbol{t}}^k\right)\\ &\geq -S'(\alpha^k)\Delta\lambda^k\left(B_n\mu p_n^k - B_t\|\boldsymbol{\beta}\|\|\boldsymbol{p}_{\boldsymbol{t}}^k\|\right)\\ &\geq -S'(\alpha^k)\Delta\lambda^k\lambda^k(\mu^2B_n - B_t) \geq 0\end{aligned}$$

if $\mu^2B_n \geq B_t$. Here, we have used the fact that $-S'(\alpha) \geq 0$ for every $\alpha \in [0,1]$ by (2.22) and (2.23), and the Cauchy–Schwarz inequality to pass to the second line. We conclude that

$$D_1 \geq -A_n'(\alpha^k)\frac{(p_n^k+\Delta p_n^k)^2}{2} - A_t'(\alpha^k)\frac{\|\boldsymbol{p}_{\boldsymbol{t}}^k+\Delta\boldsymbol{p}_{\boldsymbol{t}}^k\|^2}{2} > D_1,$$

which is a contradiction.

□

Proposition 2.6 justifies the choice of the damage functions $A_s(\alpha) = R_s\,S(\alpha)$, $s \in \{n,t\}$, (2.19). Indeed, for the statement 1) the hypothesis $S'(\alpha) = -\infty$ is essential, while in the proof of the statement 3) we have used the fact that $S'(\alpha) < 0$ and $S''(\alpha) > 0$ for $\alpha \in [0,1)$. Notice that the simultaneous evolution of plasticity and damage, although is not compulsory, would greatly simplify the discretization scheme, see Section 2.3.6, and, as a consequence, improve the model robustness since the two charge surfaces (i.e., the yield criteria for plasticity and for damage) can be seen as a single one.

2.3.4 Numerical results on typical tests

The aim of this section is to describe the behavior of the proposed model on some typical tests. For sake of simplicity, we will suppose that $\delta_{t_2} = 0$, and the only possibly nonzero tangential component will be denoted with the subscript t, as in the case $d = 2$. Moreover, from now on, we will assume that $\mu^2B_n \geq B_t$ in order to ensure the simultaneous evolution of plasticity and damage, thanks to Proposition 2.6.

Shear test with fixed compression

We assume that, at the beginning, a compression is applied until the normal stress reaches the value $-\sigma_n^{\text{com}}$, where $\sigma_n^{\text{com}} > 0$, while $\sigma_t = 0$. Then, the test consists in increasing the tangential displacement δ_t maintaining the normal stress constant. At the end of the compression stage, the values of the state variables and of the thermodynamical forces are:

$$\boldsymbol{\delta} = \begin{pmatrix}0\\ -\sigma_n^{\text{com}}/K_n\end{pmatrix}, \qquad \boldsymbol{p} = \boldsymbol{0}, \qquad \alpha = 0,$$

$$\boldsymbol{\sigma} = \boldsymbol{X} = \begin{pmatrix}0\\ -\sigma_n^{\text{com}}\end{pmatrix}, \qquad Y = 0.$$

At the beginning of the shear test, $f_{\boldsymbol{X}}(\boldsymbol{X}) = -\mu\sigma_n^{\text{com}} - \bar{c} < 0$ by (2.29), and the behavior remains elastic until the boundary of the reversibility domain $\mathbb{K}_{\boldsymbol{X}}$ is reached, i.e., when

$$\delta_t = \delta_t^{\text{crit}} := \frac{\mu\sigma_n^{\text{com}} + \bar{c}}{K_t}.$$

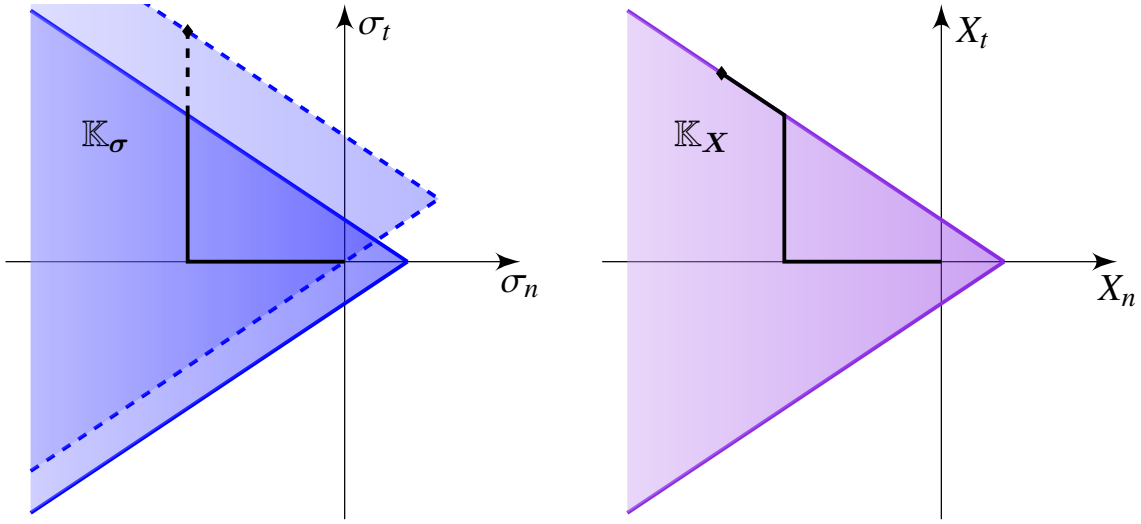


Figure 2.11 – Evolution on the stress space and on the plasticity generalized force space during a shear test with fixed compression.

In particular, during this elastic phase, the evolution of the tangential stress is linear, i.e., $\sigma_t = K_t\delta_t$, and the normal displacement jump δ_n remains constant. If we proceed with the shear test, plasticity and damage start to evolve. We can find their values thanks to the flow rule and to the plasticity and damage criteria. Assuming that $\lambda = 0$ at the beginning, the flow rule (2.30) simply becomes $\boldsymbol{p} = \lambda(\mu, 1)^\top$, since $X_t/\|X_t\| = 1$. Then, the plasticity and damage criteria (2.29) and (2.33) can be written as

$$K_t(\delta_t - \lambda) - A_t(\alpha)\lambda + \mu(-\sigma_n^{\text{com}} - \mu A_n(\alpha)\lambda) - \bar{c} = 0, \tag{2.36}$$

$$-\left(\mu^2 A_n'(\alpha) + A_t'(\alpha)\right)\frac{\lambda^2}{2} = D_1. \tag{2.37}$$

From these relations, we can express $\delta_t(\alpha)$, $\delta_n(\alpha)$ and $\sigma_t(\alpha)$ in a parametric way. By (2.37), we reach

$$\lambda(\alpha) = \sqrt{\frac{2D_1}{-(\mu^2 A_n'(\alpha) + A_t'(\alpha))}} = \sqrt{\frac{2D_1}{-(\mu^2 B_n + B_t)S'(\alpha)}},$$

$\alpha \in [0, 1]$, where $S(\alpha)$ is defined by (2.23). Then, using (2.36) and the constitutive relation (2.27), we get

$$\delta_t(\alpha) = \frac{K_t + (\mu^2 B_n + B_t)S(\alpha)}{K_t}\sqrt{\frac{2D_1}{-(\mu^2 B_n + B_t)S'(\alpha)}} + \frac{\mu\sigma_n^{\text{com}} + \bar{c}}{K_t}, \tag{2.38}$$

$$\sigma_t(\alpha) = K_t\left(\delta_t(\alpha) - \lambda(\alpha)\right) = \sqrt{2D_1(\mu^2 B_n + B_t)} \cdot \frac{S(\alpha)}{\sqrt{-S'(\alpha)}} + \mu\sigma_n^{\text{com}} + \bar{c}, \tag{2.39}$$

and

$$\delta_n(\alpha) = -\frac{\sigma_n^{\text{com}}}{K_n} + \mu\lambda(\alpha) = -\frac{\sigma_n^{\text{com}}}{K_n} + \mu\sqrt{\frac{2D_1}{-(\mu^2 B_n + B_t)S'(\alpha)}}, \tag{2.40}$$

$\alpha \in [0, 1]$.

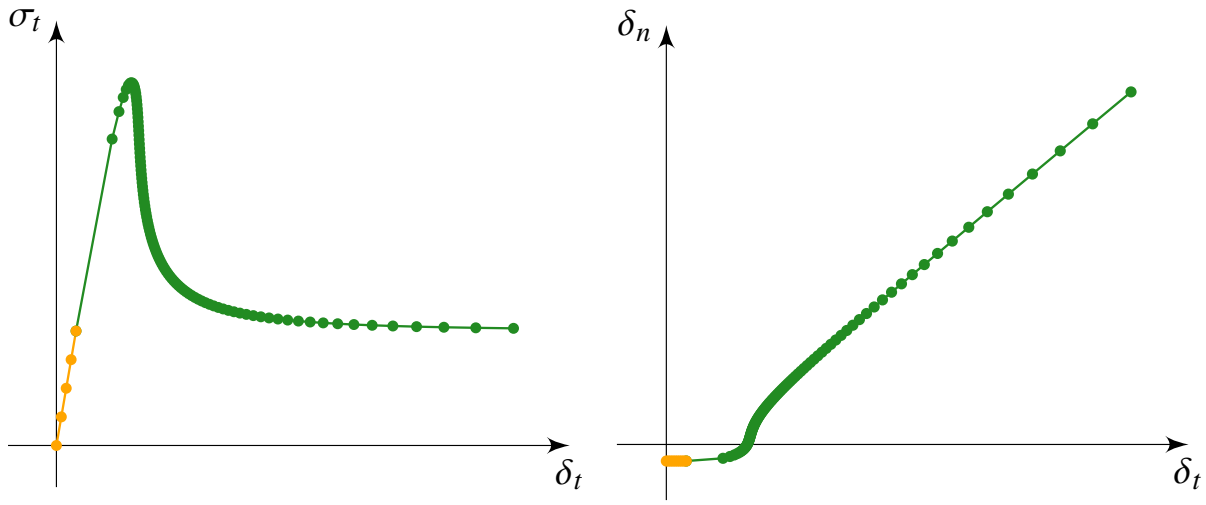


Figure 2.12 – Illustration of evolution of the shear stress σ_t (*left*) and the normal displacement jump δ_n (*right*) during a shear test with fixed compression. The phases without and with evolution of plasticity and damage are represented in *orange* and in *green*, respectively.

Remark 2.7 (Continuity of the evolution)**.** *It is easy to verify that the evolution is continuous between the elastic phase and the phase with plasticity and damage. Indeed, since*

$$\lim_{\alpha\to 0^+} \frac{1}{\sqrt{-S'(\alpha)}} = 0 \qquad \textit{and} \qquad \lim_{\alpha\to 0^+} \frac{S(\alpha)}{\sqrt{-S'(\alpha)}} = 0,$$

we have

$$\lim_{\alpha\to 0^+} \delta_t(\alpha) = \frac{\mu\sigma_n^{\mathrm{com}} + \bar{c}}{K_t} = \delta_t^{\mathrm{crit}}, \qquad \lim_{\alpha\to 0^+} \sigma_t(\alpha) = \mu\sigma_n^{\mathrm{com}} + \bar{c} = K_t\delta_t^{\mathrm{crit}},$$

and

$$\lim_{\alpha\to 0^+} \delta_n(\alpha) = -\frac{\sigma_n^{\mathrm{com}}}{K_n}.$$

Figure 2.12 shows the evolution of the tangential stress σ_t and of the normal displacement jump δ_n. In particular, the different phases of evolution are identified by different colors: the elastic phase is represented in orange, whereas the phase with evolution of plasticity and damage in green. The main properties of these curves can be also checked analytically:

- *During the initial elastic phase σ_t has a linear evolution and δ_n is constant.*
- *If the quantity $\mu^2 B_n + B_t$ is sufficiently small, there is no snap-back.* Given the derivative of the tangential jump δ_t (2.38) with respect to the damage variable α

$$\delta_t'(\alpha) = \frac{K_t S''(\alpha) + (\mu^2 B_n + B_t)\big[S(\alpha)S''(\alpha) - 2(S'(\alpha))^2\big]}{K_t}\sqrt{\frac{2D_1}{(\mu^2 B_n + B_t)(-S'(\alpha))^3}},$$

there is no snap-back if and only if $\delta_t'(\alpha) > 0$ for all $\alpha \in (0,1)$. This is equivalent to

$$K_t > (\mu^2 B_n + B_t)\left(\frac{2(S'(\alpha))^2}{S''(\alpha)} - S(\alpha)\right) = (\mu^2 B_n + B_t)R(\alpha) \qquad \forall \alpha \in (0,1),$$

according to the definition of $R(\alpha)$ (2.23). Using (2.25), it is sufficient to impose the following condition:

$$K_t > (\mu^2 B_n + B_t) \max_{\alpha\in[0,1]} R(\alpha). \tag{2.41}$$

Remark 2.8 (Consequence of the no-snap-back condition)**.** *If we assume that* (2.41) *holds, and therefore that there is no snap-back for shear evolution with fixed compression, then the second convexity condition in* (2.24) *of Proposition 2.4 is automatically satisfied.*

- *The maximum value of σ_t divides hardening from softening behavior.* Noticing that

$$\frac{\mathrm{d}\sigma_t}{\mathrm{d}\delta_t} = \frac{\sigma_t'(\alpha)}{\delta_t'(\alpha)},$$

and assuming that there is no snap-back, the hardening/softening behavior is determined by the sign of $\sigma_t'(\alpha)$. In particular,

$$\begin{aligned}\sigma_t'(\alpha) &= \sqrt{2D_1(\mu^2 B_n + B_t)}\,\frac{-2(S'(\alpha))^2 + S(\alpha)S''(\alpha)}{2\sqrt{(-S'(\alpha))^3}} \\ &= \sqrt{2D_1(\mu^2 B_n + B_t)}\,\frac{S''(\alpha)}{2\sqrt{(-S'(\alpha))^3}}\,R(\alpha),\end{aligned}$$

which vanishes in

$$\alpha_{\max} := \frac{1}{m_1 - m_2}\left(-m_2 + \sqrt{\frac{m_1 m_2}{m_1 - m_2 + 1}}\right), \tag{2.42}$$

is positive for $\alpha \in (0, \alpha_{\max})$, and negative for $\alpha \in (\alpha_{\max}, 1)$.

Remark 2.9 (Link between hardening/softening behavior and convexity of ψ_{har})**.** *The value of damage that identifies the maximum value of the shear stress, $\alpha_{\max}$* (2.42)*, coincides with the value that determines the interval in which the hardening part of the energy density ψ_{har} is convex, $\bar{\alpha}$* (2.26)*. Indeed, for both cases, we have to study the sign of the function $R(\alpha)$* (2.23)*. As a consequence, ψ_{har} is convex only during the hardening evolution.*

- *Dilatancy shows up in the evolution of δ_n.* By derivation of (2.40), we have

$$\delta_n'(\alpha) = \mu\sqrt{\frac{2D_1}{(\mu^2 B_n + B_t)(-S'(\alpha))^3}}\,S''(\alpha) > 0 \qquad \forall \alpha \in (0,1).$$

- *σ_t has a flat asymptotic behavior.* Starting from the plasticity criterion (2.36), we have

$$\lambda = \frac{K_t\delta_t - \mu\sigma_n^{\text{com}} - \bar{c}}{K_t + (\mu^2 B_n + B_t)S(\alpha)},$$

and, using the constitutive relation (2.27), we achieve that in the neighborhood of $\alpha = 1$

$$\sigma_t = K_t\left(\frac{(\mu^2 B_n + B_t)S(\alpha)\delta_t + \mu\sigma_n^{\text{com}} + \bar{c}}{K_t + (\mu^2 B_n + B_t)S(\alpha)}\right) \sim \mu\sigma_n^{\text{com}} + \bar{c}.$$

In particular, the asymptotic value of the shear test is the same value which is reached at the end of the elastic phase. This is a consequence of the fact that the reversibility domain $\mathbb{K}_{\boldsymbol{\sigma}}$ asymptotically returns to its initial position in the stress space.

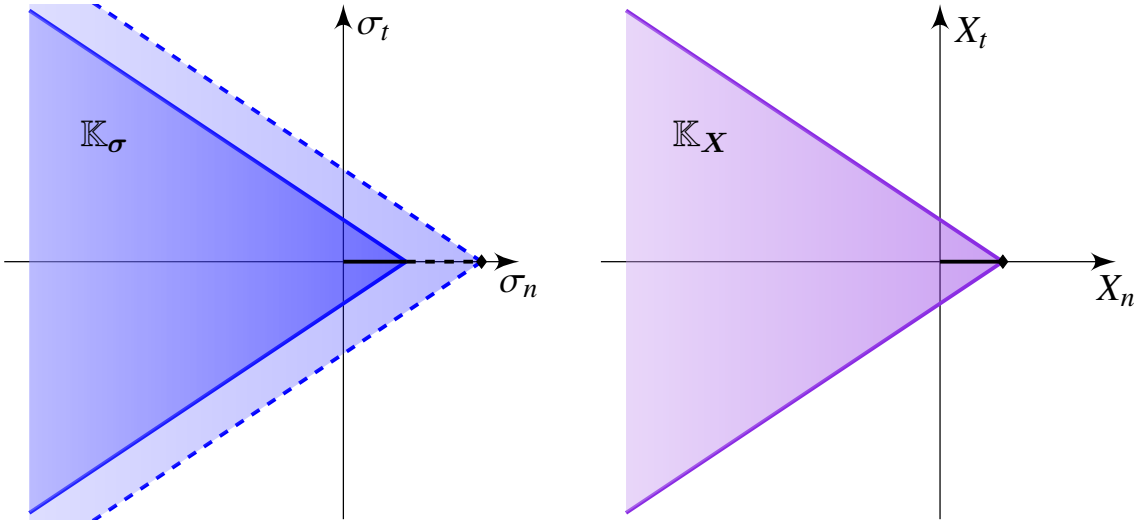


Figure 2.13 – Evolution on the stress space and on the plasticity generalized force space during a traction test.

- *The height of the peak of σ_t (i.e., the cohesion) depends on $\mu^2 B_n + B_t$, $\alpha_{\max}$ and D_1.* Thanks to the previous properties and the explicit expression for σ_t (2.39), the height of the peak is

$$C = \sigma_t(\alpha_{\max}) - \left(\mu\sigma_n^{\text{com}} + \bar{c}\right) = \sqrt{2D_1(\mu^2 B_n + B_t)} \cdot \frac{S(\alpha_{\max})}{\sqrt{-S'(\alpha_{\max})}} \tag{2.43}$$

- *δ_n has a linear asymptotic behavior.* Proceeding as before, we get

$$\delta_n = -\frac{\sigma_n^{\text{com}}}{K_n} + \mu\frac{K_t\delta_t - \mu\sigma_n^{\text{com}}}{K_t + (\mu^2 B_n + B_t)S(\alpha)} \sim \mu\delta_t,$$

  in the neighborhood of $\alpha = 1$. This is a consequence of the normal flow rule and the fact that σ_t is asymptotically flat.

Remark 2.10 (Dilatancy)**.** *The dilatancy evolution observed during shear tests on real joints, Figure 2.5, appears to be saturated for high shear values. This phenomenon can not be recovered with the proposed model, as the asymptotic behavior of δ_n is governed exclusively by the flow p_n because of the definition of normal stress (see (2.2)). This experimental non conformity is revealed further for cyclic loading.*

Traction test

During a traction test, we control the increasing evolution of the normal displacement jump δ_n, maintaining constant the tangential displacement, i.e., $\delta_t = 0$. If the residual cohesion $\bar{c} > 0$, then the evolution begins with a linear phase, with $\sigma_n = K_n\delta_n$ and $\delta_t = 0$. This phase stops when the angular point of the domain is reached, i.e.,

$$\delta_n = \delta_n^{\text{crit}} := \frac{\bar{c}}{\mu K_n}.$$

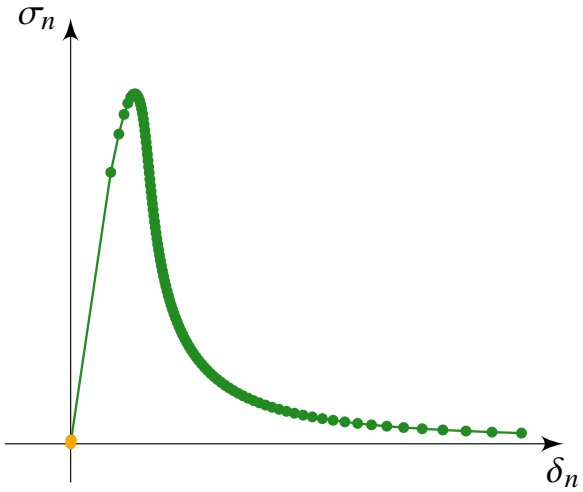


Figure 2.14 – Illustration of evolution of the traction stress σ_n during a traction test. The phases without and with evolution of plasticity and damage are represented in *orange* and in *green*, respectively.

Besides this point, we can find the evolution of the state variables and thermodynamical forces imposing $(X_n, X_t) = (\bar{c}/\mu, 0)$ and the damage criterion (2.33):

$$\begin{cases} \sigma_n = K_n(\delta_n - p_n), \\ \sigma_t = K_t(\delta_t - p_t), \end{cases} \qquad \begin{cases} X_n = K_n(\delta_n - p_n) - A_n(\alpha)p_n = \dfrac{\bar{c}}{\mu}, \\ X_t = K_t(\delta_t - p_t) - A_t(\alpha)p_t = 0, \end{cases}$$

$$Y = -A_n'(\alpha)\frac{{p_n}^2}{2} - A_t'(\alpha)\frac{{p_t}^2}{2} = D_1.$$

Indeed, from these relations we obtain

$$p_t = 0, \qquad \sigma_t = 0,$$

$$p_n(\alpha) = \sqrt{\frac{2D_1}{-B_n S'(\alpha)}},$$

$$\delta_n(\alpha) = \frac{K_n + B_n S(\alpha)}{K_n}\sqrt{\frac{2D_1}{-B_n S'(\alpha)}} + \frac{\bar{c}}{\mu K_n}, \tag{2.44}$$

and

$$\sigma_n(\alpha) = \sqrt{2D_1 B_n}\frac{S(\alpha)}{\sqrt{-S'(\alpha)}} + \frac{\bar{c}}{\mu}. \tag{2.45}$$

The evolution of the normal stress σ_n is presented by Figure 2.14, where, as for the shear test, the possibly elastic phase is shown in orange and the phase with evolution of plasticity and damage in green. The main properties of a traction test with the proposed model are:

- *After an elastic phase, the evolution of σ_n represents the movement of the vertex of the reversibility domain $\mathbb{K}_{\boldsymbol{\sigma}}$.*

- *If B_n is sufficiently small, there is no snap-back.* Deriving (2.44) with respect to α, we get

$$\delta_n'(\alpha) = \frac{K_n S''(\alpha) + B_n\left[S(\alpha)S''(\alpha) - 2(S'(\alpha))^2\right]}{K_t}\sqrt{\frac{2D_1}{B_n(-S'(\alpha))^3}}.$$

  Then, using the definition of $R(\alpha)$ (2.23), and (2.25), a sufficient condition to have no snap-back is:

$$K_n > B_n \max_{\alpha\in[0,1]} R(\alpha). \tag{2.46}$$

Remark 2.11 (Consequence of no-snap-back condition)**.** *Notice that assuming that* (2.46) *holds, then the first convexity condition in* (2.24) *of Proposition 2.4 is automatically satisfied. It is remarkable to observe that combining this result with Remark 2.8 we obtain the following statement:*

Proposition 2.12. *Let $\psi(\boldsymbol{\delta}, \boldsymbol{p}, \alpha)$ be defined by* (2.18)*, and the plasticity and damage criteria by* (2.29) *and* (2.33)*, respectively. Then, assuming that the (sufficient) no-snap-back conditions* (2.41) *and* (2.46) *hold, ψ is convex with respect to $(\boldsymbol{p}, \alpha)$.*

- *The maximum value of σ_n divides hardening from softening behavior.* Assuming that the is no snap-back, the hardening/softening behavior is established by the sign of

$$\sigma_t'(\alpha) = \sqrt{2D_1B_n}\,\frac{S''(\alpha)}{2\sqrt{(-S'(\alpha))^3}}\,R(\alpha).$$

  It vanishes in $\alpha_{\max}$, is positive for $\alpha \in (0, \alpha_{\max})$, and negative for $\alpha \in (\alpha_{\max}, 1)$, where $\alpha_{\max}$ is again defined by (2.42).

- *σ_n has a flat asymptotic behavior.* From (2.45), we get

$$\sigma_n \sim \frac{\bar{c}}{\mu}$$

  in the neighborhood of $\alpha = 1$. This represents the fact that the reversibility domain $\mathbb{K}_{\boldsymbol{\sigma}}$ asymptotically returns to its initial position.

- *The height of the peak of σ_n (i.e., the maximum tensile strength) depends on B_n, $\alpha_{\max}$, and D_1.* From the previous properties, using (2.45), the height of the peak is

$$T := \sigma_n(\alpha_{\max}) - \frac{\bar{c}}{\mu} = \sqrt{2D_1B_n}\,\frac{S(\alpha_{\max})}{\sqrt{-S'(\alpha_{\max})}}. \tag{2.47}$$

2.3.5 Influence of parameters on the evolution

The model we have proposed in this chapter has nine parameters:

$$K_n, K_t, B_n, B_t, m_1, m_2, \mu, \bar{c}, D_1.$$

In this subsection, we analyze the influence of these parameters and propose a possible way to fit them.

Rigidity coefficients

The rigidity coefficients K_n and K_t control the behavior during the elastic phases of evolution. They can be fitted via compression and shear tests. In practice, the normal rigidity is chosen sufficiently big to penalize interpenetration phenomena but sufficiently small to have convergence. For example, the default values in `JOINT_MECA_RUPT` and `JOINT_MECA_FROT` are $K_n = K_t = 3$ TPa/m [49].

Friction coefficient and residual cohesion

The friction coefficient $\mu > 0$ and the residual cohesion $\bar{c} \geq 0$ fix the shape of the "initial" (i.e., $\alpha = 0$) and of the "asymptotic" (i.e., $\alpha \approx 1$) reversibility domain in the stress space $\mathbb{K}_{\boldsymbol{\sigma}}$ via the convex function $f_{\boldsymbol{\sigma}}$ (2.32). The latter can be possibly determined experimentally with shear tests with fixed compression. For joint models, usually $\mu \approx 1$. Furthermore, a zero residual cohesion ($\bar{c} = 0$) can be chosen in order to have zero residual tensile strength, i.e., $\sigma_n \sim 0$ asymptotically, during traction tests.

Damage related parameters

All the other parameters, i.e., B_n, B_t, m_1, m_2, and D_1, are related to the evolution of the damage state variable α and, in particular, to the shape of the peaks during shear and traction tests, Figure 2.12 and Figure 2.14. The idea here is to create a link between some of these parameters and some quantities that can be measure experimentally. In particular, given the cohesion C and the maximum tensile strength T, i.e., the height of the peaks during shear and traction tests, we can fix the values of B_n and B_t using (2.43) and (2.47):

$$B_n = \frac{T^2}{2D_1} \cdot \frac{-S'(\alpha_{\max})}{(S(\alpha_{\max}))^2} \tag{2.48}$$

and

$$B_t = \frac{C^2}{2D_1} \cdot \frac{-S'(\alpha_{\max})}{(S(\alpha_{\max}))^2} - \mu^2 B_n = \frac{C^2 - \mu^2 T^2}{2D_1} \cdot \frac{-S'(\alpha_{\max})}{(S(\alpha_{\max}))^2}. \tag{2.49}$$

Remark 2.13 (Conditions on the values of T and C)**.** *The values of cohesion and maximum tensile strength can be fixed independently if and only if*

$$\mu T \leq C \leq \sqrt{2}\mu T. \tag{2.50}$$

This is a consequence of imposing $B_t \geq 0$ and $\mu^2 B_n \geq B_t$ (sufficient condition for the simultaneous evolution of plasticity and damage by Proposition 2.6): the first constrain implies $C \geq \mu T$, while the second $C \leq \sqrt{2}\mu T$. In practice, if (2.50) is not satisfied by the values recovered experimentally, we can redefine the value of T as

$$T = \frac{C}{\mu} \qquad or \qquad T = \frac{C}{\sqrt{2}\mu},$$

accordingly to the inequality violated by the experimental values.

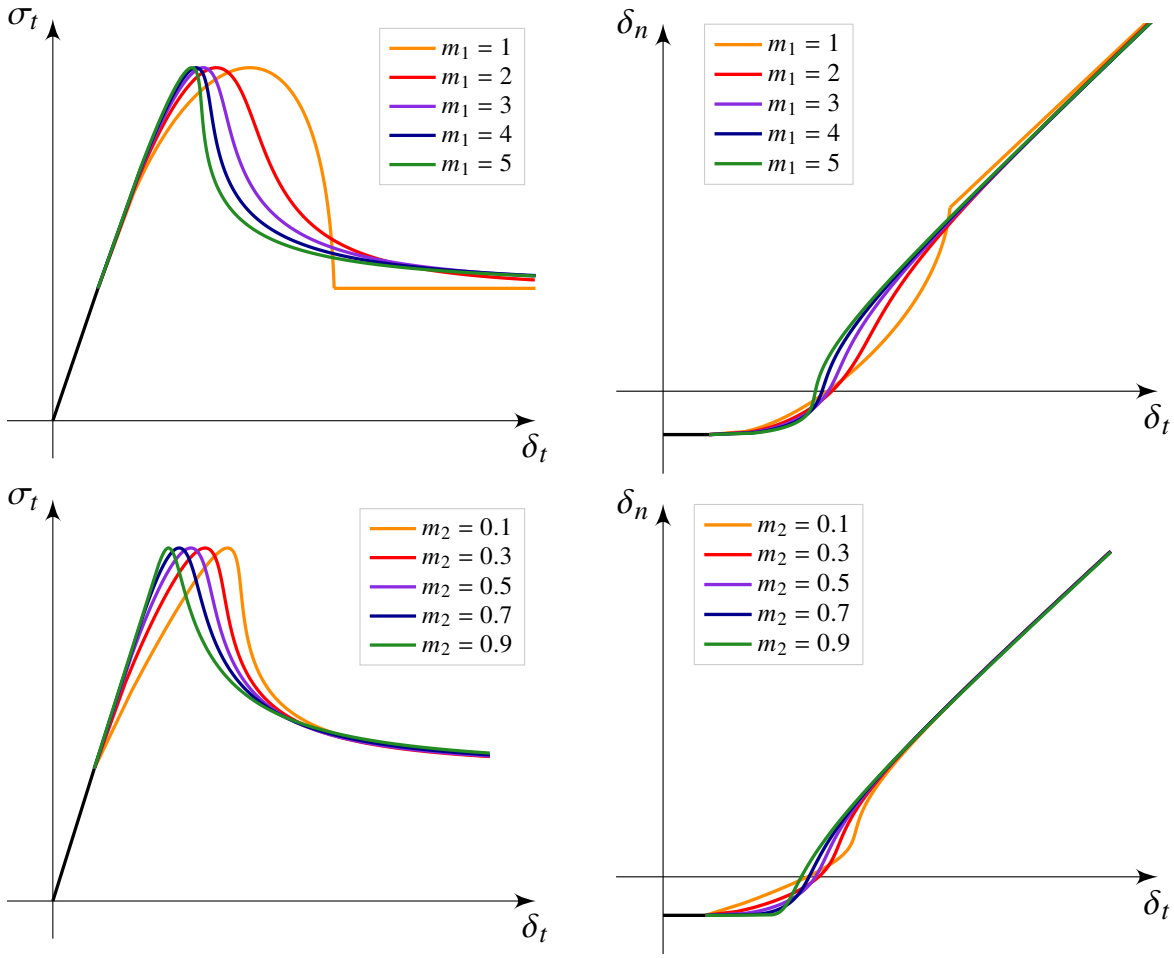


Figure 2.15 – Influence of the value of m_1 (*top*) and m_2 (*down*) on the evolution of the tangential stress σ_t (*left*) and normal displacement δ_n (*right*) during a shear test with fixed compression.

Notice that, once we have fixed B_n and B_t through (2.48) and (2.49), the no-snap-back conditions (2.41) and (2.46) can be rewritten as a condition on the parameter D_1:

$$D_1 > \max\left\{\frac{T^2}{2K_n}, \frac{C^2}{2K_t}\right\} \frac{-S'(\alpha_{\max})}{(S(\alpha_{\max}))^2} \max_{\alpha\in[0,1]} R(\alpha).$$

Figure 2.15 and 2.16 shows the influence of the parameters m_1, m_2 and D_1 on the shape of the peak in shear stress evolution (*left*) and of the dilatancy evolution (*right*), assuming that the value of cohesion C is fixed. In particular, it is evident from Figure 2.16 that D_1 has to be chosen sufficiently big in order to avoid a snap-back phenomenon. As default values for m_1 and m_2, we suggest 3 and 0.5, respectively.

Remark 2.14 (Choice of the damage functions)**.** *In the paper [94], the choice of the damage functions were motivated only from the conditions $A_s(0) = +\infty$ and $A_s(1) = 0$, $s \in \{n, t\}$, which imply that plasticity and damage start at the same time, and that we loose stiffness in the joint when it is fully damage recovering a perfect elasto-plastic model. The simpliest*

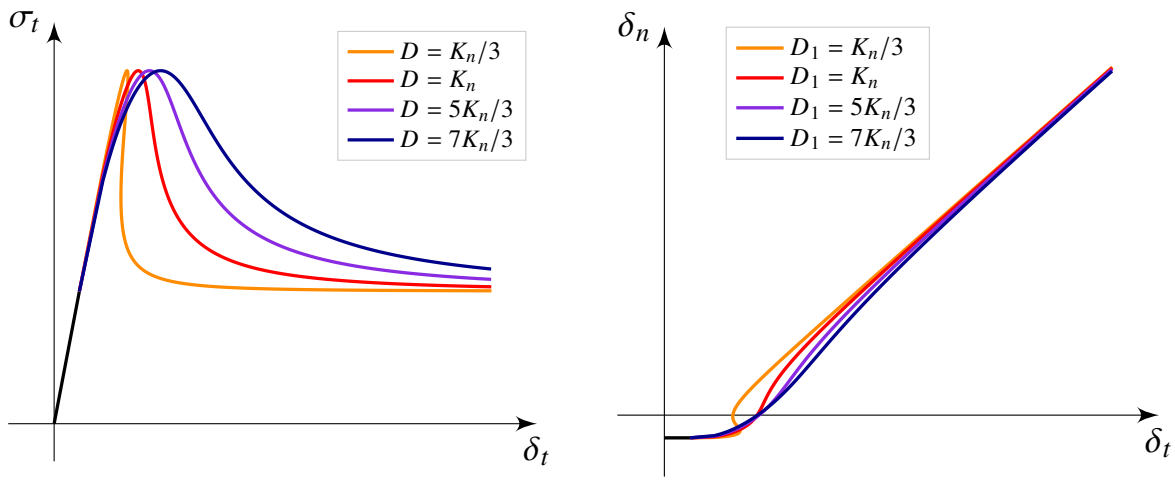


Figure 2.16 – Influence of the value of D_1 on the evolution of the tangential stress σ_t (*left*) and normal displacement (*right*) during a shear test with fixed compression.

function we can consider in this sense is

$$\gamma \frac{(1-\alpha)^{m_1}}{\alpha^{m_2}} \qquad \text{with } \gamma, m_1, m_2 > 0.$$

In addition, the assumption $m_2 < 1 < m_1$ *comes from the conditions* $A_s'(\alpha) < 0$ *and* $A_s''(\alpha) > 0$*. Notice that we can also recover this assumption directly from the considerations of the evolution during a shear test with fixed compression (or, equivalently, during a traction test). Indeed, from (2.38) and (2.39) we get*

1) the evolution is continuous if and only if

$$\lim_{\alpha\to 0^+} \frac{1}{\sqrt{-S'(\alpha)}} = 0 \qquad \text{and} \qquad \lim_{\alpha\to 0^+} \frac{S(\alpha)}{\sqrt{-S'(\alpha)}} = 0;$$

the first condition is automatically satisfy by the hypothesis $m_2 > 0$*, while the second one holds if and only if* $m_2 < 1$*;*

2) $\lim_{\alpha\to 1^-} \delta_t(\alpha) = +\infty$ *if and only if*

$$\lim_{\alpha\to 1^-} \frac{1}{\sqrt{-S'(\alpha)}} = +\infty$$

if and only if $m_1 > 1$*. In particular, Figure 2.15 show that in the case* $m_1 = 1$ *the damage is completely evolved, i.e.,* $\alpha = 1$*, for a finite value of* δ_t*.*

2.3.6 Implementation with incremental evolution

In this section we propose some implementation notes with the incremental approach used for programming a new constitutive relation in `code_aster` (this approach holds in the context either of joints or geomaterials). In particular, the integration of a constitutive law is

composed of two steps: after an elastic prediction, if the plasticity criterion is not satisfied, a correction is performed with the evolution of plasticity and damage state variables. The technique is similar to the one used to implement `JOINT_MECA_FROT` [49], but while in the latter case the singularity of the reversibility domain is considered implicitly by imposing a maximum value for the normal stress (2.11a), here we show how to treat it using the normal flow rule. The description is technical, and for a first reading we suggest to take a look to the schematic summary in Algorithm 2.1.

In the incremental approach, given a state variable or a termodynamical force z, its value at the previous and current state, and its increment are denoted with z^-, z, and $\Delta z := z - z^-$, respectively. The input values are the state variables at the previous step and the increment of displacement jump $\Delta\boldsymbol{\delta} = (\Delta\delta_n, \Delta\boldsymbol{\delta}_t)$, while the output are the current values $\boldsymbol{p}$, α, $\boldsymbol{\sigma}$, $\boldsymbol{X}$, Y. In particular, the thermodynamical forces can be easily computed by their definition once we have the current state variables:

$$\begin{cases} \sigma_n = K_n(\delta_n - p_n^- - \Delta p_n), & \text{(2.51a)} \\ \boldsymbol{\sigma}_t = K_t(\boldsymbol{\delta}_t - \boldsymbol{p}_t^- - \Delta\boldsymbol{p}_t), & \text{(2.51b)} \\ X_n = K_n(\delta_n - p_n^- - \Delta p_n) - A_n(\alpha^- + \Delta\alpha)(p_n^- + \Delta p_n), & \text{(2.51c)} \\ \boldsymbol{X}_t = K_t(\boldsymbol{\delta}_t - \boldsymbol{p}_t^- - \Delta\boldsymbol{p}_t) - A_t(\alpha^- + \Delta\alpha)(\boldsymbol{p}_t^- + \Delta\boldsymbol{p}_t), & \text{(2.51d)} \\ Y = -A_n'(\alpha^- + \Delta\alpha)\dfrac{(p_n^- + \Delta p_n)^2}{2} - A_t'(\alpha^- + \Delta\alpha)\dfrac{(p_t^- + \Delta p_t)^2}{2}. & \text{(2.51e)} \end{cases}$$

In the discretized framework, the plasticity and damage criteria (2.29) and (2.33), and the flow rule (2.30) become

$$\begin{cases} f_{\boldsymbol{X}}(\boldsymbol{X}) = \|\boldsymbol{X}_t\| + \mu X_n - \bar{c} \le 0, & \text{(2.52a)} \\ \Delta\boldsymbol{p} = \Delta\lambda \cdot \dfrac{\partial f_{\boldsymbol{X}}}{\partial \boldsymbol{X}}, & \text{(2.52b)} \\ f_{\boldsymbol{X}}(\boldsymbol{X})\,\Delta\lambda = 0,\ \Delta\lambda \ge 0, & \text{(2.52c)} \\ Y = -A_n'(\alpha^- + \Delta\alpha)\dfrac{(p_n^- + \Delta p_n)^2}{2} - A_t'(\alpha^- + \Delta\alpha)\dfrac{\|\boldsymbol{p}_t^- + \Delta\boldsymbol{p}_t\|^2}{2} \le D_1, & \text{(2.52d)} \\ \Delta\alpha \ge 0, & \text{(2.52e)} \\ (Y - D_1)\,\Delta\alpha = 0. & \text{(2.52f)} \end{cases}$$

In particular, we recall that (2.52c) translates as follows: either $\Delta\lambda = 0$ and we are in the reversibility domain $\mathbb{K}_{\boldsymbol{X}}$, or $\Delta\lambda > 0$, plasticity evolves, and $\boldsymbol{X} \in \partial\mathbb{K}_{\boldsymbol{X}}$. In a similar way, (2.52f) establishes that either $\Delta\alpha = 0$ and $Y \le D_1$, or $\Delta\alpha > 0$, i.e., damage evolves, and $Y = D_1$. Finally, since we work under the assumption of simultaneous evolution of plasticity and damage, $\Delta\lambda > 0$ if and only if $\Delta\alpha > 0$.

Remark 2.15 (Hypotheses of simultaneous evolution)**.** *We will always assume the simultaneous evolution of plasticity and damage, imposing $\mu^2 B_n \ge B_t$ (see Proposition 2.6). As a consequence*

$$\Delta\lambda > 0 \quad \Longleftrightarrow \quad \Delta\alpha > 0.$$

At first, an elastic prediction is performed: we compute

$$\boldsymbol{X}^{\text{pred}} := \boldsymbol{X}^- + \mathbb{E}\Delta\boldsymbol{\delta} = \mathbb{E}(\boldsymbol{\delta} - \boldsymbol{p}^-) - \mathbb{H}(\alpha^-)\boldsymbol{p}^-,$$

recalling that the elastic and hardening matrices $\mathbb{E}$ and $\mathbb{H}(\alpha^-)$ are defined by (2.17). Then we check whether the plasticity criterion (2.52a) is satisfied or not. If

$$\|\boldsymbol{X}_{\boldsymbol{t}}^{\text{pred}}\| + \mu X_n^{\text{pred}} - \bar{c} \leq 0, \tag{2.53}$$

then the elastic prediction is the solution. As a consequence,

$$\begin{cases} \Delta \boldsymbol{p} = \boldsymbol{0}, \\ \Delta \alpha = 0, \end{cases} \qquad \text{and} \qquad \begin{cases} \boldsymbol{\sigma} = \boldsymbol{\sigma}^- + \mathbb{E}\Delta\boldsymbol{\delta} = \mathbb{E}(\boldsymbol{\delta} - \boldsymbol{p}^-), \\ \boldsymbol{X} = \boldsymbol{X}^{\text{pred}}, \\ Y = Y^-. \end{cases}$$

Otherwise, plasticity and damage evolve, i.e., $\Delta\lambda > 0$ and $\Delta\alpha > 0$. Initially, we assume that $\boldsymbol{X_t} \neq \boldsymbol{0}$. Therefore, the plastic flow rule (2.52b) is

$$\begin{cases} \Delta p_n = \mu \Delta\lambda, & \text{(2.54a)} \\ \Delta \boldsymbol{p}_t = \Delta\lambda \dfrac{\boldsymbol{X}_t}{\|\boldsymbol{X}_t\|}. & \text{(2.54b)} \end{cases}$$

Remark 2.16 (Damage variable notation)**.** *As we will see later on, while the increment of plasticity $\Delta\boldsymbol{p}$ can be computed explicitly, the increment of damage $\Delta\alpha$ is computed implicitly by solving a nonlinear equation. In addition, since this nonlinear equation only depends on $\alpha^- + \Delta\alpha$, we can solve the equation finding directly the current value of damage $\alpha \in (\alpha^-, 1]$. For this reason, from now on we simply write α instead of $\alpha^- + \Delta\alpha$.*

Considering (2.54b), using the definition of $\boldsymbol{X_t}$ (2.51d), and multiplying both sides by $\|\boldsymbol{X_t}\|$, we have

$$\begin{aligned} \Delta\boldsymbol{p_t} \left\| K_t(\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\boldsymbol{p_t}) - A_t(\alpha)(\boldsymbol{p_t^-} + \Delta\boldsymbol{p_t}) \right\| \\ = \Delta\lambda \left(K_t(\boldsymbol{\delta}_t - \boldsymbol{p}_t^- - \Delta\boldsymbol{p}_t) - A_t(\alpha)(\boldsymbol{p}_t^- + \Delta\boldsymbol{p}_t) \right). \end{aligned}$$

With the aim of rewriting the plasticity criterion as an equation that does not depend on $\Delta\boldsymbol{p}_t$ but only on $\Delta\lambda$, we divide the terms that contain this increment from those that do not contain it:

$$\begin{aligned} \Delta\boldsymbol{p}_t \left(\left\| K_t(\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\boldsymbol{p_t}) - A_t(\alpha)(\boldsymbol{p_t^-} + \Delta\boldsymbol{p_t}) \right\| + K_t\Delta\lambda + A_t(\alpha)\Delta\lambda \right) \\ = \Delta\lambda \left(K_t(\boldsymbol{\delta_t} - \boldsymbol{p_t^-}) - A_t(\alpha)\boldsymbol{p_t^-} \right). \end{aligned}$$

Now, taking the norm of each side and observing that $\|\Delta\boldsymbol{p_t}\| = \Delta\lambda$, we obtain

$$\left\| K_t(\boldsymbol{\delta}_t - \boldsymbol{p}_t^- - \Delta\boldsymbol{p}_t) - A_t(\alpha)(\boldsymbol{p}_t^- + \Delta\boldsymbol{p}_t) \right\| = \left\| K_t(\boldsymbol{\delta}_t - \boldsymbol{p}_t^-) - A_t(\alpha)\boldsymbol{p}_t^- \right\| - K_t\Delta\lambda - A_t(\alpha)\Delta\lambda.$$

As a consequence, defining the vector

$$\boldsymbol{W} := K_t(\boldsymbol{\delta_t} - \boldsymbol{p_t^-}) - A_t(\alpha)\boldsymbol{p_t^-}, \tag{2.55}$$

and using (2.54a), the plasticity criterion (2.52a) can be rewritten as follows:

$$\|\boldsymbol{W}\| - K_t\Delta\lambda - A_t(\alpha)\Delta\lambda + \mu K_n(\delta_n - p_n^- - \mu\Delta\lambda) - \mu A_n(\alpha)(p_n^- + \mu\Delta\lambda) - \bar{c} = 0. \tag{2.56}$$

This is a linear equation of $\Delta\lambda$ that does not depend on $\Delta\boldsymbol{p_t}$. Moreover, considering again (2.54b) and denoting by $\uparrow\uparrow$ the collinearity relation between vectors, we have

$$\Delta\boldsymbol{p_t} \uparrow\uparrow \boldsymbol{X_t} = K_t(\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\boldsymbol{p_t}) - A_t(\alpha)(\boldsymbol{p_t^-} + \Delta\boldsymbol{p_t}),$$

and this implies

$$\Delta\boldsymbol{p_t} \uparrow\uparrow K_t(\boldsymbol{\delta_t} - \boldsymbol{p_t^-}) - A_t(\alpha)\boldsymbol{p_t^-} = \boldsymbol{W}.$$

As a consequence, we can write

$$\Delta\boldsymbol{p_t} = \Delta\lambda \frac{\boldsymbol{W}}{\|\boldsymbol{W}\|}. \tag{2.57}$$

Notice that the right term does not depend on $\Delta\boldsymbol{p_t}$ anymore. Using this relation, the damage criterion can be rewritten as

$$-A_n'(\alpha)\left(p_n^- + \mu\Delta\lambda\right)^2 - A_t'(\alpha)\left\|\boldsymbol{p_t^-} + \Delta\lambda \cdot \frac{\boldsymbol{W}}{\|\boldsymbol{W}\|}\right\|^2 = 2D_1. \tag{2.58}$$

Therefore, we have obtained two equations (2.56) and (2.58) with unknowns $\Delta\lambda$ and α. From (2.56) we can easily obtain the expression of $\Delta\lambda$, since it is a linear equation

$$\Delta\lambda = \frac{\|\boldsymbol{W}\| + \mu K_n(\delta_n - p_n^-) - \mu A_n(\alpha)p_n^- - \bar{c}}{\mu^2 K_n + K_t + \mu^2 A_n(\alpha) + A_t(\alpha)}. \tag{2.59}$$

Moreover, from (2.58) we get

$$\begin{aligned} -\left(\mu^2 A_n'(\alpha) + A_t'(\alpha)\right)\Delta\lambda^2 - 2\left(\mu A_n'(\alpha)p_n^- + A_t'(\alpha)\boldsymbol{p_t^-} \cdot \frac{\boldsymbol{W}}{\|\boldsymbol{W}\|}\right)\Delta\lambda - \\ - A_n'(\alpha)(p_n^-)^2 - A_t'(\alpha)\|\boldsymbol{p_t^-}\|^2 - 2D_1 = 0. \end{aligned} \tag{2.60}$$

Since the quadratic coefficient is positive and the constant coefficient is negative, the equation has always only one positive solution:

$$\Delta\lambda = \frac{\mu A_n'(\alpha)p_n^- + A_t'(\alpha)\boldsymbol{p_t^-} \cdot {}^{W}/_{\|W\|} + \sqrt{\text{disc}}}{-\left(\mu^2 A_n'(\alpha) + A_t'(\alpha`)\right)} \tag{2.61}$$

where the discriminant is

$$\begin{aligned} \text{disc} := \left(\mu A_n'(\alpha)p_n^- + A_t'(\alpha)\boldsymbol{p_t^-} \cdot \frac{\boldsymbol{W}}{\|\boldsymbol{W}\|}\right)^2 - \\ - \left(\mu^2 A_n'(\alpha) + A_t'(\alpha)\right)\left(A_n'(\alpha)(p_n^-)^2 + A_t'(\alpha)\|\boldsymbol{p_t^-}\|^2 + 2D_1\right). \end{aligned} \tag{2.62}$$

Remark 2.17 (Sign of the constant coefficient of (2.60))**.** *In order to prove that the constant coefficient of the quadratic equation (2.60) is negative, it is sufficient to use the fact that A_s', $s \in \{n, t\}$, is an increasing function (in particular, $A_s'(\alpha) > A_s'(\alpha^-)$ since $\alpha > \alpha^-$), and the damage threshold criterion at the previous step:*

$$-A_n'(\alpha)(p_n^-)^2 - A_t'(\alpha)\|\boldsymbol{p_t^-}\|^2 - 2D_1 < -A_n'(\alpha^-)(p_n^-)^2 - A_t'(\alpha^-)\|\boldsymbol{p_t^-}\|^2 - 2D_1 = 0$$

when $\alpha^- > 0$. For $\alpha^- = 0$ we have

$$-A_n'(\Delta\alpha)(p_n^-)^2 - A_t'(\Delta\alpha)\|\boldsymbol{p_t^-}\|^2 - 2D_1 = -2D_1 < 0$$

because $\boldsymbol{p}^- = \boldsymbol{0}$ (plasticity and damage start at the same time by Proposition 2.6).

Combining (2.59) with (2.61), we achieve the following nonlinear equation with the unknown $\alpha \in (\alpha^-, 1]$:

$$\frac{\|\boldsymbol{W}\| + \mu K_n(\delta_n - p_n^-) - \mu A_n(\alpha) p_n^- - \bar{c}}{\mu^2 K_n + K_t + \mu^2 A_n(\alpha) + A_t(\alpha)} = \frac{\mu A_n'(\alpha) p_n^- + A_t'(\alpha) \boldsymbol{p}_{\boldsymbol{t}}^- \cdot \boldsymbol{w}/\|\boldsymbol{w}\| + \sqrt{\text{disc}}}{-\left(\mu^2 A_n'(\alpha) + A_t'(\alpha)\right)}. \tag{2.63}$$

Once obtained the new value of damage α by solving (2.63), we have to verify that it is the actual solution, i.e., that $\boldsymbol{X} \in \partial\mathbb{K}$. With this purpose, we compute $\Delta\lambda$ from (2.59), Δp_n from the flow rule (2.54a), and finally X_n by its definition (2.51c). If $X_n \leq \bar{c}/\mu$, then we have found the solution and we can compute also $\Delta \boldsymbol{p}_{\boldsymbol{t}}$ from (2.57), and $\boldsymbol{\sigma}$, $\boldsymbol{X}_{\boldsymbol{t}}$, and Y from (2.51). Otherwise, $X_n > \bar{c}/\mu$ implies that we made the wrong assumption $\boldsymbol{X}_{\boldsymbol{t}} \neq \boldsymbol{0}$, and $\boldsymbol{X} = (\bar{c}/\mu, \boldsymbol{0})$. As a consequence, the system to solve becomes

$$\begin{cases} \Delta\lambda > 0, & \text{(2.64a)} \\ \Delta p_n = \mu\Delta\lambda \quad \text{and} \quad \|\Delta \boldsymbol{p}_{\boldsymbol{t}}\| \leq \Delta\lambda, & \text{(2.64b)} \\ X_n = \dfrac{\bar{c}}{\mu} \quad \text{and} \quad \boldsymbol{X}_{\boldsymbol{t}} = \boldsymbol{0}, & \text{(2.64c)} \\ \Delta\alpha > 0, & \text{(2.64d)} \\ Y = -A_n'(\alpha)\dfrac{(p_n^- + \Delta p_n)^2}{2} - A_t'(\alpha)\dfrac{\|\boldsymbol{p}_{\boldsymbol{t}}^- + \Delta \boldsymbol{p}_{\boldsymbol{t}}\|^2}{2} = D_1. & \text{(2.64e)} \end{cases}$$

Equation (2.64c) can be explicitly written as the following system

$$\begin{cases} X_n = K_n(\delta_n - p_n^- - \Delta p_n) - A_n(\alpha)(p_n^- + \Delta p_n) = \dfrac{\bar{c}}{\mu}, \\ \boldsymbol{X}_{\boldsymbol{t}} = K_t(\boldsymbol{\delta}_{\boldsymbol{t}} - \boldsymbol{p}_{\boldsymbol{t}}^- - \Delta \boldsymbol{p}_{\boldsymbol{t}}) - A_t(\alpha)(\boldsymbol{p}_{\boldsymbol{t}}^- + \Delta \boldsymbol{p}_{\boldsymbol{t}}) = \boldsymbol{0}, \end{cases}$$

from which we easily obtain

$$\begin{cases} \Delta p_n = \dfrac{K_n \delta_n}{K_n + A_n(\alpha)} - p_n^- - \dfrac{\bar{c}}{\mu(K_n + A_n(\alpha))}, \\ \Delta \boldsymbol{p}_{\boldsymbol{t}} = \dfrac{K_t \boldsymbol{\delta}_{\boldsymbol{t}}}{K_t + A_t(\alpha)} - \boldsymbol{p}_{\boldsymbol{t}}^-. \end{cases} \tag{2.65}$$

Inserting the latter into (2.64e) we obtain the following nonlinear equation in α:

$$-A_n'(\alpha)\left(\frac{K_n \delta_n}{K_n + A_n(\alpha)} - \frac{\bar{c}}{\mu(K_n + A_n(\alpha))}\right)^2 - A_t'(\alpha)\left(\frac{K_t\|\boldsymbol{\delta}_{\boldsymbol{t}}\|}{K_t + A_t(\alpha)}\right)^2 = 2D_1. \tag{2.66}$$

Solving this equation, we find the new value of damage $\alpha \in (\alpha^-, 1]$. Then, we can compute $\Delta \boldsymbol{p}$ from (2.65), and $\boldsymbol{\sigma}$ from (2.51a)–(2.51b).

Algorithm 2.1 summarize the procedure we have just describe. In particular, for simplicity of notation, we define the two nonlinear functions which correspond to the nonlinear equations

(2.63) and (2.66):

$$\begin{aligned} F(\alpha) \coloneqq & \left[\|\boldsymbol{W}\| + \mu K_n(\delta_n - p_n^-) - \mu A_n(\alpha) p_n^- - \bar{c}\right] \left(\mu^2 A_n'(\alpha) + A_t'(\alpha)\right) \\ & + \left[\mu A_n'(\alpha) p_n^- + A_t'(\alpha) \boldsymbol{p_t^-} \cdot \frac{\boldsymbol{W}}{\|\boldsymbol{W}\|} + \sqrt{\text{disc}}\right] \\ & \cdot \left(\mu^2 K_n + K_t + \mu^2 A_n(\alpha) + A_t(\alpha)\right), \end{aligned} \tag{2.67}$$

and

$$G(\alpha) \coloneqq A_n'(\alpha) \left(\frac{K_n \delta_n}{K_n + A_n(\alpha)} - \frac{\bar{c}}{\mu(K_n + A_n(\alpha))}\right)^2 + A_t'(\alpha) \left(\frac{K_t \|\boldsymbol{\delta_t}\|}{K_t + A_t(\alpha)}\right)^2 + 2D_1. \tag{2.68}$$

Algorithm 2.1 Implementation constitutive relation

Input: $\boldsymbol{\delta}, \boldsymbol{p}^-, \alpha^-$
compute $\boldsymbol{X}^{\text{pred}} = \boldsymbol{X}^- + \mathbb{E}\Delta\boldsymbol{\delta}$ (→ elastic prediction)
if $\|\boldsymbol{X}_{\boldsymbol{t}}^{\text{pred}}\| + \mu X_n^{\text{pred}} - \bar{c} \leq 0$ **then**
    The elastic prediction is the solution
    set $\boldsymbol{X} = \boldsymbol{X}^{\text{pred}}$ and $Y = Y^-$
    compute $\boldsymbol{\sigma} = \boldsymbol{\sigma}^- + \mathbb{E}\Delta\boldsymbol{\delta}$
else (→ correction)
    solve the nonlinear equation $F(\alpha) = 0$, with $F(\alpha)$ defined by (2.67)
    compute $\Delta\lambda$ from (2.59), Δp_n from (2.54a), and X_n from (2.51c)
    if $X_n \leq \bar{c}/\mu$ **then**
        The solution has been found
        compute $\Delta\boldsymbol{p_t}$ from (2.57) and $\boldsymbol{X_t}$ from (2.51d)
    else (→ singularity)
        set $\boldsymbol{X} = (\bar{c}/\mu, \boldsymbol{0})$
        solve the nonlinear equation $G(\alpha) = 0$, with $G(\alpha)$ defined by (2.68)
        compute $\Delta\boldsymbol{p}$ from (2.65)
    end if
    compute $\boldsymbol{\sigma}$ and Y from (2.51)
end if

2.3.7 Conclusion

The main contribution of this chapter consists in the development and analysis of one of the possible regularisations of the model proposed in [94]. We have adapted the latter volumetric constitutive relation to the context of joint modeling, and we have proposed its numerical implementation. The convexity of the surface energy density function ψ has been studied in detail, Proposition 2.4. In Proposition 2.6, we have found a condition that guarantees the simultaneous evolution of plasticity and damage, and, finally, we have studied theoretically the evolution for typical tests, Section 2.3.4, which allows to establish the values for experimental

fitting relations. See a brief discussion regarding the influence of the nine parameters of the model in Section 2.3.5.

Some good properties follow from the choice of the damage functions $A_n(\alpha)$ and $A_t(\alpha)$ (2.19): the surface energy density ψ is convex with respect to the damage variable α; asymptotically, we recover an elasto-plastic model, since the hardening term vanished for $\alpha \to 1$; the fact that $A_s'(0) = +\infty$, $s \in \{n, t\}$ guarantees that plasticity and damage start at the same time; during a shear test or a traction test, the evolution is continuous, there is a peak of the stress followed by a softening phase, and two suitable conditions ensure the absence of snap-back. In particular, we have observed that the no-snap-back conditions (2.41) and (2.46) automatically guarantee the convexity of ψ with respect to the whole dissipative state variable $(\boldsymbol{p}, \alpha)$, see Proposition 2.12. There is only one property that we would like to improve about this model: the absence of saturation of dilatancy during monotonic and cyclic shear tests. This motivates the analysis of other models, notably the study of the influence of hyperelasticity in the following chapter.

3 Hyperelasticity

The first part of this chapter is based on a paper entitled "Hyperelastic nature of the Hoek–Brown criterion" [61] and submitted for publication. The preprint is available in Hal: https://hal.archives-ouvertes.fr/hal-03501788.

We analyze the influence of hyperelasticity on the plastic behavior of materials. A specific class of what we call hyperbolic elasticity arises from theoretical considerations as a straight consequence of plastic invariance of the elastic domain. The latter property of fixed residual plasticity is observed experimentally for many geomaterials. We superimpose hyperelastic effects on the plastic Drucker–Prager constitutive relation, widely used in geoscience. Curiously, we found that the hyperbolic nonlinearity, introduced through the Standard Generalized Material formalism, curves the initially linear surface to a quadratic one, that is assimilated to the generalized Hoek–Brown criterion. We conclude then that one possible justification of the empirical Hoek–Brown fit is the material hyperelastic nature. In the second part of the chapter, we adapt these considerations to a hyperelastic constitutive relation for joint modeling.

Contents

3.1 Introduction

One of the key points in the accurate description of rocks is the precise identification of their elastic domain (or yield surface/criterion). Due to the softening behavior, not only rocks but also all analogous materials like concrete, clay, soil or even ice [130] are commonly classified by their resistance to various mixed-mode loading. This approach is closely related to the basic safety rules in industrial applications, where it is often considered that geomaterials could exhibit unstable failure once the critical loading is reached [72].

Throughout the last century, specific testing machines and corresponding measurement protocols were established by national and international committees in order to harmonize and standardize the experimental characterization of the above-cited brittle materials. To mention only a few, the Brazilian tensile [21], oedometric, uni- and tri-axial compression are all well-documented experiments that are routinely executed to catalog material strength by spotting just some points of their multidimensional yield surface. This reduced "single point" vision of material resistance is increasingly scrutinized in recent times, and multiple evolutions have been adopted. For instance, constant improvements of finite element software enable new kinds of modeling, with loadings going beyond the elastic domain to explore a more subtle post-peak behavior. In the corresponding mechanical tests, the response of the material subjected to a set of pre-established loadings is analyzed during both the elastic and softening phase, enabling the full model parameter fitting. Some more sophisticated hybrid measurement techniques are also proposed, where the loading pass is adapted during the test execution. A single, all-in-one experiment, replaces the classical set with the same goal of full model identification [76]. While the complexity of post-peak description could be reached through various theoretical formalisms, most of them still rely on the initial yield surface definition, and the question of this elastic domain shape remains the cornerstone of any non-linear model identification. Even if considerable progress was made in recent decades, this precise identification of yield surface remains nowadays a rather challenging task [83].

A common feature of geomaterials is their strong resistance to compressive loading. For some large scale structures, like hydraulic dams or underground tunnel excavations, the construction material is naturally submitted to high levels of compression. For others, like nuclear confinement buildings or bridges, civil engineering constitutive concrete parts are pre-loaded to reach artificially an initial compression state by supplementary constraint of tension reinforcing steel tendons. In both cases, the property of higher compressive resistance is exploited on the industrial level with the aim of increasing global structure robustness.

According to the physical origin of geomaterials, a large variety of criteria defining the elastic domain are employed in order to model their mechanical behavior. The simplest surface, admitting infinite resistance in compression, is the linear cone-shaped one. It was first introduced more than a century ago and, depending on whether it is written in principal stresses or with help of rotational invariants, it's commonly called either Mohr–Coulomb [31, 95] or Drucker–Prager [50] criterion. Straight relation between shear and compressive loadings, which is the main signature of these linear criteria, has allowed the development of various constitutive relations based on the same simplified dependence [80, 3].

In the middle of the last century, with numerous large infrastructural projects ongoing, more complex criteria emerged as further experimental data became available. For wider ranges of loadings, the friction-type shear dependency seemed to be deflecting from the linear curve. Logically, a quadratic relation was to be explored first. Back in 1924, assuming the hypothesis of crack propagation via rapid growth of randomly distributed micro-flaws, Griffith had already obtained a theoretical justification of the parabolic yield shape [64]. Inspired by Griffith's model, Fairhurst proposed its empirical extension validated on the tensile Brazilian test [59]. In this spirit, in 1980 Evert Hoek and Edwin T. Brown [71] came up with a new particular shape of quadratic nonlinear criterion. It reproduced the Mohr–Coulomb type singularity for weak tensile loading simultaneously taking into account the reduction of shear resistance for high compression. Purely empirical, the Hoek–Brown criterion was originally obtained for intact rocks by two parameters fitting of the results of triaxial tests. Validated during the following years on a wider experimental database, the criterion was used extensively in the design of underground excavations [72].

It should be underlined that not only previously cited [64, 59, 71], but many other authors (e.g. [106]) kept the number of model parameters reduced, so that a better fit was obtained by the new curve's form itself, rather than by the addition of supplementary fitting variables. Consequently, the final expression of the failure criteria, being the pure result of a trial error process, appears quite artificial at first glance. *In this chapter we try to establish a possible hyperelastic link between the whole class of quadratic Hoek–Brown type criteria and their linear Mohr–Coulomb (Drucker–Prager) counterparts.* The single hypothesis of the existence of a stable failure surface under free-energy type description generates a subclass of quadratic yield surfaces from the linear relation of generalized plastic force. The Hoek–Brown relation is seen then as a consequence of the simultaneous presence of both plasticity and nonlinear elasticity phenomena in the geomaterial under investigation. The numerical results presented in Section 3.4 show the nonlinear influence of hyperelasticity in both hydrostatic triaxial case and uniaxial compression one with a confining pressure. Notably, during cyclic uniaxial tests with a confining pressure, a progressive saturation of dilatancy can be observed. This argument is then adapted to the joint modeling. Again, it is possible to determine a connection between a class of linear yield criteria in the plasticity generalized force space and a class of quadratic ones in the stress space.

3.2 Main idea in the nutshell

In this section, we summarize the main idea of the chapter reducing as far as possible technical details and complex notation that are required for a rigorous theoretical description.

Under the small deformation hypothesis, the thermodynamical description of continuum

mechanics relies on the Helmholtz free energy density definition. For isothermal evolution in presence of plasticity, this function depends on at least two state variables: the total strain ε and the plastic strain $\boldsymbol{p}$, i.e., $\phi(\varepsilon, \boldsymbol{p})$. The Cauchy stress tensor is then the dual conjugate to the total strain $\boldsymbol{\sigma} = \partial\phi/\partial\varepsilon$, while the energy response of the system on plastic strain evolution is captured through the generalized plastic force $\boldsymbol{X} = -\partial\phi/\partial\boldsymbol{p}$. Most elasto-plastic models admit additive separation of elastic and plastic strains. This leads to the simplest quadratic form of free energy describing residual plastic state :

$$\phi(\varepsilon, \boldsymbol{p}) = \frac{1}{2}\mathbb{E}(\varepsilon - \boldsymbol{p})^2, \tag{3.1}$$

where $\mathbb{E}$ is the constant elastic modulus. Notice that we adopt abridged notations for tensor operations. In particular, in this linear elastic case, the generalized plastic force and Cauchy stress are equal:

$$\boldsymbol{\sigma} = \boldsymbol{X} = \mathbb{E}(\varepsilon - \boldsymbol{p}). \tag{3.2}$$

The reversible elastic domain, characterized by the absence of plasticity evolution, is defined in the space of generalized force: $\boldsymbol{p} = \text{const} \Rightarrow \boldsymbol{X} \in \mathbb{K}_{\boldsymbol{X}}$. As we have mentioned in the introduction, considerable experimental efforts are focused on the identification of the elastic domain. In perfect plasticity hypothesis, the domain is fixed during evolution and it could consequently be considered as main material property. On further stage of model development, a plastic variable evolution is introduced, as it is commonly done either trough dissipation potential [66], plastic potential [121] or directly via explicit flow rule. Among the many possible existing extensions of the basic perfect plasticity constitutive model, we focus this work on the modification of the shape of the elastic domain generated by hyperelasticity [105].

Unlike Cauchy elasticity, that postulates linear (Hooke's law) or non linear straight stress-strain relationship, hyperelasticity admits the existence of scalar strain energy density function from which the stress is derived: $\boldsymbol{\sigma} = \partial\phi/\partial\varepsilon$. In this sense, the hyperelastic formalism is conservative and is compatible with a general thermodynamical description. Usually written for large strain [105], it is easily adapted to the small strain hypothesis [103].

Let us suppose that, in (3.1), the elastic modulus is some function of the total infinitesimal strain, i.e., $\mathbb{E} = \mathbb{E}(\varepsilon)$. The stress-strain relationship becomes nonlinear and the stress is no longer equal to the generalized plastic force:

$$\boldsymbol{X} = \mathbb{E}(\varepsilon - \boldsymbol{p}) \neq \boldsymbol{\sigma}.$$

Formal writing of the expression for the stress gives:

$$\boldsymbol{\sigma} = \mathbb{E}(\varepsilon - \boldsymbol{p}) + \frac{\partial\mathbb{E}}{2\partial\varepsilon}(\varepsilon - \boldsymbol{p})^2 = \boldsymbol{X} + \frac{\partial\mathbb{E}}{2\mathbb{E}^2\partial\varepsilon}\boldsymbol{X}^2. \tag{3.3}$$

If the elastic domain is fixed in the space of generalized force $\boldsymbol{X}$, in the presence of hyperelasticity it becomes strain dependent in the space of stresses $\boldsymbol{\sigma}$: $\mathbb{K}_{\boldsymbol{\sigma}}$. For small loading ($\boldsymbol{X} \ll 1$) the quadratic term $\sim \boldsymbol{X}^2$ could be considered as a nonlinear cinematic hardening, but in general the equation (3.3) states the nonlinear modification of the initial elastic domain $\mathbb{K}_{\boldsymbol{\sigma}} \neq \mathbb{K}_{\boldsymbol{X}}$. For plastifying cyclic loading, the initial yield surface is modified on each back-and-forth loop creating a mechanism similar to those introduced earlier by different authors,

like bounding surface in [36] or parent/child surfaces in Hujeux constitutive behaviour [8]. Therefore, the presence of hyperelasticity in plastic materials introduces not only a nonlinear stress-strain relationship, but also a modification of elastic domain represented in stress space.

Nonlinear behaviors of rocks, clays and soils are historically well known, but are often considered less relevant than other major mechanical phenomena, like fracture, plasticity, dilatancy etc. For most geomaterials the reversible elastic domain is comparatively small and hard to quantify experimentally. Nevertheless, some recent well documented experiments manage to separate and evaluate the elastic nonlinearity itself. For example, in [79] the authors conducted locally cyclic loadings on sand that revealed its hyperelastic properties. Even if this new experiment sheds light on this longtime forgotten phenomenon, the general hyperelastic coupling term is too complex to be fully identified. Natural question arise on how to handle all possible couplings in the multidimensional tensor relation $\mathbb{E}(\boldsymbol{\varepsilon})$. In this chapter, it is shown that hyperelastic coupling term can be simplified (3.4) enabling its analysis with the currently accessible experimental databases (e.g. [79]).

We introduce a particular class of hyperelasto-plastic materials that have fixed yield criterion in the stress space. This yield criterion can be considered either as initial elastic domain or as residual yield surface, for example for fully damaged elasto-plastic coupled to damage behavior [94]. As the elastic domain in the space of generalized force $\boldsymbol{X}$ is supposed to be fixed, it would also stay fixed in the space of stresses $\boldsymbol{\sigma}$ if and only if the coupling term in (3.3) is constant:

$$\frac{\partial \mathbb{E}}{2\mathbb{E}^2 \partial \boldsymbol{\varepsilon}} = \text{const.}$$

This condition defines one particular class of hyperelastic materials that keep the elastic domain fixed. This equation states that the material compliance should be linear with total strain:

$$\frac{\partial \mathbb{E}}{2\mathbb{E}^2 \partial \boldsymbol{\varepsilon}} = -\frac{\partial \mathbb{E}^{-1}}{2\partial \boldsymbol{\varepsilon}} = \text{const} \qquad \Longrightarrow \qquad \mathbb{E}^{-1} \sim \boldsymbol{\varepsilon}. \tag{3.4}$$

This formal writing of a tensor based expression is certainly over-simplified, but it allows us to propose a sub-class of hyperelastic materials that has fixed yield surface. For instance, we could reach this condition by setting both Lamé coefficients as hyperbolic functions of the strain trace. This particular case of hyperelasticity and subsequent constitutive relation obtained under the assumption of Standard Generalized Materials [66, 53] are analyzed in detail in the next section. In particular, we will show that the model constructed starting from a Drucker–Prager (linear) plasticity criterion in the generalized force space will lead to a nonlinear hyperelastic constitutive law with a quadratic Hoek–Brown type domain in the stress space.

3.3 Hyperelasticity with fixed plastic yield surface

Inspired by the ideas presented shortly above, in this section we display how to derive the elasto-plastic model with a three-dimensional Hoek–Brown–type yield criterion, adding a specific type of hyperelasticity to the initially linear plasticity criterion.

Above and throughout this section, we use the usual notation of mechanics: $\boldsymbol{u}$ is the unknown displacement field, $\boldsymbol{\varepsilon} \in \mathbb{R}^{3\times3}_{\text{sym}}$ is the strain tensor, i.e., the symmetric part of the

gradient of $\boldsymbol{u}$, $\boldsymbol{p} \in \mathbb{R}^{3\times3}_{\text{sym}}$ is the plasticity component of the strain tensor, i.e., $\boldsymbol{p} = \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\text{el}}$, $\boldsymbol{\sigma} \in \mathbb{R}^{3\times3}_{\text{sym}}$ is the stress tensor, and $\boldsymbol{X} \in \mathbb{R}^{3\times3}_{\text{sym}}$ is the thermodynamical force associated with plasticity, i.e., the plastic dual variable. Strain and plasticity tensors are the state variables of our model and, for simplicity, they are decomposed into their volumetric and deviatoric parts:

$$\boldsymbol{\varepsilon} = \frac{1}{3}\text{Tr}\,\boldsymbol{\varepsilon}\,\mathbf{I}_2 + \boldsymbol{\varepsilon}^D \qquad \text{and} \qquad \boldsymbol{p} = \frac{1}{3}\text{Tr}\,\boldsymbol{p}\,\mathbf{I}_2 + \boldsymbol{p}^D.$$

We adopt the usual convention in mechanics of continuous media for the sign of strain and stress, i.e., the stress is positive in traction and negative in compression.

3.3.1 Brief history of quadratic yield criteria

We start first with a brief description of the history of the introduction of quadratic yield surfaces, what we call "Hoek–Brown–type" yield surfaces.

Back in 1924, while studying the mechanical behavior of glasses, Griffith [64] was the first to derive from theoretical considerations a quadratic multi-axial criterion for fracture:

$$(\sigma_1 - \sigma_3)^2 \sim (\sigma_1 + \sigma_3)$$

where σ_1 and σ_3 the major and minor principal stresses, respectively.

Forty years later, Fairhust [59] attempted to empirically extend the work of Griffith for the domain of high compression suitable for rock behavior analysis.

Finally, in 1980 Hoek and Brown [71] obtained a criterion shape that convinced many generations of geomaterial scientists. The original Hoek–Brown criterion is still widely used in rock mechanics and, for intact rocks, it can be written as:

$$\sigma_1 = \sigma_3 + C_0\sqrt{m_i\frac{\sigma_3}{C_0} + 1},$$

where C_0 is the uniaxial compressive strength and m_i is a material constant for the intact rock. For more details about these constants, we refer to [73, 74], where a generalized version of the criterion involving the geological strength index (GSI) is also proposed and analyzed. Furthermore, the evolution of the Hoek–Brown criterion in the literature is summarized in the article [75].

Since many papers have exhibited the strong influence of the intermediate principal stress σ_2 (see, e.g., [85, 52] and the references therein), different three-dimensional extensions based on the Hoek–Brown criterion have been developed [109, 129, 111, 85]. In particular, the generalized form of the Pan–Hudson criterion proposed by X. D. Pan and J. Hudson [106] reads, for an intact rock, as :

$$\frac{3}{2C_0}\left\|\boldsymbol{\sigma}^D\right\|^2 + \frac{\sqrt{3}m_{\text{i}}}{2\sqrt{2}}\left\|\boldsymbol{\sigma}^D\right\| + m_{\text{i}}\sigma_m - C_0 = 0,$$

where σ_m and $\boldsymbol{\sigma}^D$ are the spherical and deviatoric part of the stress tensor, respectively, i.e.,

$$\boldsymbol{\sigma} = \sigma_m\,\mathbf{I}_2 + \boldsymbol{\sigma}^D, \qquad \text{where } \sigma_m = \frac{1}{3}\text{Tr}\,\boldsymbol{\sigma}, \tag{3.5}$$

and $\left\|\boldsymbol{\sigma}^D\right\| := \sqrt{\boldsymbol{\sigma}^D : \boldsymbol{\sigma}^D}$. Here, $\mathbf{I}_2$ denotes the second order identity tensor, and the double dot product is simply the double contraction operation for second order tensors. All the described criteria are parabolic and can be summarized as various choices of constants in the general expression:

$$f_{\boldsymbol{\sigma}} = A\left\|\boldsymbol{\sigma}^D\right\|^2 + B\left\|\boldsymbol{\sigma}^D\right\| + C\sigma_m - D = 0.$$

In this chapter we don't suppose, but derive quadratic yield criterion of this type under the assumption of Standard Generalized Materials [66] in a variational framework in presence of hyperelasticity. In particular, as we have already mentioned, we will show that this model can be constructed by starting from and elasto-plastic model with a Drucker–Prager (linear) plasticity criterion and introducing hyperbolic hyperelastic dependence in some material parameters. This will lead to a nonlinear elastic constitutive relation with Hoek–Brown type yield surface.

3.3.2 Energy density and derivation of the stress tensor

We begin our analysis by precising the tensor notation for the perfect elasto-plastic model already introduced in the previous section. In particular, the considered free energy density corresponds to the non-hardening version of [94, Equation (11)], where the material is assumed to be isotropic:

$$\begin{aligned}\phi(\boldsymbol{\varepsilon}, \boldsymbol{p}) &= \frac{1}{2}\mathbb{E}(\boldsymbol{\varepsilon} - \boldsymbol{p}) : (\boldsymbol{\varepsilon} - \boldsymbol{p}) \\ &= \frac{1}{2}K(\operatorname{Tr}\boldsymbol{\varepsilon} - \operatorname{Tr}\boldsymbol{p})^2 + \mu(\boldsymbol{\varepsilon}^D - \boldsymbol{p}^D) : (\boldsymbol{\varepsilon}^D - \boldsymbol{p}^D).\end{aligned} \tag{3.6}$$

Here, the action of the fourth order elasticity tensor $\mathbb{E}$ is described by $\mathbb{E}\boldsymbol{\tau} = \lambda \operatorname{Tr}\boldsymbol{\tau}\, \mathbf{I}_2 + 2\mu\boldsymbol{\tau}$ for any second order tensor $\boldsymbol{\tau}$, $\lambda > 0$ is the Lamé parameter, $\mu > 0$ is the shear modulus, and $K = \lambda + 2\mu/3$ is the compressibility modulus of the material. The dual variables $\boldsymbol{\sigma}$ and $\boldsymbol{X}$ can be obtained by deriving the energy density (3.6) with respect to the state variables $\boldsymbol{\varepsilon}$ and $\boldsymbol{p}$:

$$\boldsymbol{\sigma} := \frac{\partial \phi}{\partial \boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}, \boldsymbol{p}) \qquad \text{and} \qquad \boldsymbol{X} := -\frac{\partial \phi}{\partial \boldsymbol{p}}(\boldsymbol{\varepsilon}, \boldsymbol{p}). \tag{3.7}$$

In particular, we recall that in this case we simply obtain (3.2).

Now, we introduce some hyperelasticity dependencies in the elasticity tensor, and in particular we suppose $\mathbb{E} = \mathbb{E}(\operatorname{Tr}\boldsymbol{\varepsilon})$. Indeed, as discussed in the previous section, material compliance need to be linear function of strain (3.4). If trace dependence is natural, the deviator one is not obvious as it wouldn't be derivable on hydrostatic axis. Here, in particular, we meet the required conditions by assuming:

$$K(\operatorname{Tr}\boldsymbol{\varepsilon}) = \frac{K_i}{2K_i\beta_m \operatorname{Tr}\boldsymbol{\varepsilon} + 1} \qquad \text{and} \qquad \mu(\operatorname{Tr}\boldsymbol{\varepsilon}) = \frac{\mu_i}{4\mu_i\beta^D \operatorname{Tr}\boldsymbol{\varepsilon} + 1}, \tag{3.8}$$

where $K_i > 0$ and $\mu_i > 0$ are the initial compressibility and shear moduli of the sound material, $\beta_m \geq 0$ and $\beta^D \geq 0$ are the hyperelastic parameters of the model. In addition, we restrict ourselves to the tests in which

$$\operatorname{Tr}\boldsymbol{\varepsilon} > \varepsilon_0 := \max\left\{-\frac{1}{2K_i\beta_m}, -\frac{1}{4\mu_i\beta^D}\right\}$$

in order to ensure the positivity of $K_0(\mathrm{Tr}\,\varepsilon)$ and $\mu_0(\mathrm{Tr}\,\varepsilon)$. Denoting with X_m and $\boldsymbol{X}^D$ the spherical and deviatoric components of $\boldsymbol{X}$ as we have done for the stress (3.5), the expression of $\boldsymbol{\sigma}$ and $\boldsymbol{X}$ can be easily obtained thanks to (3.7):

$$\begin{cases} \sigma_m = \dfrac{K_i}{2K_i\beta_m \mathrm{Tr}\,\varepsilon + 1}(\mathrm{Tr}\,\varepsilon - \mathrm{Tr}\,\boldsymbol{p}) - \dfrac{K_i^2\beta_m}{(2K_i\beta_m \mathrm{Tr}\,\varepsilon + 1)^2}(\mathrm{Tr}\,\varepsilon - \mathrm{Tr}\,\boldsymbol{p})^2 \\ \qquad\qquad - \dfrac{4\mu_i^2\beta^D}{(4\mu_i\beta^D \mathrm{Tr}\,\varepsilon + 1)^2}(\varepsilon^D - \boldsymbol{p}^D) : (\varepsilon^D - \boldsymbol{p}^D), \\ \boldsymbol{\sigma}^D = \dfrac{2\mu_i}{4\mu_i\beta^D \mathrm{Tr}\,\varepsilon + 1}(\varepsilon^D - \boldsymbol{p}^D), \end{cases} \tag{3.9}$$

and

$$\begin{cases} X_m = \dfrac{K_i}{2K_i\beta_m \mathrm{Tr}\,\varepsilon + 1}(\mathrm{Tr}\,\varepsilon - \mathrm{Tr}\,\boldsymbol{p}), \\ \boldsymbol{X}^D = \dfrac{2\mu_i}{4\mu_i\beta^D \mathrm{Tr}\,\varepsilon + 1}(\varepsilon^D - \boldsymbol{p}^D). \end{cases} \tag{3.10}$$

As a consequence, $\boldsymbol{\sigma}$ and $\boldsymbol{X}$ are connected through the following relation:

$$\boldsymbol{\sigma} = \boldsymbol{X} - \left(\beta_m X_m^2 + \beta^D \boldsymbol{X}^D : \boldsymbol{X}^D\right)\mathbf{I}_2. \tag{3.11}$$

Notice that this relation only involves the hyperelastic parameters β_m and β^D.

3.3.3 The plasticity criterion

For our model, we consider the Drucker–Prager criterion in the $\boldsymbol{X}$–space:

$$f_{\boldsymbol{X}}(\boldsymbol{X}) := \frac{1}{\sqrt{6}}\left\|\boldsymbol{X}^D\right\| + aX_m - b = 0, \tag{3.12}$$

where $a > 0$ and $b \geq 0$, and we recall that $\|\boldsymbol{\tau}\| := \sqrt{\boldsymbol{\tau} : \boldsymbol{\tau}}$. The corresponding elastic domain $\mathbb{K}_{\boldsymbol{X}} := \{\boldsymbol{X}^* \in \mathbb{R}^{3\times3}_{\mathrm{sym}} : f_{\boldsymbol{X}}(\boldsymbol{X}^*) \leq 0\}$ is a convex cone with a singular point at $\left(\frac{b}{a}, \mathbf{0}\right)$, and the behavior remains elastic while $f_{\boldsymbol{X}}(\boldsymbol{X}) < 0$. Moreover, consistently with the Standard Generalized Materials framework, we consider an associative model, i.e., we assume that the plasticity evolution follows the normality rule. As a consequence, in the points in which the boundary is smooth, i.e., where the function $f_{\boldsymbol{X}}$ is differentiable, we have

$$\dot{\boldsymbol{p}} = \dot{\lambda}\left(\frac{1}{\sqrt{6}}\frac{\boldsymbol{X}^D}{\|\boldsymbol{X}^D\|} + \frac{a}{3}\mathbf{I}_2\right), \qquad \dot{\lambda} \geq 0. \tag{3.13}$$

For further details, we refer the reader to [94, Section 2]. In addition, if $b \neq 0$, we assume

$$\beta_m \leq \frac{a}{2b}. \tag{3.14}$$

Combining (3.12) with (3.11), one can obtain the explicit expression of the plasticity criterion in the stress space:

$$f_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}) := \frac{1}{6}\left(\beta_m + 6a^2\beta^D\right)\left\|\boldsymbol{\sigma}^D\right\|^2 + \frac{1}{\sqrt{6}}\left(a - 2\beta_m b\right)\left\|\boldsymbol{\sigma}^D\right\| + a^2\sigma_m - b\left(a - \beta_m b\right) = 0. \tag{3.15}$$

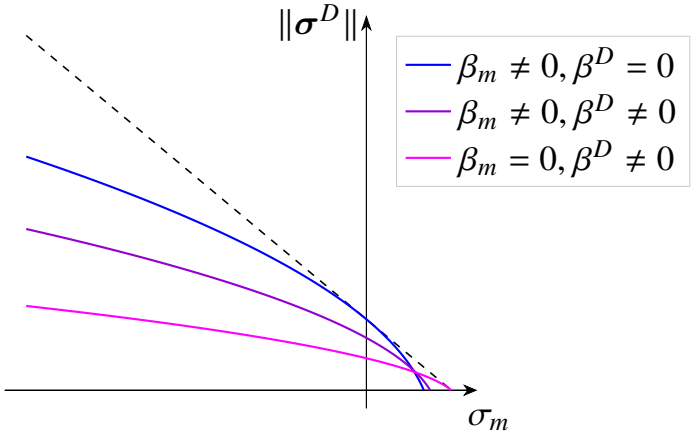


Figure 3.1 – Transformation of the failure criterion for different values of the hyperelastic coefficients β_m and β^D. The black dashed line represents the linear failure criterion (3.12).

Therefore, starting from the linear failure criterion (3.12) for the generalized force $\boldsymbol{X}$, we obtain a quadratic failure criterion for the stress tensor $\boldsymbol{\sigma}$. The corresponding reversibility domain $\mathbb{K}_{\boldsymbol{\sigma}} := \{\boldsymbol{\sigma}^* \in \mathbb{R}^{3\times3}_{\text{sym}} : f_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}^*) \leq 0\}$ is a convex cone with parabolic boundary and a singular point in $\left(\frac{b(a-\beta_m b)}{a^2}, \mathbf{0}\right)$. Notice that, since (3.15) does not depend on $\boldsymbol{\varepsilon}$ or $\boldsymbol{p}$, $\mathbb{K}_{\boldsymbol{\sigma}}$ remains fixed for any type of evolution. Figure 3.1 show the transformation of the failure criterion from linear to quadratic for three different choices of the hyperelasticity parameters (β_m, β^D).

Remark 3.1 (Motivations of assumption (3.14))**.** *Assumption* (3.14) *ensures that the domain in the stress space is convex and contains the origin* $\mathbf{0}$ *of the space. Furthermore, with this condition, the quadratic failure criterion can be written as*

$$\frac{1}{\sqrt{6}}\left\|\boldsymbol{\sigma}^D\right\| = \frac{-(a-2\beta_m b) + a\sqrt{1-4\sigma_m(\beta_m+6a^2\beta^D)+24\beta^D b(a-\beta_m b)}}{2(\beta_m+6a^2\beta^D)}.$$

3.3.4 Implementation considerations

In this section we propose some considerations about the implementation of the constitutive relation determined by (3.9), the plasticity criterion (3.12), and the normal flow rule (3.13).

With the aim of computing the evolution of a material point under the hypothesis of a quasi-static nonlinear problem, we discretize the problem using an incremental approach. For simplicity, for any tensor or scalar variable x, its value at the previous equilibrium state, at the current equilibrium state, and its increment are denoted by x^-, x and $\Delta x := x - x^-$, respectively. Then, in order to solve locally the evolution of a point we have to satisfy the

following equations:

$$\boldsymbol{\sigma} = \mathbb{E}(\mathrm{Tr}\,\varepsilon)(\varepsilon - \boldsymbol{p}) + \frac{1}{2}\frac{\partial \mathbb{E}(\mathrm{Tr}\,\varepsilon)}{\partial \mathrm{Tr}\,\varepsilon}(\varepsilon - \boldsymbol{p}) : (\varepsilon - \boldsymbol{p}), \qquad \text{(definition of } \boldsymbol{\sigma}) \tag{3.16}$$

$$\boldsymbol{X} = \mathbb{E}(\mathrm{Tr}\,\varepsilon)(\varepsilon - \boldsymbol{p}), \qquad \text{(definition of } \boldsymbol{X}) \tag{3.17}$$

$$f_{\boldsymbol{X}}(\boldsymbol{X}) \leq 0, \qquad \text{(yield criterion)} \tag{3.18}$$

$$\Delta \boldsymbol{p} = \Delta\lambda \frac{\partial f_{\boldsymbol{X}}(\boldsymbol{X})}{\partial \boldsymbol{X}}, \Delta\lambda \geq 0, \qquad \text{(flow rule)} \tag{3.19}$$

$$f_{\boldsymbol{X}}(\boldsymbol{X})\Delta\lambda = 0. \qquad \text{(plastic evolution)} \tag{3.20}$$

Here, ε, $\boldsymbol{p}^-$, and, as a consequence, $\boldsymbol{\sigma}^-$ and $\boldsymbol{X}^-$ are known, and the goal is to find the current values $\boldsymbol{p}$, $\boldsymbol{\sigma}$ and $\boldsymbol{X}$. The implementation of this system is composed by two main blocks which correspond to the phase of elastic prediction and to the phase of correction with elasto-plastic evolution.

Remark 3.2 (Stress computation)**.** *The stress tensor $\boldsymbol{\sigma}$ can be computed directly from $\boldsymbol{X}$ thanks to the relation* (3.11)*, instead of using the explicit expression* (3.16)*. As a consequence, for simplicity, in the following we will only consider* (3.17)–(3.20)*, for which the unknowns are $\Delta\boldsymbol{p}$, $\boldsymbol{X}$, and $\Delta\lambda$.*

During the phase of elastic prediction, we define

$$\boldsymbol{X}^{\mathrm{pred}} := \mathbb{E}(\mathrm{Tr}\,\varepsilon)(\varepsilon - \boldsymbol{p}^-).$$

If $f_{\boldsymbol{X}}(\boldsymbol{X}^{\mathrm{pred}}) \leq 0$, then the solution is simply

$$(\Delta\boldsymbol{p}, \boldsymbol{X}, \Delta\lambda) = (\boldsymbol{0}, \boldsymbol{X}^{\mathrm{pred}}, 0).$$

Otherwise, we have to find $(\Delta\boldsymbol{p}, \boldsymbol{X}, \Delta\lambda) \in \mathbb{R}^{3\times3}_{\mathrm{sym}} \times \mathbb{R}^{3\times3}_{\mathrm{sym}} \times \mathbb{R}^+$ such that

$$\begin{cases} \boldsymbol{X} = \mathbb{E}(\mathrm{Tr}\,\varepsilon)(\varepsilon - \boldsymbol{p}^- - \Delta\boldsymbol{p}), \\ \Delta\boldsymbol{p} = \Delta\lambda \dfrac{\partial f_{\boldsymbol{X}}(\boldsymbol{X})}{\partial \boldsymbol{X}}, \\ f_{\boldsymbol{X}}(\boldsymbol{X}) = 0. \end{cases} \tag{3.21}$$

Remark 3.3. *The system* (3.21) *with unknowns $(\Delta\boldsymbol{p}, \boldsymbol{X}, \Delta\lambda) \in \mathbb{R}^{3\times3}_{\mathrm{sym}} \times \mathbb{R}^{3\times3}_{\mathrm{sym}} \times \mathbb{R}^+$ has a symmetric jacobian:*

$$\tilde{\mathbb{J}} = \begin{pmatrix} \mathbb{E}(Tr\,\varepsilon) & \mathbf{I}_4 & \boldsymbol{0} \\ \mathbf{I}_4 & -\Delta\lambda \dfrac{\partial^2 f_{\boldsymbol{X}}(\boldsymbol{X})}{\partial \boldsymbol{X}^2} & -\dfrac{\partial f_{\boldsymbol{X}}(\boldsymbol{X})}{\partial \boldsymbol{X}} \\ \boldsymbol{0} & -\dfrac{\partial f_{\boldsymbol{X}}(\boldsymbol{X})}{\partial \boldsymbol{X}} & 0 \end{pmatrix},$$

where $\mathbf{I}_4$ denotes the fourth order identity tensor. As a consequence, we can easily define an energy function $\tilde{\phi}(\Delta\boldsymbol{p}, \boldsymbol{X}, \Delta\lambda)$ such that solving the problem (3.21) *is equivalent to find an extrema of $\tilde{\phi}$. In particular,*

$$\tilde{\phi}(\Delta\boldsymbol{p}, \boldsymbol{X}, \Delta\lambda) = \frac{1}{2}\mathbb{E}(Tr\,\varepsilon)\Delta\boldsymbol{p} : \Delta\boldsymbol{p} + \Delta\boldsymbol{p} : (\boldsymbol{X} - \mathbb{E}(Tr\,\varepsilon)(\varepsilon + \boldsymbol{p}^-)) - \Delta\lambda f_{\boldsymbol{X}}(\boldsymbol{X})$$

By defining the elastic part of the strain tensor as $\varepsilon^{\mathrm{el}} := \varepsilon - \boldsymbol{p}$, we have

$$\begin{cases} \boldsymbol{X} = \mathbb{E}(\mathrm{Tr}\,\varepsilon)(\varepsilon^{\mathrm{el},-} + \Delta\varepsilon^{\mathrm{el}}), & (3.22a)\\ \Delta\varepsilon^{\mathrm{el}} = \Delta\varepsilon - \Delta\boldsymbol{p}, & (3.22b)\\ \Delta\boldsymbol{p} = \Delta\lambda\left(\dfrac{1}{\sqrt{6}}\dfrac{\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}}{\|\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}\|} + \dfrac{a}{3}\mathbf{I}_2\right), & (3.22c)\\ \dfrac{2}{\sqrt{6}}\mu(\mathrm{Tr}\,\varepsilon)\|\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}\| + aK(\mathrm{Tr}\,\varepsilon)\mathrm{Tr}\,(\varepsilon^{\mathrm{el},-} + \Delta\varepsilon^{\mathrm{el}}) - b = 0. & (3.22d) \end{cases}$$

Notice that the last two equations does not depend explicitly on $\boldsymbol{X}$ anymore, and the latter can be easily computed once we have the increment $\Delta\varepsilon^{\mathrm{el}}$. As a consequence, substituting (3.22c) into (3.22b) the problem is reduced to: Find $(\Delta\varepsilon^{\mathrm{el}}, \Delta\lambda) \in \mathbb{R}^{3\times3}_{\mathrm{sym}} \times \mathbb{R}^{+}$ such that

$$\begin{cases} \Delta\varepsilon^{\mathrm{el}} - \Delta\varepsilon + \Delta\lambda\left(\dfrac{1}{\sqrt{6}}\dfrac{\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}}{\|\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}\|} + \dfrac{a}{3}\mathbf{I}_2\right) = 0,\\ \dfrac{2}{\sqrt{6}}\mu(\mathrm{Tr}\,\varepsilon)\|\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}\| + aK(\mathrm{Tr}\,\varepsilon)\mathrm{Tr}\,(\varepsilon^{\mathrm{el},-} + \Delta\varepsilon^{\mathrm{el}}) - b = 0. \end{cases}$$

This nonlinear problem can be solved with some iterative methods like the Newton method, for which we have to compute the Jacobian matrix:

$$\mathbb{J} = \begin{pmatrix} \mathbb{J}_1 & \mathbb{J}_2 \\ \mathbb{J}_3 & 0 \end{pmatrix},$$

where

$$\mathbb{J}_1 := \mathbf{I}_4 + \frac{\Delta\lambda}{\sqrt{6}}\left[\left(\mathbf{I}_4 - \frac{1}{3}\mathbf{I}_2 \otimes \mathbf{I}_2\right)\frac{1}{\|\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}\|} - \frac{(\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}) \otimes (\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}})}{\|\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}\|^3}\right],$$

$$\mathbb{J}_2 := \frac{1}{\sqrt{6}}\frac{\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}}{\|\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}\|} + \frac{a}{3}\mathbf{I}_2,$$

$$\mathbb{J}_3 := \frac{2}{\sqrt{6}}\mu(\mathrm{Tr}\,\varepsilon)\frac{\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}}{\|\varepsilon^{\mathrm{el,D},-} + \Delta\varepsilon^{\mathrm{el,D}}\|} + aK(\mathrm{Tr}\,\varepsilon)\mathbf{I}_2,$$

and $\otimes$ denotes the tensor product.

3.4 Numerical results

The aim of this section is to show a panel of examples of evolution using the proposed hyperelastic model in some typical test cases. In particular, we want to exhibit the influence of the hyperelastic parameters and to provide a comparison with the linear elasto-plastic model. The results are obtained with the open source code generation tool `mfront` (see [70] and also `http://tfel.sourceforge.net`). For all tests we start from a natural reference configuration with $\boldsymbol{p} = \mathbf{0}$, and we set the Poisson parameter $\nu = 0.3$, which corresponds to the compressibility modulus $K_i = 5E/6$ and to the shear modulus $\mu_i = 5E/13$.

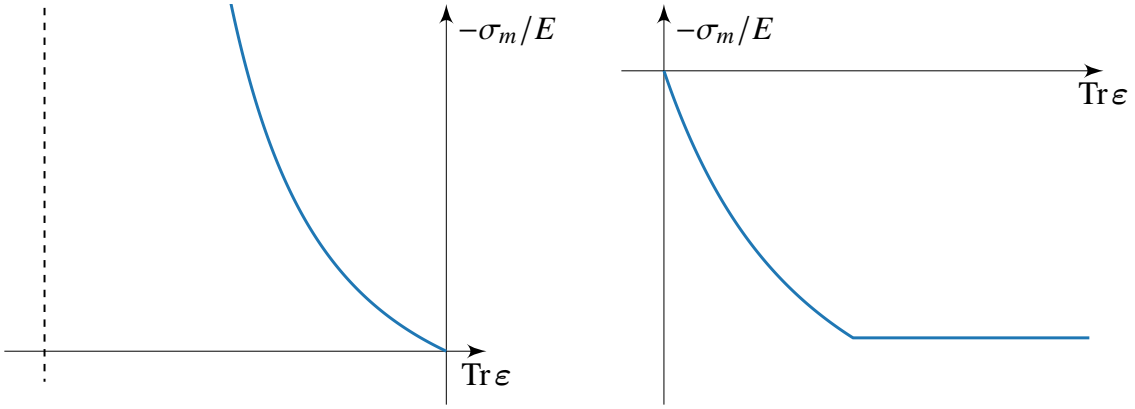


Figure 3.2 – Graphs of hydrostatic triaxial test: compression (*left*) and traction (*right*).

3.4.1 Hydrostatic triaxial tests

During a hydrostatic triaxial test, the stress remains on the hydrostatic axis throughout the loading. This corresponds to the assumption that only the spherical part of the stress evolves, i.e., $\boldsymbol{\sigma} = \bar{\sigma}\mathbf{I}_2$, with $\bar{\sigma} < 0$ for compression tests, and $\bar{\sigma} > 0$ for traction tests. For simplicity, we suppose that the lower boundary for the evolution of the volumetric part of the strain is $\varepsilon_0 = -\dfrac{1}{2K_i\beta_m}$, i.e., $K_i\beta_m \geq 2\mu_i\beta^D$. From the second equation of (3.9) we have immediately that $\boldsymbol{\varepsilon}^D = \mathbf{0}$. The behavior is always hyperelastic, i.e., $\boldsymbol{\sigma} \in \mathrm{Int}\mathbb{K}_{\boldsymbol{\sigma}}$, during a hydrostatic compression test. In particular, the evolution is simply described by

$$\boldsymbol{\sigma} = \sigma_m\mathbf{I}_2 = \left[\frac{K_i}{2K_i\beta_m\mathrm{Tr}\,\varepsilon + 1}\mathrm{Tr}\,\varepsilon - \frac{K_i^2\beta_m}{(2K_i\beta_m\mathrm{Tr}\,\varepsilon + 1)^2}(\mathrm{Tr}\,\varepsilon)^2\right]\mathbf{I}_2,$$

or

$$\boldsymbol{\varepsilon} = \frac{\mathrm{Tr}\,\boldsymbol{\varepsilon}}{3}\mathbf{I}_2 = \left[\frac{1}{6K_i\beta_m}\left(-1 + \frac{1}{\sqrt{1 - 4\beta_m\sigma_m}}\right)\right]\mathbf{I}_2.$$

In a traction test, there is an initial hyperelastic behavior until σ_m reaches his maximum value $\dfrac{b(a - \beta_m b)}{a^2}$ determined by the plasticity criterion (3.15). Then, the spherical part of the stress remains constant and plasticity evolves. One can see the corresponding graphs of the evolution of the hydrostatic stress σ_m as function of $\mathrm{Tr}\,\boldsymbol{\varepsilon}$ in Figure 3.2, in the case of compression (*left*) and traction (*right*) with parameters $a = 1$, $b = E$, $\beta_m = 1/(5E)$, and $\beta^D \leq 13\beta_m/12$.

3.4.2 Uniaxial compression test with a confining pressure

A unilateral compression test with a confining pressure is divided in two phases:

- at first, a hydrostatic compression is performed reaching the value of pressure p_0,
- then, we compress along the z–axis maintaining the lateral pressure constant.

As we have already seen in the previous subsection, during the confining phase the behavior is elastic, i.e., $\boldsymbol{p} = \mathbf{0}$, and only the hydrostatic component evolves, i.e., $\boldsymbol{\sigma}^D = \boldsymbol{\varepsilon}^D = \mathbf{0}$. At the end of this stage $\boldsymbol{\sigma} = -p_0 \mathbf{I}_2$ and $\boldsymbol{\varepsilon} = \bar{\varepsilon}\, \mathbf{I}_2$, where $p_0 > 0$ and

$$\bar{\varepsilon} := \frac{1}{6K_i\beta_m}\left(-1 + \frac{1}{\sqrt{4\beta_m p_0 + 1}}\right).$$

Then, during the second phase of the loading, we prescribe the evolution of ε_z with $\dot{\varepsilon}_z < 0$, while the lateral pressure stays equal to p_0:

$$\boldsymbol{\sigma} = \begin{pmatrix} -p_0 & 0 & 0 \\ 0 & -p_0 & 0 \\ 0 & 0 & \sigma_z \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_x & 0 & 0 \\ 0 & \varepsilon_x & 0 \\ 0 & 0 & \varepsilon_z \end{pmatrix}.$$

In addition, in this case we have $\sigma_m = \frac{1}{3}(\sigma_z - 2p_0)$ and $\|\boldsymbol{\sigma}^D\| = \sqrt{\frac{2}{3}}|p_0 + \sigma_z|$, and we define the deviatoric stress $q := p_0 + \sigma_z$. After an initial elastic phase, the boundary of the reversibility domain $\mathbb{K}_{\boldsymbol{\sigma}}$ is reached, and, as a consequence, plasticity starts to evolve following the normality rule.

Figure 3.3 shows the evolution curves for this kind of test, highlighting with different colors the stages of confining compression, hyperelastic, and plastic evolution. The parameters are given by:

$$a = 0.25, \qquad b = \frac{E}{10}, \qquad \beta_m = \frac{1}{4E}, \qquad \beta^D = \frac{3}{10E}, \qquad p_0 = \frac{E}{10}. \tag{3.23}$$

The nonlinear influence is evident especially during the hyperelastic phase (in *green*) in Figures 3.3b and 3.3c. During the plastic phase the average stress σ_m and the deviatoric stress q stay constant (see Figures 3.3a and 3.3c), since in the model we propose the is no hardening coefficient. Moreover, Figure 3.3b displays the presence of dilatancy without saturation, which is a consequence of the normal flow rule for the evolution of plasticity.

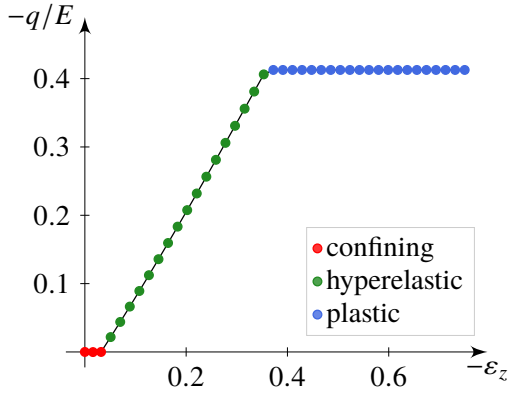
3.4.3 Cyclic test

In this subsection we consider a cyclic test in which at first we perform a hydrostatic compression, and then we decrease and increase cyclically the axial strain ε_x maintaining the lateral pressure constant. The parameters are again fixed by (3.23). The results are displayed in Figure 3.4. Notice that the main influence of hyperelasticity is the progressive accomodation of the values. This is particularly relevant for the saturation of dilatancy, Figure 3.4b.

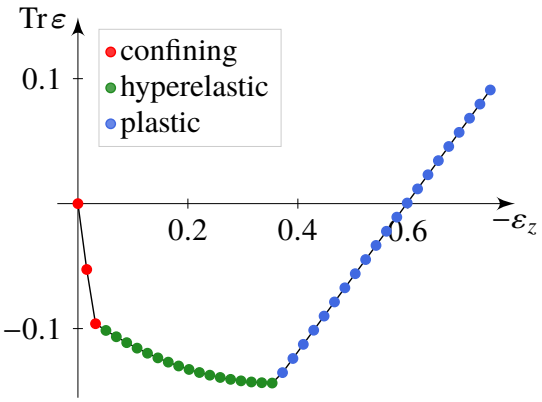
3.4.4 Radial loadings

In this last numerical example, we consider some radial tests, i.e., loadings in which the ratio $\eta := \frac{q}{\sigma_m}$ is constant during a compression. Assuming that

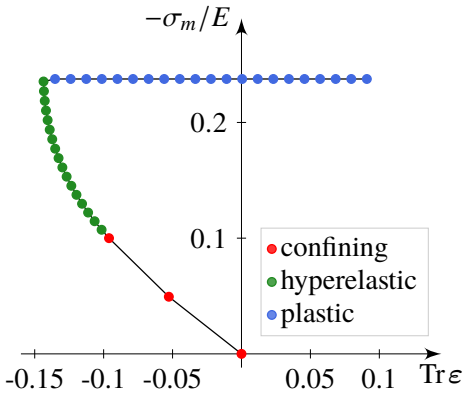
$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_x & 0 & 0 \\ 0 & \sigma_x & 0 \\ 0 & 0 & \sigma_z \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_x & 0 & 0 \\ 0 & \varepsilon_x & 0 \\ 0 & 0 & \varepsilon_z \end{pmatrix},$$



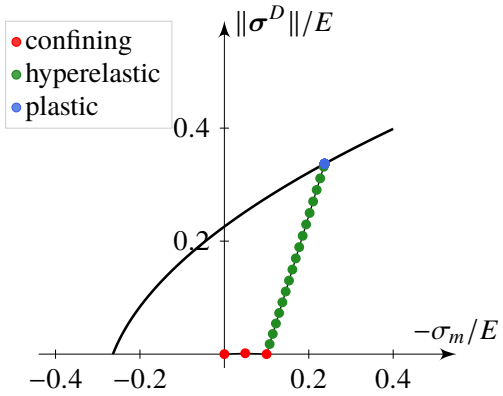
(a) Deviatoric stress $q = \sigma_x - \sigma_z$ vs. axial strain ε_z.



(b) Volumetric strain $\mathrm{Tr}\,\varepsilon$ vs. axial strain ε_z.



(c) Average stress σ_m vs. volumetric strain $\mathrm{Tr}\,\varepsilon$.



(d) Evolution in the stress domain.

Figure 3.3 – Graphs of uniaxial compression test with a confining pressure. The different stages of the evolution are shown with different colors: hydrostatic confining (*red*), hyperelastic (*green*), and plastic (*blue*) stage.

we have

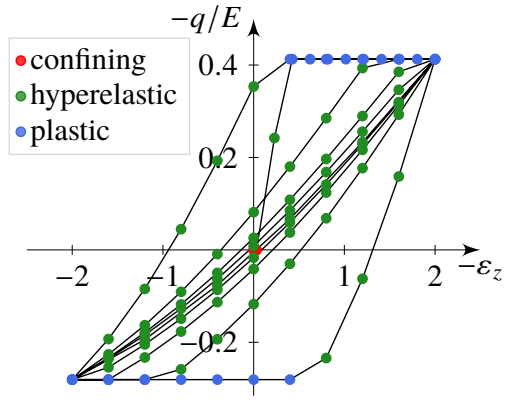
$$\eta = 3\frac{\sigma_z - \sigma_x}{2\sigma_x + \sigma_z} = \pm\sqrt{\frac{3}{2}}\frac{\|\boldsymbol{\sigma}^D\|}{\sigma_m} = const,$$

which corresponds to

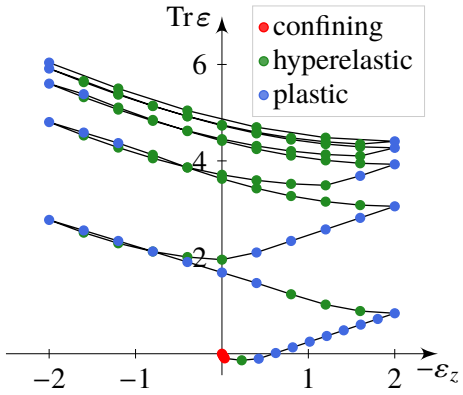
$$\sigma_z = \frac{3 + 2\eta}{3 - \eta}\sigma_x, \qquad \eta \in \mathbb{R}.$$

Figure 3.5 shows the elastic radial evolution for five different values for η and the parameters (3.23). The case $\eta = 0$ (*violet*) corresponds to a hydrostatic compression, i.e., $\|\boldsymbol{\sigma}^D\| = \mathbf{0}$ and $\boldsymbol{\sigma}$ lies on the hydrostatic axis in Figure 3.5c. For all the other cases, one can notice a nonlinear behavior in the $(q, \mathrm{Tr}\,\varepsilon)$–space, Figure 3.5a. This kind of behavior can be observed for some geomaterial, we cite for example [88, Figure 9]. In the stress space, the evolution is linear, Figure 3.5c, and the angle θ with the hydrostatic axis depends on the absolute value of η:

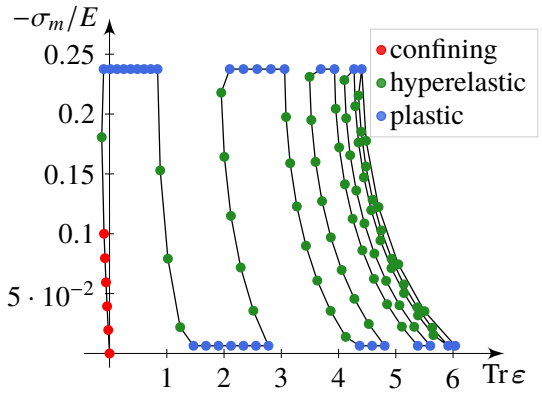
$$\tan\theta = -\sqrt{\frac{2}{3}}|\eta|.$$



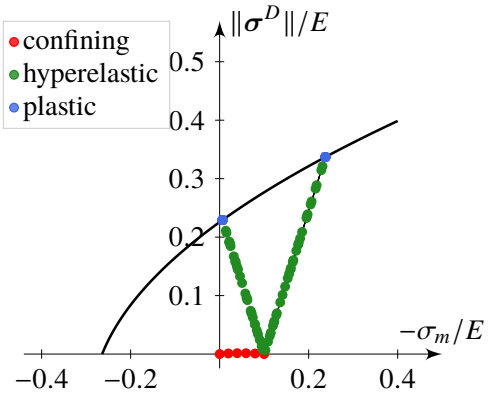
(a) Deviatoric stress $q = \sigma_x - \sigma_z$ vs. axial strain ε_z.



(b) Volumetric strain $\mathrm{Tr}\,\varepsilon$ vs. axial strain ε_z.



(c) Average stress σ_m vs. volumetric strain $\mathrm{Tr}\,\varepsilon$.



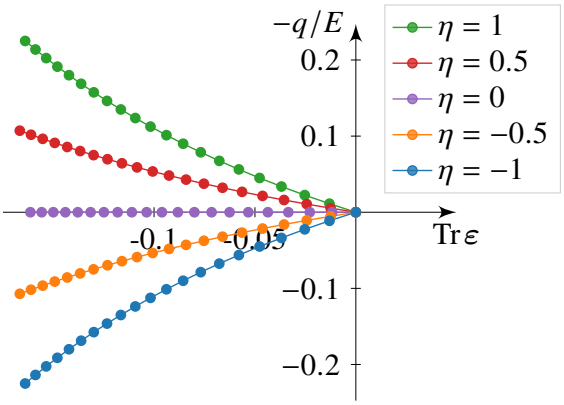
(d) Evolution in the stress domain.

Figure 3.4 – Graphs of cyclic test. The different stages of the evolution are shown with different colors: hydrostatic confining (*red*), hyperelastic (*green*), and plastic (*blue*) stage.

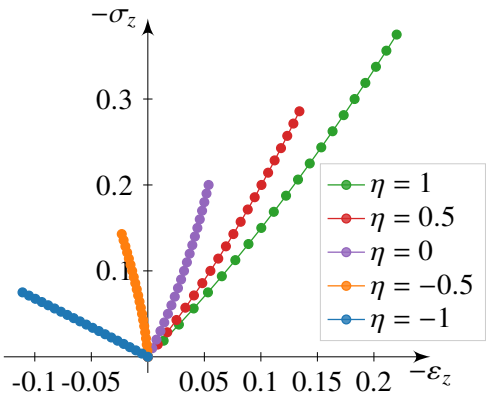
3.5 Application to the joint model

The model we have presented in Section 3.3 can be easily adapted to the joint modeling, that we have described in the previous chapter. In this section we show explicitly this adaptation.

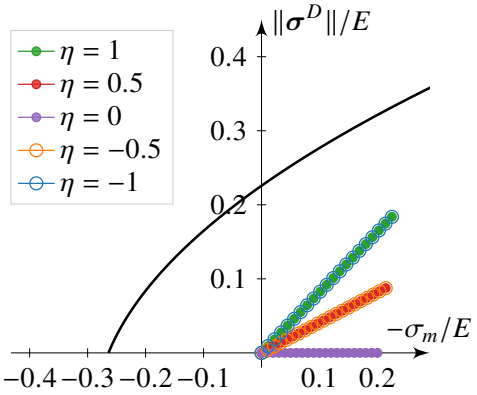
We recall briefly the context of this model, see Section 2.1 and Subsection 2.2.1 for further details. We consider an elastic object represented by the domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, with one discontinuity Γ_C, Figure 2.3a. On Γ_C the jump of displacement is defined by $\boldsymbol{\delta} = (\delta_n, \delta_{t_1}, \delta_{t_2})^\top := -[\![\boldsymbol{u}]\!]$, where $\boldsymbol{u} \colon \Omega \to \mathbb{R}^d$ is the displacement of the domain Ω, and the behavior of the interface is determined by a constitutive relation $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\delta})$ which can possibly depend on other state variables. Here, $\boldsymbol{\sigma} = (\sigma_n, \sigma_{t_1}, \sigma_{t_2})^\top$ represent the cohesive force on the interface. Moreover, we denote with $\boldsymbol{p} = (p_n, p_{t_1}, p_{t_2})^\top$ the plastic component of the displacement jump, and we search the displacement $\boldsymbol{u}$ that minimizes the total energy $E_{\mathrm{tot}}(\boldsymbol{u})$, sum of the elastic energy of the domain, the surface energy and the work of the external forces, see (2.6). In particular, we consider an elasto-plastic model in which the surface energy is expressed by $\Psi(\boldsymbol{\delta}, \boldsymbol{p}) := \int_{\Gamma_C} \psi(\boldsymbol{\delta}, \boldsymbol{p}) \, \mathrm{d}\Gamma$, where $\psi(\boldsymbol{\delta}, \boldsymbol{p})$ is surface energy density function on the interface Γ_C.



(a) Deviatoric stress q vs. volumetric strain $\mathrm{Tr}\,\varepsilon$.



(b) Axial stress σ_z vs. axial strain ε_z.



(c) Evolution in the stress domain.

Figure 3.5 – Graphs of radial loadings with different values of $\eta = q/\sigma_m$.

3.5.1 Surface energy density

In the model we propose the surface energy density is given by:

$$\begin{aligned}\psi(\boldsymbol{\delta}, \boldsymbol{p}) &= \frac{1}{2}(\boldsymbol{\delta} - \boldsymbol{p})^\top \mathbb{E}(\delta_n)\,(\boldsymbol{\delta} - \boldsymbol{p}) \\ &= K_n(\delta_n)\,\frac{(\delta_n - p_n)^2}{2} + K_t(\delta_n)\,\frac{\|\boldsymbol{\delta_t} - \boldsymbol{p_t}\|^2}{2}.\end{aligned} \tag{3.24}$$

Here, $\|\cdot\|$ is the Euclidean norm of $\mathbb{R}^{d-1}$,

$$\mathbb{E}(\delta_n) := \begin{pmatrix} K_n(\delta_n) & 0 & 0 \\ 0 & K_t(\delta_n) & 0 \\ 0 & 0 & K_t(\delta_n) \end{pmatrix} \tag{3.25}$$

is the nonlinear elastic matrix, and $K_n(\delta_n)$ and $K_t(\delta_n)$ are the normal and tangential rigidity coefficients. We assume that the expressions of these coefficients are similar to those of $K(\mathrm{Tr}\,\varepsilon)$ and $\mu(\mathrm{Tr}\,\varepsilon)$ (3.8) of the previous section:

$$K_n(\delta_n) = \frac{K_{n,0}}{2K_{n,0}\beta_n\delta_n + 1} \qquad \text{and} \qquad K_t(\delta_n) = \frac{K_{t,0}}{2K_{t,0}\beta_t\delta_n + 1}, \tag{3.26}$$

where $K_{n,0}$ and $K_{t,0}$ are two positive parameters that represent the normal and tangential rigidity coefficient for $\delta_n = 0$, respectively, and $\beta_n \geq 0$ and $\beta_t \geq 0$ are some hyperelastic coefficients. We suppose that at least one between β_n and β_t is greater than zero. In the following, we restrict ourselves to evolutions that satisfy

$$\delta_n > \delta_0 := \max\left\{-\frac{1}{2K_{n,0}\beta_n}, -\frac{1}{2K_{t,0}\beta_t}\right\}, \tag{3.27}$$

in order to ensure $K_n(\delta_n) > 0$ and $K_t(\delta_n) > 0$. Physically, δ_0 can represent the limiting value for the normal displacement δ_n which identifies the admissible compressive domain.

Remark 3.4. *As we will see in the Subsection 3.5.3, from considerations about compression tests, we should also ensure the condition*

$$\beta_t \leq \frac{K_{n,0}}{K_{t,0}}\beta_n,$$

which is equivalent to enforcing

$$\delta_0 = -\frac{1}{2K_{n,0}\beta_n}.$$

Proposition 3.5 (Convexity of ψ)**.** *The surface energy density $\psi(\boldsymbol{\delta}, \boldsymbol{p})$ defined by (3.24) is convex with respect to $(\boldsymbol{\delta}, \boldsymbol{p})$ in the semispace defined by (3.27).*

Proof. The Hessian matrix of $\psi(\boldsymbol{\delta}, \boldsymbol{p})$ is

$$\mathbb{H}_\psi = \begin{pmatrix} \dfrac{\partial^2\psi}{\partial\delta_n^2} & \dfrac{\partial^2\psi}{\partial\delta_n\partial\boldsymbol{\delta_t}} & \dfrac{\partial^2\psi}{\partial\delta_n\partial p_n} & \dfrac{\partial^2\psi}{\partial\delta_n\partial\boldsymbol{p_t}} \\ \dfrac{\partial^2\psi}{\partial\delta_n\partial\boldsymbol{\delta_t}} & K_t(\delta_n)\mathbf{I}_2 & \mathbf{0} & -K_t(\delta_n)\mathbf{I}_2 \\ \dfrac{\partial^2\psi}{\partial\delta_n\partial p_n} & \mathbf{0} & K_n(\delta_n) & \mathbf{0} \\ \dfrac{\partial^2\psi}{\partial\delta_n\partial\boldsymbol{p_t}} & -K_t(\delta_n)\mathbf{I}_2 & \mathbf{0} & K_t(\delta_n)\mathbf{I}_2 \end{pmatrix},$$

where

$$\begin{cases} \dfrac{\partial^2\psi}{\partial{\delta_n}^2} = K_n(\delta_n) + 2\dfrac{\partial K_n(\delta_n)}{\partial\delta_n}(\delta_n - p_n) + \dfrac{\partial^2 K_n(\delta_n)}{\partial\delta_n^2}\dfrac{(\delta_n - p_n)^2}{2} + \dfrac{\partial^2 K_t(\delta_n)}{\partial\delta_n^2}\dfrac{\|\boldsymbol{\delta_t} - \boldsymbol{p_t}\|^2}{2}, \\ \dfrac{\partial^2\psi}{\partial\delta_n\partial\boldsymbol{\delta_t}} = \dfrac{\partial K_t(\delta_n)}{\partial\delta_n}(\boldsymbol{\delta_t} - \boldsymbol{p_t}), \\ \dfrac{\partial^2\psi}{\partial\delta_n\partial p_n} = -K_n - \dfrac{\partial K_n(\delta_n)}{\partial\delta_n}(\delta_n - p_n), \\ \dfrac{\partial^2\psi}{\partial\delta_n\partial\boldsymbol{p}_t} = -\dfrac{\partial^2\psi}{\partial\delta_n\partial\boldsymbol{\delta_t}} = -\dfrac{\partial K_t(\delta_n)}{\partial\delta_n}(\boldsymbol{\delta_t} - \boldsymbol{p_t}), \end{cases}$$

and $\mathbf{I}_2$ is the identity matrix of $\mathbb{R}^{2\times 2}$. Furthermore, the derivatives of $K_s(\delta_n)$, $s \in \{n, t\}$, can be easily obtained recalling their definition (3.26):

$$\frac{\partial K_s(\delta_n)}{\partial \delta_n} = -\frac{2K_{s,0}^2\beta_s}{(2K_{s,0}\beta_s\delta_n + 1)^2} = -2\beta_s \left(K_s(\delta_n)\right)^2, \tag{3.28}$$

$$\frac{\partial^2 K_s(\delta_n)}{\partial \delta_n^2} = \frac{8K_{s,0}^3\beta_s^2}{(2K_{s,0}\beta_s\delta_n + 1)^3} = 8\beta_s^2 \left(K_s(\delta_n)\right)^3. \tag{3.29}$$

By applying an extension of the Sylvester's criterion [100, Chapter 2], $\mathbb{H}_\psi$ is positive semidefinite if and only if all its principal minors are nonnegative. We start by considering the diagonal elements of $\mathbb{H}_\psi$. Notice that, using (3.28)-(3.29), the first term can be rewritten as follows:

$$\frac{\partial^2 \psi}{\partial \delta_n^2} = K_n(\delta_n)\left(2\beta_n K_n(\delta_n)(\delta_n - p_n) - 1\right)^2 + 8\beta_t^2 \left(K_t(\delta_n)\right)^3 \frac{\|\boldsymbol{\delta_t} - \boldsymbol{p_t}\|^2}{2}.$$

As a consequence, all the diagonal terms of $\mathbb{H}_\psi$ are nonnegative, since $K_s(\delta_n) > 0$, $s \in \{n, t\}$, for δ_n satisfying (3.27).

Using again (3.28)-(3.29), one can show that all the other principal minors are nonnegatives in the semispace determined by (3.27). We show two examples. At first, consider the 3rd minor obtained removing the last three columns:

$$\begin{aligned}
\det &\begin{pmatrix} \dfrac{\partial^2 \psi}{\partial \delta_n^2} & \dfrac{\partial K_t(\delta_n)}{\partial \delta_n}(\boldsymbol{\delta_t} - \boldsymbol{p_t}) \\ \dfrac{\partial K_t(\delta_n)}{\partial \delta_n}(\boldsymbol{\delta_t} - \boldsymbol{p_t}) & K_t(\delta_n)\mathbf{I}_2 \end{pmatrix} \\
&= \frac{\partial^2 \psi}{\partial \delta_n^2}\left(K_t(\delta_n)\right)^2 - K_t(\delta_n)\left(\frac{\partial K_t(\delta_n)}{\partial \delta_n}\right)^2 \|\boldsymbol{\delta_t} - \boldsymbol{p_t}\|^2 \\
&= \left(K_n(\delta_n) + 2\frac{\partial K_n(\delta_n)}{\partial \delta_n}(\delta_n - p_n) + \frac{\partial^2 K_n(\delta_n)}{\partial \delta_n^2}\frac{(\delta_n - p_n)^2}{2}\right)\left(K_t(\delta_n)\right)^2 \\
&\quad + K_t(\delta_n)\left(K_t(\delta_n)\frac{\partial^2 K_t(\delta_n)}{\partial \delta_n^2} - 2\left(\frac{\partial K_t(\delta_n)}{\partial \delta_n}\right)^2\right)\frac{\|\boldsymbol{\delta_t} - \boldsymbol{p_t}\|^2}{2} \\
&= K_n(\delta_n)\left(2\beta_n K_n(\delta_n)(\delta_n - p_n) - 1\right)^2 \left(K_t(\delta_n)\right)^2 \geq 0
\end{aligned}$$

for δ_n satisfying (3.27). Notice that we have used (3.28)-(3.29) for obtaining the last line.

Then, we analyze the 4th minor obtained removing the 2nd, 3rd, 5th and 6th column:

$$\begin{aligned}
\det\begin{pmatrix} \frac{\partial^2\psi}{\partial\delta_n^2} & -K_n - \frac{\partial K_n(\delta_n)}{\partial\delta_n}(\delta_n - p_n) \\ -K_n - \frac{\partial K_n(\delta_n)}{\partial\delta_n}(\delta_n - p_n) & K_n(\delta_n) \end{pmatrix} \\
&= \frac{\partial^2\psi}{\partial\delta_n^2}K_n(\delta_n) + \left(K_n - \frac{\partial K_n(\delta_n)}{\partial\delta_n}(\delta_n - p_n)\right)^2 \\
&= \left(K_n(\delta_n) + 2\frac{\partial K_n(\delta_n)}{\partial\delta_n}(\delta_n - p_n) + \frac{\partial^2 K_n(\delta_n)}{\partial\delta_n^2}\,\frac{(\delta_n - p_n)^2}{2}\right)K_n(\delta_n) \\
&\quad + \frac{\partial^2 K_t(\delta_n)}{\partial\delta_n^2}\,\frac{\|\boldsymbol{\delta_t} - \boldsymbol{p_t}\|^2}{2}K_n(\delta_n) \\
&= 8K_n(\delta_n)\beta_t^2\,(K_t(\delta_n))^3\,\frac{\|\boldsymbol{\delta_t} - \boldsymbol{p_t}\|^2}{2} \geq 0.
\end{aligned}$$

□

3.5.2 Stress derivation and plasticity criterion

Adapting the SGM formalism, the stress vector $\boldsymbol{\sigma}$ and the thermodynamical force related to plasticity $\boldsymbol{X}$ are recovered by derivation of the surface energy density function:

$$\boldsymbol{\sigma} = (\sigma_n, \sigma_{t_1}, \sigma_{t_2})^\top := \frac{\partial\psi}{\partial\boldsymbol{\delta}} \qquad \text{and} \qquad \boldsymbol{X} = (X_n, X_{t_1}, X_{t_2})^\top := -\frac{\partial\psi}{\partial\boldsymbol{p}}. \tag{3.30}$$

Explicitly, using (3.24), (3.26) and (3.28), we get

$$\begin{cases}
\sigma_n = \dfrac{K_{n,0}}{2K_{n,0}\beta_n\delta_n + 1}(\delta_n - p_n) - \beta_n\dfrac{K_{n,0}^2}{(2K_{n,0}\beta_n\delta_n + 1)^2}(\delta_n - p_n)^2 \\
\qquad\quad - \beta_t\dfrac{K_{t,0}^2}{(2K_{t,0}\beta_t\delta_n + 1)^2}\|\boldsymbol{\delta_t} - \boldsymbol{p_t}\|^2, & \text{(3.31a)} \\
\boldsymbol{\sigma_t} = \dfrac{K_{t,0}}{2K_{t,0}\beta_t\delta_n + 1}(\boldsymbol{\delta_t} - \boldsymbol{p_t}), & \text{(3.31b)}
\end{cases}$$

and

$$\begin{cases}
X_n = \dfrac{K_{n,0}}{2K_{n,0}\beta_n\delta_n + 1}(\delta_n - p_n), \\
\boldsymbol{X_t} = \dfrac{K_{t,0}}{2K_{t,0}\beta_t\delta_n + 1}(\boldsymbol{\delta_t} - \boldsymbol{p_t}).
\end{cases} \tag{3.32}$$

The connection between $\boldsymbol{\sigma}$ and $\boldsymbol{X}$ is then similar to (3.11):

$$\begin{cases}
\sigma_n = X_n - \beta_n X_n^2 - \beta_t\,\|\boldsymbol{X}_t\|^2, & \text{(3.33a)} \\
\boldsymbol{\sigma}_t = \boldsymbol{X}_t. & \text{(3.33b)}
\end{cases}$$

Also in this model for joints, we consider a linear failure criterion in the generalized force space:

$$f_{\boldsymbol{X}}(\boldsymbol{X}) \coloneqq \|\boldsymbol{X}_t\| + \bar{a}X_n - \bar{b} = 0, \tag{3.34}$$

where $\bar{a} > 0$ and $\bar{b} \geq 0$. If $\bar{b} \neq 0$, we also suppose

$$\beta_n \leq \frac{\bar{a}}{2\bar{b}}. \tag{3.35}$$

As a consequence, the reversibility domain in the space of generalized force $\boldsymbol{X}$ is

$$\mathbb{K}_{\boldsymbol{X}} \coloneqq \left\{\boldsymbol{X}^* \in \mathbb{R}^d \ : \ \|\boldsymbol{X}_t\| + \bar{a}X_n - \bar{b} \leq 0\right\}. \tag{3.36}$$

The plasticity flow fulfills the normality rule to the reversibility domain $\mathbb{K}_{\boldsymbol{X}}$, i.e.,

$$\dot{\boldsymbol{p}} = \dot{\lambda}\frac{\partial f_{\boldsymbol{X}}}{\partial \boldsymbol{X}} \qquad \text{with} \quad \begin{cases} f_{\boldsymbol{X}}(\boldsymbol{X})\,\dot{\lambda} = 0, \\ \dot{\lambda} \geq 0. \end{cases} \tag{3.37}$$

In particular, outside of the singular point we get

$$\begin{cases} \dot{p}_n = \bar{a}\dot{\lambda}, \\ \dot{\boldsymbol{p}}_{\boldsymbol{t}} = \dot{\lambda}\dfrac{\boldsymbol{X}_{\boldsymbol{t}}}{\|\boldsymbol{X}_{\boldsymbol{t}}\|}, \end{cases} \tag{3.38}$$

whereas in the vertex $(\bar{b}/\bar{a}, \boldsymbol{0})$ of $\mathbb{K}_{\boldsymbol{X}}$ the rate of plasticity $\dot{\boldsymbol{p}}$ belongs to the cone of outer normals to $\partial\mathbb{K}_{\boldsymbol{X}}$.

Finally, using the relation (3.33), we achieve the corresponding failure criterion in the stress space:

$$f_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}) \coloneqq (\beta_n + \beta_t\bar{a}^2)\,\|\boldsymbol{\sigma}_t\|^2 + (\bar{a} - 2\beta_n\bar{b})\,\|\boldsymbol{\sigma}_t\| + \bar{a}^2\sigma_n - \bar{b}(\bar{a} - \beta_n\bar{b}) = 0. \tag{3.39}$$

Again, we automatically obtain a quadratic criterion in the stress space. The reversibility domain in the stress space determined by (3.39), i.e., $\mathbb{K}_{\boldsymbol{\sigma}} \coloneqq \{\boldsymbol{\sigma}^* \in \mathbb{R}^d \ : \ f_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}) \leq 0\}$, is a cone with parabolic boundary, axis $\boldsymbol{\sigma_t} = \boldsymbol{0}$ and vertex in $\left(\dfrac{\bar{b}(\bar{a} - \beta_n\bar{b})}{\bar{a}^2}, \boldsymbol{0}\right)$. Notice that this criterion only depends on the parameters β_n, β_t, $\bar{a}$, and $\bar{b}$, i.e., the domain $\mathbb{K}_{\boldsymbol{\sigma}}$ remains fixed during all types of evolution.

Remark 3.6 (Parameters of the linear criterion)**.** *In absence of hyperelasticity, the parameters $\bar{a}$ and $\bar{b}$ may be interpreted in the classical way: $\bar{a}$ being a friction coefficient, and $\bar{b}$ being a residual cohesion, see (2.29) in the previous chapter. On the other hand, the presence of hyperelasticity that we address in this chapter alternates the initial physical meaning of parameters for the same kind of criterion (3.34). We underline the fact that now the the shape of $\mathbb{K}_{\boldsymbol{\sigma}}$ (i.e., quadratic) is different from the shape of $\mathbb{K}_{\boldsymbol{X}}$ (i.e., linear), and, in practice, it is possible to determine experimentally only the first one. As a consequence, it would be better to fix $f_{\boldsymbol{\sigma}}$, and then to recover the corresponding criterion in the space of generalize force $f_{\boldsymbol{X}}$. In Section A.2, we describe how $\bar{a}$, $\bar{b}$, β_n and β_t can be chosen starting from a general parabolic domain in the stress space.*

Remark 3.7 (Motivation of assumption (3.35))**.** *If* (3.35) *holds, then the elastic domain* $\mathbb{K}_{\boldsymbol{\sigma}}$ *is convex and contains the origin* $\mathbf{0}$*, and the criterion* (3.39) *can be rewritten as*

$$\|\boldsymbol{\sigma_t}\| = \frac{-(\bar{a} - 2\beta_n \bar{b}) + \bar{a}\sqrt{1 - 4(\beta_n + \bar{a}^2\beta_t)\sigma_n + 4\beta_t \bar{b}(\bar{a} - \beta_n \bar{b})}}{2(\beta_n + \bar{a}^2\beta_t)}. \tag{3.40}$$

Remark 3.8. *The restriction* (3.27) *for the value of* δ_n *can be removed modifying the expression of* $K_n(\delta_n)$ *and* $K_t(\delta_n)$*. For example, one can define*

$$K_s(\delta_n) = \begin{cases} -2\beta_s K_{s,0}^2 \delta_n + K_{n,0} & \text{if } \delta_n < 0, \\ \dfrac{K_{s,0}}{2K_{s,0}\beta_s \delta_n + 1} & \text{if } \delta_n \geq 0. \end{cases}$$

However, in this case, the plastic criterion in the stress space depends on δ_n*, for* $\delta_n < 0$*, i.e., its shape changes with the evolution of the normal displacement.*

3.5.3 Numerical results on typical tests

As we have done in Section 2.3.4 for the model coupling plasticity and damage, we want to show and analyze the behavior that is possible to obtain with the hyperelastic constitutive relation we have just described on standard tests. We recall that, for sake of simplicity, we suppose $\delta_{t_2} = 0$, or, in other words, we consider a 2-dimensional model.

Compression/Traction test

During a compression or a traction test, the tangential displacement is maintained constant, i.e., $\delta_t = 0$, and we control the evolution of the normal displacement δ_n. While the vector $\boldsymbol{\sigma}$ lies inside of the elastic domain $\mathbb{K}_{\boldsymbol{\sigma}}$, the evolution is hyperelastic. This is the case for compression tests and for the first phase of traction tests if $\bar{b} > 0$. In particular, from (3.31a) we recover explicitly the relation between the normal displacement δ_n and the normal stress σ_n:

$$\sigma_n = \frac{K_{n,0}}{2K_{n,0}\beta_n \delta_n + 1}\delta_n - \beta_n \frac{K_{n,0}^2}{(2K_{n,0}\beta_n \delta_n + 1)^2}\delta_n^2. \tag{3.41}$$

Notice that, for a compression test, the normal displacement δ_n cannot decrease beyond the limiting value $(2K_{n,0}\beta_n)^{-1}$. We recall that this is already guaranteed by the assumption (3.27).

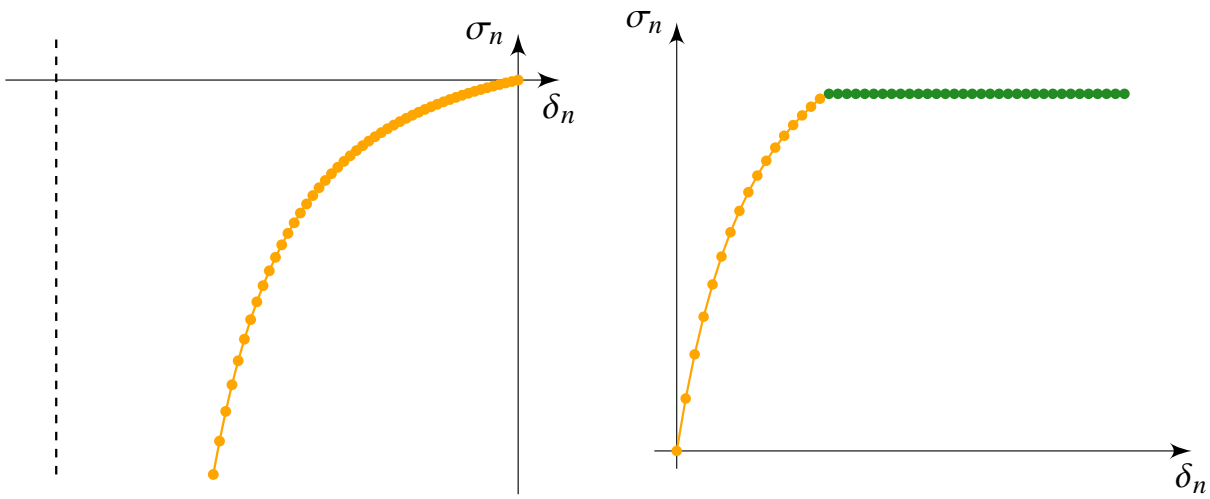
Remark 3.9 (Consideration on the limiting value δ_0 (3.27))**.** *If* $K_{t,0}\beta_t > K_{n,0}\beta_n$*, then the condition* (3.27) *implies that the normal stress has a lower bound value:*

$$\sigma_n > -\frac{K_{n,0}}{2(K_{t,0}\beta_n - K_{n,0}\beta_n)} - \beta_n \left(\frac{K_{n,0}}{2(K_{t,0}\beta_n - K_{n,0}\beta_n)}\right)^2.$$

In order to admit any value of compression σ_n*, we suggest to choose the values of parameters satisfying*

$$K_{t,0}\beta_t \leq K_{n,0}\beta_n.$$

As a consequence $\delta_0 = (2K_{n,0}\beta_n)^{-1}$*, recalling that, by definition,* δ_0 *is the limiting value for the evolution of* δ_n*, see* (3.27)*.*



(a) Compression test

(b) Traction test

Figure 3.6 – Illustration of evolution of the normal stress σ_n during a compression test (*left*) and a traction test (*right*). The phases without and with evolution of plasticity are represented in *orange* and in *green*, respectively.

During a traction test, the value of normal stress increases until it reaches the vertex of $\mathbb{K}_{\boldsymbol{\sigma}}$, i.e., $\sigma_n = \dfrac{\bar{b}(\bar{a} - \beta_n \bar{b})}{\bar{a}}$. Then, since $\mathbb{K}_{\boldsymbol{\sigma}}$ is fixed in the stress space, σ_n remains constant and the plastic component p_n evolves. Figure 3.6 shows the evolution of the normal stress for a compression (*left*) and a traction (*right*) test. In particular, we observe the strong nonlinear behavior while $\boldsymbol{\sigma}$ lies in the interior of the elastic domain $\mathbb{K}_{\boldsymbol{\sigma}}$.

Shear test with fixed compression

At first, we enforce a compression $-\sigma_n^{\text{com}} < 0$, maintaining $\sigma_t = 0$. Notice that, from (3.41), we get the value of normal displacement at the end of compressive stage:

$$\delta_n = \frac{1}{2K_{n,0}\beta_n}\left(-1 + \frac{1}{\sqrt{4\beta_n \sigma_n^{\text{com}} + 1}}\right).$$

Then, we increase the tangential displacement δ_t enforcing $\sigma_n = -\sigma_n^{\text{com}}$. After an initial hyperelastic phase, which ends when the stress vector reaches the boundary of the reversibility domain $\mathbb{K}_{\boldsymbol{\sigma}}$, plasticity evolves following the flow rule (3.38). Figure 3.7 displays the evolution of σ_t and δ_n, with

$$\frac{K_n}{K_t} = \frac{\beta_t}{\beta_n} = \frac{3}{2}.$$

The color *orange* represents the initial nonlinear elastic phase, and the color *green* the phase with evolution of plasticity. We present here the main properties of these evolutions and we compare them with those of a standard associate elasto-plastic model:

- *The evolution of σ_t and δ_n is nonlinear during the elastic phase.* In particular, dilatancy shows up also during this first phase. This differs from a standard elasto-plastic model,

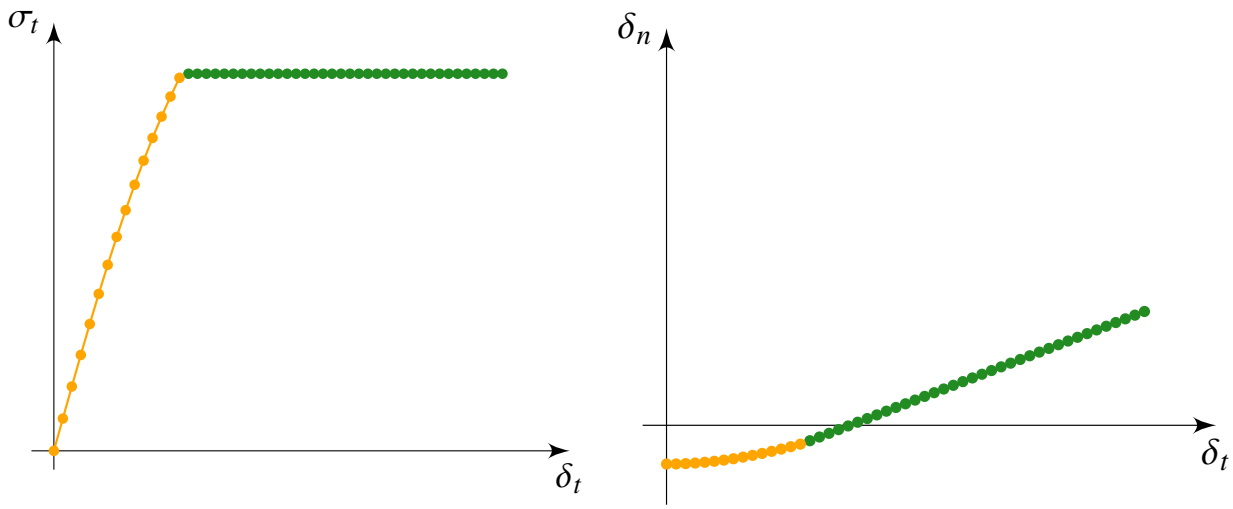


Figure 3.7 – Illustration of evolution of the shear stress σ_t (*left*) and the normal displacement jump δ_n (*right*) during a shear test with fixed compression. The phases without and with evolution of plasticity are represented in *orange* and in *green*, respectively.

in which the normal displacement δ_n remains constant during the elastic phase and dilatancy only appears with the evolution of plasticity. This is true also for the model presented in the previous chapter, see Section 2.3.4 and Figure 2.11.

- *The tangential stress σ_t remains constant while plasticity evolves.* This is an immediate consequence of three facts: 1) the flow rule (3.37) establishes that plasticity can only evolve when $\boldsymbol{X} \in \partial\mathbb{K}_{\boldsymbol{X}}$, i.e., $\boldsymbol{\sigma} \in \partial\mathbb{K}_{\boldsymbol{\sigma}}$; 2) the elastic domain $\mathbb{K}_{\boldsymbol{\sigma}}$ determined by (3.39) is fixed; 3) during the shear test, the normal stress σ_n is constant. The value of σ_t can be computed using (3.40), i.e.,

$$\sigma_t = \sigma_t^{com} := \frac{-(\bar{a} - 2\beta_n\bar{b}) + \bar{a}\sqrt{1 + 4(\beta_n + \bar{a}^2\beta_t)\sigma_n^{com} + 4\beta_t\bar{b}(\bar{a} - \beta_n\bar{b})}}{2(\beta_n + \bar{a}^2\beta_t)}.$$

- *During the plastic phase, δ_n has linear behavior. In particular, the slope of its evolution is related to the normal vector to $\partial\mathbb{K}_{\boldsymbol{\sigma}}$.* It is possible to show (see the following Remark 3.10) that

$$\begin{aligned}\delta_n &= \frac{\bar{a}^2}{2(\beta_n + \bar{a}^2\beta_t)\sigma_t^{\text{com}} + \bar{a} - 2\beta_n\bar{b}}\,\delta_t \\ &\quad - \frac{(\bar{a}^2K_{n,0} + K_{t,0})\sigma_t^{\text{com}} + K_{t,0}\bar{b}}{K_{n,0}K_{t,0}(2(\beta_n + \bar{a}^2\beta_t)\sigma_t^{\text{com}} + \bar{a} - 2\beta_n\bar{b})} \\ &= \frac{\bar{a}^2}{2(\beta_n + \bar{a}^2\beta_t)\sigma_t^{\text{com}} + \bar{a} - 2\beta_n\bar{b}}\,\delta_t + \text{const},\end{aligned} \tag{3.42}$$

and, by derivation of $f_{\boldsymbol{\sigma}}$ (3.39), the vector

$$\left(\frac{\bar{a}^2}{2(\beta_n + \bar{a}^2\beta_t)\sigma_t^{\text{com}} + \bar{a} - 2\beta_n\bar{b}}, 1\right)$$

is normal to $\partial\mathbb{K}_{\boldsymbol{\sigma}}$ in the boundary point $(\sigma_n^{\text{com}}, \sigma_t^{\text{com}})$. As a consequence, the slope of the evolution of δ_n coincides with the ratio between the components of the normal vector to $\partial\mathbb{K}_{\boldsymbol{\sigma}}$. We recall that, in a standard associative elasto-plastic model, the slope of the evolution of δ_n is $\bar{a}$, i.e., it is connected with the normal vector to $\mathbb{K}_{\boldsymbol{X}}$. In particular,

$$\frac{\bar{a}^2}{2(\beta_n + \bar{a}^2\beta_t)\sigma_t^{\text{com}} + \bar{a} - 2\beta_n\bar{b}} < \bar{a}$$

by the assumption (3.35) and by the fact that $\sigma_t^{\text{com}} > \bar{b}$ since we are in compression and $\delta_t \geq 0$.

Remark 3.10 (Derivation of (3.42))**.** *As we have already pointed out, $\sigma_t = \sigma_t^{\text{com}}$ during the plastic phase. Then, from (3.31b), we obtain the expression of the tangential plasticity, which depends on both normal and tangential displacement:*

$$p_t = \delta_t - \frac{2K_{t,0}\beta_t\delta_n + 1}{K_{t,0}}\,\sigma_t^{\text{com}}. \tag{3.43}$$

Notice that, since $\sigma_t^{\text{com}} \geq 0$, and $X_t = \sigma_t$ by (3.33a), the plasticity criterion in the generalized force space (3.34) can be written as

$$\sigma_t^{\text{com}} + \bar{a}\frac{K_{n,0}}{2K_{n,0}\beta_n\delta_n + 1}(\delta_n - p_n) - \bar{b} = 0. \tag{3.44}$$

Furthermore, since we are considering a monotonic shear test, the flow rule (3.38) implies $\boldsymbol{p} = \lambda(\bar{a}, 1)^\top$, i.e., $p_n = \bar{a}p_t$, assuming $\lambda = 0$ at the beginning. By combining this with (3.43) and (3.44), we get

$$\sigma_t^{\text{com}} + \bar{a}\frac{K_{n,0}}{2K_{n,0}\beta_n\delta_n + 1}\left(\delta_n - \bar{a}\delta_t + \bar{a}\frac{2K_{t,0}\beta_t\delta_n + 1}{K_{t,0}}\,\sigma_t^{\text{com}}\right) - \bar{b} = 0,$$

from which we easily recover (3.42).

Figure 3.8 shows the evolution of the stress $\boldsymbol{\sigma}$ and of the generalized forces $\boldsymbol{X}$ inside the corresponding domains $\mathbb{K}_{\boldsymbol{\sigma}}$ and $\mathbb{K}_{\boldsymbol{X}}$. As expected, $\boldsymbol{\sigma}$ moves on a vertical line, while $\boldsymbol{X}$ moves on a nonlinear curve whose explicit expression can be achieved from (3.33a):

$$\beta_n X_n^2 + \beta_t X_t^2 - X_n + \sigma_n^{\text{com}} = 0.$$

Motivated by the results obtained in Subsection 3.4.3, we consider also a cycling shear test with fixed compression, focusing on the evolution of the normal displacement δ_n. Figure 3.9 compares the results obtained under the same loading of δ_t, i.e., $\delta_t \in [-\tilde{\delta}_t, \tilde{\delta}_t]$, with two different constitutive relations: the hyperelasto-plastic model presented in this chapter (*left*) and a standard associative elasto-plastic one (*right*). The scale of both graphs has been changed in order to capture the main phenomenon, i.e., while in the second case the normal displacement continues to increase, the addition of hyperelasticity in the rigidity coefficients allows one to obtain the saturation of dilatancy with cyclic loads.

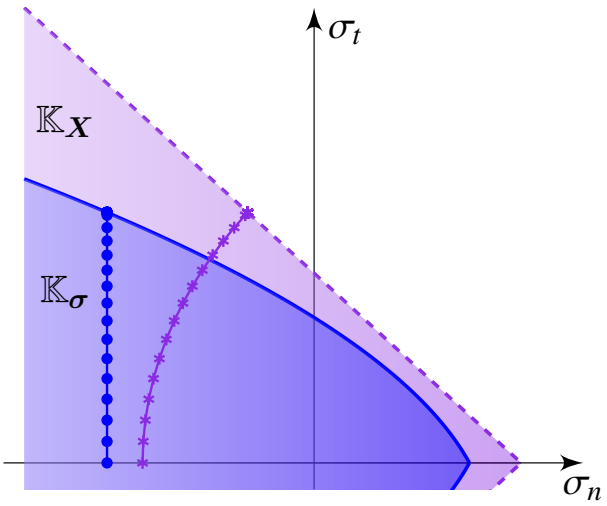
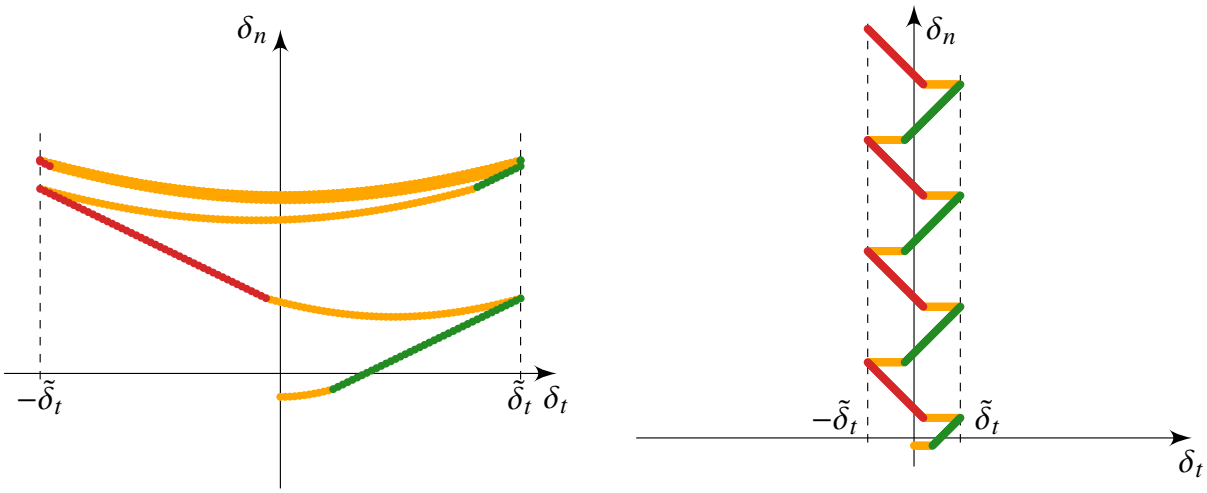


Figure 3.8 – Evolution of the stress $\boldsymbol{\sigma}$ and the generalized force $\boldsymbol{X}$ during a shear test with fixed compression.



(a) $(\beta_n, \beta_t) \neq (0, 0)$ (b) $(\beta_n, \beta_t) = (0, 0)$

Figure 3.9 – Evolution of normal displacement δ_n during a cyclic shear test with fixed compression adopting a hyperelasto-plastic model (*left*) and with a standard associative elasto-plastic model (*right*), under the same loading, i.e., $\delta_t \in [-\tilde{\delta}_t, \tilde{\delta}_t]$, suitably scaled. The elastic phases are represented in *orange*, while the plastic phases are represented in *green* or *red* accordingly to the sign of $\dot{p}_t$.

3.5.4 Implementation with incremental evolution

We present the incremental approach for implementing the hyperelastic constitutive relation. This method is summarized by Algorithm 3.1. Since the hyperelastic terms, i.e., the rigidity coefficients, only involve the normal displacement δ_n (in particular, does not involve the plasticity variable), the procedure is similar to the one described in section 2.3.6. Again, for any quantity z: z^-, z and Δz denote its value at the previous state, at the current state and their difference, respectively. The input values are the state variables at the previous step and the increment of displacement jump $\Delta\boldsymbol{\delta} = (\Delta\delta_n, \Delta\boldsymbol{\delta_t})$, and the output are the current values $\boldsymbol{p}$, $\boldsymbol{\sigma}$, and $\boldsymbol{X}$. Notice that, once we have the current value of plasticity, we can compute the value of the stress and the generalized force directly by the discrete version of (3.31) and (3.32):

$$\begin{cases} \sigma_n = K_n(\delta_n)(\delta_n - p_n^- - \Delta p_n) + \dfrac{\partial K_n(\delta_n)}{\partial \delta_n}\dfrac{(\delta_n - p_n^- - \Delta p_n)^2}{2} \\ \qquad + \dfrac{\partial K_t(\delta_n)}{\partial \delta_n}\dfrac{\|\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\boldsymbol{p_t}\|^2}{2}, & (3.45a) \\ \boldsymbol{\sigma_t} = K_t(\delta_n)(\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\boldsymbol{p_t}), & (3.45b) \\ X_n = K_n(\delta_n)(\delta_n - p_n^- - \Delta p_n), & (3.45c) \\ \boldsymbol{X_t} = K_t(\delta_n)(\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\boldsymbol{p_t}). & (3.45d) \end{cases}$$

In the discretized framework, the plasticity criterion (3.34) and the flow rule (3.37) become

$$\begin{cases} f_{\boldsymbol{X}}(\boldsymbol{X}) = \|\boldsymbol{X_t}\| + \bar{a}X_n - \bar{b} \le 0, & (3.46a) \\ \Delta\boldsymbol{p} = \Delta\lambda \cdot \dfrac{\partial f_{\boldsymbol{X}}}{\partial \boldsymbol{X}}, & (3.46b) \\ f_{\boldsymbol{X}}(\boldsymbol{X})\Delta\lambda = 0,\ \Delta\lambda \ge 0. & (3.46c) \end{cases}$$

During the elastic prediction, we compute

$$\boldsymbol{X}^{\text{pred}} := \mathbb{E}(\delta_n)(\boldsymbol{\delta} - \boldsymbol{p}^-),$$

where the nonlinear elastic matrix $\mathbb{E}(\delta_n)$ is again defined by (3.25). If

$$f_{\boldsymbol{X}}(\boldsymbol{X}^{\text{pred}}) = \|\boldsymbol{X}_{\boldsymbol{t}}^{\text{pred}}\| + \bar{a}X_n^{\text{pred}} - \bar{b} \le 0, \tag{3.47}$$

then the elastic prediction is the solution, and consequently $\Delta\boldsymbol{p} = \boldsymbol{0}$ and

$$\begin{cases} \sigma_n = K_n(\delta_n)(\delta_n - p_n^-) + \dfrac{\partial K_n(\delta_n)}{\partial \delta_n}\dfrac{(\delta_n - p_n^-)^2}{2} + \dfrac{\partial K_t(\delta_n)}{\partial \delta_n}\dfrac{\|\boldsymbol{\delta_t} - \boldsymbol{p_t^-}\|^2}{2}, & (3.48a) \\ \boldsymbol{\sigma_t} = \boldsymbol{X}_{\boldsymbol{t}}^{\text{pred}}, & (3.48b) \\ \boldsymbol{X} = \boldsymbol{X}^{\text{pred}}. & (3.48c) \end{cases}$$

Otherwise, plasticity evolves, i.e., $\Delta\lambda > 0$. Assuming that $\boldsymbol{X_t} \neq \boldsymbol{0}$, the discrete version of the plasticity flow rule (3.38) is

$$\begin{cases} \Delta p_n = \bar{a}\,\Delta\lambda & (3.49a) \\ \Delta\boldsymbol{p_t} = \Delta\lambda\dfrac{\boldsymbol{X_t}}{\|\boldsymbol{X_t}\|}. & (3.49b) \end{cases}$$

Considering (3.49b), using the definition of $\boldsymbol{X_t}$ (3.45d), and multiplying both sides by $\|\boldsymbol{X_t}\|$, we have

$$\Delta\boldsymbol{p_t}\|K_t(\delta_n)(\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\boldsymbol{p_t})\| = \Delta\lambda K_t(\delta_n)(\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\boldsymbol{p_t}),$$

from which we obtain

$$\Delta\boldsymbol{p_t}\Big(\|K_t(\delta_n)(\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\boldsymbol{p_t})\| + \Delta\lambda K_t(\delta_n)\Big) = \Delta\lambda K_t(\delta_n)(\boldsymbol{\delta_t} - \boldsymbol{p_t^-}).$$

Taking the norm of each side and noticing that $\|\Delta\boldsymbol{p_t}\| = \Delta\lambda$

$$\|K_t(\delta_n)(\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\boldsymbol{p_t})\| = \|K_t(\delta_n)(\boldsymbol{\delta_t} - \boldsymbol{p_t^-})\| - K_t(\delta_n)\Delta\lambda.$$

As a consequence, using the definition of X_n (3.45c), the plasticity criterion (3.46a) can be rewritten as

$$\|K_t(\delta_n)(\boldsymbol{\delta_t} - \boldsymbol{p_t^-})\| - K_t(\delta_n)\Delta\lambda + \bar{a}K_n(\delta_n)(\delta_n - p_n^- - \bar{a}\Delta\lambda) - \bar{b} = 0.$$

From the latter relation we easily obtain

$$\Delta\lambda = \frac{\|K_t(\delta_n)(\boldsymbol{\delta_t} - \boldsymbol{p_t^-})\| + \bar{a}K_n(\delta_n)(\delta_n - p_n^-) - \bar{b}}{\bar{a}^2K_n(\delta_n) + K_t(\delta_n)}. \tag{3.50}$$

Now, we compute the normal component of generalized force X_n inserting $\Delta p_n = \bar{a}\Delta\lambda$ in (3.45c). If $X_n \leq \bar{b}/\bar{a}$, then we have found the right value of the increment $\Delta\lambda$ and the value of $\Delta\boldsymbol{p_t}$ is computed with the following trick: denoting by $\uparrow\uparrow$ the collinearity relation between vectors and observing that

$$\Delta\boldsymbol{p_t} \uparrow\uparrow \boldsymbol{X_t} = K_t(\delta_n)(\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\boldsymbol{p_t}) \uparrow\uparrow (\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\boldsymbol{p_t}),$$

since $K_t(\delta_n) > 0$, we have

$$\Delta\boldsymbol{p_t} \uparrow\uparrow (\boldsymbol{\delta_t} - \boldsymbol{p_t^-}),$$

and, as a consequence, (3.49b) becomes

$$\Delta\boldsymbol{p_t} = \Delta\lambda\,\frac{\boldsymbol{\delta_t} - \boldsymbol{p_t^-}}{\|\boldsymbol{\delta_t} - \boldsymbol{p_t^-}\|}. \tag{3.51}$$

Finally, the stress $\boldsymbol{\sigma}$ and the tangential components $\boldsymbol{X_t}$ are computed from (3.45). If $X_n > \bar{b}/\bar{a}$, it means that we have made the wrong assumption $\boldsymbol{X_t} \neq \boldsymbol{0}$, i.e., $\boldsymbol{X} = (\bar{b}/\bar{a}, \boldsymbol{0})$. The system to solve is then

$$\begin{cases} \Delta\lambda > 0 & \text{(3.52a)} \\ \Delta p_n = \mu\Delta\lambda \quad \text{and} \quad \|\Delta\boldsymbol{p_t}\| \leq \Delta\lambda & \text{(3.52b)} \\ X_n = \dfrac{\bar{b}}{\bar{a}} \quad \text{and} \quad \boldsymbol{X_t} = \boldsymbol{0} & \text{(3.52c)} \end{cases}$$

From the definition of $\boldsymbol{X}$ (3.45c)-(3.45d) and (3.52c), one can easy obtain

$$\begin{cases} \Delta p_n = \delta_n - p_n^- - \dfrac{\bar{b}}{\bar{a}K_n(\delta_n)}, \\ \Delta\boldsymbol{p_t} = \boldsymbol{\delta_t} - \boldsymbol{p_t^-}. \end{cases} \tag{3.53}$$

Inserting the latter expressions into (3.45a)-(3.45b), we get

$$\boldsymbol{\sigma} = \left(\frac{\bar{b}}{\bar{a}^2}(\bar{a} - \beta_n\bar{b}), \boldsymbol{0}\right). \tag{3.54}$$

Algorithm 3.1 Implementation constitutive relation

Input: $\boldsymbol{\delta}, \boldsymbol{p}^-$
compute $\boldsymbol{X}^{\text{pred}} = \mathbb{E}(\delta_n)(\boldsymbol{\delta} - \boldsymbol{p}^-)$ ($\longrightarrow$ elastic prediction)
if $\|\boldsymbol{X}_t^{\text{pred}}\| + \bar{a} X_n^{\text{pred}} - \bar{b} \leq 0$ **then**
    The elastic prediction is the solution
    set $\boldsymbol{X} = \boldsymbol{X}^{\text{pred}}$
    compute $\boldsymbol{\sigma}$ from (3.48a)-(3.48b)
else ($\longrightarrow$ correction)
    compute $\Delta\lambda$ from (3.50), Δp_n from (3.49a), and X_n from (3.45c)
    if $X_n \leq \bar{b}/\bar{a}$ **then**
        The solution has been found
        compute $\Delta \boldsymbol{p}_t$ from (3.51) and $\boldsymbol{X}_t$ from (3.45d)
    else ($\longrightarrow$ singularity)
        set $\boldsymbol{X} = (\bar{b}/\bar{a}, \boldsymbol{0})$
        compute $\Delta \boldsymbol{p}$ from (3.53)
    end if
    compute $\boldsymbol{\sigma}$ (3.45a)-(3.45b)
end if

3.5.5 Numerical results on a dam

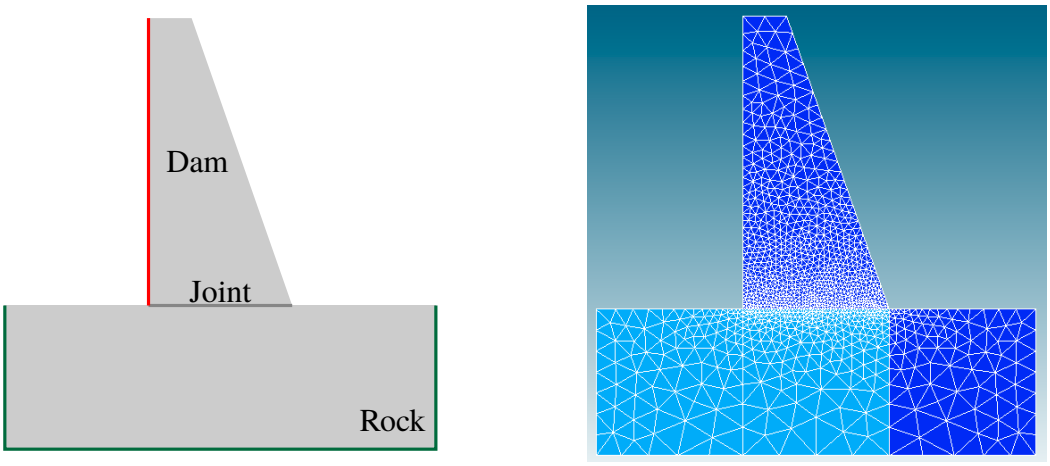
In this section we display some numerical results performed with the software `code_aster` (https://code-aster.org). The aim is to study the influence of initial normal and tangential coefficient on the compressive and opening profile of the interface, and to compare the hyperelasto-plastic model we propose with a standard associative elasto-plastic model.

We consider the 2D dam model shown by Figure 3.10a. In particular, the dam is represented by a trapezoid whose height is 10 m, and basis lengths are 5 m and 1.5 m, while the rock foundation is represented by a rectangle whose length is 15 m and height is 5 m. We assume that the green boundary part is blocked, and we enforce some weight in the dam and in the rock, and some horizontal pressure $g_N(y)$ (due to the presence of water) on the red part of the boundary. For simplicity, we consider an elastic constitutive relation for the dam and the rock, i.e.,

$$\boldsymbol{\sigma} = \lambda \mathrm{Tr}\,\boldsymbol{\varepsilon}\, \mathbf{I}_d + 2\mu\boldsymbol{\varepsilon} = \frac{\nu E}{(1+\nu)(1-2\nu)} \mathrm{Tr}\,\boldsymbol{\varepsilon}\, \mathbf{I}_d + \frac{E}{1+\nu}\boldsymbol{\varepsilon}.$$

The domain for the dam and the rock is discretized with the triangular mesh displayed by Figure 3.10b. Between the dam and the rock there is a joint which is discretized by some joint elements with zero initial width, see Figure 2.4. We compare two constitutive relations for this joint: a standard associated (i.e., with normal flow rule) elasto-plastic relation, and the hyperelasto-plastic relation introduced in this section. Notice that the former relation can be recover directly from the latter by choosing $(\beta_n, \beta_t) = (0, 0)$. The implementation of the joint relations follows the incremental evolution described in the previous subsection and summarized by Algorithm 3.1. Furthermore, the tangent matrix is implemented as detailed in Annex A.1.2.

We consider three different states:



(a) Model

(b) Mesh

Figure 3.10 – Illustration of the dam.

Parameter name	Notation	Value
Physical parameters for concrete, rock and water		
Concrete and rock Young modulus	E	40 GPa
Concrete and rock Poisson coefficient	ν	0.2
Concrete density	ρ_b	$2.3 \cdot 10^3$ kg/m^3
Water density	ρ_e	10^3 kg/m^3
Rock density	ρ_r	0 kg/m^3
Parameters for the elasto-plastic joint model		
Tensile strength	R_t	10^4 Pa
Friction coefficient	μ	1
Normal rigidity coefficient	K_n	10^{12} Pa/m
Tangential rigidity coefficient	K_t	10^{12} Pa/m
Parameters for hyperelasto-plastic joint model		
Normal hyperelasticity coefficient	β_n	$2.34 \cdot 10^{-5}$ Pa^{-1}
Tangential hyperelasticity coefficient	β_t	$9.375 \cdot 10^{-8}$ Pa^{-1}
Linear coefficient for $\mathbb{K}_X$	$\bar{a}$	≈ 2.53
Constant coefficient for $\mathbb{K}_X$	$\bar{b}$	$\approx 4.038 \cdot 10^4$ Pa

Table 3.1 – Parameters notation and value

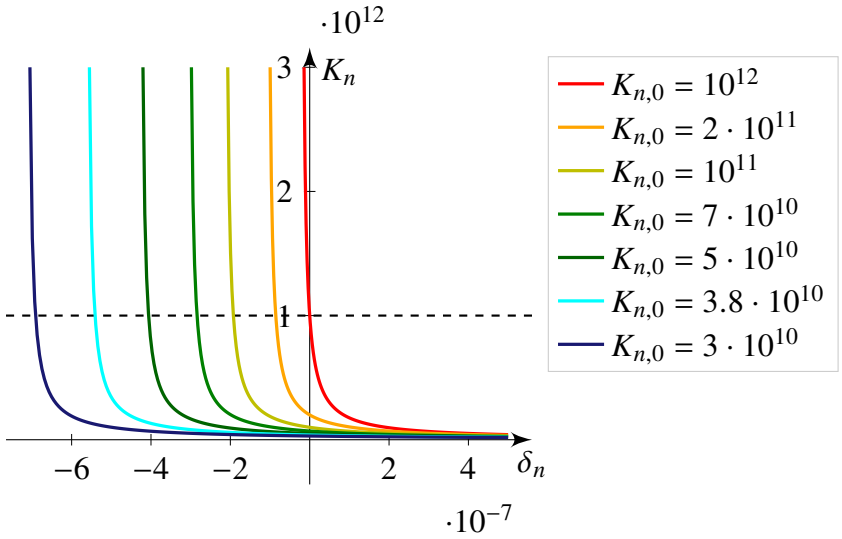


Figure 3.11 – Profile of the normal rigidity coefficient $K_n(\delta_n)$ (3.26) varying the value of the parameter $K_{n,0}$.

(1) configuration with absence of water, i.e., $g_N(y) = 0$ for $y \in [0, 10]$;

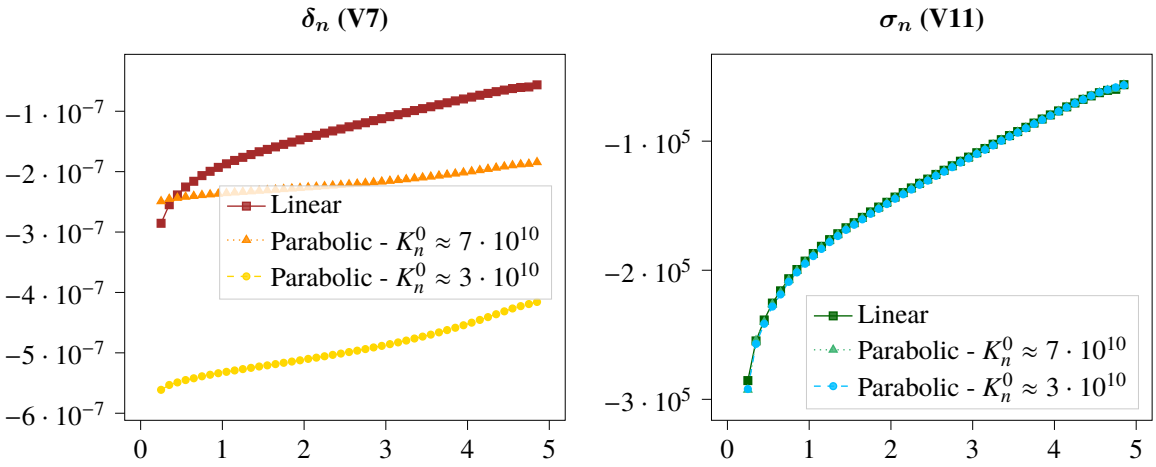
(2) configuration with $h = 9\,\text{m}$ of water and $g_N(y) = \max\{0, \rho_e g(h - y)\}$ for $y \in [0, 10]$, where ρ_e is the density of the water and g is the gravitational acceleration;

(3) configuration with $h = 9\,\text{m}$ of water, $g_N(y) = \max\{0, \rho_e g(h - y)\}$ for $y \in [0, 10]$, and enforced pressure inside the joint $p(x) = \dfrac{5 - x}{5}\rho_e g h$ for $x \in [0, 5]$.

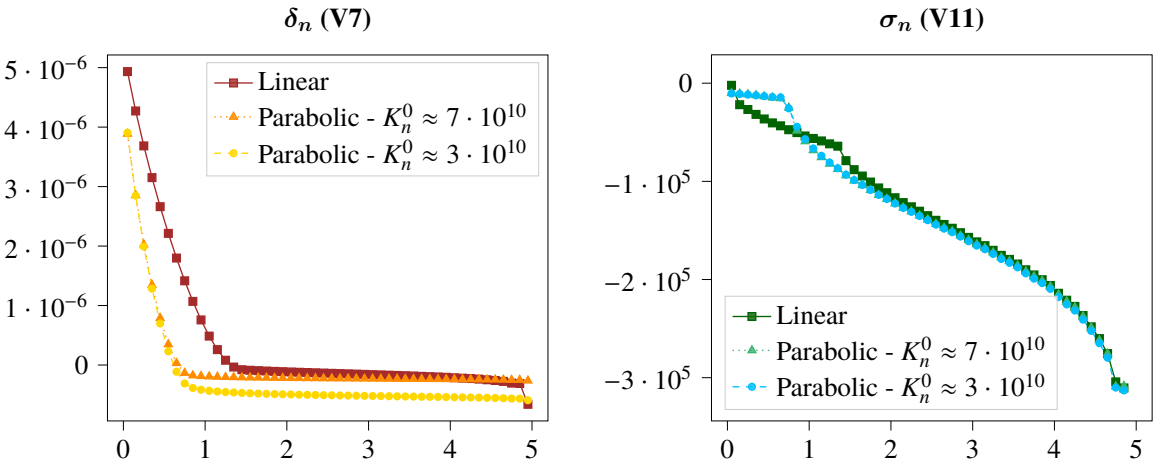
In particular, in the last configuration, we enforce the pressure p inside the joint using the Terzaghi principle [120] adapted to joint models, i.e., the normal stress is modified as follows:

$$\sigma_n^{\text{eff}}(\boldsymbol{\delta}, p) = \sigma_n^{\text{meca}}(\boldsymbol{\delta}) - p.$$

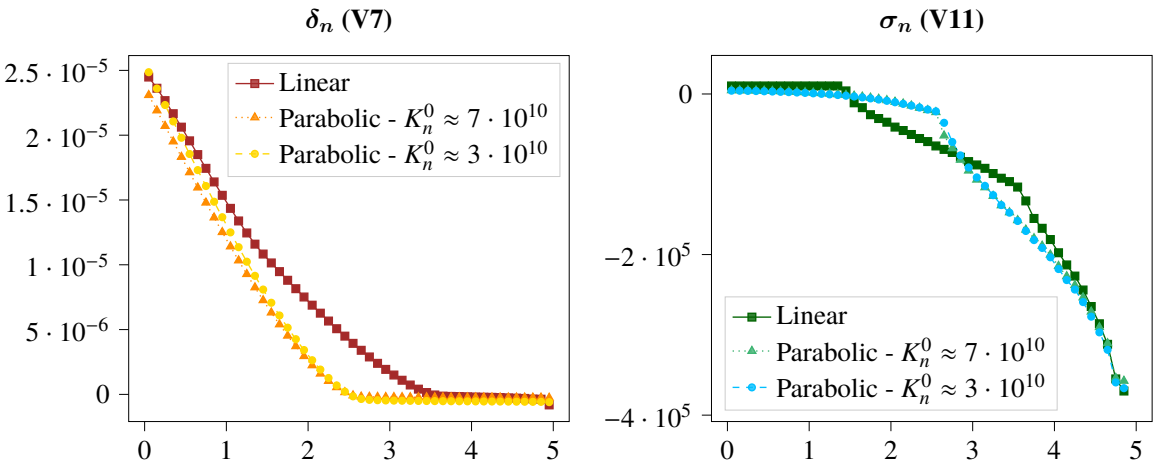
We compute the normal displacement jump δ_n and the normal stress σ_n on the interface, and, for the hyperelasto-plastic constitutive relation, two values for the initial rigidity coefficients are considered: $K_{n,0} = K_{t,0} \approx 7 \cdot 10^{10}$ and $K_{n,0} = K_{t,0} \approx 3 \cdot 10^{10}$. Figure 3.11 shows the influence of the initial value $K_{n,0}$ on the profile of the function $K_n(\delta_n)$. The other parameters and their values are summarized by Table 3.1. The results are illustrated by Figure 3.12. The influence of $K_{n,0}$ (and $K_{t,0}$) in the hyperelastic model is more evident in the profile of δ_n (plots on the *left*). In particular, we can observe a great difference of values in compression (the smaller the rigidity, the more important is the level of interpenetration of the roughness, Figure (3.12a)), while the opening behavior seems to be similar. Furthermore, in compression the values of δ_n with the standard model are for the most part bigger than those obtained with hyperelastic model. This holds also for the opening profile, i.e., the opening of the joint is smaller in the hyperelastic case. Regarding σ_n (plots on the *right* of Figure 3.12), the profile is almost identical when $\delta_n < 0$, and it slightly differs when $\delta_n > 0$.



(a) Configuration (1), i.e., case without lateral pressure.



(b) Configuration (2), i.e., case with lateral pressure.



(c) Configuration (3), i.e., case with lateral pressure and enforced pressure inside the joint.

Figure 3.12 – Comparison of vertical displacement δ_n (*left*) and normal stress σ_n (*right*) obtained with a standard elasto-plastic model and the hyperelasto-plastic model with two different values of $K_{n,0}$ (and $K_{t,0}$).

3.6 Conclusion

In this chapter, we have studied the influence of hyperelasticity on the classical perfect plasticity constitutive behavior in both contexts of geomaterials and joints modeling. It has been shown that, for a particular class of hyperelasticity, the latter phenomenon creates a link between a linear criterion in the generalized force space and a quadratic one in the stress space. On one hand, as a first consequence the Hoek–Brown criterion, empirically found long ago, appears to be the natural choice for geomaterials that exhibit nonlinear elastic behavior. On the other hand, the simple fact of observing fixed in stress-space yield surface (independent of plastification level) can point out the hyperbolic elasticity relation, which is easier to fit to experimental data [79].

Furthermore, we have numerically implemented an example of this hyperelasto-plastic model for geomaterials written in the formalism of Standard Generalized Materials. The model has been constructed in the spirit of [53] from an energy function with hyperbolic elastic dependencies in the compressibility and shear moduli. The mechanical properties of the presented model have been investigated in detail. Most notably, the obtained model reveals accommodation of dilatancy during uniaxial compression tests with a confining pressure. Being common for many geomaterials and experimentally observed, this saturation of dilatancy constitutes another indirect proof of the importance of hyperelasticity.

Finally, the hyperelasto-plastic model for joints, constructed from a surface energy density function introducing a nonlinear dependence on the rigidity coefficients, has been implemented. The results of compression and traction tests show the strong nonlinearity due to hyperelasticity, while the results of a shear test with fixed compression show that the slope of dilatancy is related to the normal to the elastic domain in the stress space and, as a consequence, its value is smaller than the slope of the corresponding model without hyperelasticity. In addition, the saturation of dilatancy can be recovered also in this context during cyclic tests.

For sure, the simplified example models defined here can hardly represent all aspects of the rather complex behavior of real geomaterials or joints. Nevertheless, it can be considered as a limiting or initial constitutive relation serving as a fundamental brick for more subtle models. In particular, this type of hyperelasticity can be added to the constitutive relations coupling plasticity and damage proposed by [94] and Chapter 2 of this manuscript to construct a richer model. We will start to address this idea in the context of joint modeling in Section 5.3.

4 A posteriori error estimate for unilateral contact problem

This chapter is based on a paper entitled "A posteriori error estimates via equilibrated stress reconstructions for contact problems approximated by Nitsche's method" [43] and published in the journal *Computers & Mathematics with Applications*. The preprint is available in ArXiv: https://arxiv.org/abs/2109.11944.

We present an a posteriori error estimate based on equilibrated stress reconstructions for the finite element approximation of a unilateral contact problem with weak enforcement of the contact conditions. We start by proving a guaranteed upper bound for the dual norm of the residual. This norm is shown to control the natural energy norm up to a boundary term, which can be removed under a saturation assumption. The basic estimate is then refined to distinguish the different components of the error, and is used as a starting point to design an algorithm including adaptive stopping criteria for the nonlinear solver and automatic tuning of a regularization parameter. We then discuss an actual way of computing the stress reconstruction based on the Arnold–Falk–Winther finite elements. Finally, we showcase the performance of our estimators on a panel of numerical tests, and we end this chapter by discussing their efficiency.

Contents

4.1 Introduction

In various fields of solid mechanics and engineering it is essential to describe contact and friction between two bodies. This is the case, e.g., when modeling foundations and joints in buildings or when considering impact problems. In this paper, we focus on a simplified unilateral contact problem without friction, for which contact is mathematically expressed by some inequalities and complementarity conditions. These constraints translate non-penetration as well as the absence of cohesive forces and friction between the two bodies. In order to deal with the above constraints from a numerical point of view, different strategies have been proposed in the literature, including penalized formulations, mixed formulations, and weak enforcement à la Nitsche. We focus on the latter, which does not require the introduction of Lagrange multipliers and results in a coercive formulation that is easy to implement. The literature on the numerical approximation of contact problems is vast, and a comprehensive state-of-the-art lies outside of the scope of the present papers. We refer to the review articles [127, 26] and references therein for a broader discussion.

Nitsche's method was originally introduced in [104] to weakly enforce Dirichlet boundary conditions. Its application to the unilateral contact problem considered in this paper was originally proposed in [28], where the well-posedness and convergence of a conforming finite element scheme are studied. Further extensions to problems involving friction or multiple bodies can be found in [29, 24, 112, 65]; we also mention [25] concerning the adaptation of these ideas to Hybrid High-Order discretizations [42, 41, 39]. A residual-based a posteriori error analysis can be found in [27] based on a saturation assumption.

In this paper we follow a different path and carry out an a posteriori error analysis based on equilibrated tractions in the spirit of the Prager–Synge equality [108] (see also [57] and [126, Chapter 7], [35, 20]). This approach, which does not require the saturation assumption when the dual norm of the residual is considered as an error measure, has also the advantage of avoiding unknown constants in the upper bound. The corresponding error estimate can additionally be refined in order to distinguish the various components of the error (discretization, linearization, regularization). This decomposition is leveraged here to design a fully adaptive resolution algorithm including an a posteriori-based stopping criterion for the nonlinear solver and the automatic tuning of the regularization parameter.

A crucial ingredient of our a posteriori analysis is a novel $\mathbb{H}(\mathbf{div})$-conforming stress reconstruction obtained from the numerical solution by solving small problems on patches around mesh vertices. In the spirit of [116, 16], we use the Arnold–Falk–Winther mixed finite element spaces with weakly enforced symmetry [6]; strong symmetry could be enforced using the Arnold–Winther finite element spaces [7] as in [116], but would come at a significantly

higher computational cost. The stress reconstruction is built so that its divergence and its normal component on the contact and Neumann portions of the domain boundary are locally in equilibrium with the volume and surface source terms, respectively (such equilibrium properties are not satisfied by stress fields resulting from $\boldsymbol{H}^1$-conforming finite element approximations). In order to distinguish the various error components, the stress reconstruction is additionally split so as to identify the contributions to the error resulting from linearization and regularization.

The main results of the paper are briefly summarized in what follows. The basic error estimate of Theorem 4.5 establishes a guaranteed upper bound for the dual norm of the residual. Such norm is shown in Theorem 4.8 to control the energy norm of the error up to a boundary term on the contact region (this term can be eliminated when a saturation assumption similar to the one used in [27] holds). A refined error estimate distinguishing the error components is derived in Theorem 4.13.

The rest of the paper is organized as follows. In Section 4.2 we describe the unilateral contact problem and its finite element approximation à la Nitsche. In Section 4.3 we derive a basic estimate for the dual norm of the residual and compare this norm with the energy norm. Section 4.4 contains a refined version of the estimate distinguishing the error components which serves as a starting point for the development of a fully adaptive algorithm. An equilibrated stress reconstruction based on the Arnold–Falk–Winther finite element is then proposed in Section 4.5. Section 4.7 briefly addresses the efficiency of the error estimators. Finally, some numerical results performed with the open source software FreeFem++ are presented in Section 4.6.

4.2 Setting

In this section we discuss the contact problem and its finite element discretization with weakly enforced contact conditions.

4.2.1 Unilateral contact problem without friction

Strong formulation

Let $d \in \{2, 3\}$ and let $\Omega \subset \mathbb{R}^d$ be a connected open subset of $\mathbb{R}^d$ representing an elastic body. We suppose that Ω is a polygon (if $d = 2$) or a polyhedron (if $d = 3$), and that its boundary $\partial\Omega$ is partitioned into three non-overlapping parts Γ_D, Γ_N, and Γ_C such that $|\Gamma_D| > 0$ and $|\Gamma_C| > 0$ ($|\cdot|$ denotes here the Hausdorff measure). In its reference configuration, the elastic body is in contact through Γ_C with a rigid foundation, and we assume that the (unknown) contact zone in the deformed configuration will be included in Γ_C. The body is clamped at Γ_D and it is subjected to a volume force $\boldsymbol{f} \in \boldsymbol{L}^2(\Omega)$ and to a surface load $\boldsymbol{g}_N \in \boldsymbol{L}^2(\Gamma_N)$ on Γ_N; see Figure 4.4 for an example.

Denoting by $\boldsymbol{n}$ the unit normal vector on $\partial\Omega$ pointing out of Ω, for any displacement field $\boldsymbol{v}\colon \Omega \to \mathbb{R}^d$ and for any density of surface forces $\boldsymbol{\sigma}(\boldsymbol{v})\boldsymbol{n}$ defined on $\partial\Omega$, we have the following (unique) decomposition into normal and tangential components:

$$\boldsymbol{v} = v^n\boldsymbol{n} + \boldsymbol{v}^t \qquad \text{and} \qquad \boldsymbol{\sigma}(\boldsymbol{v})\boldsymbol{n} = \sigma^n(\boldsymbol{v})\boldsymbol{n} + \boldsymbol{\sigma}^t(\boldsymbol{v}). \tag{4.1}$$

Remark 4.1 (Notation for normal and tangential components)**.** *In this chapter, we maintain the notation of the submitted paper [43] for the normal and tangential components on the boundary $\partial\Omega$, i.e., we use the superscripts v^n and $\boldsymbol{v}^t$, instead of the subscripts v_n and $\boldsymbol{v}_t$.*

We consider the following problem: Find the displacement field $\boldsymbol{u}\colon \Omega \to \mathbb{R}^d$ such that

$$\begin{aligned}
\mathbf{div}\,\boldsymbol{\sigma}(\boldsymbol{u}) + \boldsymbol{f} &= \mathbf{0} && \text{in } \Omega, && (4.2a)\\
\boldsymbol{\sigma}(\boldsymbol{u}) &= \mathbb{E}\,\boldsymbol{\varepsilon}(\boldsymbol{u}) && \text{in } \Omega, && (4.2b)\\
\boldsymbol{u} &= \mathbf{0} && \text{on } \Gamma_D, && (4.2c)\\
\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n} &= \boldsymbol{g}_N && \text{on } \Gamma_N, && (4.2d)\\
u^n \le 0,\ \sigma^n(\boldsymbol{u}) \le 0,\ \sigma^n(\boldsymbol{u})u^n &= 0 && \text{on } \Gamma_C, && (4.2e)\\
\boldsymbol{\sigma}^{\boldsymbol{t}}(\boldsymbol{u}) &= \mathbf{0} && \text{on } \Gamma_C, && (4.2f)
\end{aligned}$$

where $\boldsymbol{\varepsilon}(\boldsymbol{v}) \coloneqq \frac{1}{2}(\boldsymbol{\nabla}\boldsymbol{v} + \boldsymbol{\nabla}\boldsymbol{v}^\top)$ is the strain tensor field, $\boldsymbol{\sigma}(\boldsymbol{v}) \in \mathbb{R}^{d\times d}_{\mathrm{sym}}$ is the Cauchy stress tensor, **div** is the divergence operator acting row-wise on tensor valued functions, and $\mathbb{E}$ is the fourth order symmetric elasticity tensor such that, for all second order tensor $\boldsymbol{\tau}$, $\mathbb{E}\boldsymbol{\tau} = \lambda \mathrm{Tr}\,(\boldsymbol{\tau})\boldsymbol{I}_d + 2\mu\boldsymbol{\tau}$, with λ and μ denoting the Lamé parameters. Here, for sake of simplicity, we assume that the behavior of the body is homogeneous isotropic linear elastic. Indeed, the focus of this work is on the presence of contact boundary conditions. Nevertheless, an extension to possibly nonlinear cases seems possible in the spirit of [16], and will make the object of future works.

Remark 4.2 (Contact conditions)**.** *The first contact condition* (4.2e) *is a complementarity condition: if, at a point $\boldsymbol{x} \in \Gamma_C$, there is no contact in the deformed configuration (i.e., $u^n < 0$), then the normal stress vanishes at that point (i.e., $\sigma^n(\boldsymbol{u}) = 0$); on the other hand, if at $\boldsymbol{x} \in \Gamma_C$ the normal stress is nonzero (i.e., $\sigma^n(\boldsymbol{u}) < 0$), then in the deformed configuration we still have contact in $\boldsymbol{x}$ (i.e., $u^n = 0$). These conditions also account for the absence of normal cohesive forces between the elastic body and the rigid foundation. The second contact condition* (4.2f) *simply establishes the absence of friction on Γ_C.*

The incorporation of standard friction models to the following a posteriori theory seems possible, but lies outside of the scope of the present paper. This subject will be addressed in future works.

Weak formulation

Let D denote a measurable set of $\mathbb{R}^d$. In what follows, D will be typically either equal to Ω or to the union of a finite subset of mesh elements. We denote by $H^s(D)$, $s \in \mathbb{R}$, the usual Sobolev space of index s on D, with the convention that $H^0(D) \coloneqq L^2(D)$, the space of square-integrable functions on D. Its vector and tensor versions are denoted respectively by $\boldsymbol{H}^s(D) \coloneqq [H^s(D)]^d$ and $\mathbb{H}^s(D) \coloneqq [H^s(D)]^{d\times d}$. We let, similarly, $\boldsymbol{L}^2(D) \coloneqq [L^2(D)]^d$ and $\mathbb{L}^2(D) \coloneqq [L^2(D)]^{d\times d}$. Moreover, $\|\cdot\|_{s,D}$ denotes the norm of $H^s(D)$ or $\boldsymbol{H}^s(D)$ according to its argument. The first subscript is omitted when $s = 0$, i.e., $\|\cdot\|_D$ is the standard norm of $L^2(D)$, $\boldsymbol{L}^2(D)$, or $\mathbb{L}^2(D)$ according to its argument. The usual inner products of these spaces are denoted by $(\cdot,\cdot)_D$, with the convention that the subscript is omitted when $D = \Omega$.

In what follows, we will also need the space $\mathbb{H}(\mathbf{div}, D)$ spanned by functions of $\mathbb{L}^2(D)$ with weak (row-wise) divergence in $\boldsymbol{L}^2(D)$.

Denote by $\boldsymbol{H}^1_D(\Omega)$ the subspace of $\boldsymbol{H}^1(\Omega)$ incorporating the Dirichlet boundary condition on Γ_D, and by $\boldsymbol{K}$ its subset spanned by admissible displacements, i.e.,

$$\boldsymbol{H}^1_D(\Omega) \coloneqq \left\{\boldsymbol{v} \in \boldsymbol{H}^1(\Omega) \ : \ \boldsymbol{v} = \boldsymbol{0} \text{ on } \Gamma_D\right\}, \qquad \boldsymbol{K} \coloneqq \left\{\boldsymbol{v} \in \boldsymbol{H}^1_D(\Omega) \ : \ v^n \le 0 \text{ on } \Gamma_C\right\}.$$

The weak formulation of problem (4.2) corresponds to the following variational inequality (see, e.g., [67]): Find $\boldsymbol{u} \in \boldsymbol{K}$ such that

$$a(\boldsymbol{u}, \boldsymbol{v} - \boldsymbol{u}) \ge L(\boldsymbol{v} - \boldsymbol{u}) \qquad \forall \boldsymbol{v} \in \boldsymbol{K}, \tag{4.3}$$

where the bilinear form $a\colon \boldsymbol{H}^1(\Omega) \times \boldsymbol{H}^1(\Omega) \to \mathbb{R}$ and the linear form $L\colon \boldsymbol{H}^1(\Omega) \to \mathbb{R}$ are defined as follows: For all $(\boldsymbol{u}, \boldsymbol{v}) \in \boldsymbol{H}^1(\Omega) \times \boldsymbol{H}^1(\Omega)$,

$$a(\boldsymbol{u}, \boldsymbol{v}) \coloneqq (\boldsymbol{\sigma}(\boldsymbol{u}), \boldsymbol{\varepsilon}(\boldsymbol{v})), \qquad L(\boldsymbol{v}) \coloneqq (\boldsymbol{f}, \boldsymbol{v}) + (\boldsymbol{g}_N, \boldsymbol{v})_{\Gamma_N}. \tag{4.4}$$

Problem (4.3) admits a unique solution by the Stampacchia theorem.

4.2.2 Discretization

Let $\{\mathcal{T}_h\}_h$ be a family of conforming triangulations of Ω, indexed by the mesh size $h \coloneqq \max_{T\in\mathcal{T}_h} h_T$, where h_T is the diameter of the element T. This family is assumed to be regular in the classical sense; see, e.g., [30, Eq. (3.1.43)]. Furthermore, each triangulation is conformal to the subdivision of the boundary into Γ_D, Γ_N, and Γ_C in the sense that the interior of a boundary edge (if $d = 2$) or face (if $d = 3$) cannot have non-empty intersection with more than one part of the subdivision. Mesh-related notations that will be used in the a posteriori error analysis are collected in Table 4.1. For the sake of simplicity, from this point on we adopt the three-dimensional terminology and speak of faces instead of edges also in dimension $d = 2$.

For any $X \in \mathcal{T}_h \cup \mathcal{F}_h$ mesh element or face, $\mathcal{P}^n(X)$ denotes the restriction to X of d-variate polynomials of total degree $\le n$, and we set $\boldsymbol{\mathcal{P}}^n(X) \coloneqq [\mathcal{P}^n(X)]^d$ and $\mathbb{P}^n(X) \coloneqq [\mathcal{P}^n(X)]^{d\times d}$. We seek the displacement in the standard Lagrange finite element space of degree $p \ge 1$ with strongly enforced boundary condition on Γ_D:

$$\boldsymbol{V}_h \coloneqq \left\{\boldsymbol{v}_h \in \boldsymbol{H}^1_D(\Omega) \ : \ \boldsymbol{v}_h|_T \in \boldsymbol{\mathcal{P}}^p(T) \text{ for any } T \in \mathcal{T}_h\right\}.$$

Denote by $[\,\cdot\,]_{\mathbb{R}^-}\colon \mathbb{R} \to \mathbb{R}^-$ the projection on the half-line of negative real numbers $\mathbb{R}^-$, i.e., $[x]_{\mathbb{R}^-} \coloneqq \frac{1}{2}(x - |x|)$ for all $x \in \mathbb{R}$. For every real number θ and every positive bounded function $\gamma\colon \Gamma_C \to \mathbb{R}^+$, we define the following linear operator [26]:

$$\begin{aligned} P^n_{\theta,\gamma}\colon \boldsymbol{W} &\to L^2(\Gamma_C) \\ \boldsymbol{v} &\mapsto \theta\sigma^n(\boldsymbol{v}) - \gamma v^n, \end{aligned} \tag{4.5}$$

where $\boldsymbol{W} \coloneqq \left\{\boldsymbol{v} \in \boldsymbol{H}^1(\Omega) \ : \ \boldsymbol{\sigma}(\boldsymbol{v})\boldsymbol{n}|_{\Gamma_C} \in \boldsymbol{L}^2(\Gamma_C)\right\}$ (notice that $\boldsymbol{V}_h \subset \boldsymbol{W}$). Assuming that $\boldsymbol{u} \in \boldsymbol{W}$, the first contact condition (4.2e) can be written as (see [33, 28]):

$$\sigma^n(\boldsymbol{u}) = [\sigma^n(\boldsymbol{u}) - \gamma u^n]_{\mathbb{R}^-} = \left[P^n_{1,\gamma}(\boldsymbol{u})\right]_{\mathbb{R}^-}. \tag{4.6}$$

Notation	Definition
$\mathcal{F}_h$	Set of faces of $\mathcal{T}_h$
$\mathcal{F}_h^b$	Set of boundary faces, i.e., $\{F \in \mathcal{F}_h \ : \ F \subset \partial\Omega\}$
$\mathcal{F}_h^D \cup \mathcal{F}_h^N \cup \mathcal{F}_h^C$	Partition of $\mathcal{F}_h^b$ induced by the boundary and contact conditions
$\mathcal{F}_h^i$	Set of interior faces, i.e., $\mathcal{F}_h \setminus \mathcal{F}_h^b$
$\mathcal{F}_T$	Set of faces of the element $T \in \mathcal{T}_h$, i.e., $\{F \in \mathcal{F}_h \ : \ F \subset \partial T\}$
$\mathcal{F}_T^\bullet, \bullet \in \{b, D, N, C\}$	$\mathcal{F}_T \cap \mathcal{F}_h^\bullet, T \in \mathcal{T}_h$
$\mathcal{V}_h$	Set of all the vertices of $\mathcal{T}_h$
$\mathcal{V}_h^b$	Set of boundary vertices, i.e., $\{\boldsymbol{a} \in \mathcal{V}_h \ : \ \boldsymbol{a} \in \partial\Omega\}$
$\mathcal{V}_h^i$	Set of interior vertices, i.e., $\mathcal{V}_h \setminus \mathcal{V}_h^b$
$\mathcal{V}_T$	Set of vertices of the element $T \in \mathcal{T}_h$, i.e., $\{\boldsymbol{a} \in \mathcal{V}_h \ : \ \boldsymbol{a} \in \partial T\}$
$\mathcal{V}_F$	Set of vertices of the mesh face $F \in \mathcal{F}_h$, i.e., $\{\boldsymbol{a} \in \mathcal{V}_h \ : \ \boldsymbol{a} \in \partial F\}$
$\omega_{\boldsymbol{a}}$	Union of the elements sharing the vertex $\boldsymbol{a} \in \mathcal{V}_h$, i.e., $\bigcup_{T \in \mathcal{T}_h, \, \boldsymbol{a} \in \partial T} T$

Table 4.1 – Mesh-related notations.

Remark 4.3 (Case $\theta = 0$)**.** *The linear operator $P^n_{\theta,\gamma}$ is well defined on $\boldsymbol{V}_h$ since it is a subspace of the space of broken polynomials. It can be easily extended to the space $\boldsymbol{H}^1_D(\Omega)$ in the case $\theta = 0$, for which $P^n_{0,\gamma}(\boldsymbol{v}) = -\gamma v^n$, as $\boldsymbol{v} \in \boldsymbol{H}^1(\Omega)$ guarantees $\boldsymbol{v}|_{\Gamma_C} \in \boldsymbol{L}^2(\Gamma_C)$ by the trace theorem. Incidentally, the symmetric and skew-symmetric variations of the Nitsche method (4.7) below (corresponding, respectively, to $\theta = 1$ and $\theta = -1$) could also be treated at the price of technicalities concerning the definition of the trace $\sigma^n(\boldsymbol{v})$ appearing in the definition (4.5) of $P^n_{\theta,\gamma}$. The details are omitted for the sake of brevity.*

From now on, $\gamma_0 > 0$ will denote a fixed constant called *Nitsche parameter*, and we suppose that γ is the positive piecewise constant function on Γ_C which satisfies: For all $T \in \mathcal{T}_h$ such that $|\partial T \cap \Gamma_C| > 0$,

$$\gamma|_{\partial T \cap \Gamma_C} = \frac{\gamma_0}{h_T}.$$

We consider the following method à la Nitsche to approximate problem (4.2), originally introduced in [28]: Find $\boldsymbol{u}_h \in \boldsymbol{V}_h$ such that

$$a(\boldsymbol{u}_h, \boldsymbol{v}_h) - \left(\left[P^n_{1,\gamma}(\boldsymbol{u}_h) \right]_{\mathbb{R}^-}, v_h^n \right)_{\Gamma_C} = L(\boldsymbol{v}_h) \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_h. \tag{4.7}$$

For the a priori analysis of the method, we refer to [28].

4.3 Basic a posteriori error estimate

In this section we derive a basic a posteriori error estimate based on the notion of equilibrated stress reconstruction.

4.3.1 Error measure

In the framework of a posteriori error estimation, the dual norm of a residual functional can be used as a measure of the error between the exact solution $\boldsymbol{u}$ of the problem and the solution $\boldsymbol{u}_h$ obtained with the finite element method. Denoting by $(\boldsymbol{H}^1_D(\Omega))^*$ the dual space of $\boldsymbol{H}^1_D(\Omega)$, for any $\boldsymbol{w}_h \in \boldsymbol{V}_h$ the *residual* $\mathcal{R}(\boldsymbol{w}_h) \in (\boldsymbol{H}^1_D(\Omega))^*$ is defined by

$$\langle \mathcal{R}(\boldsymbol{w}_h), \boldsymbol{v} \rangle \coloneqq L(\boldsymbol{v}) - a(\boldsymbol{w}_h, \boldsymbol{v}) + \left(\left[P^n_{1,\gamma}(\boldsymbol{w}_h) \right]_{\mathbb{R}^-}, v^n \right)_{\Gamma_C} \qquad \forall \boldsymbol{v} \in \boldsymbol{H}^1_D(\Omega), \tag{4.8}$$

where $\langle \cdot, \cdot \rangle$ denotes the duality pairing between $\boldsymbol{H}^1_D(\Omega)$ and $(\boldsymbol{H}^1_D(\Omega))^*$. We equip $\boldsymbol{H}^1_D(\Omega)$ with the following mesh-dependent norm:

$$|||\boldsymbol{v}|||^2 \coloneqq \|\boldsymbol{\nabla} \boldsymbol{v}\|^2 + |\boldsymbol{v}|^2_{C,h} \qquad \forall \boldsymbol{v} \in \boldsymbol{H}^1_D(\Omega), \tag{4.9}$$

where

$$|\boldsymbol{v}|^2_{C,h} \coloneqq \sum_{F \in \mathcal{F}^C_h} \frac{1}{h_F} \|\boldsymbol{v}\|^2_F \qquad \forall \boldsymbol{v} \in \boldsymbol{H}^1_D(\Omega). \tag{4.10}$$

It is easy to show that $|\cdot|_{C,h}$ is subadditive and absolutely homogeneous, i.e., it is a seminorm. As a consequence, also $|||\cdot|||$ is subadditive and absolutely homogeneous. Moreover, if we suppose $|||\boldsymbol{v}||| = 0$, then both $\|\boldsymbol{\nabla} \boldsymbol{v}\|$ and $|\boldsymbol{v}|_{C,h}$ are zero, and this implies $\boldsymbol{v} = \boldsymbol{0}$ by the Friedrichs inequality in $\boldsymbol{H}^1_D(\Omega)$, showing that $|||\cdot|||$ is indeed a norm on $\boldsymbol{H}^1_D(\Omega)$.

The dual norm of the residual of a function $\boldsymbol{w}_h \in \boldsymbol{V}_h$ on the normed space $(\boldsymbol{H}^1_D(\Omega), |||\cdot|||)$ is given by

$$|||\mathcal{R}(\boldsymbol{w}_h)|||_* \coloneqq \sup_{\boldsymbol{v} \in \boldsymbol{H}^1_D(\Omega),\, |||\boldsymbol{v}|||=1} \langle \mathcal{R}(\boldsymbol{w}_h), \boldsymbol{v} \rangle . \tag{4.11}$$

In what follows, the quantity $|||\mathcal{R}(\boldsymbol{u}_h)|||_*$ will be used as a measure of the error committed approximating the exact solution $\boldsymbol{u}$ with $\boldsymbol{u}_h$.

4.3.2 A posteriori error estimate

We start this section by introducing the concept of equilibrated stress reconstruction and the definition of five error estimators.

Definition 4.4 (Equilibrated stress reconstruction)**.** We will call *equilibrated stress reconstruction* any second order tensor $\boldsymbol{\sigma}_h$ such that:

1. $\boldsymbol{\sigma}_h \in \mathbb{H}(\mathbf{div}, \Omega)$,
2. $(\mathbf{div}\, \boldsymbol{\sigma}_h + \boldsymbol{f}, \boldsymbol{v})_T = 0$ for every $\boldsymbol{v} \in \boldsymbol{\mathcal{P}}^0(T)$ and every $T \in \mathcal{T}_h$,
3. $(\boldsymbol{\sigma}_h \boldsymbol{n})|_F \in \boldsymbol{L}^2(F)$ for every $F \in \mathcal{F}^N_h \cup \mathcal{F}^C_h$, and $(\boldsymbol{\sigma}_h \boldsymbol{n}, \boldsymbol{v})_F = (\boldsymbol{g}_N, \boldsymbol{v})_F$ for every $\boldsymbol{v} \in \boldsymbol{\mathcal{P}}^0(F)$ and every $F \in \mathcal{F}^N_h$,
4. $\boldsymbol{\sigma}^{\boldsymbol{t}}_h = \boldsymbol{0}$ on Γ_C.

Given an equilibrated stress reconstruction $\boldsymbol{\sigma}_h$, for every element $T \in \mathcal{T}_h$, we define the following local error estimators:

$$\eta_{\mathrm{osc},T} \coloneqq \frac{h_T}{\pi}\|\boldsymbol{f} + \mathbf{div}\,\boldsymbol{\sigma}_h\|_T, \qquad \text{(oscillation)} \tag{4.12a}$$

$$\eta_{\mathrm{str},T} \coloneqq \|\boldsymbol{\sigma}_h - \boldsymbol{\sigma}(\boldsymbol{u}_h)\|_T, \qquad \text{(stress)} \tag{4.12b}$$

$$\eta_{\mathrm{Neu},T} \coloneqq \sum_{F\in\mathcal{F}_T^N} C_{t,T,F} h_F^{1/2}\|\boldsymbol{g}_N - \boldsymbol{\sigma}_h\boldsymbol{n}\|_F, \qquad \text{(Neumann)} \tag{4.12c}$$

$$\eta_{\mathrm{cnt},T} \coloneqq \sum_{F\in\mathcal{F}_T^C} h_F^{1/2}\left\|\left[P^n_{1,\gamma}(\boldsymbol{u}_h)\right]_{\mathbb{R}^-} - \sigma_h^n\right\|_F. \qquad \text{(contact)} \tag{4.12d}$$

Here, $C_{t,T,F}$ is the constant of the trace inequality $\|v - \overline{v}_F\|_F \le C_{t,T,F} h_F^{1/2}\|\boldsymbol{\nabla} v\|_T$ with $\overline{v}_F \coloneqq \frac{1}{|F|}\int_F v$ valid for every $v \in H^1(T)$ and $F \in \mathcal{F}_T$; see [126, Theorem 4.6.3] or [40, Section 1.4].

The estimator $\eta_{\mathrm{osc},T}$ represents the residual of the force balance equation (4.2a) inside the element T, $\eta_{\mathrm{str},T}$ the difference between the Cauchy stress tensor computed from the approximate solution and the equilibrated stress reconstruction, $\eta_{\mathrm{Neu},T}$ the residual of the Neumann boundary condition (4.2d), and $\eta_{\mathrm{cnt},T}$ the residual of the normal condition (4.2e) on the contact boundary.

Theorem 4.5 (A posteriori error estimate for the dual norm of the residual)**.** *Let $\boldsymbol{u}_h$ be the solution of (4.7), $\mathcal{R}(\boldsymbol{u}_h)$ the residual defined by (4.8), and $\boldsymbol{\sigma}_h$ an equilibrated stress reconstruction in the sense of Definition 4.4. Then,*

$$|||\mathcal{R}(\boldsymbol{u}_h)|||_* \le \left(\sum_{T\in\mathcal{T}_h}\left(\left(\eta_{\mathrm{osc},T} + \eta_{\mathrm{str},T} + \eta_{\mathrm{Neu},T}\right)^2 + \left(\eta_{\mathrm{cnt},T}\right)^2\right)\right)^{1/2}.$$

Proof. Thanks to the regularity of $\boldsymbol{\sigma}_h$ and of its normal trace (see Properties 1. and 3. in Definition 4.4), the following Green formula holds:

$$(\boldsymbol{\sigma}_h, \boldsymbol{\nabla} v) = -(\mathbf{div}\,\boldsymbol{\sigma}_h, v) + (\boldsymbol{\sigma}_h\boldsymbol{n}, v)_{\Gamma_N} + (\sigma_h^n, v^n)_{\Gamma_C} \qquad \forall v \in \boldsymbol{H}^1_D(\Omega), \tag{4.13}$$

where we have also used the decomposition (4.1) of the normal stress reconstruction $\boldsymbol{\sigma}_h\boldsymbol{n}$ into normal and tangential components on the contact boundary Γ_C, and the fact that $\boldsymbol{\sigma}_h^t|_{\Gamma_C} = \mathbf{0}$ thanks to Property 4. in Definition 4.4. Now, fix $v \in \boldsymbol{H}^1_D(\Omega)$ such that $|||v|||^2 = \|\nabla v\|^2 + |v|^2_{C,h} = 1$ and consider the argument of the supremum in the definition (4.11) of the dual norm of the residual. Expanding $L(\cdot)$ and $a(\cdot,\cdot)$ according to their definition (4.4), adding and subtracting the term $(\boldsymbol{\sigma}_h, \boldsymbol{\nabla} v)$, using the symmetry of $\boldsymbol{\sigma}(\boldsymbol{u}_h)$, and applying Green's formula (4.13), we obtain

$$\begin{aligned}
\langle\mathcal{R}(\boldsymbol{u}_h), v\rangle &= (\boldsymbol{f}, v) + (\boldsymbol{g}_N, v)_{\Gamma_N} - (\boldsymbol{\sigma}(\boldsymbol{u}_h), \boldsymbol{\varepsilon}(v)) + \left(\left[P^n_{1,\gamma}(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}, v^n\right)_{\Gamma_C} \\
&\quad + (\boldsymbol{\sigma}_h, \boldsymbol{\nabla} v) - (\boldsymbol{\sigma}_h, \boldsymbol{\nabla} v) \\
&= (\boldsymbol{f} + \mathbf{div}\,\boldsymbol{\sigma}_h, v) + (\boldsymbol{\sigma}_h - \boldsymbol{\sigma}(\boldsymbol{u}_h), \boldsymbol{\nabla} v) + (\boldsymbol{g}_N - \boldsymbol{\sigma}_h\boldsymbol{n}, v)_{\Gamma_N} \\
&\quad + \left(\left[P^n_{1,\gamma}(\boldsymbol{u}_h)\right]_{\mathbb{R}^-} - \sigma_h^n, v^n\right)_{\Gamma_C} \\
&=: \mathfrak{T}_1 + \cdots + \mathfrak{T}_4.
\end{aligned}$$

We estimate each term separately. Denoting by $\boldsymbol{\Pi}_T^0$ the L^2-orthogonal projection onto $\boldsymbol{\mathcal{P}}^0(T)$, and using Property 2. of Definition 4.4 with test function $\boldsymbol{\Pi}_T^0\boldsymbol{v} \in \boldsymbol{\mathcal{P}}^0(T)$, the Cauchy-Schwarz inequality, and the Poincaré inequality $\|\boldsymbol{v} - \boldsymbol{\Pi}_T^0\boldsymbol{v}\|_T \le h_T\pi^{-1}\|\boldsymbol{\nabla}\boldsymbol{v}\|_T$ (see for example [107, 12]), the first term becomes

$$\begin{aligned}\mathfrak{T}_1 &= \sum_{T\in\mathcal{T}_h}(\boldsymbol{f} + \mathbf{div}\,\boldsymbol{\sigma}_h, \boldsymbol{v} - \boldsymbol{\Pi}_T^0\boldsymbol{v})_T \le \sum_{T\in\mathcal{T}_h}\|\boldsymbol{f} + \mathbf{div}\,\boldsymbol{\sigma}_h\|_T\|\boldsymbol{v} - \boldsymbol{\Pi}_T^0\boldsymbol{v}\|_T \\ &\le \sum_{T\in\mathcal{T}_h}\frac{h_T}{\pi}\|\boldsymbol{f} + \mathbf{div}\,\boldsymbol{\sigma}_h\|_T\|\boldsymbol{\nabla}\boldsymbol{v}\|_T = \sum_{T\in\mathcal{T}_h}\eta_{\mathrm{osc},T}\|\boldsymbol{\nabla}\boldsymbol{v}\|_T.\end{aligned}$$

For the second term, we simply use the Cauchy-Schwarz inequality:

$$\mathfrak{T}_2 \le \sum_{T\in\mathcal{T}_h}\|\boldsymbol{\sigma}_h - \boldsymbol{\sigma}(\boldsymbol{u}_h)\|_T\|\boldsymbol{\nabla}\boldsymbol{v}\|_T = \sum_{T\in\mathcal{T}_h}\eta_{\mathrm{str},T}\|\boldsymbol{\nabla}\boldsymbol{v}\|_T.$$

Denoting by $\boldsymbol{\Pi}_F^0$ the L^2-orthogonal projection onto $\boldsymbol{\mathcal{P}}^0(F)$, and using Property 3. of Definition 4.4 with $\boldsymbol{\Pi}_F^0\boldsymbol{v} \in \boldsymbol{\mathcal{P}}^0(F)$ as a test function, the Cauchy-Schwarz inequality, and the trace inequality $\|\boldsymbol{v} - \boldsymbol{\Pi}_F^0\boldsymbol{v}\|_F \le C_{t,T,F}h_F^{1/2}\|\boldsymbol{\nabla}\boldsymbol{v}\|_T$, $F \subset \partial T$, we have

$$\begin{aligned}\mathfrak{T}_3 &= \sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_T^N}(\boldsymbol{g}_N - \boldsymbol{\sigma}_h\boldsymbol{n}, \boldsymbol{v} - \boldsymbol{\Pi}_F^0\boldsymbol{v})_F \le \sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_T^N}\|\boldsymbol{g}_N - \boldsymbol{\sigma}_h\boldsymbol{n}\|_F\|\boldsymbol{v} - \boldsymbol{\Pi}_F^0\boldsymbol{v}\|_F \\ &\le \sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_T^N}C_{t,T,F}h_F^{1/2}\|\boldsymbol{g}_N - \boldsymbol{\sigma}_h\boldsymbol{n}\|_F\|\boldsymbol{\nabla}\boldsymbol{v}\|_T = \sum_{T\in\mathcal{T}_h}\eta_{\mathrm{Neu},T}\|\boldsymbol{\nabla}\boldsymbol{v}\|_T,\end{aligned}$$

where we recall that, for any $T \in \mathcal{T}_h$, $\mathcal{F}_T^N$ collects the Neumann faces of T contained in $\mathcal{F}_h^N$. Finally, we consider the term on Γ_C. We define, for all $T \in \mathcal{T}_h$, the local counterpart of the seminorm (4.10)

$$|\boldsymbol{v}|_{C,T}^2 \coloneqq \sum_{F\in\mathcal{F}_T^C}\frac{1}{h_F}\|\boldsymbol{v}\|_F^2,$$

where $\mathcal{F}_T^C$ is the (possibly empty) set collecting the contact faces of T contained in $\partial T \cap \Gamma_C$. Then, using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}\mathfrak{T}_4 &\le \sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_T^C}\left\|\left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-} - \sigma_h^n\right\|_F\|v^n\|_F \le \sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_T^C}h_F^{1/2}\left\|\left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-} - \sigma_h^n\right\|_F|\boldsymbol{v}|_{C,T} \\ &= \sum_{T\in\mathcal{T}_h}\eta_{\mathrm{cnt},T}\,|\boldsymbol{v}|_{C,T}\,.\end{aligned}$$

Let, for the sake of brevity, $\eta_{a,T} \coloneqq \eta_{\mathrm{osc},T} + \eta_{\mathrm{str},T} + \eta_{\mathrm{Neu},T}$ for any element $T \in \mathcal{T}_h$. Combining

the above results and applying the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
|||\mathcal{R}(\boldsymbol{u}_h)|||_* &\leq \sup_{\boldsymbol{v}\in\boldsymbol{H}^1_D(\Omega),\,|||\boldsymbol{v}|||=1} \left\{ \sum_{T\in\mathcal{T}_h} \left(\eta_{a,T} \|\boldsymbol{\nabla} \boldsymbol{v}\|_T + \eta_{\mathrm{cnt},T}\, |\boldsymbol{v}|_{C,T} \right) \right\} \\
&\leq \sup_{\boldsymbol{v}\in\boldsymbol{H}^1_D(\Omega),\,|||\boldsymbol{v}|||=1} \left\{ \left(\sum_{T\in\mathcal{T}_h} \left(\left(\eta_{a,T}\right)^2 + \left(\eta_{\mathrm{cnt},T}\right)^2 \right) \right)^{1/2} \left(\sum_{T\in\mathcal{T}_h} \left(\|\boldsymbol{\nabla} \boldsymbol{v}\|^2_T + |\boldsymbol{v}|^2_{C,T} \right) \right)^{1/2} \right\} \\
&= \left(\sum_{T\in\mathcal{T}_h} \left(\left(\eta_{\mathrm{osc},T} + \eta_{\mathrm{str},T} + \eta_{\mathrm{Neu},T}\right)^2 + \left(\eta_{\mathrm{cnt},T}\right)^2 \right) \right)^{1/2}.
\end{aligned}$$

□

Remark 4.6 (A posteriori error estimate for stress reconstructions with contact friction estimator)**.** *Even without Property 4. in Definition 4.4 one can easily obtain an a posteriori error estimate similarly to Theorem 4.5: introducing a fifth local estimator*

$$\eta_{\mathrm{fric},T} \coloneqq \sum_{F\in\mathcal{F}_T^C} h_F^{1/2} \|\boldsymbol{\sigma}_h^{\boldsymbol{t}}\|_F, \qquad \textit{(friction)}$$

which represents the residual of the tangential condition (4.2f) on the contact boundary, one gets

$$|||\mathcal{R}(\boldsymbol{u}_h)|||_* \leq \left(\sum_{T\in\mathcal{T}_h} \left(\left(\eta_{\mathrm{osc},T} + \eta_{\mathrm{str},T} + \eta_{\mathrm{Neu},T}\right)^2 + \left(\eta_{\mathrm{cnt},T} + \eta_{\mathrm{fric},T}\right)^2 \right) \right)^{1/2}.$$

4.3.3 Comparison between the residual dual norm and the energy norm

The goal of this section is to compare the dual norm $|||\mathcal{R}(\boldsymbol{u}_h)|||_*$ with the energy norm $\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}}$ of the error, where

$$\|\boldsymbol{v}\|^2_{\mathrm{en}} \coloneqq a(\boldsymbol{v},\boldsymbol{v}) = (\boldsymbol{\sigma}(\boldsymbol{v}), \boldsymbol{\varepsilon}(\boldsymbol{v})) \qquad \forall \boldsymbol{v} \in \boldsymbol{H}^1_D(\Omega). \tag{4.14}$$

Remark 4.7 (Coercivity of the bilinear form a)**.** *The bilinear form $a(\cdot,\cdot)$ on the space $(\boldsymbol{H}^1_D(\Omega), \|\cdot\|_{1,\Omega})$ is elliptic with a constant α which depends on the Lamé parameter μ and on the Korn constant C_K:*

$$\alpha \|\boldsymbol{v}\|^2_{1,\Omega} \leq a(\boldsymbol{v},\boldsymbol{v}) = \|\boldsymbol{v}\|^2_{\mathrm{en}} \qquad \forall \boldsymbol{v} \in \boldsymbol{H}^1_D(\Omega). \tag{4.15}$$

Throughout the rest of this section, we adopt the following shorthand notation: For every $a, b \in \mathbb{R}$, we write $a \lesssim b$ for $a \leq Cb$ with $C > 0$ independent of the mesh size h and of the Nitsche parameter γ_0.

Theorem 4.8 (Control of the energy norm)**.** *Assume that the solution $\boldsymbol{u}$ of the continuous problem (4.2) belongs to $\boldsymbol{H}^{\frac{3}{2}+\nu}(\Omega)$ for some $\nu > 0$, and let $\boldsymbol{u}_h \in \boldsymbol{V}_h$ be the solution of the discrete problem (4.7). Then,*

$$\alpha^{1/2} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}} \lesssim |||\mathcal{R}(\boldsymbol{u}_h)|||_* + \sum_{F\in\mathcal{F}_h^C} \frac{1}{h_F^{1/2}} \left\| \sigma^n(\boldsymbol{u}) - \left[P^n_{1,\gamma}(\boldsymbol{u}_h) \right]_{\mathbb{R}^-} \right\|_F. \tag{4.16}$$

Furthermore, if the saturation assumption (see [26, 27])

$$\left\| \left(\frac{\gamma_0}{\gamma} \right)^{1/2} \sigma^n(\boldsymbol{u} - \boldsymbol{u}_h) \right\|_{\Gamma_C} \lesssim \|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega} \tag{4.17}$$

holds and γ_0 is sufficiently large, then

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}} \lesssim |||\mathcal{R}(\boldsymbol{u}_h)|||_*. \tag{4.18}$$

Remark 4.9 (Role of the regularity assumption)**.** *In Theorem 4.8, the solution of the contact problem (4.2) $\boldsymbol{u}$ is supposed to be sufficiently regular in order to ensure that the normal component of the Cauchy stress tensor is square-integrable on the contact boundary Γ_C. Using a trace theorem, this regularity assumption implies $\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n} \in \boldsymbol{H}^r(\Gamma_C) \subset \boldsymbol{L}^2(\Gamma_C)$. Furthermore, notice that, since we only need the latter condition near the contact boundary Γ_C, the regularity assumption on $\boldsymbol{u}$ of Theorem 4.8 can be possibly relaxed.*

Remark 4.10 (Saturation assumption)**.** *The basic error estimate in Theorem 4.5 is obtained without the need for a saturation assumption (only the norm comparisons (4.18) and (4.31) in Theorems 4.8 and 4.11 below, respectively, are obtained under this assumption (4.17)). A posteriori error estimates for similar problems that do not require the saturation assumption have also been recently derived in [65, 20].*

Proof of Theorem 4.8. The proof adapts the ideas of [27, Theorem 3.5].

1) Proof of (4.16). Let $\boldsymbol{v}_h \in \boldsymbol{V}_h$. Using the definition (4.14) of the energy norm and the bilinearity of $a(\,\cdot\,,\,\cdot\,)$, we can write

$$\begin{aligned} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}}^2 &= a(\boldsymbol{u} - \boldsymbol{u}_h, \boldsymbol{u} - \boldsymbol{u}_h) \\ &= a(\boldsymbol{u}, \boldsymbol{u} - \boldsymbol{u}_h) - a(\boldsymbol{u}_h, \boldsymbol{u} - \boldsymbol{v}_h) - a(\boldsymbol{u}_h, \boldsymbol{v}_h - \boldsymbol{u}_h). \end{aligned} \tag{4.19}$$

For the term $a(\boldsymbol{u}, \boldsymbol{u} - \boldsymbol{u}_h)$, we first use the definition (4.4) of the bilinear form $a(\,\cdot\,,\,\cdot\,)$ followed by the symmetry of the Cauchy stress tensor $\boldsymbol{\sigma}(\boldsymbol{u})$ to replace $\boldsymbol{\varepsilon}(\boldsymbol{u} - \boldsymbol{u}_h)$ with $\boldsymbol{\nabla}(\boldsymbol{u} - \boldsymbol{u}_h)$, then an integration by parts, and, finally, the fact that $\boldsymbol{u}$ satisfies (4.2) a.e. to infer:

$$\begin{aligned} a(\boldsymbol{u}, \boldsymbol{u} - \boldsymbol{u}_h) &= (\boldsymbol{\sigma}(\boldsymbol{u}), \boldsymbol{\varepsilon}(\boldsymbol{u} - \boldsymbol{u}_h)) = (\boldsymbol{\sigma}(\boldsymbol{u}), \boldsymbol{\nabla}(\boldsymbol{u} - \boldsymbol{u}_h)) \\ &= -(\mathbf{div}\, \boldsymbol{\sigma}(\boldsymbol{u}), \boldsymbol{u} - \boldsymbol{u}_h) + (\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n}, \boldsymbol{u} - \boldsymbol{u}_h)_{\partial\Omega} \\ &= (\boldsymbol{f}, \boldsymbol{u} - \boldsymbol{u}_h) + (\boldsymbol{g}_N, \boldsymbol{u} - \boldsymbol{u}_h)_{\Gamma_N} + (\sigma^n(\boldsymbol{u}), u^n - u_h^n)_{\Gamma_C}. \end{aligned} \tag{4.20}$$

Notice that, in the last term, only the normal component of the traction appears as $\boldsymbol{\sigma}^t(\boldsymbol{u}) = \boldsymbol{0}$ on Γ_C by (4.2f).

Concerning the term $a(\boldsymbol{u}_h, \boldsymbol{v}_h - \boldsymbol{u}_h)$ in (4.19), since $\boldsymbol{u}_h$ solves (4.7) and $\boldsymbol{v}_h - \boldsymbol{u}_h \in \boldsymbol{V}_h$, we have

$$a(\boldsymbol{u}_h, \boldsymbol{v}_h - \boldsymbol{u}_h) = (\boldsymbol{f}, \boldsymbol{v}_h - \boldsymbol{u}_h) + (\boldsymbol{g}_N, \boldsymbol{v}_h - \boldsymbol{u}_h)_{\Gamma_N} + \left(\left[P_{1,\gamma}^n(\boldsymbol{u}_h) \right]_{\mathbb{R}^-}, v_h^n - u_h^n \right)_{\Gamma_C}, \tag{4.21}$$

where we have additionally expanded the linear form $L(\,\cdot\,)$ according to its definition (4.4).

Plugging (4.20) and (4.21) into (4.19), grouping the terms involving $\boldsymbol{f}$ and $\boldsymbol{g}_N$, and using the definition (4.4) of the bilinear form $a(\cdot,\cdot)$ to rewrite the term $a(\boldsymbol{u}_h, \boldsymbol{u}-\boldsymbol{v}_h)$ in (4.19), we then obtain

$$\begin{aligned}\|\boldsymbol{u}-\boldsymbol{u}_h\|_{\mathrm{en}}^2 &= (\boldsymbol{f}, \boldsymbol{u}-\boldsymbol{v}_h) + (\boldsymbol{g}_N, \boldsymbol{u}-\boldsymbol{v}_h)_{\Gamma_N} + (\sigma^n(\boldsymbol{u}), u^n - u_h^n)_{\Gamma_C} \\ &\quad - (\boldsymbol{\sigma}(\boldsymbol{u}_h), \boldsymbol{\varepsilon}(\boldsymbol{u}-\boldsymbol{v}_h)) - \left(\left[P_{1,\gamma}^n(u_h)\right]_{\mathbb{R}^-}, v_h^n - u_h^n\right)_{\Gamma_C} \\ &\overset{=}{} \mathfrak{T}_1 + \mathfrak{T}_2\end{aligned} \tag{4.22}$$

where, recalling the definition (4.8) of the residual,

$$\mathfrak{T}_1 \coloneqq \langle \mathcal{R}(\boldsymbol{u}_h), \boldsymbol{u}-\boldsymbol{v}_h\rangle, \qquad \mathfrak{T}_2 \coloneqq \left(\sigma^n(\boldsymbol{u}) - \left[P_{1,\gamma}^n(u_h)\right]_{\mathbb{R}^-}, u^n - u_h^n\right)_{\Gamma_C}.$$

Notice that the reformulation of $\mathfrak{T}_1$ in terms of the residual $\mathcal{R}(\boldsymbol{u}_h)$ is a consequence of (4.8) and of the definition (4.4) of the linear form $L(\cdot)$ and of the bilinear form $a(\cdot,\cdot)$.

For the first term, we can write, by definition (4.11) of the dual norm,

$$\mathfrak{T}_1 \le |||\boldsymbol{u}-\boldsymbol{v}_h||| \, |||\mathcal{R}(\boldsymbol{u}_h)|||_*. \tag{4.23}$$

We now want to show that $|||\boldsymbol{u}-\boldsymbol{v}_h||| \lesssim \|\boldsymbol{u}-\boldsymbol{u}_h\|_{\mathrm{en}}$ for a properly selected function $\boldsymbol{v}_h$. From now on, we fix $\boldsymbol{v}_h = \boldsymbol{u}_h + \mathcal{I}_h(\boldsymbol{u}-\boldsymbol{u}_h)$, where $\mathcal{I}_h\colon \boldsymbol{H}_D^1(\Omega) \to \boldsymbol{V}_h$ is the quasi-interpolation operator defined in [13, Eq. (4.11)], whose main properties are summarized in [27, Lemma 2.1]. We analyze separately the two parts composing the norm $|||\cdot|||$ (see (4.9)). For the $\boldsymbol{H}^1$-seminorm, we use, in this order, the triangle inequality, the choice of $\boldsymbol{v}_h$, the boundedness in the $\boldsymbol{H}^1$-norm of the operator $\mathcal{I}_h$ (i.e., $\|\mathcal{I}_h\boldsymbol{v}\|_{1,\Omega} \lesssim \|\boldsymbol{v}\|_{1,\Omega}$ for every $\boldsymbol{v} \in \boldsymbol{H}_D^1(\Omega)$), and the ellipticity (4.15) of the bilinear form $a(\cdot,\cdot)$ to write:

$$\begin{aligned}\|\boldsymbol{\nabla}(\boldsymbol{u}-\boldsymbol{v}_h)\| &\le \|\boldsymbol{u}-\boldsymbol{v}_h\|_{1,\Omega} \le \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega} + \|\boldsymbol{u}_h-\boldsymbol{v}_h\|_{1,\Omega} \\ &= \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega} + \|\mathcal{I}_h(\boldsymbol{u}-\boldsymbol{u}_h)\|_{1,\Omega} \\ &\lesssim \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega} \le \alpha^{-1/2}\|\boldsymbol{u}-\boldsymbol{u}_h\|_{\mathrm{en}}.\end{aligned} \tag{4.24}$$

Next, using the definition (4.10) of the seminorm $|\cdot|_{C,h}$, the choice of $\boldsymbol{v}_h$, and the ellipticity (4.15) of $a(\cdot,\cdot)$, we obtain:

$$\begin{aligned}|\boldsymbol{u}-\boldsymbol{v}_h|_{C,h}^2 &= \sum_{F\in\mathcal{F}_h^C} \frac{1}{h_F}\|\boldsymbol{u}-\boldsymbol{v}_h\|_F^2 = \sum_{F\in\mathcal{F}_h^C} \frac{1}{h_F}\|\boldsymbol{u}-\boldsymbol{u}_h-\mathcal{I}_h(\boldsymbol{u}-\boldsymbol{u}_h)\|_F^2 \\ &\lesssim \sum_{F\in\mathcal{F}_h^C} \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\tilde{\omega}_F}^2 \lesssim \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}^2 \le \alpha^{-1}\|\boldsymbol{u}-\boldsymbol{u}_h\|_{\mathrm{en}}^2,\end{aligned} \tag{4.25}$$

where, to pass to the second line, we have used the following trace approximation property of $\mathcal{I}_h$ (see [27, Lemma 2.1]): For all $F \in \mathcal{F}_h$,

$$\|\boldsymbol{v}-\mathcal{I}_h\boldsymbol{v}\|_F \lesssim h_F^{1/2}\|\boldsymbol{v}\|_{1,\tilde{\omega}_F} \qquad \forall \boldsymbol{v} \in \boldsymbol{H}_D^1(\Omega),$$

with $\tilde{\omega}_F$ standing for the union of the mesh elements sharing at least one vertex with F, see Figure 4.1. Recalling the definition (4.9) of the triple norm, squaring (4.24) and summing it to (4.25), and taking the square root of the resulting inequality, we conclude that

$$|||\boldsymbol{u}-\boldsymbol{v}_h||| = \left(\|\boldsymbol{\nabla}(\boldsymbol{u}-\boldsymbol{v}_h)\|^2 + |\boldsymbol{u}-\boldsymbol{v}_h|_{C,h}^2\right)^{1/2} \lesssim \alpha^{-1/2}\|\boldsymbol{u}-\boldsymbol{u}_h\|_{\mathrm{en}}.$$

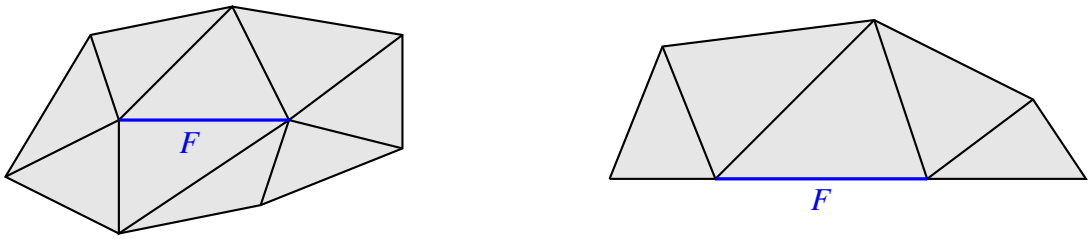


Figure 4.1 – Illustration of $\tilde{\omega}_F$ for $F \in \mathcal{F}_h^i$ (*left*) and for $F \in \mathcal{F}_h^b$ (*right*).

Combining this bound with (4.23), we obtain

$$\mathfrak{T}_1 \lesssim \alpha^{-1/2} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}} |||\mathcal{R}(\boldsymbol{u}_h)|||_*. \tag{4.26}$$

We now consider the term $\mathfrak{T}_2$. Using the Cauchy-Schwarz and trace inequalities, we have

$$\begin{aligned}
\mathfrak{T}_2 &= \sum_{F\in\mathcal{F}_h^C} \left(\sigma^n(\boldsymbol{u}) - \left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}, u^n - u_h^n\right)_F \\
&\lesssim \left(\sum_{F\in\mathcal{F}_h^C} \frac{1}{h_F} \left\|\sigma^n(\boldsymbol{u}) - \left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}\right\|_F^2\right)^{1/2} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega} \\
&\lesssim \alpha^{-1/2}\left(\sum_{F\in\mathcal{F}_h^C} \frac{1}{h_F} \left\|\sigma^n(\boldsymbol{u}) - \left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}\right\|_F^2\right)^{1/2} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}}.
\end{aligned} \tag{4.27}$$

Inserting (4.26) and (4.27) into (4.22), we obtain (4.16).

2) Proof of (4.18). For the second part of the theorem, we work under the saturation assumption (4.17). Using the contact condition $\sigma^n(\boldsymbol{u}) = \left[P_{1,\gamma}^n(\boldsymbol{u})\right]_{\mathbb{R}^-}$ (see (4.6)), and the definition (4.5) of the operator $P_{1,\gamma}^n$, we have

$$\begin{aligned}
\mathfrak{T}_2 &= \left(\left[P_{1,\gamma}^n(\boldsymbol{u})\right]_{\mathbb{R}^-} - \left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}, u^n - u_h^n\right)_{\Gamma_C} \\
&= \left(\left[P_{1,\gamma}^n(\boldsymbol{u})\right]_{\mathbb{R}^-} - \left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}, \frac{1}{\gamma}\left[\gamma(u^n - u_h^n) - \sigma^n(\boldsymbol{u} - \boldsymbol{u}_h) + \sigma^n(\boldsymbol{u} - \boldsymbol{u}_h)\right]\right)_{\Gamma_C} \\
&= -\left(\left[P_{1,\gamma}^n(\boldsymbol{u})\right]_{\mathbb{R}^-} - \left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}, \frac{1}{\gamma}\left(P_{1,\gamma}^n(\boldsymbol{u}) - P_{1,\gamma}(\boldsymbol{u}_h)\right)\right)_{\Gamma_C} \\
&\quad + \left(\left[P_{1,\gamma}^n(\boldsymbol{u})\right]_{\mathbb{R}^-} - \left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}, \frac{1}{\gamma}\sigma^n(\boldsymbol{u} - \boldsymbol{u}_h)\right)_{\Gamma_C}.
\end{aligned} \tag{4.28}$$

Due to the fact that $a[a]_{\mathbb{R}^-} = ([a]_{\mathbb{R}^-})^2$ and $a[b]_{\mathbb{R}^-} \le [a]_{\mathbb{R}^-}[b]_{\mathbb{R}^-}$, it follows that

$$(a-b)([a]_{\mathbb{R}^-} - [b]_{\mathbb{R}^-}) = a[a]_{\mathbb{R}^-} + b[b]_{\mathbb{R}^-} - a[b]_{\mathbb{R}^-} - b[a]_{\mathbb{R}^-} \ge ([a]_{\mathbb{R}^-} - [b]_{\mathbb{R}^-})^2$$

for every $a, b \in \mathbb{R}$. Using the latter inequality with $(a, b) = (P_{1,\gamma}^n(\boldsymbol{u}), P_{1,\gamma}^n(\boldsymbol{u}_h))$ for the first term in (4.28) and the Cauchy-Schwarz inequality for the second one, we have

$$\begin{aligned}
\mathfrak{T}_2 \le &-\left\|\gamma^{-1/2}\left(\left[P_{1,\gamma}^n(\boldsymbol{u})\right]_{\mathbb{R}^-} - \left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}\right)\right\|_{\Gamma_C}^2 \\
&+ \left\|\gamma^{-1/2}\left(\left[P_{1,\gamma}^n(\boldsymbol{u})\right]_{\mathbb{R}^-} - \left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}\right)\right\|_{\Gamma_C} \left\|\gamma^{-1/2}\sigma^n(\boldsymbol{u} - \boldsymbol{u}_h)\right\|_{\Gamma_C}.
\end{aligned}$$

We continue using the generalized Young inequality $ab \le a^2 + b^2/4$ for the second term followed by the saturation assumption (4.17) to write:

$$\mathfrak{T}_2 \le \frac{1}{4\gamma_0} \left\| \left(\frac{\gamma_0}{\gamma}\right)^{1/2} \sigma^n(\boldsymbol{u} - \boldsymbol{u}_h) \right\|_{\Gamma_C}^2 \lesssim \frac{1}{4\gamma_0} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega}^2 \le \frac{1}{4\gamma_0\alpha} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}}^2 . \tag{4.29}$$

Combining (4.22), (4.26), and (4.29) we finally get, for a suitable real number $C > 0$,

$$\left(\alpha^{1/2} - \frac{C}{4\gamma_0\alpha}\right) \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}}^2 \le C \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}} |||\mathcal{R}(\boldsymbol{u}_h)|||_*$$

and, taking γ_0 sufficiently large,

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}} \lesssim |||\mathcal{R}(\boldsymbol{u}_h)|||_*,$$

thus concluding the proof of (4.18). □

Theorem 4.11 (Control of the dual norm of the residual)**.** *Assume that the solution $\boldsymbol{u}$ of the continuous problem (4.2) belongs to $\boldsymbol{H}^{\frac{3}{2}+\nu}(\Omega)$ for some $\nu > 0$, and let $\boldsymbol{u}_h \in \boldsymbol{V}_h$ be the solution of the discrete problem (4.7). Then,*

$$|||\mathcal{R}(\boldsymbol{u}_h)|||_* \le (d\lambda + 4\mu)^{1/2} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}} + \left(\sum_{F \in \mathcal{F}_h^C} h_F \left\| \sigma^n(\boldsymbol{u}) - \left[P_{1,\gamma}^n(\boldsymbol{u}_h) \right]_{\mathbb{R}^-} \right\|_F^2 \right)^{1/2} . \tag{4.30}$$

Moreover, if the saturation assumption (4.17) holds, then

$$|||\mathcal{R}(\boldsymbol{u}_h)|||_* \lesssim \left[(d\lambda + 4\mu)^{1/2} + \alpha^{-1/2} \right] \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}} + \gamma_0 \left(\sum_{F \in \mathcal{F}_h^C} \frac{1}{h_F} \|\boldsymbol{u} - \boldsymbol{u}_h\|_F^2 \right)^{1/2} . \tag{4.31}$$

Proof. 1) Proof of (4.30). By definition (4.8) of the residual together with (4.2) (valid almost everywhere), and Green's formula, it holds: For any $\boldsymbol{v} \in \boldsymbol{H}_D^1(\Omega)$,

$$\begin{aligned} \langle \mathcal{R}(\boldsymbol{u}_h), \boldsymbol{v} \rangle &= (\boldsymbol{f}, \boldsymbol{v}) + (\boldsymbol{g}_N, \boldsymbol{v})_{\Gamma_N} - (\boldsymbol{\sigma}(\boldsymbol{u}_h), \boldsymbol{\varepsilon}(\boldsymbol{v})) + \left(\left[P_{1,\gamma}^n(\boldsymbol{u}_h) \right]_{\mathbb{R}^-}, v^n \right)_{\Gamma_C} \\ &= (\boldsymbol{\sigma}(\boldsymbol{u} - \boldsymbol{u}_h), \boldsymbol{\varepsilon}(\boldsymbol{v})) - \left(\sigma^n(\boldsymbol{u}) - \left[P_{1,\gamma}^n(\boldsymbol{u}_h) \right]_{\mathbb{R}^-}, v^n \right)_{\Gamma_C} . \end{aligned}$$

Then, using the symmetry of the Cauchy stress tensor $\boldsymbol{\sigma}(\boldsymbol{u} - \boldsymbol{u}_h)$, the Cauchy-Schwarz inequality, and the definition (4.9) of the norm $|||\boldsymbol{v}|||$, and additionally observing that

$$\|\boldsymbol{\sigma}(\boldsymbol{u} - \boldsymbol{u}_h)\| \le (d\lambda + 4\mu)^{1/2} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{en}} ,$$

we have

$$\begin{aligned}\langle \mathcal{R}(\boldsymbol{u}_h), \boldsymbol{v}\rangle &\le \|\boldsymbol{\sigma}(\boldsymbol{u}-\boldsymbol{u}_h)\|\,\|\boldsymbol{\nabla}\boldsymbol{v}\| + \sum_{F\in\mathcal{F}_h^C}\left\|\sigma^n(\boldsymbol{u}) - \left[P^n_{1,\gamma}(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}\right\|_F \|v^n\|_F \\ &\le (d\lambda+4\mu)^{1/2}\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{\mathrm{en}}\,\|\boldsymbol{\nabla}\boldsymbol{v}\| + \sum_{F\in\mathcal{F}_h^C} h_F^{1/2}\left\|\sigma^n(\boldsymbol{u}) - \left[P^n_{1,\gamma}(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}\right\|_F \frac{1}{h_F^{1/2}}\|\boldsymbol{v}\|_F \\ &\le \left[(d\lambda+4\mu)^{1/2}\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{\mathrm{en}} + \left(\sum_{F\in\mathcal{F}_h^C} h_F\left\|\sigma^n(\boldsymbol{u}) - \left[P^n_{1,\gamma}(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}\right\|_F^2\right)^{1/2}\right]|||\boldsymbol{v}|||.\end{aligned}$$

By definition (4.11) of the dual norm, this yields (4.30).

2) Proof of (4.31). Under the saturation assumption (4.17), starting from (4.30) and using (4.6), we obtain, for all $F\in\mathcal{F}_h^C$,

$$\begin{aligned}h_F\left\|\sigma^n(\boldsymbol{u}) - \left[P^n_{1,\gamma}(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}\right\|_F^2 &= h_F\left\|\left[P^n_{1,\gamma}(\boldsymbol{u})\right]_{\mathbb{R}^-} - \left[P^n_{1,\gamma}(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}\right\|_F^2 \\ &\le h_F\left\|P^n_{1,\gamma}(\boldsymbol{u}) - P^n_{1,\gamma}(\boldsymbol{u}_h)\right\|_F^2 \lesssim h_F\,\|\boldsymbol{\sigma}(\boldsymbol{u}-\boldsymbol{u}_h)\|_F^2 + h_F\,\|\gamma(\boldsymbol{u}-\boldsymbol{u}_h)\|_F^2,\end{aligned}$$

where we have applied the property $([a]_{\mathbb{R}^-} - [b]_{\mathbb{R}^-})^2 \le (a-b)^2$ valid for any $a,b\in\mathbb{R}$, with $(a,b) = (P^n_{1,\gamma}(\boldsymbol{u}), P^n_{1,\gamma}(\boldsymbol{u}_h))$ to pass to the second line and the triangle inequality to conclude. Then, using the saturation assumption (4.17) together with the ellipticity property (4.15) and the choice of γ, we obtain:

$$\begin{aligned}&\left(\sum_{F\in\mathcal{F}_h^C} h_F\left\|\sigma^n(\boldsymbol{u}) - \left[P^n_{1,\gamma}(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}\right\|_F^2\right)^{1/2} \\ &\quad\lesssim \left(\sum_{F\in\mathcal{F}_h^C} h_F\,\|\boldsymbol{\sigma}(\boldsymbol{u}-\boldsymbol{u}_h)\|_F^2 + \sum_{F\in\mathcal{F}_h^C} h_F\,\|\gamma(\boldsymbol{u}-\boldsymbol{u}_h)\|_F^2\right)^{1/2} \\ &\quad\le \left(\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_T^C} h_T\,\|\boldsymbol{\sigma}(\boldsymbol{u}-\boldsymbol{u}_h)\|_F^2\right)^{1/2} + \left(\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_T^C} h_F\left(\frac{\gamma_0}{h_T}\right)^2\|\boldsymbol{u}-\boldsymbol{u}_h\|_F^2\right)^{1/2} \\ &\quad\le \left\|\left(\frac{\gamma_0}{\gamma}\right)^{1/2}\boldsymbol{\sigma}(\boldsymbol{u}-\boldsymbol{u}_h)\right\|_{\Gamma_C} + \gamma_0\left(\sum_{F\in\mathcal{F}_h^C}\frac{1}{h_F}\|\boldsymbol{u}-\boldsymbol{u}_h\|_F^2\right)^{1/2} \\ &\quad\lesssim \alpha^{-1/2}\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{\mathrm{en}} + \gamma_0\left(\sum_{F\in\mathcal{F}_h^C}\frac{1}{h_F}\|\boldsymbol{u}-\boldsymbol{u}_h\|_F^2\right)^{1/2}.\end{aligned}$$

Combining this bound with (4.30), we obtain (4.31). □

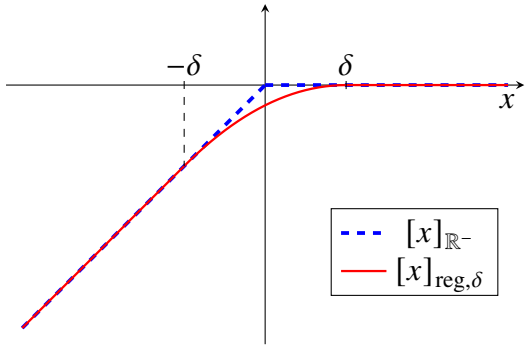


Figure 4.2 – Comparison between the projection operator $[x]_{\mathbb{R}^-}$ (*blue*) and the regularized operator $[x]_{\mathrm{reg},\delta}$ (*red*).

4.4 Identification of the error components

We consider the resolution of the (nonlinear) discrete problem (4.7) with an iterative method in which, at each iteration $k \geq 1$, the nonlinear term $\left[P^n_{1,\gamma}(\,\cdot\,)\right]_{\mathbb{R}^-}$ is replaced by a linear approximation $P^{k-1}_{\mathrm{lin}}(\,\cdot\,)$. A new approximation of the discrete solution is then obtained solving the following problem: Find $\boldsymbol{u}_h^k \in \boldsymbol{V}_h$ such that

$$a(\boldsymbol{u}_h^k, \boldsymbol{v}_h) - \left(P^{k-1}_{\mathrm{lin}}(\boldsymbol{u}_h^k), v_h^n\right)_{\Gamma_C} = L(\boldsymbol{v}_h) \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_h. \tag{4.32}$$

The linearized operator $P^{k-1}_{\mathrm{lin}}(\,\cdot\,)$ is based on the following regularization of the projection $[\,\cdot\,]_{\mathbb{R}^-}$: Given a real number $\delta > 0$ (representing the amount of regularization),

$$[x]_{\mathrm{reg},\delta} := \begin{cases} x & \text{if } x \leq -\delta \\ -\dfrac{1}{4\delta}x^2 + \dfrac{1}{2}x - \dfrac{\delta}{4} & \text{if } |x| < \delta \\ 0 & \text{if } x \geq \delta. \end{cases}$$

Figure 4.2 shows the graphs of the projection operator $[\cdot]_{\mathbb{R}^-}$ and of the regularized operator $[\cdot]_{\mathrm{reg},\delta}$. Notice that they coincide for $|x| \geq \delta$ and $[\cdot]_{\mathrm{reg},\delta}$ belongs to $C^1(\mathbb{R})$ (but not to $C^2(\mathbb{R})$). The linearized operator $P^{k-1}_{\mathrm{lin}}(\,\cdot\,)$ is obtained setting, for any $\boldsymbol{w}_h \in \boldsymbol{V}_h$,

$$\begin{aligned} P^{k-1}_{\mathrm{lin},\delta}(\boldsymbol{w}_h) &:= \left[P^n_{1,\gamma}(\boldsymbol{u}_h^{k-1})\right]_{\mathrm{reg},\delta} + \left.\frac{\partial \left[P^n_{1,\gamma}(\boldsymbol{v})\right]_{\mathrm{reg},\delta}}{\partial \boldsymbol{v}}\right|_{\boldsymbol{v}=\boldsymbol{u}_h^{k-1}} \cdot (\boldsymbol{w}_h - \boldsymbol{u}_h^{k-1}) \\ &= \left[P^n_{1,\gamma}(\boldsymbol{u}_h^{k-1})\right]_{\mathrm{reg},\delta} + \left.\frac{\mathrm{d}\,[x]_{\mathrm{reg},\delta}}{\mathrm{d}x}\right|_{x=P^n_{1,\gamma}(\boldsymbol{u}_h^{k-1})} \left(P^n_{1,\gamma}(\boldsymbol{w}_h) - P^n_{1,\gamma}(\boldsymbol{u}_h^{k-1})\right). \end{aligned} \tag{4.33}$$

Here, we add the subscript δ to emphasize that the linear operator depends on the choice of this parameter. The refined error estimate presented in the following section enables an automatic tuning of δ. Besides this regularization technique, there are different options to solve the discrete problem (4.7), e.g., with semismooth Newton methods [112, 35]. However, in this work, the presence of regularization is not an issue since the corresponding parameter δ is automatically tuned by the adaptive Algorithm 4.1 discussed below.

4.4.1 A posteriori error estimate distinguishing the error components

We present in this section an error estimate for $\boldsymbol{u}_h^k$ which enables one to identify and separate the different components of the error. The estimate hinges on the following assumption:

Assumption 4.12 (Decomposition of the stress reconstruction)**.** *Let $\boldsymbol{\sigma}_h^k$ be an equilibrated stress reconstruction in the sense of Definition 4.4. Then, $\boldsymbol{\sigma}_h^k$ can be decomposed into three parts*

$$\boldsymbol{\sigma}_h^k = \boldsymbol{\sigma}_{h,\mathrm{dis}}^k + \boldsymbol{\sigma}_{h,\mathrm{reg}}^k + \boldsymbol{\sigma}_{h,\mathrm{lin}}^k, \tag{4.34}$$

where $\boldsymbol{\sigma}_{h,\mathrm{dis}}^k$ represents discretization*, $\boldsymbol{\sigma}_{h,\mathrm{reg}}^k$ represents* regularization*, and $\boldsymbol{\sigma}_{h,\mathrm{lin}}^k$ represents* linearization*.*

In Section 4.5 we will show how to obtain an equilibrated stress reconstruction which satisfies this assumption. Finally, we introduce the following local error estimators: For every element $T \in \mathcal{T}_h$,

$$\eta_{\mathrm{osc},T}^k \coloneqq \frac{h_T}{\pi}\|\boldsymbol{f} + \mathbf{div}\,\boldsymbol{\sigma}_h^k\|_T, \qquad \text{(oscillation)} \tag{4.35a}$$

$$\eta_{\mathrm{str},T}^k \coloneqq \|\boldsymbol{\sigma}_{h,\mathrm{dis}}^k - \boldsymbol{\sigma}(\boldsymbol{u}_h^k)\|_T, \qquad \text{(stress)} \tag{4.35b}$$

$$\eta_{\mathrm{reg1},T}^k \coloneqq \|\boldsymbol{\sigma}_{h,\mathrm{reg}}^k\|_T \quad \text{and} \quad \eta_{\mathrm{reg2},T}^k \coloneqq \sum_{F\in\mathcal{F}_T^C} h_F^{1/2}\|\sigma_{h,\mathrm{reg}}^{k,n}\|_F, \qquad \text{(regularization)} \tag{4.35c}$$

$$\eta_{\mathrm{lin1},T}^k \coloneqq \|\boldsymbol{\sigma}_{h,\mathrm{lin}}^k\|_T \quad \text{and} \quad \eta_{\mathrm{lin2},T}^k \coloneqq \sum_{F\in\mathcal{F}_T^C} h_F^{1/2}\|\sigma_{h,\mathrm{lin}}^{k,n}\|_F, \qquad \text{(linearization)} \tag{4.35d}$$

$$\eta_{\mathrm{Neu},T}^k \coloneqq \sum_{F\in\mathcal{F}_T^N} C_{t,T,F} h_F^{1/2}\|\boldsymbol{g}_N - \boldsymbol{\sigma}_h^k\boldsymbol{n}\|_F, \qquad \text{(Neumann)} \tag{4.35e}$$

$$\eta_{\mathrm{cnt},T}^k \coloneqq \sum_{F\in\mathcal{F}_T^C} h_F^{1/2}\left\|\left[P_{1,\gamma}^n(\boldsymbol{u}_h^k)\right]_{\mathbb{R}^-} - \sigma_{h,\mathrm{dis}}^{k,n}\right\|_F. \qquad \text{(contact)} \tag{4.35f}$$

The corresponding global error estimators are defined setting

$$\eta_\bullet^k \coloneqq \left(\sum_{T\in\mathcal{T}_h}\left(\eta_{\bullet,T}^k\right)^2\right)^{1/2}. \tag{4.36}$$

Theorem 4.13 (A posteriori error estimate distinguishing the error components)**.** *Let $\boldsymbol{u}_h^k \in \boldsymbol{V}_h$ be the solution of the linearized problem (4.32) with $P_{lin,\delta}(\,\cdot\,)$ defined by (4.33), and let $\mathcal{R}(\boldsymbol{u}_h^k)$ be the residual of $\boldsymbol{u}_h^k$ defined by (4.8). Then, under Assumption 4.12, it holds*

$$\begin{aligned}&|||\mathcal{R}(\boldsymbol{u}_h^k)|||_*\\ &\le \left[\sum_{T\in\mathcal{T}_h}\left(\left(\eta_{\mathrm{osc},T}^k + \eta_{\mathrm{str},T}^k + \eta_{\mathrm{reg1},T}^k + \eta_{\mathrm{lin1},T}^k + \eta_{\mathrm{Neu},T}^k\right)^2 + \left(\eta_{\mathrm{cnt},T}^k + \eta_{\mathrm{reg2},T}^k + \eta_{\mathrm{lin2},T}^k\right)^2\right)\right]^{1/2}\end{aligned} \tag{4.37}$$

and, as a result,

$$|||\mathcal{R}(\boldsymbol{u}_h^k)|||_* \le \left[\left(\eta_{\mathrm{osc}}^k + \eta_{\mathrm{str}}^k + \eta_{\mathrm{reg1}}^k + \eta_{\mathrm{lin1}}^k + \eta_{\mathrm{Neu}}^k\right)^2 + \left(\eta_{\mathrm{cnt}}^k + \eta_{\mathrm{reg2}}^k + \eta_{\mathrm{lin2}}^k\right)^2\right]^{1/2}. \tag{4.38}$$

Proof. Proceeding as in the proof of Theorem 4.5, we immediately get

$$
\begin{aligned}
&|||\mathcal{R}(\boldsymbol{u}_h^k)|||_* \\
&\le \left[\sum_{T\in\mathcal{T}_h}\left(\left(\eta_{\mathrm{osc},T}^k + \|\boldsymbol{\sigma}_h^k - \boldsymbol{\sigma}(\boldsymbol{u}_h^k)\|_T + \eta_{\mathrm{Neu},T}^k\right)^2 + \left(\sum_{F\in\mathcal{F}_T^C} h_F^{1/2}\left\|\left[P_{1,\gamma}^n(\boldsymbol{u}_h^k)\right]_{\mathbb{R}^-} - \sigma_h^{k,n}\right\|_F\right)^2\right)\right]^{1/2}.
\end{aligned}
$$

Decomposing $\boldsymbol{\sigma}_h^k$ according to (4.34) and using the triangle inequality, we arrive at (4.37). Finally, (4.38) is obtained from (4.37) applying twice the inequality $\sum_{T\in\mathcal{T}_h}\left(\sum_{i=1}^m a_{i,T}\right)^2 \le \left(\sum_{i=1}^m a_i\right)^2$ valid for all families of nonnegative real numbers $(a_{i,T})_{1\le i\le m, T\in\mathcal{T}_h}$ with $a_i := \left(\sum_{T\in\mathcal{T}_h} a_{i,T}^2\right)^{1/2}$ for all $1 \le i \le m$. □

4.4.2 Fully adaptive algorithm

We propose an adaptive algorithm based on the error estimators (4.35) and (4.36), and on the result of Theorem 4.13. Denote by $\gamma_{\mathrm{reg}}, \gamma_{\mathrm{lin}} \in (0,1)$ two user-dependent parameters that represent the relative magnitude of the regularization and linearization errors with respect to the total error. Moreover, we define the following local estimators:

$$
\eta_{\mathrm{reg},T}^k := \eta_{\mathrm{reg1},T}^k + \eta_{\mathrm{reg2},T}^k, \qquad \eta_{\mathrm{lin},T}^k := \eta_{\mathrm{lin1},T}^k + \eta_{\mathrm{lin2},T}^k.
$$

The corresponding global counterparts are given by (4.36) with $\bullet \in \{\mathrm{reg}, \mathrm{lin}\}$. With these estimators and the parameters γ_{reg}, γ_{lin}, we define stopping criteria for the regularization and linearization loops, respectively, so that both the parameter δ and the number of Newton iterations on every mesh refinement iteration will be fixed automatically by the adaptive algorithm. For all $T \in \mathcal{T}_h$, the total error estimator is given by

$$
\begin{aligned}
\eta_{\mathrm{tot},T}^k := \Big[&\left(\eta_{\mathrm{osc},T}^k + \eta_{\mathrm{str},T}^k + \eta_{\mathrm{reg1},T}^k + \eta_{\mathrm{lin1},T}^k + \eta_{\mathrm{Neu},T}^k\right)^2 \\
&+ \left(\eta_{\mathrm{cnt},T}^k + \eta_{\mathrm{reg2},T}^k + \eta_{\mathrm{lin2},T}^k\right)^2\Big]^{1/2}.
\end{aligned} \tag{4.39}
$$

Algorithm 4.1 Adaptive algorithm

1: **choose** an initial function $\boldsymbol{u}_h^0 \in \mathbf{V}_h$, $\delta > 0$, $\gamma_{\mathrm{reg}}, \gamma_{\mathrm{lin}} \in (0,1)$
2: **repeat** {mesh refinement loop}
3: **repeat** {regularization loop}
4: **set** $k = 0$
5: **repeat** {Newton linearization loop}
6: **set** $k = k + 1$
7: **setup** the operator $P_{\mathrm{lin},\delta}^{k-1}$ and the linear system
8: **compute** $\boldsymbol{u}_h^k$, $\boldsymbol{\sigma}_h^k$, and the local and global estimators
9: **until** $\eta_{\mathrm{lin}}^k \le \gamma_{\mathrm{lin}}(\eta_{\mathrm{osc}}^k + \eta_{\mathrm{str}}^k + \eta_{\mathrm{Neu}}^k + \eta_{\mathrm{cnt}}^k)$
10: **decrease** δ (e.g. $\delta = \delta/2$)
11: **until** $\eta_{\mathrm{reg}}^k \le \gamma_{\mathrm{reg}}(\eta_{\mathrm{osc}}^k + \eta_{\mathrm{str}}^k + \eta_{\mathrm{Neu}}^k + \eta_{\mathrm{cnt}}^k + \eta_{\mathrm{lin}}^k)$
12: **set** δ at its previous value (e.g. $\delta = 2\delta$)
13: **refine** the elements of the mesh where $\eta_{\mathrm{tot},T}^k$ is higher
14: **update** data
15: **until** $\eta_{\mathrm{tot},T}^k$ is distributed evenly over the mesh

Remark 4.14 (Local stopping criteria)**.** *The stopping criteria in Lines 9 and 11 can alternatively be enforced locally inside each element:*

$$\eta_{\mathrm{lin},T}^k \le \gamma_{\mathrm{lin},T}\big(\eta_{\mathrm{osc},T}^k + \eta_{\mathrm{str},T}^k + \eta_{\mathrm{Neu},T}^k + \eta_{\mathrm{cnt},T}^k\big) \qquad \forall T \in \mathcal{T}_h, \tag{4.40a}$$

$$\eta_{\mathrm{reg},T}^k \le \gamma_{\mathrm{reg},T}\big(\eta_{\mathrm{osc},T}^k + \eta_{\mathrm{str},T}^k + \eta_{\mathrm{Neu},T}^k + \eta_{\mathrm{cnt},T}^k + \eta_{\mathrm{lin},T}^k\big) \qquad \forall T \in \mathcal{T}_h, \tag{4.40b}$$

where the parameters $\gamma_{\mathrm{lin},T}, \gamma_{\mathrm{reg},T} \in (0,1)$ can possibly vary element by element; see, e.g., [78] and also the discussion in [44, Section 4.1].

4.5 Equilibrated stress reconstructions

We first show how to construct an equilibrated stress reconstruction $\boldsymbol{\sigma}_h$ that satisfies the conditions of Definition 4.4, then modify the construction to match Assumption 4.12.

4.5.1 Basic equilibrated stress reconstruction

Following the path of [16], we construct $\boldsymbol{\sigma}_h$ patchwise around the mesh vertices using Arnold–Falk–Winther mixed finite element spaces [6], which are based on a stress tensor constructed in the Brezzi–Douglas–Marini space (see, e.g., [19, Chapter 3] and [14, Chapter 2]), along with Lagrange multipliers that enforce a weak symmetry constraint. From this point on, $\mathcal{V}_h^D$ will denote the set collecting all mesh vertices which lie on some Dirichlet boundary face. Notice that $\mathcal{V}_h^D$ also contains the vertices lying at the intersection between Γ_D and $\Gamma_\bullet$, $\bullet \in \{N, C\}$.

For any element $T \in \mathcal{T}_h$, and any integer $q \ge 1$, we set

$$\boldsymbol{\Sigma}_T \coloneqq \mathbb{P}^q(T), \qquad \boldsymbol{U}_T \coloneqq \boldsymbol{\mathcal{P}}^{q-1}(T), \qquad \boldsymbol{\Lambda}_T \coloneqq \left\{\boldsymbol{\mu} \in \mathbb{P}^{q-1}(T) \ : \ \boldsymbol{\mu} = -\boldsymbol{\mu}^T\right\}.$$

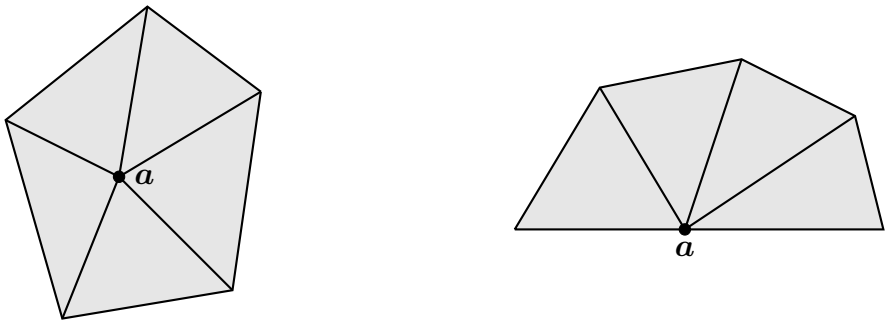


Figure 4.3 – Illustration of a patch $\omega_{\boldsymbol{a}}$ around an inner node $\boldsymbol{a} \in \mathcal{V}_h^i$ (*left*), and around a boundary node $\boldsymbol{a} \in \mathcal{V}_h^b$ (*right*).

At the global level, we define the following spaces

$$\begin{aligned}
\boldsymbol{\Sigma}_h &:= \{\boldsymbol{\tau}_h \in \mathbb{H}(\mathbf{div}, \Omega) \ : \ \boldsymbol{\tau}_h|_T \in \boldsymbol{\Sigma}_T \text{ for any } T \in \mathcal{T}_h\}, \\
\boldsymbol{U}_h &:= \left\{\boldsymbol{v}_h \in \boldsymbol{L}^2(\Omega) \ : \ \boldsymbol{v}_h|_T \in \boldsymbol{U}_T \text{ for any } T \in \mathcal{T}_h\right\}, \\
\boldsymbol{\Lambda}_h &:= \left\{\boldsymbol{\mu}_h \in \mathbb{L}^2(\Omega) \ : \ \boldsymbol{\mu}_h|_T \in \boldsymbol{\Lambda}_T \text{ for any } T \in \mathcal{T}_h\right\}.
\end{aligned}$$

Notice that $\boldsymbol{\Sigma}_h \subset \mathbb{H}(\mathbf{div}, \Omega)$ implies that its elements have continuous normal components across interfaces [39, Lemma 1.17]. Now, let $q = p$, let $\boldsymbol{u}_h$ be the solution of (4.7), and fix a mesh vertex $\boldsymbol{a}$. We denote by $\omega_{\boldsymbol{a}}$ the patch around the node $\boldsymbol{a}$, see Figure 4.3, by $\boldsymbol{n}_{\omega_{\boldsymbol{a}}}$ the normal unit outward vector on its boundary $\partial\omega_{\boldsymbol{a}}$ and by $\psi_{\boldsymbol{a}}$ the hat function associated with $\boldsymbol{a}$. On any patch $\omega_{\boldsymbol{a}}$ we then define the following spaces:

$$\boldsymbol{\Sigma}_h^{\boldsymbol{a}} := \begin{cases} \left\{\boldsymbol{\tau}_h \in \boldsymbol{\Sigma}_h(\omega_{\boldsymbol{a}}) \ : \ \boldsymbol{\tau}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \mathbf{0} \text{ on } \partial\omega_{\boldsymbol{a}} \setminus \Gamma_D\right\} & \text{if } \boldsymbol{a} \in \mathcal{V}_h^b \\ \left\{\boldsymbol{\tau}_h \in \boldsymbol{\Sigma}_h(\omega_{\boldsymbol{a}}) \ : \ \boldsymbol{\tau}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \mathbf{0} \text{ on } \partial\omega_{\boldsymbol{a}}\right\} & \text{otherwise,} \end{cases} \tag{4.41}$$

$$\boldsymbol{\Sigma}_{h,N,C}^{\boldsymbol{a}} := \begin{cases} \left\{\boldsymbol{\tau}_h \in \boldsymbol{\Sigma}_h(\omega_{\boldsymbol{a}}) \ : \ \boldsymbol{\tau}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \mathbf{0} \text{ on } \partial\omega_{\boldsymbol{a}} \setminus \partial\Omega, \right. \\ \quad \boldsymbol{\tau}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}}(\psi_{\boldsymbol{a}} \boldsymbol{g}_N) \text{ on } \partial\omega_{\boldsymbol{a}} \cap \Gamma_N, \text{ and} \\ \quad \left. \boldsymbol{\tau}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}}\left(\psi_{\boldsymbol{a}} \left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-} \boldsymbol{n}\right) \text{ on } \partial\omega_{\boldsymbol{a}} \cap \Gamma_C\right\} & \text{if } \boldsymbol{a} \in \mathcal{V}_h^b \\ \boldsymbol{\Sigma}_h^{\boldsymbol{a}} & \text{otherwise,} \end{cases} \tag{4.42}$$

$$\boldsymbol{U}_h^{\boldsymbol{a}} := \begin{cases} \boldsymbol{U}_h(\omega_{\boldsymbol{a}}) & \text{if } \boldsymbol{a} \in \mathcal{V}_h^D \\ \left\{\boldsymbol{v}_h \in \boldsymbol{U}_h(\omega_{\boldsymbol{a}}) \ : \ (\boldsymbol{v}_h, \boldsymbol{z})_{\omega_{\boldsymbol{a}}} = 0 \text{ for any } \boldsymbol{z} \in \boldsymbol{RM}^d\right\} & \text{otherwise,} \end{cases} \tag{4.43}$$

$$\boldsymbol{\Lambda}_h^{\boldsymbol{a}} := \boldsymbol{\Lambda}_h(\omega_{\boldsymbol{a}}). \tag{4.44}$$

Above, $\boldsymbol{\Sigma}_h(\omega_{\boldsymbol{a}})$, $\boldsymbol{U}_h(\omega_{\boldsymbol{a}})$, and $\boldsymbol{\Lambda}_h(\omega_{\boldsymbol{a}})$ denote the restrictions of the spaces $\boldsymbol{\Sigma}_h$, $\boldsymbol{U}_h$ and $\boldsymbol{\Lambda}_h$ to the subdomain $\omega_{\boldsymbol{a}}$, respectively. Moreover, $\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}$ is the space of normal traces on the patch boundary $\partial\omega_{\boldsymbol{a}}$ of elements in $\boldsymbol{\Sigma}_h(\omega_{\boldsymbol{a}})$, i.e., it is the space of vector-valued broken polynomials of total degree $\leq p$ on the set of boundary faces of the patch, while $\boldsymbol{RM}^d$ is space of rigid-body motions, i.e., $\boldsymbol{RM}^2 := \left\{\boldsymbol{b} + c(x_2, -x_1)^\top \ : \ \boldsymbol{b} \in \mathbb{R}^2, c \in \mathbb{R}\right\}$ and $\boldsymbol{RM}^3 := \left\{\boldsymbol{b} + \boldsymbol{c} \times \boldsymbol{x} \ : \ \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3\right\}$.

Remark 4.15 (Boundary condition for the reconstruction on internal vertices)**.** *In the definition* (4.41) *of* $\boldsymbol{\Sigma}_h^{\boldsymbol{a}}$*, we distinguish between boundary and internal vertices in order to ensure, in the case* $\boldsymbol{a} \in \mathcal{V}_h^i$*, zero normal components on the whole boundary of the patch* $\omega_{\boldsymbol{a}}$ *even if* $|\partial\omega_{\boldsymbol{a}} \cap \Gamma_D| > 0$*.*

Construction 4.16 (Basic equilibrated stress reconstruction)**.** *Let, for any vertex* $\boldsymbol{a} \in \mathcal{V}_h$, $(\boldsymbol{\sigma}_h^{\boldsymbol{a}}, \boldsymbol{r}_h^{\boldsymbol{a}}, \boldsymbol{\lambda}_h^{\boldsymbol{a}}) \in \boldsymbol{\Sigma}_{h,N,C}^{\boldsymbol{a}} \times \boldsymbol{U}_h^{\boldsymbol{a}} \times \boldsymbol{\Lambda}_h^{\boldsymbol{a}}$ *be the solution to the following problem:*

$$(\boldsymbol{\sigma}_h^{\boldsymbol{a}}, \boldsymbol{\tau}_h)_{\omega_{\boldsymbol{a}}} + (\boldsymbol{r}_h^{\boldsymbol{a}}, \mathbf{div}\, \boldsymbol{\tau}_h)_{\omega_{\boldsymbol{a}}} + (\boldsymbol{\lambda}_h^{\boldsymbol{a}}, \boldsymbol{\tau}_h)_{\omega_{\boldsymbol{a}}} = (\psi_{\boldsymbol{a}} \boldsymbol{\sigma}(\boldsymbol{u}_h), \boldsymbol{\tau}_h)_{\omega_{\boldsymbol{a}}} \tag{4.45a}$$

$$(\mathbf{div}\, \boldsymbol{\sigma}_h^{\boldsymbol{a}}, \boldsymbol{v}_h)_{\omega_{\boldsymbol{a}}} = (-\psi_{\boldsymbol{a}} \boldsymbol{f} + \boldsymbol{\sigma}(\boldsymbol{u}_h) \boldsymbol{\nabla} \psi_{\boldsymbol{a}}, \boldsymbol{v}_h)_{\omega_{\boldsymbol{a}}} \tag{4.45b}$$

$$(\boldsymbol{\sigma}_h^{\boldsymbol{a}}, \boldsymbol{\mu}_h)_{\omega_{\boldsymbol{a}}} = 0 \tag{4.45c}$$

for all $(\boldsymbol{\tau}_h, \boldsymbol{v}_h, \boldsymbol{\mu}_h) \in \boldsymbol{\Sigma}_h^{\boldsymbol{a}} \times \boldsymbol{U}_h^{\boldsymbol{a}} \times \boldsymbol{\Lambda}_h^{\boldsymbol{a}}$. *Extending* $\boldsymbol{\sigma}_h^{\boldsymbol{a}}$ *by zero outside the patch* $\omega_{\boldsymbol{a}}$*, we set* $\boldsymbol{\sigma}_h \coloneqq \sum_{\boldsymbol{a} \in \mathcal{V}_h} \boldsymbol{\sigma}_h^{\boldsymbol{a}}$.

By definition of the space $\boldsymbol{\Sigma}_{h,N,C}^{\boldsymbol{a}}$, a homogeneous Neumann boundary condition is enforced on the whole boundary of $\omega_{\boldsymbol{a}}$ for interior vertices and on $\partial\omega_{\boldsymbol{a}} \setminus \partial\Omega$ for boundary vertices. In particular, for boundary vertices in $\mathcal{V}_h^b \setminus \mathcal{V}_h^D$, a possibly non homogeneous Neumann boundary condition is enforced on the boundary faces of the patch. Therefore, when $\boldsymbol{a} \in \mathcal{V}_h^i$ or $\boldsymbol{a} \in \mathcal{V}_h^b \setminus \mathcal{V}_h^D$, the right hand side of (4.45b) has to verify the following Neumann compatibility condition:

$$\begin{aligned}(-\psi_{\boldsymbol{a}} \boldsymbol{f} + \boldsymbol{\sigma}(\boldsymbol{u}_h) \boldsymbol{\nabla} \psi_{\boldsymbol{a}}, \boldsymbol{z})_{\omega_{\boldsymbol{a}}} \\ = \left(\Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} (\psi_{\boldsymbol{a}} \boldsymbol{g}_N), \boldsymbol{z}\right)_{\partial\omega_{\boldsymbol{a}} \cap \Gamma_N} + \left(\Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} \left(\psi_{\boldsymbol{a}} \left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-} \boldsymbol{n}\right), \boldsymbol{z}\right)_{\partial\omega_{\boldsymbol{a}} \cap \Gamma_C}\end{aligned} \tag{4.46}$$

for any $\boldsymbol{z} \in \boldsymbol{RM}^d$. Fixing a rigid-body motion $\boldsymbol{z}$, it is possible to check that (4.46) holds by taking $\psi_{\boldsymbol{a}} \boldsymbol{z}$ as test function in (4.7). The following Lemma lists the main properties of the tensor $\boldsymbol{\sigma}_h$ resulting from Construction 4.16. In particular, it shows that $\boldsymbol{\sigma}_h$ satisfies all the conditions of Definition 4.4, i.e., it is an equilibrated stress reconstruction.

Lemma 4.17 (Properties of $\boldsymbol{\sigma}_h$)**.** *Let* $\boldsymbol{\sigma}_h$ *be defined by Construction 4.16. Then, it holds*

1) $\boldsymbol{\sigma}_h \in \mathbb{H}(\mathbf{div}, \Omega)$*;*

2) *For every* $T \in \mathcal{T}_h$ *and every* $\boldsymbol{v}_T \in \boldsymbol{\mathcal{P}}^{p-1}(T)$, $(\mathbf{div}\, \boldsymbol{\sigma}_h + \boldsymbol{f}, \boldsymbol{v}_T)_T = 0$*;*

3) *For every* $F \in \mathcal{F}_h^N$ *and every* $\boldsymbol{v}_F \in \boldsymbol{\mathcal{P}}^p(F)$, $(\boldsymbol{\sigma}_h \boldsymbol{n}, \boldsymbol{v}_F)_F = (\boldsymbol{g}_N, \boldsymbol{v}_F)_F$*;*

4) *For every* $F \in \mathcal{F}_h^C$ *and every* $\boldsymbol{v}_F \in \boldsymbol{\mathcal{P}}^p(F)$,

$$(\boldsymbol{\sigma}_h \boldsymbol{n}, \boldsymbol{v}_F)_F = \left(\left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-} \boldsymbol{n}, \boldsymbol{v}_F\right)_F = \left(\left[P_{1,\gamma}^n(\boldsymbol{u}_h)\right]_{\mathbb{R}^-}, v_F^n\right)_F.$$

Proof. 1) By definition, $\boldsymbol{\sigma}_h^{\boldsymbol{a}} \in \mathbb{H}(\mathbf{div}, \omega_{\boldsymbol{a}})$ for any $\boldsymbol{a} \in \mathcal{V}_h$. Due to the no-flux boundary condition on internal faces enforced in the local problem on $\omega_{\boldsymbol{a}}$, the extension of $\boldsymbol{\sigma}_h^{\boldsymbol{a}}$ by zero outside the patch is in $\mathbb{H}(\mathbf{div}, \Omega)$ and, as a consequence, $\boldsymbol{\sigma}_h \in \mathbb{H}(\mathbf{div}, \Omega)$.

2) First, we check that, for any $\boldsymbol{a} \in \mathcal{V}_h$, equation (4.45b) holds for every $\boldsymbol{v}_h \in \boldsymbol{U}_h(\omega_{\boldsymbol{a}})$. If $\boldsymbol{a} \in \mathcal{V}_h^D$, this is trivial since $\boldsymbol{U}_h^{\boldsymbol{a}} = \boldsymbol{U}_h(\omega_{\boldsymbol{a}})$. If, on the other hand, $\boldsymbol{a} \in \mathcal{V}_h \setminus \mathcal{V}_h^D$, it is sufficient to use the fact that $\boldsymbol{U}_h^{\boldsymbol{a}} = (\boldsymbol{RM}^d)^{\perp}$ (with orthogonal taken with respect to the $\boldsymbol{L}^2(\omega_{\boldsymbol{a}})$-product) along with the Green formula, the definition (4.42) of $\boldsymbol{\Sigma}_h^{\boldsymbol{a}}$, the Neumann compatibility condition (4.46), and (4.45c).

Now, fix $T \in \mathcal{T}_h$ and let $\boldsymbol{v}_T \in \boldsymbol{\mathcal{P}}^{p-1}(T)$. Extending $\boldsymbol{v}_T$ by zero outside of T, we have $\boldsymbol{v}_T \in \boldsymbol{U}_h(\omega_{\boldsymbol{a}})$ for all $\boldsymbol{a} \in \mathcal{V}_T$. Indeed, by definition, $\boldsymbol{U}_h(\omega_{\boldsymbol{a}})$ is composed by piecewise polynomials of degree at most $p-1$ that can be chosen independently inside each element of the patch. Summing (4.45b) over $\boldsymbol{a} \in \mathcal{V}_T$ we obtain:

$$0 = \sum_{\boldsymbol{a}\in\mathcal{V}_T} \left[(\mathbf{div}\, \boldsymbol{\sigma}_h^{\boldsymbol{a}}, \boldsymbol{v}_T)_{\omega_{\boldsymbol{a}}} + (\psi_{\boldsymbol{a}} \boldsymbol{f}, \boldsymbol{v}_T)_{\omega_{\boldsymbol{a}}} - (\boldsymbol{\sigma}(\boldsymbol{u}_h)\boldsymbol{\nabla}\psi_{\boldsymbol{a}}, \boldsymbol{v}_T)_{\omega_{\boldsymbol{a}}} \right] = (\mathbf{div}\, \boldsymbol{\sigma}_h + \boldsymbol{f}, \boldsymbol{v}_T)_T.$$

Here, we have used the fact that $\boldsymbol{\sigma}_h|_T = \sum_{\boldsymbol{a}\in\mathcal{V}_T} \boldsymbol{\sigma}_h^{\boldsymbol{a}}|_T$ and $\sum_{\boldsymbol{a}\in\mathcal{V}_T} \psi_{\boldsymbol{a}} = 1$ over T (so that, in particular, $\sum_{\boldsymbol{a}\in\mathcal{V}_T} \boldsymbol{\nabla}\psi_{\boldsymbol{a}} \equiv 0$).

3,4) We only detail the proof of 3) as that of 4) is similar. Let $F \in \mathcal{F}_h^N$ and let $\boldsymbol{v}_F$ be a polynomial defined on F from the discrete normal trace space $(\boldsymbol{\Sigma}_h \boldsymbol{n})|_F$, i.e., a polynomial of total degree at most p. Then, by the definition of $\boldsymbol{\Sigma}_{h,N,C}^{\boldsymbol{a}}$ (4.42),

$$(\boldsymbol{\sigma}_h \boldsymbol{n}, \boldsymbol{v}_F)_F = \sum_{\boldsymbol{a}\in\mathcal{V}_F} (\boldsymbol{\sigma}_h^{\boldsymbol{a}} \boldsymbol{n}, \boldsymbol{v}_F)_F = \sum_{\boldsymbol{a}\in\mathcal{V}_F} (\psi_{\boldsymbol{a}} \boldsymbol{g}_N, \boldsymbol{v}_F)_F = (\boldsymbol{g}_N, \boldsymbol{v}_F)_F.$$

□

4.5.2 Stress reconstruction distinguishing the error components

In Construction 4.16 we used the solution $\boldsymbol{u}_h$ of the nonlinear problem (4.7) to reconstruct an equilibrated stress $\boldsymbol{\sigma}_h$. However, as argued in Section 4.4, in practice we only dispose of an approximated solution obtained by means of a linearization method. Let $k \geq 1$ be an integer and let $\boldsymbol{u}_h^k$ be the solution of the linearized problem (4.32) with operator $P_{\mathrm{lin},\delta}(\,\cdot\,)$ defined by (4.33). Then, for any boundary vertex $\boldsymbol{a} \in \mathcal{V}_h^b$, we set

$$\begin{aligned}
\boldsymbol{\Sigma}_{h,N,C,\mathrm{dis}}^{\boldsymbol{a},k} \coloneqq \{\boldsymbol{\tau}_h \in \boldsymbol{\Sigma}_h(\omega_{\boldsymbol{a}}) \;:\; & \boldsymbol{\tau}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \boldsymbol{0} \text{ on } \partial\omega_{\boldsymbol{a}} \setminus \partial\Omega, \\
& \boldsymbol{\tau}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} (\psi_{\boldsymbol{a}} \boldsymbol{g}_N) \text{ on } \partial\omega_{\boldsymbol{a}} \cap \Gamma_N \text{ and} \\
& \boldsymbol{\tau}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} \left(\psi_{\boldsymbol{a}} \left[P_{1,\gamma}^n(\boldsymbol{u}_h^k) \right]_{\mathbb{R}^-} \boldsymbol{n} \right) \text{ on } \partial\omega_{\boldsymbol{a}} \cap \Gamma_C\}, \\
\boldsymbol{\Sigma}_{h,N,C,\mathrm{reg}}^{\boldsymbol{a},k} \coloneqq \{\boldsymbol{\tau}_h \in \boldsymbol{\Sigma}_h(\omega_{\boldsymbol{a}}) \;:\; & \boldsymbol{\tau}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \boldsymbol{0} \text{ on } \partial\omega_{\boldsymbol{a}} \setminus \partial\Omega \text{ and on } \partial\omega_{\boldsymbol{a}} \cap \Gamma_N, \text{ and} \\
& \boldsymbol{\tau}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} \left(\psi_{\boldsymbol{a}} \left(\left[P_{1,\gamma}^n(\boldsymbol{u}_h^k) \right]_{\mathrm{reg},\delta} - \left[P_{1,\gamma}^n(\boldsymbol{u}_h^k) \right]_{\mathbb{R}^-} \right) \boldsymbol{n} \right) \text{ on } \partial\omega_{\boldsymbol{a}} \cap \Gamma_C\}, \\
\boldsymbol{\Sigma}_{h,N,C,\mathrm{lin}}^{\boldsymbol{a},k} \coloneqq \{\boldsymbol{\tau}_h \in \boldsymbol{\Sigma}_h(\omega_{\boldsymbol{a}}) \;:\; & \boldsymbol{\tau}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \boldsymbol{0} \text{ on } \partial\omega_{\boldsymbol{a}} \setminus \partial\Omega \text{ and on } \partial\omega_{\boldsymbol{a}} \cap \Gamma_N, \text{ and} \\
& \boldsymbol{\tau}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} \left(\psi_{\boldsymbol{a}} \left(P_{\mathrm{lin}}^{k-1}(\boldsymbol{u}_h^k) - \left[P_{1,\gamma}^n(\boldsymbol{u}_h^k) \right]_{\mathrm{reg},\delta} \right) \boldsymbol{n} \right) \text{ on } \partial\omega_{\boldsymbol{a}} \cap \Gamma_C\},
\end{aligned}$$

and, for any internal vertex $\boldsymbol{a} \in \mathcal{V}_h^i$, $\boldsymbol{\Sigma}_{h,N,C,\bullet}^{\boldsymbol{a},k} \coloneqq \boldsymbol{\Sigma}_h^{\boldsymbol{a}}$ (see (4.41)) for $\bullet \in \{\mathrm{dis}, \mathrm{reg}, \mathrm{lin}\}$. Moreover, let $\boldsymbol{y}^k, \tilde{\boldsymbol{y}}^k \in \boldsymbol{RM}^d$ be such that, for all $\boldsymbol{z} \in \boldsymbol{RM}^d$,

$$\begin{aligned}
(\boldsymbol{y}^k, \boldsymbol{z})_{\omega_{\boldsymbol{a}}} = {} & (-\psi_{\boldsymbol{a}} \boldsymbol{f} + \boldsymbol{\sigma}(\boldsymbol{u}_h^k)\boldsymbol{\nabla}\psi_{\boldsymbol{a}}, \boldsymbol{z})_{\omega_{\boldsymbol{a}}} - \left(\Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} (\psi_{\boldsymbol{a}} \boldsymbol{g}_N), \boldsymbol{z} \right)_{\partial\omega_{\boldsymbol{a}} \cap \Gamma_N} \\
& - \left(\Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} \left(\psi_{\boldsymbol{a}} \left[P_{1,\gamma}^n(\boldsymbol{u}_h^k) \right]_{\mathbb{R}^-} \boldsymbol{n} \right), \boldsymbol{z} \right)_{\partial\omega_{\boldsymbol{a}} \cap \Gamma_C}, \\
(\tilde{\boldsymbol{y}}^k, \boldsymbol{z})_{\omega_{\boldsymbol{a}}} = {} & \left(\Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} \left(\psi_{\boldsymbol{a}} \left(\left[P_{1,\gamma}^n(\boldsymbol{u}_h^k) \right]_{\mathbb{R}^-} - \left[P_{1,\gamma}^n(\boldsymbol{u}_h^k) \right]_{\mathrm{reg},\delta} \right) \boldsymbol{n} \right), \boldsymbol{z} \right)_{\partial\omega_{\boldsymbol{a}} \cap \Gamma_C}
\end{aligned}$$

if $\boldsymbol{a} \in \mathcal{V}_h^b$, and $\boldsymbol{y}^k = \tilde{\boldsymbol{y}}^k = \boldsymbol{0}$ if $\boldsymbol{a} \in \mathcal{V}_h^i$.

Construction 4.18 (Equilibrated stress reconstruction distinguishing the error components)**.** *Let, for* $\bullet \in \{\mathrm{dis}, \mathrm{reg}, \mathrm{lin}\}$ *and any vertex* $\boldsymbol{a} \in \mathcal{V}_h$, $(\boldsymbol{\sigma}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{r}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{\lambda}_{h,\bullet}^{\boldsymbol{a},k}) \in \boldsymbol{\Sigma}_{h,N,C,\bullet}^{\boldsymbol{a},k} \times \boldsymbol{U}_h^{\boldsymbol{a}} \times \boldsymbol{\Lambda}_h^{\boldsymbol{a}}$ *be the solution to the following problem:*

$$
\begin{aligned}
(\boldsymbol{\sigma}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{\tau}_h)_{\omega_{\boldsymbol{a}}} + (\boldsymbol{r}_{h,\bullet}^{\boldsymbol{a},k}, \mathbf{div}\, \boldsymbol{\tau}_h)_{\omega_{\boldsymbol{a}}} + (\boldsymbol{\lambda}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{\tau}_h)_{\omega_{\boldsymbol{a}}} &= (\boldsymbol{\tau}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{\tau}_h)_{\omega_{\boldsymbol{a}}} && \forall \boldsymbol{\tau}_h \in \boldsymbol{\Sigma}_h^{\boldsymbol{a}},\\
(\mathbf{div}\, \boldsymbol{\sigma}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{v}_h)_{\omega_{\boldsymbol{a}}} &= (\boldsymbol{v}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{v}_h)_{\omega_{\boldsymbol{a}}} && \forall \boldsymbol{v}_h \in \boldsymbol{U}_h^{\boldsymbol{a}},\\
(\boldsymbol{\sigma}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{\mu}_h)_{\omega_{\boldsymbol{a}}} &= 0 && \forall \boldsymbol{\mu}_h \in \boldsymbol{\Lambda}_h^{\boldsymbol{a}},
\end{aligned}
$$

where

$$
\boldsymbol{\tau}_{h,\bullet}^{\boldsymbol{a},k} := \begin{cases} \psi_{\boldsymbol{a}} \boldsymbol{\sigma}(\boldsymbol{u}_h^k) & \text{if } \bullet = \mathrm{dis}, \\ 0 & \text{if } \bullet \in \{\mathrm{reg}, \mathrm{lin}\}, \end{cases} \qquad \boldsymbol{v}_{h,\bullet}^{\boldsymbol{a},k} := \begin{cases} -\psi_{\boldsymbol{a}} \boldsymbol{f} + \boldsymbol{\sigma}(\boldsymbol{u}_h^k) \boldsymbol{\nabla} \psi_{\boldsymbol{a}} - \boldsymbol{y}^k & \text{if } \bullet = \mathrm{dis}, \\ -\tilde{\boldsymbol{y}}^k & \text{if } \bullet = \mathrm{reg}, \\ \boldsymbol{y}^k + \tilde{\boldsymbol{y}}^k & \text{if } \bullet = \mathrm{lin}. \end{cases}
$$

Extending $\boldsymbol{\sigma}_{h,\bullet}^{\boldsymbol{a},k}$ *by zero outside the patch* $\omega_{\boldsymbol{a}}$*, we set* $\boldsymbol{\sigma}_{h,\bullet}^k := \sum_{\boldsymbol{a} \in \mathcal{V}_h} \boldsymbol{\sigma}_{h,\bullet}^{\boldsymbol{a},k}$*, and we define* $\boldsymbol{\sigma}_h^k := \boldsymbol{\sigma}_{h,\mathrm{dis}}^k + \boldsymbol{\sigma}_{h,\mathrm{reg}}^k + \boldsymbol{\sigma}_{h,\mathrm{lin}}^k$.

By definition, $\boldsymbol{y}^k$ and $\tilde{\boldsymbol{y}}^k$ ensure that the forcing terms $\boldsymbol{v}_{h,\bullet}^{\boldsymbol{a},k}$ satisfy the following Neumann compatibility conditions for $\boldsymbol{a} \in \mathcal{V}_h^b \setminus \mathcal{V}_h^D$:

$$
\begin{aligned}
(\boldsymbol{v}_{h,\mathrm{dis}}^{\boldsymbol{a},k}, \boldsymbol{z})_{\omega_{\boldsymbol{a}}} &= \left(\Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} (\psi_{\boldsymbol{a}} \boldsymbol{g}_N), \boldsymbol{z}\right)_{\partial\omega_{\boldsymbol{a}} \cap \Gamma_N} + \left(\Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} \left(\psi_{\boldsymbol{a}} \left[P_{1,\gamma}^n(\boldsymbol{u}_h^k)\right]_{\mathbb{R}^-} \boldsymbol{n}\right), \boldsymbol{z}\right)_{\partial\omega_{\boldsymbol{a}} \cap \Gamma_C},\\
(\boldsymbol{v}_{h,\mathrm{reg}}^{\boldsymbol{a},k}, \boldsymbol{z})_{\omega_{\boldsymbol{a}}} &= \left(\Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} \left(\psi_{\boldsymbol{a}} \left(\left[P_{1,\gamma}^n(\boldsymbol{u}_h^k)\right]_{\mathrm{reg},\delta} - \left[P_{1,\gamma}^n(\boldsymbol{u}_h^k)\right]_{\mathbb{R}^-}\right) \boldsymbol{n}\right), \boldsymbol{z}\right)_{\partial\omega_{\boldsymbol{a}} \cap \Gamma_C},\\
(\boldsymbol{v}_{h,\mathrm{lin}}^{\boldsymbol{a},k}, \boldsymbol{z})_{\omega_{\boldsymbol{a}}} &= \left(\Pi_{\boldsymbol{\Sigma}_h \boldsymbol{n}_{\omega_{\boldsymbol{a}}}} \left(\psi_{\boldsymbol{a}} \left(P_{\mathrm{lin}}^{k-1}(\boldsymbol{u}_h^k) - \left[P_{1,\gamma}^n(\boldsymbol{u}_h^k)\right]_{\mathrm{reg},\delta}\right) \boldsymbol{n}\right), \boldsymbol{z}\right)_{\partial\omega_{\boldsymbol{a}} \cap \Gamma_C}
\end{aligned}
$$

for any $\boldsymbol{z} \in \boldsymbol{RM}^d$. The obtained tensor $\boldsymbol{\sigma}_h^k$ is an equilibrated stress reconstruction in the sense of Definition 4.4, and in particular it satisfies the properties stated by the following lemma whose proof is similar to that of Lemma 4.17 and is therefore omitted for the sake of conciseness.

Lemma 4.19 (Properties of $\boldsymbol{\sigma}_h^k$)**.** *Let* $\boldsymbol{\sigma}_h^k$ *be defined by Construction 4.18. Then*

1) $\boldsymbol{\sigma}_{h,\mathrm{dis}}^k, \boldsymbol{\sigma}_{h,\mathrm{reg}}^k, \boldsymbol{\sigma}_{h,\mathrm{lin}}^k, \boldsymbol{\sigma}_h^k \in \mathbb{H}(\mathbf{div}, \Omega)$;

2) *For every* $T \in \mathcal{T}_h$ *and every* $\boldsymbol{v}_T \in \boldsymbol{\mathcal{P}}^{p-1}(T)$, $(\mathbf{div}\, \boldsymbol{\sigma}_h^k + \boldsymbol{f}, \boldsymbol{v}_T)_T = 0$;

3) *For every* $F \in \mathcal{F}_h^N$ *and every* $\boldsymbol{v}_F \in \boldsymbol{\mathcal{P}}^p(F)$, $(\boldsymbol{\sigma}_h^k \boldsymbol{n}, \boldsymbol{v}_F)_F = (\boldsymbol{g}_N, \boldsymbol{v}_F)_F$;

4) *For every* $F \in \mathcal{F}_h^C$ *and every* $\boldsymbol{v}_F \in \boldsymbol{\mathcal{P}}^p(F)$,

$$
(\boldsymbol{\sigma}_{h,\mathrm{dis}}^k \boldsymbol{n}, \boldsymbol{v}_F)_F = \left(\left[P_{1,\gamma}^n(\boldsymbol{u}_h^k)\right]_{\mathbb{R}^-} \boldsymbol{n}, \boldsymbol{v}_F\right)_F,
$$

$$(\boldsymbol{\sigma}^k_{h,\mathrm{reg}}\boldsymbol{n},\boldsymbol{v}_F)_F = \left(\left(\left[P^n_{1,\gamma}(\boldsymbol{u}^k_h)\right]_{\mathrm{reg},\delta} - \left[P^n_{1,\gamma}(\boldsymbol{u}^k_h)\right]_{\mathbb{R}^-}\right)\boldsymbol{n},\boldsymbol{v}_F\right)_F,$$

and

$$(\boldsymbol{\sigma}^k_{h,\mathrm{lin}}\boldsymbol{n},\boldsymbol{v}_F)_F = \left(\left(P^{k-1}_{\mathrm{lin}}(\boldsymbol{u}^k_h) - \left[P^n_{1,\gamma}(\boldsymbol{u}^k_h)\right]_{\mathrm{reg},\delta}\right)\boldsymbol{n},\boldsymbol{v}_F\right)_F.$$

Remark 4.20 (Validity of Property 4. in Definition 4.4 and of Assumption 4.12)**.** *The fourth property of the previous lemma implies that* $(\boldsymbol{\sigma}^k_{h,\bullet}\boldsymbol{n})|_F$ *has the same direction as the normal vector* $\boldsymbol{n}$*, and, as a consequence,* $\boldsymbol{\sigma}^{k,\boldsymbol{t}}_{h,\bullet} = \boldsymbol{0}$ *on* $F \in \mathcal{F}^C_h$*, for* $\bullet \in \{\mathrm{dis},\mathrm{reg},\mathrm{lin}\}$*. Moreover, by definition,* $\boldsymbol{\sigma}^k_h$ *is the sum of three tensors representing discretization, regularization and linearization, respectively. Therefore,* $\boldsymbol{\sigma}^k_h$ *is an equilibrated stress reconstruction in the sense of Definition 4.4 that additionally satisfies Assumption 4.12.*

Remark 4.21 (Alternative expressions of local estimators)**.** *Thanks to Lemma 4.19, we can rewrite the oscillation (4.35a), Neumann (4.35e), and contact (4.35f) estimators as follows:*

$$\eta^k_{\mathrm{osc},T} = \frac{h_T}{\pi}\left\|\boldsymbol{f} - \boldsymbol{\Pi}^{p-1}_T\boldsymbol{f}\right\|_T, \qquad \eta^k_{\mathrm{Neu},T} = \sum_{F\in\mathcal{F}^C_T} C_{t,T,F}h^{1/2}_F\left\|\boldsymbol{g}_N - \boldsymbol{\Pi}^p_F\boldsymbol{g}_N\right\|_F,$$

$$\eta^k_{\mathrm{cnt},T} = \sum_{F\in\mathcal{F}^C_T} h^{1/2}_F\left\|\left[P^n_{1,\gamma}(\boldsymbol{u}^k_h)\right]_{\mathbb{R}^-} - \Pi^p_F\left[P^n_{1,\gamma}(\boldsymbol{u}^k_h)\right]_{\mathbb{R}^-}\right\|_F,$$

where $\boldsymbol{\Pi}^{p-1}_T$*,* $\boldsymbol{\Pi}^p_F$*, and* Π^p_F *denote the* L^2*-orthogonal projectors on the polynomial spaces* $\boldsymbol{\mathcal{P}}^{p-1}(T)$*,* $\boldsymbol{\mathcal{P}}^p(F)$*, and* $\mathcal{P}^p(F)$*, respectively.*

4.6 Numerical results

We present numerical cases to validate the a posteriori error estimate of Theorem 4.13 and show its use in the framework of the adaptive Algorithm 4.1. The simulations are performed with the open source finite element library FreeFem++ (see [68, 69] and also `https://freefem.org/`). The implementation was inspired by the work Zuqi Tang for the Laplace problem (`https://who.rocq.inria.fr/Zuqi.Tang/Freefem++/Laplace.html`). With FreeFem++ we can implement the discrete formulation of the problem (4.7) in a form which is very similar to the mathematical language, the patches for defining the local problems are obtained combining the hat function and a truncation command called `trunc`, and the local estimators are obtained by defining a discontinuous piecewise constant function on the mesh. In addition, we have implemented the adaptive refinement using a command of FreeFem++ that provides a conforming mesh refining only a subset of elements given by the user, combined with a trick in order to have at each refinement step a regular mesh. We will use the notion of local and global total estimator: $\eta_{\mathrm{tot},T}$ defined by (4.39) and

$$\eta_{\mathrm{tot}} := \left(\sum_{T\in\mathcal{T}_h}(\eta_{\mathrm{tot},T})^2\right)^{1/2}.$$

For the sake of brevity, above and throughout this section we omit the superscript k which identifies the step of the Newton method.

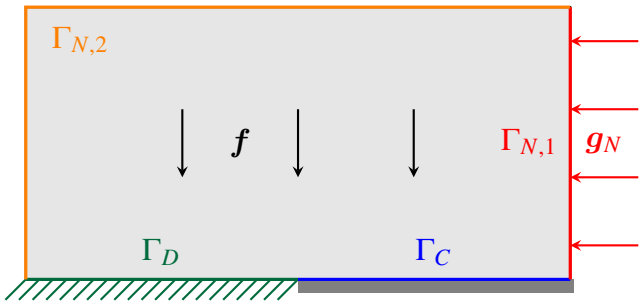
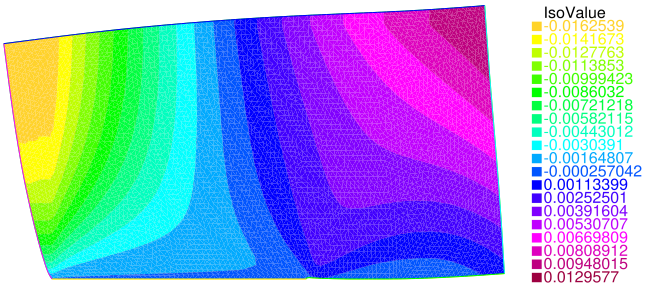
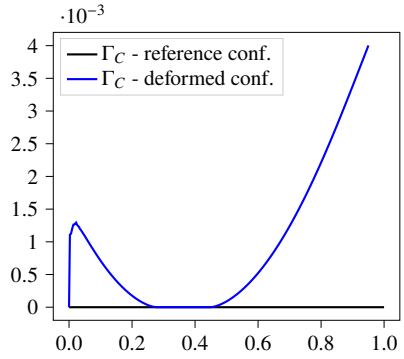


Figure 4.4 – Illustration of the rectangular domain with the subdivision of the boundary as $\partial\Omega = \Gamma_D \dot\cup (\Gamma_{N,1} \dot\cup \Gamma_{N,2}) \dot\cup \Gamma_C$.



(a) Vertical displacement in the deformed domain (amplification factor = 5).



(b) Displacement of the contact boundary.

Figure 4.5 – Vertical displacement in the deformed configuration (*left*), and representation of the contact boundary part Γ_C in the reference (*black*) and deformed (*blue*) configuration (*right*).

We consider a body that, in its reference configuration, occupies the rectangular domain $\Omega = (-1,1) \times (0,1)$ (see Figure 4.4), with mechanical parameters $E = 1$ and $\nu = 0.3$, corresponding to Lamé coefficients $\mu \approx 0.385$ and $\lambda \approx 0.577$. The body is subjected to a weight force $\boldsymbol{f} = (0, -0.01)$. Homogeneous Dirichlet boundary conditions are enforced on $\Gamma_D = (-1,0) \times \{0\}$, and the body in its undeformed configuration is in contact with a rigid horizontal interface on $\Gamma_C = (0,1) \times \{0\}$. The Nitsche parameter is $\gamma_0 = 100E$ (see [29]), whereas the regularization parameter which defines the operator $[\,\cdot\,]_{\mathrm{reg},\delta}$ is $\delta = E/100$. A pressure $\boldsymbol{g}_N = (-0.0275, 0)$ acts on the right side of the body $\Gamma_{N,1} = \{1\} \times (0,1)$, and the rest of the boundary is free, i.e., $\boldsymbol{g}_N = \boldsymbol{0}$ on $\Gamma_{N,2} = \{-1\} \times (0,1) \cup (-1,1) \times \{1\}$. Since a closed-form solution is not available for this configuration, we take as reference solution the function $\bar{\boldsymbol{u}}_h$ computed solving (4.7) with Lagrange $\mathcal{P}^2$ finite elements on a fine mesh ($h \approx 0.0084$). To compute the approximated solution $\boldsymbol{u}_h$, we use Lagrange $\mathcal{P}^1$ elements (while this choice is known to lock in the quasi-incompressible limit, it is admissible for the set of parameters considered here and is compatible with the use of the lowest-order mixed finite elements available in FreeFem++ to compute the equilibrated stress reconstructions). In the deformed configuration, the body is in contact with the rigid foundation in a non-empty interval $I_C \subset \Gamma_C$ which is approximately $(0.279, 0.447)$. Figure 4.5a shows the vertical displacement in the deformed domain with an amplification factor equal to 5. Moreover, in

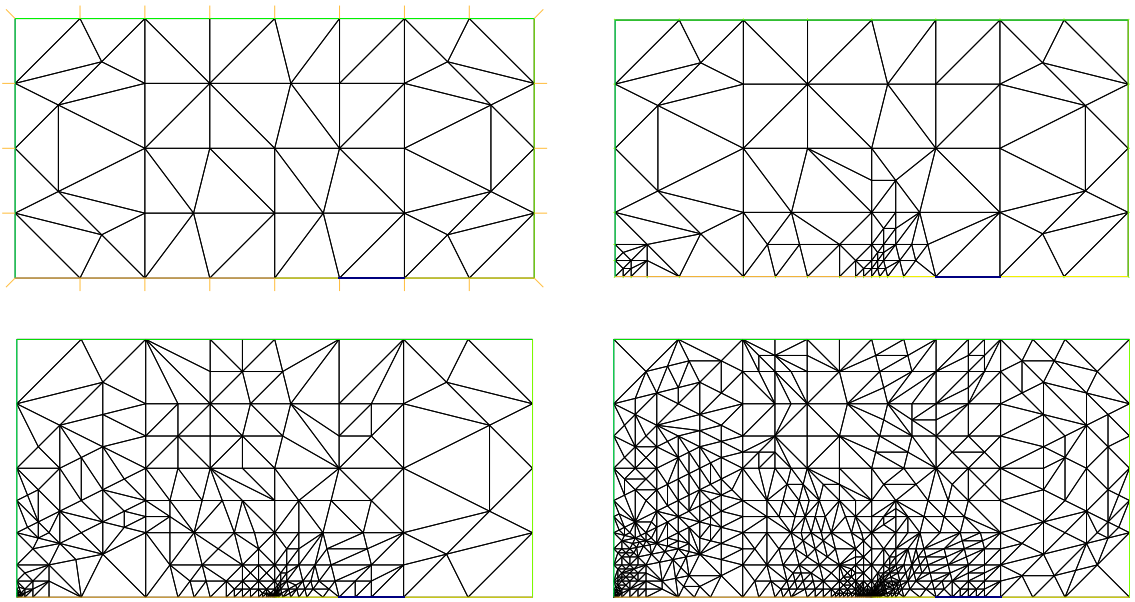


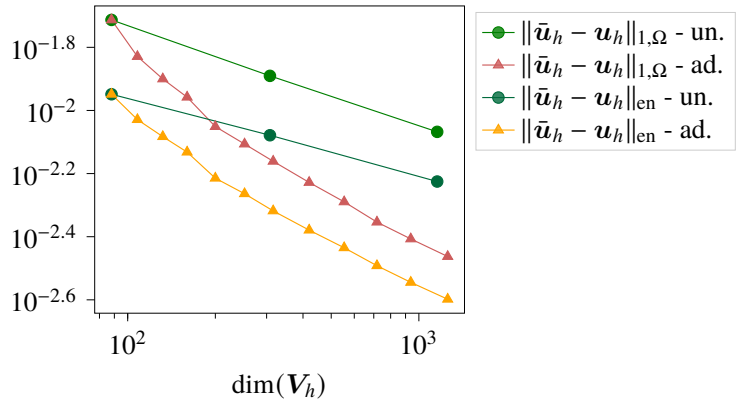
Figure 4.6 – Initial mesh and adaptively refined mesh after 3, 7 and 11 steps, respectively. The edges $F \subset \Gamma_C$ that include the contact interval I_C are highlighted in *blue*.

Figure 4.5b, which display the contact boundary part Γ_C in the reference domain (black) and in the deformed domain (blue), we can easily identify the contact interval I_C.

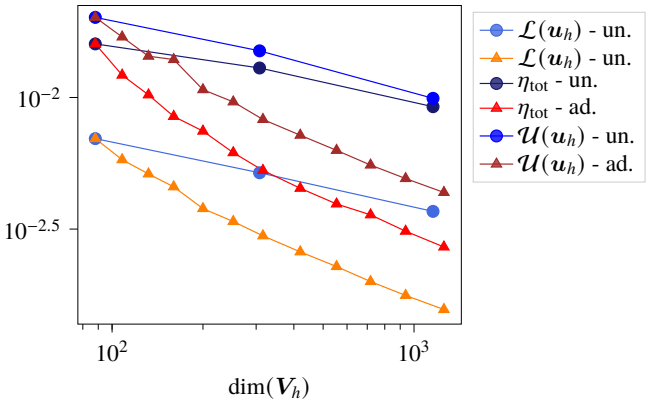
We refine adaptively the initial mesh following the distribution of the local total estimator $\eta_{\mathrm{tot},T}$, refining only the elements in which the value of this estimator is higher. In particular, at each refinement step, the 6% elements with larger estimated error are refined, i.e., are subdivided into four sub-triangles dividing each edge by two. Figure 4.6 shows the initial mesh and the result of adaptive refinement after 3, 7, and 11 steps, respectively. We remark that the refinement is concentrated on the endpoints of Γ_D (singularities due to the homogeneous Dirichlet conditions) and near the contact interval I_C. In particular, the strongest singularity is at the point where Γ_D and Γ_C meet, while the singularity at the corner point between Γ_D and Γ_N seems to be weaker. Figure 4.7a compares the convergence on uniformly and adaptively refined meshes for the H^1-norm $\|\bar{\boldsymbol{u}}_h - \boldsymbol{u}_h\|_{1,\Omega}$ and for the energy norm $\|\bar{\boldsymbol{u}}_h - \boldsymbol{u}_h\|_{\mathrm{en}}$, showing the corresponding curves as functions of the number of degrees of freedom. In particular, the uniform refinement is performed dividing all triangles of the mesh into four sub-triangles. The adaptive approach provides better convergence rates, i.e., adopting it we can achieve a fixed level of precision with fewer degrees of freedom. The asymptotic convergence rates are approximately 0.309 and 0.255 in the uniform case, and 0.450 and 0.449 in the adaptive case, for the H^1-norm and energy norm, respectively (the optimal convergence rate for smooth solutions is 0.5).

We recall that the measure of the error is the dual norm $|||\mathcal{R}(\boldsymbol{u}_h)|||_*$ defined by (4.11), which is not computable (it can be, however, approximated through an elliptic lifting). As a consequence, recalling Theorems 4.8 and 4.11, we compare the global total estimator η_{tot} with the following quantities:

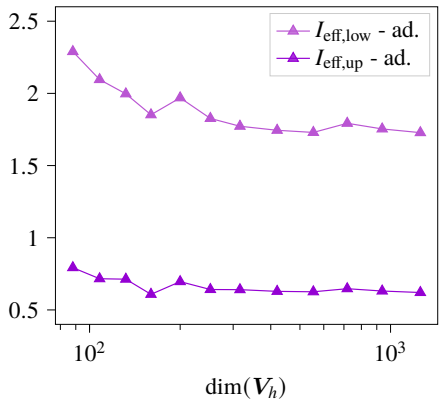
$$\mathcal{L}(\boldsymbol{u}_h) \coloneqq \mu^{1/2} \|\bar{\boldsymbol{u}}_h - \boldsymbol{u}_h\|_{\mathrm{en}}$$



(a) $\|\bar{u}_h - u_h\|_{1,\Omega}$ and $\|\bar{u}_h - u_h\|_{\text{en}}$.



(b) η_{tot}, $\mathcal{L}(u_h)$ and $\mathcal{U}(u_h)$.



(c) $I_{\text{eff,low}}$ and $I_{\text{eff,up}}$.

Figure 4.7 – Comparison between the uniform case (circles) and the adaptive one (triangles) for the H^1-norm $\|\bar{u}_h - u_h\|_{1,\Omega}$ and the energy norm $\|\bar{u}_h - u_h\|_{\text{en}}$ (*top*), and for the global total estimator η_{tot}, $\mathcal{L}(u_h)$ and $\mathcal{U}(u_h)$ (*bottom-left*). Corresponding effectivity indices $I_{\text{eff,low}}$ and $I_{\text{eff,up}}$ (*bottom-right*).

and

$$\mathcal{U}(u_h) \coloneqq (d\lambda + 4\mu)^{1/2} \|\bar{u}_h - u_h\|_{\text{en}} + \left(\sum_{F \in \mathcal{F}_h^C} h_F \left\| \sigma^n(\bar{u}_h) - \left[P^n_{1,\gamma}(u_h) \right]_{\mathbb{R}^-} \right\|_F^2 \right)^{1/2}.$$

In particular, the latter incorporates an additional error component on the contact interface. Furthermore, we define the two following effectivity indices:

$$I_{\text{eff,low}} \coloneqq \frac{\eta_{\text{tot}}}{\mathcal{L}(u_h)} = \frac{\eta_{\text{tot}}}{\mu^{1/2} \|\bar{u}_h - u_h\|_{\text{en}}} \qquad \text{and} \qquad I_{\text{eff,up}} \coloneqq \frac{\eta_{\text{tot}}}{\mathcal{U}(u_h)}.$$

The results are illustrated in Figure 4.7b and 4.7c. The total estimator always remains between the energy norm rescaled by a Lamé parameter $\mathcal{L}(u_h)$ and the upper bound for the dual residual norm $\mathcal{U}(u_h)$, i.e., $I_{\text{eff,low}} > 1$ and $I_{\text{eff,up}} < 1$. Figure 4.8 shows the distribution of the local total estimator $\eta_{\text{tot},T}$ at each mesh refinement step in both the uniform and adaptive

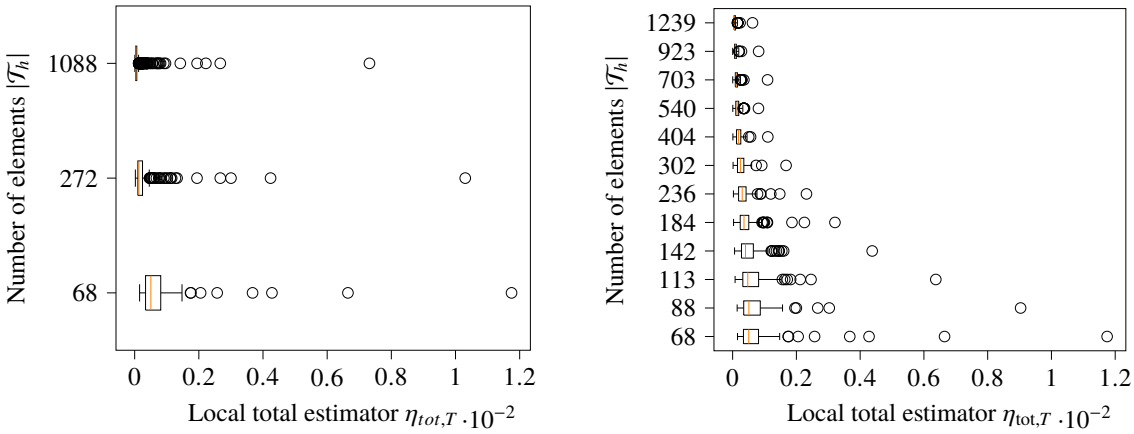


Figure 4.8 – Evolution of the distribution of the local total estimator $\eta_{\mathrm{tot},T}$ over the mesh with uniform refinement (*left*) and adaptive refinement (*right*). The right panel shows that the interval $(\min_{T\in\mathcal{T}_h} \eta_{\mathrm{tot},T}, \max_{T\in\mathcal{T}_h} \eta_{\mathrm{tot},T})$ shrinks much faster in the adaptively refined case than in the uniformly refined one. The labels of the y-axis report the number of elements of the corresponding mesh.

	Initial	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}	11^{th}
N_{reg}	7	0	1	0	0	0	0	0	0	0	0	0
N_{lin}	26	2	4	5	3	4	4	4	5	8	8	7

Table 4.2 – Number of regularization iterations N_{reg} and Newton iterations N_{lin} at each refinement step of the Algorithm 4.1.

frameworks. Here, we use boxplots to see where values concentrate. With the adaptive approach, all the local estimators $\{\eta_{\mathrm{tot},T}\}_{T\in\mathcal{T}_h}$ are contained in an interval that becomes smaller and smaller at each refinement iteration, and the decrease of the maximum value is significantly faster than in the uniformly refined computation. Indeed, in the latter there is always a value which is much bigger than the others even if the number of degrees of freedom and the number of elements are high (in the last case, $|V_h| = 1156$ and $|\mathcal{T}_h| = 1088$), showing that the error concentrates in specific areas. Figure 4.9 compares the selection of triangles to refine (highlighted in green) using the distribution of the energy norm $\|\bar{\boldsymbol{u}}_h - \boldsymbol{u}_h\|_{\mathrm{en},T}$ (*left*) and the total estimator $\eta_{\mathrm{tot},T}$ (*right*) for the initial mesh and the adaptively refined mesh after 6 and 10 steps. The sets of selected triangles are concentrated in the same zones.

Finally, we apply the fully adaptive Algorithm 4.1 with $\gamma_{\mathrm{reg}} = 0.04$ and $\gamma_{\mathrm{lin}} = 0.08$. The initial regularization parameter δ is taken equal to the Young modulus E and, at each step in which the global stopping criterion shown in Line 11 of Algorithm 4.1 is not satisfied, we divide it by 2. Table 4.2 contains the number of regularization and Newton iterations, denoted by N_{reg} and N_{lin} respectively, and Figure 4.10a displays the curves of the different global estimators for 11 adaptive refinement steps as functions of the degrees of freedom. The same estimators are shown in Figure 4.10b and 4.10c as functions of the Newton iterations for the 3rd and 9th adaptively refined meshes. A circle underlines the step (5th and 8th, respectively) at which the global stopping criterion of Line 9 is reached. At this step, the regularization

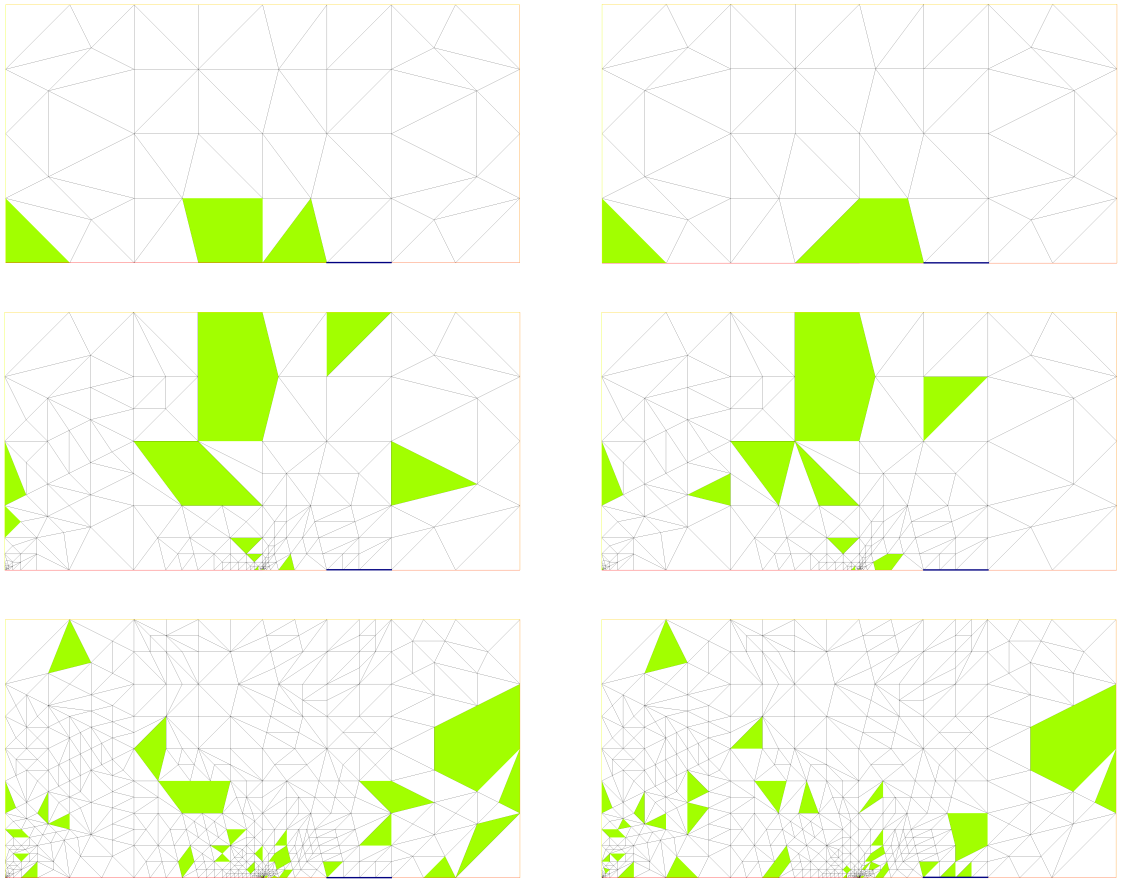
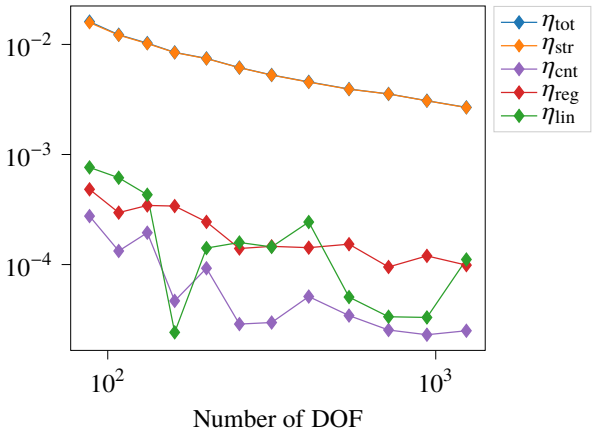


Figure 4.9 – Triangles to refine following the distribution of $\|\bar{\boldsymbol{u}}_h - \boldsymbol{u}_h\|_{\text{en}}$ (*left*) and of η_{tot} (*left*) for the initial mesh (*top*), and adaptively refined mesh after 6 and 10 steps (*middle* and *bottom*, respectively). The edges $F \subset \Gamma_C$ that include the contact interval I_C are highlighted in *blue*.

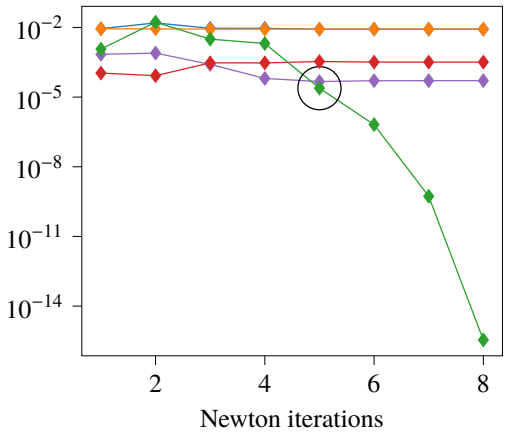
estimator satisfies the global stopping criterion of Line 11, and the other ones have already stabilized.

4.7 Efficiency of the local and global estimators

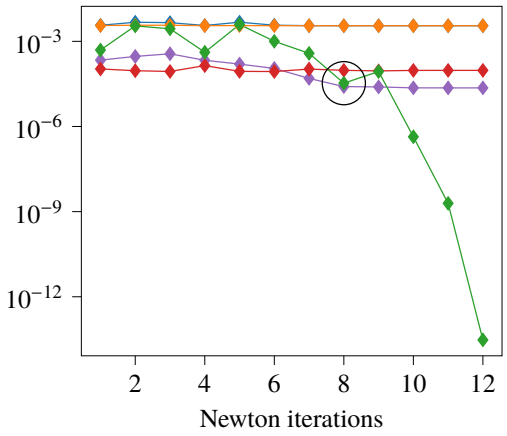
Using the stress reconstruction described in Section 4.5, we exhibit the demonstration of local and global efficiency of the estimators defined by (4.35) and (4.36). From now on, we adopt the notation of [122] for unions of elements of the triangulation, cf. Table 4.3. Finally, we remind that, as in Subsection 4.3.3, the notation $a \lesssim b$ stands for $a \leq Cb$, where $C > 0$ is a constant which is independent of the mesh size h and of the Nitsche parameter γ_0.



(a) Global estimators with the stopping criteria.



(b) 3rd adaptively refined mesh.



(c) 9th adaptively refined mesh.

Figure 4.10 – Global estimators η_{tot}, η_{str}, η_{lin}, η_{reg} and η_{cnt} as function of the number of degrees of freedom using the global stopping criteria of Algorithm 4.1 (*top*), and as function of Newton iterations for the 3rd and 9th adaptively refined mesh (*bottom-left* and *bottom-right*, respectively).

4.7.1 Local efficiency

With the aim of showing the efficiency of the local estimators (4.35) we introduce, for all $T \in \mathcal{T}_h$, the *local residual* defined as follows: For all $\boldsymbol{w}_h \in \boldsymbol{V}_h$ and all $\boldsymbol{v} \in \boldsymbol{H}^1_D(\tilde{\omega}_T)$

$$\langle \mathcal{R}_{\mathcal{T}_T}(\boldsymbol{w}_h), \boldsymbol{v} \rangle_{\tilde{\omega}_T} \coloneqq (\boldsymbol{f}, \boldsymbol{v})_{\tilde{\omega}_T} + (\boldsymbol{g}_N, \boldsymbol{v})_{\partial\tilde{\omega}_T \cap \Gamma_N} - (\boldsymbol{\sigma}(\boldsymbol{w}_h), \boldsymbol{\varepsilon}(\boldsymbol{v}))_{\tilde{\omega}_T} + \left(\left[P^n_{1,\gamma}(\boldsymbol{w}_h) \right]_{\mathbb{R}^-}, v^n \right)_{\partial\tilde{\omega}_T \cap \Gamma_C},$$

where

$$\boldsymbol{H}^1_D(\tilde{\omega}_T) \coloneqq \left\{ \boldsymbol{v} \in \boldsymbol{H}^1(\tilde{\omega}_T) \ : \ \boldsymbol{v} = \boldsymbol{0} \text{ on } \partial\tilde{\omega}_T \cap \Gamma_D \text{ and on } \partial\tilde{\omega}_T \cap \Omega \right\}.$$

Letting

$$|||\boldsymbol{v}|||_{\tilde{\omega}_T} \coloneqq \left(\|\boldsymbol{\nabla}\boldsymbol{v}\|^2_{\tilde{\omega}_T} + |\boldsymbol{v}|^2_{C,\tilde{\omega}_T} \right)^{1/2} = \left(\|\boldsymbol{\nabla}\boldsymbol{v}\|^2_{\tilde{\omega}_T} + \sum_{F \in \mathcal{F}^C_{\mathcal{T}_T}} \frac{1}{h_F} \|\boldsymbol{v}\|^2_F \right)^{1/2},$$

Notation	Definition
$\omega_{\boldsymbol{a}}$	Patch around the node $\boldsymbol{a}$, i.e., union of all elements having $\boldsymbol{a}$ as a node, Figure 4.3
$\mathcal{T}_{\boldsymbol{a}}$	Set of elements corresponding to the patch $\omega_{\boldsymbol{a}}$
ω_F	Union of all elements having F as a face, Figure 4.11
$\tilde{\omega}_F$	Union of all elements sharing at least one vertex with F, Figure 4.1
$\tilde{\omega}_T$	Union of all elements sharing at least one vertex with T, Figure 4.12
$\mathcal{T}_T$	Set of elements corresponding to $\tilde{\omega}_T$

Table 4.3 – Notation for unions and sets of elements

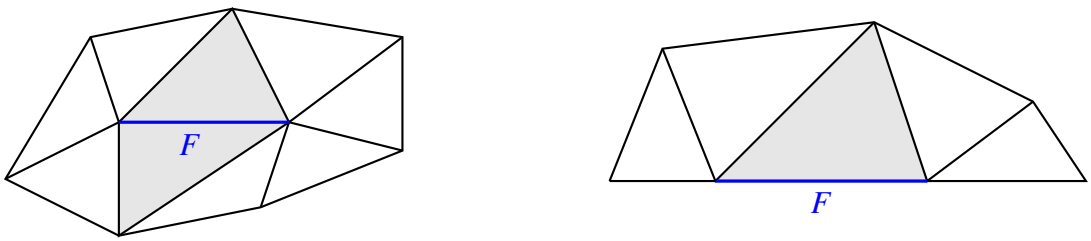


Figure 4.11 – Illustration of ω_F for $F \in \mathcal{F}_h^i$ (*left*) and for $F \in \mathcal{F}_h^b$ (*right*).

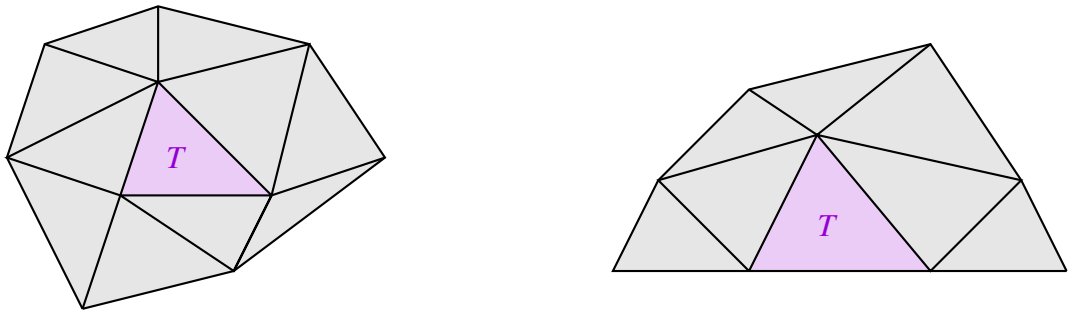


Figure 4.12 – Illustration of $\tilde{\omega}_T$ for $T \in \mathcal{T}_h$ such that $\mathcal{F}_T^b = \emptyset$ (*left*) and that $\mathcal{F}_T^b \neq \emptyset$ (*right*).

the corresponding dual norm of the residual for a function $\boldsymbol{w}_h \in \boldsymbol{V}_h$ is

$$|||\mathcal{R}_{\mathcal{T}_T}(\boldsymbol{w}_h)|||_{*,\tilde{\omega}_T} = \sup_{\boldsymbol{v}\in\boldsymbol{H}^1_D(\tilde{\omega}_T),\ |||\boldsymbol{v}|||_{\tilde{\omega}_T}=1} \langle \mathcal{R}_{\mathcal{T}_T}(\boldsymbol{w}_h), \boldsymbol{v}\rangle_{\tilde{\omega}_T}.$$

Following the path of [122], we introduce, for any element $T \in \mathcal{T}_h$, a local estimator residual-based defined on the patch $\tilde{\omega}_T$:

$$\begin{aligned}\eta^k_{\sharp,T} \coloneqq &\left(\sum_{T'\in\mathcal{T}_T} h^2_{T'} \left\|\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}(\boldsymbol{u}^k_h) + \boldsymbol{\Pi}^p_{T'}\boldsymbol{f}\right\|^2_{T'}\right)^{1/2} + \left(\sum_{F\in\mathcal{F}^i_{\mathcal{T}_T}} h_F \left\|[\![\boldsymbol{\sigma}(\boldsymbol{u}^k_h)\boldsymbol{n}_F]\!]\right\|^2_F\right)^{1/2} + \\ &+\left(\sum_{F\in\mathcal{F}^N_{\mathcal{T}_T}} h_F \left\|\boldsymbol{\sigma}(\boldsymbol{u}^k_h)\boldsymbol{n} - \boldsymbol{\Pi}^{p+1}_F \boldsymbol{g}_N\right\|^2_F\right)^{1/2} + \\ &+\left(\sum_{F\in\mathcal{F}^C_{\mathcal{T}_T}} h_F \left\|\boldsymbol{\sigma}(\boldsymbol{u}^k_h)\boldsymbol{n} - \Pi^{p+1}_F \left[P^n_{1,\gamma}(\boldsymbol{u}^k_h)\right]_{\mathbb{R}^-}\boldsymbol{n}\right\|^2_F\right)^{1/2}\end{aligned} \tag{4.47}$$

Lemma 4.22 (Control of the residual-based estimator $\eta_{\sharp,T}$)**.** *For any element $T \in \mathcal{T}_h$,*

$$\eta^k_{\sharp,T} \lesssim |||\mathcal{R}_{\mathcal{T}_T}(\boldsymbol{u}^k_h)|||_{*,\tilde{\omega}_T} + \eta^k_{\mathrm{osc},\mathcal{T}_T} + \eta^k_{\mathrm{Neu},\mathcal{T}_T} + \eta^k_{\mathrm{cnt},\mathcal{T}_T}, \tag{4.48}$$

where

$$\eta^k_{\bullet,\mathcal{T}_T} \coloneqq \left(\sum_{T'\in\mathcal{T}_T} \left(\eta^k_{\bullet,T'}\right)^2\right)^{1/2} \qquad \text{with } \bullet \in \{\mathrm{osc}, \mathrm{Neu}, \mathrm{cnt}\}.$$

Before giving a proof of Lemma 4.22, we recall that the *bubble function* of an element T or of an face F is defined respectively by

$$\psi_T \coloneqq \alpha_T \prod_{\boldsymbol{a}\in\mathcal{V}_T} \psi_{\boldsymbol{a}} \in \mathcal{P}^{d+1}(T), \qquad \psi_F \coloneqq \alpha_F \prod_{\boldsymbol{a}\in\mathcal{V}_F} \psi_{\boldsymbol{a}} \in \mathcal{P}^d(\omega_F), \tag{4.49}$$

where the constants α_T and α_F are determined by the conditions $\max_{\boldsymbol{x}\in T}\psi_T(\boldsymbol{x}) = 1$ and $\max_{\boldsymbol{x}\in F}\psi_F(\boldsymbol{x}) = 1$. Moreover, we will use the five properties of ψ_T and ψ_F stated in [122, Equation (4)] that we quote here for simplicity:

$$\|\boldsymbol{v}\|^2_T \lesssim (\psi_T\boldsymbol{v},\boldsymbol{v})_T \le \|\boldsymbol{v}\|^2_T, \tag{4.50a}$$

$$\|\psi_T\boldsymbol{v}\|_{1,T} \lesssim \frac{1}{h_T}\|\boldsymbol{v}\|_T, \tag{4.50b}$$

$$\|\varphi\|^2_F \lesssim (\psi_F\varphi,\varphi)_F \le \|\varphi\|^2_F, \tag{4.50c}$$

$$\|\psi_F\varphi\|_{1,\omega_F} \lesssim \frac{1}{h_F^{1/2}}\|\varphi\|_F, \tag{4.50d}$$

$$\|\psi_F\varphi\|_{\omega_F} \lesssim h_F^{1/2}\|\varphi\|_F, \tag{4.50e}$$

where $T \in \mathcal{T}_h$, $F \in \mathcal{F}_h$, $\boldsymbol{v}$ and φ are d-valued polynomials of degree at most r defined on T and ω_F, respectively. The hidden constants depend only on the polynomial degree r and on the shape parameter ρ.

Remark 4.23 (Extension of (4.50a) and (4.50c))**.** *Following the path of [55], it is possible to show that for any $\mathcal{S} \subseteq \mathcal{T}_h$ and for any $\boldsymbol{v} \in \boldsymbol{\mathcal{P}}^r(\mathcal{S})$*

$$\left(\sum_{T\in\mathcal{S}} h_T^2 \|\boldsymbol{v}\|_T^2\right)^{1/2} \lesssim \sup_{\substack{\boldsymbol{w}\in\boldsymbol{\mathcal{P}}^r(\mathcal{S}),\\ \|\boldsymbol{w}\|_{\mathcal{S}}=1}} \sum_{T\in\mathcal{S}} (\boldsymbol{v}, h_T\psi_T\boldsymbol{w})_T. \tag{4.51}$$

where $\|\boldsymbol{w}\|_{\mathcal{S}} \coloneqq \left(\sum_{T\in\mathcal{S}}\|\boldsymbol{w}\|_T^2\right)^{1/2}$. In a similar way, for any $\mathcal{E} \subseteq \mathcal{F}_h$ and for any $\boldsymbol{\varphi} \in \boldsymbol{\mathcal{P}}^r(\mathcal{E})$

$$\left(\sum_{F\in\mathcal{E}} h_F \|\boldsymbol{\varphi}\|_F^2\right)^{1/2} \lesssim \sup_{\substack{\boldsymbol{\phi}\in\boldsymbol{\mathcal{P}}^r(\mathcal{E}),\\ \|\boldsymbol{\phi}\|_{\mathcal{E}}=1}} \sum_{F\in\mathcal{E}} (\boldsymbol{\varphi}, h_F^{1/2}\psi_F\boldsymbol{\phi})_F \tag{4.52}$$

where $\|\boldsymbol{\phi}\|_{\mathcal{E}} \coloneqq \left(\sum_{F\in\mathcal{E}}\|\boldsymbol{\phi}\|_F^2\right)^{1/2}$.

Proof of Lemma 4.22. Let us fix an element $T \in \mathcal{T}_h$. We analyse each term of $\eta_{\sharp,T}^k$ (4.47) separately. For simplicity, we denote them with $\mathcal{J}_1$, $\mathcal{J}_2$, $\mathcal{J}_3$, and $\mathcal{J}_4$, respectively. The idea is to use (4.51) with $\mathcal{S} = \mathcal{T}_T$ for $\mathcal{J}_1$ and (4.52) with $\mathcal{E} = \mathcal{F}_{\mathcal{T}_T}^i, \mathcal{F}_{\mathcal{T}_T}^N, \mathcal{F}_{\mathcal{T}_T}^C$ for $\mathcal{J}_2$, $\mathcal{J}_3$, $\mathcal{J}_4$, respectively. Since $(\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}(\boldsymbol{u}_h^k))|_{T'} + \boldsymbol{\Pi}_{T'}^p\boldsymbol{f}|_{T'} \in \boldsymbol{\mathcal{P}}^p(T')$ for every $T' \in \mathcal{T}_T$, by (4.51)

$$\mathcal{J}_1 \lesssim \sup_{\substack{\boldsymbol{w}\in\boldsymbol{\mathcal{P}}^p(\mathcal{T}_T),\\ \|\boldsymbol{w}\|_{\tilde{\omega}_T}=1}} \sum_{T'\in\mathcal{T}_T} (\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}(\boldsymbol{u}_h^k) + \boldsymbol{\Pi}_{T'}^p\boldsymbol{f}, h_{T'}\psi_{T'}\boldsymbol{w})_{T'}. \tag{4.53}$$

Fix $\boldsymbol{w} \in \boldsymbol{\mathcal{P}}^p(\mathcal{T}_T)$ with $\|\boldsymbol{w}\|_{\tilde{\omega}_T} = 1$, and define $\boldsymbol{\lambda}|_{T'} \coloneqq h_{T'}\psi_{T'}\boldsymbol{w}|_{T'}$ for every $T' \in \mathcal{T}_T$. Notice that $\boldsymbol{\lambda} \in \boldsymbol{\mathcal{P}}^{p+d+1}(\mathcal{T}_T) \cap \boldsymbol{H}_D^1(\tilde{\omega}_T)$ (cf. (4.49)). Then, using the Green formula on each element $T' \in \mathcal{T}_T$, the definition of the residual, and the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}
&\sum_{T'\in\mathcal{T}_T} (\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}(\boldsymbol{u}_h^k) + \boldsymbol{\Pi}_{T'}^p\boldsymbol{f}, h_{T'}\psi_{T'}\boldsymbol{w})_{T'} \lesssim \\
&\quad\lesssim |||\mathcal{R}_{\mathcal{T}_T}(\boldsymbol{u}_h^k)|||_{*,\tilde{\omega}_T} |||\boldsymbol{\lambda}|||_{\tilde{\omega}_T} + \left(\sum_{T'\in\mathcal{T}_T} h_{T'}^2 \left\|\boldsymbol{f} - \boldsymbol{\Pi}_{T'}^{p-1}\boldsymbol{f}\right\|_{T'}^2\right)^{1/2} \left(\sum_{T'\in\mathcal{T}_T} \|\psi_{T'}\boldsymbol{w}\|_{T'}^2\right)^{1/2}.
\end{aligned} \tag{4.54}$$

Here, we have also used the fact that $\|\boldsymbol{f} - \boldsymbol{\Pi}_{T'}^p\boldsymbol{f}\|_{T'} \le 2\|\boldsymbol{f} - \boldsymbol{\Pi}_{T'}^{p-1}\boldsymbol{f}\|_{T'}$ for $p > 0$. By the definition of $\boldsymbol{\lambda}$, leveraging the properties (4.50a) and (4.50b), along with the fact that $\|\boldsymbol{w}\|_{\tilde{\omega}_T} = 1$, it is possible to show that

$$|||\boldsymbol{\lambda}|||_{\tilde{\omega}_T} \lesssim 1 \qquad \text{and} \qquad \left(\sum_{T'\in\mathcal{T}_T} \|\psi_{T'}\boldsymbol{w}\|_{T'}^2\right)^{1/2} \lesssim 1,$$

and, combining (4.53) and (4.54) with the estimation above, we obtain

$$\mathcal{J}_1 \lesssim |||\mathcal{R}_{\mathcal{T}_T}(\boldsymbol{u}_h^k)|||_{*,\tilde{\omega}_T} + \eta_{\mathrm{osc},\mathcal{T}_T}^k. \tag{4.55}$$

Now, we analyse for instance the term $\mathcal{J}_4$, while $\mathcal{J}_2$ and $\mathcal{J}_3$ can be treated in a similar way. Since $\boldsymbol{\sigma}(\boldsymbol{u}_h^k)\boldsymbol{n} - \Pi_F^{p+1}\left[P_{1,\gamma}^n(\boldsymbol{u}_h^k)\right]_{\mathbb{R}^-}\boldsymbol{n} \in \boldsymbol{\mathcal{P}}^{p+1}(F)$ for every $F \in \mathcal{F}_{\mathcal{T}_T}^C$, by (4.52) we have

$$\mathcal{J}_4 \lesssim \sup_{\substack{\boldsymbol{\phi}\in\boldsymbol{\mathcal{P}}^{p+1}(\mathcal{F}_{\mathcal{T}_T}^C),\\ \|\boldsymbol{\phi}\|_{\mathcal{F}_{\mathcal{T}_T}^C}=1}} \sum_{F\in\mathcal{F}_{\mathcal{T}_T}^C} \left(\boldsymbol{\sigma}(\boldsymbol{u}_h^k)\boldsymbol{n} - \Pi_F^{p+1}\left[P_{1,\gamma}^n(\boldsymbol{u}_h^k)\right]_{\mathbb{R}^-}\boldsymbol{n}, h_F^{1/2}\psi_F\boldsymbol{\phi}\right)_F. \tag{4.56}$$

Fix $\boldsymbol{\phi} \in \boldsymbol{\mathcal{P}}^{p+1}(\mathcal{F}_{\mathcal{T}_T}^C)$ with $\|\boldsymbol{\phi}\|_{\mathcal{F}_{\mathcal{T}_T}^C} = 1$. Moreover, let $\boldsymbol{\lambda} \in \boldsymbol{\mathcal{P}}^{p+d+1}(\mathcal{T}_T) \cap \boldsymbol{H}_D^1(\tilde{\omega}_T)$ such that $\boldsymbol{\lambda}|_F = h_F^{1/2}\psi_F\boldsymbol{\phi}|_F$ for every $F \in \mathcal{F}_{\mathcal{T}_T}^C$, and that vanishes outside of $\bigcup_{F\in\mathcal{F}_{\mathcal{T}_T}^C}\omega_F$. Then,

$$\begin{aligned}
&\sum_{F\in\mathcal{F}_h^C}\left(\boldsymbol{\sigma}(\boldsymbol{u}_h^k)\boldsymbol{n} - \Pi_F^{p+1}\left[P_{1,\gamma}^n(\boldsymbol{u}_h^k)\right]_{\mathbb{R}^-}\boldsymbol{n}, h_F^{1/2}\psi_F\boldsymbol{\phi}\right)_F \\
&\qquad = -\langle\mathcal{R}_{\mathcal{T}_T}(\boldsymbol{u}_h^k), \boldsymbol{\lambda}\rangle_{\tilde{\omega}_T} + \sum_{T'\in\mathcal{T}_T}(\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}(\boldsymbol{u}_h^k) + \boldsymbol{f}, \boldsymbol{\lambda})_{T'} \\
&\qquad\quad + \sum_{F\in\mathcal{F}_{\mathcal{T}_T}^N}\left(\left[P_{1,\gamma}^n(\boldsymbol{u}_h^k)\right]_{\mathbb{R}^-}\boldsymbol{n} - \Pi_F^{p+1}\left[P_{1,\gamma}^n(\boldsymbol{u}_h^k)\right]_{\mathbb{R}^-}\boldsymbol{n}, \boldsymbol{\lambda}\right)_F \\
&\qquad \lesssim |||\mathcal{R}_{\mathcal{T}_T}(\boldsymbol{u}_h^k)|||_{*,\tilde{\omega}_T}|||\boldsymbol{\lambda}|||_{\tilde{\omega}_T} + (\mathcal{J}_1 + \eta_{\mathrm{osc},\mathcal{T}_T}^k)\left(\sum_{T'\in\mathcal{T}_T}\frac{1}{h_{T'}^2}\|\boldsymbol{\lambda}\|_{T'}^2\right)^{1/2} \\
&\qquad\quad + \eta_{\mathrm{cnt},\mathcal{T}_T}^k\left(\sum_{F\in\mathcal{F}_{\mathcal{T}_T}^N}\frac{1}{h_F}\|\boldsymbol{\lambda}\|_F^2\right)^{1/2}.
\end{aligned} \tag{4.57}$$

Exploiting the properties (4.50c), (4.50d) and (4.50e), it is possible to show that

$$|||\boldsymbol{\lambda}|||_{\tilde{\omega}_T} \lesssim 1, \qquad \left(\sum_{T'\in\mathcal{T}_T}\frac{1}{h_{T'}^2}\|\boldsymbol{\lambda}\|_{T'}^2\right)^{1/2} \lesssim 1 \qquad \text{and} \qquad \left(\sum_{F\in\mathcal{F}_{\mathcal{T}_T}^C}\frac{1}{h_F}\|\boldsymbol{\lambda}\|_F^2\right)^{1/2} \lesssim 1,$$

and, combining (4.56), (4.57), and (4.55), we conclude

$$\mathcal{J}_4 \lesssim |||\mathcal{R}_{\mathcal{T}_T}(\boldsymbol{u}_h^k)|||_{*,\tilde{\omega}_T} + \eta_{\mathrm{osc},\mathcal{T}_T}^k + \eta_{\mathrm{cnt},\mathcal{T}_T}^k.$$

In addition, it is possible to obtain

$$\mathcal{J}_2 \lesssim |||\mathcal{R}_{\mathcal{T}_T}(\boldsymbol{u}_h^k)|||_{*,\tilde{\omega}_T} + \eta_{\mathrm{osc},\mathcal{T}_T}^k,$$

$$\mathcal{J}_3 \lesssim |||\mathcal{R}_{\mathcal{T}_T}(\boldsymbol{u}_h^k)|||_{*,\tilde{\omega}_T} + \eta_{\mathrm{osc},\mathcal{T}_T}^k + \eta_{\mathrm{Neu},\mathcal{T}_T}^k.$$

and combining all the results obtained, we reach the thesis (4.48). □

Lemma 4.24 (Control of the local flux estimator)**.** *Assume $d = 2$. Let $\boldsymbol{u}_h^k \in \boldsymbol{V}_h$, let $\boldsymbol{\sigma}_h^k$ be the stress reconstruction defined by Construction 4.18, and let $\eta_{\sharp,T}^k$ be the local residual-based estimator defined by (4.47). Then, for every element $T \in \mathcal{T}_h$,*

$$\eta_{\mathrm{flux},T}^k \lesssim \eta_{\sharp,T}^k. \tag{4.58}$$

In order to prove this lemma for $d = 2$, and, in particular, to obtain the relation (4.59), as in [16], we introduce the following nonconforming space [5]:

$$\boldsymbol{M}_T := \begin{cases} \{\boldsymbol{m} \in \boldsymbol{\mathcal{P}}^{p+2}(T) \ : \ \boldsymbol{m}|_F \in \boldsymbol{\mathcal{P}}(F) \text{ for any } F \in \mathcal{F}_T\} & \text{if } p \text{ is even,} \\ \{\boldsymbol{m} \in \boldsymbol{\mathcal{P}}^{p+2}(T) \ : \ \boldsymbol{m}|_F \in \boldsymbol{\mathcal{P}}^p(F) \oplus \tilde{\boldsymbol{\mathcal{P}}}^{p+2}(F) \text{ for any } F \in \mathcal{F}_T\} & \text{if } p \text{ is odd,} \end{cases}$$

where $\tilde{\boldsymbol{\mathcal{P}}}^{p+2}(F)$ is the $L^2(F)$-orthogonal complement of $\boldsymbol{\mathcal{P}}^{p+1}(F)$ in $\boldsymbol{\mathcal{P}}^{p+2}(F)$. On the patch $\omega_{\boldsymbol{a}}$ we define the spaces

$$\begin{aligned} \boldsymbol{M}_h(\omega_{\boldsymbol{a}}) := \big\{ & \boldsymbol{m}_h \in \boldsymbol{L}^2(\omega_{\boldsymbol{a}}) \ : \ \boldsymbol{m}_h|_T \in \boldsymbol{M}_T \text{ for any } T \in \mathcal{T}_{\boldsymbol{a}}, \\ & (\llbracket \boldsymbol{m}_h \rrbracket, \boldsymbol{v}_h)_F = 0 \text{ for any } \boldsymbol{v}_h \in \boldsymbol{\mathcal{P}}^p(F) \text{ and for any } F \in \mathcal{F}_{\boldsymbol{a}} \setminus \mathcal{F}_h^b \big\}, \\ \boldsymbol{M}_h^{\boldsymbol{a}} := \big\{ & \boldsymbol{m}_h \in \boldsymbol{M}_h(\omega_{\boldsymbol{a}}) \ : \ (\boldsymbol{m}_h, \boldsymbol{z})_{\omega_{\boldsymbol{a}}} = 0 \text{ for any } \boldsymbol{z} \in \boldsymbol{RM}^2 \big\} \end{aligned}$$

if $\boldsymbol{a} \in \mathcal{V}_h^i$ or $\boldsymbol{a} \in \mathcal{V}_h^b \setminus \mathcal{V}_h^D$, and

$$\boldsymbol{M}_h^{\boldsymbol{a}} := \big\{ \boldsymbol{m}_h \in \boldsymbol{M}_h(\omega_{\boldsymbol{a}}) \ : \ (\boldsymbol{m}_h, \boldsymbol{v}_h)_F = 0 \text{ for any } \boldsymbol{v}_h \in \boldsymbol{\mathcal{P}}^p(F) \text{ and for any } F \in \mathcal{F}_{\boldsymbol{a}} \cap \mathcal{F}_h^D \big\}$$

if $\boldsymbol{a} \in \mathcal{V}_h^D$.

Proof of Lemma 4.24. Let $d = 2$ and fix $T \in \mathcal{T}_h$. Combining the definition of the local flux estimator with the triangle inequality, we directly get

$$\eta_{\mathrm{str},T}^k = \left\| \boldsymbol{\sigma}_h^k - \boldsymbol{\sigma}(\boldsymbol{u}_h^k) \right\|_T \leq \sum_{\boldsymbol{a} \in \mathcal{V}_T} \left\| \boldsymbol{\sigma}_h^{\boldsymbol{a},k} - \psi_{\boldsymbol{a}} \boldsymbol{\sigma}(\boldsymbol{u}_h^k) \right\|_{\omega_{\boldsymbol{a}}}.$$

Adapting the argument of [16, Section 4.4] to our problem with Neumann and contact boundary conditions, it is possible to show that

$$\left\| \boldsymbol{\sigma}_h^{\boldsymbol{a},k} - \psi_{\boldsymbol{a}} \boldsymbol{\sigma}(\boldsymbol{u}_h^k) \right\|_{\omega_{\boldsymbol{a}}} \lesssim \sup_{\substack{\boldsymbol{m}_h \in \boldsymbol{M}_h^{\boldsymbol{a}}, \\ \|\boldsymbol{\nabla}_h \boldsymbol{m}_h\|_{\omega_{\boldsymbol{a}}} = 1}} \left(\boldsymbol{\sigma}_h^{\boldsymbol{a},k} - \psi_{\boldsymbol{a}} \boldsymbol{\sigma}(\boldsymbol{u}_h^k), \boldsymbol{\nabla}_h \boldsymbol{m}_h \right)_{\omega_{\boldsymbol{a}}}, \tag{4.59}$$

for any $\boldsymbol{a} \in \mathcal{V}_h$, where $\boldsymbol{\nabla}_h \boldsymbol{m}_h$ is the broken gradient of $\boldsymbol{m}_h$, i.e., the function in $\mathbb{L}^2(\omega_{\boldsymbol{a}})$ such that $(\boldsymbol{\nabla}_h \boldsymbol{m}_h)|_T = \boldsymbol{\nabla}(\boldsymbol{m}_h|_T)$ for any $T' \in \mathcal{T}_{\boldsymbol{a}}$. In particular, the corresponding result in the cited paper is [16, Equation (4.32)]. Then, fixing a vertex $\boldsymbol{a} \in \mathcal{V}_h$ and a function $\boldsymbol{m}_h \in \boldsymbol{M}_h^{\boldsymbol{a}}$ such that $\|\boldsymbol{\nabla}_h \boldsymbol{m}_h\|_{\omega_{\boldsymbol{a}}} = 1$, we have

$$\begin{aligned} (\boldsymbol{\sigma}_h^{\boldsymbol{a},k} - \psi_{\boldsymbol{a}} \boldsymbol{\sigma}(\boldsymbol{u}_h^k), \boldsymbol{\nabla}_h \boldsymbol{m}_h)_{\omega_{\boldsymbol{a}}} &= \sum_{T' \in \mathcal{T}_{\boldsymbol{a}}} (\boldsymbol{\sigma}_h^{\boldsymbol{a},k} - \psi_{\boldsymbol{a}} \boldsymbol{\sigma}(\boldsymbol{u}_h^k), \boldsymbol{\nabla}_h \boldsymbol{m}_h)_{T'} \\ &= \underbrace{- \sum_{T' \in \mathcal{T}_{\boldsymbol{a}}} \left(\boldsymbol{\nabla} \cdot (\boldsymbol{\sigma}_h^{\boldsymbol{a},k} - \psi_{\boldsymbol{a}} \boldsymbol{\sigma}(\boldsymbol{u}_h^k)), \boldsymbol{m}_h \right)_{T'}}_{=: \mathcal{I}_1} + \underbrace{\sum_{F \in \mathcal{F}_{\boldsymbol{a}}^i} \left(\llbracket \psi_{\boldsymbol{a}} \boldsymbol{\sigma}(\boldsymbol{u}_h^k) \boldsymbol{n}_F \rrbracket, \boldsymbol{m}_h \right)_F}_{=: \mathcal{I}_2} + \\ &\quad + \underbrace{\sum_{F \in \mathcal{F}_{\boldsymbol{a}}^N} \left(\boldsymbol{\Pi}_F^p (\psi_{\boldsymbol{a}} \boldsymbol{g}_N) - \psi_{\boldsymbol{a}} \boldsymbol{\sigma}(\boldsymbol{u}_h^k) \boldsymbol{n}, \boldsymbol{m}_h \right)_F}_{=: \mathcal{I}_3} + \\ &\quad + \underbrace{\sum_{F \in \mathcal{F}_{\boldsymbol{a}}^C} \left(\boldsymbol{\Pi}_F^p \left(\psi_{\boldsymbol{a}} \left[P_{1,\gamma}^n(\boldsymbol{u})_h^k \right] \boldsymbol{n} \right) - \psi_{\boldsymbol{a}} \boldsymbol{\sigma}(\boldsymbol{u}_h^k) \boldsymbol{n}, \boldsymbol{m}_h \right)_F}_{=: \mathcal{I}_4}. \end{aligned}$$

Here, we have use the Green formula, the properties of M_h^a, the fact that $\boldsymbol{\sigma}_h^{a,k} \in \boldsymbol{\Sigma}_{h,N,C}^{a,k}(\boldsymbol{u}_h^k)$ (cf. (4.42)), and that $\psi_a = 0$ on $\partial\omega_a \setminus \partial\Omega$. The first two terms can be treated as in [16, Proof of Theorem 4.7], obtaining

$$\mathcal{I}_1 \lesssim \left[\sum_{T'\in\mathcal{T}_a} h_{T'}^2 \left\|\boldsymbol{\Pi}_{T'}^p \boldsymbol{f} + \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}(\boldsymbol{u}_h^k)\right\|_{T'}^2\right]^{1/2} \|\boldsymbol{\nabla}_h \boldsymbol{m}_h\|_{\omega_a}$$

and

$$\mathcal{I}_2 \lesssim \left[\sum_{F\in\mathcal{F}_a^i} h_F \left\|[\![\boldsymbol{\sigma}(\boldsymbol{u}_h^k)\boldsymbol{n}_F]\!]\right\|_F^2\right]^{1/2} \|\boldsymbol{\nabla}_h \boldsymbol{m}_h\|_{\omega_a}.$$

In a similar way, using the Cauchy-Schwarz inequality, the discrete trace inequality $\|\boldsymbol{m}_h\|_F \lesssim h_F^{-1/2}\|\boldsymbol{m}_h\|_{T'}$ and the Poincaré/Friedrichs inequality [125] (if $\boldsymbol{a} \notin \mathcal{V}_h^D / \boldsymbol{a} \in \mathcal{V}_h^D$) together with the definition of M_h^a, we have

$$\begin{aligned}
\mathcal{I}_3 &= \sum_{F\in\mathcal{F}_a^N} \left(\psi_a(\boldsymbol{g}_N - \boldsymbol{\sigma}(\boldsymbol{u}_h^k)\boldsymbol{n}), \boldsymbol{\Pi}_F^p \boldsymbol{m}_h\right)_F = \sum_{F\in\mathcal{F}_a^N} \left(\boldsymbol{\Pi}_F^{p+1}\boldsymbol{g}_N - \boldsymbol{\sigma}(\boldsymbol{u}_h^k)\boldsymbol{n}, \psi_a \boldsymbol{\Pi}_F^p \boldsymbol{m}_h\right)_F \\
&\le \left[\sum_{F\in\mathcal{F}_a^N} h_F \left\|\psi_a\left(\boldsymbol{\Pi}_F^{p+1}\boldsymbol{g}_N - \boldsymbol{\sigma}(\boldsymbol{u}_h^k)\boldsymbol{n}\right)\right\|_F^2\right]^{1/2} \left[\sum_{F\in\mathcal{F}_a^N} \frac{1}{h_F}\|\boldsymbol{m}_h\|_F^2\right]^{1/2} \\
&\lesssim \left[\sum_{F\in\mathcal{F}_a^N} h_F \left\|\boldsymbol{\Pi}_F^{p+1}\boldsymbol{g}_N - \boldsymbol{\sigma}(\boldsymbol{u}_h^k)\boldsymbol{n}\right\|_F^2\right]^{1/2} \|\boldsymbol{\nabla}_h \boldsymbol{m}_h\|_{\omega_a}
\end{aligned}$$

In addition, it is possible to obtain

$$\mathcal{I}_4 \lesssim \left[\sum_{F\in\mathcal{F}_a^C} h_F \left\|\boldsymbol{\Pi}_F^{p+1}\left(\left[P_{1,\gamma}(\boldsymbol{u}_h^k)\right]_{\mathbb{R}^-}\boldsymbol{n}\right) - \boldsymbol{\sigma}(\boldsymbol{u}_h^k)\boldsymbol{n}\right\|_F^2\right]^{1/2} \|\boldsymbol{\nabla}_h \boldsymbol{m}_h\|_{\omega_a}$$

Combining these results with (4.59) yields (4.58). □

Remark 4.25 (Case $d = 3$)**.** *For the case $d = 3$, it is more difficult to find a space M_h^a with the features that allow us to recover the relation (4.59). For this reason, in this paper we show the proof of Lemma 4.24 only for $d = 2$.*

It is possible to prove this result by naturally adapting the approach of [16, Theorem 4.7]. For simplicity, we will omit the proof.

Theorem 4.26 (Local efficiency)**.** *Assume $d = 2$. Let $\boldsymbol{u}_h^k \in \boldsymbol{V}_h$ and let $\boldsymbol{\sigma}_h^k$ be the stress reconstruction of Construction 4.18, and assume that the local stopping criteria (4.40a) and (4.40b) hold. Then, for every element $T \in \mathcal{T}_h$,*

$$\begin{aligned}
&\eta_{\mathrm{osc},T}^k + \eta_{\mathrm{str},T}^k + \eta_{\mathrm{Neu},T}^k + \eta_{\mathrm{cnt},T}^k + \eta_{\mathrm{lin},T}^k + \eta_{\mathrm{reg},T}^k \\
&\qquad\qquad \lesssim |||\mathcal{R}_{\mathcal{T}_T}(\boldsymbol{u}_h^k)|||_{*,\tilde{\omega}_T} + \eta_{\mathrm{osc},\mathcal{T}_T}^k + \eta_{\mathrm{Neu},\mathcal{T}_T}^k + \eta_{\mathrm{cnt},\mathcal{T}_T}^k. \qquad (4.60)
\end{aligned}$$

Proof. The thesis follows using the local stopping criteria and combining the results of Lemmas 4.22 and 4.24. □

In the right-hand side of the relation (4.60), there are two "contact" terms: $\eta^k_{\mathrm{cnt},\mathcal{T}_T}$ (residual contact term) and $\eta^k_{\mathrm{disc},\mathcal{T}_T}$. It is possible to get rid of the second one, making additional assumptions on the parameters $\gamma_{\mathrm{reg},T}$ and $\gamma_{\mathrm{lin},T}$ of the adaptive algorithm 4.1, as shown in the following theorem.

4.7.2 Global efficiency

In this brief subsection, after showing the global versions of Lemma 4.22 and Lemma 4.24, we present the result of global efficiency of the estimators (4.36).

Lemma 4.27. *For any* $k \geq 1$

$$\eta^k_\sharp \lesssim |||\mathcal{R}(\boldsymbol{u}^k_h)|||_* + \eta^k_{\mathrm{osc}} + \eta^k_{\mathrm{Neu}} + \eta^k_{\mathrm{cnt}}. \tag{4.61}$$

Proof of Lemma 4.27. We have

$$\eta^k_\sharp \lesssim \underbrace{\left(\sum_{T\in\mathcal{T}_h} h_T^2 \left\|\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}(\boldsymbol{u}^k_h) + \boldsymbol{\Pi}^p_T \boldsymbol{f}\right\|^2_T\right)^{1/2}}_{=:\mathcal{L}_1} + \underbrace{\left(\sum_{F\in\mathcal{F}^i_h} h_F \left\|[\![\boldsymbol{\sigma}(\boldsymbol{u}^k_h)\boldsymbol{n}_F]\!]\right\|^2_F\right)^{1/2}}_{=:\mathcal{L}_2} +$$
$$+ \underbrace{\left(\sum_{F\in\mathcal{F}^N_h} h_F \left\|\boldsymbol{\sigma}(\boldsymbol{u}^k_h)\boldsymbol{n} - \boldsymbol{\Pi}^{p+1}_F \boldsymbol{g}_N\right\|^2_F\right)^{1/2}}_{=:\mathcal{L}_3} +$$
$$+ \underbrace{\left(\sum_{F\in\mathcal{F}^C_h} h_F \left\|\boldsymbol{\sigma}(\boldsymbol{u}^k_h)\boldsymbol{n} - \boldsymbol{\Pi}^{p+1}_F \left(\left[P^n_{1,\gamma}(\boldsymbol{u}^k_h)\right]_{\mathbb{R}^-}\boldsymbol{n}\right)\right\|^2_F\right)^{1/2}}_{=:\mathcal{L}_4}$$

Proceeding as in the proof of Lemma 4.22, it is possible to show that

$$\mathcal{L}_1 \lesssim |||\mathcal{R}(\boldsymbol{u}^k_h)|||_* + \eta^k_{\mathrm{osc}}, \qquad \mathcal{L}_2 \lesssim |||\mathcal{R}(\boldsymbol{u}^k_h)|||_* + \eta^k_{\mathrm{osc}},$$
$$\mathcal{L}_3 \lesssim |||\mathcal{R}(\boldsymbol{u}^k_h)|||_* + \eta^k_{\mathrm{osc}} + \eta^k_{\mathrm{Neu}}, \qquad \mathcal{L}_4 \lesssim |||\mathcal{R}(\boldsymbol{u}^k_h)|||_* + \eta^k_{\mathrm{osc}} + \eta^k_{\mathrm{cnt}}.$$

□

Lemma 4.28. *For any* $k \geq 1$,

$$\eta^k_{\mathrm{str}} \lesssim \eta^k_\sharp.$$

Proof. It is an immediate consequence of Lemma 4.24. □

Theorem 4.29 (Global efficiency)**.** *Let $\boldsymbol{u}_h^k \in \boldsymbol{V}_h$ and let $\boldsymbol{\sigma}_h^k$ be defined by Construction 4.18, assume that the global stopping criteria of Line 9 and 11 hold. Then*

$$\eta_{\mathrm{osc}}^k + \eta_{\mathrm{str}}^k + \eta_{\mathrm{Neu}}^k + \eta_{\mathrm{cnt}}^k + \eta_{\mathrm{lin}}^k + \eta_{\mathrm{reg}}^k \lesssim |||\mathcal{R}(\boldsymbol{u}_h^k)|||_* + \eta_{\mathrm{osc}}^k + \eta_{\mathrm{Neu}}^k + \eta_{\mathrm{cnt}}^k.$$

Proof. It is sufficient to use the global stopping criteria and to combine the results of Lemma 4.27 and Lemma 4.28. □

4.8 Conclusion

We have developed an a posteriori error estimation based on the equilibrated stress reconstruction for the unilateral contact problem without friction discretized with a Nitsche-based method. The obtained estimators are then compared for performing an adaptive refinement of the mesh and for defining two stopping criteria which fix automatically the number of Newton iterations and the value of a regularization parameter. The proposed numerical example compares the results obtained with a uniform refinement and an adaptive one, and displays that the latter provides better convergence rates. In addition, it has been shown that, adopting the linearization stopping criteria, a small number of Newton iterations are performed by the algorithm saving computational time. This work can be the starting point for an a posteriori error analysis for a wider range of problems, including problems with cohesive forces on the contact zone.

5 Extension of the a posteriori analysis and unified joint model

Contents

5.1 Introduction

In the context of dam modeling, different phenomena influence the behavior of an interface, notably friction, adhesion, rupture. As a consequence, with the aim of performing industrial studies, the contact problem presented in the previous chapter has to be enriched by introducing more complex constitutive relations for the interfaces of the structure. In this chapter, we propose some considerations regarding the extension of both a posteriori error estimation and constitutive relations analysis.

We start presenting an a posteriori error analysis for the contact problem between an elastic body and a rigid surface in which we add some cohesive forces on the contact surface, which is denoted with Γ_C. Notice that, in general, the constitutive relation involves both normal and tangential components of the surface density stress on Γ_C. As a consequence, some models with friction (e.g., Tresca friction [24]) are also included in this analysis. After an a posteriori error estimate result identifying the different components of the error, we explain how to adapt the equilibrated stress reconstruction of Section 4.5.2 by modifying the definition of the spaces used to construct the local problems for the stress reconstruction. The main ideas of the extension to the more general problem of a cohesive interface between two elastic bodies are also presented in Subsection 5.2.2.

Then, motivated by the results obtained by the joint constitutive relations presented in the Chapters 2 and 3, we conclude this work by proposing and analyzing a unified model for joints under the hypothesis of Standard Generalized Materials. The constitutive relation is

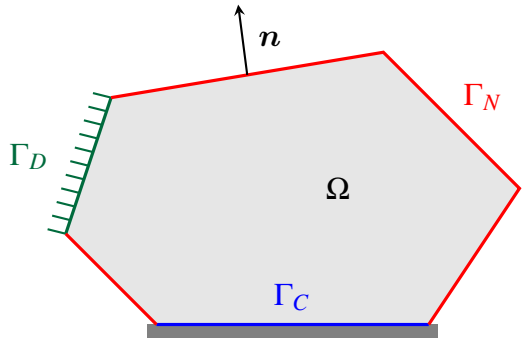


Figure 5.1 – Example of reference configuration for the cohesive contact problem. Some Dirichlet, Neumann and cohesive contact conditions are enforced on Γ_D (in *green*), on Γ_N (in *red*), and on Γ_C (in *blue*), respectively.

derived from a surface energy density which is the sum of a hyperelasto-plastic term and a hardening term which couples plasticity and damage.

5.2 A posteriori error estimation for problems with cohesive forces

The goal of this section is to propose how to extend the a posterior error analysis presented in the Chapter 4 to more complex problems, trying to maintain the same notation where possible. We consider two problems with a cohesive interface Γ_C: in the first case, Γ_C represents the contact zone between an elastic domain and a rigid surface, and in the second one, represents the joint inside an elastic domain (or, equivalently, the contact between two elastic bodies).

5.2.1 Cohesive contact problem with a rigid surface

Let $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, be a connected open domain with elastic behavior and that, in its reference configuration, is in contact with a rigid surface on a portion of this boundary, denoted with Γ_C, Figure 5.1. We suppose that, between the elastic body and the rigid foundation, there are some cohesive forces that possibly include friction. The behavior of this contact zone is determined by a constitutive relation which links the displacement $\boldsymbol{u}$ with the cohesive force $\boldsymbol{\sigma}\,\boldsymbol{n}$ on Γ_C, where $\boldsymbol{n}$ denotes the unit normal vector on $\partial\Omega$ pointing out of Ω. The rest of the boundary $\partial\Omega$ of the domain is divided into two parts Γ_D and Γ_N on which we enforce some (homogeneous) Dirichlet and (possibly non-homogeneous) Neumann boundary conditions. An example of domain is shown by Figure 5.1, in which different color identifies the subdivision of the boundary $\partial\Omega$. We denote with $\boldsymbol{\varepsilon}(\boldsymbol{v}) := \frac{1}{2}(\boldsymbol{\nabla v} + \boldsymbol{\nabla v}^\top)$ the strain tensor field, with $\boldsymbol{\sigma}(\boldsymbol{v}) \in \mathbb{R}^{d\times d}_{\text{sym}}$ the Cauchy stress tensor, with **div** the divergence operator acting row-wise on tensor valued functions, and with $\mathbb{E}$ the fourth order symmetric elasticity tensor such that, for all second-order tensor $\boldsymbol{\tau}$, $\mathbb{E}\boldsymbol{\tau} = \lambda \mathrm{Tr}\,(\boldsymbol{\tau})\boldsymbol{I}_d + 2\mu\boldsymbol{\tau}$, where λ and μ are the Lamé parameters. Assuming that the body is subjected to a volume force $\boldsymbol{f} \in \boldsymbol{L}^2(\Omega)$ and to a surface load $\boldsymbol{g}_N \in \boldsymbol{L}^2(\Gamma_N)$, the contact problem with cohesive forces is:

Find the displacement field $\boldsymbol{u}\colon \Omega \to \mathbb{R}^d$ such that

$$\begin{aligned} \mathbf{div}\,\boldsymbol{\sigma}(\boldsymbol{u}) + \boldsymbol{f} &= \mathbf{0} && \text{in } \Omega, && (5.1a)\\ \boldsymbol{\sigma}(\boldsymbol{u}) &= \mathbb{E}\,\boldsymbol{\varepsilon}(\boldsymbol{u}) && \text{in } \Omega, && (5.1b)\\ \boldsymbol{u} &= \mathbf{0} && \text{on } \Gamma_D, && (5.1c)\\ \boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n} &= \boldsymbol{g}_N && \text{on } \Gamma_N, && (5.1d)\\ \boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n} &= \boldsymbol{F}(\boldsymbol{u}) && \text{on } \Gamma_C, && (5.1e) \end{aligned}$$

where $\boldsymbol{F}$ is a function that fixes the interface constitutive relation on Γ_C and could also depend on other state variables. Introducing the natural decomposition into normal and tangential component

$$\boldsymbol{v} = v_n\boldsymbol{n} + \boldsymbol{v}_t \qquad \text{and} \qquad \boldsymbol{\sigma}(\boldsymbol{v})\boldsymbol{n} = \sigma_n(\boldsymbol{v})\boldsymbol{n} + \boldsymbol{\sigma}_t(\boldsymbol{v}), \tag{5.2}$$

for any displacement $\boldsymbol{v}\colon \Omega \to \mathbb{R}^d$ and for any surface force density $\boldsymbol{\sigma}(\boldsymbol{v})\boldsymbol{n}$, (5.1e) can be written as

$$\begin{cases} \sigma_n(\boldsymbol{u}) = F_n(\boldsymbol{u}) \\ \boldsymbol{\sigma}_t(\boldsymbol{u}) = \boldsymbol{F}_t(\boldsymbol{u}) \end{cases} \qquad \text{on } \Gamma_C.$$

Remark 5.1 (Comparison with the contact problem (4.2))**.** *The problem* (5.1) *differs from the contact problem* (4.2) *analyzed in the Chapter 4 only for the condition on the interface* Γ_C (5.1e)*, which generalizes the conditions* (4.2e)-(4.2f)*.*

Remark 5.2 (Notation for the joint constitutive relation)**.** *In the following, in order to distinguish the constitutive relations that determine the behavior of the domain and of the interface zone respectively, we will use the notation* $\boldsymbol{\sigma}(\boldsymbol{u})$ *in the first case and* $\boldsymbol{F}(\boldsymbol{u})$ *in the second one.*

Denoting by $\boldsymbol{H}^1_D(\Omega)$ the subspace of $\boldsymbol{H}^1(\Omega)$ incorporation the Dirichlet boundary condition (5.1c), and using a standard technique (i.e., multiplying both sides of (5.1a) by a test function $\boldsymbol{v} \in \boldsymbol{H}^1_D(\Omega)$, applying the Green formula, and using the boundary condition (5.1d)), we obtain the following problem: Find $\boldsymbol{u} \in \boldsymbol{H}^1_D(\Omega)$ such that

$$(\boldsymbol{\sigma}(\boldsymbol{u}), \boldsymbol{\varepsilon}(\boldsymbol{v})) - (\boldsymbol{F}(\boldsymbol{u}), \boldsymbol{v})_{\Gamma_C} = (\boldsymbol{f}, \boldsymbol{v}) + (\boldsymbol{g}_N, \boldsymbol{v})_{\Gamma_N} \qquad \forall \boldsymbol{v} \in \boldsymbol{H}^1_D(\Omega). \tag{5.3}$$

Let $\{\mathcal{T}_h\}_h$ be a family of conforming triangulations of Ω, indexed by the mesh size $h := \max_{T\in\mathcal{T}_h} h_T$, where h_T is the diameter of the element T. Each triangulation is assumed to be regular in the classical sense (see, e.g., [30, Eq. (3.1.43)]) and conformal to the subdivision of the boundary into Γ_D, Γ_N, and Γ_C (i.e., the interior of a boundary face cannot have non-empty intersection with more than one part of the subdivision). In our analysis, we will use again the mesh-related notations contained in Table 4.1.

On the mesh $\mathcal{T}_h$, we introduce the standard Lagrange finite element space of degree $p \geq 1$ with strongly enforced Dirichlet boundary condition:

$$\boldsymbol{V}_h := \left\{ \boldsymbol{v}_h \in \boldsymbol{H}^1_D(\Omega) \ : \ \boldsymbol{v}_h|_T \in \boldsymbol{\mathcal{P}}^p(T) \text{ for any } T \in \mathcal{T}_h \right\},$$

where $\boldsymbol{\mathcal{P}}^p(T) := [\mathcal{P}^p(T)]^d$, and $\mathcal{P}^p(T)$ denotes the restriction to X of d-variate polynomials of total degree $\leq p$. The discrete version of (5.3) is then: Find $\boldsymbol{u}_h \in \boldsymbol{V}_h$ such that

$$(\boldsymbol{\sigma}(\boldsymbol{u}_h), \boldsymbol{\varepsilon}(\boldsymbol{v}_h)) - (\boldsymbol{F}(\boldsymbol{u}_h), \boldsymbol{v}_h)_{\Gamma_C} = (\boldsymbol{f}, \boldsymbol{v}_h) + (\boldsymbol{g}_N, \boldsymbol{v}_h)_{\Gamma_N} \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_h. \tag{5.4}$$

In practice, this discrete problem is solved with a iterative method, e.g., the Newton method. At each iteration $k \geq 1$, we have the linearized problem: Find $\boldsymbol{u}_h^k \in \boldsymbol{V}_h$ such that

$$(\boldsymbol{\sigma}(\boldsymbol{u}_h^k), \boldsymbol{\varepsilon}(\boldsymbol{v}_h)) - (\boldsymbol{F}_{\text{lin}}^{k-1}(\boldsymbol{u}_h^k), \boldsymbol{v}_h)_{\Gamma_C} = (\boldsymbol{f}, \boldsymbol{v}_h) + (\boldsymbol{g}_N, \boldsymbol{v}_h)_{\Gamma_N} \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_h, \tag{5.5}$$

where $\boldsymbol{F}_{\text{lin}}^{k-1}(\boldsymbol{u}_h^k)$ is a linear approximation of $\boldsymbol{F}(\boldsymbol{u}_h^k)$. Assuming that the function is differentiable, we set

$$\boldsymbol{F}_{\text{lin}}^{k-1}(\boldsymbol{u}_h^k) \coloneqq \boldsymbol{F}(\boldsymbol{u}_h^{k-1}) + \frac{\partial \boldsymbol{F}(\boldsymbol{u}_h^{k-1})}{\partial \boldsymbol{u}_h^{k-1}}(\boldsymbol{u}_h^k - \boldsymbol{u}_h^{k-1}).$$

Otherwise, we can use a regularization technique as in Chapter 4 or another iterative method like semi-smooth versions Newton method. For sake of simplicity, from now on we suppose that $\boldsymbol{F}$ represents a differentiable hyperelastic (i.e., reversible) constitutive relation.

Remark 5.3 (Case of nonlinear domain)**.** *For sake of simplicity, we assume that the behavior of the body represented by Ω is homogeneous isotropic linear elastic (5.1b). We recall that, in the case of possible extensions to nonlinear behaviors of Ω, a linear approximation of $\boldsymbol{\sigma}(\boldsymbol{u}_h^k)$ has also to be introduced in the first term of (5.5). In addition, also the extension of the following a posteriori error analysis seems possible in the latter case in the spirit of [16].*

A posteriori error estimations via stress reconstruction

In the spirit of Section 4.3, we introduce a residual function starting from the discrete formulation (5.4), we define the notion of equilibrated stress reconstruction, and we propose an a posteriori error estimate for the error measured through a dual norm of the residual.

Denoting by $(\boldsymbol{H}_D^1(\Omega))^*$ the dual space of $\boldsymbol{H}_D^1(\Omega)$, for any $\boldsymbol{w}_h \in \boldsymbol{V}_h$ the *residual* $\mathcal{R}(\boldsymbol{w}_h) \in (\boldsymbol{H}_D^1(\Omega))^*$ is defined by

$$\langle \mathcal{R}(\boldsymbol{w}_h), \boldsymbol{v} \rangle \coloneqq (\boldsymbol{f}, \boldsymbol{v}) + (\boldsymbol{g}_N, \boldsymbol{v})_{\Gamma_N} - (\boldsymbol{\sigma}(\boldsymbol{w}_h), \boldsymbol{\varepsilon}(\boldsymbol{v})) + (\boldsymbol{F}(\boldsymbol{w}_h), \boldsymbol{v})_{\Gamma_C}, \tag{5.6}$$

where $\langle \cdot, \cdot \rangle$ is the duality pairing between $\boldsymbol{H}_D^1(\Omega)$ and $(\boldsymbol{H}_D^1(\Omega))^*$. The error committed approximating the exact solution $\boldsymbol{u}$ with $\boldsymbol{u}_h$ is then:

$$|||\mathcal{R}(\boldsymbol{w}_h)|||_* \coloneqq \sup_{\boldsymbol{v} \in \boldsymbol{H}_D^1(\Omega),\, |||\boldsymbol{v}|||=1} \langle \mathcal{R}(\boldsymbol{w}_h), \boldsymbol{v} \rangle, \tag{5.7}$$

where, as in the previous chapter (see (4.9) and (4.10)),

$$|||\boldsymbol{v}|||^2 \coloneqq \|\boldsymbol{\nabla} \boldsymbol{v}\|^2 + |\boldsymbol{v}|_{C,h}^2 \coloneqq \|\boldsymbol{\nabla} \boldsymbol{v}\|^2 + \sum_{F \in \mathcal{F}_h^C} \frac{1}{h_F} \|\boldsymbol{v}\|_F^2 \qquad \forall \boldsymbol{v} \in \boldsymbol{H}_D^1(\Omega). \tag{5.8}$$

We call *equilibrated stress reconstruction* any second order tensor $\boldsymbol{\sigma}_h$ that satisfies the properties 1., 2. and 3. of Definition 4.4, i.e., any $\mathbb{H}(\mathbf{div})$-conforming second order whose normal trace is square-integrable on each face $F \in \mathcal{F}_h^N \cup \mathcal{F}_h^C$, and that is locally in equilibrium with the volumetric and surface force $\boldsymbol{f}$ and $\boldsymbol{g}_N$. Moreover, we assume that, at each step $k \geq 1$ of the iterative method, we compute an equilibrated stress reconstruction $\boldsymbol{\sigma}_h^k$ which can be decomposed into two parts representing *discretization* and *linearization*, respectively:

$$\boldsymbol{\sigma}_h^k = \boldsymbol{\sigma}_{h,\text{dis}}^k + \boldsymbol{\sigma}_{h,\text{lin}}. \tag{5.9}$$

Remark 5.4. *In the case in which the function is not differentiable in one point and we apply a regularization technique as in Section 4.4, we would assume to have also a* regularization *part* $\boldsymbol{\sigma}^k_{h,\mathrm{reg}}$*, see Assumption 4.12.*

Theorem 5.5 (A posteriori error estimation distinguishing the error components)**.** *Let* $\boldsymbol{u}^k_h \in \boldsymbol{V}_h$ *be the solution of the linearized problem (5.5),* $\mathcal{R}(\boldsymbol{u}^k_h)$ *be the residual defined by (5.6), and* $\boldsymbol{\sigma}^k_h$ *be an equilibrated stress reconstruction which can be decomposed as (5.9). Then,*

$$\begin{aligned}&|||\mathcal{R}(\boldsymbol{u}^k_h)|||_* \\ &\quad\le \left[\sum_{T\in\mathcal{T}_h}\left((\eta^k_{\mathrm{osc,T}}+\eta^k_{\mathrm{str,T}}+\eta^k_{\mathrm{lin1,T}}+\eta^k_{\mathrm{Neu,T}})^2+(\eta^k_{\mathrm{nor,T}}+\eta^k_{\mathrm{tan,T}}+\eta^k_{\mathrm{lin2,T}})^2\right)\right]^{1/2},\end{aligned} \tag{5.10}$$

where $\eta_{\mathrm{osc,T}}$*,* $\eta_{\mathrm{str,T}}$*, and* $\eta_{\mathrm{Neu,T}}$ *are defined again by (4.35), while the remaining estimators are defined by:*

$$\eta_{\mathrm{lin1},T} \coloneqq \|\boldsymbol{\sigma}^k_{h,\mathrm{lin}}\|_T, \qquad \text{(linearization)} \tag{5.11a}$$

$$\eta_{\mathrm{lin2},T} \coloneqq \sum_{F\in\mathcal{F}_h^C} h_F^{1/2}\left(\|\sigma^k_{h,\mathrm{lin},n}\|_F+\|\boldsymbol{\sigma}^k_{h,\mathrm{lin},\boldsymbol{t}}\|_F\right), \qquad \text{(linearization)} \tag{5.11b}$$

$$\eta_{\mathrm{nor,T}} \coloneqq \sum_{F\in\mathcal{F}_T^C} h_F^{1/2}\left\|F_n(\boldsymbol{u}^k_h)-\sigma^k_{h,\mathrm{dis},n}\right\|_F, \qquad \text{(normal)} \tag{5.11c}$$

$$\eta_{\mathrm{tan,T}} \coloneqq \sum_{F\in\mathcal{F}_T^C} h_F^{1/2}\left\|\boldsymbol{F}_{\boldsymbol{t}}(\boldsymbol{u}^k_h)-\boldsymbol{\sigma}^k_{h,\mathrm{dis},\boldsymbol{t}}\right\|_F. \qquad \text{(tangential)} \tag{5.11d}$$

Moreover, if we define the global error estimators by

$$\eta^k_\bullet \coloneqq \left(\sum_{T\in\mathcal{T}_h}\left(\eta^k_{\bullet,T}\right)^2\right)^{1/2}, \tag{5.12}$$

then

$$|||\mathcal{R}(\boldsymbol{u}^k_h)|||_* \le \left[(\eta^k_{\mathrm{osc}}+\eta^k_{\mathrm{str}}+\eta^k_{\mathrm{reg1}}+\eta^k_{\mathrm{Neu}})^2+(\eta^k_{\mathrm{nor}}+\eta^k_{\mathrm{tan}}+\eta^k_{\mathrm{lin2}})^2\right]^{1/2}. \tag{5.13}$$

Proof. Fixing an element $\boldsymbol{v}\in\boldsymbol{H}^1_D(\Omega)$ such that $|||\boldsymbol{v}|||=1$ and proceeding as in the proof of Theorem 4.5, we obtain

$$\begin{aligned}\langle\mathcal{R}(\boldsymbol{u}^k_h),\boldsymbol{v}\rangle &= (\boldsymbol{f}+\mathbf{div}\,\boldsymbol{\sigma}^k_h,\boldsymbol{v})+(\boldsymbol{\sigma}^k_h-\boldsymbol{\sigma}(\boldsymbol{u}^k_h),\boldsymbol{\nabla}\boldsymbol{v})+(\boldsymbol{g}_N-\boldsymbol{\sigma}^k_h\boldsymbol{n},\boldsymbol{v})_{\Gamma_N}\\ &\quad+(\boldsymbol{F}(\boldsymbol{u}^k_h)-\boldsymbol{\sigma}^k_h\boldsymbol{n},\boldsymbol{v})_{\Gamma_C}.\end{aligned} \tag{5.14}$$

As regards the first three term, from the proof of Theorem 4.5 and using the decomposition of $\boldsymbol{\sigma}^k_h$ (5.9), we get

$$\begin{aligned}&(\boldsymbol{f}+\mathbf{div}\,\boldsymbol{\sigma}^k_h,\boldsymbol{v})+(\boldsymbol{\sigma}^k_h-\boldsymbol{\sigma}(\boldsymbol{u}^k_h),\boldsymbol{\nabla}\boldsymbol{v})+(\boldsymbol{g}_N-\boldsymbol{\sigma}^k_h\boldsymbol{n},\boldsymbol{v})_{\Gamma_N}\\ &\qquad\le\sum_{T\in\mathcal{T}_h}\left(\eta^k_{\mathrm{osc,T}}+\|\boldsymbol{\sigma}^k_h-\boldsymbol{\sigma}(\boldsymbol{u}^k_h)\|_T+\eta^k_{\mathrm{Neu,T}}\right)\|\boldsymbol{\nabla}\boldsymbol{v}\|_T\\ &\qquad\le\sum_{T\in\mathcal{T}_h}\left(\eta^k_{\mathrm{osc,T}}+\eta^k_{\mathrm{str,T}}+\eta^k_{\mathrm{lin1,T}}+\eta^k_{\mathrm{Neu,T}}\right)\|\boldsymbol{\nabla}\boldsymbol{v}\|_T.\end{aligned}$$

Now, let us consider the last term of (5.14). Applying the Cauchy-Schwarz inequality, the decomposition of the reconstruction $\boldsymbol{\sigma}_h^k$ (5.9), the triangle inequality and the definition of the estimators $\eta_{\mathrm{lin2},T}$ (5.11b), $\eta_{\mathrm{nor},T}$ (5.11c), and $\eta_{\mathrm{tan},T}$ (5.11d), we get

$$\begin{aligned}
(\boldsymbol{F}(\boldsymbol{u}_h^k) - \boldsymbol{\sigma}_h^k \boldsymbol{n}, \boldsymbol{v})_{\Gamma_C} &\le \sum_{T\in\mathcal{T}_h} \sum_{F\in\mathcal{F}_T^C} \left\|\boldsymbol{F}(\boldsymbol{u}_h^k) - \boldsymbol{\sigma}_h^k \boldsymbol{n}\right\|_F \|\boldsymbol{v}\|_F \\
&\le \sum_{T\in\mathcal{T}_h} \sum_{F\in\mathcal{F}_T^C} h_F^{1/2} \left\|\boldsymbol{F}(\boldsymbol{u}_h^k) - \boldsymbol{\sigma}_h^k \boldsymbol{n}\right\|_F |\boldsymbol{v}|_{C,T} \\
&\le \sum_{T\in\mathcal{T}_h} \sum_{F\in\mathcal{F}_T^C} h_F^{1/2} \left(\left\|\boldsymbol{F}(\boldsymbol{u}_h^k) - \boldsymbol{\sigma}_{h,\mathrm{dis}}^k \boldsymbol{n}\right\|_F + \left\|\boldsymbol{\sigma}_{h,\mathrm{lin}}^k\right\|_F\right) |\boldsymbol{v}|_{C,T} \\
&\le \sum_{T\in\mathcal{T}_h} \left(\eta_{\mathrm{nor,T}}^k + \eta_{\mathrm{tan,T}}^k + \eta_{\mathrm{lin2,T}}^k\right) |\boldsymbol{v}|_{C,T},
\end{aligned}$$

where $|\boldsymbol{v}|_{C,T}$ is the local counterpart of (4.10), i.e.,

$$|\boldsymbol{v}|_{C,T}^2 \coloneqq \sum_{F\in\mathcal{F}_T^C} \frac{1}{h_F} \|\boldsymbol{v}\|_F^2 \qquad \forall T \in \mathcal{T}_h.$$

Let $\eta_{a,T}^k \coloneqq \eta_{\mathrm{osc,T}}^k + \eta_{\mathrm{str,T}}^k + \eta_{\mathrm{lin1,T}}^k + \eta_{\mathrm{Neu,T}}^k$ and $\eta_{b,T}^k \coloneqq \eta_{\mathrm{nor,T}}^k + \eta_{\mathrm{tan,T}}^k + \eta_{\mathrm{lin2,T}}^k$. Recalling the definition of the dual norm of the residual (5.7) and of the norm $|||\cdot|||$ (5.8), and combining the above results and using the Cauchy-Schwarz inequality, we conclude

$$\begin{aligned}
|||\mathcal{R}(\boldsymbol{u}_h^k)|||_* &\le \sup_{\boldsymbol{v}\in\boldsymbol{H}_D^1(\Omega),\, |||\boldsymbol{v}|||=1} \left\{\sum_{T\in\mathcal{T}_h} \left(\eta_{a,T}^k \|\boldsymbol{\nabla}\boldsymbol{v}\|_T + \eta_{b,T}^k |\boldsymbol{v}|_{C,T}\right)\right\} \\
&\le \sup_{\boldsymbol{v}\in\boldsymbol{H}_D^1(\Omega),\, |||\boldsymbol{v}|||=1} \left\{\left(\sum_{T\in\mathcal{T}_h} \left((\eta_{a,T}^k)^2 + (\eta_{b,T}^k)^2\right)\right)^{1/2} \left(\sum_{T\in\mathcal{T}_h} \left(\|\boldsymbol{\nabla}\boldsymbol{v}\|_T^2 + |\boldsymbol{v}|_{C,T}^2\right)\right)^{1/2}\right\} \\
&= \left[\sum_{T\in\mathcal{T}_h} \left((\eta_{\mathrm{osc},T}^k + \eta_{\mathrm{str},T}^k + \eta_{\mathrm{lin1},T}^k + \eta_{\mathrm{Neu},T}^k)^2 + (\eta_{\mathrm{nor,T}}^k + \eta_{\mathrm{tan,T}}^k + \eta_{\mathrm{lin2,T}}^k)^2\right)\right]^{1/2}.
\end{aligned}$$

Finally, (5.13) is obtained from (5.10) applying twice the inequality $\sum_{T\in\mathcal{T}_h} \left(\sum_{i=1}^m a_{i,T}\right)^2 \le \left(\sum_{i=1}^m a_i\right)^2$ valid for all families of nonnegative real numbers $(a_{i,T})_{1\le i\le m, T\in\mathcal{T}_h}$ with $a_i \coloneqq \left(\sum_{T\in\mathcal{T}_h} a_{i,T}^2\right)^{1/2}$ for all $1 \le i \le m$. □

Adaptive resolution algorithm

Similarly to Subsection 4.4.2, we introduce an adaptive algorithm based on the local and global estimators involved in the error estimations (5.10) and (5.13), that fixes automatically the number of Newton iterations on every mesh refinement iteration. In particular, the stopping criterion for the linearization loop depends on a parameter $\gamma_{\mathrm{lin}} \in (0, 1)$ given by the user. For any element $T \in \mathcal{T}_h$, we define the local total estimator

$$\eta_{\mathrm{tot},T}^k \coloneqq \left[\left(\eta_{\mathrm{osc},T}^k + \eta_{\mathrm{str},T}^k + \eta_{\mathrm{lin1},T}^k + \eta_{\mathrm{Neu},T}^k\right)^2 + \left(\eta_{\mathrm{nor,T}}^k + \eta_{\mathrm{tan,T}}^k + \eta_{\mathrm{lin2,T}}^k\right)^2\right]^{1/2},$$

and the following local and total estimators

$$\eta^k_{\mathrm{lin},T} := \eta^k_{\mathrm{lin1},T} + \eta^k_{\mathrm{lin2},T}, \qquad \eta^k_{\mathrm{lin}} := \sum_{T\in\mathcal{T}_h} \eta^k_{\mathrm{lin},T}.$$

Algorithm 5.1 Adaptive algorithm

1: **choose** an initial function $\boldsymbol{u}_h^0 \in \mathbf{V}_h$, $\gamma_{\mathrm{lin}} \in (0,1)$
2: **repeat** {mesh refinement loop}
3: **set** $k = 0$
4: **repeat** {Newton linearization loop}
5: **set** $k = k + 1$
6: **setup** the operator $\boldsymbol{F}^{k-1}_{\mathrm{lin}}$ and the linear system
7: **compute** $\boldsymbol{u}^k_h$, $\boldsymbol{\sigma}^k_h$, and the local and global estimators
8: **until** $\eta^k_{\mathrm{lin}} \le \gamma_{\mathrm{lin}}(\eta^k_{\mathrm{osc}} + \eta^k_{\mathrm{str}} + \eta^k_{\mathrm{Neu}} + \eta^k_{\mathrm{nor}} + \eta^k_{\mathrm{tan}})$
9: **refine** the elements of the mesh where $\eta^k_{\mathrm{tot,T}}$ is higher
10: **update** data
11: **until** $\eta^k_{\mathrm{tot,T}}$ is distributed evenly over the mesh

As pointed out by Remark 4.14, it is possible to replace the global stopping criterion of Line 8 with its local version

$$\eta^k_{\mathrm{lin},T} \le \gamma_{\mathrm{lin},T}(\eta^k_{\mathrm{osc},T} + \eta^k_{\mathrm{str},T} + \eta^k_{\mathrm{Neu},T} + \eta^k_{\mathrm{nor},T} + \eta^k_{\mathrm{tan},T}) \qquad \forall T \in \mathcal{T}_h, \tag{5.15}$$

where $\gamma_{\mathrm{lin},T} \in (0,1)$ for any element T.

Stress reconstruction

At each Newton iteration, the stress reconstruction $\boldsymbol{\sigma}^k_h$ can be obtained assembling the solutions of local problems defined on patches $\omega_{\boldsymbol{a}}$ around the vertices of the mesh using the Brezzi–Douglas–Marini space.

We proceed as in Section 4.5. In particular, the spaces $\boldsymbol{\Sigma}^{\boldsymbol{a}}_h$, $\boldsymbol{U}^{\boldsymbol{a}}_h$ and $\boldsymbol{\Lambda}^{\boldsymbol{a}}_h$, are still defined by (4.41), (4.43) and (4.44), respectively. In addition, we set

$$\begin{aligned}
\boldsymbol{\Sigma}^{\boldsymbol{a},k}_{h,N,C,\mathrm{dis}} := \{\boldsymbol{\tau}_h \in \boldsymbol{\Sigma}_h(\omega_{\boldsymbol{a}}) \;:\; & \boldsymbol{\tau}_h\boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \boldsymbol{0} \text{ on } \partial\omega_{\boldsymbol{a}} \setminus \partial\Omega,\\
& \boldsymbol{\tau}_h\boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \Pi_{\boldsymbol{\Sigma}_h\boldsymbol{n}_{\omega_{\boldsymbol{a}}}}(\psi_{\boldsymbol{a}}\boldsymbol{g}_N) \text{ on } \partial\omega_{\boldsymbol{a}} \cap \Gamma_N \text{ and}\\
& \boldsymbol{\tau}_h\boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \Pi_{\boldsymbol{\Sigma}_h\boldsymbol{n}_{\omega_{\boldsymbol{a}}}}\big(\psi_{\boldsymbol{a}}\boldsymbol{F}(\boldsymbol{u}^k_h)\big) \text{ on } \partial\omega_{\boldsymbol{a}} \cap \Gamma_C\},\\
\boldsymbol{\Sigma}^{\boldsymbol{a},k}_{h,N,C,\mathrm{lin}} := \{\boldsymbol{\tau}_h \in \boldsymbol{\Sigma}_h(\omega_{\boldsymbol{a}}) \;:\; & \boldsymbol{\tau}_h\boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \boldsymbol{0} \text{ on } \partial\omega_{\boldsymbol{a}} \setminus \partial\Omega \text{ and on } \partial\omega_{\boldsymbol{a}} \cap \Gamma_N, \text{ and}\\
& \boldsymbol{\tau}_h\boldsymbol{n}_{\omega_{\boldsymbol{a}}} = \Pi_{\boldsymbol{\Sigma}_h\boldsymbol{n}_{\omega_{\boldsymbol{a}}}}\big(\psi_{\boldsymbol{a}}\big(\boldsymbol{F}^{k-1}_{\mathrm{lin}}(\boldsymbol{u}^k_h) - \boldsymbol{F}(\boldsymbol{u}^k_h)\big)\big) \text{ on } \partial\omega_{\boldsymbol{a}} \cap \Gamma_C\},
\end{aligned}$$

for any $\boldsymbol{a} \in \mathcal{V}^b_h$, and $\boldsymbol{\Sigma}^{\boldsymbol{a},k}_{h,N,C,\bullet} = \boldsymbol{\Sigma}^{\boldsymbol{a}}_h$ for $\bullet \in \{\mathrm{dis},\mathrm{lin}\}$ for any $\boldsymbol{a} \in \mathcal{V}^i_h$. Let $\boldsymbol{y}^k \in \boldsymbol{RM}^d$ be such that, for all $\boldsymbol{z} \in \boldsymbol{RM}^d$,

$$\begin{aligned}
(\boldsymbol{y}^k, \boldsymbol{z})_{\omega_{\boldsymbol{a}}} = {} & (-\psi_{\boldsymbol{a}}\boldsymbol{f} + \boldsymbol{\sigma}(\boldsymbol{u}^k_h)\boldsymbol{\nabla}\psi_{\boldsymbol{a}}, \boldsymbol{z})_{\omega_{\boldsymbol{a}}} - \left(\Pi_{\boldsymbol{\Sigma}_h\boldsymbol{n}_{\omega_{\boldsymbol{a}}}}(\psi_{\boldsymbol{a}}\boldsymbol{g}_N), \boldsymbol{z}\right)_{\partial\omega_{\boldsymbol{a}}\cap\Gamma_N}\\
& - \left(\Pi_{\boldsymbol{\Sigma}_h\boldsymbol{n}_{\omega_{\boldsymbol{a}}}}\big(\psi_{\boldsymbol{a}}\boldsymbol{F}(\boldsymbol{u}^k_h)\big), \boldsymbol{z}\right)_{\partial\omega_{\boldsymbol{a}}\cap\Gamma_C},
\end{aligned}$$

if $\boldsymbol{a} \in \mathcal{V}_h^b$, and $\boldsymbol{y}^k = \tilde{\boldsymbol{y}}^k = 0$ if $\boldsymbol{a} \in \mathcal{V}_h^i$.

Construction 5.6 (Equilibrated stress reconstruction distinguishing the error components)**.** *Let, for* $\bullet \in \{\mathrm{dis}, \mathrm{lin}\}$ *and any vertex* $\boldsymbol{a} \in \mathcal{V}_h$, $(\boldsymbol{\sigma}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{r}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{\lambda}_{h,\bullet}^{\boldsymbol{a},k}) \in \boldsymbol{\Sigma}_{h,N,C,\bullet}^{\boldsymbol{a},k} \times \boldsymbol{U}_h^{\boldsymbol{a}} \times \boldsymbol{\Lambda}_h^{\boldsymbol{a}}$ *be the solution to the following problem:*

$$
\begin{aligned}
(\boldsymbol{\sigma}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{\tau}_h)_{\omega_{\boldsymbol{a}}} + (\boldsymbol{r}_{h,\bullet}^{\boldsymbol{a},k}, \mathbf{div}\, \boldsymbol{\tau}_h)_{\omega_{\boldsymbol{a}}} + (\boldsymbol{\lambda}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{\tau}_h)_{\omega_{\boldsymbol{a}}} &= (\boldsymbol{\tau}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{\tau}_h)_{\omega_{\boldsymbol{a}}} && \forall \boldsymbol{\tau}_h \in \boldsymbol{\Sigma}_h^{\boldsymbol{a}}, \\
(\mathbf{div}\, \boldsymbol{\sigma}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{v}_h)_{\omega_{\boldsymbol{a}}} &= (\boldsymbol{v}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{v}_h)_{\omega_{\boldsymbol{a}}} && \forall \boldsymbol{v}_h \in \boldsymbol{U}_h^{\boldsymbol{a}}, \\
(\boldsymbol{\sigma}_{h,\bullet}^{\boldsymbol{a},k}, \boldsymbol{\mu}_h)_{\omega_{\boldsymbol{a}}} &= 0 && \forall \boldsymbol{\mu}_h \in \boldsymbol{\Lambda}_h^{\boldsymbol{a}},
\end{aligned}
$$

where

$$
\boldsymbol{\tau}_{h,\bullet}^{\boldsymbol{a},k} \coloneqq \begin{cases} \psi_{\boldsymbol{a}} \boldsymbol{\sigma}(\boldsymbol{u}_h^k) & \text{if } \bullet = \mathrm{dis}, \\ 0 & \text{if } \bullet \in \{\mathrm{lin}\}, \end{cases} \qquad \boldsymbol{v}_{h,\bullet}^{\boldsymbol{a},k} \coloneqq \begin{cases} -\psi_{\boldsymbol{a}} \boldsymbol{f} + \boldsymbol{\sigma}(\boldsymbol{u}_h^k) \boldsymbol{\nabla} \psi_{\boldsymbol{a}} - \boldsymbol{y}^k & \text{if } \bullet = \mathrm{dis}, \\ \boldsymbol{y}^k & \text{if } \bullet = \mathrm{lin}. \end{cases}
$$

Extending $\boldsymbol{\sigma}_{h,\bullet}^{\boldsymbol{a},k}$ *by zero outside the patch* $\omega_{\boldsymbol{a}}$*, we set* $\boldsymbol{\sigma}_{h,\bullet}^k \coloneqq \sum_{\boldsymbol{a} \in \mathcal{V}_h} \boldsymbol{\sigma}_{h,\bullet}^{\boldsymbol{a},k}$*, and we define* $\boldsymbol{\sigma}_h^k \coloneqq \boldsymbol{\sigma}_{h,\mathrm{dis}}^k + \boldsymbol{\sigma}_{h,\mathrm{lin}}^k$.

The following lemma sums up the main properties of the stress reconstruction $\boldsymbol{\sigma}_h^k$ obtained with the Construction 5.6. We omit the proof since it is similar to that of Lemma 4.17.

Lemma 5.7 (Properties of $\boldsymbol{\sigma}_h^k$)**.** *Let* $\boldsymbol{\sigma}_h^k$ *be defined by Construction 4.18. Then*

1) $\boldsymbol{\sigma}_{h,\mathrm{dis}}^k, \boldsymbol{\sigma}_{h,\mathrm{lin}}^k, \boldsymbol{\sigma}_h^k \in \mathbb{H}(\mathbf{div}, \Omega)$;

2) *For every* $T \in \mathcal{T}_h$ *and every* $\boldsymbol{v}_T \in \boldsymbol{\mathcal{P}}^{p-1}(T)$, $(\mathbf{div}\, \boldsymbol{\sigma}_h^k + \boldsymbol{f}, \boldsymbol{v}_T)_T = 0$;

3) *For every* $F \in \mathcal{F}_h^N$ *and every* $\boldsymbol{v}_F \in \boldsymbol{\mathcal{P}}^p(F)$, $(\boldsymbol{\sigma}_h^k \boldsymbol{n}, \boldsymbol{v}_F)_F = (\boldsymbol{g}_N, \boldsymbol{v}_F)_F$;

4) *For every* $F \in \mathcal{F}_h^C$ *and every* $\boldsymbol{v}_F \in \boldsymbol{\mathcal{P}}^p(F)$,

$$
(\boldsymbol{\sigma}_{h,\mathrm{dis}}^k \boldsymbol{n}, \boldsymbol{v}_F)_F = \left(\boldsymbol{F}(\boldsymbol{u}_h^k), \boldsymbol{v}_F\right)_F,
$$

and

$$
(\boldsymbol{\sigma}_{h,\mathrm{lin}}^k \boldsymbol{n}, \boldsymbol{v}_F)_F = \left(\boldsymbol{F}_{\mathrm{lin}}^{k-1}(\boldsymbol{u}_h^k) - \boldsymbol{F}(\boldsymbol{u}_h^k), \boldsymbol{v}_F\right)_F.
$$

Remark 5.8 (Alternative expressions of local estimators)**.** *The previous lemma allows one to rewrite the oscillation* (4.35a)*, Neumann* (4.35e)*, normal* (5.11c)*, and tangential* (5.11d) *as follows:*

$$
\eta_{\mathrm{osc},T}^k = \frac{h_T}{\pi} \left\| \boldsymbol{f} - \boldsymbol{\Pi}_T^{p-1} \boldsymbol{f} \right\|_T, \quad \eta_{\mathrm{Neu},T}^k = \sum_{F \in \mathcal{F}_T^C} C_{t,T,F} h_F^{1/2} \left\| \boldsymbol{g}_N - \boldsymbol{\Pi}_F^p \boldsymbol{g}_N \right\|_F,
$$

$$
\eta_{\mathrm{nor},T}^k = \sum_{F \in \mathcal{F}_T^C} h_F^{1/2} \left\| F_n(\boldsymbol{u}_h^k) - \Pi_F^p F_n(\boldsymbol{u}_h^k) \right\|_F, \quad \eta_{\mathrm{nor},T}^k = \sum_{F \in \mathcal{F}_T^C} h_F^{1/2} \left\| \boldsymbol{F}_{\boldsymbol{t}}(\boldsymbol{u}_h^k) - \boldsymbol{\Pi}_F^p \boldsymbol{F}_{\boldsymbol{t}}(\boldsymbol{u}_h^k) \right\|_F,
$$

where $\boldsymbol{\Pi}_T^{p-1}$*,* $\boldsymbol{\Pi}_F^p$*, and* Π_F^p *denote the* L^2*-orthogonal projectors on the polynomial spaces* $\boldsymbol{\mathcal{P}}^{p-1}(T)$*,* $\boldsymbol{\mathcal{P}}^p(F)$*, and* $\mathcal{P}^p(F)$*, respectively.*

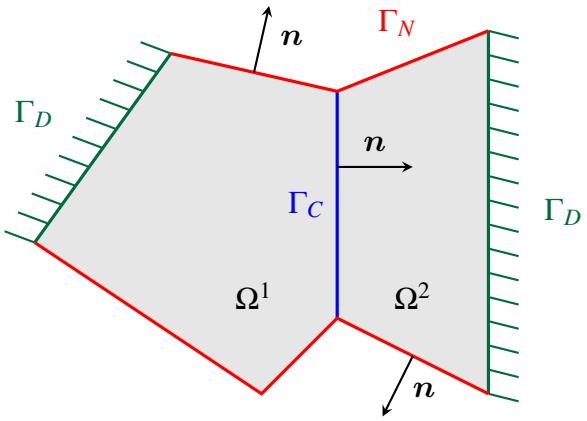


Figure 5.2 – Example of configuration for the two bodies interface problem. Some Dirichlet, Neumann and cohesive contact conditions are enforced on Γ_D (in green), on Γ_N (in red), and on Γ_C (in blue), respectively.

Efficiency

The approach of Section 4.7 can be adapted to prove the following results:

Theorem 5.9 (Local efficiency)**.** *Assume $d = 2$, let $\boldsymbol{u}_h^k \in \boldsymbol{V}_h$, let $\boldsymbol{\sigma}_h^k$ be the stress reconstruction of Construction 5.6, and assume that the local stopping criterion (5.15) holds. Then, for every element $T \in \mathcal{T}_h$,*

$$\eta_{\mathrm{osc},T}^k + \eta_{\mathrm{str},T}^k + \eta_{\mathrm{Neu},T}^k + \eta_{\mathrm{nor},T}^k + \eta_{\mathrm{tan},T}^k + \eta_{\mathrm{lin},T}^k \lesssim |||\mathcal{R}_{\mathcal{T}}(\boldsymbol{u}_h^k)|||_{*,\tilde{\omega}_T} + \eta_{\mathrm{osc},\mathcal{T}_T}^k + \eta_{\mathrm{Neu},\mathcal{T}_T}^k + \eta_{\mathrm{nor},\mathcal{T}_T}^k + \eta_{\mathrm{tan},\mathcal{T}_T}^k .$$

Theorem 5.10 (Global efficiency)**.** *Assume $d = 2$, let $\boldsymbol{u}_h^k \in \boldsymbol{V}_h$ and let $\boldsymbol{\sigma}_h^k$ be defined by Construction 5.6, assume that the global stopping criteria of Line 8 holds. Then,*

$$\eta_{\mathrm{osc}}^k + \eta_{\mathrm{str}}^k + \eta_{\mathrm{Neu}}^k + \eta_{\mathrm{nor}}^k + \eta_{\mathrm{tan}}^k + \eta_{\mathrm{lin}}^k \lesssim |||\mathcal{R}(\boldsymbol{u}_h^k)|||_* + \eta_{\mathrm{osc}}^k + \eta_{\mathrm{Neu}}^k + \eta_{\mathrm{nor}}^k + \eta_{\mathrm{tan}}^k.$$

Remark 5.11 (Extension to general constitutive relation)**.** *The considerations we have done can be extended also to quasi-static problems in which the constitutive relation (5.1e) is not reversible. This is the case in which $\boldsymbol{F}$ depends on some historical variable, like plasticity or damage. However, the results of local and global efficiency of the estimators are more questionable, since the contribution of the local estimators η_{nor}^k and η_{tan}^k could be not negligible for strongly non-regular evolutions at the contact level. We leave further investigation about the subject for future works.*

5.2.2 Two bodies interface problem with cohesive forces

In this section, we consider a general interface problem with cohesive forces determined by a constitutive relation, as presented at the beginning of Chapter 2.

Let $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, be a connected open domain, and let Γ_C be an interface dividing Ω into two subdomains Ω^1 and Ω^2, i.e., $\Omega \setminus \Gamma_C = \Omega^1 \cup \Omega^2$, that represent two elastic bodies. In this context, $\boldsymbol{n}$ denotes either the unit normal vector on $\partial\Omega$ pointing out of Ω, or the unit

normal vector pointing from Ω^1 towards Ω^2, see Figure 5.2. Choosing a local coordinate system $(\boldsymbol{n}, \boldsymbol{t}_1, \boldsymbol{t}_2)$ on the interface Γ_C, the displacement jump is defined by

$$\boldsymbol{\delta} = (\delta_n, \delta_{t_1}, \delta_{t_2}) := -[\![\boldsymbol{u}]\!] = -(\boldsymbol{u}^1 - \boldsymbol{u}^2),$$

where $\boldsymbol{u}\colon \Omega \to \mathbb{R}^d$ is the displacement of Ω, $\boldsymbol{u}^1 := \boldsymbol{u}|_{\Omega^1}$ and $\boldsymbol{u}^2 := \boldsymbol{u}|_{\Omega^2}$. Furthermore, the behavior of the interface is determinate by a relation connecting the displacement jump $\boldsymbol{\delta}$ and the surface force $\boldsymbol{\sigma}\,\boldsymbol{n}$. The boundary $\partial\Omega$ is divided into two non-overlapping part Γ_D and Γ_N in which we enforce some Dirichlet and Neumann conditions, respectively. We suppose that the domain is subjected to a load force $\boldsymbol{f} \in \boldsymbol{L}^2(\Omega \setminus \Gamma_C) := \boldsymbol{L}^2(\Omega^1) \times \boldsymbol{L}^2(\Omega^2)$, and to a surface load $\boldsymbol{g}_N \in \boldsymbol{L}^2(\Gamma_N)$. As before, $\boldsymbol{\varepsilon}(\boldsymbol{v}) := \frac{1}{2}(\boldsymbol{\nabla v} + \boldsymbol{\nabla v}^\top)$ is the strain tensor field, $\boldsymbol{\sigma}(\boldsymbol{v}) \in \mathbb{R}^{d\times d}_{\text{sym}}$ is the Cauchy stress tensor, **div** is the divergence operator acting row-wise on tensor valued functions, $\mathbb{E}$ is the fourth order symmetric elasticity tensor. Notice that now we admit that the Lamé parameters that characterize the elasticity tensor $\mathbb{E}$ could be different on the two subdomains Ω^1 and Ω^2. We consider the following interface problem with cohesive forces is: Find the displacement field $\boldsymbol{u}$ such that

$$\begin{aligned}
\mathbf{div}\,\boldsymbol{\sigma}(\boldsymbol{u}) + \boldsymbol{f} &= \boldsymbol{0} && \text{in } \Omega \setminus \Gamma_C, && (5.16a)\\
\boldsymbol{\sigma}(\boldsymbol{u}) &= \mathbb{E}\,\boldsymbol{\varepsilon}(\boldsymbol{u}) && \text{in } \Omega \setminus \Gamma_C, && (5.16b)\\
\boldsymbol{u} &= \boldsymbol{0} && \text{on } \Gamma_D, && (5.16c)\\
\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n} &= \boldsymbol{g}_N && \text{on } \Gamma_N, && (5.16d)\\
\boldsymbol{\sigma}(\boldsymbol{\delta})\boldsymbol{n} &= \boldsymbol{F}(\boldsymbol{\delta}) && \text{on } \Gamma_C, && (5.16e)
\end{aligned}$$

where $\boldsymbol{F}$ is a function which fix the interface constitutive relation on Γ_C. For sake of simplicity, we suppose that (5.16e) can be written as a differentiable hyperelastic (i.e., reversible) relation.

We define the space $\boldsymbol{V} := \boldsymbol{H}^1_D(\Omega^1) \times \boldsymbol{H}^1_D(\Omega^2)$, where $\boldsymbol{H}^1_D(\Omega^i)$ denotes the subspace of $\boldsymbol{H}^1(\Omega^i)$ incorporating the Dirichlet boundary condition on Γ_D, $i \in \{1, 2\}$. Using a standard technique (i.e., multiplying by a test function $\boldsymbol{v} \in \boldsymbol{V}$, applying the Green formula and the conditions (2.3d)-(2.3e)), we obtain the following problem: Find $\boldsymbol{u} \in \boldsymbol{V}$ such that

$$(\boldsymbol{\sigma}(\boldsymbol{u}), \boldsymbol{\varepsilon}(\boldsymbol{v}))_{\Omega\setminus\Gamma_C} + (\boldsymbol{F}(\boldsymbol{\delta}), \boldsymbol{\delta}^{\boldsymbol{v}})_{\Gamma_C} = (\boldsymbol{f}, \boldsymbol{v})_{\Omega\setminus\Gamma_C} + (\boldsymbol{g}_N, \boldsymbol{v})_{\Gamma_N} \qquad \forall \boldsymbol{v} \in \boldsymbol{V},$$

where $\boldsymbol{\delta}^{\boldsymbol{v}} := -[\![\boldsymbol{v}]\!] = -(\boldsymbol{v}^1 - \boldsymbol{v}^2)$.

Let $\{\mathcal{T}_h\}_h$ be a family of conforming triangulation of Ω such that each mesh $\mathcal{T}_h$ is conformal to the subdivision of the boundary into Γ_D and Γ_n, and of the domain into Ω^1 and Ω^2, i.e., $\mathcal{T}_h = \mathcal{T}_h^1 \cup \mathcal{T}_h^2$, where $\mathcal{T}_h^i$ is a triangulation of Ω^i, $i \in \{1, 2\}$. Defining

$$\boldsymbol{V}_h := \boldsymbol{V}_h^1 \times \boldsymbol{V}_h^2, \qquad \text{where} \qquad \boldsymbol{V}_h^i := \left\{\boldsymbol{v}_h^i \in \boldsymbol{H}^1_D(\Omega^i) \;:\; \boldsymbol{v}_h^i|_T \in \boldsymbol{\mathcal{P}}^p(T) \text{ for any } T \in \mathcal{T}_h^i\right\},$$

$i \in \{1, 2\}$, the discrete formulation reads: Find $\boldsymbol{u}_h \in \boldsymbol{V}_h$ such that

$$(\boldsymbol{\sigma}(\boldsymbol{u}_h), \boldsymbol{\varepsilon}(\boldsymbol{v}_h))_{\Omega\setminus\Gamma_C} + (\boldsymbol{F}(\boldsymbol{\delta}_h), \boldsymbol{\delta}_h^{\boldsymbol{v}})_{\Gamma_C} = (\boldsymbol{f}, \boldsymbol{v}_h)_{\Omega\setminus\Gamma_C} + (\boldsymbol{g}_N, \boldsymbol{v}_h)_{\Gamma_N} \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_h, \quad (5.17)$$

where $\boldsymbol{\delta}_h := -[\![\boldsymbol{u}_h]\!]$ and $\boldsymbol{\delta}_h^{\boldsymbol{v}} := -[\![\boldsymbol{v}_h]\!]$. This problem can be solved in practice adopting the Newton method. At each step $k \geq 1$, we get: Find $\boldsymbol{u}_h^k \in \boldsymbol{V}_h$ such that

$$(\boldsymbol{\sigma}(\boldsymbol{u}_h^k), \boldsymbol{\varepsilon}(\boldsymbol{v}_h)) + (\boldsymbol{F}_{\text{lin}}^{k-1}(\boldsymbol{\delta}_h^k), \boldsymbol{\delta}_h^{\boldsymbol{v}})_{\Gamma_C} = (\boldsymbol{f}, \boldsymbol{v}_h) + (\boldsymbol{g}_N, \boldsymbol{v}_h)_{\Gamma_N} \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_h, \qquad (5.18)$$

where

$$\boldsymbol{F}_{\text{lin}}^{k-1}(\boldsymbol{\delta}_h^k) \coloneqq \boldsymbol{F}(\boldsymbol{\delta}_h^{k-1}) + \frac{\partial \boldsymbol{F}(\boldsymbol{\delta}_h^{k-1})}{\partial \boldsymbol{\delta}_h^{k-1}}(\boldsymbol{\delta}_h^k - \boldsymbol{\delta}_h^{k-1}).$$

A posteriori error estimation via stress reconstruction

Before showing the a posteriori error estimation, we have to define the two main ingredients: the measure of the error we commit when the exact solution $\boldsymbol{u}$ is approximated with $\boldsymbol{u}_h^k$, and the notion of equilibrated stress reconstruction.

As in the Subsection 5.2.1, the error is measured with the dual norm of a residual operator defined directly from the problem formulation:

$$|||\mathcal{R}(\boldsymbol{w}_h)|||_* \coloneqq \sup_{\boldsymbol{v}\in\boldsymbol{V},\, |||\boldsymbol{v}|||=1} \langle \mathcal{R}(\boldsymbol{w}_h), \boldsymbol{v}\rangle\,, \tag{5.19}$$

where $\boldsymbol{w}_h \in \boldsymbol{V}_h$,

$$\langle \mathcal{R}(\boldsymbol{w}_h), \boldsymbol{v}\rangle \coloneqq (\boldsymbol{f}, \boldsymbol{v})_{\Omega\setminus\Gamma_C} + (\boldsymbol{g}_N, \boldsymbol{v})_{\Gamma_N} - (\boldsymbol{\sigma}(\boldsymbol{u}), \boldsymbol{\varepsilon}(\boldsymbol{v}))_{\Omega\setminus\Gamma_C} - (\boldsymbol{F}(\boldsymbol{\delta}), \boldsymbol{\delta}^{\boldsymbol{v}})_{\Gamma_C}, \tag{5.20}$$

for any $\boldsymbol{v} \in \boldsymbol{V}$, is the *residual*, and

$$|||\boldsymbol{v}|||^2 \coloneqq \|\boldsymbol{\nabla}\boldsymbol{v}\|^2_{\Omega\setminus\Gamma_C} + |\boldsymbol{v}|^2_{C,h}\,, \qquad \text{with} \quad |\boldsymbol{v}|^2_{C,h} \coloneqq \sum_{F\in\mathcal{F}_h^C} \frac{1}{h_F}\|\boldsymbol{\delta}^{\boldsymbol{v}}\|^2_F.$$

Definition 5.12 (Equilibrated stress reconstruction)**.** A *equilibrated stress reconstruction* is any second-order tensor $\boldsymbol{\sigma}_h$ such that:

1. $\boldsymbol{\sigma}_h \in \mathbb{H}(\mathbf{div}, \Omega)$,
2. $(\mathbf{div}\,\boldsymbol{\sigma}_h + \boldsymbol{f}, \boldsymbol{v})_T = 0$ for every $\boldsymbol{v} \in \boldsymbol{\mathcal{P}}^0(T)$ and every $T \in \mathcal{T}_h$,
3. $(\boldsymbol{\sigma}_h\boldsymbol{n})|_F \in \boldsymbol{L}^2(F)$ for every $F \in \mathcal{F}_h^N \cup \mathcal{F}_h^C$, and $(\boldsymbol{\sigma}_h\boldsymbol{n}, \boldsymbol{v})_F = (\boldsymbol{g}_N, \boldsymbol{v})_F$ for every $\boldsymbol{v} \in \boldsymbol{\mathcal{P}}^0(F)$ and every $F \in \mathcal{F}_h^N$.

Assumption 5.13 (Decomposition of the stress reconstruction)**.** *Let $k \geq 1$ and let $\boldsymbol{\sigma}_h^k$ be an equilibrated stress reconstruction in the sense of Definition 5.12. Then, $\boldsymbol{\sigma}_h^k$ can be decomposed into two part representing* discretization *and* linearization*:*

$$\boldsymbol{\sigma}_h^k = \boldsymbol{\sigma}_{h,\text{dis}}^k + \boldsymbol{\sigma}_{h,\text{lin}}^k.$$

Theorem 5.14 (Basic a posteriori error estimate)**.** *Let $\boldsymbol{u}_h^k \in \boldsymbol{V}_h$ be the solution of the linearized problem (5.18), $\mathcal{R}(\boldsymbol{u}_h^k)$ be the residual defined by (5.20), and $\boldsymbol{\sigma}_h^k$ be an equilibrated stress reconstruction satisfying Assumption 5.13. Then,*

$$|||\mathcal{R}(\boldsymbol{u}_h^k)|||_* \leq \left[\sum_{T\in\mathcal{T}_h}\left((\eta^k_{\text{osc,T}} + \eta^k_{\text{str,T}} + \eta^k_{\text{lin1,T}} + \eta^k_{\text{Neu,T}})^2 + (\eta^k_{\text{nor,T}} + \eta^k_{\text{tan,T}} + \eta^k_{\text{lin2,T}})^2\right)\right]^{1/2},$$

where $\eta_{\mathrm{osc,T}}$, $\eta_{\mathrm{str,T}}$, *and* $\eta_{\mathrm{Neu,T}}$ *are defined by* (4.35), $\eta_{\mathrm{lin1},T}$, *and* $\eta_{\mathrm{lin2},T}$ *by* (5.11a)–(5.11b), *and*

$$\eta_{\mathrm{nor,T}} \coloneqq \sum_{F\in\mathcal{F}_T^C} h_F^{1/2}\left\|F_n(\boldsymbol{\delta}_h^k) - \sigma_{h,\mathrm{dis},n}^k\right\|_F, \qquad \text{and} \qquad \eta_{\mathrm{tan,T}} \coloneqq \sum_{F\in\mathcal{F}_T^C} h_F^{1/2}\left\|\boldsymbol{F}_{\boldsymbol{t}}(\boldsymbol{\delta}_h^k) - \boldsymbol{\sigma}_{h,\mathrm{dis},\boldsymbol{t}}^k\right\|_F.$$

Moreover,

$$|||\mathcal{R}(\boldsymbol{u}_h^k)|||_* \leq \left[(\eta_{\mathrm{osc}}^k + \eta_{\mathrm{str}}^k + \eta_{\mathrm{reg1}}^k + \eta_{\mathrm{Neu}}^k)^2 + (\eta_{\mathrm{nor}}^k + \eta_{\mathrm{tan}}^k + \eta_{\mathrm{lin2}}^k)^2\right]^{1/2},$$

where the estimators $\eta_\bullet^k$ *are the global counterpart of* $\eta_{\bullet,T}^k$, *see* (5.12).

Proof. It is possible to proceed in the same spirit of the proofs of Theorems 4.5 and 5.5. We only remark that, in this context, the Green formula (4.13) becomes

$$(\boldsymbol{\sigma}_h, \boldsymbol{\nabla}\boldsymbol{v}) + (\mathbf{div}\,\boldsymbol{\sigma}_h, \boldsymbol{v}) = (\boldsymbol{\sigma}_h\boldsymbol{n}, \boldsymbol{v})_{\Gamma_N} - (\sigma_{h,n}, \delta_n^{\boldsymbol{v}})_{\Gamma_C} - (\boldsymbol{\sigma}_{h,\boldsymbol{t}}, \boldsymbol{\delta}_{\boldsymbol{t}}^{\boldsymbol{v}})_{\Gamma_C} \qquad \forall \boldsymbol{v} \in \boldsymbol{V}.$$

□

Remark 5.15 (Adaptive algorithm)**.** *Algorithm 5.1 can be easily applied also in this context, recalling that, although the notation is the same, the definition of most of the elements (e.g.,* $\boldsymbol{u}_h^k$, $\boldsymbol{\sigma}_h^k$, $\boldsymbol{F}_{\mathrm{lin}}^{k-1}$, *estimators) has changed.*

Remark 5.16 (Stress reconstruction)**.** *In this work, we do not present an explicitly the construction of* $\boldsymbol{\sigma}_h^k$, *or the proof of efficiency of the estimators for this context. However, the idea is to use the same approach of the Subsection 5.2.1.*

5.3 Unified nonlinear model for interfaces

In the previous section, we have shown how to extend the a posteriori error analysis via equilibrated stress reconstruction to cohesive interface problems. We have also briefly indicated that we have to choose carefully the constitutive relation, because strongly non-regular solutions could bring some difficulties for the efficiency results, see Remark 5.11. With this aim, we focus on constitutive relations coming from the Standard Generalized Materials framework, presented in Subsection 2.1.2, adapted to joint modeling, since, in the context of geomaterials, it leads to robust numerical results. We conclude this work, then, with a discussion about a constitutive relation that unifies the joint models presented in the Chapters 2 and 3.

As in the Subsection 5.2.2, $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, is a connected open domain, Γ_C is an interface dividing Ω into two subdomains Ω^1 and Ω^2, $\boldsymbol{n}$ is the unit normal vector pointing from Ω^1 towards Ω^2, and the displacement jump is defined by $\boldsymbol{\delta} \coloneqq -[\![\boldsymbol{u}]\!] = -(\boldsymbol{u}^1 - \boldsymbol{u}^2)$, see Figure 5.2 or 2.3a. The behavior of the interface zone is determined by a constitutive relation $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\delta})$. The latter is derived using the formalism of Standard Generalized Materials applied to joints, i.e., we have to define 1) the state variables, 2) a surface energy density function, and 3) a reversibility domain for each dissipative state variable.

Starting from a standard elasto-plastic model we add:

- a hardening/softening term which couples plasticity and damage, as in Chapter 2;
- a nonlinear hyperelastic dependence on the normal displacement δ_n in the normal and tangential rigidity coefficients, as in Chapter 3.

As a consequence, the state variables are the displacement jump $\boldsymbol{\delta} \in \mathbb{R}^d$, its plastic component $\boldsymbol{p} \in \mathbb{R}^d$ and the damage variable $\alpha \in [0,1]$, and the surface energy density function is

$$\begin{aligned}\psi(\boldsymbol{\delta}, \boldsymbol{p}, \alpha) &= \frac{1}{2}(\boldsymbol{\delta} - \boldsymbol{p})^\top \mathbb{E}(\delta_n)(\boldsymbol{\delta} - \boldsymbol{p}) + \frac{1}{2}\boldsymbol{p}^\top \mathbb{H}(\alpha)\boldsymbol{p} \\ &= K_n(\delta_n)\frac{(\delta_n - p_n)^2}{2} + K_t(\delta_n)\frac{\|\boldsymbol{\delta_t} - \boldsymbol{p_t}\|^2}{2} + A_n(\alpha)\frac{p_n^2}{2} + A_t(\alpha)\frac{\|\boldsymbol{p_t}\|^2}{2},\end{aligned} \tag{5.21}$$

where $\|\cdot\|$ is the Euclidean norm of $\mathbb{R}^{d-1}$,

$$\mathbb{E}(\delta_n) := \begin{pmatrix} K_n(\delta_n) & 0 & 0 \\ 0 & K_t(\delta_n) & 0 \\ 0 & 0 & K_t(\delta_n) \end{pmatrix}, \qquad \mathbb{H} := \begin{pmatrix} A_n(\alpha) & 0 & 0 \\ 0 & A_t(\alpha) & 0 \\ 0 & 0 & A_t(\alpha) \end{pmatrix}.$$

In particular, the rigidity coefficients are again given by (3.26), and the damage functions by (2.19), i.e.,

$$\begin{cases} K_n(\delta_n) = \dfrac{K_{n,0}}{2K_{n,0}\beta_n\delta_n + 1}, \\ K_t(\delta_n) = \dfrac{K_{t,0}}{2K_{t,0}\beta_t\delta_n + 1}, \end{cases} \quad \text{and} \quad \begin{cases} A_n(\alpha) = B_n\dfrac{(1-\alpha)^{m_1}}{\alpha^{m_2}}, \\ A_t(\alpha) = B_t\dfrac{(1-\alpha)^{m_1}}{\alpha^{m_2}}. \end{cases} \tag{5.22}$$

We recall that $K_{n,0}$ and $K_{t,0}$ are two positive parameters representing the normal and tangential rigidity coefficient for $\delta_n = 0$, respectively, $\beta_n \geq 0$ and $\beta_t \geq 0$ are the hyperelastic coefficients, $B_n, B_t > 0$, and $0 < m_2 < 1 < m_1$ are the damage parameters. The stress $\boldsymbol{\sigma}$, and the generalized forces $\boldsymbol{X}$ and Y are obtained by derivation of the surface energy density function ψ (5.21):

$$\begin{cases} \sigma_n = \dfrac{K_{n,0}}{2K_{n,0}\beta_n\delta_n + 1}(\delta_n - p_n) - \beta_n\dfrac{K_{n,0}^2}{(2K_{n,0}\beta_n\delta_n + 1)^2}(\delta_n - p_n)^2 \\ \qquad\quad - \beta_t\dfrac{K_{t,0}^2}{(2K_{t,0}\beta_t\delta_n + 1)^2}\|\boldsymbol{\delta_t} - \boldsymbol{p_t}\|^2, \end{cases} \tag{5.23a}$$

$$\boldsymbol{\sigma_t} = \frac{K_{t,0}}{2K_{t,0}\beta_t\delta_n + 1}(\boldsymbol{\delta_t} - \boldsymbol{p_t}), \tag{5.23b}$$

$$\begin{cases} X_n = \dfrac{K_{n,0}}{2K_{n,0}\beta_n\delta_n + 1}(\delta_n - p_n) - A_n(\alpha)p_n, \\ \boldsymbol{X_t} = \dfrac{K_{t,0}}{2K_{t,0}\beta_t\delta_n + 1}(\boldsymbol{\delta_t} - \boldsymbol{p_t}) - A_t(\alpha)\boldsymbol{p_t}, \end{cases} \tag{5.24}$$

and

$$Y = -A_n'(\alpha)\frac{p_n^2}{2} - A_t'(\alpha)\frac{\|\boldsymbol{p_t}\|^2}{2}.$$

Combining (5.23) and (5.24), we achieve the relation between the stress and the plasticity generalized force:

$$\begin{cases} \sigma_n = X_n + A_n(\alpha)p_n - \beta_n(X_n + A_n(\alpha)p_n)^2 - \beta_t \left\|\boldsymbol{X}_t + A_t(\alpha)\boldsymbol{p_t}\right\|^2, & \text{(5.25a)} \\ \boldsymbol{\sigma}_t = \boldsymbol{X}_t + A_t(\alpha)\boldsymbol{p_t}. & \text{(5.25b)} \end{cases}$$

In addition, we suppose that the plasticity criterion in the space of plasticity generalized forces $\boldsymbol{X}$ is linear, i.e., of Drucker–Prager type:

$$f_{\boldsymbol{X}}(\boldsymbol{X}) := \|\boldsymbol{X}_t\| + \bar{a}X_n - \bar{b} = 0, \tag{5.26}$$

where $\bar{a} > 0$ and $\bar{b} \geq 0$. Then, the corresponding reversibility domain,

$$\mathbb{K}_{\boldsymbol{X}} := \left\{\boldsymbol{X}^* \in \mathbb{R}^d \ : \ \|\boldsymbol{X}_t\| + \bar{a}X_n - \bar{b} \leq 0\right\},$$

is a cone with axis $\boldsymbol{X_t} = \boldsymbol{0}$ and vertex $(\bar{b}/\bar{a}, \boldsymbol{0})$. The evolution of plasticity follows the normal flow rule, which is given explicitly again by (3.37), in addition with (3.38) outside of the singular point of $\partial\mathbb{K}_{\boldsymbol{X}}$. Furthermore, using (5.25) combined with (5.26), we achieve the expression of the plasticity criterion in the stress space:

$$\begin{aligned} f_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}) := \beta_n\|\boldsymbol{\sigma_t} + A_t(\alpha)\boldsymbol{p_t}\|^2 + \bar{a}^2\beta_t\|\boldsymbol{\sigma_t}\|^2 + \left(\bar{a} - 2\beta_n(\bar{b} + \bar{a}A_n(\alpha)p_n)\right)\|\boldsymbol{\sigma_t} + A_t(\alpha)\boldsymbol{p_t}\| \\ + \bar{a}\sigma_n - (\bar{b} + \bar{a}A_n(\alpha)p_n)\left(\bar{a} - \beta_n(\bar{b} + \bar{a}A_n(\alpha)p_n)\right) = 0 \end{aligned}$$

Notice that, in the initial configuration (i.e., $\alpha = 0$) and in the asymptotic one (i.e., $\alpha \approx 1$), we recover the pure hyperelasto-plastic model presented and analyzed in Chapter 3. In particular, the analysis on initial hyperelastic evolutions (e.g., compression tests, first stage of traction and shear tests) is still valid, and the considerations on the phases in which plasticity evolves in the hyperelasto-plastic model (e.g., constant evolution of stress in both traction and shear tests, reduction of the slope of dilatancy, saturation of dilatancy in cyclic shear loading), in this unified constitutive relation, hold asymptotically.

Finally, the damage criterion and corresponding reversibility domain are defined as in Chapter 2 (see (2.34)):

$$f_Y(Y) := Y - D_1 \qquad \text{and} \qquad \mathbb{K}_Y = \left\{Y^* \ : \ Y^* \leq D_1\right\}.$$

Remark 5.17 (Simultaneous evolution of plasticity and damage)**.** *The statements of Proposition 2.6 remain valid also for this unified model if we replace the friction coefficient μ of 3) with the linear coefficient $\bar{a}$. Indeed, in the proof, we have only used elements of the model that have not been changed: the damage criterion, the definition of Y and of the damage functions $A_s(\alpha)$, $s \in \{n, t\}$, and the normal flow rule.*

5.3.1 Numerical results on typical tests

In this subsection, we show briefly the response during standard test of the unified constitutive relation. For simplicity, we use the notation of the case $d = 2$, recalling that the results remain valid also for the case $d = 3$ with $\delta_{t_2} = 0$.

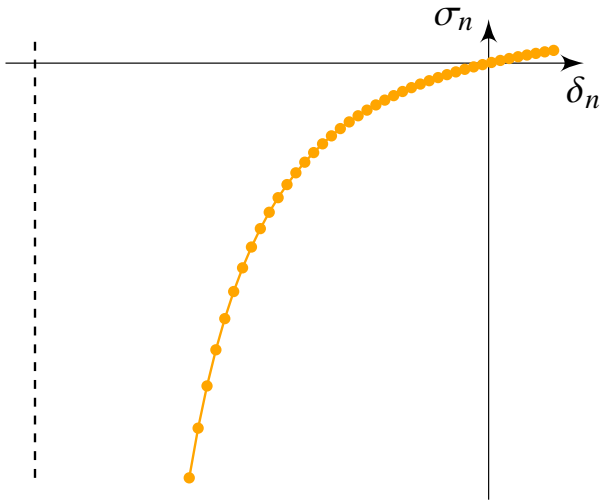


Figure 5.3 – Example of (hyperelastic) evolution of the normal stress σ_n during a compression test.

Compression test

During compression tests, we preserve the main properties of the hyperelasto-plastic model of Section 3.5: the behavior is pure hyperelastic since $\boldsymbol{\sigma} \in \operatorname{Int} \mathbb{K}_{\boldsymbol{\sigma}}$, there is a lower limiting value for the normal displacement δ_n, and the evolution is strongly nonlinear and characterized by the following relation:

$$\sigma_n = \frac{K_{n,0}}{2K_{n,0}\beta_n\delta_n + 1}\delta_n - \beta_n \frac{K_{n,0}^2}{(2K_{n,0}\beta_n\delta_n + 1)^2}\delta_n^2, \qquad \text{with } \delta_n > \frac{1}{2K_{n,0}\beta_n}.$$

Figure 3.6 show the evolution of the normal stress σ_n in compression. In addition, in this figure, we have also extended the curve for a small traction assuming that the behavior is still hyperelastic ($\bar{b} > 0$ and $\sigma_n \leq {}^{\bar{b}(\bar{a}-\beta_n\bar{b})}/_{\bar{a}^2}$), in order to point out that the evolution at the origin is continuous and differentiable.

Shear test with fixed compression

After an initial compressive phase that ends when $\sigma_n = -\sigma_n^{\text{com}}$, i.e., when

$$\delta_n = \frac{1}{2K_{n,0}\beta_n}\left(-1 + \frac{1}{\sqrt{4\beta_n\sigma_n^{\text{com}} + 1}}\right),$$

we increase monotonically the tangential displacement δ_t. We have a nonlinear elastic phase, followed by a phase in which plasticity and damage evolve. Figure 5.4 represents these phases in *orange* and *green*, respectively. In particular, the hardening and softening behavior of the shear stress is a direct consequence of the movement of the elastic domain $\mathbb{K}_{\boldsymbol{\sigma}}$. In addition, the plot on the *right* of Figure 5.4 also compares the evolution of dilatancy with and without hyperelasticity (i.e., $(\beta_n, \beta_t) \neq (0,0)$ and $(\beta_n, \beta_t) = (0,0)$, respectively, the latter case is represented in *red*), pointing out that the introduction of hyperelasticity enables the reduction of its asymptotic slope. One can also show that the saturation of dilatancy can be

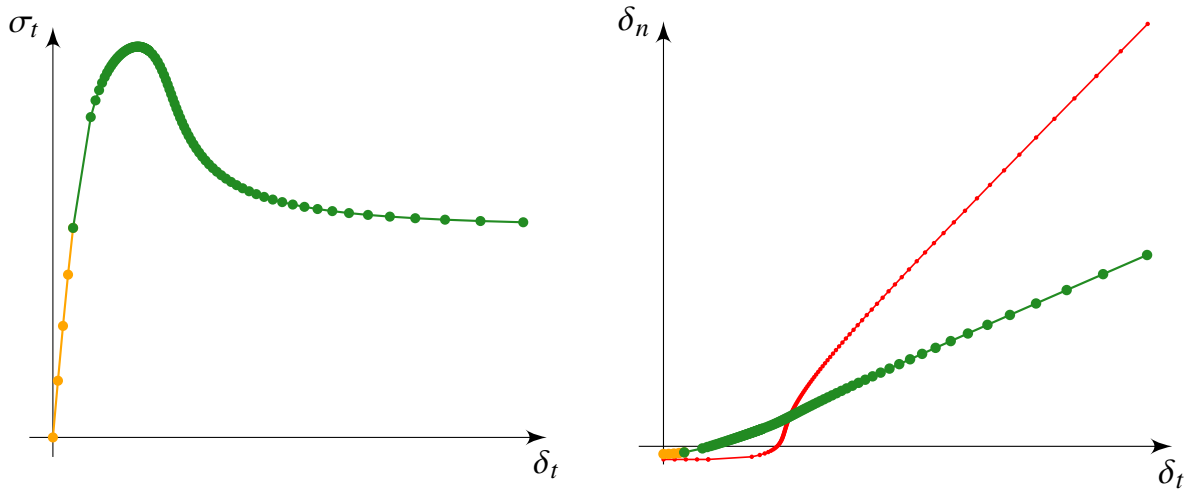


Figure 5.4 – Example of evolution of the shear stress σ_t (*left*) and the normal displacement δ_n (*right*) during a shear test with fixed compression. The phases without and with evolution of plasticity are represented in *orange* and in *green*, respectively. In particular, the plot on the *right* also compares the evolution of dilatancy obtained with and without hyperelasticity; the latter case corresponds to $(\beta_n, \beta_t) = (0, 0)$, and it is represented in *red*.

recover during cycling loadings, as in Figure 3.9. In general, it is not easy to write explicitly the expressions of $\delta_t(\alpha)$, $\sigma_t(\alpha)$, and $\delta_n(\alpha)$ because of the coupling between normal and tangential displacement components in σ_t (5.23a). As a consequence, a rigorous parametric analysis, especially on the parameters related to damage, is difficult to perform and we leave it for future developments of this work.

Traction test

Figure 5.5 shows the evolution of the normal stress during a traction test, in which we increase the value of δ_n maintaining $\delta_t = 0$. If $\bar{b} > 0$, there is an initial hyperelastic phase that stops when σ_n reaches the vertex of the domain $\mathbb{K}_{\boldsymbol{\sigma}}$, i.e., when $\sigma_n = \dfrac{\bar{b}(\bar{a} - \beta_n \bar{b})}{\bar{a}^2}$. At this point the value of the normal displacement is

$$\delta_n = \frac{1}{2K_{n,0}\beta_n}\left(-1 + \frac{\bar{a}}{\sqrt{4\beta_n \bar{b}(\bar{a} - \beta_n \bar{b}) + \bar{a}^2}}\right).$$

Then, plasticity and damage evolve, and the expressions can be derived from the following relations:

$$\begin{cases} \sigma_n = X_n + A_n(\alpha)p_n - \beta_n(X_n + A_n(\alpha)p_n)^2, & \text{(5.27a)} \\ X_n = \dfrac{K_{n,0}}{2K_{n,0}\beta_n\delta_n + 1}(\delta_n - p_n) - A_n(\alpha)p_n = \dfrac{\bar{b}}{\bar{a}}, & \text{(5.27b)} \\ Y = -A_n'(\alpha)\dfrac{p_n^2}{2} = D_1. & \text{(5.27c)} \end{cases}$$

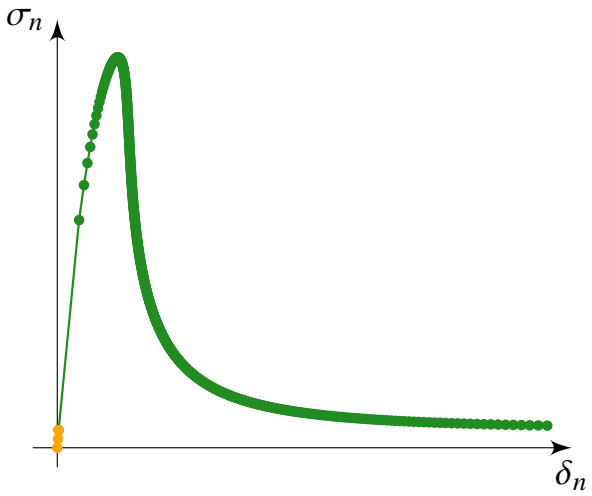


Figure 5.5 – Example of evolution of the normal stress σ_n during a traction test. The phases without and with evolution of plasticity and damage are represented in *orange* and in *green*, respectively.

From (5.27c) we get

$$p_n(\alpha) = \sqrt{\frac{2D_1}{-B_n S'(\alpha)}}, \tag{5.28}$$

recalling that $S(\alpha) := \dfrac{A_n(\alpha)}{B_n}$. Inserting (5.28) into (5.27b) and (5.27a), we achieve the parametric expression of the normal displacement and stress:

$$\delta_n(\alpha) = \frac{\bar{a}(K_{n,0} + B_n S(\alpha))\sqrt{2D_1} + \bar{b}\sqrt{-B_n S'(\alpha)}}{K_{n,0}((\bar{a} - 2\beta_n \bar{b})\sqrt{-B_n S'(\alpha)} - 2\beta_n \bar{a} B_n S(\alpha)\sqrt{2D_1})},$$

and

$$\sigma_n(\alpha) = \frac{\bar{b}}{\bar{a}}\left(1 - \beta_n \frac{\bar{b}}{\bar{a}}\right) + \left(1 - 2\beta_n \frac{\bar{b}}{\bar{a}}\right)\sqrt{2D_1 B_n}\frac{S(\alpha)}{\sqrt{-S'(\alpha)}} + 2\beta_n D_1 B_n \frac{(S(\alpha))^2}{S'(\alpha)}.$$

Remark 5.18 (Implementation notes)**.** *The unified model we have presented can be implemented in a finite element software with an incremental approach using the joint elements. In particular, it is possible to extend the argument of Subsection 2.3.6. Notice that the difference between the two models with plasticity-damage coupling involves only one state variable, i.e., the normal displacement jump δ_n, which is given as input of the algorithm, see Algorithm 2.1. As a consequence, it is sufficient to:*

1. *modify the expression of the normal stress, i.e., substitute (2.51a) with*

$$\begin{aligned}\sigma_n =& K_n(\delta_n)(\delta_n - p_n^- - \Delta p_n) + \frac{\partial K_n(\delta_n)}{\partial \delta_n}\frac{(\delta_n - p_n^- - \Delta p_n)^2}{2} \\ &+ \frac{\partial K_t(\delta_n)}{\partial \delta_n}\frac{\|\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta \boldsymbol{p_t}\|^2}{2};\end{aligned}$$

2. *replace all the occurrence of K_n and K_t respectively with $K_n(\delta_n)$ and $K_t(\delta_n)$ defined by* (5.22).

However, we remark that, although this adaptation seems to be easy, the computation of the tangent matrix would be more complex due to the presence of hyperelasticity.

5.4 Conclusion

In the first part of this chapter, we have extended the a posteriori error analysis presented in Chapter 4. In particular, for the contact problem between an elastic domain and a rigid surface with reversible cohesive forces, it is still possible to develop an adaptive algorithm based on the equilibrated stress reconstruction with a stopping criterion for the number of Newton iterations and to show local and global efficiency. Then, we have shown the main ideas for adapting this analysis also to the two bodies problem with a cohesive interface. We have also briefly indicated that the choice of constitutive relation is an important task to ensure the efficiency results, see Remark 5.11. With this aim, we have focused on constitutive relations coming from the Standard Generalized Materials framework (see Subsection 2.1.2) adapted to joint modeling, since, in the context of geomaterials, it leads to robust numerical results. In particular, we have enriched the model coupling plasticity and damage, analyzed Chapter 2, adding a nonlinear hyperbolic dependence in the elastic matrix $\mathbb{E}(\delta_n)$, in order to recover the main properties of the hyperelasto-plastic relation of Chapter 3. Notably, starting from a linear elastic domain (of Drucker–Prager type) in the space of plasticity generalized force, we recover a parabolic one (of Hoek–Brown type) in the stress space. Due to the presence of a hardening term, the latter is not fixed during the evolution of plasticity and damage, and this produces peaks of stress during shear and traction tests. In addition, the introduction of hyperelasticity reduces the asymptotic slope of dilatancy and enables to recover its saturation during cycling loadings. This initial analysis might be extended with a further investigation of the parametric influence, and with some simulations on dams models.

A Additional considerations on the constitutive relations

This appendix is devoted to the computation of the explicit expressions of the tangent matrix for the different joint models presented in this work, and to a brief analysis of some parameters of the hyperelastic joint model.

Contents

A.1 Computation of tangent matrices

As we have discuss in Section 2.2.1, Newton's method is classically used to solve nonlinear quasi-static problems, since they lead to search the zero of a nonlinear function. A crucial ingredient for having good convergence properties is thus the tangent matrix, which is obtained by derivation of the stress vector $\boldsymbol{\sigma}$:

$$\mathbb{T} := \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} = \begin{pmatrix} \dfrac{\partial \sigma_n(\boldsymbol{\delta})}{\partial \delta_n} & \dfrac{\partial \sigma_n(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}_t} \\ \dfrac{\partial \boldsymbol{\sigma}_t(\boldsymbol{\delta})}{\partial \delta_n} & \dfrac{\partial \boldsymbol{\sigma}_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}_t} \end{pmatrix}$$

Recalling that the constitutive relation $\boldsymbol{\sigma}(\boldsymbol{\delta})$ can depend on additional state variables $\boldsymbol{\alpha}$, in general, the tangent matrix as the following expression:

$$\frac{\partial \boldsymbol{\sigma}(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} = \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\delta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\delta}} + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\delta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \cdot \frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\delta}}.$$

For simplicity of notation, we will write the argument of the function only for distinguishing the derivatives in which we consider separately another state variable. In `code_aster`

(https://code-aster.org), the constitutive relations for joints are implemented with an incremental approach and the tangent matrix is obtained in a implicit way. For any state variable or thermodynamical force z, z^- and z denote its value at the previous and current state, respectively, and Δz its increment.

In the following sections, we show the explicit expression of these derivatives for the models considered in Chapters 2 and 3. For each case, we distinguish the elastic phase from the one in which also the dissipative variables evolve.

A.1.1 Chapter 2: Model coupling plasticity and damage

In this model, the dissipative variables consist in the plastic component $\boldsymbol{p}$ and in the damage variable α. In particular, we recall that the assumption of simultaneous evolution of plasticity and damage is made and that the incremental approach is summarized by Algorithm 2.1.

If the evolution is elastic, i.e., $\Delta \boldsymbol{p} = \boldsymbol{0}$ and $\Delta\alpha = 0$ and the elastic prediction is the solution, the tangent matrix simply consists in the elastic matrix $\mathbb{E}$ (2.17):

$$\mathbb{E} = \begin{pmatrix} K_n & 0 & 0 \\ 0 & K_t & 0 \\ 0 & 0 & K_t \end{pmatrix}.$$

Otherwise, its computation is more complex due to the implicit definition of the current damage value α.

At first, we assume that the new value of damage α is given by solving the nonlinear equation (2.63). The stress vector in this case is computed as follows:

$$\begin{cases} \sigma_n = K_n \left(\delta_n - p_n^- - \mu\Delta\lambda\right), \\ \boldsymbol{\sigma_t} = K_t \left(\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta\lambda \dfrac{K_t(\boldsymbol{\delta_t} - \boldsymbol{p_t^-}) - A_t(\alpha)\boldsymbol{p_t^-}}{\|K_t(\boldsymbol{\delta_t} - \boldsymbol{p_t^-}) - A_t(\alpha)\boldsymbol{p_t^-}\|}\right), \end{cases}$$

where

$$\begin{cases} \Delta\lambda = \dfrac{\|K_t(\boldsymbol{\delta_t} - \boldsymbol{p_t^-}) - A_t(\alpha)\boldsymbol{p_t^-}\| + \mu K_n(\delta_n - p_n^-) - \mu A_n(\alpha) p_n^- - \bar{c}}{\mu^2 K_n + K_t + \mu^2 A_n(\alpha) + A_t(\alpha)}, \\ F(\alpha) = 0, \end{cases}$$

with F defined by (2.67). Therefore,

$$\begin{cases} \dfrac{\partial\sigma_n}{\partial\delta_n} = K_n\left(1 - \mu\dfrac{\partial\Delta\lambda}{\partial\delta_n}\right), \\ \dfrac{\partial\sigma_n}{\partial\boldsymbol{\delta_t}} = -\mu K_n \dfrac{\partial\Delta\lambda}{\partial\boldsymbol{\delta_t}}, \\ \dfrac{\partial\boldsymbol{\sigma_t}}{\partial\delta_n} = -K_t\left[\dfrac{\partial\Delta\lambda}{\partial\delta_n}\cdot\dfrac{\boldsymbol{W}}{\|\boldsymbol{W}\|} + \dfrac{\Delta\lambda}{\|\boldsymbol{W}\|}\left(\mathbf{I}_{d-1} - \dfrac{\boldsymbol{W}\otimes\boldsymbol{W}}{\|\boldsymbol{W}\|^2}\right)\dfrac{\partial\boldsymbol{W}}{\partial\delta_n}\right], \\ \dfrac{\partial\boldsymbol{\sigma_t}}{\partial\boldsymbol{\delta_t}} = K_t\left[\mathbf{I}_{d-1} - \dfrac{\partial\Delta\lambda}{\partial\boldsymbol{\delta_t}}\otimes\dfrac{\boldsymbol{W}}{\|\boldsymbol{W}\|} - \dfrac{\Delta\lambda}{\|\boldsymbol{W}\|}\left(\mathbf{I}_{d-1} - \dfrac{\boldsymbol{W}\otimes\boldsymbol{W}}{\|\boldsymbol{W}\|^2}\right)\dfrac{\partial\boldsymbol{W}}{\partial\boldsymbol{\delta_t}}\right], \end{cases}$$

where $\otimes$ denotes the dyadic product of two vectors and, as before, $\mathbf{I}_{d-1}$ in this context denotes the identity matrix of the space $\mathbb{R}^{(d-1)\times(d-1)}$. The needed derivatives to complete the computation are provided here, recalling the definition of $\boldsymbol{W}$ (2.55):

$$
\begin{cases}
\dfrac{\partial \Delta\lambda(\delta_n)}{\partial \delta_n} = \dfrac{\mu K_n}{K_t + \mu^2 K_n + A_t(\alpha) + \mu^2 A_n(\alpha)} \\
\qquad\qquad - \dfrac{\partial \Delta\lambda(\delta_n,\alpha)}{\partial \alpha}\, \dfrac{\partial F(\delta_n,\alpha)}{\partial \delta_n} \left(\dfrac{\partial F(\delta_n,\alpha)}{\partial \alpha}\right)^{-1}, \\
\dfrac{\partial \Delta\lambda(\boldsymbol{\delta_t})}{\partial \boldsymbol{\delta_t}} = \dfrac{K_t}{K_t + \mu^2 K_n + A_t(\alpha) + \mu^2 A_n(\alpha)} \cdot \dfrac{\boldsymbol{W}}{\|\boldsymbol{W}\|}, \\
\qquad\qquad - \dfrac{\partial \Delta\lambda(\boldsymbol{\delta_t},\alpha)}{\partial \alpha}\, \dfrac{\partial F(\boldsymbol{\delta_t},\alpha)}{\partial \boldsymbol{\delta_t}} \left(\dfrac{\partial F(\boldsymbol{\delta_t},\alpha)}{\partial \alpha}\right)^{-1}, \\
\dfrac{\partial \Delta\lambda}{\partial \alpha} = \Bigg[- \left(A_t'(\alpha)\boldsymbol{p_t^-} \cdot \dfrac{\boldsymbol{W}}{\|\boldsymbol{W}\|} + \mu A_n'(\alpha) p_n^- \right) \left(\mu^2 K_n + K_t + \mu^2 A_n(\alpha) + A_t(\alpha)\right) \\
\qquad - \left(\|\boldsymbol{W}\| + \mu K_n(\delta_n - p_n^-) - \mu A_n(\alpha) p_n^- - \bar{c} \right) \left(\mu^2 A_n'(\alpha) + A_t'(\alpha)\right) \Bigg] \\
\qquad \cdot \dfrac{1}{\left(\mu^2 K_n + K_t + \mu^2 A_n(\alpha) + A_t(\alpha)\right)^2},
\end{cases}
$$

$$
\begin{cases}
\dfrac{\partial \boldsymbol{W}}{\partial \delta_n} = A_t'(\alpha)\boldsymbol{p_t^-}\, \dfrac{\partial F(\delta_n,\alpha)}{\partial \delta_n} \left(\dfrac{\partial F(\delta_n,\alpha)}{\partial \alpha}\right)^{-1}, \\
\dfrac{\partial \boldsymbol{W}}{\partial \boldsymbol{\delta_t}} = K_t + A_t'(\alpha)\boldsymbol{p_t^-} \otimes \dfrac{\partial F(\boldsymbol{\delta_t},\alpha)}{\partial \boldsymbol{\delta_t}} \left(\dfrac{\partial F(\boldsymbol{\delta_t},\alpha)}{\partial \alpha}\right)^{-1},
\end{cases}
$$

and finally, from (2.67) and the definition of the discriminant (2.62),

$$
\begin{cases}
\dfrac{\partial F}{\partial \delta_n} = \mu K_n \left(\mu^2 A_n'(\alpha) + A_t'(\alpha)\right), \\
\dfrac{\partial F}{\partial \boldsymbol{\delta_t}} = K_t \dfrac{\boldsymbol{W}}{\|\boldsymbol{W}\|} \left(\mu^2 A_n'(\alpha) + A_t'(\alpha)\right) + \left[1 + \dfrac{1}{\sqrt{\text{disc}}} \left(\mu A_n'(\alpha) p_n^- + A_t'(\alpha)\boldsymbol{p_t^-} \cdot \dfrac{\boldsymbol{W}}{\|\boldsymbol{W}\|} \right) \right] \\
\qquad \cdot \left(\mu^2 K_n + K_t + \mu^2 A_n(\alpha) + A_t(\alpha)\right) \dfrac{K_t}{\|\boldsymbol{W}\|} \left(\mathbf{I}_{d-1} - \dfrac{\boldsymbol{W}\otimes\boldsymbol{W}}{\|\boldsymbol{W}\|^2} \right) A_t'(\alpha)\boldsymbol{p_t^-}, \\
\dfrac{\partial F}{\partial \alpha} = \left(\|\boldsymbol{W}\| + \mu K_n(\delta_n - p_n^-) - \mu A_n(\alpha) p_n^- - \bar{c} \right) \left(\mu^2 A_n''(\alpha) + A_t''(\alpha)\right) + \\
\qquad + \left[\left(1 + \dfrac{1}{\sqrt{\text{disc}}} \left(\mu A_n'(\alpha) p_n^- + A_t'(\alpha)\boldsymbol{p_t^-} \cdot \dfrac{\boldsymbol{W}}{\|\boldsymbol{W}\|} \right) \right) \left(\mu A_n''(\alpha) p_n^- \right.\right. \\
\qquad \left. + A_t''(\alpha)\boldsymbol{p_t^-} \cdot \dfrac{\boldsymbol{W}}{\|\boldsymbol{W}\|} - \dfrac{\left(A_t'(\alpha)\right)^2}{\|\boldsymbol{W}\|} \boldsymbol{p_t^-} \left(\mathbf{I}_{d-1} - \dfrac{\boldsymbol{W}\otimes\boldsymbol{W}}{\|\boldsymbol{W}\|^2} \right) \boldsymbol{p_t^-} \right) \\
\qquad - \dfrac{1}{2\sqrt{\text{disc}}} \Big(\left(\mu^2 A_n''(\alpha) + A_t''(\alpha)\right) \left(A_n'(\alpha)(p_n^n)^2 + A_t'(\alpha)\|\boldsymbol{p_t^-}\|^2 + 2D_1 \right) \\
\qquad + \left(\mu^2 A_n'(\alpha) + A_t'(\alpha)\right) \left(A_n''(\alpha)(p_n^n)^2 + A_t''(\alpha)\|\boldsymbol{p_t^-}\|^2 \right) \Big) \Big] \\
\qquad \cdot \left(K_t + \mu^2 K_n + A_t(\alpha) + \mu^2 A_n(\alpha)\right) + \sqrt{\text{disc}} \left(A_t'(\alpha) + \mu^2 A_n'(\alpha) \right).
\end{cases}
$$

Now, we assume that the new value of damage α is given by solving the nonlinear equation (2.66). Then, the current stress value is:

$$\begin{cases} \sigma_n = K_n(\delta_n - p_n^- - \Delta p_n), \\ \boldsymbol{\sigma_t} = K_t(\boldsymbol{\delta_t} - \boldsymbol{p_t^-} - \Delta \boldsymbol{p_t}), \end{cases}$$

where

$$\begin{cases} \Delta p_n = \dfrac{\mu K_n \delta_n - \bar{c}}{\mu(K_n + A_n(\alpha))} - p_n^-, \\ \Delta \boldsymbol{p_t} = \dfrac{K_t \boldsymbol{\delta_t}}{K_t + A_t(\alpha)} - \boldsymbol{p_t^-}, \\ G(\alpha) = 0, \end{cases}$$

with G defined by (2.68). Therefore,

$$\begin{cases} \dfrac{\partial \sigma_n}{\partial \delta_n} = K_n \left(1 - \dfrac{\partial \Delta p_n}{\partial \delta_n}\right), \\ \dfrac{\partial \sigma_n}{\partial \boldsymbol{\delta_t}} = -K_n \dfrac{\partial \Delta p_n}{\partial \boldsymbol{\delta_t}}, \\ \dfrac{\partial \boldsymbol{\sigma_t}}{\partial \delta_n} = -K_t \dfrac{\partial \Delta \boldsymbol{p_t}}{\partial \delta_n}, \\ \dfrac{\partial \boldsymbol{\sigma_t}}{\partial \boldsymbol{\delta_t}} = K_t \left(\mathbf{I}_{d-1} - \dfrac{\partial \Delta \boldsymbol{p_t}}{\partial \boldsymbol{\delta_t}}\right). \end{cases}$$

The expression of the other derivatives are

$$\begin{cases} \dfrac{\partial \Delta p_n(\delta_n)}{\partial \delta_n} = \dfrac{K_n}{K_n + A_n(\alpha)} + \dfrac{(\mu K_n \delta_n - \bar{c}) A_n'(\alpha)}{\mu(K_n + A_n(\alpha))^2} \dfrac{\partial G(\delta_n, \alpha)}{\partial \delta_n} \left(\dfrac{\partial G(\delta_n, \alpha)}{\partial \alpha}\right)^{-1}, \\ \dfrac{\partial \Delta p_n(\boldsymbol{\delta_t})}{\partial \boldsymbol{\delta_t}} = \dfrac{(\mu K_n \delta_n - \bar{c}) A_n'(\alpha)}{\mu(K_n + A_n(\alpha))^2} \dfrac{\partial G(\boldsymbol{\delta_t}, \alpha)}{\partial \boldsymbol{\delta_t}} \left(\dfrac{\partial G(\boldsymbol{\delta_t}, \alpha)}{\partial \alpha}\right)^{-1}, \end{cases}$$

$$\begin{cases} \dfrac{\partial \Delta \boldsymbol{p_t}(\delta_n)}{\partial \delta_n} = \dfrac{K_t A_t'(\alpha) \boldsymbol{\delta_t}}{(K_t + A_t(\alpha))^2} \dfrac{\partial G(\delta_n, \alpha)}{\partial \delta_n} \left(\dfrac{\partial G(\delta_n, \alpha)}{\partial \alpha}\right)^{-1}, \\ \dfrac{\partial \Delta \boldsymbol{p_t}(\boldsymbol{\delta_t})}{\partial \boldsymbol{\delta_t}} = \dfrac{K_t \mathbf{I}_{d-1}}{K_t + A_t(\alpha)} + \dfrac{K_t A_t'(\alpha) \boldsymbol{\delta_t}}{(K_t + A_t(\alpha))^2} \otimes \dfrac{\partial G(\boldsymbol{\delta_t}, \alpha)}{\partial \boldsymbol{\delta_t}} \left(\dfrac{\partial G(\boldsymbol{\delta_t}, \alpha)}{\partial \alpha}\right)^{-1}, \end{cases}$$

and finally, from (2.68)

$$\begin{cases} \dfrac{\partial G}{\partial \delta_n} = 2A_n'(\alpha) \dfrac{K_n(\mu K_n \delta_n - \bar{c})}{\mu(K_n + A_n(\alpha))^2}, \\ \dfrac{\partial G}{\partial \boldsymbol{\delta_t}} = 2A_t'(\alpha) \dfrac{K_t^2 \boldsymbol{\delta_t}}{(K_t + A_t(\alpha))^2}, \\ \dfrac{\partial G}{\partial \alpha} = \left[A_n''(\alpha)(K_n + A_n(\alpha)) - 2(A_n'(\alpha))^2\right] \dfrac{(\mu K_n \delta_n - \bar{c})^2}{\mu^2(K_n + A_n(\alpha))^3} + \\ \qquad + \left[A_t''(\alpha)(K_t + A_t(\alpha)) - 2(A_t'(\alpha))^2\right] \dfrac{(K_t \|\boldsymbol{\delta_t}\|)^2}{(K_t + A_t(\alpha))^3}. \end{cases}$$

A.1.2 Chapter 3: Hyperelasto-plastic model

We recall that in this model we have only one dissipative state variable: the plasticity vector $\boldsymbol{p}$. The incremental approach for the hyperelastic constitutive relation is summarize by Algorithm 3.1.

If the evolution is elastic, i.e., $\Delta\boldsymbol{p} = \mathbf{0}$, the tangent matrix is simply obtained by derivation of (3.48a)-(3.48b):

$$\begin{pmatrix} \dfrac{\partial\sigma_n}{\partial\delta_n} & \dfrac{\partial K_t(\delta_n)}{\partial\delta_n}(\boldsymbol{\delta}_t - \boldsymbol{p}_t^-)^T \\ \dfrac{\partial K_t(\delta_n)}{\partial\delta_n}(\boldsymbol{\delta}_t - \boldsymbol{p}_t^-) & K_t(\delta_n)\mathbf{I}_2 \end{pmatrix},$$

where

$$\frac{\partial\sigma_n}{\partial\delta_n} = K_n(\delta_n) + 2\,\frac{\partial K_n(\delta_n)}{\partial\delta_n}(\delta_n - p_n^-) + \frac{\partial^2 K_n(\delta_n)}{\partial\delta_n^2}\,\frac{(\delta_n - p_n^-)^2}{2} + \frac{\partial^2 K_t(\delta_n)}{\partial\delta_n^2}\,\frac{(\boldsymbol{\delta}_t - \boldsymbol{p}_t^-)^2}{2}.$$

In particular, we recall that

$$\begin{aligned} K_s(\delta_n) &= \frac{K_{s,0}}{2K_{s,0}\beta_s\delta_n + 1}, \\ \frac{\partial K_s(\delta_n)}{\partial\delta_n} &= -\frac{2K_{s,0}^2\beta_s}{(2K_{s,0}\beta_s\delta_n + 1)^2} = -2\beta_s\left(K_s(\delta_n)\right)^2, \\ \frac{\partial^2 K_s(\delta_n)}{\partial\delta_n^2} &= \frac{8K_{s,0}^3\beta_s^2}{(2K_{s,0}\beta_s\delta_n + 1)^3} = 8\beta_s^2\left(K_s(\delta_n)\right)^3. \end{aligned}$$

for $s \in \{n, t\}$.

Otherwise, we have to consider also the contribution of $\Delta\boldsymbol{p}$. More in details if $\boldsymbol{X}_t \neq \mathbf{0}$, the tangent matrix is obtained by derivation of (3.45a)-(3.45b), where $\Delta\boldsymbol{p}$ is given by the flow rule (2.54a)-(3.49b), and $\Delta\lambda$ by (3.50). We then have

$$\begin{cases} \dfrac{\partial\sigma_n}{\partial\delta_n} = K_n(\delta_n)\left(1 - \dfrac{\partial\Delta p_n}{\partial\delta_n}\right) + \dfrac{\partial K_n(\delta_n)}{\partial\delta_n}(\boldsymbol{\delta}_t - \boldsymbol{p}_t^- - \Delta\boldsymbol{p}_t)\left(2 - \dfrac{\partial\Delta p_n}{\partial\delta_n}\right) \\ \qquad + \dfrac{\partial^2 K_n(\delta_n)}{\partial\delta_n^2}\,\dfrac{(\delta_n - p_n^- - \Delta p_n)^2}{2} - \dfrac{\partial K_t(\delta_n)}{\partial\delta_n}(\boldsymbol{\delta}_t - \boldsymbol{p}_t^- - \Delta\boldsymbol{p}_t)\dfrac{\partial\Delta\boldsymbol{p}_t}{\partial\delta_n} \\ \qquad + \dfrac{\partial^2 K_t(\delta_n)}{\partial\delta_n^2}\,\dfrac{(\boldsymbol{\delta}_t - \boldsymbol{p}_t^- - \Delta\boldsymbol{p}_t)^2}{2}, \\ \dfrac{\partial\sigma_n}{\partial\boldsymbol{\delta}_t} = -\left(K_n(\delta_n) + \dfrac{\partial K_n(\delta_n)}{\partial\delta_n}(\delta_n - p_n^- - \Delta p_n)\right)\dfrac{\partial\Delta p_n}{\partial\boldsymbol{\delta}_t} \\ \qquad + \dfrac{\partial K_t(\delta_n)}{\partial\delta_n}(\boldsymbol{\delta}_t - \boldsymbol{p}_t^- - \Delta\boldsymbol{p}_t)\left(\mathbf{I}_2 - \dfrac{\partial\Delta\boldsymbol{p}_t}{\partial\boldsymbol{\delta}_t}\right), \\ \dfrac{\partial\boldsymbol{\sigma}_t}{\partial\delta_n} = \dfrac{\partial K_t(\delta_n)}{\partial\delta_n}(\boldsymbol{\delta}_t - \boldsymbol{p}_t^- - \Delta\boldsymbol{p}_t) - K_t(\delta_n)\dfrac{\partial\Delta\boldsymbol{p}_t}{\partial\delta_n}, \\ \dfrac{\partial\boldsymbol{\sigma}_t}{\partial\boldsymbol{\delta}_t} = K_t(\delta_n)\left(\mathbf{I}_2 - \dfrac{\partial\Delta\boldsymbol{p}_t}{\partial\boldsymbol{\delta}_t}\right). \end{cases}$$

Moreover,

$$\frac{\partial \Delta\lambda}{\partial \delta_n} = \frac{1}{(\bar{a}^2 K_n(\delta_n) + K_t(\delta_n))^2}\Bigg[\left(\frac{\partial K_t(\delta_n)}{\partial \delta_n}\|\boldsymbol{\delta}_t - \boldsymbol{p}_t^-\| + \bar{a}\frac{\partial K_n(\delta_n)}{\partial \delta_n}(\delta_n - p_n^-) + \bar{a}K_n(\delta_n)\right)$$
$$\cdot\left(\bar{a}^2 K_n(\delta_n) + K_t(\delta_n)\right) - \left(K_t(\delta_n)\|\boldsymbol{\delta}_t - \boldsymbol{p}_t^-\| + \bar{a}K_n(\delta_n)(\delta_n - p_n^-) - \bar{b}\right)$$
$$\cdot\left(\bar{a}^2\frac{\partial K_n(\delta_n)}{\partial \delta_n} + \frac{\partial K_t(\delta_n)}{\partial \delta_n}\right)\Bigg],$$

$$\frac{\partial \Delta\lambda}{\partial \boldsymbol{\delta}_t} = \frac{K_t(\delta_n)}{\bar{a}^2 K_n(\delta_n) + K_t(\delta_n)}\frac{\boldsymbol{\delta}_t - \boldsymbol{p}_t^-}{\|\boldsymbol{\delta}_t - \boldsymbol{p}_t^-\|},$$

and, consequently,

$$\frac{\partial \Delta p_n}{\partial \boldsymbol{\delta}} = \bar{a}\frac{\partial \Delta\lambda}{\partial \boldsymbol{\delta}}, \qquad \frac{\partial \Delta \boldsymbol{p}_t}{\partial \delta_n} = \frac{\partial \Delta\lambda}{\partial \delta_n}\frac{\boldsymbol{\delta}_t - \boldsymbol{p}_t^-}{\|\boldsymbol{\delta}_t - \boldsymbol{p}_t^-\|},$$

and

$$\mathbf{I}_2 - \frac{\partial \Delta \boldsymbol{p}_t}{\partial \boldsymbol{\delta}_t} = \frac{1}{\bar{a}^2 K_n(\delta_n) + K_t(\delta_n)}\Bigg[\left(\bar{a}^2 K_n(\delta_n) - \frac{\bar{a}K_n(\delta_n)(\delta_n - p_n^-) - \bar{b}}{\|\boldsymbol{\delta}_t - \boldsymbol{p}_t^-\|}\right)\mathbf{I}_2 +$$
$$+ \frac{\bar{a}K_n(\delta_n)(\delta_n - p_n^-) - \bar{b}}{\|\boldsymbol{\delta}_t - \boldsymbol{p}_t^-\|^3}(\boldsymbol{\delta}_t - \boldsymbol{p}_t^-)\otimes(\boldsymbol{\delta}_t - \boldsymbol{p}_t^-)\Bigg].$$

A.2 Hyperelastic model: from parabolic to linear domain

In this section, we show some considerations on the parameters of the hyperelasto-plastic model for joints described in Section 3.5. In particular, we want to complete the argument of Remark 3.6, explaining how to fix the parameters $\bar{a}$, $\bar{b}$, β_n and β_t, starting from a general parabolic domain in the stress space.

Let the plasticity criterion in the stress space be defined by

$$\sigma_n = A\|\boldsymbol{\sigma_t}\|^2 + B\|\boldsymbol{\sigma_t}\| + C, \tag{A.1}$$

where $A < 0$, $B \leq 0$, and $C \geq 0$. These assumptions guarantee that the corresponding reversibility domain $\mathbb{K}_{\boldsymbol{\sigma}}$ is convex and contains the half-line $(\boldsymbol{\sigma_t} = \mathbf{0}) \cup (\sigma_n \leq 0)$. The parameters A, B, and C can be possibly be fitted through experimental tests on physical joints. Comparing (A.1) with (3.39), we obtain the following relations

$$\begin{cases} A = -\dfrac{\beta_n + \bar{a}^2\beta_t}{\bar{a}^2}, & \text{(A.2a)} \\ B = -\dfrac{\bar{a} - 2\beta_n\bar{b}}{\bar{a}^2}, & \text{(A.2b)} \\ C = \dfrac{\bar{b}(\bar{a} - \beta_n\bar{b})}{\bar{a}^2}. & \text{(A.2c)} \end{cases}$$

We consider separately different situations:

- If $B = 0$, then $\bar{a} = 2\beta_n \bar{b}$, and, consequently, by (A.2c)

$$\beta_n = \frac{1}{4C}.$$

Moreover, by (A.2a), we obtain

$$\begin{cases} \bar{a} = \sqrt{-\dfrac{1}{4C(A+\beta_t)}}, \\ \bar{b} = \dfrac{\bar{a}}{2\beta_n} = \dfrac{1}{16}\sqrt{-\dfrac{1}{C^3(A+\beta_t)}}, \\ 0 \le \beta_t < -A. \end{cases}$$

- If $C = 0$, then by (A.2c) $\bar{b} = 0$, since $\bar{a} - 2\beta_n \bar{b} \ge 0$. As a consequence, by (A.2a) and (A.2b), we obtain

$$\begin{cases} \bar{a} = -\dfrac{1}{B}, \\ \beta_t = -A - B^2\beta_n, \\ 0 \le \beta_n \le -\dfrac{A}{B^2}. \end{cases}$$

- If $B \neq 0$ and $\beta_n = 0$, then

$$\begin{cases} \bar{a} = -\dfrac{1}{B}, \\ \bar{b} = -\dfrac{C}{B}, \\ \beta_t = -A. \end{cases}$$

- Finally, if $B \neq 0$, $C \neq 0$ and $\beta_n \neq 0$, we have

$$\begin{cases} \bar{a} = \sqrt{-\dfrac{\beta_n}{A+\beta_t}} = -\dfrac{\sqrt{1-4\beta_n C}}{B}, \\ \bar{b} = \dfrac{\bar{a}}{2\beta_n}(B\bar{a}+1) = -\dfrac{\sqrt{1-4\beta_n C}}{2\beta_n B}\left(1-\sqrt{1-4\beta_n C}\right), \\ \beta_t = \dfrac{A+\beta_n(B^2-4AC)}{4\beta_n C - 1}. \end{cases}$$

In order to ensure the fact that $\bar{a}$ is real and that β_t is positive, we have the conditions

$$\beta_n < \frac{1}{4C} \qquad \text{and} \qquad \beta_n \le -\frac{A}{B^2 - 4AC}.$$

Note that, since $B \neq 0$, it is sufficient to impose only the second condition.

Bibliography

[1] M. Ainsworth and J. Oden. A posteriori error estimators for second order elliptic systems part 2. An optimal order process for calculating self-equilibrating fluxes. *Computers & Mathematics with Applications*, 26(9):75–87, 1993.

[2] M. Ainsworth and J. T. Oden. *A Posteriori Error Estimation in Finite Element Analysis*. Wiley, 2000.

[3] L. R. Alejano and A. Bobet. Drucker–Prager criterion. *Rock Mechanics and Rock Engineering*, 45:995–999, 2012.

[4] P. F. Antonietti, L. Beirão da Veiga, C. Lovadina, and M. Verani. Hierarchical a posteriori error estimators for the mimetic discretization of elliptic problems. *SIAM Journal on Numerical Analysis*, 51:654–675, 2013.

[5] T. Arbogast and Z. Chen. On the implementation of mixed methods as nonconforming methods for second-order elliptic problems. *Mathematics of Computation*, 64(211):943–972, 1995.

[6] D. N. Arnold, R. S. Falk, and R. Winther. Mixed finite element methods for linear elasticity with weakly imposed symmetry. *Mathematics of Computation*, 76(260):1699–1723, 2007.

[7] D. N. Arnold and R. Winther. Mixed finite elements for elasticity. *Numerische Mathematik*, 92:401–419, 2002.

[8] D. Aubry, J. C. Hujeux, F. Lassoudiere, and Y. Meimon. A double memory model with multiple mechanisms for cyclic soil behavior. In *Proceedings of International Symposium on Numerical Models in Geomechanics*, pages 3–13, 1982.

[9] S. Bandis, A. Lumsden, and N. Barton. Experimental studies of scale effects on the shear behaviour of rock joints. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, 18(1):1–21, 1981.

[10] R. E. Bank and R. K. Smith. A posteriori error estimates based on hierarchical bases. *SIAM Journal on Numerical Analysis*, 30(4):921–935, 1993.

[11] G. I. Barenblatt. The mathematical theory of equilibrium cracks in brittle fracture. *Advances in Applied Mechanics*, 7:55–129, 1962.

[12] M. Bebendorf. A note on the Poincaré inequality for convex domains. *Zeitschrift für Analysis und ihre Anwendungen*, 22(4):751–756, 2003.

[13] C. Bernardi and V. Girault. A local regularization operator for triangular and quadrilateral finite elements. *SIAM Journal of Numerical Analysis*, 35(5):1893–1916, 1998.

[14] D. Boffi, F. Brezzi, and M. Fortin. *Mixed Finite Element Methods and Applications*, volume 44 of *Springer Series in Computational Mathematics*. Springer Science & Business Media, 2013.

[15] M. Botti, D. A. Di Pietro, and P. Sochala. A Hybrid High-Order method for nonlinear elasticity. *SIAM Journal on Numerical Analysis*, 55(6):2687–2717, 2017.

[16] M. Botti and R. Riedlbeck. Equilibrated stress tensor reconstruction and a posteriori error estimation for nonlinear elasticity. *Computional Methods in Applied Mathematics*, 20:39–59, 2020.

[17] D. Braess and J. Schöberl. Equilibrated residual error estimator for edge elements. *Mathematics Of Computation*, 262:651–672, 2008.

[18] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer Science & Business Media, 2008.

[19] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*, volume 15 of *Springer Series in Computational Mathematics*. Springer-Verlag, 1991.

[20] D. Capatina and R. Luce. Local flux reconstruction for a frictionless unilateral contact problem. In *Numerical Mathematics and Advanced Applications ENUMATH 2019*, volume 139 of *Lecture Notes in Computational Science and Engineering*, pages 235–243, 2021.

[21] F. L. L. B. Carneiro. A new method to determine the tensile strength of concrete. In *Proceedings of the 5th meeting of the Brazilian Association for Technical Rules ("Associação Brasileire de Normas Técnicas—ABNT")*, pages 126–129, September 1943.

[22] C. Carstensen. A unifying theory of a posteriori finite element error control. *Numerische Mathematik*, 100:617–637, 2005.

[23] C. Carstensen, M. Eigel, R. H. W. Hoppe, and C. Löbhard. A review of unified a posteriori finite element error control. *Numerical Mathematics: Theory, Methods and Applications*, 5(4):509—-558, 2012.

[24] F. Chouly. An adaptation of Nitsche's method to the Tresca friction problem. *Journal of Mathematical Analysis and Applications*, 411:329–339, 2014.

[25] F. Chouly, A. Ern, and N. Pignet. A Hybrid High-Order discretization combined with Nitsche's method for contact and Tresca friction in small strain elasticity. *SIAM Journal on Scientific Computing*, 42:2300–2324, 2020.

[26] F. Chouly, M. Fabre, P. Hild, R. Mlika, J. Pousin, and Y. Renard. An overview of recent results on Nitsche's method for contact problems. *Geometrically Unfitted Finite Element Methods and Applications*, 121:93–141, 2017.

[27] F. Chouly, M. Fabre, P. Hild, J. Pousin, and Y. Renard. Residual-based a posteriori error estimation for contact problems approximated by Nitsche's method. *IMA Journal of Numerical Analysis*, 38:921–954, 2018.

[28] F. Chouly and P. Hild. A Nitsche-based method for unilateral contact problems: numerical analysis. *SIAM Journal on Numerical Analysis*, 51(2):1295–1307, 2013.

[29] F. Chouly, P. Hild, and Y. Renard. Symmetric and non-symmetric variants of Nitsche's method for contact problems in elasticity: theory and numerical experiments. *Mathematics of Computation*, 84:1089–1112, 2015.

[30] P. G. Ciarlet. *The finite element method for elliptic problems*, volume 40 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002. Reprint of the 1978 original [North-Holland, Amsterdam; MR0520174 (58 #25001)].

[31] C. A. Coulomb. Essai sur une application des règles des maximis et minimis à quelquels problèmes de statique, relatifs à l'architecture. *Academie Royale Des Sciences Paris Mem. Math. Phys.*, 7:343–382, 1776.

[32] E. Creusé and S. Nicaise. A posteriori error estimator based on gradient recovery by averaging for convection-diffusion-reaction problems approximated by discontinuous Galerkin methods. *IMA Journal of Numerical Analysis*, 33(1):212–241, 2013.

[33] A. Curnier and P. Alart. A generalized Newton method for contact problems with friction. *Journal de Mécanique Théorique et Appliquée*, 7:67–82, 1988.

[34] J. Dabaghi. *Estimations d'erreur a posteriori pour des inégalités variationnelles: application à un écoulement diphasique en milieu poreux*. PhD thesis, Sorbonne Université, June 2019.

[35] J. Dabaghi, V. Martin, and M. Vohralík. Adaptive inexact semismooth newton methods for the contact problem between two membranes. *Journal of Scientific Computing*, 84(28), 2020.

[36] Y. F. Dafalias and E. P. Popov. A model of nonlinearly hardening materials for complex loading. *Acta Mechanica*, 21:173–192, 1975.

[37] P. Destuynder and B. Métivet. Explicit error bounds in a conforming finite element method. *Mathematics Of Computation*, 68:1379–1396, 1999.

[38] G. Deveze and G. Coubard. Développement d'une base de données sur la résistance à la traction de l'interface béton-rocher. 2015.

[39] D. A. Di Pietro and J. Droniou. *The Hybrid High-Order method for polytopal meshes*, volume 19 of *Modeling, Simulation and Application*. Springer International Publishing, 2020.

[40] D. A. Di Pietro and A. Ern. *Mathematical aspects of discontinuous Galerkin methods*, volume 69 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Heidelberg, 2012.

[41] 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.

[42] D. A. Di Pietro, A. Ern, and S. Lemaire. An arbitrary-order and compact-stencil discretization of diffusion on general meshes based on local reconstruction operators. *Comput. Meth. Appl. Math.*, 14(4):461–472, 2014.

[43] D. A. Di Pietro, I. Fontana, and K. Kazymyrenko. A posteriori error estimates via equilibrated stress reconstructions for contact problems approximated by Nitsche's method. *Computers & Mathematics with Applications*, 111:61–80, 2022.

[44] 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. *Mathematics of Computation*, 84(291):153–186, 2015.

[45] P. Divoux. *Modélisation du comportement hydro-mécanique des discontinuités dans les structures et les fondations rocheuses. Application aux barrages en béton*. PhD thesis, Université Joseph-Fourier – Grenoble I, October 1997.

[46] Documentation Code_Aster: R3.06.09. *Éléments finis de joint mécaniques et éléments finisde joint couplés hydromécanique*, October 2014.

[47] Documentation Code_Aster: R3.06.13. *Éléments finis d'interface mixte pour des modèles de zone cohésive (`xxx_INTERFACE` et `xxx_INTERFACE_S`)*, September 2012.

[48] Documentation Code_Aster: R5.03.01. *Algorithme non linéaire quasi-statique (`STAT_NON_LINE`)*, September 2018.

[49] Documentation Code_Aster: R7.01.25. *Lois de comportement des joints des barrages: `JOINT_MECA_RUPT` et `JOINT_MECA_FROT`*, May 2019.

[50] D. C. Drucker and W. Prager. Soil mechanics and plastic analysis or limit design. *Quarterly of Applied Mathematics*, 10:157–165, 1952.

[51] D. S. Dugdale. Yielding of steel sheets containing slits. 8:100–104, 1960.

[52] E. Eberhardt. The Hoek-Brown failure criterion. *Rock Mechanics and Rock Engineering*, 45:981–988, 2012.

[53] I. Einav, G. Houlsby, and G. Nguyen. Coupled damage and plasticity models derived from energy and dissipation potentials. *International Journal of Solids and Structures*, 44(7):2487–2508, 2007.

[54] L. El Alaoui and A. Ern. Residual and hierarchical a posteriori error estimates for nonconforming mixed finite element methods. *M2AN Mathematical Modelling and Numerical Analysis*, 38(6):903–929, 2004.

[55] 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. *Computer Methods in Applied Mechanics and Engineering*, 200:2782–2795, 2011.

[56] B. El Merabi. *Comportement mécanique des joints cohésifs de béton-granite au niveau de l'interface barrage-fondation : influence géométrique et mécanique des aspérités*. PhD thesis, Université Grenoble Alpes, January 2018.

[57] A. Ern and M. Vohralík. Polynomial-degree-robust a posteriori estimates in a unified setting for conforming, nonconforming, discontinuous Galerkin, and mixed discretizations. *SIAM Journal on Numerical Analysis*, 53(2):1058–1081, 2015.

[58] R. Eymard, T. Gallouet, and R. Herbin. Finite volume methods. In *Solution of Equation in Rn (Part 3), Techniques of Scientific Computing (Part 3)*, volume 7 of *Handbook of Numerical Analysis*, pages 713–1020. Elsevier, 2000.

[59] C. Fairhurst. On the validity of the 'Brazilian' test for brittle materials. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, 1(4):535–546, 1964.

[60] F. Fierro and A. Veeser. A posteriori error estimators, gradient recovery by averaging, and superconvergence. *Numerische Mathematik*, 103:267–298, 2006.

[61] I. Fontana, K. Kazymyrenko, and D. A. Di Pietro. Hyperelastic nature of the Hoek–Brown criterion. Submitted.

[62] G. Francfort and J.-J. Marigo. Revisiting brittle fracture as an energy minimization problem. *Journal of the Mechanics and Physics of Solids*, 46(8):1319–1342, 1998.

[63] P. Germain, Q. S. Nguyen, and P. Suquet. Continuum thermodynamics. *Journal of Applied Mechanics*, 105:1011–1020, 1983.

[64] A. Griffith. The theory of rupture. In *Proc. Ist. Int. Congr. Appl. Mech.*, pages 54–63, 1924.

[65] T. Gustafsson, R. Stenberg, and J. Videman. On Nitsche's method for elastic contact problems. *SIAM Journal on Scientific Computing*, 42(2):B425–B446, 2020.

[66] B. Halphen and Q. S. Nguyen. Sur les matériaux standard généralisés. *Journal de Mécanique*, 14(1):39–63, January 1975.

[67] J. Haslinger, I. Hlaváček, and J. Nečas. Numerical methods for unilateral problems in solid mechanics. In *Finite Element Methods (Part 2), Numerical Methods for Solids (Part2)*, volume 4 of *Handbook of Numerical Analysis*, pages 313–485. Elsevier, 1996.

[68] F. Hecht. New development in FreeFem++. *J. Numer. Math.*, 20(3-4):251–265, 2012.

[69] F. Hecht. *FreeFEM Documentation*, Release 4.6 edition, July 2020.

[70] T. Helfer and É. Castelier. Le générateur de code `mfront`: présentation générale et application aux propriétés matériau et aux modèles, 2013.

[71] E. Hoek and E. T. Brown. Empirical strength criterion for rock masses. *Journal of the Geotechnical Engineering Division*, 106:1013–1035, 1980.

[72] E. Hoek and E. T. Brown. *Underground Excavation in Rock*. E & FN Spon, 1982.

[73] E. Hoek and E. T. Brown. Practical estimates of rock mass strenght. *International Journal of Rock Mechanics and Mining Sciences*, 34(8):1165–1186, 1997.

[74] E. Hoek and E. T. Brown. The Hoek–Brown failure criterion and GSI – 2018 edition. *Journal of Rock Mechanics and Geotechnical Engineering*, 11:445–463, 2019.

[75] E. Hoek and P. Marinos. A brief history of the development of the Hoek–Brown failure criterion. *Solis and Rocks*, 30(2):85–92, 2007.

[76] C. Jailin, A. Carpiuc, K. Kazymyrenko, M. Poncelet, H. Leclerc, F. Hild, and S. Roux. Virtual hybrid test control of sinuous crack. *Journal of the Mechanics and Physics of Solids*, 102:239–256, 2017.

[77] Q. Jiang, L. Song, F. Yan, C. Liu, B. Yang, and J. Xiong. Experimental investigation of anisotropic wear damage for natural joints under direct shearing test. *International Journal of Geomechanics*, 20:04020015, 2020.

[78] P. Jiránek, Z. Strakoš, and M. Vohralík. A posteriori error estimates including algebraic error and stopping criteria for iterative solvers. *SIAM Journal on Scientific Computing*, 32(3):1567–1590, 2010.

[79] L. Knittel, T. Wichtmann, A. Niemunis, G. Huber, E. Espino, and T. Triantafyllidis. Pure elastic stiffness of sand represented by response envelopes derived from cyclic triaxial tests with local strain measurements. *Acta Geotechnica*, 15:2075–2088, 2020.

[80] J. F. Labuz and A. Zang. Mohr–Coulomb failure criterion. *Rock Mechanics and Rock Engineering*, 45:975–979, 2012.

[81] P. Ladeveze and D. Leguillon. Error estimate procedure in the finite element method and applications. *SIAM Journal on Numerical Analysis*, 20(3):485–509, 1983.

[82] J. Laverne. *Formulation Energétique de la Rupture par des Modèles de Forces Cohésives: Considérations Théoriques et Implantations Numériques*. PhD thesis, Université Paris XIII, November 2004.

[83] S.-K. Lee, Y.-C. Song, and S.-H. Han. Biaxial behavior of plain concrete of nuclear containment building. *Nuclear Engineering and Design*, 227(2):143–153, 2004.

[84] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2002.

[85] H. Li, T. Guo, Y. Nan, and B. Han. A simplified three–dimensional extension of Hoek–Brown strenght. *Journal of Rock Mechanics and Geotechnical Engineering*, 13:568–578, 2021.

[86] E. Lorentz. A mixed interface finite element for cohesive zone models. *Computer Methods in Applied Mechanics and Engineering*, 198(2):302–317, December 2008.

[87] R. Luce and B. Wohlmuth. A local a posteriori error estimator based on equilibrated fluxes. *SIAM Journal on Numerical Analysis*, 42(4):1394–1414, 2004.

[88] M. P. Luong. Stress–strain aspects of cohesionless soils under cyclic and transient loading. In *Proceedings of the International Symposium on Soils under Cyclic and Transient Loading*, pages 315–324. A. A. Balkema, January 1980.

[89] D. Mao, L. Shen, and A. Zhou. Adaptive finite element algorithms for eigenvalue problems based on local averaging type a posteriori error estimates. *Advances in Computational Mathematics*, 25:135–160, 2006.

[90] J.-J. Marigo. Formulation d'une loi d'endommagement d'un matériau élastique. *Comptes rendus de l'Académie des sciences. Série II*, 292:1309–1312, 1981.

[91] J.-J. Marigo. Constitutive relations in plasticity, damage and fracture mechanics based on a work property. *Nuclear Engineering and Design*, 114:249–272, 1989.

[92] J.-J. Marigo. From Clausius–Duhem and Drucker–Ilyushin inequalities to standard materials. In *Continuum Thermomechanics*, Solid Mechanics and Its Applications, pages 289–300. Springer Netherlands, 2002.

[93] J.-J. Marigo. *Plasticité et Rupture*. École Polytechnique, 2016.

[94] J.-J. Marigo and K. Kazymyrenko. A micromechanical inspired model for the coupled to damage elasto-plastic behavior of geomaterials under compression. *Mechanics & Industry*, 20, April 2019.

[95] O. Mohr. Welche umstände bedingen die elastizitätsgrenze und den bruch eines materials? *Zeitschrift des Vereines deutscher Ingenieure*, 44:1524–1530, 1900.

[96] J. J. Moreau. Sur les lois de frottement, de plasticité et de viscosité. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 271:608–611, 1970.

[97] J. J. Moreau. Fonctions de résistance et fonctions de dissipation, 1971. Article dans "Séminaire d'analyse convexe", Montpellier, exposé n°6.

[98] H. Mouzannar. *Caractérisation de la résistance au cisaillement et comportement des interfaces entre béton et fondation rocheuse des structures hydrauliques*. PhD thesis, Université de Lyon, September 2016.

[99] J. Muralha, G. Grasselli, and B. Tatone. ISRM suggested method for laboratory determination of the shear strength of rock joints: Revised version. *Rock Mechanics and Rock Engineering*, 47:291–302, 2014.

[100] K. Murota. *Discrete Convex Analysis*, volume 10 of *Monographs on Discrete Mathematics and Applications*. SIAM Society for Industrial and Applied Mathematics, 2003.

[101] 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.

[102] Q. S. Nguyen and H. D. Bui. Sur les matériaux élastoplastiques à écrouissage positif ou négatif. *Journal de Mécanique*, 13(2), June 1974.

[103] A. Niemunis and M. Cudny. On hyperelasticity for clays. *Computers and Geotechnics*, 23:221–236, 1998.

[104] J. Nitsche. Über ein variationsprinzip zur lösung von Dirichlet-problemen bei verwendung von teilräumen, die keinen randbedingungen unterworfen sind. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 36:9–15, 1971.

[105] R. W. Ogden. *Non-linear elastic deformations*. Ellis Horwood series in mathematics and its applications. E. Horwood Halsted Press, 1984.

[106] X. D. Pan and J. Hudson. A simplified three-dimensional Hoek-Brown yield criterion. *Proc Symposium on Rock Mechanics and Power Plants*, pages 95–103, 1988.

[107] L. E. Payne and H. F. Weinberger. An optimal poincaré inequality for convex domains. *Archive for Rational Mechanics and Analysis*, 5:286–292, 1960.

[108] W. Prager and J. L. Synge. Approximations in elasticity based on the concept of function space. *Quarterly of Applied Mathematics*, 5:241–269, 1947.

[109] S. Priest. Three-dimensional failure criteria based on the Hoek-Brown criterion. *Rock Mechanics and Rock Engineering*, 45:989–993, 2012.

[110] A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*, volume 23 of *Springer Series in Computational Mathematics*. Springer, 1994.

[111] S. Raude, F. Laigle, R. Giot, and R. Fernandes. A unified thermoplastic/viscoplastic costitutive model for geomaterials. *Acta Geotechnica*, 11:849–869, 2016.

[112] Y. Renard. Generalized Newton's methods for the approximation and resolution of frictional contact problems in elasticity. *Computer Methods in Applied Mechanics and Engineering*, 256:38–55, 2013.

[113] S. Repin and J. Valdman. Functional a posteriori error estimates for incremental models in elasto-plasticity. *Central European Journal of Mathematics*, 7(3):506–519, 2009.

[114] S. I. Repin. *A Posteriori Estimates for Partial Differential Equations*, volume 4 of *Radon Series on Computational and Applied Mathematics*. De Gruyter, 2008.

[115] R. Riedlbeck. *Algorithmes adaptatifs pour la poromécanique et la poro-plasticité*. PhD thesis, Université de Montpellier, November 2017.

[116] R. Riedlbeck, D. A. Di Pietro, and A. Ern. Equilibrated stress tensor 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.

[117] D. Saiang, L. Malmgren, and E. Nordlund. Laboratory tests on shotcrete-rock joints in direct shear, tension and compression. *Rock Mechanics and Rock Engineering*, 38:275–297, 2005.

[118] A. Signorini. Questioni di elasticità non linearizzata e semilinearizzata. *Rendiconti di Matematica e delle sue Applicazioni*, 18:95–139, 1959.

[119] G. Temporelli. *Da Molare al Vajoint. Storie di dighe*. Erga edizioni, 2011.

[120] K. Terzaghi. *Theoretical Soil Mechanics*. John Wiley & Sons, 1943.

[121] P. A. Veermer and R. de Borst. Non-associated plasticity for soils, concrete and rock. volume 29, 1984.

[122] R. Verfürth. A review of a posteriori error estimation techniques for elasticity problems. *Computer Methods in Applied Mechanics and Engineering*, 176:419–440, 1999.

[123] R. Verfürth. *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*. Wiley, June 1996.

[124] R. Verfürth. *A posteriori error estimation techniques for finite element methods*. Numerical Mathematics and Scientific Computation. Oxford University Press, 2013.

[125] M. Vohralík. On the discrete Poincaré-Friedrichs inequalities for nonconforming approximations of the Sobolev space H1. *Numerical Functional Analysis and Optimization*, 26:925–952, 2005.

[126] M. Vohralík. *A posteriori error estimates for efficiency and error control in numerical simulations*. UPMC Sorbonne Universités, February 2015.

[127] B. Wohlmuth. Variationally consistent discretization schemes and numerical algorithms for contact problems. *Acta Numerica*, 20:569–734, 2011.

[128] S. Yousef. *Étude d'estimations d'erreur a posteriori et d'adaptivité basée sur des critères d'arrêt et raffinement de maillages pour des problèmes d'écoulements multiphasiques et thermiques*. PhD thesis, Université Pierre et Marie Curie, 2013.

[129] Q. Zhang, H. Zhu, and L. Zhang. Modification of a generalized three–dimensional Hoek–Brown strenght criterion. *Internation Journal of Rock Mechanics and Mining Sciences*, 59:80–96, 2013.

[130] Z. Zhou, W. Ma, S. Zhang, Y. Mu, S. Zhao, and G. Li. Yield surface evolution for columnar ice. *Results in Physics*, 6:851–859, 2016.