Nonconforming discretizations of a poromechanical model on general meshes

Simon Lemaire

Thèse

présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PARIS-EST

École doctorale : MSTIC

Mention : MATHÉMATIQUES APPLIQUÉES

par

Simon Lemaire

Discrétisations non-conformes d'un modèle poromécanique sur maillages généraux

Soutenue publiquement le 12 décembre 2013 devant le jury composé de

M. Léo Agélas	Invité	IFP Énergies nouvelles
M. François Bouchut	Examinateur	Université Paris-Est Marne-la-Vallée
M. Daniele A. Di Pietro	Examinateur	Université Montpellier 2
M. Robert Eymard	Directeur de thèse	Université Paris-Est Marne-la-Vallée
M. Roland Masson	Président du jury	Université de Nice Sophia-Antipolis
M. Jan M. Nordbotten	Rapporteur	Université de Bergen, Norvège
M. Martin Vohralík	Rapporteur	INRIA Paris-Rocquencourt

à ma mère Danielle et à mon père Dominique

Remerciements

Il est des pages d'un manuscrit qui seront davantage parcourues que les autres, et celles-ci en font précisément partie. Il convient donc de les soigner, et de les coucher avec la justesse et l'exhaustivité qu'un travail de thèse se doit d'honorer.

Au moment de remercier, en premier lieu, ceux qui ont initié et guidé ce travail durant ces trois années, l'angoisse m'assaille. Tel un collégien face à l'épreuve d'histoire du brevet, je n'arrive plus à associer de dates précises aux différents monarques de la frise chronologique. Mais par où donc commencer ? Par le début peut-être : une proposition de thèse déposée auprès de la direction scientifique de l'IFP (en toute rigueur IFP Énergies nouvelles, le lecteur me passera le raccourci...) par le chef du département de mathématiques appliquées, Roland Masson. Je suis alors en stage de fin d'études dans ce même département, l'occasion est belle, nous faisons affaire. La conjoncture politique au sein de l'IFP me volera Roland quelques mois plus tard, après une courte mais très appréciable collaboration dont je ne peux que le remercier. Un changement d'école doctorale plus tard, je fais alors la connaissance de Daniele Di Pietro, devenu par décret mon promoteur officiel. Nous modifions quelque peu le cap, Daniele déborde d'idées, d'énergie, a toujours du temps pour répondre aux nombreuses interrogations qui me viennent à l'esprit, qu'elles soient mathématiques ou existentielles (surtout). Et puis le sens de l'humour sicilien est plus qu'à mon goût, j'apprécie en tous points cette collaboration (qui n'a d'ailleurs jamais vraiment cessé). Mais un beau matin, rattrapé par ses velléités académiques, Daniele s'en va. Je fais alors la connaissance de Léo Agélas, dernier promoteur de cette thèse, qui par sa bienveillance, sa discrétion (gage de confiance), sa gentillesse et ses qualités mathématiques, permit d'arriver sans encombre au dénouement. Dénouement qui, il est important de le souligner, n'aurait jamais vu le jour sans l'encadrement sur la durée de Robert Eymard, directeur de thèse académique, et seul monarque à n'avoir jamais abdiqué ou été chassé. Présent du début à la fin, d'autant plus dans les grands moments de doutes et d'angoisses, toujours disponible et arrangeant, me témoignant une confiance absolue à la fois dans la liberté accordée et dans sa volonté de me voir faire toujours davantage, la précision, la rigueur et la perspicacité de sa réflexion n'ont d'égales que ses qualités humaines. Nos discussions à Montrouge resteront pour moi comme de très bons souvenirs.

Quelques mots pour remercier très chaleureusement Martin Vohralík, pour m'avoir fait l'honneur de rapporter ce manuscrit, et pour ses remarques toujours très pertinentes et constructives. Also a few words in English to express my huge gratitude to Jan Nordbotten, who has accepted as well to be one referee, and who managed to be present for the defense. Our discussions have always been of great interest to me, and this is a complete privilege for me having you as a jury member. Je remercie également vivement François Bouchut d'avoir accepté d'être membre de ce jury, et Roland et Daniele de faire le déplacement depuis leurs contrées quelque peu lointaines.

S'il est vrai que je n'avais pas de bureau attitré à l'université, la raison en est que ma thèse s'est entièrement déroulée à l'IFP. Il y a de ce fait quelques personnes que je souhaiterais remercier. Un merci spécial à Zakia Benjelloun-Touimi, chef du département de mathématiques appliquées, qui a toujours su trouver des crédits, et ce même lorsqu'il n'y en avait plus. Merci à Virginie, la secrétaire de département la plus efficace du monde. Une pensée également pour Sylvie C., Audrey et Christiane. Merci à tous les collègues mathématiciens et informaticiens, avec mentions spéciales pour Huy, Isabelle, Stéphane, Julien, Sylvie P., Delphine, Frédéric, Françoise, Anthony, Aziz, Benjamin, Long, Jean-Yves, Steven, Christophe D., Alessio, Romain. Une mention très spéciale pour Jean-Louis, avec qui une discussion cinéma ne peut se refuser (sauf quand c'est la troisième de la journée). Merci également aux collègues géomécaniciens, Daniele C. et Nicolas, une pensée pour Mathilde. J'en arrive tout naturellement à mes compagnons de galère, collègues et pour beaucoup amis thésards. Il y a d'abord les modèles, ceux qui étaient déjà grands lorsque j'étais encore petit, Hoël (te concernant j'ai toujours l'impression d'être petit en fait), Cindy (modèle de réussite à part entière, l'humilité en plus, je n'oublierai pas nos conférences communes, un certain buisson non plus), Florian (si Karl Lagerfeld avait fait des mathématiques), Xavier (imperturbable, merci pour ta patience quand je te spammais). Un double merci au passage à Nataliya. Ensuite il y a la génération volée, Eugenio (le seul coiffeur docteur en mathématiques), Antoine (qui a raccroché les crampons), Carole (notre jeune maman). Comment ne pas parler d'Eugenio et Carole, avec qui j'ai partagé mon bureau pendant deux ans ? Merci pour tous ces moments partagés, les chorégraphies du vendredi soir d'Eugenio, les séances shopping en ligne ou les minutes fille de Carole, et tout le reste, votre bonne humeur, votre amitié, merci d'avoir été là pendant les périodes troubles et d'avoir supporté mes humeurs changeantes. J'en arrive à ma génération, que j'ai partagée, en plus de mon bureau la dernière année, avec Soleiman, compagnon fort agréable et impressionnant de calme et de sérénité. Je n'oublierai pas tes leçons d'arabe littéraire ! Une pensée pour Brahim, François, Anthony, Noelia, Ratiba (fan de Beyoncé et toujours de très bonne compagnie), et enfin Franck, mon compagnon d'aventures en Autriche, Slovaquie, au Mexique, en Italie. Je crois qu'on a tous les deux de bons souvenirs de nos voyages, c'est toujours un plaisir de partir avec toi ! Une mention spéciale à Rezki, cuisinier hors pair. J'en arrive à la jeune génération, Claire (nos conversations du midi vont te manquer tu verras), Tassadit (bonne chance avec les BN), Caroline (la transfuge), Thibaut (capable de parler plus vite que les idées ne s'enchaînent dans sa tête, nos conversations du midi vont me manquer), Pierre C. (Montreuillois convaincu), Huong (la relève). Dernier mais non des moindres, Mohamed, rencontré lorsque je n'étais encore qu'un stagiaire enfermé dans la cave du bâtiment Isabey, on ne s'est pas quittés depuis. Coach sportif sans pitié la semaine (j'en profite pour délivrer une mention sportive à Yacine et aux deux Frédéric), m'obligeant à enchaîner les tractions tout en écoutant du Sean Paul, faisant fi de mes plaintes répétées, compagnon de sorties le week end. Que de bons souvenirs, merci à toi pour tout.

J'en arrive aux (autres) amis, de tous horizons ceux-là, ma bouffée d'oxygène, précieux à mes yeux car présents à mes côtés comme pour me rappeler que la vie ne se résume pas aux mathématiques. Il y a d'abord François L., ami d'enfance, on ne s'est pas quittés depuis le collège, toujours prêt à faire un Bizkit quand je descends trois jours dans mon Aveyron natal. Ton amitié m'est précieuse. Une petite pensée pour mes compagnons de prépa, que j'ai (pratiquement tous) perdus de vue, je ne m'imaginais pas à l'époque que le chemin de l'école était encore long. J'en arrive à la bande de l'ENSTA. Parmi eux, Maxence, thésard également (sur la fin), que de bons moments partagés, entre voyages (Écosse, Égypte), repas gargantuesques, cueillette des champignons, plongées, discussions (plus ou moins) philosophiques ; moments partagés avec Baptiste également, le sauveur du Ben Nevis, toujours présent quand ça n'allait pas, merci, et pourtant certains diront que tu es le pire.

Une pensée pour nos deux Munichois, Raphaël P. et Anne-Céline (docteurs à présent), merci pour les bons moments et les cartes postales ! Une pensée pour Raphaël G., futur docteur (en médecine). Comment ne pas parler de Mathieu M. (un poke à Cricri), ami fidèle, futur docteur, compagnon de soirées ou de footing, un vrai Breton, tenant bien la marée, et toujours là quand ça ne va pas. Une petite pensée enfin pour P.-E., Mathieu S. et Valentin, trois fameux lurons avec qui les soirées partagées se sont désormais faites rares. J'en arrive à la joyeuse bande de Lamarck. Un sacré leader, Benjamin B., désormais Berlinois mais Parisien de cœur, on se connaît depuis le lycée, on s'est perdus de vue, retrouvés par hasard dans un RER A bondé, tout est reparti comme si rien ne s'était jamais arrêté. J'ai rencontré par ton biais de sacrés drilles : Jeff (un Sosh, un style, un grain), John (la bohème), Jonquille (au sommet du check), Anne-Lise (and So What?), Kevin (attention à tes clés), Vincent (*Que sais-je ?*). Que de bons souvenirs, à arpenter Paris, promener Olivier ou braquer un bus, merci à tous, une soirée ensemble me fait l'effet d'une semaine de vacances. Benjamin, rendez-vous à Londres. Autre bande, autre ambiance. Un grand merci à Anne-Sophie, Sabrina, Raphaël L. et Mohamed, pour les afterworks, les soirées chez Anne-So, les nombreux Rosa Bonheur, le voyage à New York que je n'ai pas fait pour rédiger ! C'est toujours un plaisir de se retrouver pour de nouvelles aventures autour d'un rosé Piscine et de Piccolini les amis. J'en arrive à mes amis plongeurs. Merci du fond du cœur à Michel, président de club en or qui a toujours eu beaucoup de bienveillance à mon égard. Merci à quelques plongeurs en particulier, Natacha (l'âge d'une dame, la fraîcheur d'une demoiselle), Françoise (un rayon de soleil), Aline (possède des parts au Corcoran's des Grands Boulevards), Marie T. (toujours pleine d'entrain), Béline (ne fais pas de thèse en archéologie jeune folle !). La plongée m'amène naturellement à Benjamin G. (déjà docteur) et Marie M. (en voie de l'être). Benjamin je t'ai connu à une période un peu trouble, tu as toujours été une oreille attentive et un ami fidèle. Je t'en suis extrêmement reconnaissant et sache que je t'estime beaucoup en tant que canard. Désormais voisins, j'ai déjà commencé à venir squatter régulièrement votre chez-vous, merci pour tous ces repas ensemble. Pensez cependant à refondre votre activité matrimoniale, ça ne fonctionne pas ! Un peu hors catégorie, j'ai également une pensée pour Mathieu P., ami fidèle, cinéphile averti, acteur dans l'âme, se posant des tonnes de questions existentielles, ce qui nous rassemble forcément. Nos soirées L'AEDLP pendant ma rédaction ont été salvatrices, merci pour cela, et tout le reste. Merci à mes amis du Sud-Ouest, Ève (aux bottes de Lara Croft) et Nathalie (je n'oublierai jamais la Mairie des Lilas), ainsi que le reste de la bande. Ève, merci pour nos discussions passionnantes, pour les séjours à Biarritz et Bayonne, et pour tout le reste. Merci à Claude G. et Pierre R., docteurs eux aussi, qui ne se reconnaîtront pas. Une pensée émue pour Aurore. Une autre pour Jessica. Merci enfin à celles ou ceux qui m'ont apporté ce dont j'avais besoin au moment où j'en avais besoin, sans le savoir forcément.

Mes remerciements les plus importants sont peut-être contenus dans ces quelques lignes. Je remercie ma famille (mes oncles, tantes, cousins, ma grand-mère Simone, ma grande-tante) pour leurs encouragements. Une pensée pour Rose, René, Marc et Henri. Merci à toi, Adélaïde, d'avoir toujours pris soin de moi comme si j'étais ton propre fils, depuis ma venue à Paris pour les concours jusqu'à aujourd'hui. Merci à toi, Candice, ma sœur dont j'admire la culture, la force de caractère et l'optimisme à toute épreuve. Tu as été docteur avant moi, tu as toujours été un modèle et donc un moteur, merci pour nos voyages partagés, nos délires, merci d'être là.

Enfin, du fond du cœur, merci Maman et Papa. Si j'écris ces lignes à l'heure qu'il est, c'est entièrement grâce à vous, à votre parcours depuis L'Arnaldesq ou Ham, à votre idée de l'éducation, aux valeurs que vous m'avez transmises, à vos sacrifices et à vos encouragements. Ces trois ans n'ont pas toujours été faciles à vivre pour vous, j'en ai conscience et m'en excuse. Mais le résultat est là, et je suis fier de vous compter parmi les présents en ce jour de soutenance, pour essayer de vous rendre (un peu) tout ce que vous m'avez donné.

Table des matières

Chapitre I

Introduction et contexte

Sommaire

Nous présentons dans cette introduction les motivations industrielles qui ont donné lieu à la rédaction de ce manuscrit, et nous détaillons le contenu des chapitres.

I.1 Contexte et objectifs de la thèse

La modélisation couplée des écoulements en milieu poreux et de la géomécanique, plus communément dénommée poromécanique, se trouve au cœur de plusieurs problématiques importantes chez IFP Énergies nouvelles (IFPEn), à la fois en simulation de réservoir, du stockage géologique du CO_2 et en modélisation de bassin. En simulation de réservoir, le couplage mécanique-écoulement [72] joue un rôle important pour l'étude des problèmes de compaction et subsidence induits par la mise en production de réservoirs peu consolidés, pour la stabilité des puits, ou encore la fracturation hydraulique. La non prise en compte de ce couplage peut aussi conduire à de mauvaises prédictions de la production. Ekofisk en Norvège ou Bachaquero au Venezuela sont de bons exemples de gisements pour lesquels la prise en compte de ce couplage est cruciale. Pour Ekofisk par exemple, l'extraction des hydrocarbures entraîne une réduction du volume poreux (l'eau injectée à très forte pression en remplacement de ces derniers conduit à une décomposition du squelette crayeux qui se recompose sous une forme plus compacte) qui provoque un phénomène de subsidence qui à terme peut endommager les équipements de puits. C'est pourquoi, après avoir déjà constaté une subsidence de 4m, les plateformes ont été rehaussées en 1987 lors d'une opération de grande envergure. Le couplage mécanique-écoulement est aussi crucial pour l'étude des risques liés à l'injection et au stockage du CO_2, comme la tenue mécanique de la couverture ou la réactivation mécanique des failles. En modélisation de bassin, la modélisation couplée de l'écoulement et de la compaction en grandes déformations est actuellement simplifiée à l'aide de modèles 1D qui ne sont pas satisfaisants dans le cas de tectoniques complexes. Des recherches sont en cours sur les lois de comportement à l'échelle des bassins qui doivent mener à terme à des modélisations 3D couplées.

En simulation de réservoir et du stockage du CO_2, le couplage est traité dans le milieu industriel par un couplage externe de codes spécialisés et très riches chacun dans leur domaine propre : le code d'écoulement polyphasique compositionnel (incluant la thermique) en milieu poreux (PumaFlow™ ou COORES™ chez IFPEn) et le code de mécanique (chez IFPEn il s'agit du code *open source* Code_Aster ou du code commercial ABAQUS®).

Les codes de mécanique utilisent des méthodes de discrétisation de type éléments finis et des maillages conformes sauf si le modèle est lui-même discontinu (contact par exemple). Les codes d'écoulement en milieu poreux utilisent des méthodes de discrétisation volumes finis (habituellement centrés) et le maillage standard est de type Corner Point Geometry (CPG) [69]. Bien que conçus à partir d'une grille hexaédrique structurée, les maillages CPG ne sont pas compatibles avec les codes éléments finis classiques pour plusieurs raisons :

- les mailles hexaédriques dégénèrent du fait des érosions (*pinch out*) en plusieurs types de mailles non-standards, pouvant présenter des faces non-planes, ou pouvant être définies par moins de 8 sommets (7, 6 ou 5 par exemple) ;
- le raffinement local (LGR pour *Local Grid Refinement*), par exemple utilisé au voisinage des puits, de régions d'intérêt, ou lorsque des fronts se propagent, est habituellement non-conforme ;
- les failles sont modélisées par des dédoublements de nœuds et glissements de nœuds le long des directrices qui génèrent des non-conformités complexes (avec trous et recouvrements).

On donne Figure I.1 un exemple schématique 2D de maillage CPG. À noter qu'en 2D les dégénérescences relatives aux érosions (ou aux failles) sont bien moins dramatiques qu'elles ne peuvent l'être en 3D.

Pour réaliser le couplage éléments finis-volumes finis, il faut donc remailler localement le maillage CPG. C'est relativement aisé dans le cas des mailles hexaédriques dégénérées et du raffinement local mais complexe à réaliser proprement dans le cas des failles. Il faut ensuite effectuer les calculs d'interpolation afférents entre maillages 3D. Enfin, le couplage externe

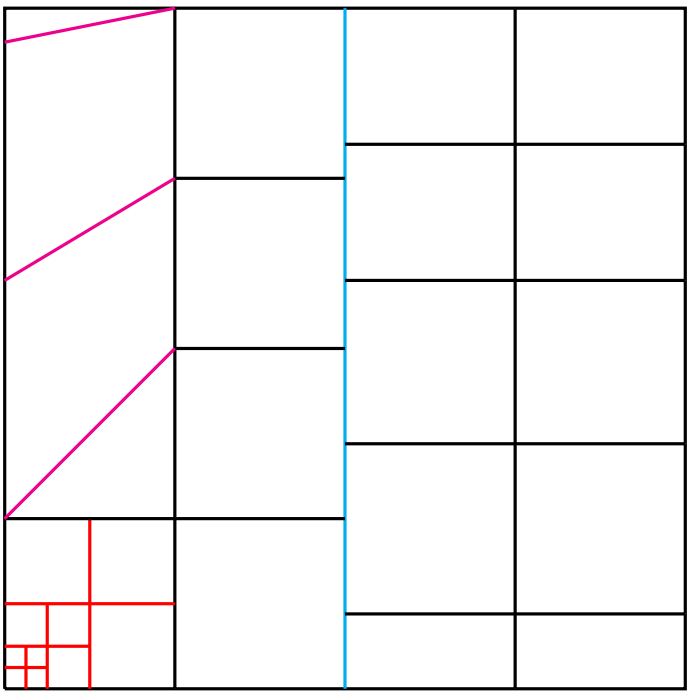


FIGURE I.1 – Exemple schématique 2D de maillage CPG avec LGR (en rouge), érosions (violet) et faille (bleu).

mécanique-écoulement est réalisé via une méthode séquentielle [52] :

- soit via une méthode itérative qui consiste (à chaque pas de temps) à résoudre à tour de rôle l'écoulement et la mécanique en en fixant l'un des deux, ceci jusqu'à convergence vers un point fixe (la convergence de certains de ces algorithmes vers la solution du problème parfaitement couplé a été récemment prouvée par Mikelić et Wheeler [59]) ;
- soit via une méthode explicite (et moins précise) qui consiste à ne faire qu'une seule itération de la procédure précédente ;
- soit via une méthode de couplage approximatif (*loosely coupling method* en anglais) qui consiste à ne résoudre la mécanique qu'après un certain nombre de résolutions de l'écoulement (et donc de pas de temps), ce qui permet de diminuer le coût de résolution mais ce qui nécessite des estimateurs fiables de quand mettre à jour la réponse mécanique.

Au final, la lourdeur des opérations géométriques et numériques nécessaires (remaillage local, interpolation 3D, couplage externe séquentiel) est une des raisons pour lesquelles les couplages mécanique-écoulement ne sont pas maîtrisés industriellement actuellement chez IFPEn.

Cette thèse se propose donc d'étudier une alternative qui consiste à traiter la mécanique à l'aide de méthodes de discrétisation non-conformes, pouvant donc être utilisées sur le même maillage que l'écoulement, typiquement de type CPG. Le raccordement aux épontes est aussi facilité par l'utilisation de maillages non-conformes. Un autre avantage des méthodes non-conformes est qu'elles prennent mieux en compte les discontinuités de propriétés qui sont importantes dans le cas des réservoirs et des bassins. On s'intéresse par ailleurs à des méthodes de discrétisation de plus bas ordre. Ce choix est justifié à la fois par l'incertitude inhérente aux données physiques que l'on incorpore à notre modèle, et à la fois par le besoin de maintenir les coûts numériques à l'intérieur de bornes acceptables. Une fois donnée une discrétisation non-conforme de la mécanique, alors cette dernière et l'écoulement peuvent être traités dans un même code, et

les opérations de remaillage et d'interpolation ne sont plus nécessaires. La résolution du système linéaire peut être assurée de manière parfaitement couplée (ce qui demande l'utilisation de solveurs élaborés pour des systèmes de grande taille), ou peut être réalisée de manière séquentielle comme expliqué plus haut. Il est important de noter pour conclure que la discrétisation de la mécanique et de l'écoulement sur une même grille est tout à fait dans l'ordre des choses étant donné que les hétérogénéités des paramètres décrivant ces deux physiques coïncident souvent (puisque dépendant tous deux du type de roche considéré).

La famille la plus générale de méthodes non-conformes est celle des méthodes de Galerkin discontinu (dG). Elles ont déjà été étudiées récemment pour la mécanique [77, 85] et pour la poroélasticité avec succès [67]. En plus d'être adaptées aux maillages généraux y compris non-conformes, elles ont l'avantage de pouvoir monter en ordre très facilement. Leur principal inconvénient est le grand nombre de degrés de liberté qu'elles engendrent et donc le coût de résolution des systèmes. On se concentre ainsi sur l'étude de méthodes moins coûteuses de type volumes finis. Très peu d'approches existent en volumes finis pour la mécanique. On peut citer [73] qui propose un schéma volumes finis centré à base d'interpolations par moindres carrés qui ne respectent pas les discontinuités. On peut également citer le travail récent de Nordbotten [64] sur l'adaptation des méthodes Volumes Finis Multi-Points (MPFA) au cas (vectoriel) de la mécanique. Les méthodes centrées (ne faisant intervenir que des inconnues de mailles) sont très peu coûteuses mais ont souvent l'inconvénient de ne pas être inconditionnellement stables (la plupart de ces méthodes ne sont pas symétriques) et de ne pas permettre de dériver facilement des critères de stabilité. Un deuxième inconvénient est qu'elles ne disposent pas d'un cadre théorique très solide dans lequel les étudier, à la différence des éléments finis non-conformes par exemple. On s'intéresse ainsi dans ce travail aux schémas Volumes Finis Hybrides (HFV) [40], issus de travaux récents sur la discrétisation des problèmes de diffusion sur maillages généraux. Ces schémas font partie d'une plus vaste famille qui est celle des méthodes Hybrides Mimétiques Mixtes (HMM) [32], regroupant dans un cadre unifié les méthodes de Volumes Finis Hybrides, de Différences Finies Mimétiques (MFD) [20, 18] et de Volumes Finis Mixtes (MFV) [31]. Les Volumes Finis Hybrides ont l'avantage de pouvoir également s'interpréter comme des éléments finis non-conformes. Cette idée est à la base du travail récent de Di Pietro [23] sur les méthodes de Galerkin centrées aux mailles (ccG), qui mélangent des concepts hérités des éléments finis et des volumes finis. Les méthodes ccG sont fondées sur la définition d'un espace polynomial incomplet sur maillages généraux, dont la construction (héritée des méthodes MPFA) ne repose que sur des inconnues de mailles, et possédant des propriétés d'approximation optimales. L'espace en question est ensuite utilisé dans des formulations discrètes inspirées des méthodes dG, où la consistance, la symétrie et la coercivité sont pénalisées directement dans la forme bilinéaire.

C'est cette vision un peu charnière, héritée d'une direction industrielle plusieurs fois changeante (Roland Masson, puis Daniele A. Di Pietro, puis Léo Agélas) combinée à une direction académique elle restée très stable (Robert Eymard), que nous allons adopter dans ce manuscrit, essayant de tirer profit des avantages de chacun de ces différents cadres.

I.2 Plan du manuscrit

Comme il convient de commencer quelque part, nous nous concentrons dans ce manuscrit sur un modèle de poroélasticité quasi-statique monophasique. Ainsi, nous considérons le cas de milieux poreux linéaires (possiblement hétérogènes), saturés par un fluide visqueux faiblement compressible, et pour lesquels les effets d'inertie sur la structure mécanique sont négligeables (le domaine est notamment considéré comme étant fixe). Ce cas que nous qualifierons d'école est bien entendu très éloigné de la réalité, mais il est à la base d'une compréhension du modèle

et de ses difficultés ainsi que d'une complexification future de celui-ci. En effet, concernant la mécanique par exemple, savoir traiter un modèle élastique (que nous supposerons d'ailleurs isotrope dans ce manuscrit) est à la base du traitement de physiques plus élaborées. Lorsque l'on dispose d'un espace discret (ou d'une formulation discrète) pour lequel (laquelle) on sait prouver une inégalité de Korn, on est assuré d'avoir une discrétisation coercive du modèle. Passer du cas isotrope au cas de lois de Hooke plus générales reposant sur des tenseurs (admissibles) de raideur d'ordre 4 se fait ensuite sans difficulté. Les modèles d'élasticité non-linéaire ou d'élastoplasticité sont également intimement liés à l'élasticité linéaire.

Le mansucrit est organisé comme suit.

Dans le Chapitre II, on présente successivement les modèles d'élasticité linéaire et de poroélasticité considérés. Une fois ceux-ci introduits, on fait un inventaire des difficultés liées à l'approximation numérique de chacun d'entre eux, avant de présenter un état de l'art documenté nous permettant de définir les orientations des chapitres suivants.

Pour l'élasticité, on aborde le problème de coercivité inhérent à une approximation non-conforme du modèle, ainsi que le problème de verrouillage numérique. Le verrouillage numérique (ou *locking*) se produit lorsqu'une contrainte d'incompressibilité est imposée à l'espace d'approximation. Si l'espace considéré n'est pas adapté à l'approximation des divergences (qui représentent les variations de volume dans le milieu), alors les résultats numériques obtenus sont de piètre qualité. C'est le cas par exemple des éléments finis de Lagrange de plus bas ordre. On fait le lien entre la bonne approximation de l'opérateur divergence et la stabilité du couplage dans l'approximation d'un problème de point-selle tel qu'un problème de Stokes. Il s'avère que la robustesse vis-à-vis du *locking* passe par la vérification d'une condition de type inf-sup (ou de manière équivalente l'existence d'un opérateur de Fortin) entre l'espace de discrétisation des déplacements et l'espace discret (qui serait l'espace des pressions dans un problème de Stokes) sur lequel projeter l'opérateur divergence. S'inspirant de cette constatation et de l'état de l'art en matière d'approximation de modèles d'élasticité, nous décidons de baser notre discrétisation sur l'élément fini de Crouzeix–Raviart [22], qui est non-conforme et qui a la bonne propriété de savoir approximer les divergences, puisque possédant ses degrés de liberté sur les faces de la maille. L'adaptation de cet élément au cas de maillages généraux est réalisée au Chapitre III par le biais d'une analogie avec les méthodes HFV.

Pour la poroélasticité, on aborde le problème du couplage mécanique-écoulement. Pour des temps courts, le terme darcéen (cf. (II.18b)) lié à la pression de pore fournit une contribution quasi-nulle au modèle, ce qui signifie que la pression n'intervient que très peu par l'intermédiaire de son gradient. Tout se passe (du moins lorsque $c_0 = 0$) comme dans un problème de Stokes, à la différence près que dès que $t > 0$, on impose des conditions de bord. Le gradient de pression ne commence à contribuer au modèle que pour des temps plus avancés et pour peu que la perméabilité du milieu ne soit pas trop faible. D'un point de vue numérique, l'approximation de ce genre de phénomène s'avère compliquée. En effet, l'existence du terme darcéen suggère de considérer des pressions discrètes qui soient au minimum affines par maille. Or, dans les premiers pas de temps, on ne peut avoir de contrôle sur la pression au mieux que via sa reconstruction, ce qui s'avère insuffisant. Ce manque de contrôle sur le gradient de pression se paie en termes d'oscillations spatiales parasites. On discute dans le Chapitre II des différentes techniques existant pour contrôler ce phénomène oscillatoire (cf. également Chapitre V).

Dans le Chapitre III, on se concentre sur l'approximation du modèle d'élasticité. On présente la construction d'un espace d'approximation sur maillages généraux, dont les propriétés

ressemblent de très près à celles de l'espace de Crouzeix–Raviart classique : propriétés d'approximation optimales (dont l'existence d'un opérateur de Fortin permettant de préserver la divergence par maille de la fonction interpolée) et de conformité faible (i.e. la continuité des fonctions au barycentre des interfaces d'un sous-maillage). La construction de cet espace se base sur le gradient Volumes Finis Hybrides et s'inspire de la philosophie des méthodes ccG.

On définit sur un sous-maillage pyramidal (que l'on prouve être régulier au sens éléments finis) fictif (dans le sens où aucune information n'a besoin d'être stockée à son sujet) du maillage initial une reconstruction affine, partant pour chaque pyramide (associée à une face de la maille) de l'inconnue de face en question, et se déplaçant selon le gradient introduit dans les méthodes HFV. On prouve que l'espace ainsi engendré, et s'apparentant à une généralisation de l'espace de Crouzeix–Raviart, possède toutes les qualités nécessaires pour l'approximation non-conforme du modèle d'élasticité, dans le sens où sa conformité faible et ses propriétés d'approximation sur maillages généraux en font un espace d'approximation à part entière, pour lequel l'existence d'un opérateur de Fortin garantit le bon traitement des divergences. On investigue également le cas d'un maillage simplicial conforme, et on prouve notamment que l'espace de Crouzeix–Raviart est inclus dans l'espace ainsi construit. De plus, l'espace ainsi construit ne diffère de l'*espace* (la notion d'espace est un peu galvaudée dans ce contexte) HFV que par le caractère affine de la reconstruction, le gradient étant identique.

C'est précisément l'introduction de cette reconstruction qui, d'une part permet d'analyser cet espace sous le nouveau jour qu'est celui des éléments finis non-conformes, et d'autre part permet d'envisager pour la discrétisation non-conforme de l'élasticité un traitement du problème de coercivité par une stabilisation des sauts, technique inspirée des méthodes dG.

Dans le Chapitre IV, on utilise l'espace précédemment construit pour approximer un modèle d'élasticité linéaire. On propose une méthode primale (par opposition à mixte) d'approximation du champ de déplacement qui est inconditionnellement stable sur maillages généraux et qui est robuste au *locking*. Le traitement du problème de coercivité est basé comme nous l'avons dit sur une stabilisation des sauts héritée des méthodes dG. Cette pénalisation permet d'obtenir une inégalité de Korn (faible) discrète qui garantit la stabilité.

Cette méthode nécessite de considérer un degré de liberté par composante du champ de déplacement pour chaque face et chaque cellule du maillage. Ce n'est donc pas la méthode la moins coûteuse que l'on puisse imaginer mais le rapport entre son coût et les propriétés qu'elle assure reste très bon en comparaison à un équivalent (en termes de propriétés) élément fini $\mathbb{P}_2^d$. Il est à noter que dans certains cas il est possible d'éliminer localement les inconnues de maille, réduisant ainsi la taille du système. On considère également le cas d'un matériau hétérogène pour lequel on propose une adaptation de la méthode initiale. On étudie par ailleurs le lien entre la méthode proposée et les méthodes volumes finis et éléments finis classiques.

Finalement, on propose une série de tests numériques attestant du bon comportement du schéma, dans le traitement de problèmes hétérogènes, de *locking*, ou d'approximation sur des grilles générales. Des comparaisons sont proposées avec une méthode élément fini $\mathbb{P}_1^d$. Tous ces tests sont réalisés en deux dimensions d'espace grâce à une implémentation prototype `C++` basée sur le cadre abstrait introduit par Di Pietro, Gratien et Prud'homme [27].

Dans le Chapitre V, on étudie la convergence d'une famille de méthodes pour le problème de poroélasticité. Cette famille de méthodes, appelée schémas Euler-Gradient, repose sur une discrétisation Euler implicite en temps et Gradient en espace. La discrétisation Gradient, introduite par Eymard *et al.* [45, 41, 33], repose sur un cadre abstrait englobant une large classe

de méthodes d'approximation pour des problèmes elliptiques linéaires ou non-linéaires (voire non-locaux). Une discrétisation Gradient est, dans sa version la plus simple, définie par la donnée de trois éléments : un espace de degrés de liberté, un opérateur de reconstruction sur cet espace (permettant de définir la reconstruction des fonctions approchées), et un opérateur gradient (également défini à partir de l'espace de degrés de liberté). Ce formalisme, qui regroupe notamment les éléments finis conformes, la plupart des éléments finis non-conformes, certaines méthodes MPFA, le schéma VAG [42, 41], ainsi que les méthodes HMM, se base sur quatre principales hypothèses à vérifier par les schémas : une hypothèse de coercivité qui s'exprime comme une inégalité de Friedrichs ou de Poincaré uniforme, une hypothèse d'approximation optimale (souvent dénommée, improprement au sens éléments finis, consistance), une hypothèse de conformité limite qui signifie que les opérateurs gradient et de reconstruction vérifient à la limite une formule de Green continue, ainsi qu'une hypothèse de compacité (qui permet de contrôler les translations en espace et qui ne sert que dans le cas non-linéaire).

Nous basant sur ce formalisme, nous définissons une discrétisation Gradient à la fois pour le déplacement et la pression, pour lesquelles nous faisons l'hypothèse supplémentaire de disposer d'une condition inf-sup sur la reconstruction de pression. Dans notre cas, l'hypothèse de compacité n'est pas nécessaire car le problème est linéaire. Nous démontrons l'existence et l'unicité de la solution du schéma Euler-Gradient ainsi établi, ainsi que sa convergence vers l'unique solution de régularité minimale (à savoir $L^2(0,T;H^1(\Omega)^d)$ pour le déplacement et $L^2(0,T;H^1(\Omega))$ pour la pression lorsque celle-ci n'intervient pas dans la dérivée en temps) du problème continu. Plus précisément, nous démontrons la convergence forte du gradient de déplacement ainsi que de la reconstruction de pression (ce dernier résultat est basé sur la condition inf-sup), ainsi que la convergence faible du gradient de pression et de la reconstruction de déplacement. Ce résultat de convergence est valable pour toutes les valeurs (admissibles) pouvant être prises par les paramètres physiques (on considère notamment le cas de matériaux potentiellement quasi-incompressibles et de zones potentiellement peu perméables dans le milieu).

Ces résultats théoriques sont validés sur une série de cas-tests bi-dimensionnels, réalisés sur la même plateforme prototype que ceux du chapitre précédent, et comparés à une méthode éléments finis $\mathbb{P}_1^d/\mathbb{P}_1$ connue pour ne pas vérifier d'inf-sup. On étudie une discrétisation Gradient en espace particulière, basée sur un traitement de l'élasticité linéaire fondé sur l'espace de Crouzeix–Raviart généralisé développé aux Chapitres III et IV, et sur un traitement de la pression fondé sur une méthode HFV (disposant donc du même type de gradient mais dont la reconstruction est constante par maille, autorisant ainsi la vérification d'une condition inf-sup telle que celle supposée dans les hypothèses). On teste d'abord le comportement du schéma dans un cas homogène à perméabilité suffisamment grande, sur les temps courts. La reconstruction de pression converge en espace mais présente des oscillations parasites assez prononcées. La vérification d'une condition inf-sup ne suffit donc pas à les éliminer. Nous donnons une explication du phénomène : la condition inf-sup vérifiée ici est très différente de celles habituellement rencontrées dans le cadre éléments finis dans le sens où la reconstruction de pression est constante par maille au lieu d'être affine. Le contrôle ne se fait donc que sur une projection de la pression affine, ce qui est insuffisant pour réduire de manière efficace les oscillations. Il semble en même temps délicat de vérifier une condition inf-sup au sens éléments finis quand les discrétisations du déplacement et de la pression sont toutes deux affines. Nous exposons d'autres techniques possibles pour pallier à ce problème, voir également les perspectives Chapitre VI. Par ailleurs, le schéma se comporte très bien en temps long, la stabilisation de la pression dûe au terme darcéen opère. On teste ensuite le comportement du schéma sur grilles générales et dans un cas hétérogène (la perméabilité est constante par morceaux) avec une zone peu perméable. Lorsque la perméabilité de cette zone est suffisamment grande, tout se passe comme dans le cas (homogène) précédent. Par contre, lorsque la perméabilité est trop basse, les résultats se dé-

gradent. La convergence de la reconstruction de pression est toujours assurée mais les résultats se détériorent avec l'augmentation du temps de simulation. La stabilisation normalement dûe au terme darcéen n'opère pas. Ces problèmes (rencontrés également avec une méthode $\mathbb{P}_1^d/\mathbb{P}_1$) semblent également provenir d'un manque de stabilisation de la pression et indiquent ainsi que la constante de stabilisation doit être proportionnelle à l'inverse de la plus petite perméabilité. Il semble que si la pression n'est pas stabilisée rapidement (par un terme de diffusion prenant de l'importance dès les premiers pas de temps ou par d'autres techniques), alors on constate une instabilité en temps long lorsque trop d'erreurs se sont ajoutées. Par ailleurs, la robustesse de la méthode sur maillages généraux est validée.

Le Chapitre VI présente quelques conclusions et les perspectives immédiates ou à plus long terme de ce travail. Notamment, il s'agira de se pencher plus avant sur des méthodes de stabilisation efficaces de l'approximation de pression dans les premiers pas de temps ou dans des zones peu perméables.

Les annexes, au nombre de trois, présentent des travaux réalisés en marge de la ligne directrice de ce manuscrit.

On introduit en Annexe A une généralisation, inspirée du Chapitre III, de l'espace de Raviart–Thomas de plus bas ordre au cas des maillages généraux. Les propriétés principales de cet espace (conformité $\boldsymbol{H}(\mathrm{div};\Omega)$ et approximation des divergences) sont dupliquées et étendues.

En Annexe B, on présente une discrétisation inf-sup stable du problème de Stokes quasi-statique basée sur un couplage de l'espace de Crouzeix–Raviart généralisé introduit au Chapitre III avec une discrétisation centrée de la pression. On s'intéresse tout particulièrement au cas du traitement numérique des larges forçages irrotationnels dans le cas où une décomposition de Helmholtz du second membre est connue au niveau continu. On montre qu'un traitement adéquat du second membre permet de s'affranchir de l'influence de sa partie irrotationnelle sur l'approximation de la vitesse (qui au niveau continu ne dépend pas de la partie irrotationnelle du terme source). On illustre ce résultat sur un cas-test 2D en utilisant la discrétisation introduite auparavant.

Enfin, en Annexe C, on présente un moyen d'obtenir une inégalité de Korn discrète et donc la coercivité d'une approximation Volumes Finis Hybrides de l'élasticité linéaire. On se place donc dans le cas où la reconstruction considérée est constante par maille, et où une stabilisation par les sauts n'est pas envisageable. La stabilisation de la forme bilinéaire passe par la réduction du nombre de degrés de liberté par interpolation de la (des) composante(s) tangentielle(s) du déplacement aux faces du maillage. Cette interpolation permet d'ajouter la rigidité nécessaire au système pour contrôler les mouvements de corps rigide, tout en permettant de garantir la robustesse au *locking* et la stabilité du couplage avec des pressions centrées grâce au fait que les inconnues normales aux faces sont préservées. Le comportement de la méthode est testé en 2D et en 3D (les tests 3D ont été réalisés par Roland Masson sur un prototype Fortran 3D) sur différentes grilles, et les résultats sont très encourageants. Le principal inconvénient de la méthode réside dans le fait qu'il n'existe pas à l'heure d'aujourd'hui de preuve qu'une inégalité de Korn est bien vérifiée sur l'espace associé. Son principal avantage réside dans le peu de degrés de liberté qu'elle engendre (un degré de liberté par face du maillage après interpolation de la (des) composante(s) tangentielle(s) et élimination locale (qui fonctionne ici dans tous les cas) des inconnues de maille).

Chapitre II

From linear elasticity to poroelasticity

Sommaire

In this chapter we present the physical model that we consider as our poroelasticity reference problem, beginning with an introduction of the linear elasticity equations. We give an overview, sometimes merely based on a heuristic approach, of the different problems related to the numerical approximation of such models. Finally, we provide a state of the art in terms of approximation, from which we take advantage to clearly define our approximation choices and the orientation of the following chapters.

From now on, we denote by Ω a bounded connected open polygonal or polyhedral subset of $\mathbb{R}^d$, where $d \in \{2, 3\}$ stands for the space dimension. Its boundary is denoted by Γ, with unit outward normal $\boldsymbol{n}$.

II.1 The linear elasticity model

II.1.1 Continuous setting

We consider a linearly elastic, isotropic, and homogeneous medium occupying the domain Ω. The linear behavior of the material implies that we enter the framework of infinitesimal strain theory, meaning in particular that the geometry and the constitutive properties of the material at each point of space can be assumed to be unchanged by the deformation (Ω is in particular a fixed domain). We also neglect the inertia effects in the structure, this is the so-called quasistatic assumption.

In these conditions, the linear elasticity problem consists in finding a vector-valued displacement field $\boldsymbol{u} : \Omega \to \mathbb{R}^d$ such that

$$\begin{aligned} -\boldsymbol{\nabla}\cdot\underline{\underline{\sigma}}(\boldsymbol{u}) &= \boldsymbol{f} \qquad && \text{in } \Omega, \\ \boldsymbol{u} &= \boldsymbol{0} && \text{on } \Gamma_{\mathrm{D}}, \\ \underline{\underline{\sigma}}(\boldsymbol{u})\boldsymbol{n} &= \boldsymbol{0} && \text{on } \Gamma_{\mathrm{N}}, \end{aligned} \tag{II.1}$$

where Γ_{D} and Γ_{N} are such that Γ_{D} has nonzero measure, $\Gamma_{\mathrm{D}} \cap \Gamma_{\mathrm{N}} = \varnothing$, and $\Gamma_{\mathrm{D}} \cup \Gamma_{\mathrm{N}} = \Gamma$. The sets Γ_{D} and Γ_{N} are respectively associated to Dirichlet and Neumann boundary conditions. We assume, and this is the sense of the assumption on the measure of Γ_{D}, that the displacement is always prescribed at least on one part of the boundary. Thus, we do not consider the pure traction problem but we may treat the pure displacement one when $\Gamma_{\mathrm{D}} = \Gamma$. For the sake of simplicity, we consider homogeneous boundary conditions. The nonhomogeneous case could be handled similarly. The vector-valued field $\boldsymbol{f} : \Omega \to \mathbb{R}^d$ is the body force per unit volume (for example the gravity) and $\underline{\underline{\sigma}}(\boldsymbol{u})$ is the Cauchy stress tensor given by Hooke's law of linear elasticity, which reads for an isotropic material:

$$\underline{\underline{\sigma}}(\boldsymbol{u}) := 2\mu\,\underline{\underline{\varepsilon}}(\boldsymbol{u}) + \lambda\boldsymbol{\nabla}\cdot\boldsymbol{u}\,\underline{\underline{I_d}}, \qquad \underline{\underline{\varepsilon}}(\boldsymbol{u}) := \frac{1}{2}(\boldsymbol{\nabla}\boldsymbol{u} + \boldsymbol{\nabla}\boldsymbol{u}^T),$$

where $\underline{\underline{\varepsilon}}(\boldsymbol{u})$ is the infinitesimal strain tensor, and $\underline{\underline{I_d}}$ the identity tensor in $\mathbb{R}^d$. The constitutive properties of the material are described by the two constants λ and μ (the fact that λ and μ are constant with respect to the space variable and the deformation is a consequence of the homogeneity and infinitesimal strain assumptions respectively), referred to as Lamé parameters and homogeneous to a pressure. Another way to describe the material makes use of its Poisson's ratio ν (dimensionless) and elastic modulus E (homogeneous to a pressure), which are related to the Lamé parameters through

$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)}, \qquad \mu = \frac{E}{2(1+\nu)}. \tag{II.2}$$

The second Lamé parameter μ, also called shear modulus (and denoted G), is strictly positive and assumed to be bounded away from zero and from infinitely large values. The first Lamé parameter λ (related to the shear modulus and to the bulk modulus K by $\lambda = K - \frac{2}{3}G$ in 3D and $\lambda = K - G$ in 2D) is also assumed to be strictly positive (physically it can possibly be negative but it is positive for most materials) and bounded away from zero, but it may take unboundedly large values. As it is associated in the model to $\boldsymbol{\nabla}\cdot\boldsymbol{u}$, which represents the variations of volume in the medium, this parameter is associated to the compressibility of the material. The case $\lambda = +\infty$ (which corresponds to a Poisson's ratio $\nu = 0.5$) occurs when an incompressible material is considered. In that case, we have $\boldsymbol{\nabla}\cdot\boldsymbol{u} = 0$ (cf. [6] for an example of approximation of such a problem). From a practical point of view, the medium is never

completely incompressible but tends to be so. Hence we do not consider the case $\lambda = +\infty$ but only the case $\lambda \to +\infty$, which describes a quasi-incompressible behavior. We will see in the next section that this limit behavior leads to numerical difficulties in the approximation of the model.

Coherently with the context, we indifferently denote by $(\cdot,\cdot)_{0,\Omega}$ the scalar products in $L^2(\Omega)$, $L^2(\Omega)^d$, and $L^2(\Omega)^{d,d}$, which are respectively defined by $(w,v)_{0,\Omega} := \int_\Omega wv\,\mathrm{d}\boldsymbol{x}$ in $L^2(\Omega)$, by $(\boldsymbol{w},\boldsymbol{v})_{0,\Omega} := \int_\Omega \boldsymbol{w}\cdot\boldsymbol{v}\,\mathrm{d}\boldsymbol{x}$ in $L^2(\Omega)^d$, and by $(\underline{\underline{w}},\underline{\underline{v}})_{0,\Omega} := \int_\Omega \underline{\underline{w}}:\underline{\underline{v}}\,\mathrm{d}\boldsymbol{x}$ in $L^2(\Omega)^{d,d}$. The corresponding norms are as well indifferently denoted by $\|\cdot\|_{0,\Omega}$. We will introduce more systematic notations in Chapter III, Section III.1.3.

In order to write the weak formulation of problem (II.1), we introduce the space

$$H^1_{\mathrm{D}}(\Omega) := \{v \in H^1(\Omega) \mid v_{|\Gamma_{\mathrm{D}}} = 0\},$$

which reduces to the classical $H^1_0(\Omega)$ space when $\Gamma_{\mathrm{D}} = \Gamma$. Thanks to Friedrichs' inequality, there holds that $\|\boldsymbol{\nabla} v\|_{0,\Omega}$ is a norm on $H^1_{\mathrm{D}}(\Omega)$. Assuming that $\boldsymbol{f} \in L^2(\Omega)^d$, the weak formulation of problem (II.1) reads: Find $\boldsymbol{u} \in H^1_{\mathrm{D}}(\Omega)^d$ such that

$$a(\boldsymbol{u},\boldsymbol{v}) = (\boldsymbol{f},\boldsymbol{v})_{0,\Omega} \qquad \forall \boldsymbol{v} \in H^1_{\mathrm{D}}(\Omega)^d, \tag{II.3}$$

where $a(\boldsymbol{w},\boldsymbol{v}) := (\underline{\underline{\sigma}}(\boldsymbol{w}),\underline{\underline{\varepsilon}}(\boldsymbol{v}))_{0,\Omega} = 2\mu(\underline{\underline{\varepsilon}}(\boldsymbol{w}),\underline{\underline{\varepsilon}}(\boldsymbol{v}))_{0,\Omega} + \lambda(\boldsymbol{\nabla}\cdot\boldsymbol{w},\boldsymbol{\nabla}\cdot\boldsymbol{v})_{0,\Omega}$. There holds as an immediate consequence, using the fact that λ is a strictly positive constant,

$$a(\boldsymbol{v},\boldsymbol{v}) \geqslant 2\mu\,\|\underline{\underline{\varepsilon}}(\boldsymbol{v})\|^2_{0,\Omega} \qquad \forall \boldsymbol{v} \in H^1_{\mathrm{D}}(\Omega)^d.$$

Hence, since μ is bounded away from zero, the well-posedness of the weak formulation (II.3) relies on Korn's inequality in $H^1_{\mathrm{D}}(\Omega)^d$ (cf., e.g., [14, Remark 1.1], or [3, Theorems 5.3.2 and 5.3.4]).

Lemma II.1 (Korn's inequality)**.** *There exists a constant $C_{\Omega,\Gamma_{\mathrm{D}}}$, whose dependencies are specified in subscript, such that*

$$\|\boldsymbol{\nabla}\boldsymbol{v}\|_{0,\Omega} \leqslant C_{\Omega,\Gamma_{\mathrm{D}}}\,\|\underline{\underline{\varepsilon}}(\boldsymbol{v})\|_{0,\Omega} \qquad \forall \boldsymbol{v} \in H^1_{\mathrm{D}}(\Omega)^d, \tag{II.4}$$

and $C_{\Omega,\Gamma_{\mathrm{D}}} = \sqrt{2}$ in the case $\Gamma_{\mathrm{D}} = \Gamma$.

This inequality is mandatory to prove the coercivity of the formulation, since it gives a control of the full gradient by its symmetric part. It implies that no rigid body motion is applied to the structure. A rigid body motion is a motion with vanishing elastic energy, i.e. of the form $\boldsymbol{v}(\boldsymbol{x}) = \boldsymbol{a} + \underline{\underline{B}}\boldsymbol{x}$, with $\boldsymbol{a} \in \mathbb{R}^d$ and $\underline{\underline{B}}$ an anti-symmetric tensor. We will see in the next section that, from a discrete point of view, find a lowest-order nonconforming approximation space satisfying that kind of inequality is not an easy task. Combining Korn's inequality (II.4) and the fact that $\|\boldsymbol{\nabla}\boldsymbol{v}\|_{0,\Omega}$ is a norm on $H^1_{\mathrm{D}}(\Omega)^d$ completes the proof of well-posedness of problem (II.3).

Remark II.1 (Pure displacement problem)**.** *In the case $\Gamma_{\mathrm{D}} = \Gamma$, the weak formulation* (II.3) *can be rewritten into the equivalent form: Find $\boldsymbol{u} \in H^1_0(\Omega)^d$ such that*

$$\tilde{a}(\boldsymbol{u},\boldsymbol{v}) = (\boldsymbol{f},\boldsymbol{v})_{0,\Omega} \qquad \forall \boldsymbol{v} \in H^1_0(\Omega)^d, \tag{II.5}$$

where $\tilde{a}(\boldsymbol{w},\boldsymbol{v}) := \mu(\boldsymbol{\nabla}\boldsymbol{w},\boldsymbol{\nabla}\boldsymbol{v})_{0,\Omega} + (\mu+\lambda)(\boldsymbol{\nabla}\cdot\boldsymbol{w},\boldsymbol{\nabla}\cdot\boldsymbol{v})_{0,\Omega}$. This comes from the relation

$$\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\boldsymbol{\varphi}^T) = \boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\cdot\boldsymbol{\varphi}\underline{\underline{I_d}}), \tag{II.6}$$

valid for any sufficiently regular function, and especially for $\boldsymbol{\varphi} \in C_c^\infty(\Omega)^d$*, which enables to state, using two integrations by parts and a density argument, that*

$$\mu(\boldsymbol{\nabla w}, \boldsymbol{\nabla v}^T)_{0,\Omega} = \mu(\boldsymbol{\nabla w}, \boldsymbol{\nabla \cdot v}\underline{\underline{I_d}})_{0,\Omega} = \mu(\boldsymbol{\nabla \cdot w}, \boldsymbol{\nabla \cdot v})_{0,\Omega}, \tag{II.7}$$

for all $\boldsymbol{w}, \boldsymbol{v} \in H_0^1(\Omega)^d$*. The same argument is used to prove Korn's inequality* (II.4) *in* $H_0^1(\Omega)^d$ *and derive the multiplicative constant. The equivalence between the two formulations only holds if (i) pure Dirichlet boundary conditions are considered (even nonhomogeneous since the weak formulation can still be written in* $H_0^1(\Omega)^d$*, the solution is then obtained up to the addition of a lifting); (ii) the material is homogeneous (with constant Lamé parameters). The advantage of such a formulation is its natural coercivity, which does not rely on Korn's inequality* (II.4)*. From a discrete point of view, the treatment of the pure displacement problem is thus much easier.*

To conclude this introduction, we state a regularity result for problem (II.1) in dimension $d = 2$. A proof of that lemma (in the cases $\Gamma_{\mathrm{D}} = \Gamma$ and $\Gamma_{\mathrm{N}} = \Gamma$, and for nonhomogeneous boundary conditions) can be found, e.g., in the classical work of Brenner and Sung [16, Lemmata 2.2 and 2.3].

Lemma II.2 (Regularity)**.** *Let* $d = 2$ *and assume that* Ω *is convex. Then, problem* (II.1) *has a unique solution* $\boldsymbol{u} \in H_{\mathrm{D}}^1(\Omega)^d \cap H^2(\Omega)^d$*. Moreover, there exists a real* $C_\mu > 0$*, only depending on* Ω *and* μ *but not on* λ*, such that, for* λ *large enough,*

$$\mathcal{N}_{\mathrm{el}}(\boldsymbol{u}) := \|\boldsymbol{u}\|_{2,\Omega} + \lambda|\boldsymbol{\nabla \cdot u}|_{1,\Omega} \leqslant C_\mu \|\boldsymbol{f}\|_{0,\Omega}. \tag{II.8}$$

The notations $|\cdot|_{1,\Omega}$ and $\|\cdot\|_{2,\Omega}$ respectively refer to the classical seminorm in $H^1(\Omega)$ and norm in $H^2(\Omega)^d$ (cf. again Section III.1.3). This a priori estimate implies that if $\lambda \to +\infty$, the divergence of the displacement field approaches zero, corresponding to a quasi-incompressible behavior (the variations of volume in the medium tend to vanish). Note that an energy estimate enables to show that $\lambda^{1/2}\|\boldsymbol{\nabla \cdot u}\|_{0,\Omega}$ is bounded independently of λ. The restriction to $d = 2$ is a priori purely theoretical and we can assume that the same result may hold in $d = 3$. This regularity result is rather comforting from a physical point of view but, more practically and as we will see in the following section, it gives a useful tool to derive discretization error estimates that are robust with respect to the first Lamé parameter λ. In the case of nonhomogeneous boundary conditions, the right-hand side of this regularity estimate is modified accordingly to take them into account, see, e.g., [16]. A generalization of this result to composite materials with piecewise constant mechanical properties is proved in [29], see Remark IV.5.

II.1.2 Numerical issues

Throughout this section, we give an overview, sometimes only based on heuristic arguments, of the different problems related to the approximation of the linear elasticity problem (II.3). The aim here is not to be completely rigorous, but to explain roughly what are the different problems, in order to define in the next section, and in the light of what already exists in the literature, the best answers to give to these issues in the following chapters. We recall, see Chapter I, that owing to the applications we aim, we consider lowest-order approximation methods, which must handle possibly fairly general meshes.

II.1.2.1 A certain lack of coercivity

As we explained in Section II.1.1, the coercivity of the weak formulation (II.3) relies on Korn's inequality (II.4) in $H_{\mathrm{D}}^1(\Omega)^d$. From a discrete point of view, any conforming approximation

of the problem based on a finite element space $\boldsymbol{U}_h \subset H^1_{\mathrm{D}}(\Omega)^d$ is coercive, in the sense that a discrete Korn's inequality holds on $\boldsymbol{U}_h$ as a consequence of (II.4). However, such approximations have two main drawbacks:

(i) first, they are completely mesh-dependent since the dimension of the local polynomial space is directly related to the shape of the element. Hence, considering fairly general meshes may dramatically complicate the computation. In most of 3D industrial codes, only tetrahedral or hexahedral elements are considered and nonmatching interfaces are merely not handled;

(ii) secondly, the discretization obtained is not robust with respect to the first Lamé parameter λ, as we will detail in Section II.1.2.2.

As a consequence, despite of their natural well-posedness, conforming finite elements are not suited at all to our needs. Hence we have to consider nonconforming approximations, and thus discrete spaces on which a discrete Korn's inequality does not necessarily hold.

Let temporarily focus on the pure traction problem, i.e. $\Gamma_{\mathrm{N}} = \Gamma$. The pure displacement problem has no interest in that context since it can be rewritten into a coercive form, see Remark II.1. The pure traction problem is well-posed in the space

$$\boldsymbol{U} := \left\{ \boldsymbol{v} \in H^1(\Omega)^d \mid \int_\Omega \boldsymbol{v} \,\mathrm{d}\boldsymbol{x} = \boldsymbol{0}, \left| \int_\Omega \boldsymbol{\nabla}\times \boldsymbol{v} \,\mathrm{d}\boldsymbol{x} \right| = 0 \right\},$$

where $\boldsymbol{\nabla}\times$ is the classical rotation operator when $d = 2$, and curl operator when $d = 3$. The well-posedness of this problem means that Korn's inequality (II.4) holds on $\boldsymbol{U}$, see [14, Remark 1.1]. Let now see what happens on a nonconforming discrete level. If we restrict ourselves to a matching simplicial mesh, it is well known, as it has been pointed out by Falk [46], that the first order nonconforming space (spanned by piecewise affine functions that are continuous at the midpoint of mesh interfaces) does not fulfill a discrete Korn's inequality. To establish this fact, we use a dimension-counting argument.

First, note that this space is the so-called lowest-order Crouzeix–Raviart space introduced in [22]. Let $\mathcal{T}_h$ be a matching simplicial discretization of the domain Ω (h classically represents the maximum diameter of the mesh elements) and let $\mathbb{CR}(\mathcal{T}_h)$ denote the Crouzeix–Raviart space on $\mathcal{T}_h$. We introduce

$$\boldsymbol{U}_h := \left\{ \boldsymbol{v}_h \in \mathbb{CR}(\mathcal{T}_h)^d \mid \int_\Omega \boldsymbol{v}_h \,\mathrm{d}\boldsymbol{x} = \boldsymbol{0}, \left| \int_\Omega \boldsymbol{\nabla}_h\times \boldsymbol{v}_h \,\mathrm{d}\boldsymbol{x} \right| = 0 \right\},$$

where $\boldsymbol{\nabla}_h\times$ is the broken rotation or curl operator (defined from the broken gradient operator $\boldsymbol{\nabla}_h$ that will be rigorously introduced in Section III.1.3).
Let henceforth assume $d = 2$. Then, $\boldsymbol{U}_h$ has dimension $2\,\mathrm{card}(\mathcal{F}_{\mathcal{T}_h}) - 3$, where $\mathcal{F}_{\mathcal{T}_h}$ is the set of edges of the mesh. Thus, the subspace of $\boldsymbol{U}_h$ with $\underline{\underline{\varepsilon_h}}(\boldsymbol{v}_h) = \underline{\underline{0}}$ (where $\underline{\underline{\varepsilon_h}}$ is the broken infinitesimal strain tensor) has dimension greater or equal to $2\,\mathrm{card}(\mathcal{F}_{\mathcal{T}_h}) - 3\,\mathrm{card}(\mathcal{T}_h) - 3$, since this relation brings at most $3\,\mathrm{card}(\mathcal{T}_h)$ additional independent constraints (the infinitesimal strain tensor is piecewise constant and symmetric). As a consequence, using Euler relations (see, e.g., [35, Lemma 1.57]), this subspace has dimension greater or equal to $\mathrm{card}(\mathcal{F}^{\mathrm{b}}_{\mathcal{T}_h}) - 3$, where $\mathcal{F}^{\mathrm{b}}_{\mathcal{T}_h}$ is the subset of boundary edges, which means that it has strictly positive dimension as soon as $\mathcal{T}_h$ consists of more than one triangle. On the other hand, the dimension of the subspace of $\boldsymbol{U}_h$ with $\boldsymbol{\nabla}_h\boldsymbol{v}_h = \underline{\underline{0}}$ is clearly zero. Hence, there must exist functions in $\boldsymbol{U}_h$ for which Korn's inequality (II.4) fails.

The conclusion is: in a lowest-order nonconforming space, there is no reason for a discrete Korn's inequality to hold. Even more, and this statement will make sense in the following,

discrete spaces to which the Crouzeix–Raviart space belongs cannot fulfill a discrete Korn's inequality. In the state of the art, see Section II.1.3, we will see that different remedies exist to prove Korn's inequality on nonconforming spaces (reduced integration techniques, jumps penalization, order increasing, rigidity adding). We will discuss the advantages and drawbacks of each of these techniques.

II.1.2.2 Quasi-incompressible materials: the locking phenomenon

Assume $\Gamma_{\mathrm{D}} = \Gamma$ and consider the linear elasticity problem (II.3). Given a conforming finite element space $\boldsymbol{U}_h \subset H_0^1(\Omega)^d$, we search for $\boldsymbol{u}_h \in \boldsymbol{U}_h$ such that

$$a(\boldsymbol{u}_h, \boldsymbol{v}_h) = (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega} \qquad \forall \boldsymbol{v}_h \in \boldsymbol{U}_h, \tag{II.9}$$

where we recall $a(\boldsymbol{w}, \boldsymbol{v}) := 2\mu(\underline{\underline{\varepsilon}}(\boldsymbol{w}), \underline{\underline{\varepsilon}}(\boldsymbol{v}))_{0,\Omega} + \lambda(\boldsymbol{\nabla}\cdot\boldsymbol{w}, \boldsymbol{\nabla}\cdot\boldsymbol{v})_{0,\Omega}$. This problem is well-posed since Korn's inequality (II.4) does hold on the conforming space $\boldsymbol{U}_h$. We introduce the energy norm $\|\boldsymbol{v}\|_{\mathrm{el}} := a(\boldsymbol{v}, \boldsymbol{v})^{1/2}$ on $H_0^1(\Omega)^d$ (the fact that $\|\cdot\|_{\mathrm{el}}$ is a norm is a consequence of the symmetry of the bilinear form a and of its coercivity expressed by Korn's inequality). If $\boldsymbol{u} \in H_0^1(\Omega)^d$ denotes the unique solution to (II.3), then Céa's lemma gives the following estimate of the discretization error:

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{el}} \leqslant \inf_{\boldsymbol{v}_h \in \boldsymbol{U}_h} \|\boldsymbol{u} - \boldsymbol{v}_h\|_{\mathrm{el}}. \tag{II.10}$$

The discretization error is bounded by the approximation error.

When λ is very large, which corresponds to a quasi-incompressible material, results of poor quality can be obtained when solving problem (II.9). More specifically, it can be observed that the material deforms as if it were much stiffer. In other words, it appears to *lock*, and hence the name of *numerical locking* for describing this phenomenon. According to the estimate (II.10), there must be in that case a problem of approximation in the discrete space we consider.

Let first give a heuristic explanation to this phenomenon. To this end, let introduce the notation $\boldsymbol{U} := H_0^1(\Omega)^d$, the space

$$\underline{\boldsymbol{U}_h} := \{\boldsymbol{w}_h \in \boldsymbol{U}_h \mid \boldsymbol{\nabla}\cdot\boldsymbol{w}_h = 0\},$$

and the norm $\|\boldsymbol{v}\|_{\boldsymbol{U}} := \|\boldsymbol{\nabla}\boldsymbol{v}\|_{0,\Omega}$ on $\boldsymbol{U}$. Formally, one sees in (II.9) that in the limit case $\lambda = +\infty$, $\boldsymbol{\nabla}\cdot\boldsymbol{u}_h = 0$ (take $\boldsymbol{v}_h = \boldsymbol{u}_h$, divide the equation by λ and let λ goes to $+\infty$). Thus, the solution $\boldsymbol{u}_h$ is constrained to lie in the limit in the space $\underline{\boldsymbol{U}_h}$. Therefore, instead of being controlled by

$$\inf_{\boldsymbol{v}_h \in \boldsymbol{U}_h} \|\boldsymbol{u} - \boldsymbol{v}_h\|_{\boldsymbol{U}},$$

the approximation properties of the space in the incompressible limit are actually given by

$$\inf_{\boldsymbol{w}_h \in \underline{\boldsymbol{U}_h}} \|\boldsymbol{u} - \boldsymbol{w}_h\|_{\boldsymbol{U}}.$$

Whereas the approximation properties of $\boldsymbol{U}_h$ are usually well-known (standard finite element space), the approximation properties of $\underline{\boldsymbol{U}_h}$ are less clear and can be very poor. The extreme case is when $\underline{\boldsymbol{U}_h}$ is reduced to $\{\boldsymbol{0}\}$: the elastic solid is then completely stuck. This is the case, for example, of the (conforming) $\mathbb{P}_1^d$ finite element space on special matching simplicial meshes, see, e.g., [15, Section 11.3]. This explains the locking problem. This phenomenon would not occur in the presence of an inequality such that, there exists $C > 0$ independent of h such that

$$\inf_{\boldsymbol{w}_h \in \underline{\boldsymbol{U}_h}} \|\boldsymbol{u} - \boldsymbol{w}_h\|_{\boldsymbol{U}} \leqslant C \inf_{\boldsymbol{v}_h \in \boldsymbol{U}_h} \|\boldsymbol{u} - \boldsymbol{v}_h\|_{\boldsymbol{U}}. \tag{II.11}$$

Indeed, in such a case, the approximation properties of $\underline{\boldsymbol{U}_h}$ would be the same as those of $\boldsymbol{U}_h$. But inequality (II.11) is not true in general. All the difficulty lies in the approximation of nontrivial (or nonconstant in the case of nonhomogeneous Dirichlet boundary conditions) fields with zero-divergence (these fields are said to be solenoidal).

For further use, we first define

$$L^2_0(\Omega) := \{q \in L^2(\Omega) \mid \int_\Omega q(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 0\}, \tag{II.12}$$

and notice that $\boldsymbol{\nabla}{\cdot}\boldsymbol{U}_h \subset L^2_0(\Omega)$ since $\boldsymbol{U}_h \subset \boldsymbol{U}$. To avoid locking, as we will detail more precisely in the state of the art, different approximation techniques, ranging from primal to mixed, have been proposed in the literature. The solution of a mixed formulation of linear elasticity is characterized as the saddle-point of a Lagrangian functional involving two or three discrete unknowns (stress, displacement, pressure-like variables...). These methods may give a good remedy to locking but are often computationally more expensive than primal ones where the displacement is the sole unknown. For that reason, we focus on primal methods. To eliminate locking, it has been proposed in the engineering literature to slightly modify the energy of the problem, using a reduced integration technique on the divergence operator. We thus consider the following modified energy

$$J_h(\boldsymbol{v}_h) := \mu \int_\Omega |\underline{\boldsymbol{\varepsilon}}(\boldsymbol{v}_h)|^2 \mathrm{d}\boldsymbol{x} + \frac{\lambda}{2} \int_\Omega (\Pi_h(\boldsymbol{\nabla}{\cdot}\boldsymbol{v}_h))^2 \mathrm{d}\boldsymbol{x} - (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega},$$

where $\Pi_h : L^2(\Omega) \to P_h$ is the L^2-orthogonal projector onto the broken polynomial space P_h (to be determined). We remind that, for $p \in L^2(\Omega)$, $\Pi_h(p)$ is characterized by $\Pi_h(p) \in P_h$ and

$$(\Pi_h(p), q_h)_{0,\Omega} = (p, q_h)_{0,\Omega} \qquad \forall q_h \in P_h.$$

When restricting Π_h to $L^2_0(\Omega)$, then $\Pi_h(p) \in P_h \cap L^2_0(\Omega)$ owing to the mean conservation property of the L^2-orthogonal projector onto broken polynomial spaces. Minimizing J_h over $\boldsymbol{U}_h$ is equivalent to solving the modified problem

$$a_h(\boldsymbol{u}_h, \boldsymbol{v}_h) = (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega} \qquad \forall \boldsymbol{v}_h \in \boldsymbol{U}_h, \tag{II.13}$$

where $a_h(\boldsymbol{w}, \boldsymbol{v}) := 2\mu(\underline{\boldsymbol{\varepsilon}}(\boldsymbol{w}), \underline{\boldsymbol{\varepsilon}}(\boldsymbol{v}))_{0,\Omega} + \lambda(\Pi_h(\boldsymbol{\nabla}{\cdot}\boldsymbol{w}), \Pi_h(\boldsymbol{\nabla}{\cdot}\boldsymbol{v}))_{0,\Omega}$. Obviously, the modification of the initial problem into (II.13) has to be paid in terms of a consistency error. As a consequence, the discretization error is no longer given by (II.10) but reads

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{el}} \leqslant C_{\mathrm{D}} \left(\inf_{\boldsymbol{v}_h \in \boldsymbol{U}_h} \|\boldsymbol{u} - \boldsymbol{v}_h\|_{\mathrm{el}} + \sup_{\boldsymbol{v}_h \in \boldsymbol{U}_h \setminus \{\boldsymbol{0}\}} \frac{|a_h(\boldsymbol{u}, \boldsymbol{v}_h) - (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega}|}{\|\boldsymbol{v}_h\|_{\mathrm{el}}} \right), \tag{II.14}$$

where $\|\boldsymbol{v}\|_{\mathrm{el}} := a_h(\boldsymbol{v}, \boldsymbol{v})^{1/2}$ and $C_{\mathrm{D}} > 0$ is a constant independent of h, λ, μ, and $\boldsymbol{u}$. Let now see how the above modification of the initial problem changes the approximation properties of the space involved in the incompressible limit case. To that extent, let rewrite the modified problem (II.13) into a mixed form. Introducing $p_h = -\lambda \Pi_h(\boldsymbol{\nabla}{\cdot}\boldsymbol{u}_h) \in P_h \cap L^2_0(\Omega)$ leads to

$$\lambda(\Pi_h(\boldsymbol{\nabla}{\cdot}\boldsymbol{u}_h), \Pi_h(\boldsymbol{\nabla}{\cdot}\boldsymbol{v}_h))_{0,\Omega} = -(p_h, \Pi_h(\boldsymbol{\nabla}{\cdot}\boldsymbol{v}_h))_{0,\Omega} = -(p_h, \boldsymbol{\nabla}{\cdot}\boldsymbol{v}_h)_{0,\Omega}.$$

Thus, solving the modified problem (II.13) is equivalent to searching for $(\boldsymbol{u}_h, p_h) \in \boldsymbol{U}_h \times P_h \cap L^2_0(\Omega)$ such that, for all $(\boldsymbol{v}_h, q_h) \in \boldsymbol{U}_h \times P_h \cap L^2_0(\Omega)$,

$$\begin{aligned} 2\mu(\underline{\boldsymbol{\varepsilon}}(\boldsymbol{u}_h), \underline{\boldsymbol{\varepsilon}}(\boldsymbol{v}_h))_{0,\Omega} - (\boldsymbol{\nabla}{\cdot}\boldsymbol{v}_h, p_h)_{0,\Omega} &= (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega}, \\ (\boldsymbol{\nabla}{\cdot}\boldsymbol{u}_h, q_h)_{0,\Omega} + \lambda^{-1}(p_h, q_h)_{0,\Omega} &= 0. \end{aligned} \tag{II.15}$$

Under this mixed form, the benefits of the reduced integration technique can now clearly be seen. Indeed, let introduce the space

$$\overline{\boldsymbol{U}_h} := \{\boldsymbol{w}_h \in \boldsymbol{U}_h \mid \Pi_h(\boldsymbol{\nabla}\cdot\boldsymbol{w}_h) = 0\}.$$

One sees that when λ goes to infinity, the solution $\boldsymbol{u}_h$ to problem (II.15), and thus to the modified problem (II.13), is constrained to lie in $\overline{\boldsymbol{U}_h}$ (instead of $\underline{\boldsymbol{U}_h}$ as for problem (II.9)). The trick is now clear. Whereas $\underline{\boldsymbol{U}_h}$ was a *hidden* and not very convenient space, the space $\overline{\boldsymbol{U}_h}$ is actually linked to the choice of the space P_h. To choose P_h, a compromise has to be found between locking and accuracy. The heuristic is the following. On the one hand, a smaller P_h makes $\overline{\boldsymbol{U}_h}$ larger and thus an inequality like (II.11) easier to obtain, which avoids locking but enforces poorly the incompressibility constraint. On the other hand, a larger P_h enforces better the incompressibility constraint but leads to introduce locking since inequality (II.11) is harder to achieve with a small $\overline{\boldsymbol{U}_h}$. The question is now: how to choose P_h from a practical point of view?

To answer that question, let first introduce the following result by Nečas, which can be found, e.g., in [63, 48].

Lemma II.3 (Surjectivity of the divergence operator)**.** *The divergence operator is surjective from* $H^1_0(\Omega)^d$ *to* $L^2_0(\Omega)$*. Thus, for all* $p \in L^2_0(\Omega)$*, there exists* $\boldsymbol{u}_{\mathrm{N}} \in H^1_0(\Omega)^d$ *such that*

$$\boldsymbol{\nabla}\cdot\boldsymbol{u}_{\mathrm{N}} = p \qquad \textit{and} \qquad \|\boldsymbol{u}_{\mathrm{N}}\|_{\boldsymbol{U}} \leqslant C_{\mathrm{N}}\|p\|_{0,\Omega},$$

where $C_{\mathrm{N}} > 0$ *only depends on* Ω*.*

This result holds true for Lipschitz domains, which is the case of polygonal or polyhedral domains. Let now define the notion of Fortin operator for our problem; see, e.g., [17].

Definition II.1 (Fortin operator)**.** We call Fortin operator an interpolator $\mathcal{I}_h : \boldsymbol{U} \to \boldsymbol{U}_h$ such that

(i) $\forall\, \boldsymbol{v} \in \boldsymbol{U}, \quad \Pi_h(\boldsymbol{\nabla}\cdot\mathcal{I}_h(\boldsymbol{v})) = \Pi_h(\boldsymbol{\nabla}\cdot\boldsymbol{v})$;

(ii) there exists $C_{\mathrm{S}} > 0$, independent of h, such that

$$\forall\, \boldsymbol{v} \in \boldsymbol{U}, \quad \|\mathcal{I}_h(\boldsymbol{v})\|_{\boldsymbol{U}} \leqslant C_{\mathrm{S}}\,\|\boldsymbol{v}\|_{\boldsymbol{U}}.$$

The Fortin operator $\mathcal{I}_h$ is designed in order to satisfy optimal approximation properties (cf. Lemma III.5) under classical requirements on the mesh sequence (cf. Section III.1). Note that the regularity $H^1(\Omega)^d$ is insufficient to define a Fortin operator using classical Lagrange interpolation on conforming finite element spaces since pointwise evaluations of functions are needed. In this case, a solution is to consider the Clément interpolator, cf. [35, Section 1.6.1]. The existence of a Fortin operator is instrumental in the proof of the following result, inspired of [17, Proposition 2.5].

Lemma II.4 (Robustness with respect to locking)**.** *Assume that* P_h *is chosen such that there exists* $\mathcal{I}_h$ *Fortin operator in the sense of Definition II.1. Let* $\boldsymbol{u} \in H^1_0(\Omega)^d$ *such that, for all* $q \in L^2_0(\Omega)$*,* $(\boldsymbol{\nabla}\cdot\boldsymbol{u}, q)_{0,\Omega} = 0$*. Then, there exists* $C > 0$*, independent of* h*, such that*

$$\inf_{\boldsymbol{w}_h \in \overline{\boldsymbol{U}_h}} \|\boldsymbol{u} - \boldsymbol{w}_h\|_{\boldsymbol{U}} \leqslant C \inf_{\boldsymbol{v}_h \in \boldsymbol{U}_h} \|\boldsymbol{u} - \boldsymbol{v}_h\|_{\boldsymbol{U}}.$$

Proof. Let $\boldsymbol{v}_h \in \boldsymbol{U}_h$ and let consider $p := \Pi_h(\boldsymbol{\nabla}\cdot(\boldsymbol{u}-\boldsymbol{v}_h)) \in P_h \cap L^2_0(\Omega)$. According to Lemma II.3, there exists $\boldsymbol{u}_{\rm N} \in H^1_0(\Omega)^d$ such that $\boldsymbol{\nabla}\cdot\boldsymbol{u}_{\rm N} = p$, and $\|\boldsymbol{u}_{\rm N}\|_{\boldsymbol{U}} \leqslant C_{\rm N}\|p\|_{0,\Omega}$ for a constant $C_{\rm N} > 0$ independent of h. The existence of the Fortin operator $\mathcal{I}_h$ gives

$$\Pi_h(\boldsymbol{\nabla}\cdot\mathcal{I}_h(\boldsymbol{u}_{\rm N})) = \Pi_h(\boldsymbol{\nabla}\cdot\boldsymbol{u}_{\rm N}) \qquad \text{with} \qquad \|\mathcal{I}_h(\boldsymbol{u}_{\rm N})\|_{\boldsymbol{U}} \leqslant C_{\rm S}\,\|\boldsymbol{u}_{\rm N}\|_{\boldsymbol{U}},$$

where $C_{\rm S} > 0$ is independent of h. Owing to the fact that $p \in P_h$, $\Pi_h(\boldsymbol{\nabla}\cdot\boldsymbol{u}_{\rm N}) = p$ and thus

$$\Pi_h(\boldsymbol{\nabla}\cdot\mathcal{I}_h(\boldsymbol{u}_{\rm N})) = \Pi_h(\boldsymbol{\nabla}\cdot(\boldsymbol{u}-\boldsymbol{v}_h)),$$

which means in particular that $\boldsymbol{w}_h := \mathcal{I}_h(\boldsymbol{u}_{\rm N}) + \boldsymbol{v}_h \in \overline{\boldsymbol{U}_h}$ since $\Pi_h(\boldsymbol{\nabla}\cdot\boldsymbol{u}) = 0$ by assumption. Then, we get

$$\|\boldsymbol{u}-\boldsymbol{w}_h\|_{\boldsymbol{U}} \leqslant \|\boldsymbol{u}-\boldsymbol{v}_h\|_{\boldsymbol{U}} + \|\mathcal{I}_h(\boldsymbol{u}_{\rm N})\|_{\boldsymbol{U}} \leqslant \|\boldsymbol{u}-\boldsymbol{v}_h\|_{\boldsymbol{U}} + C_{\rm S}C_{\rm N}\|p\|_{0,\Omega}.$$

Owing to the definition of the L^2-orthogonal projector, we have

$$\|p\|_{0,\Omega} \leqslant \|\boldsymbol{\nabla}\cdot(\boldsymbol{u}-\boldsymbol{v}_h)\|_{0,\Omega} \leqslant C_{\rm B}\|\boldsymbol{u}-\boldsymbol{v}_h\|_{\boldsymbol{U}},$$

where $C_{\rm B} > 0$ is a constant independent of h. The conclusion follows with $C = 1 + C_{\rm B}C_{\rm S}C_{\rm N}$. □

When $\lambda = +\infty$, the solution $\boldsymbol{u} \in H^1_0(\Omega)^d$ of problem (II.3) is constrained to lie in the space

$$\{\boldsymbol{v} \in H^1_0(\Omega)^d \mid (\boldsymbol{\nabla}\cdot\boldsymbol{v}, q)_{0,\Omega} = 0, \forall\, q \in L^2_0(\Omega)\}.$$

This can be seen by rewriting (II.3) under an equivalent mixed form, just as we did for the modified problem (II.13), and by letting λ go to infinity. Thus, up to a choice of P_h such that a Fortin operator exists, Lemma II.4 guarantees that the approximation (II.13) of the linear elasticity problem will not lock in the quasi-incompressible limit. However, it still does not really help choosing P_h. The following remark does, cf. [17, Proposition II.8].

Remark II.2 (Link with a discrete inf-sup condition)**.** *Let P_h be given. Let introduce the bilinear form $b(\boldsymbol{v}_h, q_h) := -(\boldsymbol{\nabla}\cdot\boldsymbol{v}_h, q_h)_{0,\Omega}$ on $\boldsymbol{U}_h \times P_h \cap L^2_0(\Omega)$. The existence of a Fortin operator is equivalent to the verification of the following discrete inf-sup condition:*

$$\forall q_h \in P_h \cap L^2_0(\Omega), \qquad \beta\|q_h\|_{0,\Omega} \leqslant \sup_{\boldsymbol{v}_h \in \boldsymbol{U}_h\setminus\{\boldsymbol{0}\}} \frac{b(\boldsymbol{v}_h, q_h)}{\|\boldsymbol{v}_h\|_{\boldsymbol{U}}},$$

where $\beta > 0$ is independent of h. Moreover, we have the relations $\beta = (C_{\rm S}C_{\rm N})^{-1}$ and $C_{\rm B} = \|b\|$.

The conclusion follows: to guarantee the locking-free aspect of the discretization (II.13), it is sufficient to choose P_h such that a discrete inf-sup condition (with a constant independent of h) holds on $\boldsymbol{U}_h \times P_h \cap L^2_0(\Omega)$.

Remark II.3 (Nonconforming approximation)**.** *Let $\boldsymbol{U}_h \not\subset \boldsymbol{U}$ be a nonconforming approximation space satisfying $\boldsymbol{\nabla}_h\cdot\boldsymbol{U}_h \subset L^2_0(\Omega)$, and let P_h be given. We consider the following problem: Find $\boldsymbol{u}_h \in \boldsymbol{U}_h$ such that*

$$a_h(\boldsymbol{u}_h, \boldsymbol{v}_h) = (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega} \qquad \forall \boldsymbol{v}_h \in \boldsymbol{U}_h, \tag{II.16}$$

where $a_h(\boldsymbol{w}, \boldsymbol{v}) := 2\mu(\underline{\underline{\varepsilon_h}}(\boldsymbol{w}), \underline{\underline{\varepsilon_h}}(\boldsymbol{v}))_{0,\Omega} + \lambda(\Pi_h(\boldsymbol{\nabla}_h\cdot\boldsymbol{w}), \Pi_h(\boldsymbol{\nabla}_h\cdot\boldsymbol{v}))_{0,\Omega} + s_h(\boldsymbol{w}, \boldsymbol{v})$, using the broken versions of the different differential operators (cf. Section III.1.3). Introducing

$$\|\boldsymbol{v}\|_{\rm el} := a_h(\boldsymbol{v}, \boldsymbol{v})^{1/2},$$

the discretization error associated to (II.16) *still can be written under the form* (II.14)*, with the slight difference that the second term in the right-hand side does not only take into account a consistency error but also now a conformity one. The bilinear form s_h is a consistent stabilization term (jumps penalization) which aims to recover coercivity on a nonconforming level. We will give more details in Chapter IV. With a slight modification of Definition II.1 to take into account the broken character of the divergence operator when applied to $\boldsymbol{U}_h$, and of the norm $\|\boldsymbol{v}\|_U := \|\boldsymbol{\nabla}_h \boldsymbol{v}\|_{0,\Omega}$ on $\boldsymbol{U} + \boldsymbol{U}_h$, it is a simple matter to show that Lemma II.4 and Remark II.2 still hold when considering a nonconforming approximation space. As a consequence, an inf-sup stable pair $(\boldsymbol{U}_h \not\subset \boldsymbol{U}, P_h \cap L^2_0(\Omega))$ will give a locking-free primal approximation of linear elasticity equations.*

We now give some examples of locking-free conforming or nonconforming pairs $(\boldsymbol{U}_h, P_h)$ (on matching simplicial meshes) that can be encountered in the literature. Let $\mathcal{T}_h$ be a matching simplicial discretization of Ω, h representing the maximum diameter of the mesh elements, and let $\mathbb{P}^0_d(\mathcal{T}_h)$ and $\mathbb{P}^1_d(\mathcal{T}_h)$ respectively denote the spaces of piecewise constant and piecewise affine functions on $\mathcal{T}_h$ (cf. again Section III.1.3).

(i) Let introduce the space

$$\boldsymbol{U}_h := \{\boldsymbol{v}_h \in \mathbb{CR}(\mathcal{T}_h)^d \mid \boldsymbol{v}_h(\overline{\boldsymbol{x}}_F) = \boldsymbol{0}, \forall F \in \mathcal{F}^{\mathrm{b}}_{\mathcal{T}_h}\},$$

where $\mathbb{CR}(\mathcal{T}_h)$ is the Crouzeix–Raviart space introduced in Section II.1.2.1, and $\overline{\boldsymbol{x}}_F$ is the barycenter of the boundary face $F \in \mathcal{F}^{\mathrm{b}}_{\mathcal{T}_h}$. First note that $\boldsymbol{\nabla}_h \cdot \boldsymbol{U}_h \subset L^2_0(\Omega)$. We consider the nonconforming approximation (II.16) on $\boldsymbol{U}_h$, and we choose Π_h as the L^2-orthogonal projector onto $P_h := \mathbb{P}^0_d(\mathcal{T}_h)$ (classically denoted Π^0_h). Note that, when applied to $\boldsymbol{\nabla}_h \cdot \boldsymbol{U}_h$, Π^0_h is actually the identity operator. Members of $\overline{\boldsymbol{U}_h}$ are thus pointwise divergence-free. It is a simple matter to prove that the resulting discretization is locking-free. As a matter of fact, it is well-known that the pair $(\boldsymbol{U}_h, P_h \cap L^2_0(\Omega))$ satisfies an inf-sup condition. In other words, there exists a Fortin operator which preserves the mean value of the divergence (and actually of the whole gradient) inside each element. For more details in the case $d = 2$, see, e.g., Brenner and Sung [16] (for the pure displacement problem under its naturally coercive formulation, cf. Remark II.1), and Hansbo and Larson [50] (for a stabilized version, obtained as a particular case of a discontinuous Galerkin method). Note that an equivalent construction exists on quadrilaterals, this is the so-called Rannacher–Turek element [71].

(ii) Let $d = 2$ and let $\mathcal{T}_{h/2}$ be the matching triangular submesh of $\mathcal{T}_h$ obtained by connecting the barycenters of edges in $\mathcal{T}_h$. Let define

$$\boldsymbol{U}_h := \{\boldsymbol{v}_h \in H^1_0(\Omega)^d \mid \boldsymbol{v}_h \in \mathbb{P}^1_d\left(\mathcal{T}_{h/2}\right)^d\}.$$

We consider the conforming approximation (II.13) on $\boldsymbol{U}_h$, and we choose Π_h as the L^2-orthogonal projector onto $P_h := \mathbb{P}^0_d(\mathcal{T}_h)$, i.e. $\Pi_h = \Pi^0_h$. Note that $\boldsymbol{\nabla} \cdot \boldsymbol{U}_h \subset \mathbb{P}^0_d\left(\mathcal{T}_{h/2}\right)$. Members of $\overline{\boldsymbol{U}_h}$ are thus discretely divergence-free. The resulting discretization is locking-free. The construction of the Fortin operator assuming a regularity $H^2(\Omega)^d$ is detailed in [16] for the pure traction problem.

(iii) Let $\boldsymbol{U}_h$ be the (conforming) $\mathbb{P}^d_2$ finite element space with vanishing boundary conditions, and let Π_h be the L^2-orthogonal projector onto the (conforming) $\mathbb{P}_1$ finite element space. Note that $\boldsymbol{\nabla} \cdot \boldsymbol{U}_h \subset \mathbb{P}^1_d(\mathcal{T}_h)$. Members of $\overline{\boldsymbol{U}_h}$ are thus discretely divergence-free. The conforming approximation (II.13) on $\boldsymbol{U}_h$ is locking-free. Indeed, denoting $P_h := \mathbb{P}_1$, it is well-known that the pair $(\boldsymbol{U}_h, P_h \cap L^2_0(\Omega))$ (the so-called Taylor–Hood element) satisfies an inf-sup condition.

Let finally see, from a practical point of view, what are the different steps to derive locking-free discretization error estimates for sufficiently regular solutions. Assume that the continuous solution satisfies $\boldsymbol{u} \in \boldsymbol{U} \cap H^2(\Omega)^d$. Locking-free discretizations satisfy an estimate of the form

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{el}} \leqslant Ch\|\boldsymbol{f}\|_{0,\Omega}, \tag{II.17}$$

where $\|\cdot\|_{\mathrm{el}}$ is the discrete energy norm, and $C > 0$ is a constant, possibly depending on μ and on the mesh regularity parameters, but independent of h, λ, and $\boldsymbol{u}$. The key point here is that the multiplicative constant in the right-hand side of (II.17) does not blow up in the limit $\lambda \to +\infty$, i.e. the method converges uniformly with respect to λ. To obtain (II.17), we prove (without any assumption on the dimension d) that there holds, with $\mathcal{N}_{\mathrm{el}}(\boldsymbol{u})$ defined by (II.8),

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{el}} \leqslant C_{\mathrm{el}} h \mathcal{N}_{\mathrm{el}}(\boldsymbol{u}),$$

where $C_{\mathrm{el}} > 0$ has the same dependencies as C. Then, the conclusion follows from a regularity result like the one stated in Lemma II.2 in the case $d = 2$ and Ω convex. This result gives both the regularity of the solution $\boldsymbol{u}$ and the uniform bound on $\mathcal{N}_{\mathrm{el}}(\boldsymbol{u})$ with respect to λ. Then, $C = C_{\mathrm{el}} C_\mu$. We assume that an equivalent result may hold true in $d = 3$. In Chapter IV, we will study the convergence of our nonconforming method by following the above steps, starting from an abstract error estimate of the form (II.14).

II.1.3 State of the art and approximation choices

In this section we analyze the existing literature regarding the approximation of the linear elasticity equations. We particularly focus on the questions of coercivity and robustness with respect to locking that we have tackled in the previous section. The aim is, under the light of what already exists, to justify the choices and the orientations of the two following chapters.

As we already noticed in Section II.1.2.1, even if conforming finite element approximations of linear elasticity are naturally well-posed, they do not fit into our specifications. The first reason is that they are not suited at all to general meshes. The second one is that they lock in the quasi-incompressible limit. On matching triangular meshes, as pointed out by Falk [46], continuous finite elements suffer a deterioration in the convergence rate as $\lambda \to +\infty$ for piecewise polynomials of degree less or equal to 3. If we focus on the lowest-order space, we have seen in example (ii) of Section II.1.2.2 that robustness can be achieved up to a reduced integration of divergence terms. However, the price to pay is a remeshing of the primal mesh which merely doubles the number of unknowns! Hence, conforming methods are definitely not a good candidate for the discretization of linear elasticity.

Let thus focus on nonconforming methods. The coercivity issue of such approximations can be fixed through various ways. As far as the lowest-order nonconfoming space (the Crouzeix–Raviart space) is concerned, according to Falk [46], coercivity can be recovered by a reduced integration of rotational terms. The price to pay is here again a remeshing of the primal mesh wich doubles (after local elimination) the number of unknowns. Another way to reach coercivity is to increase the order of approximation. Indeed, nonconforming piecewise quadratic and cubic finite elements provide stable (and robust with respect to locking) discretizations, see again Falk [46]. However, owing to the applications we aim, we only deal with lowest-order approximations. Another technique to obtain coercivity is to add rigidity to the system by reducing the number of degrees of freedom. We refer to Appendix C where we present a Hybrid Finite Volume (HFV) method (on general 2D and 3D meshes) where the tangential component(s) of the displacement on mesh faces is (are) interpolated by using normal unknowns belonging to a stencil of neighboring faces. The linear interpolation is second order accurate in order to preserve the order of approximation of the scheme. Note that cell unknowns can also be

globally eliminated in that case since they only depend on a stencil of neighboring normal face unknowns. The robustness with respect to locking is granted as soon as the same holds for the HFV method without interpolation (which is the case, see Section III.3.2 for the construction of a Fortin operator on the corresponding space), since normal displacements on mesh faces are kept as degrees of freedom. The numerical tests are convincing. However, this technique presents two drawbacks. First, we did not manage to write a general proof attesting the unconditional stability of such a method. Secondly, even if the number of unknowns is reduced in comparison with the same method without interpolation (the only unknowns left are the normal displacements on mesh faces), the computational costs can be prohibitive on fine meshes because of the large stencil of neighboring faces that we have to consider for the interpolation. This increases the calculation (owing to the resolution of local problems) and assembling times, and deteriorates the matrix conditioning. The last technique to reach coercivity is inspired from discontinuous Galerkin (dG) methods. This technique uses a (consistent) stabilization of the bilinear form by least-square jumps penalization. Coercivity then results from the application of a weak Korn's inequality holding for piecewise H^1 vector fields, see Brenner [14]. In [49, 50], Hansbo and Larson design a lowest-order dG method on matching triangular meshes for quasi-incompressible linear elasticity which does not lock. By restricting the dG method to the Crouzeix–Raviart space, they derive a stabilized version of the lowest-order nonconforming method. The coercivity of this stabilized (and locking-free) Crouzeix–Raviart approximation is thus guaranteed by penalizing the jumps of discrete functions on mesh interfaces. Actually, dG methods are not optimal at all to approximate quasi-incompressible linear elasticity, in the sense that they imply a large number of degrees of freedom that are not necessary, since, at last, the Fortin operator involved is the Crouzeix–Raviart one. Note that Di Pietro and Nicaise [29] have proposed a locking-free dG method (on matching simplicial meshes) for linear elasticity with piecewise constant mechanical properties. The stabilization by jumps penalization is a good remedy but it has two drawbacks. The first one is that a notion of gradient-based affine reconstruction must be defined to give a sense to the jumps. This notion does not necessarily have sense for every nonconforming space, we think in particular to HFV methods (see Eymard, Gallouët and Herbin [39, 40]), but also to Mimetic Finite Difference (MFD) methods (see Brezzi, Lipnikov et al. [20, 18, 19]), and to Mixed Finite Volume (MFV) ones (see Droniou and Eymard [30]). These methods are closely related, as it has been investigated in [32], and have the particularity of considering constant reconstructions, with gradient operator and reconstruction only linked by a discrete Friedrichs' inequality and a limit-conformity assumption. The second drawback of the jumps penalization is that it enlarges the stencil as the jumps couple the unknowns between neighboring elements. The calculation (evaluations on quadrature nodes) and assembling times are also increased. Nevertheless, jumps penalization remains, from our point of view and after comparison, perhaps the best solution to ensure coercivity. Note that more general Hooke's laws, featuring fourth-order stiffness tensors satisfying certain symmetry and positive-definiteness properties in order for the problem to be well-posed, can be considered. Also in that case, jumps penalization guarantees the well-posedness on a discrete level.

As far as numerical locking is concerned, one classical way of circumventing the problem is the use of a mixed formulation, where the solution is characterized as the saddle-point of a Lagrangian functional involving two or three discrete unknowns (stress, displacement, pressure-like variables...). The resulting methods converge uniformly in λ, but are often computationally more expensive than primal methods where the displacement is the sole unknown. In this context, we recall, e.g., the PEERS method of Arnold, Brezzi and Douglas [4], the mixed method of Stenberg [75], and the mixed methods of Chavan, Lamichhane and Wohlmuth [21], and Lamichhane and Stephan [57]. All these methods require matching tetrahedral (triangular) or hexahedral (quadrilateral) meshes. General meshes matching regularity assumptions that are

similar to the ones we will consider, cf. Section III.1, have been considered by Beirão da Veiga [7], who introduces a mixed MFD method which does not lock in the quasi-incompressible limit. The problem of locking has also been addressed without resorting to mixed formulations, and several methods can be found in the literature. We can cite, e.g., the nonconforming methods of Falk [46], and the p-version method of Vogelius [78]. In this work we take inspiration, in particular, from the classical paper of Brenner and Sung [16], where the authors propose a locking-free method on matching triangular meshes based on the Crouzeix–Raviart element, see example (i) of Section II.1.2.2. The coercivity issue is here circumvented by considering the pure displacement problem and the naturally coercive form (II.5). Another source of inspiration is the work of Hansbo and Larson [49, 50], that we already detailed above. All these works require matching simplicial meshes.

Primal methods on general meshes have also been investigated. Beirão da Veiga, Brezzi and Marini [8] propose a virtual element (VE) discretization of linear elasticity which does not lock in the quasi-incompressible limit. In the finite volume sphere, we can also cite the work of Krell and coworkers on Discrete Duality Finite Volume (DDFV) schemes for the steady Stokes problem with variable viscosity (which arises for non-Newtonian fluids), in two and three space dimensions [55, 56]. DDFV schemes are staggered discretizations in the sense that the different discrete unknowns are located on different nodes. When considering a variable viscosity, one needs to derive a discrete Korn's inequality to ensure the coercivity of the diffusion operator. This is done in the work of Krell and coworkers by mimicking, on the discrete level, the relation (II.6) and the integration by parts formula. However, the spaces used to approximate the velocity and the pressure do not fulfill a discrete inf-sup condition on every kind of meshes, as recently investigated in [13]. In the context of the Stokes problem, the stability is recovered by penalizing the mass conservation equation, but in the context of quasi-incompressible elasticity, even if they are coercive, DDFV methods cannot ensure the existence of a Fortin operator on general meshes. Still in the finite volume framework, we can cite the work of Beirão da Veiga et al. on MFD methods on polyhedral meshes for the steady Stokes problem with variable (and possibly fourth-order tensorial) viscosity, in two and three space dimensions [9, 10]. For the velocity, nodal unknowns allow to recover a discrete Korn's inequality, while normal face unknowns enable to guarantee the existence of a Fortin operator. The coupled discretization with piecewise constant pressures is thus inf-sup stable and coercive, with few degrees of freedom involved. This method can be immediately extended to incompressible linear elasticity under its mixed form, or to a coercive and locking-free primal method for (possibly quasi-incompressible) elasticity by an element-wise condensation of pressure-like terms, that is to say by introducing a projection on the divergence operator. Finally, we can cite the work of Nordbotten [64], who proposes several vectorial Multi-Point Flux Approximation (MPFA) methods for linear elasticity (with general Hooke's laws), which only involve cell unknowns for the components of the displacement. These methods apply to general meshes in two and three space dimensions, are locally conservative, computationally cheap, and give good results for heterogeneous media and challenging grids, but rather poor results for quasi-incompressible materials. Other drawbacks are the rather complex local calculations needed in the construction of the method, as well as the lack of theoretical framework to study such approximations, whose stability properties are not shown. In addition, these methods are often nonsymmetric, which implies the use of more complicate solvers like GMRES or BiCGStab, instead of a simpler conjugate gradient method.

In this work, we aim to design a lowest-order, primal, symmetric (nonconforming) discretization of linear elasticity on general meshes, which is unconditionally coercive and robust with repect to locking. For that purpose, we take inspiration from the works of Brenner and Sung [16] and Hansbo and Larson [50]. In Chapter III, we build a nonconforming lowest-order discrete

space, which can be seen as an extension of the Crouzeix–Raviart space to general meshes, and which has the desired properties

(i) of approximation and weak conformity (in this case, the continuity of mean values at interfaces);

(ii) of existence of a Fortin operator.

These properties guarantee the robustness with repect to locking of any discretization of linear elasticity based on that space. The construction of the space is inspired from HFV and cell-centered Galerkin (ccG) methods, see Di Pietro [23] for the latter (ccG brings the useful notion of gradient-based affine reconstruction for finite volume methods). In Chapter IV, we apply this new space to the approximation of the elasticity equations, where we treat the coercivity issue by jumps penalization. We also investigate the local conservativity properties of the method.

II.2 Biot's consolidation model

II.2.1 Continuous setting

From now on, let Ω represent a linearly elastic porous medium saturated by a slightly compressible and viscous fluid, in which inertia effects in the elastic structure are negligible. This poroelasticity model is referred to as *quasistatic* and *single-phase*, in the sense that

- *quasistatic*: the acceleration term is neglected in the momentum balance as the inertia effects in the elastic structure are negligible;
- *single-phase*: the medium is saturated by a (slightly compressible) fluid.

Given a simulation time $T > 0$, the poroelasticity problem, see the pioneering works of Biot [12] and von Terzaghi [81], consists in finding a vector-valued displacement field $\boldsymbol{u} : \Omega \times (0,T] \to \mathbb{R}^d$, and a scalar-valued pore pressure $p : \Omega \times (0,T] \to \mathbb{R}$, such that

$$
\begin{aligned}
-\boldsymbol{\nabla}\cdot\underline{\boldsymbol{\sigma}}(\boldsymbol{u}) + \alpha\boldsymbol{\nabla}p &= \boldsymbol{f} && \text{in } \Omega\times(0,T], && \text{(II.18a)}\\
\partial_t(\alpha\boldsymbol{\nabla}\cdot\boldsymbol{u} + c_0\,p) - \boldsymbol{\nabla}\cdot(\kappa\boldsymbol{\nabla}p) &= h && \text{in } \Omega\times(0,T], && \text{(II.18b)}\\
\boldsymbol{u} &= \boldsymbol{0} && \text{on } \Gamma\times(0,T], && \text{(II.18c)}\\
\kappa\boldsymbol{\nabla}p\cdot\boldsymbol{n} &= 0 && \text{on } \Gamma\times(0,T], && \text{(II.18d)}\\
\langle p\rangle_\Omega &= 0 && \text{in } (0,T], && \text{(II.18e)}\\
(\alpha\boldsymbol{\nabla}\cdot\boldsymbol{u} + c_0\,p)(\cdot,0) &= \beta && \text{in } \Omega, && \text{(II.18f)}
\end{aligned}
$$

where

$$
\underline{\boldsymbol{\sigma}}(\boldsymbol{u}) := 2\mu\,\underline{\boldsymbol{\varepsilon}}(\boldsymbol{u}) + \lambda\boldsymbol{\nabla}\cdot\boldsymbol{u}\,\underline{\underline{I_d}}, \qquad \underline{\boldsymbol{\varepsilon}}(\boldsymbol{u}) := \frac{1}{2}(\boldsymbol{\nabla}\boldsymbol{u} + \boldsymbol{\nabla}\boldsymbol{u}^T), \tag{II.19}
$$

and where we have introduced the time-dependent notation $\langle\psi\rangle_\Omega := \frac{1}{|\Omega|}\int_\Omega \psi(\boldsymbol{x},\cdot)\mathrm{d}\boldsymbol{x}$.
This saddle-point-type model is valid under the following assumptions:

(i) infinitesimal strain theory;

(ii) small variations of porosity;

(iii) small relative variations of the fluid density with respect to a uniform equilibrium value.

The mechanical behavior of the material is described through Hooke's law (II.19), valid for isotropic linearly elastic materials, see Section II.1.1. More general Hooke's laws, involving fourth-order stiffness tensors satisfying certain symmetry and positive-definiteness properties in order for the problem to be well-posed, could be considered. Here, the material is also assumed to be homogeneous, and hence the Lamé parameters are constant in the whole medium. The

second Lamé parameter μ remains bounded, whereas λ may take unboundedly large values in the case of a quasi-incompressible material ($\lambda \to +\infty$). In that model, $\underline{\underline{\sigma}}(\boldsymbol{u})$ is called the effective stress tensor, while the total stress tensor is actually defined as $\underline{\underline{\sigma}}^{\mathrm{T}}(\boldsymbol{u}) := \underline{\underline{\sigma}}(\boldsymbol{u}) - \alpha p \underline{\underline{I_d}}$, and is such that $-\boldsymbol{\nabla}\cdot\underline{\underline{\sigma}}^{\mathrm{T}}(\boldsymbol{u}) = \boldsymbol{f}$.

The mechanical equilibrium of the coupled solid-fluid system is described by Equation (II.18a), where $\boldsymbol{f} : \Omega \times (0,T] \to \mathbb{R}^d$ is the body force per unit volume (for example the gravity). The coefficient $\alpha > 0$ (dimensionless) is the so-called Biot–Willis coefficient (sometimes denoted b, see [36]), which symmetrically quantifies

- the variation of stress induced by an increment of fluid pressure for a constant pore volume ($\alpha\boldsymbol{\nabla}p$ in (II.18a));
- the amount of fluid that can be forced into the medium by a mechanical variation of pore volume for a constant fluid pressure ($\alpha\boldsymbol{\nabla}\cdot\boldsymbol{u}$ in (II.18b)).

This coefficient is usually close to unity and we will take it equal to one in the following. This convention is adopted, e.g., in [82, 62, 74, 2].

The continuity Equation (II.18b) is the mass balance of fluid. The volume of fluid both depends on the pressure-dependent part $c_0\, p$ and on a part depending on the mechanical variations of pore volume (for constant fluid pressure) $\alpha\boldsymbol{\nabla}\cdot\boldsymbol{u}$. The coefficient $c_0 \geqslant 0$, which is homogeneous to the inverse of a pressure, is the so-called constrained specific storage coefficient (linked to Biot modulus $M > 0$ by $c_0 := \frac{1}{M}$). This coefficient is a measure of both

- the amount of fluid that can be forced into the medium by pressure increments if the fluid is assumed to be incompressible (this measure is directly linked to the compressibility of the structure);
- the amount of fluid that can be forced into the medium by pressure increments for a constant pore volume (this last measure is directly linked to the compressibility of the fluid, and vanishes for an incompressible fluid).

In some applications like consolidation processes (of clay for example), the fluid is considered to be incompressible, and the elastic structure to have very low sensitiveness to pressure increments for the range of pressures considered. Hence, the constrained specific storage coefficient is assumed to be very small, and is merely neglected in this model. In that case, the volume of fluid only depends on the mechanical variations of pore volume (for constant fluid pressure). This model is referred to as Biot's consolidation model. From a numerical point of view, and as we will detail further, the correct approximation of Biot's consolidation model is actually more involved than the one of the poroelasticity problem, and the extension to this latter is in fact straightforward as soon as Biot's consolidation model is correctly treated. Hence, we will focus in this work on the theoretical study of that special case.

The fluid flow in the porous medium follows Darcy's law, with velocity given by

$$\boldsymbol{v} = -\kappa\boldsymbol{\nabla}p,$$

where we have neglected the gravity effects on the fluid. The scalar-valued field $h : \Omega\times(0,T] \to \mathbb{R}$ is a source term, which satisfies $\langle h\rangle_\Omega = 0$ in $(0,T]$, and which is often taken equal to zero in consolidation models. The scalar-valued field κ is the mobility of the fluid, i.e. the ratio between the (scalar-valued) permeability of the medium and the constant dynamic viscosity of the fluid. The mobility is homogeneous to the ratio of a velocity with a force per unit volume and satisfies

$$0 < \underline{\kappa} \leqslant \kappa(\boldsymbol{x}) \leqslant \overline{\kappa} < +\infty \qquad \text{for a.e. } \boldsymbol{x} \in \Omega. \tag{II.20}$$

The mobility remains bounded away from infinitely large values but may take (strictly positive) arbitrarily small values in poorly permeable regions ($\underline{\kappa} \to 0^+$). Besides, we assume for theoretical needs (cf. Chapter V) that $\kappa \in W^{1,\infty}(\Omega)$, and that the mobility satisfies: there exists

$C_{\mathrm{m}} > 1$ such that

$$\|\kappa^{1/2}\|_{W^{1,\infty}(\Omega)} \leqslant C_{\mathrm{m}}\, \underline{\kappa}^{1/2}, \tag{II.21}$$

which is equivalent to infer an upper bound on the permeability contrast and on the variation scale of this latter.

To close the model, we prescribe boundary conditions on the displacement and pressure for $t > 0$, as well as we enforce in (II.18f) an initial condition $\beta : \Omega \to \mathbb{R}$ such that $\langle \beta \rangle_\Omega = 0$ on the quantity $(\alpha \boldsymbol{\nabla}\cdot\boldsymbol{u} + c_0\, p)$. To model the incompressible response of the solid-fluid aggregate in the beginning of the consolidation process, β is often taken equal to zero in consolidation models. For the sake of simplicity, we prescribe a homogeneous Dirichlet boundary condition (II.18c) on the displacement and a homogeneous Neumann boundary condition (II.18d) on the pressure, which models an impermeable boundary ($\boldsymbol{v}\cdot\boldsymbol{n} = 0$). The condition of zero mean value (II.18e) on the pressure ables to close the model. Other boundary conditions could be handled with slight modifications.

From now on, we focus on Biot's consolidation problem since we will study its numerical approximation in Chapter V. Thus, we assume $c_0 = 0$ in (II.18b) and (II.18f). We also assume that $\alpha = 1$, and that, for obvious physical reasons, either λ may take unboundedly large values (quasi-incompressible behavior), either $\underline{\kappa}$ may tend to vanish (local quasi-impermeable behavior). When dealing with quasi-incompressible behaviors ($\lambda \to +\infty$), we further assume that the quantity $\lambda^{1/2}\|\beta\|_{0,\Omega}$ is bounded independently of λ.

In order to write the weak formulation of Biot's consolidation problem, let first introduce the space $\overline{H^1}(\Omega) := H^1(\Omega) \cap L^2_0(\Omega)$, where $L^2_0(\Omega)$ has been defined in (II.12). In the sequel, we focus on solutions with low regularity, i.e. $L^2(0,T;H^1_0(\Omega)^d)$ for the displacement, and $L^2(0,T;\overline{H^1}(\Omega))$ for the pressure. Note that in the case $c_0 \neq 0$, we would have considered pressures belonging to $H^1(0,T;L^2_0(\Omega)) \cap L^2(0,T;\overline{H^1}(\Omega))$. For a function ψ defined a.e. on the space-time cylinder $\Omega \times (0,T]$, we consider ψ as a function of the time variable with values in a Hilbert space V spanned by functions of the space variable, in such a way that

$$\psi : (0,T] \ni t \mapsto \psi(t) \equiv \psi(\cdot,t) \in V, \qquad \text{for a.e. } t \in (0,T].$$

Let $\boldsymbol{U} := H^1_0(\Omega)^d$, $P := \overline{H^1}(\Omega)$, and let denote by $C^\infty_c((0,T))$ the space of time bump functions. We recall that, thanks to Friedrichs' and Poincaré inequalities, $\|\boldsymbol{\nabla}\boldsymbol{v}\|_{0,\Omega}$ is a norm on $\boldsymbol{U}$, as well as $\|\boldsymbol{\nabla} q\|_{0,\Omega}$ is a norm on P. For an initial datum $\beta \in L^2_0(\Omega)$, and source terms $\boldsymbol{f} \in H^1(0,T;L^2(\Omega)^d)$, $h \in L^2(0,T;L^2_0(\Omega))$, we consider the following weak formulation of problem (II.18) with $c_0 = 0$ and $\alpha = 1$: Find $\boldsymbol{u} \in L^2(0,T;\boldsymbol{U})$ and $p \in L^2(0,T;P)$ such that

$$\int_0^T \tilde{a}(\boldsymbol{u}(t),\boldsymbol{v})\,\varphi(t)\,\mathrm{d}t + \int_0^T b(\boldsymbol{v},p(t))\,\varphi(t)\,\mathrm{d}t = \int_0^T (\boldsymbol{f}(t),\boldsymbol{v})_{0,\Omega}\,\varphi(t)\,\mathrm{d}t \qquad \forall \boldsymbol{v} \in \boldsymbol{U}, \forall \varphi \in C^\infty_c((0,T)), \tag{II.22a}$$

$$\int_0^T b(\boldsymbol{u}(t),q)\,\varphi'(t)\,\mathrm{d}t + \int_0^T c(p(t),q)\,\varphi(t)\,\mathrm{d}t = \int_0^T (h(t),q)_{0,\Omega}\,\varphi(t)\,\mathrm{d}t \qquad \forall q \in P, \forall \varphi \in C^\infty_c((0,T)), \tag{II.22b}$$

$$((\boldsymbol{\nabla}\cdot\boldsymbol{u})(0),q)_{0,\Omega} = (\beta,q)_{0,\Omega} \qquad \forall q \in P, \tag{II.22c}$$

where $\tilde{a}(\boldsymbol{w},\boldsymbol{v}) := \mu(\boldsymbol{\nabla}\boldsymbol{w},\boldsymbol{\nabla}\boldsymbol{v})_{0,\Omega} + (\mu+\lambda)(\boldsymbol{\nabla}\cdot\boldsymbol{w},\boldsymbol{\nabla}\cdot\boldsymbol{v})_{0,\Omega}$, $b(\boldsymbol{v},q) := -(\boldsymbol{\nabla}\cdot\boldsymbol{v},q)_{0,\Omega}$, and $c(r,q) := (\kappa\boldsymbol{\nabla} r,\boldsymbol{\nabla} q)_{0,\Omega}$. Some remarks are in order. First of all, note the use of the naturally coercive linear elasticity bilinear form $\tilde{a}$. The use of such a convenient alternative formulation is justified by Remark II.1.

Remark II.4 (Initial datum)**.** *Owing to the surjectivity of the divergence operator stated in Lemma II.3, the $L^2_0(\Omega)$ initial datum on the divergence of the displacement can be expressed as*

$$\exists \boldsymbol{u}^{(0)} \in \boldsymbol{U} \quad \text{such that} \quad \beta = \boldsymbol{\nabla}\cdot\boldsymbol{u}^{(0)}, \quad \text{with } \|\boldsymbol{\nabla}\boldsymbol{u}^{(0)}\|_{0,\Omega} \leqslant C_{\mathrm{N}}\|\beta\|_{0,\Omega}, \tag{II.23}$$

where $C_{\mathrm{N}} > 0$ is the constant (only depending on Ω) introduced in Lemma II.3. Hence, we can define an initial displacement field $\boldsymbol{u}^{(0)} \in \boldsymbol{U}$.

Remark II.5 (Regularity)**.** *If $(\boldsymbol{u}, p) \in L^2(0,T;\boldsymbol{U}) \times L^2(0,T;P)$ satisfies Equation* (II.22b)*, then $t \mapsto ((\boldsymbol{\nabla}\cdot\boldsymbol{u})(t), q)_{0,\Omega} \in H^1((0,T))$ for all $q \in P$. Indeed, an integration by parts gives*

$$-\partial_t\left(b(\boldsymbol{u}, q)\right) + c(p, q) = (h, q)_{0,\Omega} \quad \forall q \in P, \quad \text{in } \mathcal{D}'((0,T)),$$

leading to the conclusion since, for all $q \in P$, $t \mapsto (h(t), q)_{0,\Omega} \in L^2((0,T))$, and $t \mapsto c(p(t), q) \in L^2((0,T))$. As a consequence, $t \mapsto ((\boldsymbol{\nabla}\cdot\boldsymbol{u})(t), q)_{0,\Omega} \in C^0([0,T])$ for all $q \in P$, hence giving a sense to (II.22c)*.*

The existence and uniqueness of a minimal regularity solution to problem (II.22) (with homogeneous permeability) has been studied by Ženíšek in [82]. This study handles the case of piecewise C^2 boundaries but does not handle the one of poorly permeable media, nor the one of quasi-incompressible materials. In this work, we prove the existence and uniqueness of the weak solution to problem (II.22) in the case of a polygonal or polyhedral domain, under the regularity assumptions on the data previously introduced, and independently of the (admissible) values that can possibly be taken by λ or $\underline{\kappa}$. The existence is proved by constructing a sequence of nonconforming approximations that converges to a solution of (II.22), see Chapter V. The uniqueness is a consequence of the following lemma, inspired from [82].

Lemma II.5 (A priori estimate)**.** *Let $(\boldsymbol{u}, p) \in L^2(0,T;\boldsymbol{U}) \times L^2(0,T;P)$ be a solution to problem* (II.22)*. Then, it satisfies the following a priori estimate:*

$$\int_0^T \tilde{a}(\boldsymbol{u}(t), \boldsymbol{u}(t))\,\mathrm{d}t + \frac{1}{2}c(z(T), z(T)) = \int_0^T (\boldsymbol{f}(t), \boldsymbol{u}(t))_{0,\Omega}\,\mathrm{d}t + \int_0^T \Big(\int_0^t h(s)\,\mathrm{d}s, p(t)\Big)_{0,\Omega}\,\mathrm{d}t + (\beta, z(T))_{0,\Omega}, \tag{II.24}$$

where $z(t) := \int_0^t p(s)\,\mathrm{d}s$, for all $t \in [0,T]$.

Proof. From Equation (II.22a), there holds for a.e. $t \in (0,T)$, and for all $\boldsymbol{v} \in \boldsymbol{U}$,

$$\tilde{a}(\boldsymbol{u}(t), \boldsymbol{v}) + b(\boldsymbol{v}, p(t)) = (\boldsymbol{f}(t), \boldsymbol{v})_{0,\Omega}.$$

Let set $\boldsymbol{v} = \boldsymbol{u}(t) \in \boldsymbol{U}$, and integrate on $(0,T)$:

$$\int_0^T \tilde{a}(\boldsymbol{u}(t), \boldsymbol{u}(t))\,\mathrm{d}t + \int_0^T b(\boldsymbol{u}(t), p(t))\,\mathrm{d}t = \int_0^T (\boldsymbol{f}(t), \boldsymbol{u}(t))_{0,\Omega}\,\mathrm{d}t. \tag{II.25}$$

Integrating by parts Equation (II.22b), and owing to Remark II.5, there holds for a.e. $s \in (0,T)$, and for all $q \in P$,

$$-\partial_t\left(b(\boldsymbol{u}(s), q)\right) + c(p(s), q) = (h(s), q)_{0,\Omega}. \tag{II.26}$$

For a given $t \in [0,T]$, let integrate the previous relation on $(0,t)$:

$$-b(\boldsymbol{u}(t), q) + \int_0^t c(p(s), q)\,\mathrm{d}s = \int_0^t (h(s), q)_{0,\Omega}\,\mathrm{d}s + (\beta, q)_{0,\Omega},$$

where we have used Remark II.5 and (II.22c). Let $q = p(t) \in P$, in such a way that $q = \partial_t z(t)$. Then, there holds

$$-b(\boldsymbol{u}(t), p(t)) + \frac{1}{2}\partial_t \left(c(z(t), z(t))\right) = (\int_0^t h(s)\,\mathrm{d}s, p(t))_{0,\Omega} + (\beta, p(t))_{0,\Omega}.$$

Integrating the last relation on $(0, T)$ gives

$$\int_0^T b(\boldsymbol{u}(t), p(t))\,\mathrm{d}t = \frac{1}{2}c(z(T), z(T)) - (\beta, z(T))_{0,\Omega} - \int_0^T (\int_0^t h(s)\,\mathrm{d}s, p(t))_{0,\Omega}\,\mathrm{d}t,$$

which, combined to (II.25), leads to the conclusion. □

Theorem II.1 (Uniqueness of the solution to (II.22))**.** *Independently of the (admissible) values that can possibly be taken by λ or $\underline{\kappa}$, whenever a solution to problem* (II.22) *exists, then it is unique and we denote it $(\boldsymbol{u}, p) \in L^2(0, T; \boldsymbol{U}) \times L^2(0, T; P)$.*

Proof. Owing to the linearity of the problem, let assume $\boldsymbol{f} \equiv \boldsymbol{0}$, $h \equiv 0$, and $\beta \equiv 0$ in (II.22), and let prove that $(\boldsymbol{u}, p) \equiv (\boldsymbol{0}, 0)$. Owing to the positivity of the term $\frac{1}{2}c(z(T), z(T))$, the estimate (II.24) combined with the fact that μ is a strictly positive constant, directly yields $\|\boldsymbol{u}\|_{L^2(0,T;H_0^1(\Omega)^d)} = 0$. This result, combined with an integration on $(0, T)$ of (II.26) where we have set $q = p(s) \in P$, and with the positivity of $\underline{\kappa}$ stated in (II.20), yields $\|p\|_{L^2(0,T;\overline{H^1}(\Omega))} = 0$, hence concluding the proof. □

II.2.2 Numerical issues

Like we did in Section II.1.2 for the linear elasticity model, we roughly detail in this paragraph the difficulties that may arise in the numerical approximation of a *quasistatic single-phase* poroelasticity problem.

These issues have two origins: first, the discretization of the linear elasticity model (coercivity and numerical locking issues), and then, the (possibly saddle-point) coupling between the flow and the mechanics (stability issues). We will not tackle again the coercivity and locking issues, and we refer the reader to Section II.1.2.

Concerning the time discretization, we will consider a first order implicit Euler method, which is the simplest and most widely used method in the literature (sometimes under its modified θ-form). Thus, we will not go further into details.

As far as the stability of the saddle-point mechanics-flow coupling is concerned, it is actually closely related to the elasticity locking phenomenon. For both of them, the difficulty lies in the approximation of the divergence operator. According to Section II.1.2.2, a locking-free discretization of elasticity is obtained as soon as there exists a discrete (pressure) space which satisfies an inf-sup condition when coupled with the displacement approximation space. In the linear elasticity context, locking is handled by projecting the discrete divergence operator onto that very space. In the context of a poroelastic displacement-pressure coupling, stability can thus be obtained by considering discrete pressure reconstructions belonging to that space, which is in fact equivalent to projecting the discrete divergence operator onto the pressure space in the coupling term. From a mathematical point of view, the inf-sup condition yields an estimate in the $L^\infty(0, T; L_0^2(\Omega))$ norm on the discrete pore pressure which is independent of $\underline{\kappa}^{-1}$.

It is of some importance to note that inf-sup stability is not needed in a somehow compressible poroelastic model (i.e. with $c_0 > 0$). As a matter of fact, in that case, the introduction of the additional term $\partial_t(c_0\, p)$ in the left-hand side of the fluid balance Equation (II.18b) directly yields a discrete $L^\infty(0, T; L_0^2(\Omega))$ estimate on the pore pressure which does not depend on $\underline{\kappa}^{-1}$

(nor on λ), and which does not hinge on the existence of a Fortin operator. The strong convergence of the approximate pressure reconstruction towards the continuous pressure can be proved in the same way as the strong convergence of the approximate gradient of the displacement.

When considering Biot's consolidation model (i.e. $c_0 = 0$), the only estimate that naturally holds on the discrete pressure is a $L^2(0,T;\overline{H^1}(\Omega))$ one, which derives from the diffusion term and which depends on $\underline{\kappa}^{-1}$. Hence, in the presence of poorly permeable regions or in the first time steps, the stability of the pressure approximation is not granted. This results in spurious spatial oscillations of the pressure approximation, see, e.g., Phillips and Wheeler [68], or Berdal Haga, Osnes and Langtangen [11]. The only way to avoid this spurious phenomenon is to introduce a stabilization on the pressure. This can be done either by stabilizing the flow Equation (II.22b), either by using approximation spaces for the displacement and pressure that actually satisfy an inf-sup condition, see Section II.2.3 for examples and discussion. A discrete inf-sup condition ensures an additional estimate on the pressure in the $L^\infty(0,T;L^2_0(\Omega))$ norm, depending on λ, but which does not depend on $\underline{\kappa}^{-1}$, see Chapter V. This makes sense from a physical point of view. Indeed, in a medium featuring very low permeability regions, an incompressible fluid cannot flow unless the material be compressible. Thus, as we explained in the previous section, when considering an incompressible fluid, the two limit cases $\underline{\kappa} \to 0^+$ and $\lambda \to +\infty$ cannot occur simultaneously. This means, from a discrete point of view, that a $L^2(0,T;L^2_0(\Omega))$ estimate holds on the approximate pressure, independently of the (admissible) values that can possibly be taken by λ or $\underline{\kappa}$, as soon as a discrete inf-sup condition is fulfilled. Indeed, when considering potentially poorly permeable regions, the material is assumed to be compressible and the estimate deriving from the inf-sup condition (which only depends on the bounded parameter λ) ensures the stability of the pressure, while in the case of a potentially quasi-incompressible material, the medium is assumed to be permeable and the stability of the approximate pressure is granted by the estimate deriving from the diffusion term (which only depends on the bounded parameter $\underline{\kappa}^{-1}$). Note finally that in Biot's consolidation model, the inf-sup condition seems mandatory to prove the strong convergence of the approximate pressure reconstruction towards the continuous pressure in the $L^2(0,T;L^2_0(\Omega))$ norm. In a way, the strong convergence of the pressure reconstruction cannot be guaranteed unless having an estimate which does not depend on $\underline{\kappa}^{-1}$.

The problem of spurious spatial oscillations of the pore pressure is actually more involved than a simple saddle-point coupling issue. The difficulty comes from the fact that, in very early times (or when the permeability is low), the pressure is quasi-$L^2_0(\Omega)$ as the diffusion term gives an almost vanishing contribution. However, as soon as $t > 0$, boundary conditions are imposed on the pore pressure hence giving necessarily to this latter a $\overline{H^1}(\Omega)$ dimension. When $c_0 = 0$, if no discrete inf-sup condition holds, the only control of pressure is given by the diffusion term, which is almost inexistant in early times. Spurious spatial oscillations then arise. If an inf-sup condition holds, then it yields a control of the pressure reconstruction, hence reducing the oscillations. Taking $c_0 > 0$ drives to the same stabilizing result. To approximate the pore pressure, one has to consider (at least, since it belongs to $\overline{H^1}(\Omega)$) a piecewise affine representation. If we consider a piecewise affine discretization of the displacement components (which is less costly than a quadratic one), then one has to derive an inf-sup condition for an equal-order approximation pair, which seems difficult. Several tricks allow to circumvent this problem. The first one is to consider a pair of spaces satisfying a (weaker) inf-sup condition (which could not be termed like that in the finite element framework), that is to say giving an estimate on a projection of the classical pressure reconstruction. We will consider that case in Chapter V but we will see that the results are not fully satisfactory. The second remedy is to add a stabilization term to the flow equation; this can be done in several ways, see Section V.4.2. The third remedy is to treat the Darcean term using a mixed method, which enables to give to

the pressure a $L^2_0(\Omega)$ dimension only (since the flux and the pressure are two different objects), and to discretize it accordingly (piecewise constant for example) using a stable coupling method with the flux. When coupled to a piecewise affine discretization of displacement, if this one is discontinuous (as in dG methods), then the coupling is stable, see Wheeler et al. [67]. The only problem of that method is that mixed methods often necessitate more unknowns than primal ones.

II.2.3 State of the art and approximation choices

In this section, we give an overview of the existing literature regarding the approximation of poroelasticity problems, and we use it to motivate our approximation choices.

There exists a wide range of poroelasticity models. They range from *dynamic* (with acceleration terms in the momentum balance) to *quasistatic*, from *multiphase compositional* to *single-phase*, they can model *multiporosity* and *multipermeability* systems, or *secondary consolidation* processes (in that case a term of the form $\lambda^*\partial_t(\boldsymbol{\nabla}\cdot\boldsymbol{u})\underline{\underline{I_d}}$ with $\lambda^* > 0$ is added to the total stress tensor $\underline{\underline{\sigma}}^{\mathrm{T}}(\boldsymbol{u})$). We focus here on the *quasistatic single-phase* model (II.18).

The mathematical issues of well-posedness of such a model have first been studied by Auriault and Sanchez–Palencia in [5]. In the later work of Showalter [74], an existence and uniqueness theory for strong in time solutions is developed, for source data assumed to be Hölder continuous in time. This work also addresses the case of Neumann-Neumann boundary regions, where the flux is prescribed both on the pressure through $\kappa\boldsymbol{\nabla}p\cdot\boldsymbol{n}$, and on the displacement through $\underline{\underline{\sigma}}(\boldsymbol{u})\boldsymbol{n}$. Often this problem is circumvented by prescribing $\underline{\underline{\sigma}}^{\mathrm{T}}(\boldsymbol{u})\boldsymbol{n}$ as the displacement flux, the problem being that this flux is usually unknown in practical problems. We can also cite the work of Ern and Meunier [36], who present an a priori analysis of the continuous problem for strong in time solutions, assuming their existence for data satisfying the assumptions introduced in Section II.2.1. Biot's consolidation problem (i.e. (II.18) with $c_0 = 0$) has been tackled by Ženíšek in [82]. The existence and uniqueness of a low regularity solution $(\boldsymbol{u}, p) \in L^2(0,T;H^1_0(\Omega)^d) \times L^2(0,T;\overline{H^1}(\Omega))$ (actually Ženíšek considers more general boundary conditions) is proved under the assumptions on the data introduced in Section II.2.1. Note however that this theory does not handle the cases of poorly permeable regions or quasi-incompressible materials.

As far as the numerical approximation is concerned, the a priori analysis of Euler-Galerkin approximations (i.e. implicit Euler in time, and continuous Galerkin in space) of Biot's consolidation problem has first been carried out by Murad, Loula and coworkers [60, 61, 62]. This analysis includes the semi-discrete and fully discrete cases, and the short- and long-time behaviors, for various stable and unstable combinations of the displacement and pore pressure approximation spaces. In [62], the dependences of the spatial error bounds with respect to time and to the meshsize are compared for combinations of finite elements ranging from unstable (equal-order Lagrangian approximations) to stable (Taylor–Hood and mini elements). It appears that the lowest equal-order Lagrangian approximation presents more singular dependence for small time than stable Taylor–Hood or mini approximations. It also appears that the consolidation process causes a regularization of the exact solution and a stabilization of the pore pressure approximation. Consequently, possible spurious spatial oscillations of the pressure field close to the origin decay in time, especially for unstable methods. Hence, after a certain time, both stable and unstable methods converge. In the work of Aguilar et al. [2], a stabilized finite element scheme is proposed to handle the problem of numerical locking in poorly permeable regions or/and in very short times. The method is based on the lowest equal-order Lagrangian approximation, and relies on a perturbation of the flow equation, with a stabilization parameter depending on the meshsize square. The resulting scheme is shown to be locking-free on numer-

ical examples, with a better robustness than inf-sup stable combinations of spaces, due to the ability of tuning the stabilization parameter according to the meshsize values. In that context, we can also cite the work of Wan, Durlofsky, Hughes and Aziz [83] on stabilized finite element methods. Another way to handle locking is to use a discontinuous Galerkin (dG) method. When using lowest equal-order discontinuous spaces for the displacement and pressure, the robustness with respect to locking can be obtained by penalizing the pressure jumps in the flow equation. This method has been tested with success by Daniele A. Di Pietro but not published yet. We can also cite the work of Phillips and Wheeler. In their first two papers [65, 66], they introduce a mixed/continuous Galerkin approximation of the poroelasticity model, which relies on a mixed discretization of the pressure and a continuous Galerkin discretization of the displacement. The semi-discrete and fully discrete cases are studied, with a time discretization using an implicit θ-scheme. Optimal error estimates are derived under strong regularity assumptions on the data and on the solutions. However, this scheme does not handle numerical locking in the sense that no $\underline{\kappa}^{-1}$-independent estimate holds on the discrete pressure when $c_0 = 0$. In [67, 68], the authors introduce a mixed/dG method for the the same problem, which turns out to handle locking in the limit $c_0 = 0$, at least for the lowest-order combination of the approximation spaces, as it is shown numerically therein. However, the robustness with respect to locking is not proved in [67], since the $L^2(0,T;L^2_0(\Omega))$ estimate on the pressure error that derive the authors comes from the $L^2(0,T;\overline{H^1}(\Omega))$ pressure error estimate, and thus depends on $\underline{\kappa}^{-1}$ (note that more general boundary conditions are actually considered in [67]). But actually, when considering the lowest-order combination of the approximation spaces, that is to say piecewise constant pressures (which is possible with a mixed scheme), and piecewise (discontinuous) affine displacements, the discrete stability is guaranteed since the two spaces satisfy an inf-sup condition (the Fortin operator is the Crouzeix–Raviart interpolator on simplices). To ensure the robustness of the coupling for any choices of the mixed/dG approximation spaces, one has to use a least-square penalization of the pressure jumps in the flow equation as we already mentioned. In the finite element sphere, we can cite as another contribution the work of Ern and Meunier [36], which details, for an Euler-Galerkin approximation, and under strong regularity assumptions in time and space on the solutions, the a priori and a posteriori analyses of problem (II.18). We can finally cite the works of Korsawe, Starke et al. [53, 54] on least-squares mixed finite element methods, and of Wheeler, Xue and Yotov [84] on the coupling of multi-point flux mixed finite element methods (for flow) with continuous Galerkin methods (for mechanics). In this last paper, the emphasis is put on the treatment of (possibly discontinuous) full tensor permeabilities, and on the use of irregular and rough tetrahedral or hexahedral grids. Note however that the grids must be matching owing to the conforming approximation of mechanics.

To the best of our knowledge, the existing literature on the approximation of the poroelasticity problem on general meshes is very poor. The limiting factor is obviously the need to design a (nonconforming) stable discretization of linear elasticity valid on general meshes (i.e. with polyhedral cells and possibly nonmatching interfaces). In addition, this discretization must be stable when coupled with cell-centered pressures. Owing to these difficulties, very few coupled finite volume approaches have been studied for poroelasticity.
In the industrial world, and this is the case in IFP Énergies nouvelles, mechanics-flow coupling are usually ensured by an external coupling of specialized and very rich codes:

- a code treating multiphase compositional Darcy flows (PumaFlow™ or COORES™ in IFP Énergies nouvelles);
- a code treating mechanics (Code_Aster or ABAQUS® in IFP Énergies nouvelles).

Mechanics is treated with conforming finite elements on matching finite element-type meshes (except if the model is itself discontinuous), whereas porous media flow codes use cell-centered finite volume methods on CPG (Corner Point Geometry) meshes [69]. CPG meshes are widely

used in reservoir simulation. They are based on a structured hexahedral grid, but are not compatible with classical finite element codes for several reasons:

- hexahedral cells may degenerate into nonstandard polyhedral cells to model the erosion of geological layers;
- vertices may be dedoubled and slide along the coordlines (i.e. straight lines orthogonal to the geological layers) to model faults, generating possibly severe nonconformities (with holes and overlapping);
- nonconforming local grid refinement (LGR) is used in near wellbore regions.

Hence, to realize the mechanics-flow coupling, one has to locally remesh (which can be intricate in the presence of faults) the CPG grid before computing the interpolation operations between 3D meshes. Then, the external coupling is realized sequentially, using coupling algorithms. There are three main types of sequential coupling methods:

- iteratively coupled: either the flow or the mechanics is solved first, then the other problem is solved using the intermediate solution. This sequential procedure is iterated at each time step until convergence within an acceptable tolerance. The converged solution is identical to the one obtained using a fully coupled approach. Examples of such techniques are drained and undrained splits (the mechanical problem is solved first), or fixed strain and fixed stress splits (the flow problem is solved first);
- explicitly coupled: this method is also called the noniterative sequential method as only one iteration is taken. This method is obviously less accurate. It can also be used as a preconditioner for a fully coupled resolution;
- loosely coupled: the coupling is resolved only after a certain number of flow time steps. This method can save computational cost but it is less accurate and requires reliable estimates of when to update the mechanical response.

The stability, accuracy and efficiency of such sequential coupling methods have been studied in detail by Kim, Tchelepi and Juanes [52] for poroelasticity and poroelastoplasticity with single-phase flow. In [59], Mikelić and Wheeler prove the convergence of the undrained split and fixed stress split methods, by exhibiting the contraction mapping constant. This constant actually tends to one as c_0 tends to zero, which means that this proof does not apply to Biot's consolidation model. In that case, it has not been proved that iterative methods converge (we bet that the proof of convergence must rely on an inf-sup condition).

Hence, industrial mechanics-flow coupling is not an easy task, since it involves local remeshing, interpolations between 3D meshes, and external sequential coupling. That is a reason why mechanics-flow couplings are not correctly handled yet in IFP Énergies nouvelles. The alternative solution is to develop a coupled finite volume approach, that is to say introduce a discretization of mechanics directly applicable on CPG meshes. With such an approach, coupling can be realized in a fully coupled way (flow and mechanics are solved simultaneously at every time step). Note that it is also possible in that case to use a sequential coupling method, the difference being now that neither interpolation nor external coupling are needed. The fully coupled method necessitates the resolution of a larger system and the use of complex solvers, but it avoids the (possibly) iterative procedure of sequential methods. The literature regarding coupled finite volume approximations of the poroelasticity problem is rather poor. We can cite the work of Shaw and Stone [73] who design a cell-centered finite volume method using interpolations, but which does not honor discontinuities. We can also cite the ongoing work of Jan Nordbotten, who designs computationally cheap cell-centered finite volume schemes for poroelasticity, based on the multi-point approximations of linear elasticity introduced in [64], cf. Section II.1.3. The main problem regarding this kind of discretizations is their conditional stability and the lack of underlying theoretical framework.

Our aim in this work is to fill the gap, by designing an unconditionally stable (and symmetric)

lowest-order discretization method for the *quasistatic single-phase* poroelasticity problem, which applies on general meshes, and whose convergence can be proved under very low regularity assumptions on the solutions and on the data. We will focus in Chapter V on the special case of Biot's consolidation problem ($c_0 = 0$), since once we have designed a robust numerical scheme for this latter, it is an easy task to extend it to the general case. The regularity on the data and on the solutions we consider is the one we introduced in Section II.2.1. In particular, we consider solutions $(\boldsymbol{u}, p) \in L^2(0,T; H_0^1(\Omega)^d) \times L^2(0,T; \overline{H^1}(\Omega))$, whose uniqueness for problem (II.22) has been proved in Theorem II.1. Their existence will be proved in a constructive way in Chapter V. These approximation choices fit into the industrial constraints we have, especially concerning the regularity of the solutions, which may be very low in practical problems. We recall that the use of lowest-order methods is justified both by the inherent uncertainty associated to physical data, and by the need to keep computational costs within affordable bounds.

To gain generality, we consider the generic framework of Gradient schemes, that has been introduced by Eymard et al. in [45, 41, 33], and which is adapted to the discretization of linear and nonlinear (possibly nonlocal) elliptic equations. This framework (coupled with a time discretization) has been used with success to approximate parabolic models such as the incompressible (immiscible) two-phase flow problem in heterogeneous porous media [44], or the Stefan problem [37]. A Gradient discretization is the data of a set of degrees of freedom and of a gradient and reconstruction operators. In order to prove the convergence of such approximations, sequences of Gradient discretizations must satisfy the following assumptions:

- coercivity: expressed as a uniform (with respect to the mesh parameter) Friedrichs' or Poincaré inequality between the gradient operator and the reconstruction;
- optimal approximation properties (also called consistency);
- limit-conformity: an integration by parts formula holds in the limit (when the mesh parameter tends to zero) between the gradient operator and the reconstruction;
- compactness: this property is only needed for nonlinear problems and ensures the control of translations.

Gradient schemes encompass a large number of well-known methods, including Galerkin methods (and in particular conforming finite elements), the Crouzeix–Raviart method, some MPFA and DDFV schemes, the HFV/MFD/MFV class of methods (cf. [32]), and the Vertex Approximate Gradient (VAG) scheme introduced by Eymard, Guichard, Herbin and Masson [42, 41, 43, 44]. As we will detail in Chapter V, the generalized Crouzeix–Raviart space introduced in Chapter III also enters the framework of Gradient discretizations, as it only differs from HFV methods through the reconstruction considered (and obviously the subgrid stabilization parameter).

In this work, we design an unconditionally stable family of Euler-Gradient approximations (i.e. implicit Euler in time and Gradient scheme in space) on possibly fairly general meshes (depending on the method used) for the saddle-point model of Biot's consolidation ($c_0 = 0$). We consider separate Gradient discretizations for displacement and pressure, whose sequences are assumed to be coupled through a uniform (with respect to the mesh parameter) inf-sup condition. This further assumption obviously reduces the field of admissible candidate methods. For data satisfying the regularity introduced in Section II.2.1, we prove the convergence of this family of Euler-Gradient approximations to the unique low regularity solution $(\boldsymbol{u}, p) \in L^2(0,T; \boldsymbol{U}) \times L^2(0,T; P)$ of problem (II.22). More precisely, we prove the strong convergence of the approximate displacement gradient and pore pressure, and the weak convergence of the approximate displacement and mobility-weighted pressure gradient. This family of approximations is also shown to be locking-free, in the sense that these convergence results are totally independent from the (admissible) values that can possibly be taken by λ or $\underline{\kappa}$. In Section V.4, numerical experiments are led on general meshes, using a discretization of mechanics based on

the generalized Crouzeix–Raviart space introduced in Chapter III, and a HFV discretization (with subgrid stabilization parameter taken equal to d) of pressure.

Chapitre III

A generalized Crouzeix–Raviart space

Sommaire

This chapter is inspired from the article [28], written with Daniele A. Di Pietro and accepted for publication in Mathematics of Computation. The aim of this chapter is to introduce a new discrete space, which can be considered as an extension of the Crouzeix–Raviart space to general meshes. We first introduce our discrete setting, including notations, discrete analysis tools, and the definition of an admissible mesh sequence. Then, we construct the space and study its conformity and approximation properties, giving a sense to its designation. We also focus on the case of a matching simplicial mesh, and finally introduce a discrete norm on that space.

III.1 Discrete setting and admissible mesh sequences

Following [24, Chapter 1] and [23, Section 1], we introduce in this section the concept of admissible mesh sequence of a bounded connected polygonal or polyhedral domain $\Omega \subset \mathbb{R}^d$ (with boundary Γ), where $d \in \{2,3\}$ stands for the space dimension. For the sake of brevity, we only give the proofs of the new results, and refer to [24, 23] for further details.

III.1.1 Shape- and contact-regularity

Let $\mathcal{H} \subset \mathbb{R}_*^+$ denote a countable set having 0 as its unique accumulation point. We consider mesh sequences $\mathcal{K}_\mathcal{H} := (\mathcal{K}_h)_{h\in\mathcal{H}}$ where, for all $h \in \mathcal{H}$, $\mathcal{K}_h$ denotes a finite collection of nonempty disjoint open polyhedra $\mathcal{K}_h = \{K\}$ such that $\overline{\Omega} = \bigcup_{K\in\mathcal{K}_h} \overline{K}$ and $h = \max_{K\in\mathcal{K}_h} h_K$ (h_K denotes here the diameter of the element $K \in \mathcal{K}_h$). We say that a hyperplanar closed connected subset F of $\overline{\Omega}$ is a mesh face if it has positive $(d-1)$-dimensional measure and if either there exist K_1, $K_2 \in \mathcal{K}_h$ such that $F \subset \partial K_1 \cap \partial K_2$ (and F is called an interface) or there exists $K \in \mathcal{K}_h$ such that $F \subset \partial K \cap \Gamma$ (and F is called a boundary face). Interfaces are collected in the set $\mathcal{F}_{\mathcal{K}_h}^{\mathrm{i}}$, boundary faces in $\mathcal{F}_{\mathcal{K}_h}^{\mathrm{b}}$ and we let $\mathcal{F}_{\mathcal{K}_h} := \mathcal{F}_{\mathcal{K}_h}^{\mathrm{i}} \cup \mathcal{F}_{\mathcal{K}_h}^{\mathrm{b}}$. The diameter of a face $F \in \mathcal{F}_{\mathcal{K}_h}$, is denoted by h_F. Moreover, we set, for all $K \in \mathcal{K}_h$, $\mathcal{F}_K := \{F \in \mathcal{F}_{\mathcal{K}_h} \mid F \subset \partial K\}$. According to the context, the notation $|\cdot|$ is used for the d- or the $(d-1)$-dimensional Lebesgue measure. In the rest of this paragraph, we discuss some fairly general regularity conditions on the mesh sequence $\mathcal{K}_\mathcal{H}$ that allow to prove basic results such as trace and inverse inequalities and polynomial approximation properties.

Definition III.1 (Shape- and contact-regularity)**.** The mesh sequence $\mathcal{K}_\mathcal{H}$ is shape- and contact-regular if for all $h \in \mathcal{H}$, $\mathcal{K}_h$ admits a matching simplicial submesh $\mathcal{T}_h$ such that

(i) *Shape-regularity.* There exists a real $\varrho_1 > 0$ independent of h such that, for all $h \in \mathcal{H}$ and all simplex $T \in \mathcal{T}_h$ of diameter h_T and inradius r_T, there holds $\varrho_1 h_T \leqslant r_T$;

(ii) *Contact-regularity.* There exists a real $\varrho_2 > 0$ independent of h such that, for all $h \in \mathcal{H}$, all $K \in \mathcal{K}_h$, and all $T \in \mathcal{T}_K := \{T \in \mathcal{T}_h \mid T \subset K\}$, there holds $\varrho_2 h_K \leqslant h_T$.

III.1.2 Admissible mesh sequences

The discrete space introduced in this work requires to identify a set of points which play a pivotal role in the construction.

Definition III.2 (Cell centers)**.** The mesh sequence $\mathcal{K}_\mathcal{H}$ admits a set of cell centers if, for all $h \in \mathcal{H}$ and all $K \in \mathcal{K}_h$, there exists a point $\boldsymbol{x}_K$ such that K is star-shaped with respect to $\boldsymbol{x}_K$ (the *cell center*) and, for all $F \in \mathcal{F}_K$, there holds,

$$d_{K,F} \geqslant \varrho_3 h_K, \tag{III.1}$$

where $d_{K,F}$ denotes the orthogonal distance between $\boldsymbol{x}_K$ and F and $\varrho_3 > 0$ is independent of h.

Let $\mathcal{K}_\mathcal{H}$ admit a set of cell centers. We define for all $h \in \mathcal{H}$ the pyramidal submesh

$$\mathcal{P}_h = \{K_F\}_{K\in\mathcal{K}_h,\, F\in\mathcal{F}_K},$$

where, for all $K \in \mathcal{K}_h$ and all $F \in \mathcal{F}_K$, K_F denotes the open pyramid of apex $\boldsymbol{x}_K$ and base F. An example of mesh $\mathcal{K}_h$ and associated pyramidal submesh $\mathcal{P}_h$ is provided in Figure III.1. Each element of $\mathcal{P}_h$ is associated to a unique element $K \in \mathcal{K}_h$ and a unique face $F \in \mathcal{F}_K$. When this link is irrelevant, the generic element of $\mathcal{P}_h$ is noted P instead of K_F. The pyramids

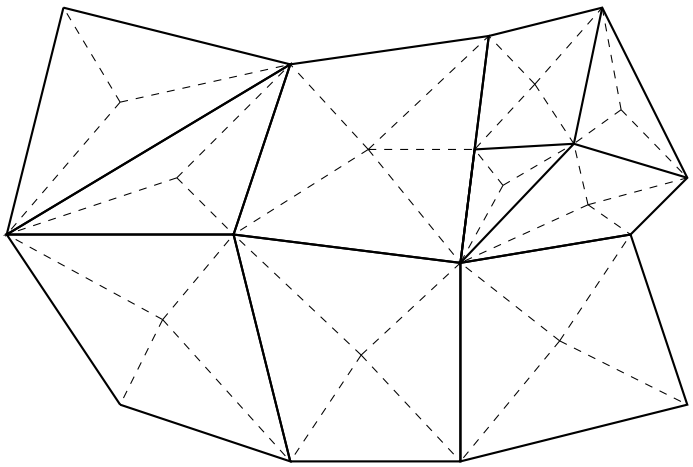


Figure III.1: Example of mesh $\mathcal{K}_h$ (solid lines) and pyramidal submesh $\mathcal{P}_h$ (dashed lines) in two dimensions.

$\{K_F\}_{K\in\mathcal{K}_h,\,F\in\mathcal{F}_K}$ are nondegenerated owing to assumption (III.1). In the two-dimensional case, $\mathcal{P}_h$ is matching and simplicial while, in higher dimension, it is in general not simplicial. Owing to the planarity of faces, there holds for all $K\in\mathcal{K}_h$ and all $F\in\mathcal{F}_K$,

$$|K_F| = \frac{|F|\,d_{K,F}}{d}. \tag{III.2}$$

The set of faces of $\mathcal{P}_h$ (including the mesh faces in $\mathcal{F}_{\mathcal{K}_h}$ as well as the lateral faces of the pyramids) is denoted by $\mathcal{F}_{\mathcal{P}_h}$ and we let $\mathcal{F}^{\rm i}_{\mathcal{P}_h} := \mathcal{F}_{\mathcal{P}_h}\backslash\mathcal{F}^{\rm b}_{\mathcal{K}_h}$ and $\mathcal{F}^{\rm b}_{\mathcal{P}_h} := \mathcal{F}^{\rm b}_{\mathcal{K}_h}$. Additionally, for all $P\in\mathcal{P}_h$, we introduce the set $\mathcal{F}_P := \{F\in\mathcal{F}_{\mathcal{P}_h}\mid F\subset\partial P\}$.

Lemma III.1 (Shape- and contact-regularity of the pyramidal submesh)**.** *Let $\mathcal{K}_\mathcal{H}$ admit a set of cell centers. Then, if $\mathcal{K}_\mathcal{H}$ is shape- and contact-regular, the same holds for $\mathcal{P}_\mathcal{H}$.*

Proof. Let $h\in\mathcal{H}$. By assumption, $\mathcal{K}_h$ admits a matching simplicial submesh $\mathcal{T}_h$. A matching simplicial submesh $\mathfrak{T}_h$ of the pyramidal submesh $\mathcal{P}_h$ can be constructed as follows: For all $K\in\mathcal{K}_h$ and all $F\in\mathcal{F}_K$ (i) a $(d-1)$-simplicial mesh $\mathfrak{S}_F$ of F is obtained taking the trace of $\mathcal{T}_h$ on F; (ii) a d-simplicial mesh $\mathfrak{T}_{K_F}$ of the pyramid K_F is then obtained connecting the (hyperplanar) elements in $\mathfrak{S}_F$ to the cell center. A matching simplicial submesh of $\mathcal{P}_h$ is obtained by setting

$$\mathfrak{T}_h := \bigcup_{K\in\mathcal{K}_h,\,F\in\mathcal{F}_K} \mathfrak{T}_{K_F}.$$

(i) *Shape-regularity.* We prove that there exists a real $\varrho_1'>0$ independent of h such that $\varrho_1' h_T\leqslant r_T$ for all $T\in\mathfrak{T}_h$. Let $K_F\in\mathcal{P}_h$ and $T\in\mathfrak{T}_{K_F}$ be given. Denoting by r_T the inradius of T, letting $\mathcal{A}_T := |\partial T|$ and $\sigma := \partial T\cap F$, there holds $d\,|T| = r_T\mathcal{A}_T = |\sigma|\,d_{K,F}$, hence

$$r_T = \frac{|\sigma|\,d_{K,F}}{\mathcal{A}_T}. \tag{III.3}$$

Since the $(d-1)$-dimensional measure of each face of T is bounded by h_K^{d-1} and T has $(d+1)$ faces, there holds $\mathcal{A}_T\leqslant(d+1)h_K^{d-1}$. Let now $S\in\mathcal{T}_h$ be the unique simplex such that $\partial S\cap F=\sigma$ and $S\subset K$. Denoting by r_σ the inradius of σ, and observing that $r_\sigma\geqslant r_S$ by a simple argument based on the Pythagorean theorem, it is inferred $|\sigma|\geqslant|\mathfrak{B}_{d-1}|\,r_\sigma^{d-1}\geqslant|\mathfrak{B}_{d-1}|\,r_S^{d-1}\geqslant$

$|\mathfrak{B}_{d-1}|\,(\varrho_1\varrho_2)^{d-1}h_K^{d-1}$ owing to the shape- and contact-regularity of $\mathcal{K}_h$ ($\mathfrak{B}_{d-1}$ denotes here the $(d-1)$-dimensional unit ball). Plugging these inequalities into (III.3), it is inferred

$$r_T \geqslant \frac{|\mathfrak{B}_{d-1}|\,(\varrho_1\varrho_2)^{d-1}}{d+1} d_{K,F} \geqslant \varrho_3 \frac{|\mathfrak{B}_{d-1}|\,(\varrho_1\varrho_2)^{d-1}}{d+1} h_T,$$

and the conclusion follows with $\varrho_1' = \varrho_3\,|\mathfrak{B}_{d-1}|\,(\varrho_1\varrho_2)^{d-1}/(d+1)$.
(ii) *Contact-regularity.* We prove that there exists a real $\varrho_2' > 0$ independent of h such that, for all $K_F \in \mathcal{P}_h$ and all $T \in \mathfrak{T}_{K_F}$, $\varrho_2' h_{K_F} \leqslant h_T$. To this end, we invoke (III.1) to infer, for all $K_F \in \mathcal{P}_h$ and all $T \in \mathfrak{T}_{K_F}$, $h_T \geqslant d_{K,F} \geqslant \varrho_3 h_K \geqslant \varrho_3 h_{K_F}$, where h_{K_F} denotes the diameter of K_F. The conclusion follows with $\varrho_2' = \varrho_3$. □

We close this section with the following definition.

Definition III.3 (Admissible mesh sequence)**.** The mesh sequence $\mathcal{K}_\mathcal{H}$ is *admissible* if it is shape- and contact-regular and it admits a set of cell centers. For an admissible mesh sequence, the reals ϱ_1, ϱ_2, and ϱ_3, are collectively referred to as *mesh regularity parameters*.

This definition encompasses fairly general meshes, featuring (possibly nonconvex) polygonal or polyhedral elements, and nonmatching interfaces.

III.1.3 Broken function spaces and polynomial approximation

For $\mathcal{S}_h \in \{\mathcal{K}_h, \mathcal{P}_h\}$ and an integer $k \geqslant 0$, we introduce the broken polynomial space

$$\mathbb{P}_d^k(\mathcal{S}_h) := \{v \in L^2(\Omega) \mid \forall S \in \mathcal{S}_h,\, v_{|S} \in \mathbb{P}_d^k(S)\},$$

where $\mathbb{P}_d^k$ denotes the space of polynomial functions of total degree at most k. Broken polynomial spaces are a special instance of broken Sobolev spaces: for an integer $l \geqslant 1$,

$$H^l(\mathcal{S}_h) := \{v \in L^2(\Omega) \mid \forall S \in \mathcal{S}_h,\, v_{|S} \in H^l(S)\}.$$

Let $K \in \mathcal{K}_h$, $P \in \mathcal{P}_h$, and $F \in \mathcal{F}_{\mathcal{P}_h}$. For $X \in \{\Omega, K, P\}$, we denote $|\cdot|_{l,X}$ and $\|\cdot\|_{l,X}$ the usual seminorm and norm on $H^l(X)$. For $\mathcal{S}_h \in \{\mathcal{K}_h, \mathcal{P}_h\}$, we define

$$|\cdot|_{l,\mathcal{S}_h} := \left(\sum_{S\in\mathcal{S}_h} |\cdot|_{l,S}^2\right)^{1/2}, \qquad \|\cdot\|_{l,\mathcal{S}_h} := \left(\sum_{S\in\mathcal{S}_h} \|\cdot\|_{l,S}^2\right)^{1/2},$$

respectively as the broken seminorm and norm on $H^l(\mathcal{S}_h)$. For $X \in \{\Omega, K, P, F\}$, we denote $(\cdot,\cdot)_{0,X}$ and $\|\cdot\|_{0,X}$ (sometimes also denoted $|\cdot|_{0,X}$) the usual scalar product and norm on $L^2(X)$. The notations remain unchanged when considering vector- or tensor-valued elements. For $\mathcal{S}_h \in \{\mathcal{K}_h, \mathcal{P}_h\}$, we finally define the broken gradient operator, denoted by $\boldsymbol{\nabla}_h$ and acting on functions $v \in H^1(\mathcal{S}_h)$, such that $(\boldsymbol{\nabla}_h v)_{|S} := \boldsymbol{\nabla}(v_{|S})$ for all $S \in \mathcal{S}_h$. We also define the broken divergence of a vector-valued field $\boldsymbol{v} \in H^1(\mathcal{S}_h)^d$ denoted by $\boldsymbol{\nabla}_h\cdot\boldsymbol{v}$, and the broken symmetric gradient $\underline{\underline{\varepsilon_h}}(\boldsymbol{v})$, respectively as the trace and as the symmetric part of the broken tensor-gradient $\boldsymbol{\nabla}_h\boldsymbol{v}$.

The shape- and contact-regularity of the mesh sequences $\mathcal{K}_\mathcal{H}$ and $\mathcal{P}_\mathcal{H}$ are instrumental to prove the following result, see [24, Lemmata 1.46 and 1.49].

Lemma III.2 (Trace inequalities). *Let $\mathcal{K}_\mathcal{H}$ be an admissible mesh sequence, and denote by $\mathcal{P}_\mathcal{H}$ the corresponding sequence of pyramidal submeshes. Then, there exist two reals $C_{\rm tr}$ and $C_{\rm tr,c}$ independent of h such that, for all $h \in \mathcal{H}$ with $\mathcal{S}_h \in \{\mathcal{K}_h, \mathcal{P}_h\}$,*

$$\forall v_h \in \mathbb{P}_d^k(\mathcal{P}_h),\ \forall P \in \mathcal{P}_h,\ \forall F \in \mathcal{F}_P, \qquad \|v_h\|_{0,F} \leqslant C_{\rm tr} h_F^{-1/2} \|v_h\|_{0,P}, \tag{III.4}$$

$$\forall v \in H^1(\mathcal{S}_h),\ \forall S \in \mathcal{S}_h,\ \forall F \in \mathcal{F}_S, \qquad \|v\|_{0,F} \leqslant C_{\rm tr,c}\big(h_S^{-1}\|v\|_{0,S}^2 + h_S|v|_{1,S}^2\big)^{1/2}. \tag{III.5}$$

For every interface $F \in \mathcal{F}_{\mathcal{S}_h}^{\rm i}$, $\mathcal{S}_h \in \{\mathcal{K}_h, \mathcal{P}_h\}$, we introduce an arbitrary but fixed ordering of the elements S_1 and S_2 such that $F \subset \partial S_1 \cap \partial S_2$ and let $\boldsymbol{n}_F := \boldsymbol{n}_{S_1,F} = -\boldsymbol{n}_{S_2,F}$, where $\boldsymbol{n}_{S_i,F}$, $i \in \{1,2\}$, denotes the unit normal to F pointing out of S_i. The orientation of the normal remains coherent when $F \in \mathcal{F}_{\mathcal{K}_h}^{\rm i}$ is regarded as an element of $\mathcal{F}_{\mathcal{P}_h}^{\rm i}$. For all $S \in \mathcal{S}_h$, we also introduce the symbol $\boldsymbol{n}_S$ to denote the vector-valued field such that $\boldsymbol{n}_{S|F} = \boldsymbol{n}_{S,F}$ for all $F \in \mathcal{F}_S$. On boundary faces $F \in \mathcal{F}_{\mathcal{P}_h}^{\rm b}$, $\boldsymbol{n}_F$ denotes the unit normal pointing out of Ω.
We next introduce jump and average trace operators that are widely used in the context of nonconforming finite element methods. For a face $F \in \mathcal{F}_{\mathcal{P}_h}^{\rm i}$ with $F \subset \partial P_1 \cap \partial P_2$ and a scalar-valued function v admitting a possibly two-valued trace on F we set,

$$[\![v]\!]_F := v_{|P_1} - v_{|P_2}, \qquad \{v\}_F := \frac{1}{2}\left(v_{|P_1} + v_{|P_2}\right).$$

If $F \in \mathcal{F}_{\mathcal{P}_h}^{\rm b}$ with $F = \partial P \cap \Gamma$, we conventionally set $[\![v]\!]_F = \{v\}_F := v_{|P}$. When applied to vector-valued functions, both the jump and average operators act component-wise. Whenever no confusion can arise, we omit the subscript F and simply write $[\![v]\!]$, $\{v\}$.

We close this section by considering polynomial approximation on admissible mesh sequences. It has been proved in [24, Lemma 1.40] that, for a shape- and contact-regular mesh sequence, the number of simplices from the submesh $\mathcal{T}_h$ contained in each element $K \in \mathcal{K}_h$ is bounded uniformly in h. This, together with the results of Dupont and Scott [34], yields the following.

Lemma III.3 (Optimal polynomial approximation). *Let $\mathcal{K}_\mathcal{H}$ denote a shape- and contact-regular mesh sequence. Then, for all $h \in \mathcal{H}$, all $K \in \mathcal{K}_h$, all polynomial degree $k \geqslant 0$, all $s \in \{0, \ldots, k+1\}$ and all $v \in H^s(K)$, there holds with Π_h^k denoting the L^2-orthogonal projector onto $\mathbb{P}_d^k(\mathcal{K}_h)$,*

$$|v - \Pi_h^k(v)|_{m,K} \leqslant C_{\rm app} h_K^{s-m} |v|_{s,K} \qquad \forall\, m \in \{0, \ldots, s\}, \tag{III.6}$$

where $C_{\rm app}$ is independent of both K and h.

We also note the following result, which is an immediate consequence of the trace inequality (III.5) with $\mathcal{S}_h = \mathcal{K}_h$ and of the approximation properties of the L^2-orthogonal projector.

Proposition III.1 (Approximation on mesh faces). *For an admissible mesh sequence $\mathcal{K}_\mathcal{H}$ there holds for all $h \in \mathcal{H}$, all $K \in \mathcal{K}_h$, all $F \in \mathcal{F}_K$, all polynomial degree $k \geqslant 0$, all $s \in \{0, \ldots, k+1\}$, and all $v \in H^s(K)$,*

$$\|v - \Pi_h^k(v)\|_{0,F} \leqslant C h_K^{s-1/2} |v|_{s,K},$$

where $C = C_{\rm tr,c} C_{\rm app}$ with $C_{\rm tr,c}$ defined as in (III.5) *and $C_{\rm app}$ as in* (III.6).

III.2 Construction of the space

The construction of our extended Crouzeix–Raviart space borrows ideas from both the cell-centered Galerkin (ccG) [23] and the Hybrid Finite Volume (HFV) [40] frameworks. Let $\mathcal{K}_h$ denote an (admissible) general polygonal or polyhedral mesh, matching the regularity requirements discussed in Section III.1. In the spirit of ccG methods, the discrete space is constructed in three steps:

(i) we fix the vector space $\mathbb{V}_h$ of face- and cell-centered degrees of freedom (DOFs) on $\mathcal{K}_h$;

(ii) we define a discrete gradient reconstruction operator $\mathfrak{G}_h$ acting on $\mathbb{V}_h$. The reconstructed gradient is piecewise constant on the (fictitious) pyramidal submesh $\mathcal{P}_h$, whose construction has been detailed in Section III.1, and it results from the sum of two terms: a consistent part depending on face unknowns only plus a subgrid correction involving both face and cell unknowns. We will see in the next section that the weak conformity of the space (here the continuity of mean values at interfaces) is ensured by finely tuning the latter contribution;

(iii) we define an affine reconstruction operator $\mathfrak{R}_h$ acting on $\mathbb{V}_h$ which maps every vector of DOFs onto a broken affine function on $\mathcal{P}_h$. This function is obtained by perturbing the (unique) face unknown associated to each pyramid with a linear correction based on the discrete gradient $\mathfrak{G}_h$. The discrete space is then defined as

$$\mathfrak{CR}(\mathcal{K}_h) := \mathfrak{R}_h(\mathbb{V}_h) \subset \mathbb{P}_d^1(\mathcal{P}_h).$$

The pyramidal submesh can be considered as fictitious in our construction in the sense that all the relevant geometric information can be computed on the primal mesh, which is therefore the only one that needs to be described and manipulated by the end-user. Note that similar ideas are used in Appendix A to construct a $\boldsymbol{H}(\mathrm{div};\Omega)$-conforming discrete space on general meshes which can be viewed as an extension of the standard lowest-order Raviart–Thomas space.

Let now enter into the details of the construction. As for HFV methods, the vector space of DOFs contains cell and face unknowns and is defined by

$$\mathbb{V}_h := \left\{ \mathrm{v}_h = \left((v_K \in \mathbb{R})_{K\in\mathcal{K}_h}, (v_F \in \mathbb{R})_{F\in\mathcal{F}_{\mathcal{K}_h}} \right) \in \mathbb{R}^{\mathcal{K}_h} \times \mathbb{R}^{\mathcal{F}_{\mathcal{K}_h}} \right\}. \tag{III.7}$$

The gradient operator generalizes the one of [40], and is composed of a consistent contribution piecewise constant on the primal mesh $\mathcal{K}_h$, plus a subgrid correction piecewise constant on the pyramidal submesh $\mathcal{P}_h$. More precisely, $\mathfrak{G}_h : \mathbb{V}_h \to \mathbb{P}_d^0(\mathcal{P}_h)^d$ realizes the mapping $\mathrm{v}_h \mapsto \mathfrak{G}_h(\mathrm{v}_h)$ with

$$\mathfrak{G}_h(\mathrm{v}_h)_{|K_F} = \boldsymbol{G}_{K_F}(\mathrm{v}_h) := \boldsymbol{G}_K(\mathrm{v}_h) + \boldsymbol{R}_{K_F}(\mathrm{v}_h), \qquad \forall K \in \mathcal{K}_h,\, F \in \mathcal{F}_K, \tag{III.8}$$

where, letting $\overline{\boldsymbol{x}}_F := \langle \boldsymbol{x} \rangle_F$ (for a function φ integrable on F, we define $\langle \varphi \rangle_F := \int_F \varphi / |F|$),

$$\boldsymbol{G}_K(\mathrm{v}_h) = \frac{1}{|K|} \sum_{F\in\mathcal{F}_K} |F|\, v_F \boldsymbol{n}_{K,F}, \quad \boldsymbol{R}_{K_F}(\mathrm{v}_h) = \frac{\eta}{d_{K,F}} \left(v_F - v_K - \boldsymbol{G}_K(\mathrm{v}_h)\cdot(\overline{\boldsymbol{x}}_F - \boldsymbol{x}_K)\right) \boldsymbol{n}_{K,F}, \tag{III.9}$$

and $\eta > 1$ is a user-dependent parameter. With a slight abuse in notation, the symbols $\boldsymbol{G}_{K_F}(\mathrm{v}_h)$, $\boldsymbol{G}_K(\mathrm{v}_h)$, and $\boldsymbol{R}_{K_F}(\mathrm{v}_h)$ will also be used to denote the corresponding constant fields on K_F, K, and K_F, respectively. The reconstruction operator $\mathfrak{R}_h : \mathbb{V}_h \to \mathbb{P}_d^1(\mathcal{P}_h)$ realizes the mapping $\mathrm{v}_h \mapsto \mathfrak{R}_h(\mathrm{v}_h)$ with

$$\mathfrak{R}_h(\mathrm{v}_h)_{|K_F}(\boldsymbol{x}) = v_F + \mathfrak{G}_h(\mathrm{v}_h)_{|K_F}\cdot(\boldsymbol{x} - \overline{\boldsymbol{x}}_F), \qquad \forall K_F \in \mathcal{P}_h,\, \forall \boldsymbol{x} \in K_F. \tag{III.10}$$

By construction, there holds $\boldsymbol{\nabla}_h \mathfrak{R}_h = \mathfrak{G}_h$. We emphasize that, in view of Lemma III.4 below, the affine reconstruction in K_F is obtained by perturbing the face unknown v_F, unlike [23], where the cell unknown v_K is used instead. We are now ready to introduce the discrete space

$$\mathfrak{CR}(\mathcal{K}_h) := \mathfrak{R}_h(\mathbb{V}_h).$$

The space $\mathfrak{CR}(\mathcal{K}_h)$ shares the same gradient operator (except concerning the value of the stabilization parameter η, see Remark III.1) as HFV methods. However, the main difference lies

in the fact that $\mathfrak{CR}(\mathcal{K}_h)$ introduces the notion of gradient-based piecewise affine reconstruction, while in the HFV spirit the reconstruction is piecewise constant and is just related to the gradient operator through a discrete Friedrichs' inequality and a limit-conformity assumption. Thus, the space $\mathfrak{CR}(\mathcal{K}_h)^d$ is much more adapted to the discretization of linear elasticity equations when one wants to recover coercivity by jumps penalization, since a notion of gradient-based piecewise affine reconstruction is needed. As we already mentioned in Section II.1.3, we have studied another technique to obtain a discrete Korn's inequality for HFV-based approximations, see Appendix C.

III.3 Conformity and approximation properties

We investigate in this section how the space $\mathfrak{CR}(\mathcal{K}_h)$ extends the weak conformity (the continuity of mean values at mesh interfaces) and approximation (including the existence of a Fortin operator) properties of the classical Crouzeix–Raviart space.

III.3.1 Weak conformity

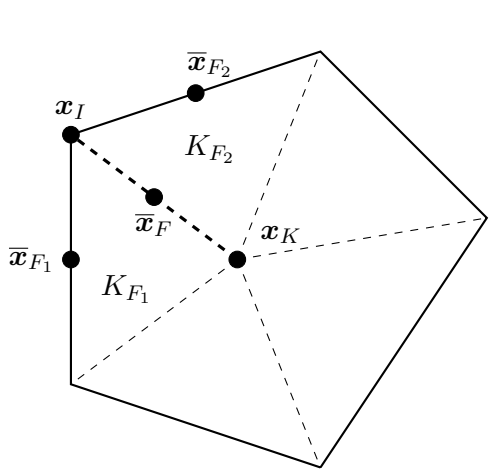
When approximating a variational problem, one has to choose a discrete approximation space in which to search the solution. Conformity measures the difference between this approximation space and the continuous one in which the variational problem is posed. When the discrete space belongs to the continuous one (like in continuous finite elements), the approximation is said to be conforming. Otherwise, the approximation is said to be nonconforming and conformity has to be ensured by other means. One solution is to impose some weak continuity constraints (pointwise for example) between elements, this is the case of nonconforming finite elements for example. This strategy gives a first order conformity error, meaning that the error decreases as h. This property is called weak conformity. In the finite volume sphere, the emphasis is put on the construction of methods that converge on general meshes. Thus, the proofs of convergence often rely on compactness arguments and usually do not pay too much importance to the study of the convergence rate. Hence, the useful conformity notion is the one of limit-conformity (see Chapter V), meaning that the conformity error tends to vanish as h tends to zero, but without any convergence rate indication. Another way to ensure conformity is to add consistency terms to the discrete bilinear form, as it is the case in dG methods. These terms make the discrete bilinear form consistent, and thus the conformity error vanish, just like in conforming approximations. Note that the term consistency error is more appropriate in this case, since the approximation space is all the same nonconforming. However, the price to pay is the addition of two other (consistent) terms in the discrete bilinear form, one to recover symmetry, and another to recover coercivity.

In this paragraph we study the weak conformity properties of $\mathfrak{CR}(\mathcal{K}_h)$. We prove that the choice $\eta = d$ in (III.9) yields the continuity of the mean values (or, equivalently, the barycentric values) of discrete functions across all the interfaces in $\mathcal{F}^{\mathrm{i}}_{\mathcal{P}_h}$ (including lateral pyramidal faces).

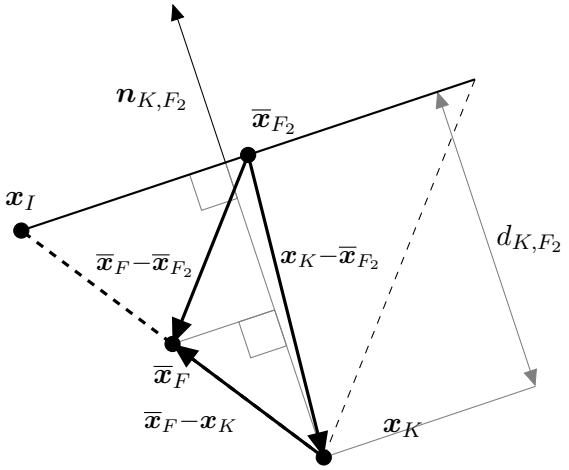
Lemma III.4 (Continuity of mean values at interfaces)**.** *Let $\mathcal{K}_h$ belong to an admissible mesh sequence and assume $\eta = d$ in* (III.9)*. Then, there holds for all $v_h \in \mathfrak{CR}(\mathcal{K}_h)$,*

$$\forall F \in \mathcal{F}^{\mathrm{i}}_{\mathcal{P}_h}, \qquad \langle [\![v_h]\!] \rangle_F = 0.$$

Proof. Let $F \in \mathcal{F}^{\mathrm{i}}_{\mathcal{P}_h}$, $\mathrm{v}_h \in \mathbb{V}_h$, and set $v_h := \mathfrak{R}_h(\mathrm{v}_h) \in \mathfrak{CR}(\mathcal{K}_h)$. We distinguish two cases. (i) If $F \in \mathcal{F}^{\mathrm{i}}_{\mathcal{K}_h}$ is a face of the primal mesh $\mathcal{K}_h$, the fact that $\langle [\![v_h]\!] \rangle_F = 0$ is an immediate consequence of choosing v_F as a starting point in (III.10). (ii) If $F \in \mathcal{F}^{\mathrm{i}}_{\mathcal{P}_h} \backslash \mathcal{F}^{\mathrm{i}}_{\mathcal{K}_h}$ is a lateral pyramidal face,



(a) Element K (solid line), elementary pyramidal submesh (dashed line), and lateral pyramidal face F shared by the pyramids K_{F_1} and K_{F_2} (thick dashed line)



(b) Zoom on the pyramid K_{F_2}

Figure III.2: Notation in two dimensions for the proof of Lemma III.4.

there exist a unique element $K \in \mathcal{K}_h$ and two faces F_1, $F_2 \in \mathcal{F}_K$ such that $F \subset \partial K_{F_1} \cap \partial K_{F_2}$ (cf. Figure III.2a). There holds for $i \in \{1,2\}$ (cf. Figure III.2b),

$$(\overline{\boldsymbol{x}}_F - \overline{\boldsymbol{x}}_{F_i})\cdot\boldsymbol{n}_{K,F_i} = (\overline{\boldsymbol{x}}_F - \boldsymbol{x}_K)\cdot\boldsymbol{n}_{K,F_i} + (\boldsymbol{x}_K - \overline{\boldsymbol{x}}_{F_i})\cdot\boldsymbol{n}_{K,F_i} = \left(\frac{d-1}{d} - 1\right) d_{K,F_i} = -\frac{d_{K,F_i}}{d},$$

where we have used the fact that $\overline{\boldsymbol{x}}_F$ is the barycenter of the $(d-1)$-simplex F to treat the term $(\overline{\boldsymbol{x}}_F-\boldsymbol{x}_K)\cdot\boldsymbol{n}_{K,F_i}$. Using the above result together with (III.9) it is inferred for $i \in \{1,2\}$,

$$\alpha_i := \boldsymbol{R}_{K_{F_i}}(\nabla_h)\cdot(\overline{\boldsymbol{x}}_F - \overline{\boldsymbol{x}}_{F_i}) = -\frac{\eta}{d}\left(v_{F_i} - v_K - \boldsymbol{G}_K(\nabla_h)\cdot(\overline{\boldsymbol{x}}_{F_i} - \boldsymbol{x}_K)\right).$$

Using the definition of the jump operator and substituting the expression (III.10) for the barycentric values $v_{h|K_{F_i}}(\overline{\boldsymbol{x}}_F)$, $i \in \{1,2\}$, we obtain

$$\begin{aligned}\langle [\![v_h]\!] \rangle_F &= v_{h|K_{F_1}}(\overline{\boldsymbol{x}}_F) - v_{h|K_{F_2}}(\overline{\boldsymbol{x}}_F) = v_{F_1} - v_{F_2} - \boldsymbol{G}_K(\nabla_h)\cdot(\overline{\boldsymbol{x}}_{F_1} - \overline{\boldsymbol{x}}_{F_2}) + \alpha_1 - \alpha_2 \\ &= \left(1 - \frac{\eta}{d}\right)\left(v_{F_1} - v_{F_2} - \boldsymbol{G}_K(\nabla_h)\cdot(\overline{\boldsymbol{x}}_{F_1} - \overline{\boldsymbol{x}}_{F_2})\right).\end{aligned}$$

The assumption $\eta = d$ finally yields $\langle [\![v_h]\!] \rangle_F = 0$, thereby concluding the proof. □

Remark III.1 (On the choice of η)**.** *The choice of η modifies the position of the continuity point along the lateral faces of the pyramids. Indeed, denoting $\boldsymbol{x}_I$ some interior point of the $(d-2)$-face of the $(d-1)$-simplex F which is shared by F_1 and F_2 (cf. Figure III.2a), let consider the point $\boldsymbol{x}_F = \beta\boldsymbol{x}_K + (1-\beta)\boldsymbol{x}_I$, for a real $\beta \in (0,1)$. Hence, $\boldsymbol{x}_F$ can be any interior point of F. Then, remarking that $(\boldsymbol{x}_F-\overline{\boldsymbol{x}}_{F_i}) = \beta(\boldsymbol{x}_K-\overline{\boldsymbol{x}}_{F_i})+(1-\beta)(\boldsymbol{x}_I-\overline{\boldsymbol{x}}_{F_i})$, and that $(\boldsymbol{x}_I-\overline{\boldsymbol{x}}_{F_i})\cdot\boldsymbol{n}_{K,F_i} = 0$ for $i \in \{1,2\}$, it is a simple matter to show, following the steps of the proof of Lemma III.4, that*

$$[\![v_h]\!]_F(\boldsymbol{x}_F) = (1 - \eta\beta)\left(v_{F_1} - v_{F_2} - \boldsymbol{G}_K(\nabla_h)\cdot(\overline{\boldsymbol{x}}_{F_1} - \overline{\boldsymbol{x}}_{F_2})\right). \tag{III.11}$$

Hence, a choice $\eta = \beta^{-1}$ ensures the continuity of functions at the interior point $\boldsymbol{x}_F$ of lateral pyramidal faces. Note that when $d = 3$, the continuity is actually ensured on a segment of

the pyramidal face which is parallel to the 1*-face to which the point* $\boldsymbol{x}_I$ *belongs. Note also that the continuity of functions at the barycenter of primal faces is unaffected by this modification of* η*, since it is a consequence of the starting point choice. Thus, any* $\eta > 1$ *ensures weak conformity properties to the space* $\mathfrak{CR}(\mathcal{K}_h)$*. In [40] for example, the choice* $\eta = d^{1/2}$ *is advocated to recover the two-point finite volume scheme on superadmissible meshes. However, the choice* $\eta = d$*, which leads to continuity at face barycenter, that is to say at the precise point where the quadrature is exact for affine functions, is (after practical comparison) the best choice in terms of discretization error.*

III.3.2 Approximation

In this section we introduce a suitable interpolator on $\mathfrak{CR}(\mathcal{K}_h)$ and study its approximation properties. Let $\mathcal{I}_h^{\mathfrak{CR}} : H^1(\Omega) \to \mathfrak{CR}(\mathcal{K}_h)$ be such that, for all $v \in H^1(\Omega)$, $\mathcal{I}_h^{\mathfrak{CR}}(v) := \mathfrak{R}_h(\mathbb{v}_h)$ with

$$\mathbb{V}_h \ni \mathbb{v}_h := \left(\left(\Pi_h^1(v)(\boldsymbol{x}_K) \right)_{K\in\mathcal{K}_h}, \left(\langle v \rangle_F \right)_{F\in\mathcal{F}_{\mathcal{K}_h}} \right), \tag{III.12}$$

where Π_h^1 denotes the L^2-orthogonal projector onto $\mathbb{P}_d^1(\mathcal{K}_h)$. When applied to vector-valued fields, $\mathcal{I}_h^{\mathfrak{CR}}$ acts component-wise.

Lemma III.5 (Approximation in $\mathfrak{CR}(\mathcal{K}_h)$)**.** *For all* $\eta > 1$ *and all* $v \in H^1(\Omega)$*, there holds with* $v_h := \mathcal{I}_h^{\mathfrak{CR}}(v) \in \mathfrak{CR}(\mathcal{K}_h)$*,*

$$\Pi_h^0(\boldsymbol{\nabla}_h v_h) = \Pi_h^0(\boldsymbol{\nabla} v), \tag{III.13}$$

where Π_h^0 *denotes the* L^2*-orthogonal projector onto* $\mathbb{P}_d^0(\mathcal{K}_h)^d$*. Moreover, there exists a real* $C > 0$ *independent of the meshsize such that, for all* $h \in \mathcal{H}$*, all* $K \in \mathcal{K}_h$*, all* $v \in H^1(\Omega) \cap H^{l+1}(\mathcal{K}_h)$*,* $l \in \{0, 1\}$*, there holds with* $v_h := \mathcal{I}_h^{\mathfrak{CR}}(v)$*,*

$$\|v - v_h\|_{0,K} + h_K \|\boldsymbol{\nabla} v - \boldsymbol{\nabla}_h v_h\|_{0,K} \leqslant C h_K^{l+1} |v|_{l+1,K}. \tag{III.14}$$

Proof. To avoid naming generic constants, we use the notation $a \lesssim b$ for the inequality $a \leqslant Cb$ with $C > 0$ independent of the meshsize.
(i) *Equality* (III.13). For a given $v \in H^1(\Omega)$, let $\mathbb{v}_h$ be defined as in (III.12). We start by noting the following orthogonality relation (cf. [40, eq. (27)]) valid for all $\mathbb{w}_h \in \mathbb{V}_h$ and all $K \in \mathcal{K}_h$:

$$\sum_{F\in\mathcal{F}_K} |K_F| \, \boldsymbol{R}_{K_F}(\mathbb{w}_h) = \boldsymbol{0}. \tag{III.15}$$

As a consequence, for all $K \in \mathcal{K}_h$ there holds,

$$\Pi_h^0(\boldsymbol{\nabla}_h v_h)_{|K} = \boldsymbol{G}_K(\mathbb{v}_h) = \frac{1}{|K|} \sum_{F\in\mathcal{F}_K} |F| \langle v \rangle_F \boldsymbol{n}_{K,F} = \frac{1}{|K|} \int_{\partial K} v \boldsymbol{n}_K = \Pi_h^0(\boldsymbol{\nabla} v)_{|K},$$

where we have used the planarity of faces and Green's formula. Relation (III.13) follows.
(ii) *Inequality* (III.14). Let $v \in H^1(\Omega) \cap H^{l+1}(\mathcal{K}_h)$ and define $\mathbb{v}_h$ as in (III.12). We first estimate $\|\boldsymbol{\nabla} v - \boldsymbol{\nabla}_h v_h\|_{0,K}$, $K \in \mathcal{K}_h$. Using (III.8), the previous point, and the triangular inequality we infer

$$\|\boldsymbol{\nabla} v - \boldsymbol{\nabla}_h v_h\|_{0,K} \leqslant \|\boldsymbol{\nabla} v - \Pi_h^0(\boldsymbol{\nabla} v)\|_{0,K} + \left(\sum_{F\in\mathcal{F}_K} |K_F| \, |\boldsymbol{R}_{K_F}(\mathbb{v}_h)|^2 \right)^{\frac{1}{2}} := \mathfrak{T}_1 + \mathfrak{T}_2.$$

Using the approximation properties of the L^2-orthogonal projector it is readily inferred $\mathfrak{T}_1 \lesssim h_K^l|v|_{l+1,K}$. To estimate the second term, we preliminarily observe that there holds for all $F \in \mathcal{F}_K$ with $w_h := \Pi_h^1(v)$,

$$\boldsymbol{R}_{K_F}(\mathbb{v}_h) = \frac{\eta}{d_{K,F}}\left(\langle v\rangle_F - w_h(\boldsymbol{x}_K) - \boldsymbol{G}_K(\mathbb{v}_h)\cdot(\overline{\boldsymbol{x}}_F - \boldsymbol{x}_K)\right)\boldsymbol{n}_{K,F} = \frac{\eta}{d_{K,F}}\left(\alpha_{K,F} + \beta_{K,F}\right)\boldsymbol{n}_{K,F}, \tag{III.16}$$

where $\alpha_{K,F} := \langle v\rangle_F - \langle w_{h|K}\rangle_F$, $\beta_{K,F} := \big(\boldsymbol{\nabla} w_{h|K} - \boldsymbol{G}_K(\mathbb{v}_h)\big)\cdot(\overline{\boldsymbol{x}}_F - \boldsymbol{x}_K)$, and, since $w_{h|K}$ is affine in K, $w_h(\boldsymbol{x}_K) = \langle w_{h|K}\rangle_F - \boldsymbol{\nabla} w_{h|K}\cdot(\overline{\boldsymbol{x}}_F - \boldsymbol{x}_K)$. There follows from equation (III.16)

$$\mathfrak{T}_2^2 \lesssim \sum_{F\in\mathcal{F}_K}\frac{|K_F|}{d_{K,F}^2}|\alpha_{K,F}|^2 + \sum_{F\in\mathcal{F}_K}\frac{|K_F|}{d_{K,F}^2}|\beta_{K,F}|^2 := \mathfrak{T}_{2,1} + \mathfrak{T}_{2,2}.$$

Using (III.2), the Cauchy–Schwarz inequality, the mesh regularity assumption (III.1), the fact that $\mathrm{card}(\mathcal{F}_K)$ is bounded uniformly in h (cf. [24, Lemma 1.41]), and Proposition III.1 it is inferred,

$$\mathfrak{T}_{2,1} = \frac{1}{d}\sum_{F\in\mathcal{F}_K}\frac{1}{d_{K,F}\,|F|}\left(\int_F v - w_h\right)^2 \leqslant \frac{1}{d\varrho_3}\sum_{F\in\mathcal{F}_K}\frac{1}{h_K}\|v - w_h\|_{0,F}^2 \lesssim h_K^{2l}|v|_{l+1,K}^2.$$

On the other hand, since $|\overline{\boldsymbol{x}}_F - \boldsymbol{x}_K| \leqslant h_K$ and both $\boldsymbol{\nabla} w_{h|K}$ and $\boldsymbol{G}_K(\mathbb{v}_h)$ are constant on K, there holds

$$\mathfrak{T}_{2,2} \leqslant \sum_{F\in\mathcal{F}_K}|K_F|\,\frac{h_K^2}{d_{K,F}^2}|\boldsymbol{\nabla} w_{h|K} - \boldsymbol{G}_K(\mathbb{v}_h)|^2 \leqslant \frac{1}{\varrho_3^2}\|\boldsymbol{\nabla} w_{h|K} - \Pi_h^0(\boldsymbol{\nabla} v)\|_{0,K}^2 \lesssim h_K^{2l}|v|_{l+1,K}^2,$$

where we have used the mesh regularity assumption (III.1) together with (III.13), and concluded using the approximation properties of the L^2-orthogonal projector. Gathering up the bounds on $\mathfrak{T}_1$ and $\mathfrak{T}_2$ it is inferred

$$\|\boldsymbol{\nabla} v - \boldsymbol{\nabla}_h v_h\|_{0,K} \lesssim h_K^l|v|_{l+1,K}. \tag{III.17}$$

To complete the proof of inequality (III.14) it only remains to estimate $\|v - v_h\|_{0,K}$. To this end, letting again $w_h := \Pi_h^1(v)$, we apply the triangular inequality to infer

$$\|v - v_h\|_{0,K} \leqslant \|v - w_h\|_{0,K} + \|w_h - v_h\|_{0,K} := \mathfrak{T}_1 + \mathfrak{T}_2.$$

The approximation properties of the L^2-orthogonal projector readily yield $\mathfrak{T}_1 \lesssim h_K^{l+1}|v|_{l+1,K}$. For the second term, we notice that for all $F \in \mathcal{F}_K$ and all $\boldsymbol{x} \in K_F$, the linearity of both $w_{h|K}$ and $v_{h|K_F}$ yields

$$w_{h|K}(\boldsymbol{x}) = \langle w_{h|K}\rangle_F + \boldsymbol{\nabla} w_{h|K}\cdot(\boldsymbol{x} - \overline{\boldsymbol{x}}_F), \quad v_{h|K_F}(\boldsymbol{x}) = \langle v\rangle_F + \boldsymbol{\nabla} v_{h|K_F}\cdot(\boldsymbol{x} - \overline{\boldsymbol{x}}_F).$$

As a consequence,

$$\|w_h - v_h\|_{0,K_F}^2 \lesssim \int_{K_F}(\langle w_{h|K} - v\rangle_F)^2 + \int_{K_F}\big[(\boldsymbol{\nabla} w_{h|K} - \boldsymbol{\nabla} v_{h|K_F})\cdot(\boldsymbol{x} - \overline{\boldsymbol{x}}_F)\big]^2 := \mathfrak{T}_{2,1} + \mathfrak{T}_{2,2}.$$

Using (III.2), the Cauchy–Schwarz inequality, and Proposition III.1 it is inferred

$$\mathfrak{T}_{2,1} = \frac{|F|\,d_{K,F}}{d}(\langle w_{h|K} - v\rangle_F)^2 \leqslant \frac{d_{K,F}}{d}\|w_{h|K} - v\|_{0,F}^2 \lesssim h_K^{2(l+1)}|v|_{l+1,K}^2.$$

Since $|\boldsymbol{x} - \overline{\boldsymbol{x}}_F| \leqslant h_K$ for all $\boldsymbol{x} \in K_F$ and both $\boldsymbol{\nabla} w_{h|K}$ and $\boldsymbol{\nabla} v_{h|K_F}$ are constant on K_F, the estimate (III.17) yields

$$\mathfrak{T}_{2,2} \leqslant h_K^2\|\boldsymbol{\nabla} w_{h|K} - \boldsymbol{\nabla} v_{h|K_F}\|_{0,K_F}^2 \lesssim h_K^{2(l+1)}|v|_{l+1,K}^2.$$

Summing over $F \in \mathcal{F}_K$, using the bounds for $\mathfrak{T}_{2,1}$ and $\mathfrak{T}_{2,2}$ together with the fact that $\mathrm{card}(\mathcal{F}_K)$ is bounded uniformly in h, it is inferred $\mathfrak{T}_2 \lesssim h_K^{l+1}|v|_{l+1,K}$, thereby yielding $\|v - v_h\|_{0,K} \lesssim h_K^{l+1}|v|_{l+1,K}$, and therefore concluding the proof. □

Let now introduce $\boldsymbol{H}(\mathrm{div};\mathcal{P}_h) := \{\boldsymbol{v} \in L^2(\Omega)^d \mid \forall P \in \mathcal{P}_h, \boldsymbol{\nabla}\cdot(\boldsymbol{v}_{|P}) \in L^2(P)\}$, and let $D_h : \boldsymbol{H}(\mathrm{div};\mathcal{P}_h) \to \mathbb{P}_d^0(\mathcal{K}_h)$ such that, for all $\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div};\mathcal{P}_h)$,

$$D_h(\boldsymbol{v}) := \Pi_h^0(\boldsymbol{\nabla}_h\cdot\boldsymbol{v}), \tag{III.18}$$

where Π_h^0 denotes the L^2-orthogonal projector onto $\mathbb{P}_d^0(\mathcal{K}_h)$. An immediate consequence of the first point in Lemma III.5 is that the discrete vector space $\mathfrak{CR}(\mathcal{K}_h)^d$ possesses the following approximation property.

Corollary III.1 (Divergence approximation)**.** *Let $\boldsymbol{v} \in H^1(\Omega)^d$ and $\boldsymbol{v}_h := \mathcal{I}_h^{\mathfrak{CR}}(\boldsymbol{v}) \in \mathfrak{CR}(\mathcal{K}_h)^d$. There holds*

$$D_h(\boldsymbol{v}_h) = D_h(\boldsymbol{v}).$$

Moreover, there exists a real $C > 0$ independent of the meshsize such that, for all $h \in \mathcal{H}$, all $K \in \mathcal{K}_h$, and all $\boldsymbol{v} \in H^1(\Omega)^d \cap \boldsymbol{H}^1(\mathrm{div};\mathcal{K}_h)$ with $\boldsymbol{H}^1(\mathrm{div};\mathcal{K}_h) := \{\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div};\Omega) \mid \boldsymbol{\nabla}_h\cdot\boldsymbol{v} \in H^1(\mathcal{K}_h)\}$ and $\boldsymbol{v}_h := \mathcal{I}_h^{\mathfrak{CR}}(\boldsymbol{v})$,

$$\|\boldsymbol{\nabla}\cdot\boldsymbol{v} - D_h(\boldsymbol{v}_h)\|_{0,K} + h_K|\boldsymbol{\nabla}\cdot\boldsymbol{v} - D_h(\boldsymbol{v}_h)|_{1,K} \leqslant Ch_K|\boldsymbol{\nabla}\cdot\boldsymbol{v}|_{1,K}.$$

According to Definition II.1, Corollary III.1 and Lemma III.5 (which gives the H^1-stability property for $l = 0$) prove that $\mathcal{I}_h^{\mathfrak{CR}}$ can play the role of a Fortin operator when considering a projector $\Pi_h := \Pi_h^0$. This means that a primal discretization of linear elasticity equations relying on the under-integrated discrete divergence operator D_h will lead to a locking-free approximation, cf. Chapters IV and V. This also means that a coupling with pressure reconstructions belonging to $\mathbb{P}_d^0(\mathcal{K}_h) \cap L_0^2(\Omega)$ will give an inf-sup stable method, cf. Chapter V for the poroelasticity problem, and Appendix B for the Stokes equations.

III.4 The matching simplicial case

In this section, we consider an (admissible) matching simplicial mesh, that we denote $\mathcal{T}_h$. We recall the notation $\mathbb{CR}(\mathcal{T}_h)$ for the classical Crouzeix–Raviart space on $\mathcal{T}_h$. We have seen in the previous section that the space $\mathfrak{CR}(\mathcal{K}_h)$ has equivalent conformity and approximation properties on general meshes than the classical Crouzeix–Raviart space. The following result establishes a link between the two spaces in the matching simplicial case.

Proposition III.2 (Link between the two spaces)**.** *For all $\eta > 1$ in* (III.9)*, there holds*

$$\mathbb{CR}(\mathcal{T}_h) \subset \mathfrak{CR}(\mathcal{T}_h).$$

Proof. Let $v_h \in \mathbb{CR}(\mathcal{T}_h)$ and set $\mathbb{v}_h := ((v_h(\boldsymbol{x}_T))_{T\in\mathcal{T}_h}, (v_h(\overline{\boldsymbol{x}}_F))_{F\in\mathcal{F}_{\mathcal{T}_h}})$. By definition there holds (cf. (III.9)) $\boldsymbol{G}_T(\mathbb{v}_h) = (\boldsymbol{\nabla}_h v_h)_{|T}$ for all $T \in \mathcal{T}_h$. Using the linearity of v_h inside each element it is inferred $\boldsymbol{R}_{T_F}(\mathbb{v}_h) = \boldsymbol{0}$, hence $\boldsymbol{G}_{T_F}(\mathbb{v}_h) = \boldsymbol{G}_T(\mathbb{v}_h) = (\boldsymbol{\nabla}_h v_h)_{|T}$ for all $F \in \mathcal{F}_T$. As a consequence, we conclude that $v_h = \mathfrak{R}_h(\mathbb{v}_h) \in \mathfrak{CR}(\mathcal{T}_h)$. □

In the matching simplicial case, the conformity properties of $\mathfrak{CR}(\mathcal{T}_h)$ are much more stronger than the ones stated in Lemma III.4.

Lemma III.6 (Conformity properties, simplicial case)**.** *For all $\eta > 1$ in* (III.9)*, there holds for all $v_h \in \mathfrak{CR}(\mathcal{T}_h)$,*

$$\forall F \in \mathcal{F}_{\mathcal{T}_h}^{\mathrm{i}}, \quad \langle [\![v_h]\!] \rangle_F = 0, \qquad \forall F \in \mathcal{F}_{\mathcal{P}_h}^{\mathrm{i}} \backslash \mathcal{F}_{\mathcal{T}_h}^{\mathrm{i}}, \quad [\![v_h]\!]_F(\boldsymbol{x}) = 0 \quad \forall \boldsymbol{x} \in F.$$

Proof. The proof of this result relies on the fact that, when considering a matching simplicial mesh, $\boldsymbol{G}_T(\mathfrak{v}_h)$ coincides by definition with the standard Crouzeix–Raviart gradient on any $T \in \mathcal{T}_h$, hence the second factor in (III.11) vanishes independently of η or $\boldsymbol{x}_F$. Hence, following the steps of the proof of Lemma III.4, the first point remains a consequence of the starting point choice, while the second is a consequence of the previous remark, using (III.11) stated in Remark III.1 (the result remains valid for $\boldsymbol{x}_F$ belonging to the boundary of F). □

It is interesting to figure out that this conformity result states the continuity of functions belonging to $\mathfrak{CR}(\mathcal{T}_h)$ on each $T \in \mathcal{T}_h$, but does not guarantee the global linearity on T. This is a consequence of the introduction of the degree of freedom v_T. If the functions of $\mathfrak{CR}(\mathcal{T}_h)$ were in addition linear on each $T \in \mathcal{T}_h$, then we would have equality between $\mathfrak{CR}(\mathcal{T}_h)$ and $\mathbb{CR}(\mathcal{T}_h)$. But this is not the case in general: to be convinced, consider a simplex T with identical face unknowns ($v_F = v$ for all $F \in \mathcal{F}_T$) and a different cell unknown ($v_T \neq v$). Then, the classical Crouzeix–Raviart reconstruction on T is a constant function of value v, while the reconstruction in $\mathfrak{CR}(\mathcal{T}_h)$ is necessarily a strictly piecewise affine (continuous) function since $v_T \neq v$. We will see through Chapter IV and Appendix B, that the classical Crouzeix–Raviart solution and the one obtained using $\mathfrak{CR}(\mathcal{T}_h)$ as a discretization space, are identical for any linear variational problem as soon as the treatment of the right-hand side does not depend on cell unknowns. In other words, on simplicial meshes, the linear system forces the subgrid correction to vanish when the right-hand side does not *see* the cell unknowns. The conformity result of Lemma III.6 is also interesting for elasticity problems since it indicates that a discrete Korn's inequality can be obtained on $\mathfrak{CR}(\mathcal{T}_h)^d$ by penalizing the jumps of functions on primal faces only.

Note finally that, as far as approximation is concerned, the proof of Lemma III.5 can be simplified exploiting the result of Proposition III.2 to infer for $v \in H^1(\Omega)$ and for all $T \in \mathcal{T}_h$,

$$\inf_{v_h \in \mathfrak{CR}(\mathcal{T}_h)} \|v - v_h\|_{1,T} \leqslant \inf_{v_h \in \mathbb{CR}(\mathcal{T}_h)} \|v - v_h\|_{1,T},$$

and conclude using the approximation properties of the standard Crouzeix–Raviart space.

III.5 Discrete H^1_{D}-norm

We recall the following notation $H^1_{\mathrm{D}}(\Omega) := \{v \in H^1(\Omega) \mid v_{|\Gamma_{\mathrm{D}}} = 0\}$, where Γ_{D} is a subset of Γ with nonzero measure accounting for Dirichlet boundary conditions in variational problems. When $\Gamma_{\mathrm{D}} = \Gamma$, then $H^1_{\mathrm{D}}(\Omega)$ reduces to $H^1_0(\Omega)$. For problems naturally set in $H^1_{\mathrm{D}}(\Omega)$, boundary conditions can be accounted for in a strong manner by introducing the following subspace of $\mathfrak{CR}(\mathcal{K}_h)$:

$$\mathfrak{CR}_{\mathrm{D}}(\mathcal{K}_h) := \mathfrak{R}_h(\mathbb{V}_{h,\mathrm{D}}), \qquad \mathbb{V}_{h,\mathrm{D}} = \{\mathfrak{v}_h \in \mathbb{V}_h \mid v_F = 0,\ \forall F \in \mathcal{F}_{\mathcal{K}_h}^{\mathrm{b,D}}\}, \tag{III.19}$$

where $\mathcal{F}_{\mathcal{K}_h}^{\mathrm{b,D}} := \{F \in \mathcal{F}_{\mathcal{K}_h}^{\mathrm{b}} \mid F \subset \Gamma_{\mathrm{D}}\}$ is a nonempty set by assumption. We also introduce the set $\mathcal{F}_{\mathcal{K}_h}^{\mathrm{b,N}} := \mathcal{F}_{\mathcal{K}_h}^{\mathrm{b}} \backslash \mathcal{F}_{\mathcal{K}_h}^{\mathrm{b,D}}$, which denotes the set of Neumann-type boundary faces. When $\Gamma_{\mathrm{D}} = \Gamma$, we prefer the following notation:

$$\mathfrak{CR}_0(\mathcal{K}_h) := \mathfrak{R}_h(\mathbb{V}_{h,0}), \qquad \mathbb{V}_{h,0} = \{\mathfrak{v}_h \in \mathbb{V}_h \mid v_F = 0,\ \forall F \in \mathcal{F}_{\mathcal{K}_h}^{\mathrm{b}}\}. \tag{III.20}$$

In the following proposition, we show that the L^2-norm of the broken gradient is a norm on $\mathfrak{CR}_{\rm D}(\mathcal{K}_h)$ (and thus on $\mathfrak{CR}_0(\mathcal{K}_h)$) by proving uniform discrete equivalence with the usual dG norm, cf. [24, Section 5.1]:

$$\|v_h\|_{\rm dG}^2 := \|\boldsymbol{\nabla}_h v_h\|_{0,\Omega}^2 + |v_h|_{\rm J,D}^2, \qquad |v_h|_{\rm J,D}^2 := \sum_{F\in\mathcal{F}_{\mathcal{P}_h}\backslash\mathcal{F}_{\mathcal{K}_h}^{\rm b,N}} \frac{1}{h_F}\|[\![v_h]\!]\|_{0,F}^2. \tag{III.21}$$

Proposition III.3 (Discrete norm)**.** *For all $\eta > 1$ in* (III.9)*, there exists a real $C > 0$ independent of the meshsize such that, for all $v_h \in \mathfrak{CR}_{\rm D}(\mathcal{K}_h)$,*

$$\|\boldsymbol{\nabla}_h v_h\|_{0,\Omega} \leqslant \|v_h\|_{\rm dG} \leqslant C\|\boldsymbol{\nabla}_h v_h\|_{0,\Omega}.$$

Proof. The notation $a \lesssim b$ stands for $a \leqslant Cb$ with $C > 0$ independent of the meshsize. Clearly, $\|\boldsymbol{\nabla}_h v_h\|_{0,\Omega} \leqslant \|v_h\|_{\rm dG}$ for all $v_h \in \mathfrak{CR}_{\rm D}(\mathcal{K}_h)$. To prove that $\|v_h\|_{\rm dG} \lesssim \|\boldsymbol{\nabla}_h v_h\|_{0,\Omega}$ for all $v_h \in \mathfrak{CR}_{\rm D}(\mathcal{K}_h)$, it suffices to show that $|v_h|_{\rm J,D} \lesssim \|\boldsymbol{\nabla}_h v_h\|_{0,\Omega}$. Let $\eta > 1$, and let $F \in \mathcal{F}_P\backslash\mathcal{F}_{\mathcal{K}_h}^{\rm b,N}$ where $P \in \mathcal{P}_h$. Owing to Remark III.1, there exists at least one interior point $\boldsymbol{x}_F \in F$ (which may depend on η) such that $[\![v_h]\!]_F(\boldsymbol{x}_F) = 0$. Owing now to the linearity of v_h inside P there holds for all $\boldsymbol{x} \in P$, $v_{h|P}(\boldsymbol{x}) = v_{h|P}(\boldsymbol{x}_F) + \boldsymbol{\nabla}(v_{h|P})\cdot(\boldsymbol{x} - \boldsymbol{x}_F)$. These two remarks together with the discrete trace inequality (III.4) yield

$$\|[\![v_h]\!]\|_{0,F} = \|[\![v_h]\!] - [\![v_h]\!](\boldsymbol{x}_F)\|_{0,F} \leqslant h_F\|[\![\boldsymbol{\nabla}_h v_h]\!]\|_{0,F} \lesssim h_F^{1/2}\sum_{P\in\mathcal{P}_F}\|\boldsymbol{\nabla}(v_{h|P})\|_{0,P}, \tag{III.22}$$

where we have set $\mathcal{P}_F := \{P \in \mathcal{P}_h \mid F \subset \partial P\}$. Using (III.21) together with (III.22) it is inferred

$$|v_h|_{\rm J,D}^2 = \sum_{F\in\mathcal{F}_{\mathcal{P}_h}\backslash\mathcal{F}_{\mathcal{K}_h}^{\rm b,N}} \frac{1}{h_F}\|[\![v_h]\!] - [\![v_h]\!](\boldsymbol{x}_F)\|_{0,F}^2 \lesssim \sum_{F\in\mathcal{F}_{\mathcal{P}_h}\backslash\mathcal{F}_{\mathcal{K}_h}^{\rm b,N}}\sum_{P\in\mathcal{P}_F}\|\boldsymbol{\nabla}(v_{h|P})\|_{0,P}^2 \lesssim \|\boldsymbol{\nabla}_h v_h\|_{0,\Omega}^2, \tag{III.23}$$

where the last bound is a consequence of the fact that the maximum number of faces of a pyramid is bounded uniformly in h since $\mathcal{P}_h$ is shape- and contact-regular, cf. Lemma III.1 and [24, Lemma 1.41]. □

We have now all the necessary tools to study the approximation of variational problems in $\mathfrak{CR}(\mathcal{K}_h)$.

Chapitre IV

A primal, coercive, and locking-free discretization of linear elasticity equations

Sommaire

This chapter is inspired from the article [28], written with Daniele A. Di Pietro and accepted for publication in Mathematics of Computation. We present a primal discretization on general meshes of the linear elasticity equations (II.1) (stemming from the model introduced in Section II.1.1), hinging on the generalized Crouzeix–Raviart space introduced in Chapter III. Coercivity is ensured through a least-square penalization of the jumps and we prove robustness with respect to the first Lamé parameter. We investigate the links of the proposed approximation with finite volume and (classical) finite element methods, and prove that all depends on the approximation of the right-hand side. Finally, we present relevant numerical examples in two space dimensions to assess the behavior of such a discretization.

IV.1 Discretization

We consider the weak formulation (II.3) of linear elasticity equations on $\boldsymbol{U} := H^1_{\mathrm{D}}(\Omega)^d$, and we further introduce the space $\boldsymbol{U}_* := \boldsymbol{U} \cap H^2(\Omega)^d$. Let $\mathcal{K}_h$ be a general polygonal or polyhedral mesh, belonging to an admissible mesh sequence in the sense of Definition III.3. We consider a primal approximation (i.e. an approximation of the displacement field only) of the problem in the space

$$\boldsymbol{U}_h := \mathfrak{CR}_{\mathrm{D}}(\mathcal{K}_h)^d,$$

with $\mathfrak{CR}_{\mathrm{D}}(\mathcal{K}_h)$ defined in (III.19). Henceforth we assume the choice $\eta = d$ in (III.9), so that the continuity of mean (or, equivalently, barycentric) values stated in Lemma III.4 holds.

We recall that the well-posedness of the continuous weak formulation (II.3) relies on Korn's inequality (II.4). Here, owing to the nonconformity of the space we consider, Korn's inequality only holds in the following weak sense (cf. [14, eq. (1.19)]).

Lemma IV.1 (Weak Korn's inequality)**.** *There exists a constant $C_{\mathrm{K}} > 0$, independent of the meshsize but possibly depending on the mesh regularity parameters, such that, for all $\boldsymbol{v} = (v_i)_{1\leqslant i\leqslant d} \in H^1(\mathcal{P}_h)^d$,*

$$\|\boldsymbol{\nabla}_h \boldsymbol{v}\|_{0,\Omega} \leqslant C_{\mathrm{K}} \left(\|\underline{\underline{\varepsilon}}_h(\boldsymbol{v})\|^2_{0,\Omega} + |\boldsymbol{v}|^2_{\mathrm{J,D}} \right)^{\frac{1}{2}}, \tag{IV.1}$$

where $|\boldsymbol{v}|^2_{\mathrm{J,D}} := \sum_{i=1}^d |v_i|^2_{\mathrm{J,D}}$ with $|v_i|^2_{\mathrm{J,D}}$ defined in (III.21)*.*

To design the discrete bilinear form for our problem, we take inspiration from [50] and consider a coercivity treatment under the form of a (consistent) least-square penalization of function jumps. More specifically, the discrete problem reads: Find $\boldsymbol{u}_h \in \boldsymbol{U}_h$ such that

$$a_h(\boldsymbol{u}_h, \boldsymbol{v}_h) = (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega} \qquad \forall \boldsymbol{v}_h \in \boldsymbol{U}_h, \tag{IV.2}$$

with symmetric discrete bilinear form a_h such that

$$a_h(\boldsymbol{w}, \boldsymbol{v}) := 2\mu(\underline{\underline{\varepsilon}}_h(\boldsymbol{w}), \underline{\underline{\varepsilon}}_h(\boldsymbol{v}))_{0,\Omega} + \lambda(D_h(\boldsymbol{w}), D_h(\boldsymbol{v}))_{0,\Omega} + 2\mu\chi \sum_{F\in\mathcal{F}_{\mathcal{P}_h}\setminus\mathcal{F}^{\mathrm{b,N}}_{\mathcal{K}_h}} h_F^{-1}([\![\boldsymbol{w}]\!], [\![\boldsymbol{v}]\!])_{0,F}, \tag{IV.3}$$

where $0 < \chi \leqslant 1$ is a user-dependent parameter. As we already explained, it is worth observing that, while the idea of penalizing jumps to recover coercivity appears natural in the present approach, this is not the case in other related frameworks for which the notion of affine reconstruction does not necessarily make sense. Note also the treatment of the divergence operator, using the discrete divergence D_h defined in Corollary III.1. The approximation properties of D_h turn out to be instrumental to ensure that λ only appears in terms of the form $\lambda|\boldsymbol{\nabla}\cdot\boldsymbol{u}|_{1,\Omega}$ (where $\boldsymbol{u}$ is the unique solution to (II.3)) in the right-hand side of the error estimate (cf., in particular, the bound for the conformity and consistency terms in the proof of Theorem IV.1).

The energy norm associated to the bilinear form a_h is

$$\|\boldsymbol{v}\|^2_{\mathrm{el}} := a_h(\boldsymbol{v}, \boldsymbol{v}) = 2\mu\|\underline{\underline{\varepsilon}}_h(\boldsymbol{v})\|^2_{0,\Omega} + \lambda\|D_h(\boldsymbol{v})\|^2_{0,\Omega} + 2\mu\chi|\boldsymbol{v}|^2_{\mathrm{J,D}}. \tag{IV.4}$$

Using weak Korn's inequality (IV.1) of Lemma IV.1, and the fact that μ is a strictly positive constant, we can state the following coercivity result.

Lemma IV.2 (Coercivity)**.** *There holds for all $\boldsymbol{v}_h \in \boldsymbol{U}_h$,*

$$a_h(\boldsymbol{v}_h, \boldsymbol{v}_h) = \|\boldsymbol{v}_h\|^2_{\mathrm{el}} \geqslant 2\mu\chi C_{\mathrm{K}}^{-2}\|\boldsymbol{\nabla}_h\boldsymbol{v}_h\|^2_{0,\Omega}.$$

Owing to Proposition III.3 and to Lax–Milgram Lemma, the well-posedness of problem (IV.2) is now straightforward. However, a few remarks are in order.

Remark IV.1 (Pure displacement problem)**.** *In the case $\Gamma_{\rm D} = \Gamma$, one may use the results of Remark II.1 and better consider the following discretization of problem* (II.1)*, based on* (II.5)*: Find $\boldsymbol{u}_h \in \mathfrak{CR}_0(\mathcal{K}_h)^d$ (cf.* (III.20)*) such that*

$$\tilde{a}_h(\boldsymbol{u}_h, \boldsymbol{v}_h) = (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega} \qquad \forall \boldsymbol{v}_h \in \mathfrak{CR}_0(\mathcal{K}_h)^d, \tag{IV.5}$$

with symmetric discrete bilinear form $\tilde{a}_h$ such that

$$\tilde{a}_h(\boldsymbol{w}, \boldsymbol{v}) := \mu(\boldsymbol{\nabla}_h \boldsymbol{w}, \boldsymbol{\nabla}_h \boldsymbol{v})_{0,\Omega} + \mu(\boldsymbol{\nabla}_h \cdot \boldsymbol{w}, \boldsymbol{\nabla}_h \cdot \boldsymbol{v})_{0,\Omega} + \lambda(D_h(\boldsymbol{w}), D_h(\boldsymbol{v}))_{0,\Omega}. \tag{IV.6}$$

In (IV.5)*, cell-centered unknowns for a given element $K \in \mathcal{K}_h$ are only linked with the face unknowns located on the boundary of K. As a result, they can be locally eliminated by taking the Schur complement of the corresponding block in the local matrix. This requires, in general, to inverse a $d \times d$ matrix. However, this cost can be further reduced by replacing in* (IV.6) *the term $\mu(\boldsymbol{\nabla}_h \cdot \boldsymbol{u}_h, \boldsymbol{\nabla}_h \cdot \boldsymbol{v}_h)_{0,\Omega}$ by $\mu(D_h(\boldsymbol{u}_h), D_h(\boldsymbol{v}_h))_{0,\Omega}$. This choice avoids, without jeopardizing the approximation, the interaction of the cell unknowns for the different components of the displacement, hence the corresponding block of the local matrix is diagonal and trivial to invert.*

Remark IV.2 (Implementation)**.** *We stress that in the case of problem* (IV.2) *it is not possible to integrate the penalty term in* (IV.3) *using the face barycenter as a quadrature point, since with a choice $\eta = d$ this would yield a vanishing contribution. A quadrature rule exact for polynomials of degree at least* 2 *must be used instead. Also, the penalty term establishes a link between the cell unknowns of neighboring elements. As a result, the stencil is no more compact and it is no longer possible to formulate the method in terms of face unknowns only as for the pure displacement problem; cf. Remark IV.1. Note finally that in the matching simplicial case, it is sufficient to penalize the function jumps on the faces of the primal mesh only, cf. Lemma III.6.*

We mention at this point the recent work of Vohralík and Wohlmuth [79, 80] which proposes efficient implementation strategies for classical nonconforming and mixed finite element approximations of diffusive problems, and addresses general meshes with a different approach.

IV.2 Error estimate

Let $\boldsymbol{U}_{*h} := \boldsymbol{U}_* + \boldsymbol{U}_h$, and extend the bilinear form a_h to $\boldsymbol{U}_{*h} \times \boldsymbol{U}_{*h}$, which consequently extends the norm $\|\cdot\|_{\rm el}$ to $\boldsymbol{U}_{*h}$.

Lemma IV.3 (Conformity and consistency errors)**.** *Let $\boldsymbol{u} \in \boldsymbol{U}$ denote the solution to* (II.3) *and further assume that $\boldsymbol{u} \in \boldsymbol{U}_*$. Then, there holds for all $\boldsymbol{v}_h \in \boldsymbol{U}_h$,*

$$a_h(\boldsymbol{u}, \boldsymbol{v}_h) = (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega} + \mathcal{E}_h(\boldsymbol{v}_h),$$

where

$$\mathcal{E}_h(\boldsymbol{v}_h) := \sum_{F \in \mathcal{F}_{\mathcal{P}_h} \setminus \mathcal{F}_{\mathcal{K}_h}^{\rm b,N}} (\underline{\sigma}(\boldsymbol{u})\boldsymbol{n}_F, [\![\boldsymbol{v}_h]\!])_{0,F} + \lambda((D_h(\boldsymbol{u}) - \boldsymbol{\nabla}\cdot\boldsymbol{u}), \boldsymbol{\nabla}_h \cdot \boldsymbol{v}_h)_{0,\Omega}. \tag{IV.7}$$

Proof. Observing that $\lambda(D_h(\boldsymbol{u}), D_h(\boldsymbol{v}_h))_{0,\Omega} = \lambda(D_h(\boldsymbol{u}), \boldsymbol{\nabla}_h{\cdot}\boldsymbol{v}_h)_{0,\Omega}$, and summing and subtracting $\lambda(\boldsymbol{\nabla}{\cdot}\boldsymbol{u}, \boldsymbol{\nabla}_h{\cdot}\boldsymbol{v}_h)_{0,\Omega}$ from the right-hand side of (IV.3) with $(\boldsymbol{w}, \boldsymbol{v}) = (\boldsymbol{u}, \boldsymbol{v}_h)$ yields

$$a_h(\boldsymbol{u}, \boldsymbol{v}_h) = (\underline{\boldsymbol{\sigma}}(\boldsymbol{u}), \boldsymbol{\nabla}_h\boldsymbol{v}_h)_{0,\Omega} + \lambda((D_h(\boldsymbol{u}) - \boldsymbol{\nabla}{\cdot}\boldsymbol{u}), \boldsymbol{\nabla}_h{\cdot}\boldsymbol{v}_h)_{0,\Omega},$$

where we have used the fact that the penalization term is consistent. Integrating by parts the first term element-wise, rearranging the boundary contributions, and using $[\![\underline{\boldsymbol{\sigma}}(\boldsymbol{u})]\!]_F\boldsymbol{n}_F = \boldsymbol{0}$ and $\{\underline{\boldsymbol{\sigma}}(\boldsymbol{u})\}_F\boldsymbol{n}_F = \underline{\boldsymbol{\sigma}}(\boldsymbol{u})\boldsymbol{n}_F$ for all $F \in \mathcal{F}^{\mathrm{i}}_{\mathcal{P}_h}$ (since $\boldsymbol{u} \in \boldsymbol{U}_*$), as well as $\{\underline{\boldsymbol{\sigma}}(\boldsymbol{u})\}_F\boldsymbol{n}_F = \boldsymbol{0}$ for all $F \in \mathcal{F}^{\mathrm{b,N}}_{\mathcal{K}_h}$, it is inferred

$$\begin{aligned}
(\underline{\boldsymbol{\sigma}}(\boldsymbol{u}), \boldsymbol{\nabla}_h\boldsymbol{v}_h)_{0,\Omega} &= -(\boldsymbol{\nabla}{\cdot}\underline{\boldsymbol{\sigma}}(\boldsymbol{u}), \boldsymbol{v}_h)_{0,\Omega} + \sum_{P\in\mathcal{P}_h}(\underline{\boldsymbol{\sigma}}(\boldsymbol{u})\boldsymbol{n}_P, \boldsymbol{v}_{h|P})_{0,\partial P}\\
&= (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega} + \sum_{F\in\mathcal{F}^{\mathrm{i}}_{\mathcal{P}_h}}([\![\underline{\boldsymbol{\sigma}}(\boldsymbol{u})]\!]\boldsymbol{n}_F, \{\boldsymbol{v}_h\})_{0,F} + \sum_{F\in\mathcal{F}_{\mathcal{P}_h}}(\{\underline{\boldsymbol{\sigma}}(\boldsymbol{u})\}\boldsymbol{n}_F, [\![\boldsymbol{v}_h]\!])_{0,F}\\
&= (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega} + \sum_{F\in\mathcal{F}_{\mathcal{P}_h}\setminus\mathcal{F}^{\mathrm{b,N}}_{\mathcal{K}_h}}(\underline{\boldsymbol{\sigma}}(\boldsymbol{u})\boldsymbol{n}_F, [\![\boldsymbol{v}_h]\!])_{0,F},
\end{aligned}$$

where we have used $-\boldsymbol{\nabla}{\cdot}\underline{\boldsymbol{\sigma}}(\boldsymbol{u}) = \boldsymbol{f}$ a.e. in Ω in the second line (equivalence between (II.3) and (II.1)). This concludes the proof. □

The first term in (IV.7) represents the conformity error while the second one is the consistency error.

We can now derive an error estimate for solutions matching the regularity $\boldsymbol{u} \in \boldsymbol{U}_*$. This additional regularity is verified, e.g., under the assumptions of Lemma II.8. In the following theorem, the weak conformity of the space (i.e. the continuity of mean values at interfaces for a choice $\eta = d$) plays an important role in estimating the boundary contribution in the conformity error.

Theorem IV.1 (Error estimate for (IV.2))**.** *Let $\boldsymbol{u} \in \boldsymbol{U}$ denote the solution to* (II.3) *and additionally assume that $\boldsymbol{u} \in \boldsymbol{U}_*$. Then, there exists $C_{\mathrm{el}} > 0$ independent of the meshsize, of λ, and of $\boldsymbol{u}$ such that, denoting by $\boldsymbol{u}_h \in \boldsymbol{U}_h$ the unique solution to* (IV.2)*, there holds*

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{el}} \leqslant C_{\mathrm{el}}h\mathcal{N}_{\mathrm{el}}(\boldsymbol{u}), \tag{IV.8}$$

where $\mathcal{N}_{\mathrm{el}}(\boldsymbol{u})$ is defined in Lemma II.2.

Proof. We note $a \lesssim b$ the inequality $a \leqslant Cb$ where $C > 0$ has the same dependence as the constant C_{el} in (IV.8). The Cauchy–Schwarz inequality yields boundedness in the form $a_h(\boldsymbol{w}, \boldsymbol{v}) \leqslant \|\boldsymbol{w}\|_{\mathrm{el}}\|\boldsymbol{v}\|_{\mathrm{el}}$ for all $(\boldsymbol{w}, \boldsymbol{v}) \in \boldsymbol{U}_{*h} \times \boldsymbol{U}_{*h}$. This, together with Lemmata IV.2 and IV.3 and the second Strang Lemma [76] (cf. also [35, Lemma 2.25]), yields:

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{el}} \lesssim \inf_{\boldsymbol{v}_h\in\boldsymbol{U}_h}\|\boldsymbol{u} - \boldsymbol{v}_h\|_{\mathrm{el}} + \sup_{\boldsymbol{v}_h\in\boldsymbol{U}_h\setminus\{\boldsymbol{0}\}}\frac{|\mathcal{E}_h(\boldsymbol{v}_h)|}{\|\boldsymbol{v}_h\|_{\mathrm{el}}} := \mathfrak{T}_1 + \mathfrak{T}_2. \tag{IV.9}$$

The first term in the right-hand side depends on the approximation properties of the discrete space, while the second is linked to the conformity and consistency errors. Let $\boldsymbol{w}_h := \mathcal{I}_h^{\mathfrak{CR}}(\boldsymbol{u}) \in \boldsymbol{U}_h$. Using Lemma III.5, Corollary III.1, and the trace inequality (III.5) with $\mathcal{S}_h = \mathcal{P}_h$ combined with Lemma III.5, respectively to treat the three terms in the right-hand side of (IV.4) with $\boldsymbol{v} = \boldsymbol{u} - \boldsymbol{w}_h$, we infer

$$\mathfrak{T}_1 \leqslant \|\boldsymbol{u} - \boldsymbol{w}_h\|_{\mathrm{el}} \lesssim h\|\boldsymbol{u}\|_{2,\Omega} + h\lambda^{1/2}|\boldsymbol{\nabla}{\cdot}\boldsymbol{u}|_{1,\Omega}. \tag{IV.10}$$

To treat the conformity and consistency errors, denote by $\mathfrak{T}_{2,1}$ and $\mathfrak{T}_{2,2}$ the two terms in the right-hand side of (IV.7). Let

$$\underline{\underline{\varphi_\mu}} := 2\mu(\underline{\underline{\varepsilon}}(\boldsymbol{u}) - \Pi_h^0(\underline{\underline{\varepsilon}}(\boldsymbol{u}))), \qquad \psi_\lambda := \lambda(\boldsymbol{\nabla}\cdot\boldsymbol{u} - \Pi_h^0(\boldsymbol{\nabla}\cdot\boldsymbol{u})).$$

Using the continuity of mean values at interfaces (since $\eta = d$) together with the fact that both $\{2\mu\Pi_h^0(\underline{\underline{\varepsilon}}(\boldsymbol{u}))\}_F$ and $\{\lambda\Pi_h^0(\boldsymbol{\nabla}\cdot\boldsymbol{u})\}_F$ are constant on every $F \in \mathcal{F}_{\mathcal{P}_h}$, it is inferred

$$\begin{aligned}\mathfrak{T}_{2,1} &= \sum_{F\in\mathcal{F}_{\mathcal{P}_h}\setminus\mathcal{F}_{\mathcal{K}_h}^{\mathrm{b,N}}} (\{\underline{\sigma}(\boldsymbol{u}) - \Pi_h^0(\underline{\sigma}(\boldsymbol{u}))\}\boldsymbol{n}_F, [\![\boldsymbol{v}_h]\!] - \langle [\![\boldsymbol{v}_h]\!]\rangle_F)_{0,F} \\ &= \sum_{F\in\mathcal{F}_{\mathcal{P}_h}\setminus\mathcal{F}_{\mathcal{K}_h}^{\mathrm{b,N}}} (\{\underline{\underline{\varphi_\mu}}\}\boldsymbol{n}_F, [\![\boldsymbol{v}_h]\!] - \langle [\![\boldsymbol{v}_h]\!]\rangle_F)_{0,F} + \sum_{F\in\mathcal{F}_{\mathcal{P}_h}\setminus\mathcal{F}_{\mathcal{K}_h}^{\mathrm{b,N}}} (\{\psi_\lambda\}\boldsymbol{n}_F, [\![\boldsymbol{v}_h]\!] - \langle [\![\boldsymbol{v}_h]\!]\rangle_F)_{0,F}.\end{aligned}$$

The Cauchy–Schwarz inequality followed by the trace inequality (III.5) with $\mathcal{S}_h = \mathcal{P}_h$, the fact that the maximum number of faces of a pyramid is bounded uniformly in h (cf. Lemma III.1), the approximation properties of the L^2-orthogonal projection, and (III.23) with $\boldsymbol{x}_F = \overline{\boldsymbol{x}}_F$ yield

$$\begin{aligned}\mathfrak{T}_{2,1} &\lesssim \left\{\sum_{P\in\mathcal{P}_h} h\left(\|\underline{\underline{\varphi_\mu}}_{|P}\|_{0,\partial P}^2 + \|\psi_{\lambda|P}\|_{0,\partial P}^2\right)\right\}^{\frac12} \times \left\{\sum_{F\in\mathcal{F}_{\mathcal{P}_h}\setminus\mathcal{F}_{\mathcal{K}_h}^{\mathrm{b,N}}} h_F^{-1}\|[\![\boldsymbol{v}_h]\!] - \langle [\![\boldsymbol{v}_h]\!]\rangle_F\|_{0,F}^2\right\}^{\frac12} \\ &\lesssim h\mathcal{N}_{\mathrm{el}}(\boldsymbol{u})\|\boldsymbol{v}_h\|_{\mathrm{el}},\end{aligned} \tag{IV.11}$$

where the bound $\|\boldsymbol{\nabla}_h\boldsymbol{v}_h\|_{0,\Omega} \lesssim \|\boldsymbol{v}_h\|_{\mathrm{el}}$ is a consequence of Lemma IV.2. Finally, using the Cauchy–Schwarz inequality together with the approximation properties of the L^2-orthogonal projection and Lemma IV.2, it is inferred

$$\mathfrak{T}_{2,2} \leqslant \lambda\|\Pi_h^0(\boldsymbol{\nabla}\cdot\boldsymbol{u}) - \boldsymbol{\nabla}\cdot\boldsymbol{u}\|_{0,\Omega}\|\boldsymbol{\nabla}_h\cdot\boldsymbol{v}_h\|_{0,\Omega} \lesssim h\lambda|\boldsymbol{\nabla}\cdot\boldsymbol{u}|_{1,\Omega}\|\boldsymbol{v}_h\|_{\mathrm{el}}. \tag{IV.12}$$

Using inequalities (IV.10), (IV.11), and (IV.12) to bound the right-hand side of (IV.9) the conclusion follows. □

We are now able to state the main result of this section.

Corollary IV.1 (Uniform convergence with respect to λ)**.** *Under the assumptions of Lemma II.2, the locking-free error estimate* (II.17) *holds true.*

Some remarks are in order.

Remark IV.3 (Use of Lemma II.2)**.** *In the proof of Theorem IV.1, the a priori bound on $\lambda|\boldsymbol{\nabla}\cdot\boldsymbol{u}|_{1,\Omega}$ is only required to bound $\mathfrak{T}_2$. For $\mathfrak{T}_1$, the weaker regularity estimate $\lambda^{1/2}|\boldsymbol{\nabla}\cdot\boldsymbol{u}|_{1,\Omega} \lesssim \|\boldsymbol{f}\|_{0,\Omega}$ is sufficient.*

Remark IV.4 (L^2-error estimate)**.** *Optimal error estimates for the L^2-error on the displacement can be derived using the Aubin–Nitsche trick based on the symmetry of the method.*

Remark IV.5 (Heterogeneous medium)**.** *We consider a material with piecewise constant (scalar-valued) mechanical properties, that is to say $\lambda, \mu \in \mathbb{P}_d^0(\mathcal{K}_h)$. The discretization $\mathcal{K}_h$ of the domain is here assumed to match the heterogeneities of the Lamé parameters. We further assume that*

$$0 < \underline{\mu} \leqslant \mu(\boldsymbol{x}) \leqslant \overline{\mu}, \qquad 0 < \underline{\lambda} \leqslant \lambda(\boldsymbol{x}) \leqslant \overline{\lambda} \qquad \text{for a.e. } \boldsymbol{x} \in \Omega,$$

where $\overline{\lambda} < +\infty$ may tend to infinity in the eventuality of a quasi-incompressible region in the material. According to the work of Di Pietro and Nicaise [29], the regularity result of Lemma II.2 can be extended to the heterogeneous case. Under appropriate assumptions on the dimension ($d = 2$), on the second Lamé parameter μ, and on the geometry (Ω convex), then problem (II.1) *has a unique solution $\boldsymbol{u} \in \boldsymbol{U} \cap H^2(\mathcal{K}_h)^d$, and there exists a positive constant $C_{\mu,\underline{\lambda}}$, only depending on Ω, μ, and $\underline{\lambda}$ (but not on $\overline{\lambda}$) such that, for $\underline{\lambda}$ large enough (which is equivalent to assuming an upper bound on the compressibility contrast),*

$$\mathcal{N}_{\rm el}(\boldsymbol{u}) := \|\boldsymbol{u}\|_{2,\mathcal{K}_h} + |\lambda \boldsymbol{\nabla}\cdot\boldsymbol{u}|_{1,\mathcal{K}_h} \leqslant C_{\mu,\underline{\lambda}} \|\boldsymbol{f}\|_{0,\Omega}. \tag{IV.13}$$

It is possible to design a locking-free method for linear elasticity equations also in the heterogeneous case. Let consider problem (IV.2)*, with a slightly different bilinear form:*

$$a_h(\boldsymbol{w},\boldsymbol{v}) := (2\mu\, \underline{\underline{\varepsilon_h}}(\boldsymbol{w}), \underline{\underline{\varepsilon_h}}(\boldsymbol{v}))_{0,\Omega} + (\lambda D_h(\boldsymbol{w}), D_h(\boldsymbol{v}))_{0,\Omega} + 2\underline{\mu}\chi \sum_{F\in\mathcal{F}_{\mathcal{P}_h}\setminus\mathcal{F}^{\rm b,N}_{\mathcal{K}_h}} h_F^{-1}([\![\boldsymbol{w}]\!],[\![\boldsymbol{v}]\!])_{0,F}, \tag{IV.14}$$

where $0 < \chi \leqslant 1$ is a user-dependent parameter. Remark that (IV.14) *reduces to* (IV.3) *in the homogeneous case. The coercivity of this formulation is expressed as in Lemma IV.2 using $\underline{\mu} > 0$. An error estimate of the form* (IV.8) *can easily be derived by remarking that $\lambda\Pi^0_h(\boldsymbol{\nabla}\cdot\boldsymbol{u}) = \overline{\Pi}^0_h(\lambda\boldsymbol{\nabla}\cdot\boldsymbol{u})$.*

Remark IV.6 (Heterogeneous medium, pure displacement problem)**.** *When considering a pure displacement problem in a heterogeneous medium, the formulation* (II.5) *is no longer equivalent to problem* (II.1)*. Thus, the bilinear form* (IV.6) *is not appropriate to discretize the problem. Taking inspiration from Remark II.1, we consider discretization* (IV.5) *with the following symmetric bilinear form:*

$$\tilde{a}_h(\boldsymbol{w},\boldsymbol{v}) := (2\mu\, \underline{\underline{\varepsilon_h}}(\boldsymbol{w}), \underline{\underline{\varepsilon_h}}(\boldsymbol{v}))_{0,\Omega} + (\lambda D_h(\boldsymbol{w}), D_h(\boldsymbol{v}))_{0,\Omega} + \underline{\mu}\left((\boldsymbol{\nabla}_h\cdot\boldsymbol{w}, \boldsymbol{\nabla}_h\cdot\boldsymbol{v})_{0,\Omega} - (\boldsymbol{\nabla}_h\boldsymbol{w}, \boldsymbol{\nabla}_h\boldsymbol{v}^T)_{0,\Omega}\right), \tag{IV.15}$$

where the stabilization is inspired from (II.7)*. This stabilization introduces a consistency error which tends to zero as h under the regularity $\boldsymbol{u} \in \boldsymbol{U} \cap H^2(\mathcal{K}_h)^d$, meaning that the convergence result* (IV.8) *is unaffected (with an update of the energy norm according to $\tilde{a}_h$ and of $\mathcal{N}_{\rm el}(\boldsymbol{u})$ according to* (IV.13)*). Remark that* (IV.15) *reduces to* (IV.6) *in the homogeneous case. The coercivity of the discretization is ensured with multiplicative constant $\underline{\mu} > 0$. The advantage of such a stabilization in comparison with a jumps penalization is that it does not enlarge the stencil since it is volumetric. Hence, cell unknowns can be eliminated, cf. Remark IV.1. Actually, this stabilization does not even involve the cell unknowns since the subgrid correction of the gradient operator has a vanishing contribution in the stabilization (remark that $(\boldsymbol{a}\otimes\boldsymbol{b}):(\boldsymbol{b}\otimes\boldsymbol{c}) = \mathrm{tr}(\boldsymbol{a}\otimes\boldsymbol{b})\mathrm{tr}(\boldsymbol{c}\otimes\boldsymbol{b})$). Note however that the interaction between cell unknowns for the different components of the displacement cannot be avoided in that case owing to the first term of the bilinear form.*

IV.3 Links with finite volume and finite element methods

In this section we investigate the links between our method and classical finite volume or finite element methods. We show (for the pure displacement problem) that the treatment of the right-hand side determines the framework to which the method belongs.

IV.3.1 Flux formulation and local conservation, the finite volume side

Let consider the pure displacement problem and its approximation (IV.5). Here, we do not make any assumption on the value of the stabilization parameter $\eta > 1$. Following [40, Section 2.4], it is possible to reformulate the discrete bilinear form (IV.6) in terms of numerical fluxes. More specifically, introducing $\mathbb{U}_h := \mathbb{V}_{h,0}^d$ where $\mathbb{V}_{h,0}$ is defined by (III.20), let $\boldsymbol{w}_h, \boldsymbol{v}_h \in \boldsymbol{U}_h$ be two discrete functions and denote by $\mathbb{w}_h = (\mathbb{w}_{h,i})_{1\leqslant i\leqslant d} \in \mathbb{U}_h$ and $\mathbb{v}_h = (\mathbb{v}_{h,i})_{1\leqslant i\leqslant d} \in \mathbb{U}_h$ the corresponding vectors of DOFs, where, for all $i \in \{1,\dots,d\}$, $\mathbb{w}_{h,i}$ and $\mathbb{v}_{h,i}$ are the vectors of DOFs associated to the i-th components of $\boldsymbol{w}_h$ and $\boldsymbol{v}_h$ respectively. We show that there exists a family of numerical fluxes $(\boldsymbol{\Phi}_{K,F}(\mathbb{w}_h))_{K\in\mathcal{K}_h,F\in\mathcal{F}_K}$ with $\boldsymbol{\Phi}_{K,F}(\mathbb{w}_h) = (\Phi_{K,F,i}(\mathbb{w}_h))_{1\leqslant i\leqslant d}$ such that

$$\tilde{a}_h(\boldsymbol{w}_h,\boldsymbol{v}_h) = \sum_{K\in\mathcal{K}_h}\sum_{F\in\mathcal{F}_K} \boldsymbol{\Phi}_{K,F}(\mathbb{w}_h)\cdot(\boldsymbol{v}_F - \boldsymbol{v}_K), \tag{IV.16}$$

with $\tilde{a}_h$ defined by (IV.6).

Proposition IV.1 (Flux formulation)**.** *For all $\boldsymbol{w}_h, \boldsymbol{v}_h \in \boldsymbol{U}_h$, the flux formulation* (IV.16) *is obtained by setting for all $K \in \mathcal{K}_h$, $F \in \mathcal{F}_K$, and $i \in \{1,\dots,d\}$,*

$$\Phi_{K,F,i}(\mathbb{w}_h) := \sum_{F'\in\mathcal{F}_K} |K_{F'}| \left[\mu \boldsymbol{G}_{K_{F'}}(\mathbb{w}_{h,i})\cdot\boldsymbol{y}_{F',F}^K + \left(\sum_{j=1}^d \mu \boldsymbol{G}_{K_{F'}}(\mathbb{w}_{h,j}) + \lambda \boldsymbol{G}_K(\mathbb{w}_{h,j}) \right) \cdot \boldsymbol{e}_j \left(\boldsymbol{y}_{F',F}^K \cdot \boldsymbol{e}_i \right) \right],$$

where $\mathbb{w}_h, \mathbb{v}_h \in \mathbb{U}_h$ are the vectors of DOFs associated to $\boldsymbol{w}_h$ and $\boldsymbol{v}_h$ respectively, $(\boldsymbol{e}_i)_{1\leqslant i\leqslant d}$ denotes the canonical basis of $\mathbb{R}^d$, and

$$\boldsymbol{y}_{F',F}^K := \begin{cases} \frac{|F|}{|K|}\boldsymbol{n}_{K,F} + \frac{\eta}{d_{K,F}}\left(1 - \frac{|F|}{|K|}\boldsymbol{n}_{K,F}\cdot(\overline{\boldsymbol{x}}_F - \boldsymbol{x}_K)\right)\boldsymbol{n}_{K,F} & \text{if } F = F', \\ \frac{|F|}{|K|}\boldsymbol{n}_{K,F} - \frac{\eta}{d_{K,F'}|K|}\,|F|\,\boldsymbol{n}_{K,F}\cdot(\overline{\boldsymbol{x}}_{F'} - \boldsymbol{x}_K)\boldsymbol{n}_{K,F'} & \text{otherwise.} \end{cases} \tag{IV.17}$$

Proof. For all $\mathbb{v}_h \in \mathbb{V}_h$, all $K \in \mathcal{K}_h$, and all $F' \in \mathcal{F}_K$, there holds with $\boldsymbol{G}_{K_{F'}}(\mathbb{v}_h)$ defined by (III.8) (cf. [40, eq. (26) et seq.]),

$$\boldsymbol{G}_{K_{F'}}(\mathbb{v}_h) = \sum_{F\in\mathcal{F}_K} (v_F - v_K)\,\boldsymbol{y}_{F',F}^K. \tag{IV.18}$$

Using (III.8) and (III.10), and observing that $\lambda(D_h(\boldsymbol{w}_h), D_h(\boldsymbol{v}_h))_{0,\Omega} = \lambda(D_h(\boldsymbol{w}_h), \boldsymbol{\nabla}_h\cdot\boldsymbol{v}_h)_{0,\Omega}$ owing to (III.18) together with the properties of the L^2-orthogonal projector, it is inferred

$$\tilde{a}_h(\boldsymbol{w}_h,\boldsymbol{v}_h) = \\ \sum_{i=1}^d \sum_{K\in\mathcal{K}_h}\sum_{F'\in\mathcal{F}_K} |K_{F'}| \left[\mu\boldsymbol{G}_{K_{F'}}(\mathbb{w}_{h,i}) + \left(\sum_{j=1}^d \mu\boldsymbol{G}_{K_{F'}}(\mathbb{w}_{h,j})\cdot\boldsymbol{e}_j + \lambda\boldsymbol{G}_K(\mathbb{w}_{h,j})\cdot\boldsymbol{e}_j \right)\boldsymbol{e}_i \right] \cdot \boldsymbol{G}_{K_{F'}}(\mathbb{v}_{h,i}).$$

The conclusion follows using the expression (IV.18) for $\boldsymbol{G}_{K_{F'}}(\mathbb{v}_{h,i})$ and exchanging the sums of indices F and F'. □

The main interest of this alternative formulation is that it allows to prove a local conservation property similar to those encountered in standard finite volume methods. Recalling the expression (IV.16) for the bilinear form $\tilde{a}_h$ and using the cell unknown to approximate the right-hand side in each element, the discrete problem (IV.5) in algebraic form reads: Find $\mathbb{u}_h \in \mathbb{U}_h$ such that for all $\mathbb{v}_h \in \mathbb{U}_h$ there holds,

$$\sum_{K\in\mathcal{K}_h}\sum_{F\in\mathcal{F}_K} \boldsymbol{\Phi}_{K,F}(\mathbb{u}_h)\cdot(\boldsymbol{v}_F - \boldsymbol{v}_K) = \sum_{K\in\mathcal{K}_h} |K|\,\boldsymbol{f}_K\cdot\boldsymbol{v}_K, \tag{IV.19}$$

where $\boldsymbol{f}_K := \frac{1}{|K|}\int_K \boldsymbol{f}\mathrm{d}\boldsymbol{x}$ for all $K \in \mathcal{K}_h$. Consider now an interface $F \in \mathcal{F}^{\mathrm{i}}_{\mathcal{K}_h}$ such that $F \subset \partial K_1 \cap \partial K_2$, and let for $i \in \{1,\dots,d\}$ $\mathbb{v}_{h,i}$ be such that $v_{F,i} = 1$, $v_{F',i} = 0$ for all $F' \in \mathcal{F}_{\mathcal{K}_h}\backslash\{F\}$, and $v_{K,i} = 0$ for all $K \in \mathcal{K}_h$, with the other components of $\mathbb{v}_h$ that are zero. There follows from (IV.16),

$$\Phi_{K_1,F,i}(\mathbb{u}_h) = -\Phi_{K_2,F,i}(\mathbb{u}_h), \tag{IV.20}$$

i.e., the method is locally conservative. An important remark is that the loading term does not appear in (IV.20) since its approximation in (IV.19) only involves cell DOFs. The method written under the form (IV.19) is the exact application of HFV to the pure displacement problem of elasticity (with the restriction that we made here no assumption on $\eta > 1$) with reduced integration of the divergence operator. The treatment (IV.19) of the right-hand side can be proved to introduce a consistency error which converges to zero as h, which means that it does not modify the error estimate (IV.8).

IV.3.2 Link with the Crouzeix–Raviart solution, the finite element side

Let consider again the pure displacement problem and its approximation (IV.5), and let $\eta > 1$. We consider a matching simplicial mesh that we denote $\mathcal{T}_h$ and we let

$$\mathbb{CR}_0(\mathcal{T}_h) := \{v_h \in \mathbb{CR}(\mathcal{T}_h) \mid v_h(\overline{\boldsymbol{x}}_F) = 0,\ \forall F \in \mathcal{F}^{\mathrm{b}}_{\mathcal{K}_h}\}, \tag{IV.21}$$

and $\hat{\boldsymbol{U}}_h := \mathbb{CR}_0(\mathcal{T}_h)^d$. We show in this section that a suitable treatment of the right-hand side allows to recover the Crouzeix–Raviart solution $\hat{\boldsymbol{u}}_h \in \hat{\boldsymbol{U}}_h$ such that (cf. [16]),

$$\tilde{a}_h(\hat{\boldsymbol{u}}_h, \boldsymbol{v}_h) = (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega} \qquad \forall \boldsymbol{v}_h \in \hat{\boldsymbol{U}}_h. \tag{IV.22}$$

Let $W(\mathcal{P}_h) := \{v \in H^1(\mathcal{P}_h) \mid \langle[\![v]\!]\rangle_F = 0 \text{ for all } F \in \mathcal{F}^{\mathrm{i}}_{\mathcal{P}_h}\}$, and denote by $\mathcal{I}_h^{\mathbb{CR}} : W(\mathcal{P}_h) \to \mathbb{CR}(\mathcal{T}_h)$ the interpolator that maps a function $v \in W(\mathcal{P}_h)$ on the function $v_h := \mathcal{I}_h^{\mathbb{CR}}(v) \in \mathbb{CR}(\mathcal{T}_h)$ such that $v_h(\overline{\boldsymbol{x}}_F) = \langle v\rangle_F$ for all $F \in \mathcal{F}_{\mathcal{T}_h}$. We consider the following variation of (IV.5): Find $\boldsymbol{u}_h \in \mathfrak{CR}_0(\mathcal{T}_h)^d$ such that

$$\tilde{a}_h(\boldsymbol{u}_h, \boldsymbol{v}_h) = (\boldsymbol{f}, \mathcal{I}_h^{\mathbb{CR}}(\boldsymbol{v}_h))_{0,\Omega} \qquad \forall \boldsymbol{v}_h \in \mathfrak{CR}_0(\mathcal{T}_h)^d, \tag{IV.23}$$

where the sole difference with respect to (IV.5) lies in the treatment of the right-hand side.

Lemma IV.4 (Relation between (IV.22) and (IV.23))**.** *There holds $\boldsymbol{u}_h = \hat{\boldsymbol{u}}_h$.*

Proof. Let $\mathbb{U}_h \ni \hat{\mathbb{u}}_h = (\hat{\mathbb{u}}_{h,i})_{1\leqslant i\leqslant d}$ be such that, for $i \in \{1,\dots,d\}$,

$$\hat{\mathbb{u}}_{h,i} := ((\hat{u}_{h,i}(\boldsymbol{x}_K))_{K\in\mathcal{K}_h}, (\hat{u}_{h,i}(\overline{\boldsymbol{x}}_F))_{F\in\mathcal{F}_{\mathcal{K}_h}}) \in \mathbb{V}_{h,0}.$$

Clearly, for all $K \in \mathcal{K}_h$, all $F \in \mathcal{F}_K$, and all $i \in \{1,\dots,d\}$, $\boldsymbol{R}_{K_F}(\hat{\mathbb{u}}_{h,i}) = \boldsymbol{0}$, hence $\boldsymbol{G}_{K_F}(\hat{\mathbb{u}}_{h,i}) = \boldsymbol{G}_K(\hat{\mathbb{u}}_{h,i}) = (\boldsymbol{\nabla}_h\hat{u}_{h,i})_{|K}$. As a consequence, $\hat{\boldsymbol{u}}_h = \mathfrak{R}_h(\hat{\mathbb{u}}_h)$. Accounting for this fact, there holds for all $\boldsymbol{v}_h \in \mathfrak{CR}_0(\mathcal{T}_h)^d$ such that $\boldsymbol{v}_h = \mathfrak{R}_h(\mathbb{v}_h)$ with $\mathbb{v}_h \in \mathbb{U}_h$,

$$\begin{aligned}
\tilde{a}_h(\hat{\boldsymbol{u}}_h, \boldsymbol{v}_h) &= \sum_{i=1}^d \sum_{K\in\mathcal{K}_h} \sum_{F\in\mathcal{F}_K} |K_F| \Big\{\mu\boldsymbol{G}_K(\hat{\mathbb{u}}_{h,i})\cdot\boldsymbol{G}_{K_F}(\mathbb{v}_{h,i}) + \mu\boldsymbol{G}_K(\hat{\mathbb{u}}_{h,i})\cdot\boldsymbol{e}_i D_{K_F}(\mathbb{v}_h) \\
&\qquad + \lambda\boldsymbol{G}_K(\hat{\mathbb{u}}_{h,i})\cdot\boldsymbol{e}_i D_K(\mathbb{v}_h)\Big\} \\
&= \sum_{i=1}^d \sum_{K\in\mathcal{K}_h} |K| \Big\{\mu\boldsymbol{G}_K(\hat{\mathbb{u}}_{h,i})\cdot\boldsymbol{G}_K(\mathbb{v}_{h,i}) + (\mu+\lambda)\boldsymbol{G}_K(\hat{\mathbb{u}}_{h,i})\cdot\boldsymbol{e}_i D_K(\mathbb{v}_h)\Big\} \\
&= \tilde{a}_h(\hat{\boldsymbol{u}}_h, \mathcal{I}_h^{\mathbb{CR}}(\boldsymbol{v}_h)) = (\boldsymbol{f}, \mathcal{I}_h^{\mathbb{CR}}(\boldsymbol{v}_h))_{0,\Omega},
\end{aligned}$$

where the first passage is a consequence of (III.15) and where we have let, for the sake of conciseness, $D_K(\mathbb{v}_h) := \sum_{j=1}^d \boldsymbol{G}_K(\mathbb{v}_{h,j})\cdot\boldsymbol{e}_j$ and $D_{K_F}(\mathbb{v}_h) := \sum_{j=1}^d \boldsymbol{G}_{K_F}(\mathbb{v}_{h,j})\cdot\boldsymbol{e}_j$. Owing to the coercivity of $\tilde{a}_h$, problem (IV.23) admits a unique solution and we therefore conclude that $\hat{\boldsymbol{u}}_h = \boldsymbol{u}_h$. □

Morally, as soon as the right-hand side does not *see* the cell unknowns, the system forces the subgrid corrections to vanish, and the two solutions coincide. When considering the more general approximation (IV.2), it is a simple matter to prove that the solution on a matching simplicial mesh coincides with the solution of the stabilized Crouzeix–Raviart method developed by Hansbo and Larson [50]. This is a consequence of Lemma III.6 and of the fact that the subgrid corrections vanish on primal faces.

IV.4 Numerical examples

In this section we provide a selection of two-dimensional numerical examples that illustrate the different results of this chapter, namely the ability of our method to treat heterogeneous media, its robustness with respect to numerical locking, and its adaptivity to fairly general meshes. When it is relevant, we propose a comparison with a conforming $\mathbb{P}_1^d$ finite element method (on a matching simplicial mesh sequence). The implementation has been realized as a 2D `C++` prototype based on recent open source libraries, and relies on the general framework recently introduced in [26, 27], to which we refer for futher details.

IV.4.1 Mesh families and error measure

We consider several two-dimensional mesh families, that are mainly inspired from the FVCA5 benchmark:

(a) a *matching triangular* mesh sequence, which will be useful for comparison with the conforming $\mathbb{P}_1^d$ finite element method; cf. Figure IV.1a;

(b) a *Cartesian* mesh sequence, which is the most widely used grid type in reservoir simulation as it forms the basis of CPG (Corner Point Geometry) meshes [69]; cf. Figure IV.1b;

(c) a *locally refined Cartesian* mesh sequence, which gives an example of nonconforming h-refinement as it can be encountered in the context of LGR (Local Grid Refinement) in specific locations (like in near wellbore regions) where the resolution needs to be increased, or when moving fronts are present; besides, it is a good benchmark to assess the correct treatment of nonmatching interfaces; cf. Figure IV.1c;

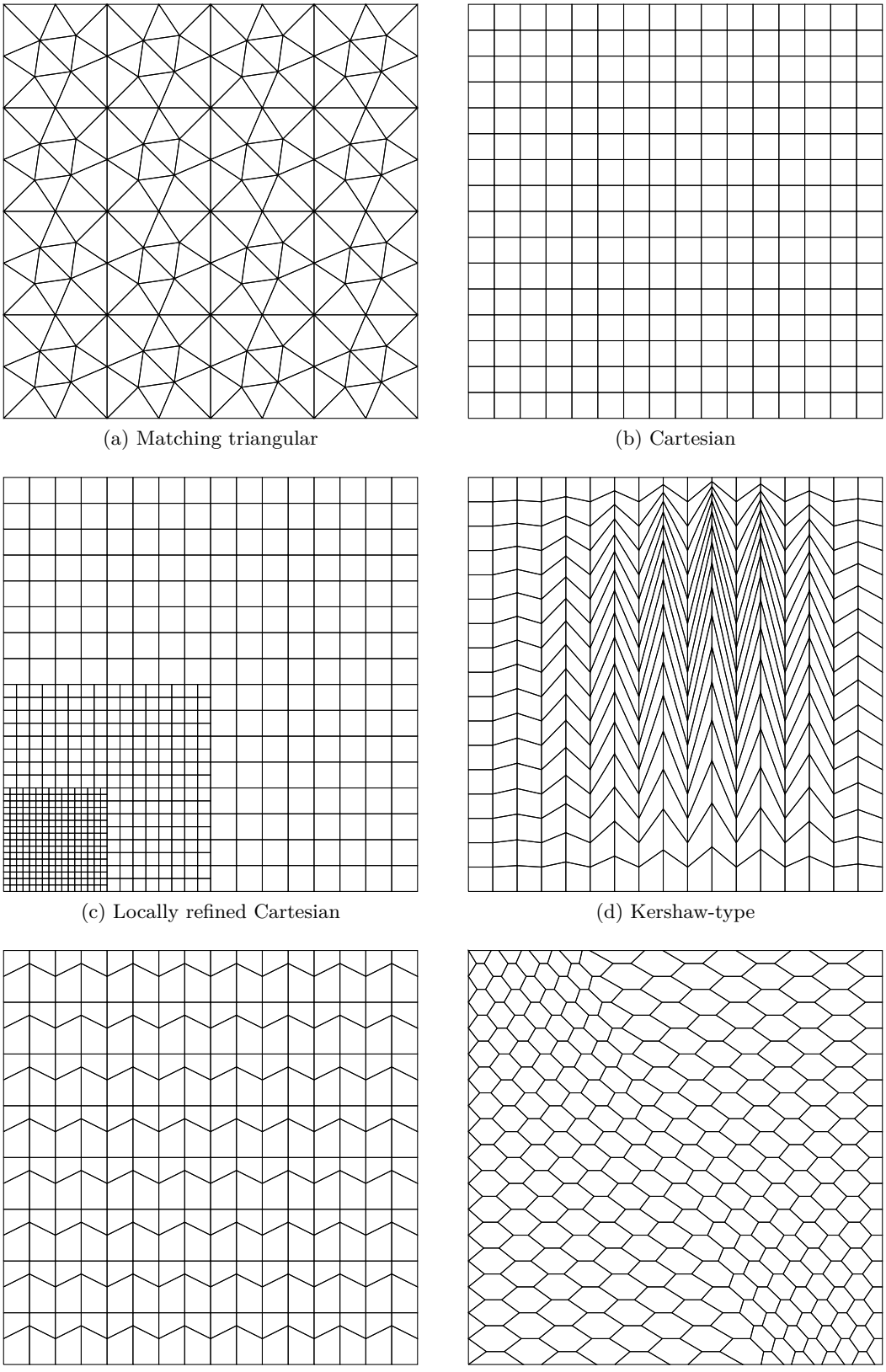
(d) a *Kershaw-type* mesh sequence, which is of great practical interest since it may represent a geological porous medium that has historically undergone non-smooth deformations toward a highly skewed state; cf. Figure IV.1d;

(e) a *trapezoidal* mesh sequence, which illustrates the case when the refined grid elements do not converge to parallelograms, meaning that the shape factor of the grids does not improve with grid refinement; cf. Figure IV.1e;

(f) a *hexagonal-dominant* mesh sequence, which is an example of challenging tilted grid featuring different polygonal elements; cf. Figure IV.1f.

The mesh families are such that their meshsize is approximately divided by two between two successive members. We test the different discretizations of the linear elasticity equations we have introduced:

(i) for (possibly) mixed-type (Dirichlet–Neumann) boundary conditions, we use discretization (IV.2), that will be denoted CRg-JS (for Jumps Stabilization), with bilinear form



(a) Matching triangular

(b) Cartesian

(c) Locally refined Cartesian

(d) Kershaw-type

(e) Trapezoidal

(f) Hexagonal-dominant

Figure IV.1: Members of the 2D mesh families for the numerical tests of Section IV.4.

- (IV.3) in the homogeneous case;
- (IV.14) in the heterogeneous case;

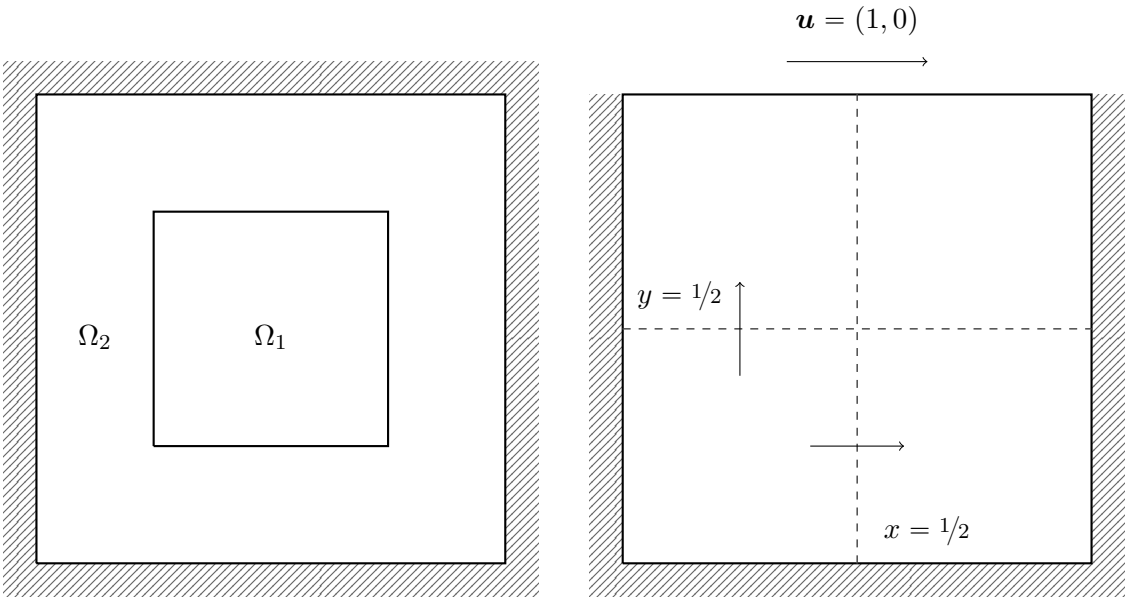
(ii) for pure Dirichlet boundary conditions (pure displacement problem), we use discretization (IV.5), that will be denoted CRg-VS (for Volumetric Stabilization), with bilinear form
- (IV.6) in the homogeneous case;
- (IV.15) in the heterogeneous case.

Note that for both methods, the test-cases presented below have been computed considering pure Dirichlet boundary conditions. However, experiments have been realized for the CRg-JS method and confirm that this latter enables to treat correctly mixed-type boundary conditions. We make the choice $\eta = d$ in (III.9) for the subgrid stabilization parameter. The right-hand side of (IV.2) and (IV.5) is approximated in a finite volume way using the cell unknown as a quadrature point, cf. (IV.19) and Section IV.3.1. The H^1 and L^2 relative errors are measured as

$$\frac{\|\boldsymbol{\nabla} \boldsymbol{u} - \boldsymbol{\nabla}_h \boldsymbol{u}_h\|_{0,\Omega}}{\|\boldsymbol{\nabla} \boldsymbol{u}\|_{0,\Omega}} \approx \frac{\left(\sum_{K\in\mathcal{K}_h} \sum_{1\leqslant i,j\leqslant d} |K| \Big((\boldsymbol{\nabla} \boldsymbol{u})_{ij}(\boldsymbol{x}_K) - \boldsymbol{G}_{K,j}(\mathbb{u}_{h,i}) \Big)^2 \right)^{1/2}}{\left(\sum_{K\in\mathcal{K}_h} |K| \, |(\boldsymbol{\nabla} \boldsymbol{u})(\boldsymbol{x}_K)|^2 \right)^{1/2}}, \tag{IV.24a}$$

$$\frac{\|\boldsymbol{u} - \boldsymbol{u}_h\|_{0,\Omega}}{\|\boldsymbol{u}\|_{0,\Omega}} \approx \frac{\left(\sum_{K\in\mathcal{K}_h} |K| \, |\boldsymbol{u}(\boldsymbol{x}_K) - \boldsymbol{u}_K|^2 \right)^{1/2}}{\left(\sum_{K\in\mathcal{K}_h} |K| \, |\boldsymbol{u}(\boldsymbol{x}_K)|^2 \right)^{1/2}}. \tag{IV.24b}$$

For the sake of simplicity, these relative errors are referred to as $\|\boldsymbol{\nabla} \boldsymbol{u} - \boldsymbol{\nabla}_h \boldsymbol{u}_h\|_{0,\Omega}$ and $\|\boldsymbol{u} - \boldsymbol{u}_h\|_{0,\Omega}$ in the plots axes.



(a) Heterogeneous test-case of Section IV.4.2 (b) Closed cavity test-case of Section IV.4.3.2

Figure IV.2: Configuration for the test-cases of Section IV.4.

IV.4.2 Heterogeneous medium

We investigate the robustness of the discretizations (IV.2)–(IV.14) and (IV.5)–(IV.15) with respect to the heterogeneity ratio of the medium. For that purpose, let consider the following manufactured solution inspired by the one proposed in [64] by Nordbotten. Let $\Omega := (0,1)^2$ such that $\Omega := \overline{\Omega_1} \cup \Omega_2$, where $\Omega_1 := (\frac{1}{4}, \frac{3}{4})^2$ and $\Omega_2 := \Omega \backslash \overline{\Omega_1}$. We consider a material with piecewise constant Lamé parameters such that

$$\lambda = \mu = \kappa \quad \text{in } \Omega_1, \qquad \lambda = \mu = 1 \quad \text{in } \Omega_2,$$

where $\kappa \in \{10^{-6}, 1, 10^6\}$ enables to vary the heterogeneity ratio, the case $\kappa = 1$ corresponding to the homogeneous case where (IV.14) reduces to (IV.3) and (IV.15) to (IV.6). An illustration of the geometry is provided in Figure IV.2a.

For this material, we consider the following solution, which honors a homogeneous Dirichlet boundary condition:

$$u_x = \frac{1}{\lambda} \sin(4\pi x) \sin(4\pi y), \qquad u_y = u_x. \tag{IV.25}$$

This solution is continuous on Ω, with continuous (independent of κ) stress. However, note that this solution does not exhibit the regularity required by Theorem IV.1. The body force is obtained by taking the divergence of the stress tensor:

$$f_x = 64\pi^2 \sin(4\pi x) \sin(4\pi y) - 32\pi^2 \cos(4\pi x) \cos(4\pi y), \qquad f_y = f_x.$$

We consider the Cartesian mesh sequence to which belongs the member of Figure IV.1b. This mesh sequence matches the heterogeneities in the mechanical properties of the material. For $\kappa \in \{10^{-6}, 1, 10^6\}$, we plot on Figure IV.3 the H^1 and L^2 relative errors (computed as in (IV.24a) and (IV.24b)) for both discretizations CRg-JS (IV.2)–(IV.14) and CRg-VS (IV.5)–(IV.15). For the discretization CRg-JS, we take a stabilization parameter $\chi = 1$ in (IV.14). We also provide a comparison with the conforming $\mathbb{P}_1^d$ finite element method on the matching triangular mesh sequence of Figure IV.1a.

The results show that both discretizations CRg-JS and CRg-VS are insensitive to the heterogeneity contrast, and exhibit a second order convergence rate in the L^2 norm. In the H^1 seminorm, both discretizations exhibit a supra-convergent behavior with a second order convergence rate, which is due to the Cartesian type of the grid sequence. First order is obtained on the matching triangular mesh sequence of Figure IV.1a (not ploted here).

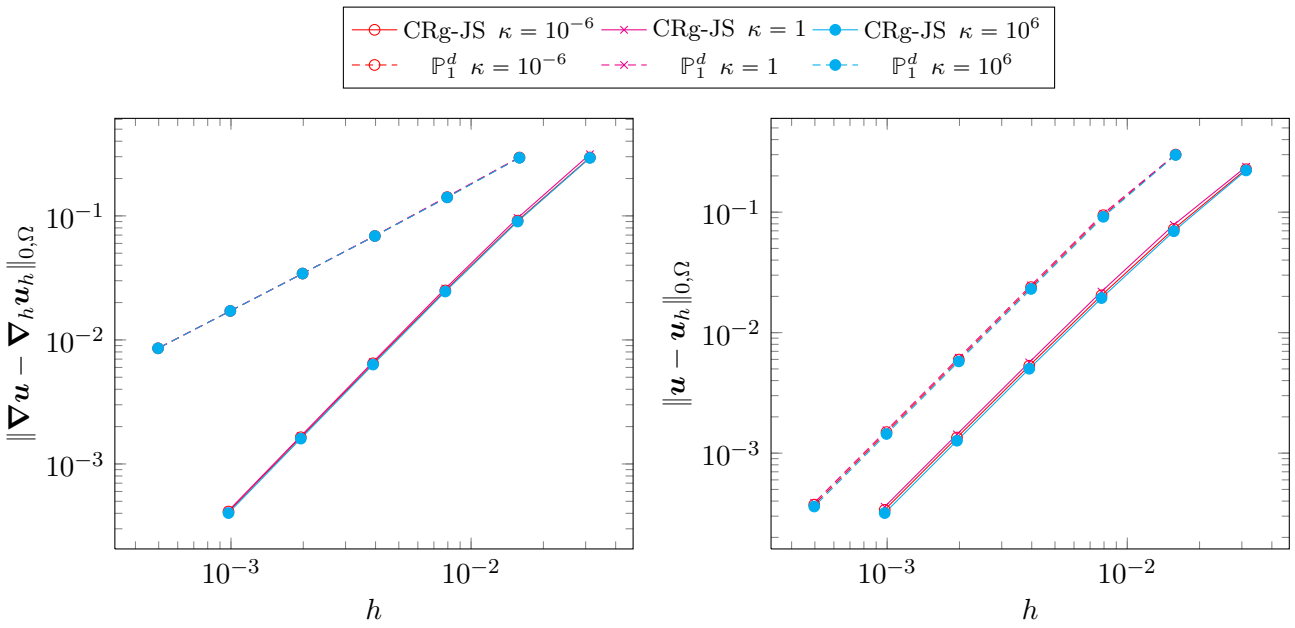
IV.4.3 Quasi-incompressible materials

We now investigate the robustness of the discretizations CRg-JS and CRg-VS with respect to the first Lamé parameter λ. Designing a relevant test-case to assess the robustness with respect to locking is not an easy task as it must satisfy several features:

- the displacement field $\boldsymbol{u}$ must be such that $|\boldsymbol{\nabla}\cdot\boldsymbol{u}| \to 0$ as $\lambda \to +\infty$ and $\boldsymbol{u}$ does not tend to a constant field;
- the body force $\boldsymbol{f}$ must be such that $\|\boldsymbol{f}\|_{0,\Omega}$ tends to a bounded constant as $\lambda \to +\infty$.

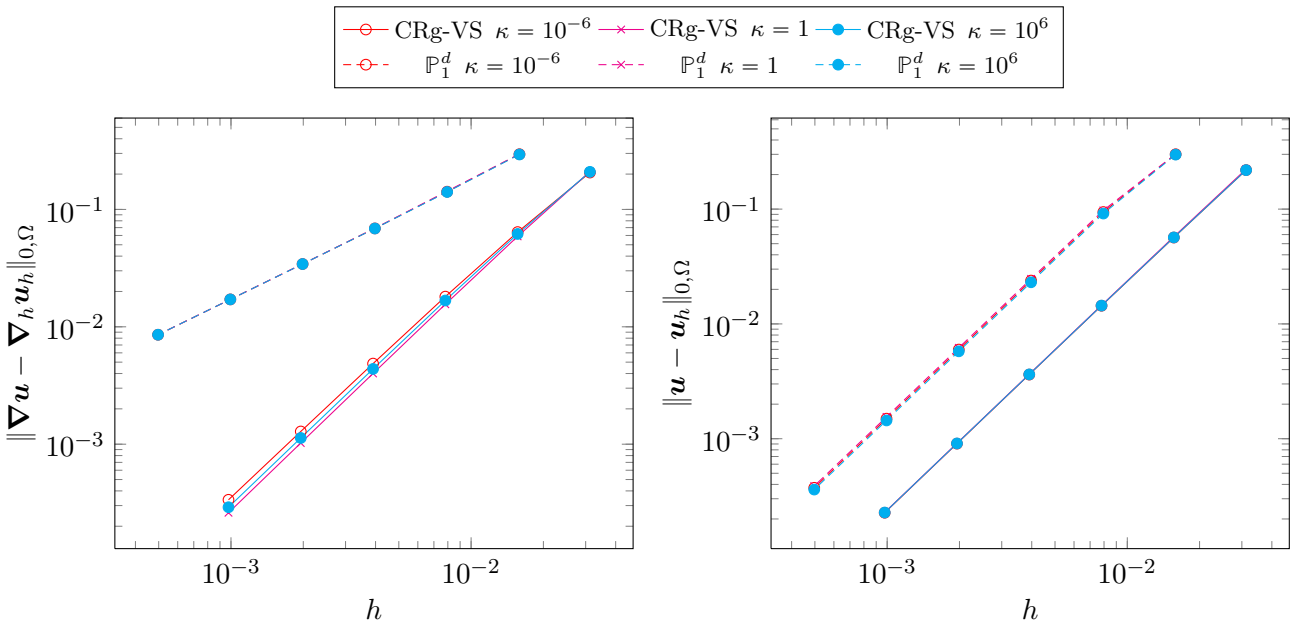
This corresponds to practically relevant cases, where for the same body force applied to materials with different compressibility, the approximation of the displacement field by usual finite element methods deteriorates as the compressibility of the material decreases. This is the sign of numerical locking, and of the fact that an estimate of the form (II.17) does not hold for such methods, since they are not able to approximate nonconstant divergence-free fields.

We study two different relevant test-cases for which we provide a comparison of the results with conforming finite elements.



(a) H^1 relative error

(b) L^2 relative error



(c) H^1 relative error

(d) L^2 relative error

Figure IV.3: Effect of the heterogeneity ratio κ on the discretizations CRg-JS and CRg-VS (solid lines) vs. $\mathbb{P}_1^d$ (dashed lines).

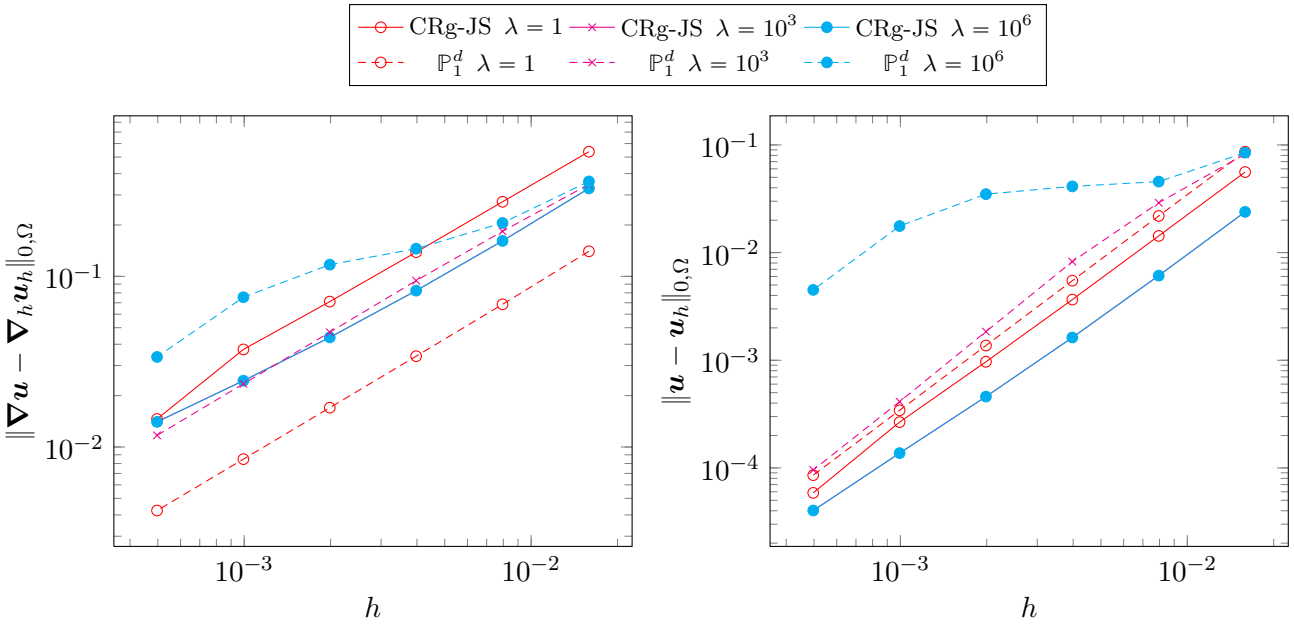
IV.4.3.1 A manufactured solution

Let $\Omega := (0,1)^2$ and let consider a homogeneous material with (constant) Lamé parameters such that $\mu = 1$ and $\lambda \in \{1, 10^3, 10^6\}$. According to the relation

$$\nu = \frac{\lambda}{2(\lambda + \mu)},$$

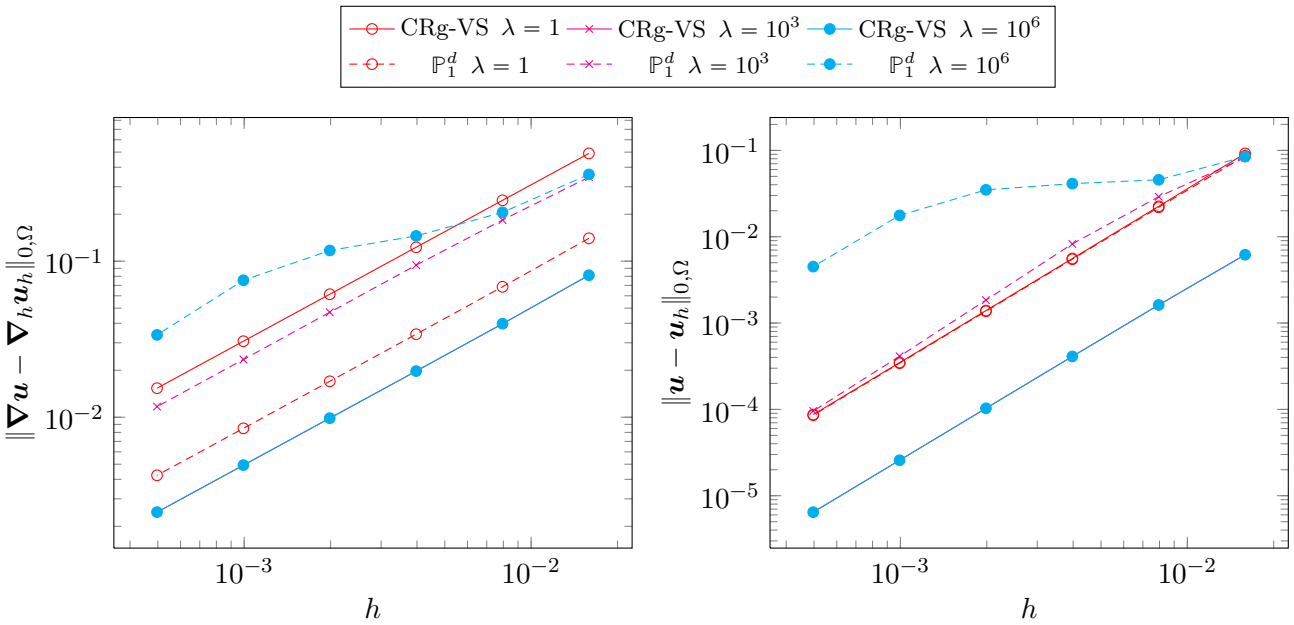
consequence of (II.2), we consider materials with Poisson's ratio $\nu \in \{0.25, 0.4995, 0.4999995\}$, i.e. $\nu \to 0.5$. We consider the following manufactured solution:

$$u_x = \cos(\frac{2\pi}{\lambda}x)\sin(2\pi y), \qquad u_y = \sin(2\pi x)\cos(\frac{2\pi}{\lambda}y).$$



(a) H^1 relative error

(b) L^2 relative error



(c) H^1 relative error

(d) L^2 relative error

Figure IV.4: Effect of the first Lamé parameter λ on the discretizations CRg-JS and CRg-VS (solid lines) vs. $\mathbb{P}_1^d$ (dashed lines).

Note that this solution exhibits the regularity required by Theorem IV.1. This solution satisfies $\boldsymbol{u} \to (\sin(2\pi y), \sin(2\pi x))$ as $\lambda \to +\infty$, and $|\nabla \cdot \boldsymbol{u}| \to 0$. The body force is obtained by

taking the divergence of the stress tensor:

$$f_x = 4\pi^2(\mu + \frac{1}{\lambda} + \frac{2\mu}{\lambda^2})\cos(\frac{2\pi}{\lambda}x)\sin(2\pi y) + 4\pi^2(1+\frac{\mu}{\lambda})\cos(2\pi x)\sin(\frac{2\pi}{\lambda}y),$$
$$f_y = 4\pi^2(\mu + \frac{1}{\lambda} + \frac{2\mu}{\lambda^2})\sin(2\pi x)\cos(\frac{2\pi}{\lambda}y) + 4\pi^2(1+\frac{\mu}{\lambda})\sin(\frac{2\pi}{\lambda}x)\cos(2\pi y).$$

In the limit $\lambda \to +\infty$, $\boldsymbol{f} \to 4\pi^2\mu\,(\sin(2\pi y), \sin(2\pi x))$, and $\|\boldsymbol{f}\|_{0,\Omega}$ remains bounded.

We consider the matching triangular mesh sequence of Figure IV.1a. For $\lambda \in \{1, 10^3, 10^6\}$, we plot on Figure IV.4 the H^1 and L^2 relative errors (computed as in (IV.24a) and (IV.24b)) for both discretizations CRg-JS (IV.2)–(IV.3) and CRg-VS (IV.5)–(IV.6), and we compare it with the conforming $\mathbb{P}_1^d$ finite element method. For the discretization CRg-JS, we take a stabilization parameter $\chi = 0.2$ in (IV.3).

As far as the reference $\mathbb{P}_1^d$ method is concerned, the results go worse as λ grows. For $\lambda = 10^6$, clear signs of numerical locking are observed, with a genuinely pronounced loss of convergence in the L^2 norm. The convergence rate is only recovered for very refined grids. For small λ, the $\mathbb{P}_1^d$ method outperforms both discretizations in the H^1 seminorm (but not in the L^2 norm). In low compressibility regimes, both discretizations show robustness with respect to numerical locking. As expected, the errors scale with the L^2 norm of the body force, which decreases in that case as λ grows. We remark that the precision of the CRg-VS method is better than the one of the CRg-JS method, but optimal convergence rates are reached for both methods in the L^2 norm and H^1 seminorm.

IV.4.3.2 The closed cavity problem

We consider the closed cavity problem of Hansbo and Larson [50]. Although this problem does not exhibit the regularity required by Theorem IV.1, it is included as it is one of the simplest benchmarks for numerical locking.

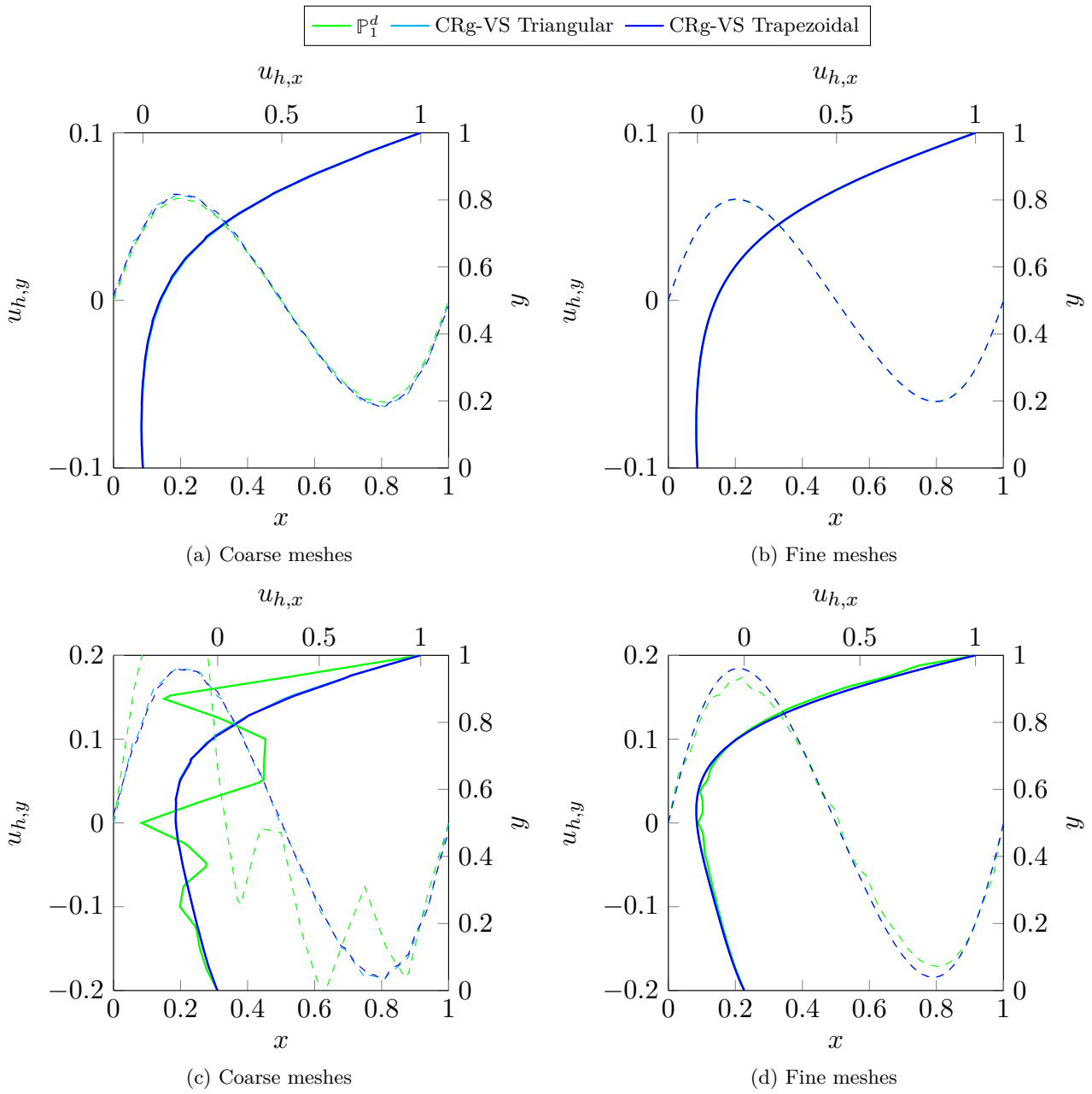
Let $\Omega := (0,1)^2$, $\boldsymbol{f} \equiv \boldsymbol{0}$, and prescribe a horizontal displacement $\boldsymbol{u} = (1,0)$ on the upper side of Ω, and $\boldsymbol{u} = \boldsymbol{0}$ on the remaining three. An illustration of the problem is provided in Figure IV.2b. We consider a homogeneous material with elastic modulus and Poisson's ratio such that $E = 1000$ and $\nu \in \{0.25, 0.4999\}$ respectively. The Lamé parameters are derived from the relations (II.2), which give $\mu \in \{400, 333\}$ and $\lambda \in \{400, 1\,666\,444\}$.

For $\nu \in \{0.25, 0.4999\}$, the discrete problem is solved on the trapezoidal mesh sequence of Figure IV.1e for the CRg-VS (IV.5)–(IV.6) method, and on the matching triangular mesh sequence of Figure IV.1a for the CRg-VS and $\mathbb{P}_1^d$ methods, the $\mathbb{P}_1^d$ method being taken as a reference. From each of the two mesh families, a coarse and a fine meshes are selected featuring roughly the same number of elements. For both values of ν, Figure IV.5 depicts the values of the horizontal approximate displacement $u_{h,x}$ along the vertical centerline $x = 1/2$ (solid lines), as well as the values of the vertical approximate displacement $u_{h,y}$ along the horizontal centerline $y = 1/2$ (dashed lines), cf. again Figure IV.2b.

For large values of the Poisson's ratio, the $\mathbb{P}_1^d$ method shows clear signs of numerical locking. For coarse meshes, the approximation is totally irrelevant. It begins to be better as the grid is refined, but still remains rather imprecise for fine grids. As ν tends to 0.5 the errors keep increasing. At the opposite, the CRg-VS method shows very good robustness with repect to locking, on both kinds of meshes.

IV.4.4 Robustness on challenging grids

We here assess the robustness of the discretizations CRg-JS (IV.2)–(IV.14) and CRg-VS (IV.5)–(IV.15) with respect to challenging grid sequences. We consider the heterogeneous test-case of



(a) Coarse meshes

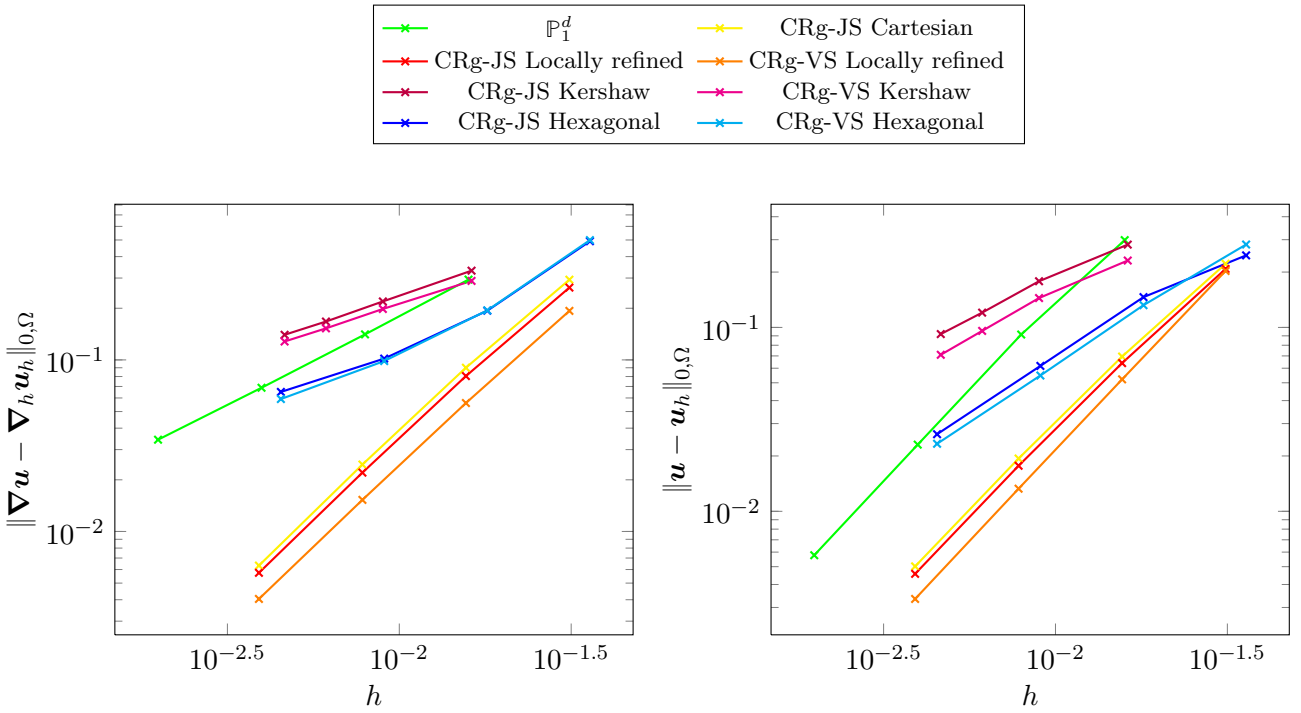
(b) Fine meshes

(c) Coarse meshes

(d) Fine meshes

Figure IV.5: Results for the closed cavity problem on a coarse and a fine mesh extracted from the mesh families of Figure IV.1e and IV.1a. *Top:* $\nu = 0.25$. *Bottom:* $\nu = 0.4999$. *Solid lines:* horizontal displacement $u_{h,x}$ along the vertical centerline. *Dashed lines:* vertical displacement $u_{h,y}$ along the horizontal centerline.

Section IV.4.2 for a heterogeneity contrast $\kappa = 10^6$. The solution is given in (IV.25). We solve the discrete problem on the Cartesian, locally refined Cartesian, Kershaw-type, and hexagonal-dominant mesh sequences of Figures IV.1b, IV.1c, IV.1d, and IV.1f, for both discretizations CRg-JS and CRg-VS, and we give a comparison with the $\mathbb{P}_1^d$ finite element method on the matching triangular mesh sequence of Figure IV.1a. For the Cartesian mesh sequence, only the results for the CRg-JS method are reminded since this case can be found on Figure IV.3. For the discretization CRg-JS, we take a stabilization parameter $\chi = 1$ in (IV.14). Note that the heterogeneity ratio κ is such that the solution (IV.25) is nonzero in the domain Ω_2 (cf. Fig-



(a) H^1 relative error

(b) L^2 relative error

Figure IV.6: Robustness of the discretizations CRg-JS and CRg-VS on challenging grids vs. $\mathbb{P}_1^d$.

ure IV.2a), to which belong most of the refined regions of the locally refined meshes. This test-case is close (on a smaller scale) from the kind of cases that can be encountered in practical reservoir simulation: a heterogeneous problem, that does not match the regularity assumptions of Theorem IV.1, to be solved on a potentially highly skewed locally refined Cartesian grid.

The results of Figure IV.6 exhibit the good behavior of both methods on challenging grids. The $\mathbb{P}_1^d$ benchmark exhibits a first order slope in the H^1 seminorm and a second order slope in the L^2 norm. On the Cartesian and locally refined Cartesian mesh sequences, both methods exhibit a supra-convergent behavior in the H^1 seminorm and a second order convergence in the L^2 norm. For Kershaw-type meshes, the convergence rate is worse than one in the H^1 seminorm, whereas order one is obtained in the L^2 norm. For hexagonal-dominant meshes, first order convergence is obtained in the H^1 seminorm and L^2 norm, with a slight loss of convergence in the H^1 seminorm for fine meshes. Note that for that kind of meshes, where the number of faces per element (and thus the number of lateral pyramidal faces) explodes, the calculation and assembling times for the CRg-JS method are a bit prohibitive for fine meshes. In addition, the conditioning of the matrix deteriorates with the explosion of its stencil, as the jumps penalization couples more and more unknowns.

Chapitre V

Convergence of Euler-Gradient approximations of Biot's consolidation problem

Sommaire

This chapter will form the body of the article [1] (still in preparation). We here tackle the numerical approximation of Biot's consolidation model, which is a particular limit case ($c_0 = 0$) of the poroelasticity problem (II.18) introduced in Section II.2. We design a family of Euler-Gradient discretizations, i.e. implicit Euler in time and Gradient scheme in space, of the weak formulation (II.22). As explained in Section II.2.3, Gradient schemes are a generic framework for the discretization of linear and nonlinear elliptic equations [45, 41, 33]. We consider separate Gradient discretizations for both displacement and pressure, that we introduce in Section V.1. Up to specific assumptions on both sequences (namely coercivity, optimal approximation, limit-conformity and inf-sup stability), we prove in Section V.2 the convergence of this Euler-Gradient family of approximations under minimal regularity assumptions on the solutions and on the data, cf. Section II.2.1. An important feature is that the convergence result is totally independent from the (admissible) values that can possibly be taken by λ or $\underline{\kappa}$. Then, in Section V.3, we give some examples of discretizations falling in this framework, and focus more particularly on a discretization of mechanics based on the generalized Crouzeix–Raviart space introduced in Chapter III, coupled to a Hybrid Finite Volume treatment of pressure. Finally, we provide in Section V.4 some numerical examples in two space dimensions and on general meshes based on the above choice.

Note that the notations used in this chapter voluntarily differ as they enter a more general framework from the ones used in Chapters III and IV.

V.1 Euler-Gradient discretization

We first introduce in this section the sufficient conditions on the Gradient (space) discretizations of pore pressure and displacement to obtain a converging approximation for problem (II.22). Then, we explicit the time-space setting before introducing the discrete problem.

V.1.1 Space discretization

It is of some importance noticing, before going further, that we do not need for the moment to define a notion of (admissible) mesh sequence.

V.1.1.1 Pore pressure

Let us begin by giving the definition of a pressure Gradient discretization for problem (II.22).

Definition V.1 (Pressure Gradient discretization)**.** A pressure Gradient discretization $\mathcal{D}^{\mathrm{p}}$ is defined by $\mathcal{D}^{\mathrm{p}} := (X^{\mathrm{p}}_{\mathcal{D},0}, \Pi^{\mathrm{p}}_{\mathcal{D}}, \boldsymbol{\nabla}^{\mathrm{p}}_{\mathcal{D}})$, where

(i) the *zero-mean* set of discrete unknowns $X^{\mathrm{p}}_{\mathcal{D},0}$ is a finite dimensional vector space on $\mathbb{R}$;

(ii) the linear mapping $\Pi^{\mathrm{p}}_{\mathcal{D}} : X^{\mathrm{p}}_{\mathcal{D},0} \to P_{\mathcal{D},0}$, where $P_{\mathcal{D},0} := P_{\mathcal{D}} \cap L^2_0(\Omega)$ with $P_{\mathcal{D}}$ finite dimensional subspace of $L^2(\Omega)$, is the reconstruction of the approximate function;

(iii) the linear mapping $\boldsymbol{\nabla}^{\mathrm{p}}_{\mathcal{D}} : X^{\mathrm{p}}_{\mathcal{D},0} \to L^2(\Omega)^d$ is the discrete gradient operator. It is chosen such that $\|\boldsymbol{\nabla}^{\mathrm{p}}_{\mathcal{D}} \cdot\|_{0,\Omega}$ is a norm on $X^{\mathrm{p}}_{\mathcal{D},0}$.

We recall that the space $L^2_0(\Omega)$ is defined in (II.12). The term *zero-mean* and the zero subscript in $X^{\mathrm{p}}_{\mathcal{D},0}$ emphasize the fact that the spatial zero-mean condition on the pore pressure (cf. model (II.18)) is strongly enforced through adequate constraints in the set of unknowns.

We now present three sufficient assumptions on sequences of pressure Gradient discretizations, that are typical of the Gradient schemes framework, in order to prove the convergence of our approximation of problem (II.22). Note that when considering nonlinear elliptic problems, a fourth assumption of compactness is often needed. Let first introduce the space

$$\boldsymbol{H}_0(\mathrm{div};\Omega) := \{\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div};\Omega) \mid \gamma_{\boldsymbol{n}}(\boldsymbol{v}) = 0\},$$

where $\gamma_{\boldsymbol{n}}(\boldsymbol{v}) \in H^{-\frac{1}{2}}(\Gamma)$ is the normal trace $\boldsymbol{v}\cdot\boldsymbol{n}_{|\Gamma}$.

Let $(\mathcal{D}^{\mathrm{p}}_m)_{m\in\mathbb{N}}$ be a sequence of pressure Gradient discretizations in the sense of Definition V.1.

Assumption V.1 (Pressure coercivity)**.** *For a given $\mathcal{D}^{\mathrm{p}}$, let define the norm of the linear mapping $\Pi^{\mathrm{p}}_{\mathcal{D}}$ as*

$$C^{\mathrm{p}}_{\mathcal{D}} := \max_{q_{\mathcal{D}} \in X^{\mathrm{p}}_{\mathcal{D},0}\setminus\{0\}} \frac{\|\Pi^{\mathrm{p}}_{\mathcal{D}} q_{\mathcal{D}}\|_{0,\Omega}}{\|\boldsymbol{\nabla}^{\mathrm{p}}_{\mathcal{D}} q_{\mathcal{D}}\|_{0,\Omega}}. \tag{V.1}$$

Then, there exists $C^{\mathrm{p}}_{\mathrm{P}} > 0$ such that $C^{\mathrm{p}}_{\mathcal{D}_m} \leqslant C^{\mathrm{p}}_{\mathrm{P}}$ for all $m \in \mathbb{N}$. The sequence $(\mathcal{D}^{\mathrm{p}}_m)_{m\in\mathbb{N}}$ is said to be coercive*.*

Remark V.1 (Discrete Poincaré inequality)**.** *One can derive from Equation* (V.1) $\|\Pi^{\mathrm{p}}_{\mathcal{D}} q_{\mathcal{D}}\|_{0,\Omega} \leqslant C^{\mathrm{p}}_{\mathcal{D}} \|\boldsymbol{\nabla}^{\mathrm{p}}_{\mathcal{D}} q_{\mathcal{D}}\|_{0,\Omega}$.

The coercivity of the pressure discretization is thus expressed as a uniform Poincaré inequality.

Assumption V.2 (Pressure approximation)**.** *For a given $\mathcal{D}^{\rm p}$, let $S^{\rm p}_{\mathcal{D}} : \overline{H^1}(\Omega) \to \mathbb{R}^+$ defined by*

$$\forall \varphi \in \overline{H^1}(\Omega), \quad S^{\rm p}_{\mathcal{D}}(\varphi) := \min_{q_{\mathcal{D}} \in X^{\rm p}_{\mathcal{D},0}} \left(\|\Pi^{\rm p}_{\mathcal{D}} q_{\mathcal{D}} - \varphi\|_{0,\Omega} + \|\boldsymbol{\nabla}^{\rm p}_{\mathcal{D}} q_{\mathcal{D}} - \boldsymbol{\nabla}\varphi\|_{0,\Omega} \right). \tag{V.2}$$

Then, for all $\varphi \in \overline{H^1}(\Omega)$, $S^{\rm p}_{\mathcal{D}_m}(\varphi) \to 0$ as $m \to +\infty$. The sequence $(\mathcal{D}^{\rm p}_m)_{m\in\mathbb{N}}$ is said to enjoy optimal approximation properties*.*

This property is usually termed as *consistency* in the Gradient schemes literature, cf. [33].

Assumption V.3 (Pressure limit-conformity)**.** *For a given $\mathcal{D}^{\rm p}$, let $W^{\rm p}_{\mathcal{D}} : \boldsymbol{H}_0(\mathrm{div};\Omega) \to \mathbb{R}^+$ defined by*

$$\forall \boldsymbol{\varphi} \in \boldsymbol{H}_0(\mathrm{div};\Omega), \quad W^{\rm p}_{\mathcal{D}}(\boldsymbol{\varphi}) := \max_{q_{\mathcal{D}} \in X^{\rm p}_{\mathcal{D},0}\setminus\{0\}} \frac{1}{\|\kappa^{1/2}\boldsymbol{\nabla}^{\rm p}_{\mathcal{D}} q_{\mathcal{D}}\|_{0,\Omega}} \left| (\kappa^{1/2}\boldsymbol{\nabla}^{\rm p}_{\mathcal{D}} q_{\mathcal{D}}, \boldsymbol{\varphi})_{0,\Omega} + (\Pi^{\rm p}_{\mathcal{D}} q_{\mathcal{D}}, \boldsymbol{\nabla}\cdot(\kappa^{1/2}\boldsymbol{\varphi}))_{0,\Omega} \right|. \tag{V.3}$$

Then, for all $\boldsymbol{\varphi} \in \boldsymbol{H}_0(\mathrm{div};\Omega)$, $W^{\rm p}_{\mathcal{D}_m}(\boldsymbol{\varphi}) \to 0$ as $m \to +\infty$. The sequence $(\mathcal{D}^{\rm p}_m)_{m\in\mathbb{N}}$ is said to be limit-conforming*.*

The limit-conformity property establishes a link between the gradient and the reconstruction operators (in addition of the one established by the discrete Poincaré inequality). It guarantees that the two operators fulfill a continuous Green's formula in the limit. In the case of a conforming finite element approximation, then $W^{\rm p}_{\mathcal{D}}(\boldsymbol{\varphi}) = 0$ for all $\boldsymbol{\varphi} \in \boldsymbol{H}_0(\mathrm{div};\Omega)$. Here, the limit-conformity assumption concerns the operator that we do consider in our model, that is $\kappa^{1/2}\boldsymbol{\nabla}^{\rm p}_{\mathcal{D}}$. We remind the reader that κ is assumed to be such that $\kappa \in W^{1,\infty}(\Omega)$, which gives a sense to $\boldsymbol{\nabla}\cdot(\kappa^{1/2}\boldsymbol{\varphi})$ in $L^2_0(\Omega)$. The additional assumption (II.21) is needed as we will detail in Section V.3 to prove that the HFV [40] method applied to pressure discretization is limit-conforming in the sense of the above definition.

V.1.1.2 Displacement

Let now turn to the definition of a displacement Gradient discretization for problem (II.22).

Definition V.2 (Displacement Gradient discretization)**.** A displacement Gradient discretization $\mathcal{D}^{\rm d}$ is defined by $\mathcal{D}^{\rm d} := (\boldsymbol{X}^{\rm d}_{\mathcal{D},0}, \Pi^{\rm d}_{\mathcal{D}}, \boldsymbol{\nabla}^{\rm d}_{\mathcal{D}})$, where

(i) the *homogeneous* set of discrete unknowns $\boldsymbol{X}^{\rm d}_{\mathcal{D},0}$ is a finite dimensional vector space on $\mathbb{R}^d$;

(ii) the linear mapping $\Pi^{\rm d}_{\mathcal{D}} : \boldsymbol{X}^{\rm d}_{\mathcal{D},0} \to L^2(\Omega)^d$ is the reconstruction of the approximate function;

(iii) the linear mapping $\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}} : \boldsymbol{X}^{\rm d}_{\mathcal{D},0} \to L^2(\Omega)^{d,d}$ is the discrete gradient operator. It is chosen such that $\|\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}} \cdot\|_{0,\Omega}$ is a norm on $\boldsymbol{X}^{\rm d}_{\mathcal{D},0}$.

The term *homogeneous* and the zero subscript in $\boldsymbol{X}^{\rm d}_{\mathcal{D},0}$ emphasize the fact that the homogeneous Dirichlet boundary condition on the displacement is strongly enforced through adequate constraints in the set of unknowns.

Taking inspiration from the previous paragraph, we introduce three equivalent sufficient assumptions on sequences of displacement Gradient discretizations in order to prove the convergence of our approximation of problem (II.22). We here add another assumption, which is not classical in the Gradient schemes literature as saddle-point problems have not yet been studied in that framework, which concerns the displacement-pressure coupling. For that purpose, let us define the discrete divergence operator $\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}}\cdot : \boldsymbol{X}^{\rm d}_{\mathcal{D},0} \to L^2_0(\Omega)$ such that, for all $\boldsymbol{v}_{\mathcal{D}} \in \boldsymbol{X}^{\rm d}_{\mathcal{D},0}$,

$$\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}}\cdot\boldsymbol{v}_{\mathcal{D}} := \mathrm{tr}(\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}}).$$

For a given pressure Gradient discretization $\mathcal{D}^{\rm p}$, we also define $\pi^{\rm p}_{\mathcal{D}} : L^2(\Omega) \to P_{\mathcal{D}}$ as the L^2-orthogonal projector onto $P_{\mathcal{D}}$. Usually, $P_{\mathcal{D}}$ is a broken polynomial space on a spatial discretization (mesh) of the domain. Hence, under classical requirements on the mesh sequence like the ones exposed in Section III.1, we can assume that the L^2-orthogonal projector has optimal approximation properties, in the sense that, for $(\mathcal{D}^{\rm p}_m)_{m\in\mathbb{N}}$ sequence of pressure Gradient discretizations, and for $\varphi \in L^2(\Omega)$,

$$\|\pi^{\rm p}_{\mathcal{D}_m}(\varphi) - \varphi\|_{0,\Omega} \to 0 \quad \text{as} \quad m \to +\infty. \tag{V.4}$$

When restricting $\pi^{\rm p}_{\mathcal{D}}$ to $L^2_0(\Omega)$, then $\pi^{\rm p}_{\mathcal{D}}(\varphi) \in P_{\mathcal{D},0}$ owing to the mean conservation property of the L^2-orthogonal projector onto broken polynomial spaces. Finally, we denote by

$$\underline{\underline{H}}(\mathrm{div};\Omega) := \left\{ \underline{\underline{\boldsymbol{v}}} \in L^2(\Omega)^{d,d} \mid \boldsymbol{\nabla}\cdot\underline{\underline{\boldsymbol{v}}} \in L^2(\Omega)^d \right\}. \tag{V.5}$$

Let $(\mathcal{D}^{\rm d}_m)_{m\in\mathbb{N}}$ be a sequence of displacement Gradient discretizations in the sense of Definition V.2.

Assumption V.4 (Displacement coercivity)**.** *For a given $\mathcal{D}^{\rm d}$, let define the norm of the linear mapping $\Pi^{\rm d}_{\mathcal{D}}$ as*

$$C^{\rm d}_{\mathcal{D}} := \max_{\boldsymbol{v}_{\mathcal{D}} \in \boldsymbol{X}^{\rm d}_{\mathcal{D},0}\setminus\{\boldsymbol{0}\}} \frac{\|\Pi^{\rm d}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}}\|_{0,\Omega}}{\|\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}}\|_{0,\Omega}}. \tag{V.6}$$

Then, there exists $C^{\rm d}_{\rm F} > 0$ such that $C^{\rm d}_{\mathcal{D}_m} \leqslant C^{\rm d}_{\rm F}$ for all $m \in \mathbb{N}$. The sequence $(\mathcal{D}^{\rm d}_m)_{m\in\mathbb{N}}$ is said to be coercive*.*

Remark V.2 (Discrete Friedrichs' inequality)**.** *Equation* (V.6) *directly yields $\|\Pi^{\rm d}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}}\|_{0,\Omega} \leqslant C^{\rm d}_{\mathcal{D}}\|\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}}\|_{0,\Omega}$.*

The coercivity of the discretization is hence defined as a uniform Friedrichs' inequality. Note that we do not assume that a discrete Korn's inequality holds since we consider in (II.22) the pure displacement problem and the naturally coercive bilinear form $\tilde{a}$.

Assumption V.5 (Displacement approximation)**.** *For a given $\mathcal{D}^{\rm d}$, let $S^{\rm d}_{\mathcal{D}} : H^1_0(\Omega)^d \to \mathbb{R}^+$ defined by*

$$\forall \boldsymbol{\varphi} \in H^1_0(\Omega)^d, \quad S^{\rm d}_{\mathcal{D}}(\boldsymbol{\varphi}) := \min_{\boldsymbol{v}_{\mathcal{D}} \in \boldsymbol{X}^{\rm d}_{\mathcal{D},0}} \left(\|\Pi^{\rm d}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}} - \boldsymbol{\varphi}\|_{0,\Omega} + \|\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}} - \boldsymbol{\nabla}\boldsymbol{\varphi}\|_{0,\Omega} \right).$$

Then, for all $\boldsymbol{\varphi} \in H^1_0(\Omega)^d$, $S^{\rm d}_{\mathcal{D}_m}(\boldsymbol{\varphi}) \to 0$ as $m \to +\infty$. The sequence $(\mathcal{D}^{\rm d}_m)_{m\in\mathbb{N}}$ is said to enjoy optimal approximation properties*.*

Assumption V.6 (Displacement limit-conformity)**.** *For a given $\mathcal{D}^{\rm d}$, let $W^{\rm d}_{\mathcal{D}} : \underline{\underline{H}}(\mathrm{div};\Omega) \to \mathbb{R}^+$ defined by*

$$\forall \underline{\underline{\boldsymbol{\varphi}}} \in \underline{\underline{H}}(\mathrm{div};\Omega), \quad W^{\rm d}_{\mathcal{D}}(\underline{\underline{\boldsymbol{\varphi}}}) := \max_{\boldsymbol{v}_{\mathcal{D}} \in \boldsymbol{X}^{\rm d}_{\mathcal{D},0}\setminus\{\boldsymbol{0}\}} \frac{1}{\|\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}}\|_{0,\Omega}} \left| (\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}}, \underline{\underline{\boldsymbol{\varphi}}})_{0,\Omega} + (\Pi^{\rm d}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}}, \boldsymbol{\nabla}\cdot\underline{\underline{\boldsymbol{\varphi}}})_{0,\Omega} \right|. \tag{V.7}$$

Then, for all $\underline{\underline{\boldsymbol{\varphi}}} \in \underline{\underline{H}}(\mathrm{div};\Omega)$, $W^{\rm d}_{\mathcal{D}_m}(\underline{\underline{\boldsymbol{\varphi}}}) \to 0$ as $m \to +\infty$. The sequence $(\mathcal{D}^{\rm d}_m)_{m\in\mathbb{N}}$ is said to be limit-conforming*.*

In order to ensure a stable coupling, we consider the following additional assumption on sequences of displacement/pressure Gradient discretizations. Let $(\mathcal{D}_m)_{m\in\mathbb{N}} := (\mathcal{D}^{\rm d}_m, \mathcal{D}^{\rm p}_m)_{m\in\mathbb{N}}$ be a sequence of displacement/pressure Gradient discretizations in the sense of Definitions V.2 and V.1.

Assumption V.7 (Displacement-pressure coupling)**.** *For all $m \in \mathbb{N}$, there exists an interpolator $\mathcal{I}^{\mathrm{d}}_{\mathcal{D}_m} : H^1_0(\Omega)^d \to \boldsymbol{X}^{\mathrm{d}}_{\mathcal{D}_m,0}$ such that*

$$\forall \boldsymbol{\varphi} \in H^1_0(\Omega)^d, \quad \pi^{\mathrm{p}}_{\mathcal{D}_m}(\boldsymbol{\nabla}^{\mathrm{d}}_{\mathcal{D}_m} \cdot \mathcal{I}^{\mathrm{d}}_{\mathcal{D}_m}(\boldsymbol{\varphi})) = \pi^{\mathrm{p}}_{\mathcal{D}_m}(\boldsymbol{\nabla} \cdot \boldsymbol{\varphi}), \tag{V.8}$$

and there exists $C_{\mathrm{S}} > 0$, independent of m, such that, for all $m \in \mathbb{N}$,

$$\forall \boldsymbol{\varphi} \in H^1_0(\Omega)^d, \quad \|\boldsymbol{\nabla}^{\mathrm{d}}_{\mathcal{D}_m} \mathcal{I}^{\mathrm{d}}_{\mathcal{D}_m}(\boldsymbol{\varphi})\|_{0,\Omega} \leqslant C_{\mathrm{S}} \|\boldsymbol{\nabla}\boldsymbol{\varphi}\|_{0,\Omega}. \tag{V.9}$$

The sequence $(\mathcal{D}_m)_{m\in\mathbb{N}}$ is said to possess a sequence of Fortin operators *in the sense of Definition II.1.*

This assumption is also adapted to the discretization of the (possibly) quasi-incompressible linear elasticity model, see Remark V.4. We make the classical further assumption that the sequence of Fortin operators enjoys *optimal approximation properties*, in the sense that

$$\forall \boldsymbol{\varphi} \in H^1_0(\Omega)^d, \quad \left(\|\Pi^{\mathrm{d}}_{\mathcal{D}_m} \mathcal{I}^{\mathrm{d}}_{\mathcal{D}_m}(\boldsymbol{\varphi}) - \boldsymbol{\varphi}\|_{0,\Omega} + \|\boldsymbol{\nabla}^{\mathrm{d}}_{\mathcal{D}_m} \mathcal{I}^{\mathrm{d}}_{\mathcal{D}_m}(\boldsymbol{\varphi}) - \boldsymbol{\nabla}\boldsymbol{\varphi}\|_{0,\Omega}\right) \to 0 \quad \text{as } m \to +\infty. \tag{V.10}$$

This (stronger) assumption (V.10) replaces Assumption V.5.

Remark V.3 (Discrete inf-sup condition)**.** *For a given $\mathcal{D} := (\mathcal{D}^{\mathrm{d}}, \mathcal{D}^{\mathrm{p}})$, let introduce the displacement-pressure coupling bilinear form on $\boldsymbol{X}^{\mathrm{d}}_{\mathcal{D},0} \times X^{\mathrm{p}}_{\mathcal{D},0}$ given by*

$$b_{\mathcal{D}}(\boldsymbol{v}_{\mathcal{D}}, q_{\mathcal{D}}) := -(\pi^{\mathrm{p}}_{\mathcal{D}}(\boldsymbol{\nabla}^{\mathrm{d}}_{\mathcal{D}} \cdot \boldsymbol{v}_{\mathcal{D}}), \Pi^{\mathrm{p}}_{\mathcal{D}} q_{\mathcal{D}})_{0,\Omega}.$$

Then, Assumption V.7 is equivalent to assuming that a discrete inf-sup condition holds for the sequence of coupling bilinear forms $(b_{\mathcal{D}_m})_{m\in\mathbb{N}}$, with a multiplicative constant independent of m.

This property would ensure in the context of a Stokes problem for example that the discretization is inf-sup stable. In the context of Biot's consolidation problem, as we already explained in Section II.2.2, this property may actually be insufficient to properly speak about inf-sup stability in the finite element sense. All depends on whether the reconstruction is gradient-based (like in finite element methods) or not (for example piecewise constant, like in most finite volume methods). To clarify the idea, let consider for the pressure discretization a HFV method, that is to say a piecewise constant reconstruction based on cell unknowns, and a gradient which is given by (III.8)–(III.9) and which depends on both face and cell unknowns (the subgrid correction establishes the link between cell and face unknowns). We assume that the components of the discrete displacement belong to the generalized Crouzeix–Raviart space introduced in Chapter III. Then, Lemma III.5 and Corollary III.1 ensure that an inf-sup condition holds for this pair of displacement/pressure spaces (since the pressure reconstruction is piecewise constant). This inf-sup condition gives a stability estimate on the approximate reconstructed pressure. This estimate turns out to give a real control on pressure in the context of a Stokes problem since the pressure only lives in $L^2_0(\Omega)$ and is fully discretized using cell unknowns. However, in the context of Biot's consolidation model, we see that this estimate only gives control on pressure cell unknowns, without any direct information on face unknowns. From a continuous point of view, in small times or in poorly permeable regions, the pore pressure is quasi-$L^2_0(\Omega)$ since the diffusion term has an almost vanishing contribution, but boundary conditions are applied to the pressure as soon as $t > 0$, hence giving to this quasi-$L^2_0(\Omega)$ object a $\overline{H^1}(\Omega)$ dimension (dimension that is fully acquired as time goes on since the Darcean diffusion term takes more and more importance). From a discrete point of view, this results in spurious spatial oscillations of the pore pressure approximation in early times or in poorly permeable regions, due to a lack of control on the pressure gradient of this (forced to be) $\overline{H^1}(\Omega)$ object.

A possible remedy is to add artificial diffusion as proposed by Aguilar et al. [2] (they add a discrete stabilization term in the flow equation whose continuous equivalent would be $\partial_t(\beta\triangle p)$, with $\beta > 0$ a user-dependent parameter depending on the meshsize). This gives a $L^\infty(0,T;\overline{H^1}(\Omega))$ control of the pressure that is independent from $\underline{\kappa}^{-1}$, but this method has the drawback of denaturing the physical model. Another remedy hinges on the verification of an inf-sup condition. This gives a $L^\infty(0,T;L^2_0(\Omega))$ control on the pressure which does not depend on $\underline{\kappa}^{-1}$. This method has been shown to be efficient in the finite element context, see Murad et al. [62]. A $L^\infty(0,T;L^2_0(\Omega))$ estimate on the reconstructed pressure, when this latter is gradient-based as in finite element methods, actually enables to have a kind of control on the gradient and thus reduces spurious oscillations in early times or in poorly permeable regions. If we apply it to our generalized Crouzeix–Raviart/HFV discretization, the inf-sup condition gives an estimate on the pressure reconstruction, which only concerns cell unknowns. This estimate obviously has an impact on the control of face unknowns since these latter are linked to cell unknowns through the subgrid correction of the gradient operator but this impact is (after numerical assessment) less important than for inf-sup stable pairs of finite elements where the pressure reconstruction is gradient-based. We obtain a method which is more stable than for an unstable pair of finite elements, but less stable than for a stable one. We then cannot really speak in that case of inf-sup stability in the finite element sense but practically this still contributes to reduce spurious oscillations of the pressure approximation. Whatever it be, from a theoretical point of view, this stability property allows to prove the strong convergence in $L^2(0,T;L^2_0(\Omega))$ of the approximate pressure reconstruction independently of the (admissible) values of $\underline{\kappa}$, wich seems difficult without.

V.1.2 Time-space discretization

As we study an elliptic-parabolic model, we must introduce a time discretization for problem (II.22). We consider a first order implicit (also known as backward) Euler discretization in time. We could as well consider a θ-scheme (which reduces to the implicit scheme for $\theta = 1$) as in [33].

Definition V.3 (Time discretization)**.** A time discretization of the interval $(0,T]$ is defined by $N \in \mathbb{N}^*$ and $\delta := (t^{(n)})_{n\in[\![0,N]\!]}$ such that

(i) $t^{(0)} = 0 < t^{(1)} \ldots < t^{(N)} = T$;

(ii) for $n \in [\![0, N-1]\!]$, we set $\delta t^{(n+1/2)} := t^{(n+1)} - t^{(n)}$ and $\delta t^{\mathrm{M}} := \max_{n\in[\![0,N-1]\!]} \delta t^{(n+1/2)}$.

Let now introduce a sufficient assumption on sequences of time discretizations in order to prove the convergence of our approximation of problem (II.22). For that purpose, let $(\delta_m)_{m\in\mathbb{N}}$ be a sequence of time discretizations in the sense of Definition V.3.

Assumption V.8 (Time consistency)**.** *The following three conditions are fulfilled:*

(i) $\delta t^{\mathrm{M}}_m \to 0$ *as* $m \to +\infty$ *(which implies* $N_m \to +\infty$*);*

(ii) $\delta t^{\mathrm{M}}_m < 1$ *for all* $m \in \mathbb{N}$*;*

(iii) there exists $C_t > 0$*, independent of* m*, such that for all* $m \in \mathbb{N}$*,*

$$\forall\, n \in [\![1, N_m - 1]\!], \qquad \frac{|\delta t_m^{(n+1/2)} - \delta t_m^{(n-1/2)}|}{\delta t_m^{(n-1/2)}} \leqslant C_t.$$

The sequence $(\delta_m)_{m\in\mathbb{N}}$ *is said to be* consistent*.*

The third assumption (iii) quantifies the relative variations of time step and is only needed for a theoretical purpose, see (iv) in the proof of Theorem V.1. In the following, for sequences of time discretizations, the dependence of N_m on m will be understood without denoting it in subscript (then N_m will always be denoted N).

We can now introduce the time-space setting.

Definition V.4 (Euler-Gradient (time-space) discretization)**.** An Euler-Gradient (time-space) discretization is given by the couple $(\delta, \mathcal{D})$, where

(i) δ is a time discretization in the sense of Definition V.3;

(ii) $\mathcal{D} := (\mathcal{D}^{\mathrm{d}}, \mathcal{D}^{\mathrm{p}})$ is a displacement/pressure Gradient discretization in the sense of Definitions V.2 and V.1.

In the following, we will consider sequences $(\delta_m, \mathcal{D}_m)_{m\in\mathbb{N}}$ with $\mathcal{D}_m := (\mathcal{D}_m^{\mathrm{d}}, \mathcal{D}_m^{\mathrm{p}})$ of Euler-Gradient (time-space) discretizations.

V.1.3 Discrete problem

We consider an Euler-Gradient approximation of problem (II.22).

For that purpose, let $(\delta, \mathcal{D})$ with $\mathcal{D} := (\mathcal{D}^{\mathrm{d}}, \mathcal{D}^{\mathrm{p}})$ denote an Euler-Gradient (time-space) discretization in the sense of Definition V.4. Let choose $\boldsymbol{u}_{\mathcal{D}}^{(0)} \in \boldsymbol{X}_{\mathcal{D},0}^{\mathrm{d}}$ satisfying (V.11c). The discrete problem reads:
Find $(\boldsymbol{u}_{\mathcal{D}}^{(n)} \in \boldsymbol{X}_{\mathcal{D},0}^{\mathrm{d}}, p_{\mathcal{D}}^{(n)} \in X_{\mathcal{D},0}^{\mathrm{p}})_{n\in[\![1,N]\!]}$ such that, for all $n \in [\![0, N-1]\!]$,

$$\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)}, \boldsymbol{v}_{\mathcal{D}}) + b_{\mathcal{D}}(\boldsymbol{v}_{\mathcal{D}}, p_{\mathcal{D}}^{(n+1)}) = (\boldsymbol{f}^{(n+1)}, \Pi_{\mathcal{D}}^{\mathrm{d}}\boldsymbol{v}_{\mathcal{D}})_{0,\Omega} \qquad \forall \boldsymbol{v}_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}^{\mathrm{d}}, \tag{V.11a}$$

$$-b_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)} - \boldsymbol{u}_{\mathcal{D}}^{(n)}, q_{\mathcal{D}}) + \delta t^{(n+1/2)}\, c_{\mathcal{D}}(p_{\mathcal{D}}^{(n+1)}, q_{\mathcal{D}}) = \delta t^{(n+1/2)}\, (h^{(n+1)}, \Pi_{\mathcal{D}}^{\mathrm{p}} q_{\mathcal{D}})_{0,\Omega} \qquad \forall q_{\mathcal{D}} \in X_{\mathcal{D},0}^{\mathrm{p}}, \tag{V.11b}$$

$$(\boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{d}} \cdot \boldsymbol{u}_{\mathcal{D}}^{(0)}, \Pi_{\mathcal{D}}^{\mathrm{p}} q_{\mathcal{D}})_{0,\Omega} = (\beta, \Pi_{\mathcal{D}}^{\mathrm{p}} q_{\mathcal{D}})_{0,\Omega} \qquad \forall q_{\mathcal{D}} \in X_{\mathcal{D},0}^{\mathrm{p}}, \tag{V.11c}$$

where $\tilde{a}_{\mathcal{D}}(\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{v}_{\mathcal{D}}) := \mu(\boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{d}}\boldsymbol{w}_{\mathcal{D}}, \boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{d}}\boldsymbol{v}_{\mathcal{D}})_{0,\Omega} + (\mu+\lambda)(\pi_{\mathcal{D}}^{\mathrm{p}}(\boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{d}}\cdot\boldsymbol{w}_{\mathcal{D}}), \pi_{\mathcal{D}}^{\mathrm{p}}(\boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{d}}\cdot\boldsymbol{v}_{\mathcal{D}}))_{0,\Omega}$, $b_{\mathcal{D}}(\boldsymbol{v}_{\mathcal{D}}, q_{\mathcal{D}}) := -(\pi_{\mathcal{D}}^{\mathrm{p}}(\boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{d}}\cdot\boldsymbol{v}_{\mathcal{D}}), \Pi_{\mathcal{D}}^{\mathrm{p}} q_{\mathcal{D}})_{0,\Omega} = -(\boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{d}}\cdot\boldsymbol{v}_{\mathcal{D}}, \Pi_{\mathcal{D}}^{\mathrm{p}} q_{\mathcal{D}})_{0,\Omega}$, and $c_{\mathcal{D}}(r_{\mathcal{D}}, q_{\mathcal{D}}) := (\kappa \boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{p}} r_{\mathcal{D}}, \boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{p}} q_{\mathcal{D}})_{0,\Omega}$.
The discrete source terms are defined as $L^2(\Omega)^d \ni \boldsymbol{f}^{(n+1)} := \boldsymbol{f}(t^{(n+1)})$, and $L_0^2(\Omega) \ni h^{(n+1)} := \frac{1}{\delta t^{(n+1/2)}} \int_{t^{(n)}}^{t^{(n+1)}} h(t)\,\mathrm{d}t$.

Remark V.4 (Quasi-incompressible materials)**.** *In order to deal with quasi-incompressible materials, i.e. $\lambda \to +\infty$, we perform static condensation on the divergence operator in $\tilde{a}_{\mathcal{D}}$, see Section II.1.2.2. It here consists in projecting the divergence operator onto the discrete (pressure) space $P_{\mathcal{D},0}$, which satisfies under Assumption V.7 a (uniform) inf-sup condition when combined with the displacement approximation space, cf. Remark V.3. Hence, deal with a sequence of displacement/pressure Gradient discretizations that possesses a sequence of Fortin operators ensures the robustness of the quasi-incompressible elasticity approximation with respect to numerical locking.*

V.2 Convergence to minimal regularity solutions

We now study the convergence of the approximation scheme (V.11), under the sufficient spatial discretization assumptions exposed in Section V.1.1 and time discretization condition of

Assumption V.8, and under minimal regularity assumptions on the continuous solution and on the data.

We begin by giving some a priori estimates on the discrete solutions, before establishing, in a second time, the convergence result.

V.2.1 A priori estimates

Let $(\delta, \mathcal{D})$ with $\mathcal{D} := (\mathcal{D}^{\mathrm{d}}, \mathcal{D}^{\mathrm{p}})$ be a given Euler-Gradient (time-space) discretization. We denote by $(\boldsymbol{u}_{\mathcal{D}}^{\delta}, p_{\mathcal{D}}^{\delta}) \in \mathbb{P}_d^0\left(\delta; \boldsymbol{X}_{\mathcal{D},0}^{\mathrm{d}}\right) \times \mathbb{P}_d^0\left(\delta; X_{\mathcal{D},0}^{\mathrm{p}}\right)$ the piecewise constant in time solutions of (V.11), with values in $\boldsymbol{X}_{\mathcal{D},0}^{\mathrm{d}}$ and $X_{\mathcal{D},0}^{\mathrm{p}}$ respectively, such that, for $n \in [\![0, N-1]\!]$,

$$\boldsymbol{u}_{\mathcal{D}|\,(t^{(n)},t^{(n+1)}]}^{\delta} := \boldsymbol{u}_{\mathcal{D}}^{(n+1)} \in \boldsymbol{X}_{\mathcal{D},0}^{\mathrm{d}}, \qquad p_{\mathcal{D}|\,(t^{(n)},t^{(n+1)}]}^{\delta} := p_{\mathcal{D}}^{(n+1)} \in X_{\mathcal{D},0}^{\mathrm{p}}.$$

Assumption V.9 (Choice of $\boldsymbol{u}_{\mathcal{D}}^{(0)}$)**.** *The initial discrete displacement $\boldsymbol{u}_{\mathcal{D}}^{(0)}$ satisfies: there exists $C_{\mathrm{I}} > 0$, independent of $\mathcal{D}$, such that*

$$\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(0)}, \boldsymbol{u}_{\mathcal{D}}^{(0)}) \leqslant C_{\mathrm{I}}\, \tilde{a}(\boldsymbol{u}^{(0)}, \boldsymbol{u}^{(0)}),$$

where $\boldsymbol{u}^{(0)} \in H_0^1(\Omega)^d$ is defined in Remark II.4.

This assumption will be verified later. In order to derive the a priori estimates, we first recall the following version of Grönwall's inequality, cf. Quarteroni and Valli [70, p. 14] or Heywood and Rannacher [51, p. 369].

Lemma V.1 (Discrete Grönwall's inequality)**.** *Let $N \in \mathbb{N}^*$. Let k and B denote two positive real numbers, and let $(a_n)_{n\geqslant 1}$, $(b_n)_{n\geqslant 1}$, and $(\delta_n)_{n\geqslant 1}$ denote three sequences of nonnegative real numbers such that*

$$a_N + k \sum_{n=1}^{N} b_n \leqslant k \sum_{n=1}^{N} \delta_n a_n + B.$$

Then, if $k\delta_n < 1$ for all $n \in [\![1, N]\!]$, there holds

$$a_N + k \sum_{n=1}^{N} b_n \leqslant B \exp\left(k \sum_{n=1}^{N} \frac{\delta_n}{1 - k\delta_n} \right).$$

In order to treat the cases of low permeability regions and quasi-incompressible materials, we pay attention, in the a priori estimates we derive, to their dependencies with respect to $\underline{\kappa}$ and λ. We recall that, for obvious physical reasons, either $\underline{\kappa}$ may tend to zero (presence of poorly permeable regions), either λ may tend to infinity (quasi-incompressible material).

Lemma V.2 (Discrete $L^\infty(0,T; H_0^1(\Omega)^d)$ displacement estimate, $L^2(0,T; \overline{H^1}(\Omega))$, $L^\infty(0,T; L_0^2(\Omega))$ pressure estimates)**.** *Let $(\delta, \mathcal{D})$ with $\mathcal{D} := (\mathcal{D}^{\mathrm{d}}, \mathcal{D}^{\mathrm{p}})$ be an Euler-Gradient (time-space) discretization in the sense of Definition V.4. We assume that δ is consistent in the sense that it satisfies* (ii) *in Assumption V.8, and we assume that $\mathcal{D}$ admits a Fortin operator in the sense of Assumption V.7. Let $(\boldsymbol{u}_{\mathcal{D}}^{\delta}, p_{\mathcal{D}}^{\delta})$ be a solution of* (V.11)*. Under the regularity assumptions on the data assumed in Section II.2.1, there holds*

- $\mu \|\boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{d}} \boldsymbol{u}_{\mathcal{D}}^{\delta}\|^2_{L^\infty(0,T;L^2(\Omega)^{d,d})} \leqslant \max_{t\in(0,T]} \tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{\delta}(t), \boldsymbol{u}_{\mathcal{D}}^{\delta}(t)) \leqslant C(C_{\mathcal{D}}^{\mathrm{d}}, C_{\mathcal{D}}^{\mathrm{p}})$,
- $\|\kappa^{1/2} \boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{\delta}\|^2_{L^2(0,T;L^2(\Omega)^d)} = \int_0^T c_{\mathcal{D}}(p_{\mathcal{D}}^{\delta}(t), p_{\mathcal{D}}^{\delta}(t))\,\mathrm{d}t \leqslant C(C_{\mathcal{D}}^{\mathrm{d}}, C_{\mathcal{D}}^{\mathrm{p}})$,

- $\|\Pi_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{\delta}\|_{L^\infty(0,T;L_0^2(\Omega))}^2 \leqslant C(C_{\mathcal{D}}^{\mathrm{d}}, \lambda)$,

where $C(\cdot,\dots)$ denotes a generic constant, depending on the data, whose dependencies with repect to possibly unbounded quantities are all precised in argument.

Proof. (i) Let $n \in [\![1, N]\!]$. Accounting for the fact that $\Pi_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{(n)} \in P_{\mathcal{D},0} \subset L_0^2(\Omega)$, and owing to the surjectivity of the divergence operator from $H_0^1(\Omega)^d$ to $L_0^2(\Omega)$, cf. Lemma II.3, there exists $\boldsymbol{v}_{\mathrm{N}}^{(n)} \in H_0^1(\Omega)^d$ such that

$$\Pi_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{(n)} = \boldsymbol{\nabla}\cdot \boldsymbol{v}_{\mathrm{N}}^{(n)}, \qquad \text{with } \|\boldsymbol{\nabla} \boldsymbol{v}_{\mathrm{N}}^{(n)}\|_{0,\Omega} \leqslant C_{\mathrm{N}} \|\Pi_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{(n)}\|_{0,\Omega}, \tag{V.12}$$

where $C_{\mathrm{N}} > 0$ is defined in Lemma II.3 and does not depend on $\mathcal{D}$ nor n. The existence of a Fortin operator $\mathcal{I}_{\mathcal{D}}^{\mathrm{d}}$ (cf. Assumption V.7) allows us to infer that, letting $\boldsymbol{X}_{\mathcal{D},0}^{\mathrm{d}} \ni \boldsymbol{v}_{\mathrm{N},\mathcal{D}}^{(n)} := \mathcal{I}_{\mathcal{D}}^{\mathrm{d}}(\boldsymbol{v}_{\mathrm{N}}^{(n)})$,

$$\forall q_{\mathcal{D}} \in X_{\mathcal{D},0}^{\mathrm{p}}, \quad b_{\mathcal{D}}(\boldsymbol{v}_{\mathrm{N}}^{(n)}, q_{\mathcal{D}}) = b_{\mathcal{D}}(\boldsymbol{v}_{\mathrm{N},\mathcal{D}}^{(n)}, q_{\mathcal{D}}), \qquad \text{with } \|\boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{d}} \boldsymbol{v}_{\mathrm{N},\mathcal{D}}^{(n)}\|_{0,\Omega} \leqslant C_{\mathrm{S}} \|\boldsymbol{\nabla} \boldsymbol{v}_{\mathrm{N}}^{(n)}\|_{0,\Omega}, \tag{V.13}$$

where $C_{\mathrm{S}} > 0$ is defined in (V.9) and does not depend on $\mathcal{D}$ nor n. Taking $q_{\mathcal{D}} = p_{\mathcal{D}}^{(n)}$ and using (V.13), (V.12), and (V.11a), we get

$$\|\Pi_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{(n)}\|_{0,\Omega}^2 = -b_{\mathcal{D}}(\boldsymbol{v}_{\mathrm{N},\mathcal{D}}^{(n)}, p_{\mathcal{D}}^{(n)}) = \tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n)}, \boldsymbol{v}_{\mathrm{N},\mathcal{D}}^{(n)}) - (\boldsymbol{f}^{(n)}, \Pi_{\mathcal{D}}^{\mathrm{d}} \boldsymbol{v}_{\mathrm{N},\mathcal{D}}^{(n)})_{0,\Omega}. \tag{V.14}$$

To estimate the first term of the right-hand side, we successively use Cauchy-Schwarz inequality ($\tilde{a}_{\mathcal{D}}$ is a symmetric positive definite bilinear form), the boundedness of $\tilde{a}_{\mathcal{D}}$, (V.13), and (V.12), to infer

$$\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n)}, \boldsymbol{v}_{\mathrm{N},\mathcal{D}}^{(n)}) \leqslant C_{\mathrm{B}}(\mu^{1/2}, \lambda^{1/2}) C_{\mathrm{S}} C_{\mathrm{N}}\, \tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n)}, \boldsymbol{u}_{\mathcal{D}}^{(n)})^{1/2} \|\Pi_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{(n)}\|_{0,\Omega}, \tag{V.15}$$

where $C_{\mathrm{B}}(\mu^{1/2}, \lambda^{1/2})$ is the boundedness constant only depending on μ and λ. Then, using successively the Cauchy-Schwarz inequality, the discrete Friedrichs' inequality of Remark V.2, (V.13), and (V.12), we infer

$$(\boldsymbol{f}^{(n)}, \Pi_{\mathcal{D}}^{\mathrm{d}} \boldsymbol{v}_{\mathrm{N},\mathcal{D}}^{(n)})_{0,\Omega} \leqslant C_{\mathcal{D}}^{\mathrm{d}} C_{\mathrm{S}} C_{\mathrm{N}} \|\boldsymbol{f}^{(n)}\|_{0,\Omega} \|\Pi_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{(n)}\|_{0,\Omega}. \tag{V.16}$$

Finally, gathering up (V.14), (V.15), and (V.16), we obtain

$$\|\Pi_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{(n)}\|_{0,\Omega} \leqslant C_{\mathrm{S}} C_{\mathrm{N}} \left(C_{\mathrm{B}}(\mu^{1/2}, \lambda^{1/2})\, \tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n)}, \boldsymbol{u}_{\mathcal{D}}^{(n)})^{1/2} + C_{\mathcal{D}}^{\mathrm{d}} \|\boldsymbol{f}^{(n)}\|_{0,\Omega} \right). \tag{V.17}$$

(ii) Testing (V.11a) with $\boldsymbol{v}_{\mathcal{D}} = \boldsymbol{u}_{\mathcal{D}}^{(n+1)} - \boldsymbol{u}_{\mathcal{D}}^{(n)}$ and summing between 0 and $N-1$ yields

$$\frac{1}{2}\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(N)}, \boldsymbol{u}_{\mathcal{D}}^{(N)}) - \frac{1}{2}\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(0)}, \boldsymbol{u}_{\mathcal{D}}^{(0)}) + \sum_{n=0}^{N-1} b_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)} - \boldsymbol{u}_{\mathcal{D}}^{(n)}, p_{\mathcal{D}}^{(n+1)}) \leqslant \sum_{n=0}^{N-1} (\boldsymbol{f}^{(n+1)}, \Pi_{\mathcal{D}}^{\mathrm{d}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)} - \boldsymbol{u}_{\mathcal{D}}^{(n)}))_{0,\Omega}, \tag{V.18}$$

where we have used the fact that

$$2\,\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)}, \boldsymbol{u}_{\mathcal{D}}^{(n+1)} - \boldsymbol{u}_{\mathcal{D}}^{(n)}) = \tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)} - \boldsymbol{u}_{\mathcal{D}}^{(n)}, \boldsymbol{u}_{\mathcal{D}}^{(n+1)} - \boldsymbol{u}_{\mathcal{D}}^{(n)}) + \tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)}, \boldsymbol{u}_{\mathcal{D}}^{(n+1)}) - \tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n)}, \boldsymbol{u}_{\mathcal{D}}^{(n)}),$$

to infer

$$\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)}, \boldsymbol{u}_{\mathcal{D}}^{(n+1)} - \boldsymbol{u}_{\mathcal{D}}^{(n)}) \geqslant \frac{1}{2}\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)}, \boldsymbol{u}_{\mathcal{D}}^{(n+1)}) - \frac{1}{2}\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n)}, \boldsymbol{u}_{\mathcal{D}}^{(n)}).$$

Letting now $q_{\mathcal{D}} = p_{\mathcal{D}}^{(n+1)}$ in (V.11b), and summing between 0 and $N-1$ leads

$$-\sum_{n=0}^{N-1} b_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)}-\boldsymbol{u}_{\mathcal{D}}^{(n)}, p_{\mathcal{D}}^{(n+1)})+\sum_{n=0}^{N-1} \delta t^{(n+1/2)}\, c_{\mathcal{D}}(p_{\mathcal{D}}^{(n+1)}, p_{\mathcal{D}}^{(n+1)}) = \sum_{n=0}^{N-1} \delta t^{(n+1/2)}\, (h^{(n+1)}, \Pi_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{(n+1)})_{0,\Omega}. \tag{V.19}$$

Summing (V.18) and (V.19) yields

$$\frac{1}{2}\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(N)}, \boldsymbol{u}_{\mathcal{D}}^{(N)}) + \sum_{n=0}^{N-1} \delta t^{(n+1/2)}\, c_{\mathcal{D}}(p_{\mathcal{D}}^{(n+1)}, p_{\mathcal{D}}^{(n+1)}) \leqslant \mathfrak{R}^{(N)}, \tag{V.20}$$

where

$$\mathfrak{R}^{(N)} := \frac{1}{2}\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(0)}, \boldsymbol{u}_{\mathcal{D}}^{(0)})+\sum_{n=0}^{N-1} (\boldsymbol{f}^{(n+1)}, \Pi_{\mathcal{D}}^{\mathrm{d}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)} - \boldsymbol{u}_{\mathcal{D}}^{(n)}))_{0,\Omega}+\sum_{n=0}^{N-1} \delta t^{(n+1/2)}\, (h^{(n+1)}, \Pi_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{(n+1)})_{0,\Omega}. \tag{V.21}$$

Let denote respectively by $\mathfrak{T}_1^{(N)}$, $\mathfrak{T}_2^{(N)}$, and $\mathfrak{T}_3^{(N)}$ the three terms in $\mathfrak{R}^{(N)}$. By discrete integration by parts, $\mathfrak{T}_2^{(N)}$ can be rewritten

$$\mathfrak{T}_2^{(N)} = (\boldsymbol{f}^{(N)}, \Pi_{\mathcal{D}}^{\mathrm{d}} \boldsymbol{u}_{\mathcal{D}}^{(N)})_{0,\Omega} - (\boldsymbol{f}^{(0)}, \Pi_{\mathcal{D}}^{\mathrm{d}} \boldsymbol{u}_{\mathcal{D}}^{(0)})_{0,\Omega} - \sum_{n=0}^{N-1} (\boldsymbol{f}^{(n+1)} - \boldsymbol{f}^{(n)}, \Pi_{\mathcal{D}}^{\mathrm{d}} \boldsymbol{u}_{\mathcal{D}}^{(n)})_{0,\Omega}. \tag{V.22}$$

Here again, we denote by $\mathfrak{T}_{2,1}^{(N)}$, $\mathfrak{T}_{2,2}^{(N)}$, and $\mathfrak{T}_{2,3}^{(N)}$ the three terms in $\mathfrak{T}_2^{(N)}$. Let estimate these different terms. According to Assumption V.9,

$$\mathfrak{T}_1^{(N)} \leqslant \frac{1}{2} C_{\mathrm{I}}\, \tilde{a}(\boldsymbol{u}^{(0)}, \boldsymbol{u}^{(0)}). \tag{V.23}$$

Using Cauchy-Schwarz inequality, Remark V.2, the fact that $\mu \|\boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{d}} \boldsymbol{v}_{\mathcal{D}}\|_{0,\Omega}^2 \leqslant \tilde{a}_{\mathcal{D}}(\boldsymbol{v}_{\mathcal{D}}, \boldsymbol{v}_{\mathcal{D}})$ for all $\boldsymbol{v}_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}^{\mathrm{d}}$, Assumption V.9, and Young's inequality, we infer

$$\mathfrak{T}_{2,2}^{(N)} \leqslant \frac{1}{2}\mu^{-1}(C_{\mathcal{D}}^{\mathrm{d}})^2 \|\boldsymbol{f}\|_{L^\infty(0,T;L^2(\Omega)^d)}^2 + \frac{1}{2} C_{\mathrm{I}}\, \tilde{a}(\boldsymbol{u}^{(0)}, \boldsymbol{u}^{(0)}). \tag{V.24}$$

Using the same first three arguments, and Young's inequality with $\varepsilon > 0$, we get

$$\mathfrak{T}_{2,1}^{(N)} \leqslant \frac{\varepsilon}{2}\, \tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(N)}, \boldsymbol{u}_{\mathcal{D}}^{(N)}) + \frac{1}{2\varepsilon}\mu^{-1}(C_{\mathcal{D}}^{\mathrm{d}})^2 \|\boldsymbol{f}\|_{L^\infty(0,T;L^2(\Omega)^d)}^2. \tag{V.25}$$

As far as $\mathfrak{T}_{2,3}^{(N)}$ is concerned, two applications of Cauchy-Schwarz inequality and some algebraic manipulations first give

$$\mathfrak{T}_{2,3}^{(N)} \leqslant \|\boldsymbol{f}'\|_{L^2(0,T;L^2(\Omega)^d)} \left(\delta t^{(1/2)} \|\Pi_{\mathcal{D}}^{\mathrm{d}} \boldsymbol{u}_{\mathcal{D}}^{(0)}\|_{0,\Omega}^2\right)^{1/2} + \|\boldsymbol{f}'\|_{L^2(0,T;L^2(\Omega)^d)} \left(\sum_{n=0}^{N-1} \delta t^{(n+1/2)} \|\Pi_{\mathcal{D}}^{\mathrm{d}} \boldsymbol{u}_{\mathcal{D}}^{(n+1)}\|_{0,\Omega}^2\right)^{1/2}.$$

Finally, the same arguments as for the proof of (V.24), and Young's inequality with $\chi > 0$, yield

$$\mathfrak{T}_{2,3}^{(N)} \leqslant \frac{\chi}{2} \sum_{n=0}^{N-1} \delta t^{(n+1/2)} \tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)}, \boldsymbol{u}_{\mathcal{D}}^{(n+1)}) + \frac{\chi+1}{2\chi}\mu^{-1}(C_{\mathcal{D}}^{\mathrm{d}})^2 \|\boldsymbol{f}'\|_{L^2(0,T;L^2(\Omega)^d)}^2 + \frac{T}{2} C_{\mathrm{I}}\, \tilde{a}(\boldsymbol{u}^{(0)}, \boldsymbol{u}^{(0)}). \tag{V.26}$$

To estimate $\mathfrak{T}_3^{(N)}$, two applications of Cauchy-Schwarz inequality first give

$$\mathfrak{T}_3^{(N)} \leqslant \|h\|_{L^2(0,T;L_0^2(\Omega))} \left(\sum_{n=0}^{N-1} \delta t^{(n+1/2)} \|\Pi_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{(n+1)}\|_{0,\Omega}^2\right)^{1/2}.$$

We establish two different estimates for $\mathfrak{T}_3^{(N)}$, depending on whether λ may take unboundedly large values (quasi-incompressible material) or $\underline{\kappa}$ may tend to zero (presence of poorly permeable regions). We recall that both cases cannot occur simultaneously.
(a) *$\underline{\kappa}$ is bounded away from zero:* Using the discrete Poincaré inequality of Remark V.1, the fact that $\underline{\kappa}\|\boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{p}} q_{\mathcal{D}}\|_{0,\Omega}^2 \leqslant c_{\mathcal{D}}(q_{\mathcal{D}}, q_{\mathcal{D}})$ for all $q_{\mathcal{D}} \in X_{\mathcal{D},0}^{\mathrm{p}}$, and Young's inequality with $\sigma > 0$, we infer

$$\mathfrak{T}_3^{(N)} \leqslant \frac{\sigma}{2} \sum_{n=0}^{N-1} \delta t^{(n+1/2)} c_{\mathcal{D}}(p_{\mathcal{D}}^{(n+1)}, p_{\mathcal{D}}^{(n+1)}) + \frac{1}{2\sigma} \underline{\kappa}^{-1} (C_{\mathcal{D}}^{\mathrm{p}})^2 \|h\|_{L^2(0,T;L_0^2(\Omega))}^2. \tag{V.27}$$

(b) *λ is bounded away from infinity:* Using (V.17) established in (i) under the assumption that $\mathcal{D}$ admits a Fortin operator, and Young's inequality with $\sigma > 0$, we infer

$$\mathfrak{T}_3^{(N)} \leqslant \frac{\sigma}{2} \sum_{n=0}^{N-1} \delta t^{(n+1/2)} \tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n+1)}, \boldsymbol{u}_{\mathcal{D}}^{(n+1)}) + C_{\mathrm{S}}^2 C_{\mathrm{N}}^2 \left(\sigma^{-1} C_{\mathrm{B}}(\mu, \lambda) + (C_{\mathcal{D}}^{\mathrm{d}})^2 \right) \|h\|_{L^2(0,T;L_0^2(\Omega))}^2 + \frac{T}{2} \|\boldsymbol{f}\|_{L^\infty(0,T;L^2(\Omega)^d)}^2, \tag{V.28}$$

where $C_{\mathrm{B}}(\mu, \lambda) := C_{\mathrm{B}}^2(\mu^{1/2}, \lambda^{1/2})$.

In both cases, we now derive the a priori estimate we look for.
(a) *$\underline{\kappa}$ is bounded away from zero:* Using (V.20), (V.21), (V.22), (V.25), (V.27), (V.26), (V.24), and (V.23) with choices of $\varepsilon, \chi, \sigma$ such that $0 < \varepsilon < 1$, $0 < \sigma < 2$, and $\chi = \min(1-\varepsilon, 2-\sigma) > 0$, we invoke Lemma V.1 with $k = 1$, $a_n := \tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(n)}, \boldsymbol{u}_{\mathcal{D}}^{(n)})$, $b_n := \delta t^{(n-1/2)} c_{\mathcal{D}}(p_{\mathcal{D}}^{(n)}, p_{\mathcal{D}}^{(n)})$, and $\delta_n := \delta t^{(n-1/2)}$ for $n \in [\![1, N]\!]$, to infer

$$\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(N)}, \boldsymbol{u}_{\mathcal{D}}^{(N)}) + \sum_{n=0}^{N-1} \delta t^{(n+1/2)}\, c_{\mathcal{D}}(p_{\mathcal{D}}^{(n+1)}, p_{\mathcal{D}}^{(n+1)}) \leqslant B_{(\mathrm{a})} \exp\left(\sum_{n=0}^{N-1} \frac{\delta t^{(n+1/2)}}{1 - \delta t^{(n+1/2)}} \right), \tag{V.29}$$

where

$$B_{(\mathrm{a})} := \frac{T+2}{\chi} C_{\mathrm{I}}\, \tilde{a}(\boldsymbol{u}^{(0)}, \boldsymbol{u}^{(0)}) + \frac{\varepsilon+1}{\chi\varepsilon} \mu^{-1} (C_{\mathcal{D}}^{\mathrm{d}})^2 \|\boldsymbol{f}\|_{L^\infty(0,T;L^2(\Omega)^d)}^2 + \frac{\chi+1}{\chi^2} \mu^{-1} (C_{\mathcal{D}}^{\mathrm{d}})^2 \|\boldsymbol{f}'\|_{L^2(0,T;L^2(\Omega)^d)}^2 + \frac{1}{\chi\sigma} \underline{\kappa}^{-1} (C_{\mathcal{D}}^{\mathrm{p}})^2 \|h\|_{L^2(0,T;L_0^2(\Omega))}^2.$$

Note that the use of Lemma V.1 with $k = 1$ is licit since the time discretization δ is consistent by assumption. Equation (V.29) provides a discrete $L^\infty(0,T;H_0^1(\Omega)^d)$ estimate on the displacement and $L^2(0,T;\overline{H^1}(\Omega))$ estimate on the pressure, with a multiplicative constant only depending on the coercivity constants $C_{\mathcal{D}}^{\mathrm{d}}$, $C_{\mathcal{D}}^{\mathrm{p}}$, and on bounded quantities (including $\underline{\kappa}^{-1}$). The fact that $\tilde{a}(\boldsymbol{u}^{(0)}, \boldsymbol{u}^{(0)})$ is bounded comes from (II.23) and from the fact that $\lambda^{1/2}\|\beta\|_{0,\Omega}$ is bounded independently of λ by assumption.
(b) *λ is bounded away from infinity:* Using now (V.20), (V.21), (V.22), (V.25), (V.28), (V.26), (V.24), and (V.23) with choices of $\varepsilon, \chi, \sigma$ such that $0 < \varepsilon < 1$, $0 < \sigma < 1-\varepsilon$, $0 < \chi < 1-\varepsilon$, and $\chi + \sigma = 1 - \varepsilon$, we invoke Lemma V.1 with the same arguments as in (iia) to infer

$$\tilde{a}_{\mathcal{D}}(\boldsymbol{u}_{\mathcal{D}}^{(N)}, \boldsymbol{u}_{\mathcal{D}}^{(N)}) + \sum_{n=0}^{N-1} \delta t^{(n+1/2)}\, c_{\mathcal{D}}(p_{\mathcal{D}}^{(n+1)}, p_{\mathcal{D}}^{(n+1)}) \leqslant B_{(\mathrm{b})} \exp\left(\sum_{n=0}^{N-1} \frac{\delta t^{(n+1/2)}}{1 - \delta t^{(n+1/2)}} \right), \tag{V.30}$$

where

$$B_{(\mathrm{b})} := \frac{T+2}{\chi+\sigma} C_{\mathrm{I}}\, \tilde{a}(\boldsymbol{u}^{(0)}, \boldsymbol{u}^{(0)}) + \left(\frac{T}{\chi+\sigma} + \frac{\varepsilon+1}{\varepsilon(\chi+\sigma)} \mu^{-1} (C_{\mathcal{D}}^{\mathrm{d}})^2 \right) \|\boldsymbol{f}\|_{L^\infty(0,T;L^2(\Omega)^d)}^2 + \frac{\chi+1}{\chi(\chi+\sigma)} \mu^{-1} (C_{\mathcal{D}}^{\mathrm{d}})^2 \|\boldsymbol{f}'\|_{L^2(0,T;L^2(\Omega)^d)}^2 + \frac{2C_{\mathrm{S}}^2 C_{\mathrm{N}}^2}{\chi+\sigma} \left(\sigma^{-1} C_{\mathrm{B}}(\mu, \lambda) + (C_{\mathcal{D}}^{\mathrm{d}})^2 \right) \|h\|_{L^2(0,T;L_0^2(\Omega))}^2.$$

Note again that the use of Lemma V.1 with $k = 1$ is licit since δ is consistent by assumption. Here, Equation (V.30) provides a discrete $L^\infty(0,T;H_0^1(\Omega)^d)$ estimate on the displacement and a $L^2(0,T;L^2(\Omega)^d)$ estimate on the product $\kappa^{1/2}$-pressure gradient, with a multiplicative constant only depending on the coercivity constant $C_{\mathcal{D}}^{\mathrm{d}}$, and on bounded quantities (including λ). (iii) Combining (V.17) with $n = N$ and (V.30), we finally get

$$\|\Pi_{\mathcal{D}}^{\mathrm{p}} p_{\mathcal{D}}^{(N)}\|_{0,\Omega}^2 \leqslant 2C_{\mathrm{S}}^2 C_{\mathrm{N}}^2 C_{\mathrm{B}}(\mu,\lambda) B_{(\mathrm{b})} \exp\left(\sum_{n=0}^{N-1} \frac{\delta t^{(n+1/2)}}{1-\delta t^{(n+1/2)}}\right) + 2C_{\mathrm{S}}^2 C_{\mathrm{N}}^2 (C_{\mathcal{D}}^{\mathrm{d}})^2 \|\boldsymbol{f}\|_{L^\infty(0,T;L^2(\Omega)^d)}^2. \tag{V.31}$$

Equation (V.31) provides a discrete $L^2(0,T;L_0^2(\Omega))$ estimate on the pressure whose multiplicative constant only depends on the coercivity constant $C_{\mathcal{D}}^{\mathrm{d}}$, on λ, and on bounded quantities. This concludes the proof. □

The a priori estimates of Lemma V.2 combined with the fact that the matrix of the (linear) discrete problem is square, guarantee the following result.

Lemma V.3 (Existence and uniqueness of the solution to (V.11))**.** *Let $(\delta, \mathcal{D})$ with $\mathcal{D} := (\mathcal{D}^{\mathrm{d}}, \mathcal{D}^{\mathrm{p}})$ be an Euler-Gradient (time-space) discretization in the sense of Definition V.4. Then, independently of the (admissible) values that can possibly be taken by λ or $\underline{\kappa}$, problem* (V.11) *admits a unique solution.*

V.2.2 Convergence result

We can now state the main result of this chapter.

Theorem V.1 (Convergence of the scheme)**.** *Let $(\delta_m, \mathcal{D}_m)_{m\in\mathbb{N}}$ with $\mathcal{D}_m := (\mathcal{D}_m^{\mathrm{d}}, \mathcal{D}_m^{\mathrm{p}})$ be a sequence of Euler-Gradient (time-space) discretizations in the sense of Definition V.4 such that*

(a) the associated sequence of time discretizations $(\delta_m)_{m\in\mathbb{N}}$ is consistent *in the sense of Assumption V.8;*

(b) the associated sequence of displacement/pressure Gradient discretizations $(\mathcal{D}_m)_{m\in\mathbb{N}}$ possesses a sequence of Fortin operators *enjoying* optimal approximation properties *in the sense of Assumption V.7 and* (V.10)*;*

(c) the associated sequences of displacement $(\mathcal{D}_m^{\mathrm{d}})_{m\in\mathbb{N}}$ and pressure $(\mathcal{D}_m^{\mathrm{p}})_{m\in\mathbb{N}}$ Gradient discretizations are coercive *(Assumptions V.4, V.1), enjoy* optimal approximation properties *(*(V.10)*, Assumption V.2), and are* limit-conforming *(Assumptions V.6, V.3).*

For any $m \in \mathbb{N}$, let $(\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}, p_{\mathcal{D}_m}^{\delta_m}) \in \mathbb{P}_d^0\left(\delta_m; \boldsymbol{X}_{\mathcal{D}_m,0}^{\mathrm{d}}\right) \times \mathbb{P}_d^0\left(\delta_m; X_{\mathcal{D}_m,0}^{\mathrm{p}}\right)$ be the solution to the scheme (V.11)*, with $\boldsymbol{u}_{\mathcal{D}_m}^{(0)}$ chosen such that $\boldsymbol{u}_{\mathcal{D}_m}^{(0)} := \mathcal{I}_{\mathcal{D}_m}^{\mathrm{d}}(\boldsymbol{u}^{(0)})$. Then, independently of the values that can possibly be taken by λ or $\underline{\kappa}$ (with the condition that both limit cases cannot occur simultaneously),*

- $\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}} \boldsymbol{u}_{\mathcal{D}_m}^{\delta_m} \to \boldsymbol{\nabla}\boldsymbol{u}$ *in* $L^2(0,T;L^2(\Omega)^{d,d})$, $\quad \lambda^{1/2}\pi_{\mathcal{D}_m}^{\mathrm{p}}(\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}} \cdot \boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}) \to \lambda^{1/2}\boldsymbol{\nabla}\cdot\boldsymbol{u}$ *in* $L^2(0,T;L_0^2(\Omega))$ *as* $m \to +\infty$,
- $\Pi_{\mathcal{D}_m}^{\mathrm{d}} \boldsymbol{u}_{\mathcal{D}_m}^{\delta_m} \rightharpoonup \boldsymbol{u}$ *in* $L^2(0,T;L^2(\Omega)^d)$ *as* $m \to +\infty$,
- $\kappa^{1/2}\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{p}} p_{\mathcal{D}_m}^{\delta_m} \rightharpoonup \kappa^{1/2}\boldsymbol{\nabla} p$ *in* $L^2(0,T;L^2(\Omega)^d)$ *as* $m \to +\infty$,
- $\Pi_{\mathcal{D}_m}^{\mathrm{p}} p_{\mathcal{D}_m}^{\delta_m} \to p$ *in* $L^2(0,T;L_0^2(\Omega))$ *as* $m \to +\infty$,

where $(\boldsymbol{u},p)\in L^2(0,T;H_0^1(\Omega)^d)\times L^2(0,T;\overline{H^1}(\Omega))$ *is the unique solution to* (II.22) *(cf. Theorem II.1).*

Proof. The proof splits into five different parts.
(i) *Choice of the initial discrete displacement:* Let $m\in\mathbb{N}$. Owing to (V.8) and (II.23), $\boldsymbol{u}_{\mathcal{D}_m}^{(0)}$ satisfies (V.11c). In addition, owing to (V.9), and to (V.8) combined with the fact that $\|\pi_{\mathcal{D}_m}^{\rm p}(\boldsymbol{\nabla}\cdot\boldsymbol{u}^{(0)})\|_{0,\Omega}\leqslant\|\boldsymbol{\nabla}\cdot\boldsymbol{u}^{(0)}\|_{0,\Omega}$, Assumption V.9 is satisfied with $C_{\rm I}=\max(C_{\rm S}^2,1)$.
(ii) *A priori estimates on the sequences of solutions:* Let $m\in\mathbb{N}$. We here denote by C a generic constant, independent of m and only depending on bounded quantities. From the first point in Lemma V.2, and owing to the coercivity assumptions on the sequences of displacement and pressure Gradient discretizations, we infer

$$\|\boldsymbol{\nabla}_{\mathcal{D}_m}^{\rm d}\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}\|_{L^2(0,T;L^2(\Omega)^{d,d})}\leqslant C,\qquad\|\lambda^{1/2}\pi_{\mathcal{D}_m}^{\rm p}(\boldsymbol{\nabla}_{\mathcal{D}_m}^{\rm d}\cdot\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m})\|_{L^2(0,T;L_0^2(\Omega))}\leqslant C. \tag{V.32}$$

From the second point in Lemma V.2, and owing again to the coercivity assumptions, we get

$$\|\kappa^{1/2}\boldsymbol{\nabla}_{\mathcal{D}_m}^{\rm p}p_{\mathcal{D}_m}^{\delta_m}\|_{L^2(0,T;L^2(\Omega)^d)}\leqslant C. \tag{V.33}$$

Finally, using the third point in Lemma V.2 when λ is bounded away from infinity, or the second one combined with the discrete Poincaré inequality of Remark V.1 otherwise (in that case $\underline{\kappa}$ is bounded away from zero by assumption), and owing again to the coercivity assumptions, we infer

$$\|\Pi_{\mathcal{D}_m}^{\rm p}p_{\mathcal{D}_m}^{\delta_m}\|_{L^2(0,T;L_0^2(\Omega))}\leqslant C. \tag{V.34}$$

(iii) *Weak convergence:* From (V.32), and the discrete Friedrichs' inequality of Remark V.2 combined with the coercivity assumptions, we infer the existence of $\overline{\boldsymbol{u}}\in L^2(0,T;L^2(\Omega)^d)$, $\overline{\underline{\boldsymbol{G}}}\in L^2(0,T;L^2(\Omega)^{d,d})$, and $\overline{D}\in L^2(0,T;L_0^2(\Omega))$ such that, up to subsequences and without any change in notations,

$$\Pi_{\mathcal{D}_m}^{\rm d}\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}\rightharpoonup\overline{\boldsymbol{u}}\qquad\text{in } L^2(0,T;L^2(\Omega)^d)\qquad\text{as } m\to+\infty, \tag{V.35a}$$

$$\boldsymbol{\nabla}_{\mathcal{D}_m}^{\rm d}\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}\rightharpoonup\overline{\underline{\boldsymbol{G}}}\qquad\text{in } L^2(0,T;L^2(\Omega)^{d,d})\qquad\text{as } m\to+\infty, \tag{V.35b}$$

$$\lambda^{1/2}\pi_{\mathcal{D}_m}^{\rm p}(\boldsymbol{\nabla}_{\mathcal{D}_m}^{\rm d}\cdot\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m})\rightharpoonup\overline{D}\qquad\text{in } L^2(0,T;L_0^2(\Omega))\qquad\text{as } m\to+\infty. \tag{V.35c}$$

Let $\underline{\underline{\varphi}}\in C_c^\infty(0,T;C_c^\infty(\Omega)^{d,d})$, hence for all $t\in(0,T]$, $\underline{\underline{\varphi}}(t)\in C_c^\infty(\Omega)^{d,d}\subset\underline{\underline{H}}(\mathrm{div};\Omega)$ (cf. (V.5)). According to (V.7) (cf. Assumption V.6), and to (V.32), there holds

$$\int_0^T\left((\boldsymbol{\nabla}_{\mathcal{D}_m}^{\rm d}\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t),\underline{\underline{\varphi}}(t))_{0,\Omega}+(\Pi_{\mathcal{D}_m}^{\rm d}\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t),\boldsymbol{\nabla}\cdot\underline{\underline{\varphi}}(t))_{0,\Omega}\right)\mathrm{d}t\leqslant CT^{1/2}\max_{t\in(0,T]}W_{\mathcal{D}_m}^{\rm d}(\underline{\underline{\varphi}}(t)).$$

Going up to the limit $m\to+\infty$, owing to the limit-conformity assumption on the sequence of displacement Gradient discretizations, and to the weak convergences (V.35a) and (V.35b), we infer that $\overline{\boldsymbol{u}}\in L^2(0,T;H_0^1(\Omega)^d)$ and that $\overline{\underline{\boldsymbol{G}}}=\boldsymbol{\nabla}\overline{\boldsymbol{u}}$. From this last result and (V.35b), it is straightforward that

$$\boldsymbol{\nabla}_{\mathcal{D}_m}^{\rm d}\cdot\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}\rightharpoonup\boldsymbol{\nabla}\cdot\overline{\boldsymbol{u}}\quad\text{in } L^2(0,T;L_0^2(\Omega))\text{ as } m\to+\infty. \tag{V.36}$$

Besides, owing to (V.36) and to the strong approximation properties of the L^2-orthogonal projector (V.4), it is also straightforward that

$$\pi_{\mathcal{D}_m}^{\rm p}(\boldsymbol{\nabla}_{\mathcal{D}_m}^{\rm d}\cdot\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m})\rightharpoonup\boldsymbol{\nabla}\cdot\overline{\boldsymbol{u}}\quad\text{in } L^2(0,T;L_0^2(\Omega))\text{ as } m\to+\infty. \tag{V.37}$$

It is now a simple matter, by letting $\lambda^{1/2}$ act on the test function while passing to the limit, to prove that $\overline{D} = \lambda^{1/2}\boldsymbol{\nabla}\cdot\overline{\boldsymbol{u}}$ in (V.35c). It is worth observing that this argument is licit for any (possibly large) value of λ.
From (V.34) and (V.33), we infer the existence of $\overline{p} \in L^2(0,T;L^2_0(\Omega))$ and $\overline{\boldsymbol{G}} \in L^2(0,T;L^2(\Omega)^d)$ such that, up to subsequences and without any change in notations,

$$\Pi^{\mathrm{p}}_{\mathcal{D}_m} p^{\delta_m}_{\mathcal{D}_m} \rightharpoonup \overline{p} \qquad \text{in } L^2(0,T;L^2_0(\Omega)) \qquad \text{as } m \to +\infty, \tag{V.38a}$$

$$\kappa^{1/2}\boldsymbol{\nabla}^{\mathrm{p}}_{\mathcal{D}_m} p^{\delta_m}_{\mathcal{D}_m} \rightharpoonup \overline{\boldsymbol{G}} \qquad \text{in } L^2(0,T;L^2(\Omega)^d) \qquad \text{as } m \to +\infty. \tag{V.38b}$$

Let $\boldsymbol{\varphi} \in C^\infty_c(0,T;C^\infty_c(\Omega)^d)$, hence for all $t \in (0,T]$, $\boldsymbol{\varphi}(t) \in C^\infty_c(\Omega)^d \subset \boldsymbol{H}_0(\mathrm{div};\Omega)$. According to (V.3) (cf. Assumption V.3), and to (V.33), there holds

$$\int_0^T \left((\kappa^{1/2}\boldsymbol{\nabla}^{\mathrm{p}}_{\mathcal{D}_m} p^{\delta_m}_{\mathcal{D}_m}(t), \boldsymbol{\varphi}(t))_{0,\Omega} + (\Pi^{\mathrm{p}}_{\mathcal{D}_m} p^{\delta_m}_{\mathcal{D}_m}(t), \boldsymbol{\nabla}\cdot(\kappa^{1/2}\boldsymbol{\varphi}(t)))_{0,\Omega} \right) \mathrm{d}t \leqslant CT^{1/2} \max_{t\in(0,T]} W^{\mathrm{p}}_{\mathcal{D}_m}(\boldsymbol{\varphi}(t)).$$

Passing to the limit $m \to +\infty$, owing to the limit-conformity assumption on the sequence of pressure Gradient discretizations, and to the weak convergences (V.38a) and (V.38b), we infer that $\overline{p} \in L^2(0,T;\overline{H^1}(\Omega))$ and that $\overline{\boldsymbol{G}} = \kappa^{1/2}\boldsymbol{\nabla}\overline{p}$.
(iv) *Identification of the limit* $(\overline{\boldsymbol{u}},\overline{p})$*:* Let $\varphi \in C^\infty([0,T])$ satisfying $\varphi(T) = 0$. For any time discretization δ of the sequence $(\delta_m)_{m\in\mathbb{N}}$, let denote for $n \in [\![0, N-1]\!]$ $\varphi^{(n+1)} := \varphi(t^{(n+1)})$, in such a way that $\varphi^{(N)} = 0$, and $\psi^{(n+1)} := \frac{1}{\delta t^{(n+1/2)}} \int_{t^{(n)}}^{t^{(n+1)}} \varphi(t)\,\mathrm{d}t$. We introduce the piecewise constant functions $\varphi^\delta \in \mathbb{P}^0_d(\delta)$, $\psi^\delta \in \mathbb{P}^0_d(\delta)$, and $\varphi'^\delta \in \mathbb{P}^0_d(\delta)$ such that, for $n \in [\![0, N-1]\!]$,

$$\varphi^\delta{}_{|(t^{(n)},t^{(n+1)}]} := \varphi^{(n+1)}, \qquad \psi^\delta{}_{|(t^{(n)},t^{(n+1)}]} := \psi^{(n+1)}, \qquad \varphi'^\delta{}_{|(t^{(n)},t^{(n+1)}]} := \frac{\varphi^{(n+2)} - \varphi^{(n+1)}}{\delta t^{(n+1/2)}},$$

with the natural definition $\varphi^{(N+1)} := 0$. Using Taylor's theorem and point (iii) of the time consistency Assumption V.8, one can prove that

$$\|\varphi^\delta - \varphi\|_{L^2((0,T))} \leqslant C_1\delta t^{\mathrm{M}}, \qquad \|\psi^\delta - \varphi\|_{L^2((0,T))} \leqslant C_2\delta t^{\mathrm{M}}, \qquad \|\varphi'^\delta - \varphi'\|_{L^2((0,T))} \leqslant C_3\delta t^{\mathrm{M}}, \tag{V.39}$$

where $C_1, C_2, C_3 > 0$ are three constants independent of δ, depending on T, and on the first derivative of φ for C_1, C_2, on C_t and on the first and second derivatives of φ for C_3. For convenience, we also introduce the piecewise constant in time functions $\boldsymbol{f}^\delta \in \mathbb{P}^0_d\left(\delta; L^2(\Omega)^d\right)$ and $h^\delta \in \mathbb{P}^0_d\left(\delta; L^2_0(\Omega)\right)$ such that, for $n \in [\![0, N-1]\!]$,

$$\boldsymbol{f}^\delta{}_{|(t^{(n)},t^{(n+1)}]} := \boldsymbol{f}^{(n+1)} \in L^2(\Omega)^d, \qquad h^\delta{}_{|(t^{(n)},t^{(n+1)}]} := h^{(n+1)} \in L^2_0(\Omega),$$

where we make use of the definitions of $\boldsymbol{f}^{(n+1)}$ and $h^{(n+1)}$ introduced in Section V.1.3.
Let $\boldsymbol{v} \in H^1_0(\Omega)^d$ and $q \in \overline{H^1}(\Omega)$. For any displacement/pressure Gradient discretization $\mathcal{D} := (\mathcal{D}^{\mathrm{d}}, \mathcal{D}^{\mathrm{p}})$ of the sequence $(\mathcal{D}_m)_{m\in\mathbb{N}}$, let denote $\boldsymbol{v}_\mathcal{D} := \mathcal{I}^{\mathrm{d}}_\mathcal{D}(\boldsymbol{v}) \in \boldsymbol{X}^{\mathrm{d}}_{\mathcal{D},0}$ and

$$X^{\mathrm{p}}_{\mathcal{D},0} \ni q_\mathcal{D} := \operatorname*{argmin}_{r_\mathcal{D}\in X^{\mathrm{p}}_{\mathcal{D},0}} \left(\|\Pi^{\mathrm{p}}_\mathcal{D} r_\mathcal{D} - q\|_{0,\Omega} + \|\boldsymbol{\nabla}^{\mathrm{p}}_\mathcal{D} r_\mathcal{D} - \boldsymbol{\nabla} q\|_{0,\Omega} \right).$$

Let $m \in \mathbb{N}$. Letting $\boldsymbol{v}_{\mathcal{D}_m} \in \boldsymbol{X}^{\mathrm{d}}_{\mathcal{D}_m,0}$ as a test function in (V.11a), multiplying by $\delta t_m^{(n+1/2)}\psi_m^{(n+1)}$, and summing between 0 and $N-1$, then taking $q_{\mathcal{D}_m} \in X^{\mathrm{p}}_{\mathcal{D}_m,0}$ as a test function in (V.11b), multiplying by $\varphi_m^{(n+1)}$, and summing between 0 and $N-1$, finally letting $q_{\mathcal{D}_m} \in X^{\mathrm{p}}_{\mathcal{D}_m,0}$ as a test

function in (V.11c), yields

$$\int_0^T \tilde{a}_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t), \boldsymbol{v}_{\mathcal{D}_m})\,\psi^{\delta_m}(t)\,\mathrm{d}t + \int_0^T b_{\mathcal{D}_m}(\boldsymbol{v}_{\mathcal{D}_m}, p_{\mathcal{D}_m}^{\delta_m}(t))\,\psi^{\delta_m}(t)\,\mathrm{d}t = \mathfrak{R}_m^{(\mathrm{a})}, \tag{V.40a}$$

$$\int_0^T b_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t), q_{\mathcal{D}_m})\,\varphi'^{\delta_m}(t)\,\mathrm{d}t + \int_0^T c_{\mathcal{D}_m}(p_{\mathcal{D}_m}^{\delta_m}(t), q_{\mathcal{D}_m})\,\varphi^{\delta_m}(t)\,\mathrm{d}t = \mathfrak{R}_m^{(\mathrm{b})}, \tag{V.40b}$$

$$(\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}}\cdot\boldsymbol{u}_{\mathcal{D}_m}^{(0)}, \Pi_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m})_{0,\Omega} = (\beta, \Pi_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m})_{0,\Omega}, \tag{V.40c}$$

where we set

$$\mathfrak{R}_m^{(\mathrm{a})} := \int_0^T (\boldsymbol{f}^{\delta_m}(t), \Pi_{\mathcal{D}_m}^{\mathrm{d}}\boldsymbol{v}_{\mathcal{D}_m})_{0,\Omega}\,\psi^{\delta_m}(t)\,\mathrm{d}t,$$

and

$$\mathfrak{R}_m^{(\mathrm{b})} := \int_0^T (h^{\delta_m}(t), \Pi_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m})_{0,\Omega}\,\varphi^{\delta_m}(t)\,\mathrm{d}t - b_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{(0)}, q_{\mathcal{D}_m})\,\varphi(\delta t_m^{(1/2)}),$$

and where we have used the following discrete integration by parts formula (accounting for the fact that $\varphi_m^{(N+1)} = 0$):

$$-\sum_{n=0}^{N-1} b_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{(n+1)} - \boldsymbol{u}_{\mathcal{D}_m}^{(n)}, q_{\mathcal{D}_m})\,\varphi_m^{(n+1)} = \sum_{n=0}^{N-1} b_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{(n+1)}, q_{\mathcal{D}_m})\,(\varphi_m^{(n+2)} - \varphi_m^{(n+1)}) + b_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{(0)}, q_{\mathcal{D}_m})\,\varphi(\delta t_m^{(1/2)}).$$

Let now go up to the limit $m \to +\infty$. Let begin by rewriting the left-hand sides of (V.40a) and (V.40b), that we respectively denote $\mathfrak{L}_m^{(\mathrm{a})}$ and $\mathfrak{L}_m^{(\mathrm{b})}$. First,

$$\begin{aligned}\mathfrak{L}_m^{(\mathrm{a})} = &\int_0^T \mu\,(\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}}\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t), \boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}}\boldsymbol{v}_{\mathcal{D}_m})_{0,\Omega}\,\varphi(t)\,\mathrm{d}t + \int_0^T \mu\,(\pi_{\mathcal{D}_m}^{\mathrm{p}}(\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}}\cdot\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t)), \boldsymbol{\nabla}\cdot\boldsymbol{v})_{0,\Omega}\,\varphi(t)\,\mathrm{d}t\\ &+ \int_0^T (\lambda^{1/2}\pi_{\mathcal{D}_m}^{\mathrm{p}}(\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}}\cdot\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t)), \lambda^{1/2}\boldsymbol{\nabla}\cdot\boldsymbol{v})_{0,\Omega}\,\varphi(t)\,\mathrm{d}t + \int_0^T (\boldsymbol{\nabla}\cdot\boldsymbol{v}, \Pi_{\mathcal{D}_m}^{\mathrm{p}} p_{\mathcal{D}_m}^{\delta_m}(t))_{0,\Omega}\,\varphi(t)\,\mathrm{d}t,\end{aligned}$$

where we have used the definition of ψ^{δ_m}, (V.8) and the fact that $\boldsymbol{v}_{\mathcal{D}_m}$ is the Fortin interpolate of $\boldsymbol{v}$, and the properties of the L^2-orthogonal projector. Then,

$$\begin{aligned}\mathfrak{L}_m^{(\mathrm{b})} = &-\int_0^T (\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}}\cdot\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t), q)_{0,\Omega}\,\varphi'(t)\,\mathrm{d}t - \Big\{\int_0^T (\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}}\cdot\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t), (\Pi_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m} - q))_{0,\Omega}\,\varphi'(t)\,\mathrm{d}t\\ &+ \int_0^T (\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}}\cdot\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t), \Pi_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m})_{0,\Omega}\,(\varphi'^{\delta_m}(t) - \varphi'(t))\,\mathrm{d}t\Big\} + \int_0^T (\kappa^{1/2}\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{p}} p_{\mathcal{D}_m}^{\delta_m}(t), \kappa^{1/2}\boldsymbol{\nabla} q)_{0,\Omega}\,\varphi(t)\,\mathrm{d}t\\ &+ \Big\{\int_0^T (\kappa^{1/2}\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{p}} p_{\mathcal{D}_m}^{\delta_m}(t), \kappa^{1/2}(\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m} - \boldsymbol{\nabla} q))_{0,\Omega}\,\varphi(t)\,\mathrm{d}t\\ &+ \int_0^T (\kappa^{1/2}\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{p}} p_{\mathcal{D}_m}^{\delta_m}(t), \kappa^{1/2}\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m})_{0,\Omega}\,(\varphi^{\delta_m}(t) - \varphi(t))\,\mathrm{d}t\Big\}.\end{aligned}$$

As $m \to +\infty$, owing to the weak convergence results (V.35b), (V.37), (V.35c), (V.38a), and to the strong convergence (V.10) of $(\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}}\boldsymbol{v}_{\mathcal{D}_m})_{m\in\mathbb{N}}$ in $L^2(\Omega)^{d,d}$, we infer

$$\mathfrak{L}_m^{(\mathrm{a})} \to \int_0^T \tilde{a}(\overline{\boldsymbol{u}}(t), \boldsymbol{v})\,\varphi(t)\,\mathrm{d}t + \int_0^T b(\boldsymbol{v}, \overline{p}(t))\,\varphi(t)\,\mathrm{d}t. \tag{V.41}$$

Concerning $\mathfrak{L}_m^{(\mathrm{b})}$, owing to the boundedness of the weakly converging sequences $(\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}} \cdot \boldsymbol{u}_{\mathcal{D}_m}^{\delta_m})_{m\in\mathbb{N}}$ and $(\kappa^{1/2}\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{p}} p_{\mathcal{D}_m}^{\delta_m})_{m\in\mathbb{N}}$ (due to (V.36), (V.38b)), and of the strongly converging ones $(\Pi_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m})_{m\in\mathbb{N}}$ and $(\kappa^{1/2}\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m})_{m\in\mathbb{N}}$ (V.2), combined with the strong convergence of the same last sequences and of $(\varphi^{\delta_m})_{m\in\mathbb{N}}$ and $(\varphi'^{\delta_m})_{m\in\mathbb{N}}$ (V.39), the terms into brackets vanish as $m \to +\infty$. Finally, the weak convergences (V.36) and (V.38b) enable to infer

$$\mathfrak{L}_m^{(\mathrm{b})} \to \int_0^T b(\overline{\boldsymbol{u}}(t), q)\, \varphi'(t)\, \mathrm{d}t + \int_0^T c(\overline{p}(t), q)\, \varphi(t)\, \mathrm{d}t. \tag{V.42}$$

Concerning the right-hand side, using the definition of h^{δ_m}, and combining the expression of $\mathfrak{R}_m^{(\mathrm{b})}$ with (V.40c) yields

$$\mathfrak{R}_m^{(\mathrm{b})} = \int_0^T (h(t), \Pi_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m})_{0,\Omega}\, \varphi^{\delta_m}(t)\, \mathrm{d}t + (\beta, \Pi_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m})_{0,\Omega}\, \varphi(\delta t_m^{(1/2)}).$$

Owing to the strong convergence of the sequences $(\Pi_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m})_{m\in\mathbb{N}}$ (V.2) and $(\varphi^{\delta_m})_{m\in\mathbb{N}}$ (V.39), and to the continuity of φ and point (i) of the time consistency Assumption V.8, we infer in the limit $m \to +\infty$ that

$$\mathfrak{R}_m^{(\mathrm{b})} \to \int_0^T (h(t), q)_{0,\Omega}\, \varphi(t)\, \mathrm{d}t + (\beta, q)_{0,\Omega}\, \varphi(0). \tag{V.43}$$

Then, using the dominated convergence theorem on vector-valued Sobolev spaces, we show that in the limit $m \to +\infty$,

$$\mathfrak{R}_m^{(\mathrm{a})} \to \int_0^T (\boldsymbol{f}(t), \boldsymbol{v})_{0,\Omega}\, \varphi(t)\, \mathrm{d}t. \tag{V.44}$$

Following Ženíšek [82, p. 205], one can prove, using (V.42) and (V.43), that

$$((\boldsymbol{\nabla}\cdot\overline{\boldsymbol{u}})(0), q)_{0,\Omega} = (\beta, q)_{0,\Omega}. \tag{V.45}$$

Finally, restricting our study to $\varphi \in C_c^\infty((0,T))$ (then $\varphi(0) = 0$), owing to (V.40), (V.41), (V.42), (V.44), (V.43), and (V.45), $(\overline{\boldsymbol{u}}, \overline{p}) \in L^2(0,T; H_0^1(\Omega)^d) \times L^2(0,T; \overline{H^1}(\Omega))$ turns out to be a solution to problem (II.22). This proves in a constructive way the existence of solutions to (II.22). Owing to the uniqueness of such solutions, cf. Theorem II.1, $(\overline{\boldsymbol{u}}, \overline{p}) = (\boldsymbol{u}, p)$, where $(\boldsymbol{u}, p)$ denotes the unique solution to problem (II.22), and the whole sequences converge.

(v) *Strong convergence:* For any time-space discretization $(\delta, \mathcal{D})$ of the sequence $(\delta_m, \mathcal{D}_m)_{m\in\mathbb{N}}$, let denote for $k \in [\![0, N-1]\!]$ $z_{\mathcal{D}}^{(k+1)} := \sum_{n=0}^k \delta t^{(n+1/2)} p_{\mathcal{D}}^{(n+1)}$, and let introduce $z_{\mathcal{D}} \in X_{\mathcal{D},0}^{\mathrm{p}}$ such that

$$X_{\mathcal{D},0}^{\mathrm{p}} \ni z_{\mathcal{D}} := z_{\mathcal{D}}^{(N)} = \int_0^T p_{\mathcal{D}}^{\delta}(t)\, \mathrm{d}t.$$

Let $m \in \mathbb{N}$, and $k \in [\![0, N-1]\!]$. Summing (V.11b) on n between 0 and k, and using (V.11c), yields, for all $q_{\mathcal{D}_m} \in X_{\mathcal{D}_m,0}^{\mathrm{p}}$,

$$-b_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{(k+1)}, q_{\mathcal{D}_m}) + c_{\mathcal{D}_m}(z_{\mathcal{D}_m}^{(k+1)}, q_{\mathcal{D}_m}) = (\int_0^{t_m^{(k+1)}} h(s)\, \mathrm{d}s, \Pi_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m})_{0,\Omega} + (\beta, \Pi_{\mathcal{D}_m}^{\mathrm{p}} q_{\mathcal{D}_m})_{0,\Omega}.$$

Letting now $q_{\mathcal{D}_m} = \delta t_m^{(k+1/2)} p_{\mathcal{D}_m}^{(k+1)}$, and summing on k between 0 and $N-1$, we infer

$$\sum_{k=0}^{N-1} \delta t_m^{(k+1/2)} \Big(-b_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{(k+1)}, p_{\mathcal{D}_m}^{(k+1)}) + c_{\mathcal{D}_m}(z_{\mathcal{D}_m}^{(k+1)}, p_{\mathcal{D}_m}^{(k+1)})\Big) =$$
$$\sum_{k=0}^{N-1} \delta t_m^{(k+1/2)} (\int_0^{t_m^{(k+1)}} h(s)\, \mathrm{d}s, \Pi_{\mathcal{D}_m}^{\mathrm{p}} p_{\mathcal{D}_m}^{(k+1)})_{0,\Omega} + (\beta, \Pi_{\mathcal{D}_m}^{\mathrm{p}} z_{\mathcal{D}_m})_{0,\Omega}. \tag{V.46}$$

Letting $\boldsymbol{v}_{\mathcal{D}_m} = \delta t_m^{(k+1/2)} \boldsymbol{u}_{\mathcal{D}_m}^{(k+1)}$ in (V.11a), and summing on k between 0 and $N-1$, we get

$$\sum_{k=0}^{N-1} \delta t_m^{(k+1/2)} \Big(\tilde{a}_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{(k+1)}, \boldsymbol{u}_{\mathcal{D}_m}^{(k+1)}) + b_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{(k+1)}, p_{\mathcal{D}_m}^{(k+1)}) \Big) = \sum_{k=0}^{N-1} \delta t_m^{(k+1/2)} (\boldsymbol{f}^{(k+1)}, \Pi_{\mathcal{D}_m}^{\mathrm{d}} \boldsymbol{u}_{\mathcal{D}_m}^{(k+1)})_{0,\Omega}. \quad \text{(V.47)}$$

Introducing the two sequences $(\boldsymbol{a}_k)_{k\in[\![1,N]\!]}$, and $(\boldsymbol{b}_k)_{k\in[\![0,N]\!]}$ with $\boldsymbol{b}_0 := \boldsymbol{0}$, such that for all $k \in [\![0, N-1]\!]$,

$$\boldsymbol{a}_{k+1} := \delta t_m^{(k+1/2)} \kappa^{1/2} \boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{p}} p_{\mathcal{D}_m}^{(k+1)}, \qquad \boldsymbol{b}_{k+1} := \kappa^{1/2} \boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{p}} z_{\mathcal{D}_m}^{(k+1)},$$

noting that $\sum_{k=0}^{N-1} \delta t_m^{(k+1/2)} c_{\mathcal{D}_m}(z_{\mathcal{D}_m}^{(k+1)}, p_{\mathcal{D}_m}^{(k+1)}) = \sum_{k=0}^{N-1} \int_\Omega \boldsymbol{b}_{k+1} \cdot \boldsymbol{a}_{k+1}$, and that $\boldsymbol{a}_{k+1} = \boldsymbol{b}_{k+1} - \boldsymbol{b}_k$ for all $k \in [\![0, N-1]\!]$, finally recalling the following inequality

$$\boldsymbol{b}_{k+1} \cdot (\boldsymbol{b}_{k+1} - \boldsymbol{b}_k) \geqslant \frac{1}{2} |\boldsymbol{b}_{k+1}|^2 - \frac{1}{2} |\boldsymbol{b}_k|^2,$$

yields

$$\sum_{k=0}^{N-1} \delta t_m^{(k+1/2)} c_{\mathcal{D}_m}(z_{\mathcal{D}_m}^{(k+1)}, p_{\mathcal{D}_m}^{(k+1)}) \geqslant \frac{1}{2} c_{\mathcal{D}_m}(z_{\mathcal{D}_m}, z_{\mathcal{D}_m}). \quad \text{(V.48)}$$

From (V.46), (V.47), and (V.48), there holds

$$\int_0^T \tilde{a}_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t), \boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t))\,\mathrm{d}t + \frac{1}{2} c_{\mathcal{D}_m}(z_{\mathcal{D}_m}, z_{\mathcal{D}_m}) \leqslant \int_0^T (\boldsymbol{f}^{\delta_m}(t), \Pi_{\mathcal{D}_m}^{\mathrm{d}} \boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t))_{0,\Omega}\,\mathrm{d}t + \sum_{k=0}^{N-1} \delta t_m^{(k+1/2)} \Big(\int_0^{t_m^{(k+1)}} h(s)\,\mathrm{d}s, \Pi_{\mathcal{D}_m}^{\mathrm{p}} p_{\mathcal{D}_m}^{(k+1)} \Big)_{0,\Omega} + (\beta, \Pi_{\mathcal{D}_m}^{\mathrm{p}} z_{\mathcal{D}_m})_{0,\Omega}. \quad \text{(V.49)}$$

From (V.38a), (V.38b), and from the identification of the limit (iv), we infer that $\kappa^{1/2} \boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{p}} z_{\mathcal{D}_m} \rightharpoonup \kappa^{1/2} \boldsymbol{\nabla} z(T)$ in $L^2(\Omega)^d$, and that $\Pi_{\mathcal{D}_m}^{\mathrm{p}} z_{\mathcal{D}_m} \rightharpoonup z(T)$ in $L_0^2(\Omega)$, as $m \to +\infty$, where we have used the notation of Lemma II.5. Then, owing to (V.35a), (V.38a), and (V.49), there holds

$$\limsup_{m\to+\infty} \int_0^T \tilde{a}_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t), \boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t))\,\mathrm{d}t \leqslant -\frac{1}{2} c(z(T), z(T)) + \int_0^T (\boldsymbol{f}(t), \boldsymbol{u}(t))_{0,\Omega}\,\mathrm{d}t + \int_0^T \Big(\int_0^t h(s)\,\mathrm{d}s, p(t) \Big)_{0,\Omega}\,\mathrm{d}t + (\beta, z(T))_{0,\Omega},$$

where the terms containing $\boldsymbol{f}$ and h have been treated using the dominated convergence theorem on vector-valued Sobolev spaces. Using the weak convergence results (V.35b) and (V.35c), and the estimate (II.24) of Lemma II.5, we infer

$$\lim_{m\to+\infty} \int_0^T \tilde{a}_{\mathcal{D}_m}(\boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t), \boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}(t))\,\mathrm{d}t = \int_0^T \tilde{a}(\boldsymbol{u}(t), \boldsymbol{u}(t))\,\mathrm{d}t. \quad \text{(V.50)}$$

This last result (V.50) shows the strong convergence of $(\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}} \boldsymbol{u}_{\mathcal{D}_m}^{\delta_m})_{m\in\mathbb{N}}$ and $(\lambda^{1/2} \pi_{\mathcal{D}_m}^{\mathrm{p}} (\boldsymbol{\nabla}_{\mathcal{D}_m}^{\mathrm{d}} \cdot \boldsymbol{u}_{\mathcal{D}_m}^{\delta_m}))_{m\in\mathbb{N}}$.

It now remains to prove the strong convergence of $(\Pi_{\mathcal{D}_m}^{\mathrm{p}} p_{\mathcal{D}_m}^{\delta_m})_{m\in\mathbb{N}}$. Let $m \in \mathbb{N}$. Following point (i) of the proof of Lemma V.2, let take $n \in [\![1, N]\!]$. There exists $\boldsymbol{v}_{\mathrm{N},m}^{(n)} \in H_0^1(\Omega)^d$ such that

$\Pi^{\rm p}_{\mathcal{D}_m} p^{(n)}_{\mathcal{D}_m} = \boldsymbol{\nabla}\cdot\boldsymbol{v}^{(n)}_{{\rm N},m}$, and $\|\boldsymbol{\nabla}\boldsymbol{v}^{(n)}_{{\rm N},m}\|_{0,\Omega} \leqslant C_{\rm N}\|\Pi^{\rm p}_{\mathcal{D}_m} p^{(n)}_{\mathcal{D}_m}\|_{0,\Omega}$, with $C_{\rm N} > 0$ defined in Lemma II.3 and independent of m, and n. Let $\boldsymbol{X}^{\rm d}_{\mathcal{D}_m,0} \ni \boldsymbol{v}^{(n)}_{{\rm N},\mathcal{D}_m} := \mathcal{I}^{\rm d}_{\mathcal{D}_m}(\boldsymbol{v}^{(n)}_{{\rm N},m})$. Then, owing to Remark V.2 combined with the coercivity assumption, and to (V.9), there holds

$$(C^{\rm d}_{\rm F})^{-1}\|\Pi^{\rm d}_{\mathcal{D}_m}\boldsymbol{v}^{(n)}_{{\rm N},\mathcal{D}_m}\|_{0,\Omega} \leqslant \|\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}_m}\boldsymbol{v}^{(n)}_{{\rm N},\mathcal{D}_m}\|_{0,\Omega} \leqslant C_{\rm S}C_{\rm N}\|\Pi^{\rm p}_{\mathcal{D}_m} p^{(n)}_{\mathcal{D}_m}\|_{0,\Omega}. \tag{V.51}$$

For any time-space discretization $(\delta,\mathcal{D})$, let introduce the piecewise constant in time function $\boldsymbol{v}^{\delta}_{{\rm N},\mathcal{D}} \in \mathbb{P}^0_d\left(\delta;\boldsymbol{X}^{\rm d}_{\mathcal{D},0}\right)$ such that, for $n \in [\![0,N-1]\!]$,

$$\boldsymbol{v}^{\delta}_{{\rm N},\mathcal{D}|\,(t^{(n)},t^{(n+1)}]} := \boldsymbol{v}^{(n+1)}_{{\rm N},\mathcal{D}}.$$

Owing to the estimates (V.34) and (V.51), valid for any admissible values of $\underline{\kappa}$ and λ, we infer the existence of $\overline{\boldsymbol{v}}_{\rm N} \in L^2(0,T;H^1_0(\Omega)^d)$ such that, as $m \to +\infty$,

$$\Pi^{\rm d}_{\mathcal{D}_m}\boldsymbol{v}^{\delta_m}_{{\rm N},\mathcal{D}_m} \rightharpoonup \overline{\boldsymbol{v}}_{\rm N} \qquad \text{in } L^2(0,T;L^2(\Omega)^d), \tag{V.52a}$$

$$\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}_m}\boldsymbol{v}^{\delta_m}_{{\rm N},\mathcal{D}_m} \rightharpoonup \boldsymbol{\nabla}\overline{\boldsymbol{v}}_{\rm N} \qquad \text{in } L^2(0,T;L^2(\Omega)^{d,d}), \tag{V.52b}$$

where we have used the limit-conformity Assumption V.6.
Taking $\varphi \in C^\infty_c(0,T;C^\infty_c(\Omega)\cap L^2_0(\Omega))$, and studying the limit of $\int_0^T (\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}_m}\cdot\boldsymbol{v}^{\delta_m}_{{\rm N},\mathcal{D}_m}(t), \pi^{\rm p}_{\mathcal{D}_m}(\varphi(t)))_{0,\Omega}\,{\rm d}t$, we infer, using (V.8), the weak convergence of $(\Pi^{\rm p}_{\mathcal{D}_m}p^{\delta_m}_{\mathcal{D}_m})_{m\in\mathbb{N}}$ (V.38a), and of $(\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}_m}\cdot\boldsymbol{v}^{\delta_m}_{{\rm N},\mathcal{D}_m})_{m\in\mathbb{N}}$ (as a direct consequence of (V.52b)), along with the strong approximation properties of the L^2-orthogonal projector (V.4), that

$$\boldsymbol{\nabla}\cdot\overline{\boldsymbol{v}}_{\rm N} = p \qquad \text{in } L^2(0,T;L^2_0(\Omega)). \tag{V.53}$$

Then, owing to (V.14),

$$\begin{aligned}\limsup_{m\to+\infty}\|\Pi^{\rm p}_{\mathcal{D}_m}p^{\delta_m}_{\mathcal{D}_m}\|^2_{L^2(0,T;L^2_0(\Omega))} &= \limsup_{m\to+\infty}\left(\int_0^T \tilde{a}_{\mathcal{D}_m}(\boldsymbol{u}^{\delta_m}_{\mathcal{D}_m}(t),\boldsymbol{v}^{\delta_m}_{{\rm N},\mathcal{D}_m}(t))\,{\rm d}t\right.\\ &\left. - \int_0^T (\boldsymbol{f}^{\delta_m}(t),\Pi^{\rm d}_{\mathcal{D}_m}\boldsymbol{v}^{\delta_m}_{{\rm N},\mathcal{D}_m}(t))_{0,\Omega}\,{\rm d}t\right) = \int_0^T \tilde{a}(\boldsymbol{u}(t),\overline{\boldsymbol{v}}_{\rm N}(t))\,{\rm d}t - \int_0^T (\boldsymbol{f}(t),\overline{\boldsymbol{v}}_{\rm N}(t))_{0,\Omega}\,{\rm d}t,\end{aligned}$$

where we have used the strong convergence of $(\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}_m}\boldsymbol{u}^{\delta_m}_{\mathcal{D}_m})_{m\in\mathbb{N}}$ and $(\lambda^{1/2}\pi^{\rm p}_{\mathcal{D}_m}(\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}_m}\cdot\boldsymbol{u}^{\delta_m}_{\mathcal{D}_m}))_{m\in\mathbb{N}}$, and the fact that $(\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}_m}\boldsymbol{v}^{\delta_m}_{{\rm N},\mathcal{D}_m})_{m\in\mathbb{N}}$, $(\lambda^{1/2}\pi^{\rm p}_{\mathcal{D}_m}(\boldsymbol{\nabla}^{\rm d}_{\mathcal{D}_m}\cdot\boldsymbol{v}^{\delta_m}_{{\rm N},\mathcal{D}_m}))_{m\in\mathbb{N}}$, and $(\Pi^{\rm d}_{\mathcal{D}_m}\boldsymbol{v}^{\delta_m}_{{\rm N},\mathcal{D}_m})_{m\in\mathbb{N}}$, weakly converge as a consequence of (V.52b) and (V.52a). The term containing $\boldsymbol{f}$ is here again handled using the dominated convergence theorem on vector-valued Sobolev spaces. Finally, from (II.22a) and (V.53), and from the weak convergence of $(\Pi^{\rm d}_{\mathcal{D}_m}\boldsymbol{v}^{\delta_m}_{{\rm N},\mathcal{D}_m})_{m\in\mathbb{N}}$, we infer

$$\begin{aligned}\|p\|^2_{L^2(0,T;L^2_0(\Omega))} \leqslant \liminf_{m\to+\infty}\|\Pi^{\rm p}_{\mathcal{D}_m}p^{\delta_m}_{\mathcal{D}_m}\|^2_{L^2(0,T;L^2_0(\Omega))} &\leqslant \limsup_{m\to+\infty}\|\Pi^{\rm p}_{\mathcal{D}_m}p^{\delta_m}_{\mathcal{D}_m}\|^2_{L^2(0,T;L^2_0(\Omega))}\\ &\leqslant -\int_0^T b(\overline{\boldsymbol{v}}_{\rm N}(t),p(t))\,{\rm d}t = \|p\|^2_{L^2(0,T;L^2_0(\Omega))}.\end{aligned} \tag{V.54}$$

Equation (V.54) shows the strong convergence of the approximate pressure reconstruction, hence concluding the proof. □

V.3 Some examples of Gradient discretizations

A large number of well-known methods can be proved to fall in the framework of Gradient schemes. Among them, we can cite the Galerkin methods (and in particular conforming finite elements), the Crouzeix–Raviart method and most of nonconforming finite element methods, some MPFA and DDFV schemes, the HFV/MFD/MFV class of methods, and the Vertex Approximate Gradient (VAG) scheme introduced by Eymard, Guichard, Herbin and Masson [42, 41, 43, 44]. By falling in the framework, we mean that all these discretization methods can be described, when applied to the approximation of a linear (or nonlinear with an additional compactness assumption) elliptic problem, through a definition like Definition V.2, and can be proved to satisfy the assumptions of coercivity (cf. Definition V.4 for example), optimal approximation or consistency (cf. Definition V.5 for example), and limit-conformity (cf., e.g., Definition V.6). For all the above cited methods, the classical proofs of these results can be found in the Gradient scheme literature, we can cite in particular [45, 41, 33]. The last reference contains in Section 5.3 a detailed study on the HFV/MFD/MFV (referred to as HMM for Hybrid Mimetic Mixed) class of methods.

In our case, since we study a saddle-point problem, we add another assumption to characterize the admissible pairs of displacement/pressure Gradient discretizations, which is the one of satisfying an inf-sup condition when coupled, cf. Assumption V.7 and Remark V.3. This obviously reduces the field of admissible methods (for which applies in particular the convergence Theorem V.1). We can cite as candidates:

- inf-sup stable pairs of conforming finite elements, for example the $\mathbb{P}_2^d/\mathbb{P}_1$ method, or the mini $\mathbb{P}_1^d-$bubble$/\mathbb{P}_1$ element; this eliminates equal-order Lagrangian approximations like the (unstable) $\mathbb{P}_1^d/\mathbb{P}_1$ method;
- the HFV/HFV method (which gives a coercive formulation of linear elasticity for the pure displacement problem);
- the HFV-I (for Interpolation)/HFV method (see Appendix C and Remark C.3) which gives a coercive formulation of linear elasticity even for mixed-type boundary conditions without the need to introduce jumps.

Note that a method using the Crouzeix–Raviart space for both the displacement and pressure discretization cannot satisfy the inf-sup assumption since the Crouzeix–Raviart pressure reconstruction is piecewise affine.

In the applications of Section V.4, we consider a generalized Crouzeix–Raviart discretization (see Chapter IV) of the displacement coupled to a HFV discretization of pressure. The first enables to take advantage of the introduction of a piecewise affine reconstruction to obtain a coercive formulation for linear elasticity in the case of mixed-type mechanical boundary conditions (thanks to jumps penalization), while the second enables (thanks to its piecewise constant reconstruction) to guarantee that an inf-sup condition holds when coupled to displacement (cf. Lemma III.5 and Corollary III.1). For both methods, we need to define a notion of admissible mesh sequence. The one of Definition III.3 is well-adapted, and for both, since the discretization is not staggered.

It is completely straightforward in the light of Chapter III to prove that the generalized Crouzeix–Raviart method (at least for $\eta = d$) defines a Gradient discretization (satisfying the assumptions of coercivity, optimal approximation, limit-conformity and existence of a Fortin operator) for the pure displacement problem (for mixed-type boundary conditions, a discrete Korn's inequality assumption would be needed to complete the framework). As far as the HFV method (which is also a Gradient discretization as we already said) is concerned, the main

features of the method (which is actually closely related to the generalized Crouzeix–Raviart method as they only differ from the kind of reconstruction considered and the stabilization parameter of the gradient) are recalled in Appendix C.

Something that has to be noted in the case of Biot's consolidation problem is that the classical assumption of limit-conformity is a bit modified for pressure since we consider as a *gradient operator* the product of the mobility square root with the pressure gradient. Then, the proof of limit-conformity for HFV discretizations has to be adapted as well. Following the steps of the proof of [33, Lemma 5.9], and denoting $h_{\mathcal{D}}$ the meshsize, we get for the HFV method

$$W_{\mathcal{D}}^{\mathrm{p}}(\boldsymbol{\varphi}) \leqslant C_1 h_{\mathcal{D}} \frac{\|\boldsymbol{\kappa}^{1/2}\boldsymbol{\varphi}\|_{W^{1,\infty}(\Omega)^d} \|\boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{p}} q_{\mathcal{D}}\|_{0,\Omega}}{\|\boldsymbol{\kappa}^{1/2}\boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{p}} q_{\mathcal{D}}\|_{0,\Omega}} \leqslant C_2 h_{\mathcal{D}} \|\boldsymbol{\varphi}\|_{W^{1,\infty}(\mathbb{R}^d)^d} \frac{\|\boldsymbol{\kappa}^{1/2}\|_{W^{1,\infty}(\Omega)}}{\underline{\kappa}^{1/2}},$$

where $\boldsymbol{\varphi} \in C_c^\infty(\mathbb{R}^d)^d$ and $C_1, C_2 > 0$ are independent of $h_{\mathcal{D}}$. The result of Assumption V.3 can then only be obtained (using the above inequality and [33, Lemma 2.9]) if (II.21) is fulfilled, hence justifying this latter assumption.

V.4 Numerical applications

We provide in this section some carefully chosen two-dimensional numerical examples in order to assess the performance of a particular Euler-Gradient approximation scheme for problem (II.22), where the displacement/pressure space discretization is handled using the generalized Crouzeix–Raviart/HFV method.

We test the behavior of the numerical scheme (on general meshes) with respect to some relevant parameters of the poroelasticity model, i.e. the constrained specific storage coefficient c_0 (we tackle in particular the limit case $c_0 = 0$ corresponding to Biot's consolidation model), the mobility κ (we focus in particular on the case of a heterogeneous field with locally small permeability $\underline{\kappa} \to 0^+$), and the time T (we particularly tackle the case of early and long times).

The focus here is neither on the influence of Lamé parameters on the approximation of mechanics (as it has already been fully assessed in Section IV.4), nor on the influence of the Biot–Willis coefficient α on the approximation of the poroelasticity problem, since from a physical point of view this parameter is often close to unity. We will take it equal to one in the whole section.

When it is relevant, we propose a comparison of the results with a conforming (unstable) finite element pair $\mathbb{P}_1^d/\mathbb{P}_1$. The implementation has been realized in the same 2D `C++` prototype as the one used for the numerical examples of Section IV.4. We recall that the implementation of this prototype relies on the general framework introduced in [26, 27].

V.4.1 Mesh families, time discretization and error measure

As far as the spatial discretization is concerned, we consider some of the two-dimensional mesh families of Section V.4:

(a) a *matching triangular* mesh sequence, which will be useful for comparison with the conforming $\mathbb{P}_1^d/\mathbb{P}_1$ finite element method; cf. Figure IV.1a;

(b) a *Cartesian* mesh sequence, as it is the most widely used grid type in reservoir simulation and as it forms the basis of CPG meshes; cf. Figure IV.1b;

(c) a *Kershaw-type* mesh sequence, which is of great practical interest as it may represent a geological porous medium that has historically undergone non-smooth deformations toward a highly skewed state; cf. Figure IV.1d.

Note that, even if it is not included in that section, the convergence of the method has also been assessed on a locally refined Cartesian mesh sequence (cf. Figure IV.1c) to test the treatment of nonmatching interfaces, on a trapezoidal mesh sequence (cf. Figure IV.1e) to test the behavior on grids whose elements do not converge to parallelograms, and on the hexagonal-dominant mesh sequence of Figure IV.1f to test the behavior on grids featuring different polygonal elements.

The linear elasticity model is discretized using the CRg-VS bilinear form (IV.6) (in its homogeneous version since we consider constant Lamé parameters), which necessitates to consider pure Dirichlet mechanical boundary conditions. The results presented in the next paragraphs have thus been computed considering pure Dirichlet boundary conditions, for both mechanics and flow. Experiments have been realized using the CRg-JS bilinear form (IV.3) to discretize linear elasticity and confirm that the CRg-JS/HFV method correctly handles the case of mixed-type (for mechanics or/and flow) boundary conditions. For both the generalized Crouzeix–Raviart and the HFV method, we make the choice $\eta = d$ in (III.9) for the subgrid stabilization parameter of their common (in the expression) gradient.

As far as the time discretization is concerned (cf. Definition V.3), we consider sequences $(\delta_m)_{m\in\mathbb{N}}$ such that, for any member $m \in \mathbb{N}$ of a sequence, the time step δt_m is uniform and such that $\delta t_m = T/N_m$. A time discretization sequence is related to its associated mesh sequence in the following way: between two successive members, when the meshsize halves, the (uniform) time step is divided by four. This enables to obtain optimal errors in the $L^2(\Omega)$-norm as they depend on the meshsize square.

As far as the approximation of the right-hand sides of (V.11a) and (V.11b) is concerned, we treat it in a finite volume way by using the cell unknown (available for both discretizations) as a quadrature point, cf. (IV.19) for an example.

Finally, for a given time-space discretization $(\delta, \mathcal{D})$ with $\mathcal{D} := (\mathcal{D}^{\mathrm{d}}, \mathcal{D}^{\mathrm{p}})$ ($h_{\mathcal{D}}$ denotes the meshsize), the relative errors are measured at the final time T such that $t^{(N)} = T$. For the displacement, it is computed in the following way

$$\frac{\|\boldsymbol{\nabla}\boldsymbol{u}(T) - \boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{d}}\boldsymbol{u}_{\mathcal{D}}^{(N)}\|_{0,\Omega}}{\|\boldsymbol{\nabla}\boldsymbol{u}(T)\|_{0,\Omega}} \approx \frac{\left(\sum_{K\in\mathcal{K}_h} |K| \left|(\boldsymbol{\nabla}\boldsymbol{u})(\boldsymbol{x}_K, T) - \boldsymbol{\nabla}_K^{\mathrm{d}}\boldsymbol{u}_{\mathcal{D}}^{(N)}\right|^2\right)^{1/2}}{\left(\sum_{K\in\mathcal{K}_h} |K| \left|(\boldsymbol{\nabla}\boldsymbol{u})(\boldsymbol{x}_K, T)\right|^2\right)^{1/2}}, \tag{V.55a}$$

$$\frac{\|\boldsymbol{u}(T) - \Pi_{\mathcal{D}}^{\mathrm{d}}\boldsymbol{u}_{\mathcal{D}}^{(N)}\|_{0,\Omega}}{\|\boldsymbol{u}(T)\|_{0,\Omega}} \approx \frac{\left(\sum_{K\in\mathcal{K}_h} |K| \left|\boldsymbol{u}(\boldsymbol{x}_K, T) - \boldsymbol{u}_K^{(N)}\right|^2\right)^{1/2}}{\left(\sum_{K\in\mathcal{K}_h} |K| \left|\boldsymbol{u}(\boldsymbol{x}_K, T)\right|^2\right)^{1/2}}, \tag{V.55b}$$

where the notations are all introduced in Definition C.1 of Appendix C. This measure of the error is the same (at time T) as the one introduced in (IV.24a)–(IV.24b) but in another framework and using different notations. For the pore pressure, the errors are computed as

$$\frac{\|\boldsymbol{\nabla}p(T) - \boldsymbol{\nabla}_{\mathcal{D}}^{\mathrm{p}}p_{\mathcal{D}}^{(N)}\|_{0,\Omega}}{\|\boldsymbol{\nabla}p(T)\|_{0,\Omega}} \approx \frac{\left(\sum_{K\in\mathcal{K}_h} |K| \left|(\boldsymbol{\nabla}p)(\boldsymbol{x}_K, T) - \boldsymbol{\nabla}_K^{\mathrm{p}}p_{\mathcal{D}}^{(N)}\right|^2\right)^{1/2}}{\left(\sum_{K\in\mathcal{K}_h} |K| \left|(\boldsymbol{\nabla}p)(\boldsymbol{x}_K, T)\right|^2\right)^{1/2}}, \tag{V.56a}$$

$$\frac{\|p(T) - \Pi_{\mathcal{D}}^{\mathrm{p}}p_{\mathcal{D}}^{(N)}\|_{0,\Omega}}{\|p(T)\|_{0,\Omega}} \approx \frac{\left(\sum_{K\in\mathcal{K}_h} |K| \left|p(\boldsymbol{x}_K, T) - p_K^{(N)}\right|^2\right)^{1/2}}{\left(\sum_{K\in\mathcal{K}_h} |K| \left|p(\boldsymbol{x}_K, T)\right|^2\right)^{1/2}}, \tag{V.56b}$$

where the notations are the same as for the displacement but in the scalar case.

For the sake of simplicity, the H^1 relative errors (V.55a) and (V.56a) are referred to as $\|\boldsymbol{\nabla} \boldsymbol{u} - \boldsymbol{\nabla}^{\mathrm{d}}_{\mathcal{D}} \boldsymbol{u}_{\mathcal{D}}\|_{0,\Omega}$ and $\|\boldsymbol{\nabla} p - \boldsymbol{\nabla}^{\mathrm{p}}_{\mathcal{D}} p_{\mathcal{D}}\|_{0,\Omega}$ in the plots axes, and the L^2 relative errors (V.55b) and (V.56b) as $\|\boldsymbol{u} - \Pi^{\mathrm{d}}_{\mathcal{D}} \boldsymbol{u}_{\mathcal{D}}\|_{0,\Omega}$ and $\|p - \Pi^{\mathrm{p}}_{\mathcal{D}} p_{\mathcal{D}}\|_{0,\Omega}$.

Note that when computing the L^2 error (at final time) for the $\mathbb{P}_1$ pore pressure approximation, we use a quadrature formula which is exact for polynomials of degree 2.

We consider in the test-cases below examples where the constrained specific storage coefficient c_0 does not vanish. Hence, we need to discretize the term $\partial_t(c_0 p)$ in (II.18b). This term adds a contribution in the weak formulation (II.22) in the left-hand side of the flow equation (II.22b) which takes the form $-\int_0^T c_0\,(p(t), q)_{0,\Omega}\,\varphi'(t)\,\mathrm{d}t$. From a discrete point of view, to take into account this contribution in (V.11), we add a term of the form

$$c_0 \left(\Pi^{\mathrm{p}}_{\mathcal{D}}(p_{\mathcal{D}}^{(n+1)} - p_{\mathcal{D}}^{(n)}), \Pi^{\mathrm{p}}_{\mathcal{D}} q_{\mathcal{D}}\right)_{0,\Omega}$$

in the left-hand side of the flow equation (V.11b). As we explained in Section II.2.2, this additional term increases the stability of the model as it gives a $L^\infty(0,T; L^2_0(\Omega))$ control on the approximate pressure reconstruction that does not depend on $\underline{\kappa}^{-1}$ and which does not hinge on an inf-sup condition. When considering a $\mathbb{P}_1^d/\mathbb{P}_1$ approximation of the poroelasticity model, we integrate this additional term with a quadrature rule exact for polynomials of degree 2.

V.4.2 Stabilization of the pore pressure approximation

We here investigate the influence of the constrained specific storage coefficient c_0 and of time on the quality of the pore pressure approximation given by the CRg-VS/HFV method. For that purpose, let $\Omega := (0,1)^2$, let $T > 0$ classicaly denote the simulation time, and let consider a homogeneous porous medium with (constant) Lamé parameters such that $\lambda = \mu = 1$ and (constant) sufficiently large permeability such that the mobility satisfies $\kappa = 1$. We recall that $\alpha = 1$. We consider the following manufactured solution:

$$u_x = e^{-t} x^2 y, \qquad u_y = -e^{-t} x y^2, \qquad p = e^{-t} \sin(x/\sqrt{2}) \sin(y/\sqrt{2}).$$

The volumetric body force $\boldsymbol{f}$ and the source term h are obtained by plugging the above solution into (II.18a) and (II.18b) respectively:

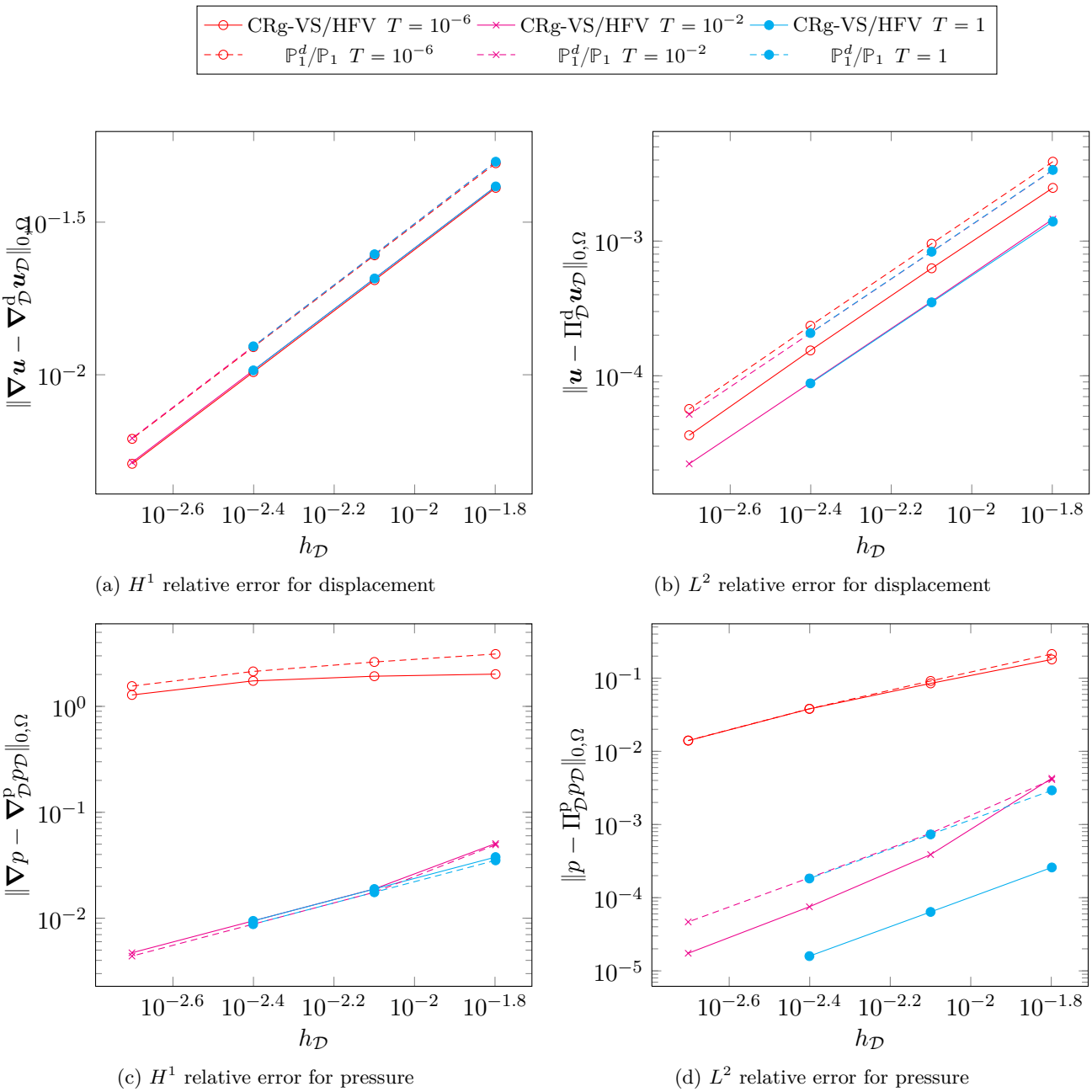
$$\begin{aligned} f_x &= \frac{\alpha}{\sqrt{2}}\, e^{-t} \cos(x/\sqrt{2}) \sin(y/\sqrt{2}) - 2\mu\, e^{-t} y, \\ f_y &= \frac{\alpha}{\sqrt{2}}\, e^{-t} \sin(x/\sqrt{2}) \cos(y/\sqrt{2}) + 2\mu\, e^{-t} x, \\ h &= (\kappa - c_0) p. \end{aligned}$$

We consider the matching triangular mesh sequence of Figure IV.1a. For $c_0 \in \{0, 1\}$, we plot on Figures V.1 and V.2 respectively, the H^1 and L^2 relative errors for displacement (computed as in (V.55)) and pressure (computed as in (V.56))

(i) at $T = 10^{-6}$ using one time step;

(ii) at $T = 10^{-2}$ using a time discretization sequence with time steps such that $\delta t_0 = 10^{-2}$ ($h_{\mathcal{D}_0} \approx 10^{-2}$) and $\delta t_m = \delta t_0 / 4^m$;

(iii) at $T = 1$ using a time discretization sequence with time steps such that $\delta t_0 = 10^{-2}$ and $\delta t_m = \delta t_0 / 4^m$.



(a) H^1 relative error for displacement

(b) L^2 relative error for displacement

(c) H^1 relative error for pressure

(d) L^2 relative error for pressure

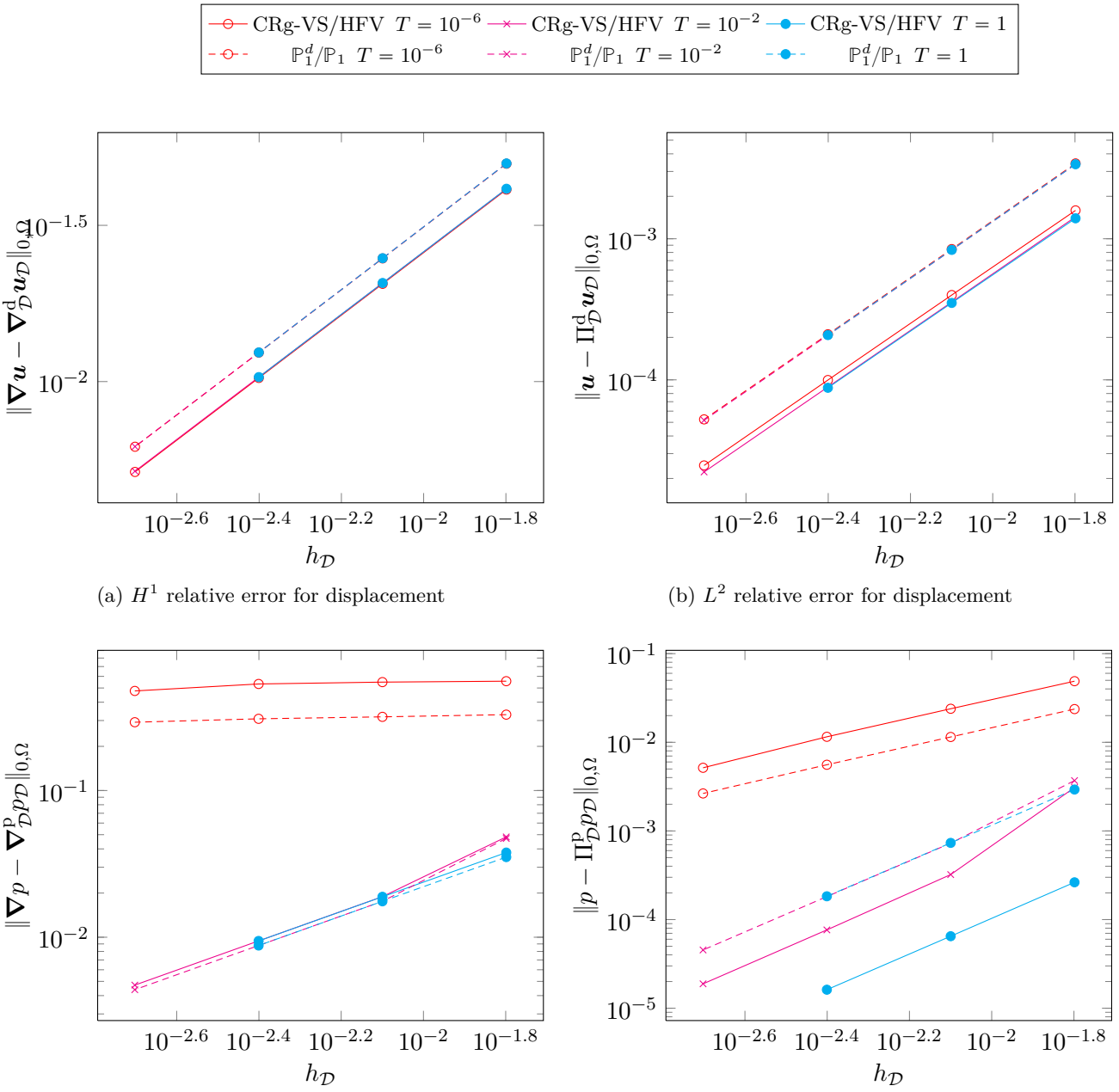
Figure V.1: Time effect on the stabilization of the pore pressure approximation for $c_0 = 0$, CRg-VS/HFV (solid lines) vs. $\mathbb{P}_1^d/\mathbb{P}_1$ (dashed lines).

We compare the results with the conforming $\mathbb{P}_1^d/\mathbb{P}_1$ finite element method.

Concerning the displacement approximation, the results are insensitive to the value of c_0 (which is not surprising) and quasi-insensitive to time. The errors in the L^2 norm and H^1 seminorm are globally better for the CRg-VS/HFV method than for the $\mathbb{P}_1^d/\mathbb{P}_1$ method.

Concerning the pore pressure approximation, some comments are in order. First note that the artefact observed between the first and the second member of the mesh sequence for both methods at $T = 10^{-2}$ only indicates that the first time step is not optimal.

For $c_0 = 0$, the approximate pressure gradient does not converge for both methods in the early time $T = 10^{-6}$. In the L^2 norm, the approximate pressure converges with order one for



(a) H^1 relative error for displacement

(b) L^2 relative error for displacement

(c) H^1 relative error for pressure

(d) L^2 relative error for pressure

Figure V.2: Time effect on the stabilization of the pore pressure approximation for $c_0 = 1$, CRg-VS/HFV (solid lines) vs. $\mathbb{P}_1^d/\mathbb{P}_1$ (dashed lines).

both methods. Hence, all happens just like if the pressure belonged to $L^2(\Omega)$. As the grid is refined, the approximate pressure gradient begins to converge and the L^2 relative error tends to converge with order more than one. At $T = 10^{-2}$ or $T = 1$, we observe the stabilization effect of the Darcean diffusion term on the pressure approximation. For both methods, the pressure gradient now converges at order one, and the reconstruction at order two. We note that the stabilization effect is stronger on the CRg-VS/HFV method since the results in the L^2 norm keep improving between $T = 10^{-2}$ and $T = 1$. For long times, the pressure reconstruction obtained with the CRg-VS/HFV method is more precise than the one obtained with the $\mathbb{P}_1^d/\mathbb{P}_1$ method.

For $c_0 = 1$, we observe the stabilization effect of that parameter. For early times, the approximate pressure gradient still does not converge for both methods but the error is less important than for $c_0 = 0$. The pressure reconstruction converges with order one and the errors are also less important. We can notice that the stabilization effect of c_0 is stronger for the $\mathbb{P}_1^d/\mathbb{P}_1$ method, which is not surprising since the reconstruction of this latter is gradient-based. We see for early times that the pressure gradient has a less convergent behavior than for the case $c_0 = 0$. This is due to the fact that, here, the term depending on c_0 has much more weight than the Darcean term, which means that all happens just like if the pressure exclusively belonged to $L^2(\Omega)$. For longer times, the stabilizing effect of c_0 is exceeded by the one of the Darcean term and results are almost similar to those obtained with $c_0 = 0$.

It is important noticing that, globally, the CRg-VS/HFV method is not really more stable in early times than the $\mathbb{P}_1^d/\mathbb{P}_1$ method, based on an unstable finite element pair. The inf-sup condition fulfilled by the CRg-VS/HFV pair is not sufficient, as we began to explain in Sections II.2.3 and V.1.1.2, to ensure the absence of spurious spatial oscillations on the pressure approximation. Nevertheless, it has a theoretical interest since it allows to prove the strong convergence of the pressure approximation (independently of $\underline{\kappa}$). Without searching for a piecewise quadratic displacement field (which is very costly), it seems difficult to design an inf-sup stable (in the finite element sense) method for Biot's consolidation problem as the pressure must be at least piecewise affine. Stabilization techniques (which do not rely on an inf-sup condition) could be considered to treat the early times spurious oscillations issue. In the context of the CRg-VS/HFV method, a stabilization inspired from Aguilar et al. [2] could be considered as it only involves the pressure gradient. Note that this kind of stabilization is very strong since it directly applies on the pressure gradient, and not on the reconstruction as inf-sup stabilizations do. To consider dG-like stabilizations, meaning stabilizations of the pressure jumps, a notion of affine reconstruction would have to be defined for the pressure approximation space, meaning that the pressure would have to belong to the generalized Crouzeix–Raviart space too. This stabilization technique has been tested with success but not published yet by Daniele A. Di Pietro for dG methods. A perspective could be to try adapt it in our case.

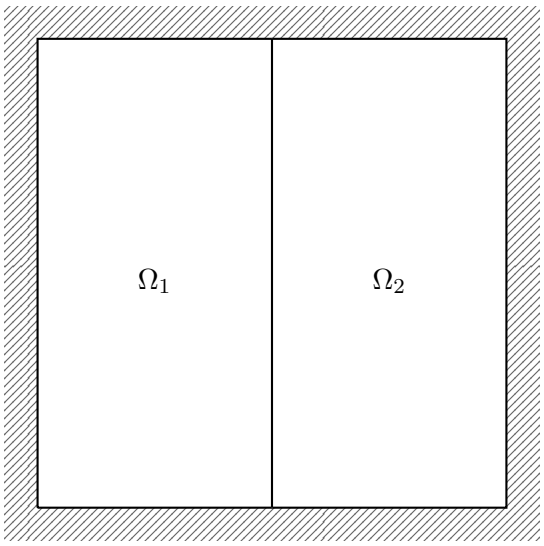
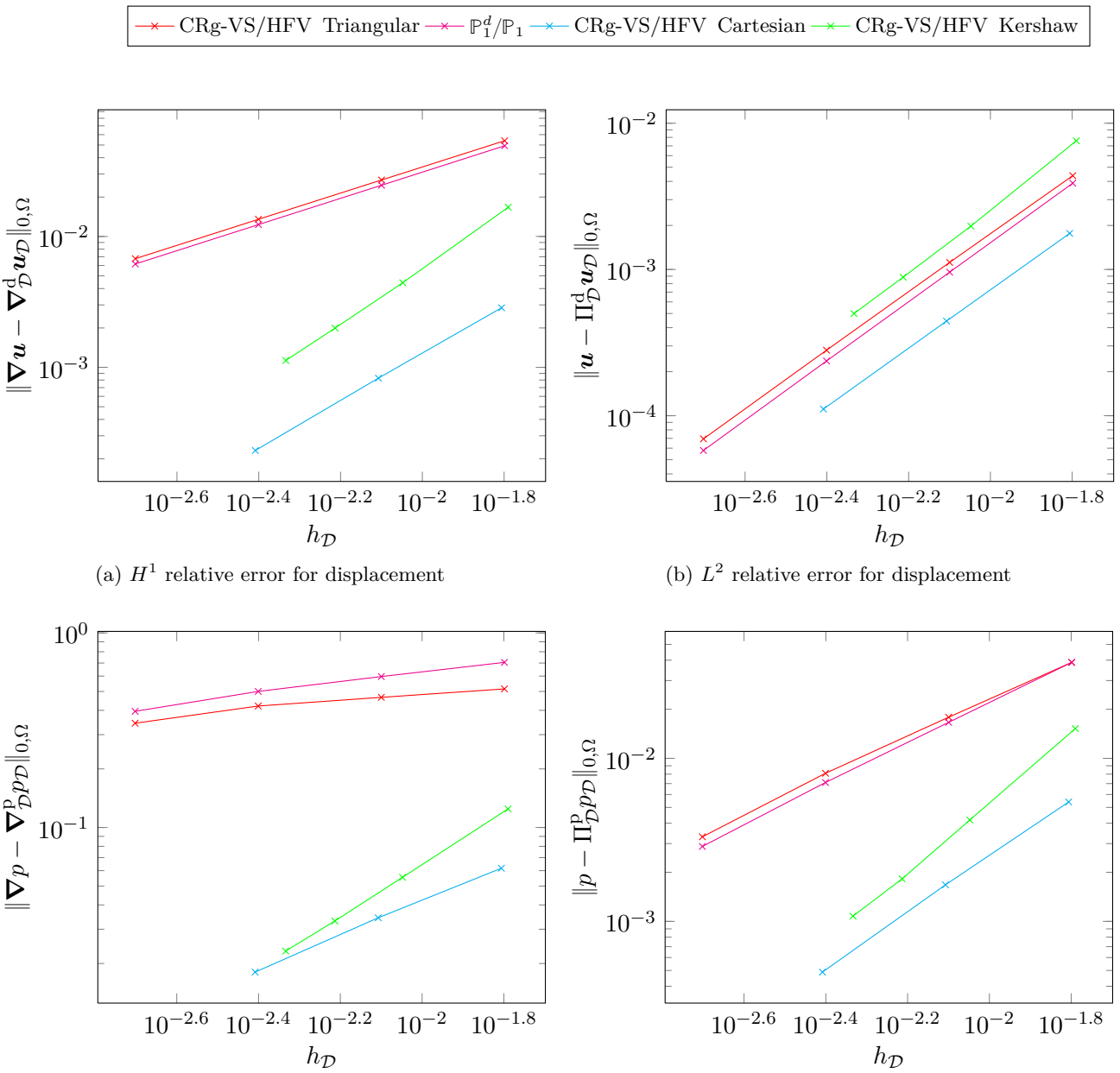


Figure V.3: Configuration of the heterogeneous test-case of Section V.4.3.



(a) H^1 relative error for displacement

(b) L^2 relative error for displacement

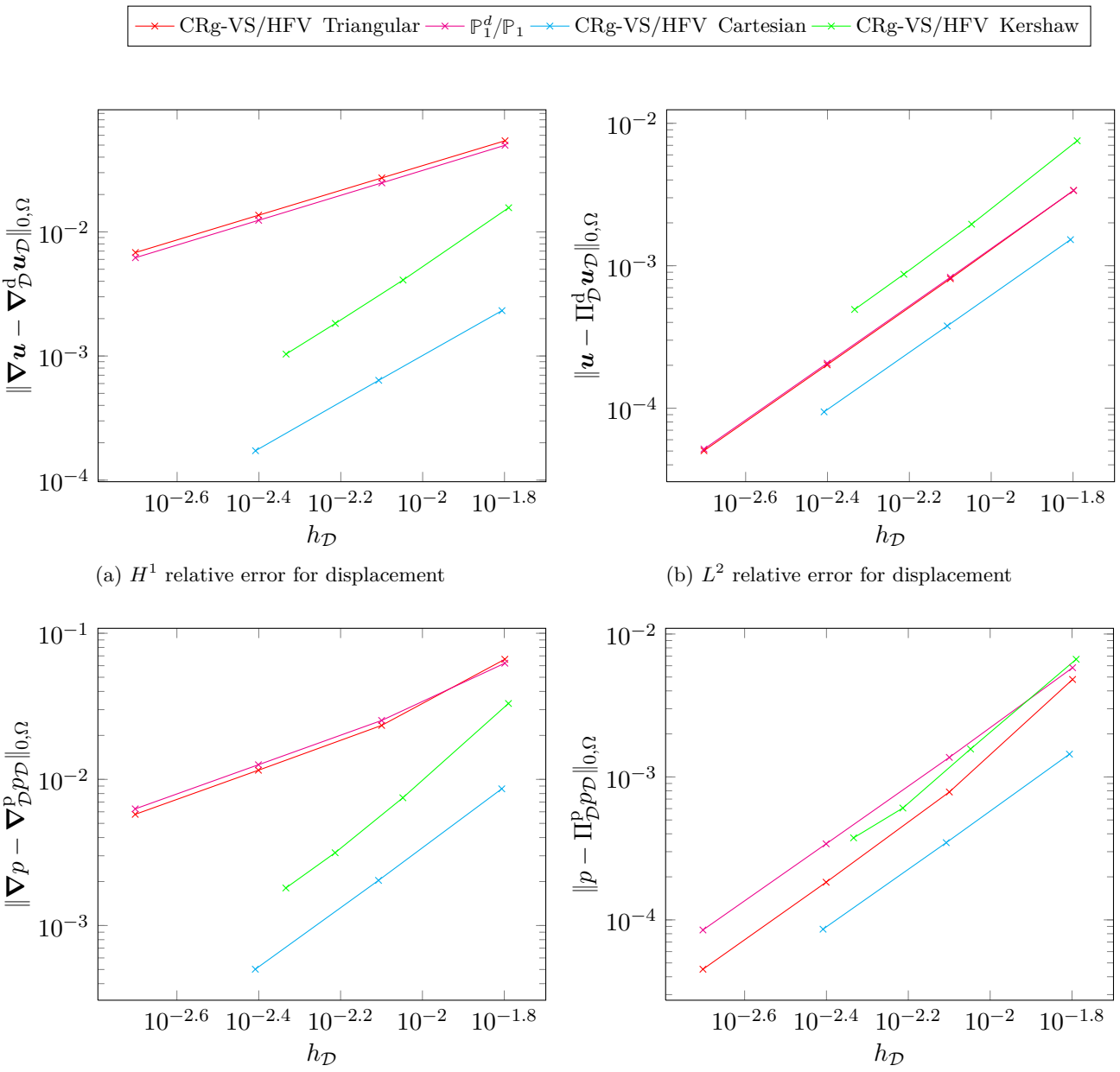
(c) H^1 relative error for pressure

(d) L^2 relative error for pressure

Figure V.4: Robustness of the discretization CRg-VS/HFV on challenging grids at $T = 10^{-6}$ for $\varepsilon = 10^{-1}$ vs. $\mathbb{P}_1^d/\mathbb{P}_1$.

V.4.3 Heterogeneous porous medium, low permeability and challenging grids

We investigate in this section the effect of poorly permeable regions in the porous medium on the approximation of the pore pressure by the CRg-VS/HFV method on potentially challenging grids. For that purpose, let consider the following manufactured solution. Let $\Omega := (0,1)^2$ such that $\overline{\Omega} := \overline{\Omega_1} \cup \overline{\Omega_2}$, where $\Omega_1 := \left(0, \frac{1}{2}\right) \times (0,1)$ and $\Omega_2 := \left(\frac{1}{2}, 1\right) \times (0,1)$. We recall that $\alpha = 1$ and that $T > 0$ denotes the simulation time, and we here assume $c_0 = 0$. We consider a porous medium with constant Lamé parameters $\lambda = \mu = 1$ and piecewise constant permeability such



(a) H^1 relative error for displacement

(b) L^2 relative error for displacement

(c) H^1 relative error for pressure

(d) L^2 relative error for pressure

Figure V.5: Robustness of the discretization CRg-VS/HFV on challenging grids at $T = 10^{-2}$ for $\varepsilon = 10^{-1}$ vs. $\mathbb{P}_1^d/\mathbb{P}_1$.

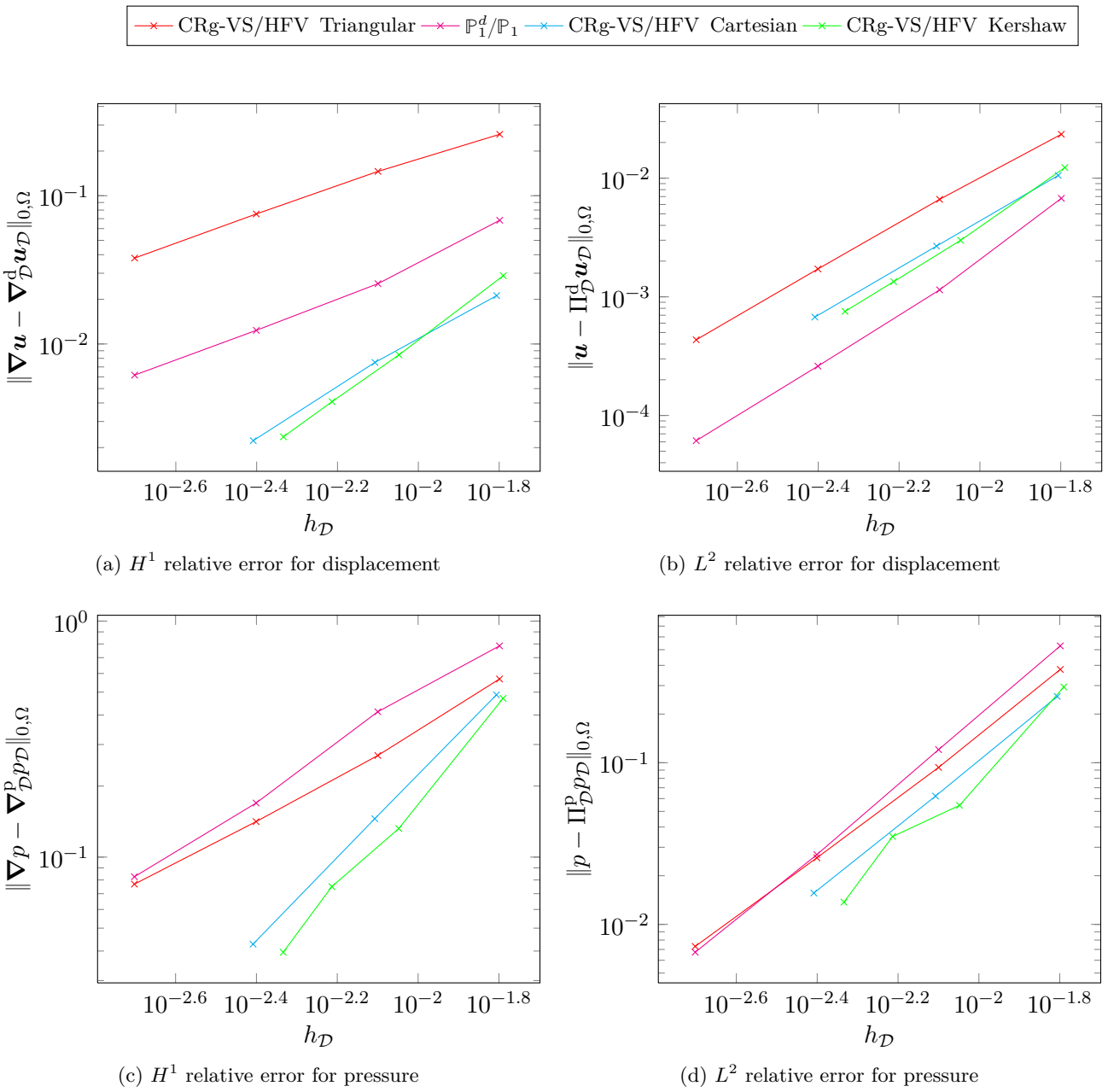
that the mobility field satisfies

$$\kappa = \varepsilon \quad \text{in } \Omega_1, \qquad \kappa = 1 \quad \text{in } \Omega_2,$$

where $\varepsilon > 0$ allows to vary the permeability contrast, the case $\varepsilon = 1$ corresponding to a homogeneous medium. An illustration of the geometry is provided in Figure V.3.

For this medium, we consider the following solution:

$$u_x = e^{-t}x^2y, \qquad u_y = -e^{-t}xy^2, \qquad p = \begin{cases} e^{-t}\cos(x - \frac{1}{2}) & \text{if } x > \frac{1}{2}, \\ e^{-t}\cos((x - \frac{1}{2})/\sqrt{\varepsilon}) & \text{otherwise.} \end{cases}$$



(a) H^1 relative error for displacement

(b) L^2 relative error for displacement

(c) H^1 relative error for pressure

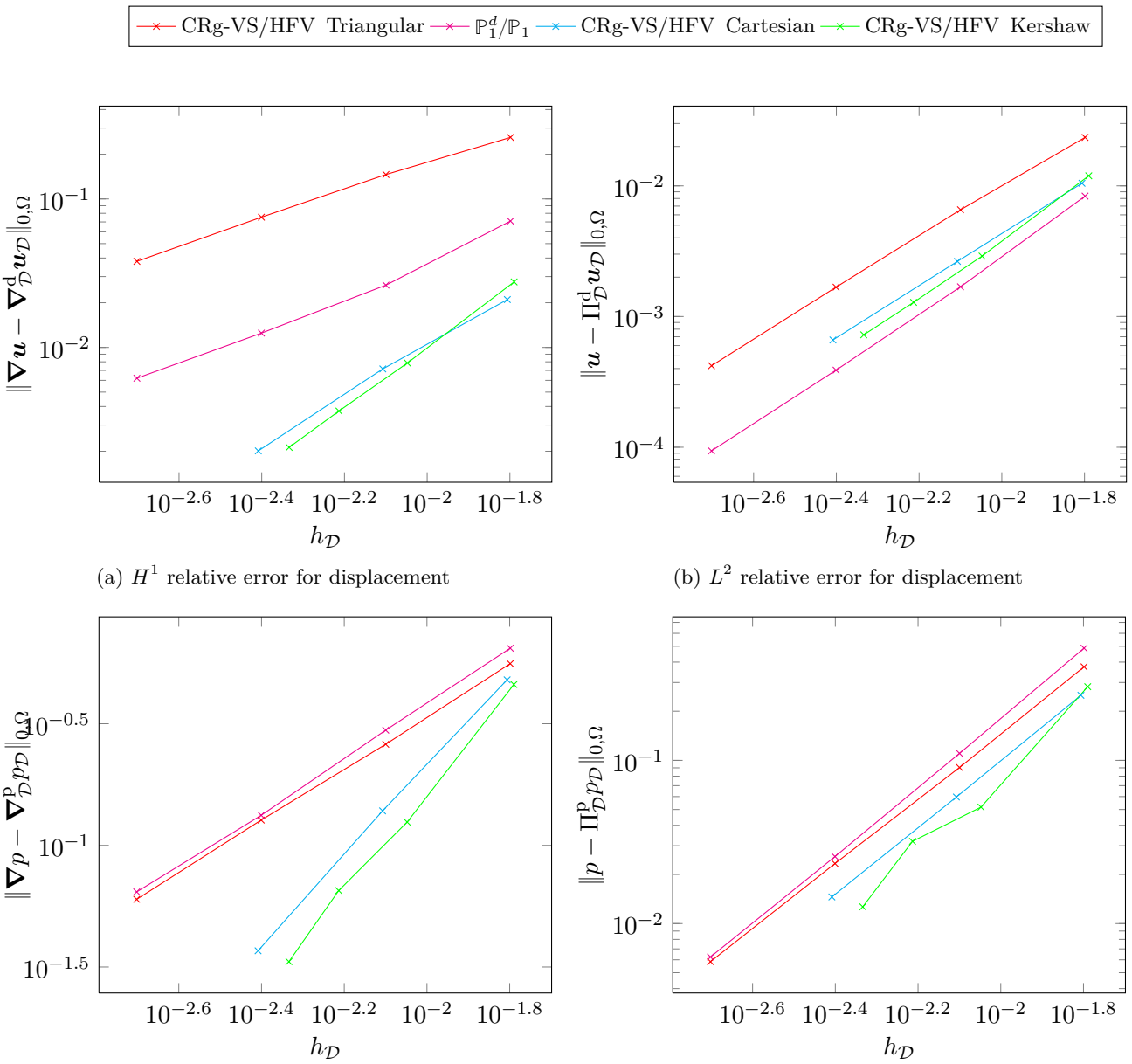
(d) L^2 relative error for pressure

Figure V.6: Robustness of the discretization CRg-VS/HFV on challenging grids at $T = 10^{-6}$ for $\varepsilon = 10^{-3}$ vs. $\mathbb{P}_1^d/\mathbb{P}_1$.

This solution is continuous on Ω, with continuous pressure gradient and pressure flux. The volumetric body force $\boldsymbol{f}$ and the source term h are obtained by plugging the solution into (II.18a) and (II.18b) respectively:

$$
\begin{aligned}
f_x &= \begin{cases} -\alpha\, e^{-t} \sin(x - \tfrac{1}{2}) - 2\mu\, e^{-t} y & \text{if } x > \tfrac{1}{2}, \\ -\tfrac{\alpha}{\sqrt{\varepsilon}}\, e^{-t} \sin((x - \tfrac{1}{2})/\sqrt{\varepsilon}) - 2\mu\, e^{-t} y & \text{otherwise,} \end{cases} \\
f_y &= 2\mu\, e^{-t} x, \\
h &= (1 - c_0) p.
\end{aligned}
$$

We consider the matching triangular (cf. Figure IV.1a), the Cartesian (cf. Figure IV.1b) and



(a) H^1 relative error for displacement

(b) L^2 relative error for displacement

(c) H^1 relative error for pressure

(d) L^2 relative error for pressure

Figure V.7: Robustness of the discretization CRg-VS/HFV on challenging grids at $T = 10^{-2}$ for $\varepsilon = 10^{-3}$ vs. $\mathbb{P}_1^d/\mathbb{P}_1$.

the Kershaw-type (cf. Figure IV.1d) mesh families. Note that the Cartesian and matching triangular mesh sequences match the heterogeneities of the medium, which is not the case of the Kershaw-type sequence. For a permeability contrast such that $\varepsilon \in \{10^{-1}, 10^{-3}\}$, and a final time $T \in \{10^{-6}, 10^{-2}\}$ (using one time step for the first, a time discretization sequence with uniform time steps such that $\delta t_0 = 10^{-2}$ and $\delta t_m = \delta t_0/4^m$ for the second), we plot on Figures V.4, V.5, and V.6, V.7 respectively, the H^1 and L^2 relative errors for displacement (computed as in (V.55)) and pressure (computed as in (V.56))

(i) for the CRg-VS/HFV method on the matching triangular sequence;

(ii) for the $\mathbb{P}_1^d/\mathbb{P}_1$ method on the matching triangular sequence;

(iii) for the CRg-VS/HFV method on the Cartesian sequence;

(iv) for the CRg-VS/HFV method on the Kershaw-type sequence.

For $\varepsilon = 10^{-1}$, the results are pretty similar to those obtained for the homogeneous test-case of Section V.4.2. Concerning the displacement approximation, the results are quasi-insensitive to time. We here notice that the errors exactly compare for both the CRg-VS/HFV and the $\mathbb{P}_1^d/\mathbb{P}_1$ methods on the matching triangular mesh sequence. On Cartesian and Kershaw-type grids, we observe a supra-convergent behavior in the H^1 seminorm. Concerning the pressure approximation, we here again observe the stabilizing effect of the Darcean term for sufficiently large times. In early times, the pressure gradient poorly converges, with order less than one but with some improvement as the grid refines, and the reconstruction with order approximately one, also tending to more as the grid refines. For sufficiently large times, the optimal orders of convergence are reached in the L^2 norm and H^1 seminorm. Supra-convergence is observed on Cartesian and Kershaw-type grids for the CRg-VS/HFV method.

In the case $\varepsilon = 10^{-3}$, the results degenerate. First of all, we notice a clear deterioration of the displacement approximation for the CRg-VS/HFV method. Concerning the pressure approximation, even in early times, both the pressure gradient and its reconstruction surprisingly converge with optimal order (a supra-convergent behavior is even observed for the Cartesian and Kershaw-type grids). These results are insensitive to time and no real difference can be noticed comparing times $T = 10^{-6}$ and $T = 10^{-2}$. Hence, the diffusion term does not have any stabilizing effect as time goes on. Even worse, the relative errors deteriorate (first for the pressure, then impacting on displacement) for both methods for large times (not plotted here). Note however that, in accordance with the convergence result of Theorem V.1, for a given T, the pressure reconstruction converges as the mesh and the time step refine. The same experiment led with an inverse mobility contrast (i.e. $\varepsilon = 10^3$) gives satisfactory results. The problem thus comes from the presence of a poorly permeable region, and here again relies on a lack of stabilization of pressure, which drives in that case to a divergent behavior for long times as the errors sum. It suggests that an efficient stabilization of the pressure approximation must be proportional to the inverse of the smaller permeability (see perspectives in Chapter VI).

Chapitre VI

Perspectives futures

Sommaire

Le bilan des développements de ce manuscrit étant présenté en introduction Section I.2, nous faisons en guise de conclusion un état des lieux des perspectives futures envisageables à ce travail, en dissociant les aspects de recherche sur les schémas, de complexification des modèles, et de validation industrielle.

VI.1 Recherche sur les schémas

Comme nous l'avons vu au Chapitre VI, une approximation du problème de poroélasticité basée sur des espaces de déplacement et de pression discrets tous deux affines par morceaux ne permet pas, sans stabilisation adéquate, d'assurer une approximation de la pression satisfaisante et sans oscillations parasites pour les temps courts ou dans les zones très peu perméables. Vérifier une condition inf-sup au sens volumes finis (c'est à dire avoir une estimation sur une projection de la pression affine) ne suffit pas à stabiliser le modèle. Les différentes pistes envisagées et pas encore testées sont les suivantes :

- monter en ordre sur l'espace de discrétisation du déplacement afin de vérifier une condition inf-sup au sens éléments finis avec l'approximation affine de la pression. Cette méthode est exclue d'office car trop coûteuse ;
- ajouter un terme de stabilisation de pression inspiré de Aguilar *et al.* [2] qui fait intervenir la dérivée en temps du Laplacien de pression et qui assure donc un contrôle de son gradient pour tous temps, avec une constante de proportionnalité dépendant du paramètre de maillage ; cette méthode semble donner de très bons résultats mais a l'inconvénient de dénaturer le modèle physique. Dans notre cas, elle aurait l'avantage de ne pas nécessiter l'introduction d'une reconstruction affine pour la pression et d'être très simple à implémenter ;
- ajouter un terme de stabilisation par les sauts comme l'a testé Daniele A. Di Pietro (non publié) dans le cadre dG ; ce terme que l'on prend inversement proportionnel à la plus petite perméabilité du milieu permet de stabiliser les zones peu perméables et les premiers pas de temps ; dans notre cas, l'application de cette méthode nécessite l'introduction d'une reconstruction affine pour la pression, ce qui peut se faire sans peine grâce à l'espace introduit au Chapitre III ;
- traiter l'écoulement grâce à une méthode mixte comme le proposent Phillips et Wheeler [67]. Cette méthode a la particularité de permettre une discrétisation constante par morceaux de la pression (puisque le flux est discrétisé séparément) qui est inf-sup stable au sens élements finis lorsqu'elle est couplée à une discrétisation affine discontinue du déplacement (dG par exemple ou dans notre cas appartenant à l'espace Crouzeix–Raviart généralisé). Le principal inconvénient de cette méthode est qu'elle nécessite l'ajout d'un autre problème mixte en plus du problème de point-selle déjà considéré, ce qui ajoute des inconnues au système et complexifie encore sa résolution.

VI.2 Complexification des modèles

Un autre angle d'amélioration et de poursuite future concerne la complexification des modèles physiques. Cette complexification peut être liée à la mécanique ou à l'écoulement darcéen.

Concernant la mécanique, le premier pas serait de considérer des lois de Hooke plus générales pour l'élasticité, avec tenseur de raideur (admissible) d'ordre 4. L'adaptation des méthodes proposées dans ce manuscrit à ce cas est immédiate. Un deuxième pas serait sans doute de considérer des modèles d'élasticité non-linéaire, puis d'introduire des lois de comportement plus compliquées, comme des modèles d'élastoplasticité avec écrouissage ou de viscoélastoplasticité. Un modèle d'élastoplasticité a commencé à être étudié durant cette thèse mais les résultats n'ont pas été concrétisés dans le temps imparti. Un dernier pas serait sûrement de considérer des modèles de fractures, avant d'avoir une représentation mécanique adaptée à la modélisation géologique.

Concernant l'écoulement, le premier pas serait de considérer un modèle diphasique immiscible car la convergence de ce dernier (non couplé à un modèle mécanique) a été récemment étudiée dans le cadre des schémas Gradient par Eymard *et al.* [44]. Ensuite, l'étape d'après serait de considérer des écoulements polyphasiques compositionnels, dont l'approximation par le schéma Vertex Approximate Gradient (VAG) a été étudiée dans [42, 43]. Le schéma VAG fait partie des schémas Gradient.

Il reste donc beaucoup de travail à faire avant de pouvoir traiter un modèle réaliste de poromécanique.

VI.3 Validation industrielle

D'un point de vue industriel, les perspectives futures sont claires et consistent à passer en trois dimensions d'espace. Les méthodes numériques présentées dans ce manuscrit sont conçues pour fonctionner en 2D et 3D mais n'ont été pour la plupart (excepté en Annexe C) testées qu'en 2D. Il convient donc de réaliser sur des cas réalistes 3D simplifiés une validation des méthodes, et une comparaison avec des éléments finis (avec remaillage local le cas échéant) pour la mécanique. Les comparaisons doivent prendre en compte le nombre d'inconnues, le conditionnement et le remplissage des matrices, ainsi que le type et la complexité des solveurs utilisés.

Bibliographie

[1] L. Agélas, R. Eymard, and S. Lemaire. Convergence of Euler-Gradient approximations of Biot's consolidation problem on general meshes. 2014. In preparation.

[2] G. Aguilar, F. Gaspar, F. Lisbona, and C. Rodrigo. Numerical stabilization of Biot's consolidation model by a perturbation on the flow equation. *Int. J. Numer. Methods Engrg.*, 75 :1282–1300, 2008.

[3] G. Allaire. *Analyse numérique et optimisation*. Les éditions de l'École Polytechnique, Palaiseau, 2009.

[4] D. N. Arnold, F. Brezzi, and J. Douglas. PEERS : A new mixed finite element for plane elasticity. *Japan Journal of Industrial and Applied Mathematics*, 1(2) :347–367, 1984.

[5] J.-L. Auriault and E. Sanchez-Palencia. Étude du comportement macroscopique d'un milieu poreux saturé déformable. *Journal de Mécanique*, 16 :576–603, 1977.

[6] F. Auricchio, L. Beirão da Veiga, A. Buffa, C. Lovadina, A. Reali, and G. Sangalli. A fully "locking-free" isogeometric approach for plane linear elasticity problems : a stream function formulation. *Comput. Methods Appl. Mech. Engrg.*, 197(1–4) :160–172, 2007.

[7] L. Beirão da Veiga. A mimetic discretization method for linear elasticity. *M2AN Math. Model. Numer. Anal.*, 44(2) :231–250, 2010.

[8] L. Beirão da Veiga, F. Brezzi, and L. D. Marini. Virtual elements for linear elasticity problems. *SIAM J. Numer. Anal.*, 51(2) :794–812, 2013.

[9] L. Beirão da Veiga, V. Gyrya, K. Lipnikov, and G. Manzini. Mimetic finite difference method for the Stokes problem on polygonal meshes. *J. Comput. Phys.*, 228(19) :7215–7232, 2009.

[10] L. Beirão da Veiga, K. Lipnikov, and G. Manzini. Error analysis for a mimetic discretization for the steady Stokes problem on polyhedral meshes. *SIAM J. Numer. Anal.*, 48(4) :1419–1443, 2010.

[11] J. Berdal Haga, H. Osnes, and H. P. Langtangen. On the causes of pressure oscillations in low-permeable and low-compressible porous media. *Int. J. Numer. Anal. Methods Geomech.*, 36(12) :1507–1522, 2012.

[12] M. A. Biot. General theory of three-dimensional consolidation. *J. Appl. Phys.*, 12(2) :155–164, 1941.

[13] F. Boyer, S. Krell, and F. Nabet. Inf-sup stability of the Discrete Duality Finite Volume method for the 2D Stokes problem. 2013. Submitted. Preprint available at `http://hal.archives-ouvertes.fr/hal-00795362`.

[14] S. C. Brenner. Korn's inequalities for piecewise H^1 vector fields. *Math. Comp.*, 73(247) :1067–1087, 2004.

[15] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[16] S. C. Brenner and L.-Y. Sung. Linear finite element methods for planar linear elasticity. *Math. Comp.*, 59(200) :321–338, 1992.

[17] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*, volume 15 of *Springer Series in Computational Mathematics*. Springer-Verlag, New York, 1991.

[18] F. Brezzi, K. Lipnikov, and M. Shashkov. Convergence of the mimetic finite difference method for diffusion problems on polyhedral meshes. *SIAM J. Numer. Anal.*, 43(5) :1872–1896, 2005.

[19] F. Brezzi, K. Lipnikov, M. Shashkov, and V. Simoncini. A new discretization methodology for diffusion problems on generalized polyhedral meshes. *Comput. Methods Appl. Mech. Engrg.*, 196(37–40) :3682–3692, 2007.

[20] F. Brezzi, K. Lipnikov, and V. Simoncini. A family of mimetic finite difference methods on polygonal and polyhedral meshes. *Math. Mod. Meths. Appli. Sci. (M3AS)*, 15(10) :1533–1551, 2005.

[21] K. S. Chavan, B. P. Lamichhane, and B. I. Wohlmuth. Locking-free finite element methods for linear and nonlinear elasticity in 2D and 3D. *Comput. Methods Appl. Mech. Engrg.*, 196 :4075–4086, 2007.

[22] M. Crouzeix and P.-A. Raviart. Conforming and nonconforming finite element methods for solving the stationary Stokes equations. *RAIRO Modél. Math. Anal. Num.*, 7(3) :33–75, 1973.

[23] D. A. Di Pietro. Cell centered Galerkin methods for diffusive problems. *M2AN Math. Model. Numer. Anal.*, 46(1) :111–144, 2012.

[24] D. A. Di Pietro and A. Ern. *Mathematical aspects of discontinuous Galerkin methods*, volume 69 of *Mathématiques & Applications*. Springer-Verlag, Berlin, 2011.

[25] D. A. Di Pietro, R. Eymard, S. Lemaire, and R. Masson. Hybrid finite volume discretization of linear elasticity models on general meshes. In *Finite Volumes for Complex Applications VI Problems & Perspectives*, volume 4, pages 331–339, Prague, 2011. Springer–Verlag.

[26] D. A. Di Pietro and J.-M. Gratien. Lowest order methods for diffusive problems on general meshes : a unified approach to definition and implementation. In *Finite Volumes for Complex Applications VI Problems & Perspectives*, volume 4, pages 803–819, Prague, 2011. Springer–Verlag.

[27] D. A. Di Pietro, J.-M. Gratien, and C. Prud'homme. A domain-specific embedded language in C++ for lowest-order discretizations of diffusive problems on general meshes. *BIT Numerical Mathematics*, 53(1) :111–152, 2013.

[28] D. A. Di Pietro and S. Lemaire. An extension of the Crouzeix–Raviart space to general meshes with application to quasi-incompressible linear elasticity and Stokes flow. *Math. Comp.*, 2013. Accepted for publication. Preprint available at `http://hal.archives-ouvertes.fr/hal-00753660`.

[29] D. A. Di Pietro and S. Nicaise. A locking-free discontinuous Galerkin method for linear elasticity in locally nearly incompressible heterogeneous media. *App. Num. Math.*, 63 :105–116, 2013.

[30] J. Droniou and R. Eymard. A mixed finite volume scheme for anisotropic diffusion problems on any grid. *Numer. Math.*, 105(1) :35–71, 2006.

[31] J. Droniou and R. Eymard. Study of the mixed finite volume method for Stokes and Navier–Stokes equations. *Numer. Methods for Partial Differential Equations*, 25(1) :137–171, 2009.

[32] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin. A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods. *Math. Mod. Meths. Appli. Sci. (M3AS)*, 20(2) :265–295, 2010.

[33] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin. Gradient schemes : a generic framework for the discretization of linear, nonlinear and nonlocal elliptic and parabolic equations. *Math. Mod. Meths. Appli. Sci. (M3AS)*, 23(13) :2395–2432, 2013.

[34] T. Dupont and R. Scott. Polynmomial approximation of functions in Sobolev spaces. *Math. Comp.*, 34(150) :441–463, 1980.

[35] A. Ern and J.-L. Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[36] A. Ern and S. Meunier. A posteriori error analysis of Euler-Galerkin approximations to coupled elliptic-parabolic problems. *M2AN Math. Model. Numer. Anal.*, 43(2) :353–375, 2009.

[37] R. Eymard, P. Féron, T. Gallouët, C. Guichard, and R. Herbin. Gradient schemes for the Stefan problem. *Int. J. Finite Vol.*, 10, 2013.

[38] R. Eymard, T. Gallouët, and R. Herbin. *The finite volume method*, volume 7 of *Handbook of Numerical Analysis*. P. G. Ciarlet and J.-L. Lions eds., North Holland, 2000.

[39] R. Eymard, T. Gallouët, and R. Herbin. A new finite volume scheme for anisotropic diffusion problems on general grids : convergence analysis. *C. R. Acad. Sci. Paris, Ser. I*, 344 :403–406, 2007.

[40] R. Eymard, T. Gallouët, and R. Herbin. Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes. SUSHI : a scheme using stabilization and hybrid interfaces. *IMA J. Num. Anal.*, 30(4) :1009–1043, 2010.

[41] R. Eymard, C. Guichard, and R. Herbin. Small-stencil 3D schemes for diffusive flows in porous media. *M2AN Math. Model. Numer. Anal.*, 46(2) :265–290, 2012.

[42] R. Eymard, C. Guichard, R. Herbin, and R. Masson. Multiphase flow in porous media using the VAG scheme. In *Finite Volumes for Complex Applications VI Problems & Perspectives*, volume 4, pages 409–417, Prague, 2011. Springer–Verlag.

[43] R. Eymard, C. Guichard, R. Herbin, and R. Masson. Vertex-centered dicretization of multiphase compositional Darcy flows on general meshes. *Comput. Geosci.*, 16(4) :987–1005, 2012.

[44] R. Eymard, C. Guichard, R. Herbin, and R. Masson. Gradient schemes for two-phase flow in heterogeneous porous media and Richards equation. *ZAMM - J. Appl. Math. Mech.*, 2013. Published online. DOI 10.1002/zamm.201200206.

[45] R. Eymard and R. Herbin. Gradient scheme approximations for diffusion problems. In *Finite Volumes for Complex Applications VI Problems & Perspectives*, volume 4, pages 439–447, Prague, 2011. Springer–Verlag.

[46] R. S. Falk. Nonconforming finite element methods for the equations of linear elasticity. *Math. Comp.*, 57(196) :529–550, 1991.

[47] K. J. Galvin, A. Linke, L. G. Rebholz, and N. E. Wilson. Stabilizing poor mass conservation in incompressible flow problems with large irrotational forcing and application to thermal convection. *Comput. Methods Appl. Mech. Engrg.*, 237–240 :166–176, 2012.

[48] V. Girault and P.-A. Raviart. *Finite element methods for Navier–Stokes equations*, volume 5 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1986. Theory and algorithms.

[49] P. Hansbo and M. G. Larson. Discontinuous Galerkin methods for incompressible and nearly incompressible elasticity by Nitsche's method. *Comput. Methods Appl. Mech. Engrg.*, 191(17–18) :1895–1908, 2002.

[50] P. Hansbo and M. G. Larson. Discontinuous Galerkin and the Crouzeix–Raviart element : application to elasticity. *M2AN Math. Model. Numer. Anal.*, 37(1) :63–72, 2003.

[51] J. G. Heywood and R. Rannacher. Finite element approximation of the nonstationary Navier–Stokes problem. Part IV : Error analysis for second-order time discretization. *SIAM J. Numer. Anal.*, 27(2) :353–384, 1990.

[52] J. Kim, H. A. Tchelepi, and R. Juanes. Stability, accuracy, and efficiency of sequential methods for coupled flow and geomechanics. *SPE Journal*, 16(2) :249–262, 2011.

[53] J. Korsawe and G. Starke. A least-squares mixed finite element method for Biot's consolidation problem in porous media. *SIAM J. Numer. Anal.*, 43(1) :318–339, 2005.

[54] J. Korsawe, G. Starke, W. Wang, and O. Kolditz. Finite element analysis of poro-elastic consolidation in porous media : Standard and mixed approaches. *Comput. Methods Appl. Mech. Engrg.*, 195(9–12) :1096–1115, 2006.

[55] S. Krell. Stabilized DDFV schemes for Stokes problem with variable viscosity on general 2D meshes. *Numer. Methods for Partial Differential Equations*, 27(6) :1666–1706, 2011.

[56] S. Krell and G. Manzini. The Discrete Duality Finite Volume method for Stokes equations on three-dimensional polyhedral meshes. *SIAM J. Numer. Anal.*, 50(2) :808–837, 2012.

[57] B. P. Lamichhane and E. P. Stephan. A symmetric mixed finite element method for nearly incompressible elasticity based on biorthogonal systems. *Numer. Methods for Partial Differential Equations*, 28(4) :1336–1353, 2012.

[58] K. Mahesh, G. Constantinescu, and P. Moin. A numerical method for large-eddy simulation in complex geometries. *J. Comput. Phys.*, 197 :215–240, 2004.

[59] A. Mikelić and M. F. Wheeler. Convergence of iterative coupling for coupled flow and geomechanics. *Comput. Geosci.*, 17(3) :455–462, 2013.

[60] M. A. Murad and A. F. D. Loula. Improved accuracy in finite element analysis of Biot's consolidation problem. *Comput. Methods Appl. Mech. Engrg.*, 95(3) :359–382, 1992.

[61] M. A. Murad and A. F. D. Loula. On stability and convergence of finite element approximations of Biot's consolidation problem. *Int. J. Numer. Methods Engrg.*, 37(4) :645–667, 1994.

[62] M. A. Murad, V. Thomée, and A. F. D. Loula. Asymptotic behavior of semidiscrete finite-element approximations of Biot's consolidation problem. *SIAM J. Numer. Anal.*, 33(3) :1065–1083, 1996.

[63] J. Nečas. Équations aux dérivées partielles. *Les Presses de l'Université de Montréal*, 1966.

[64] J. M. Nordbotten. Cell-centered finite volume discretizations for deformable porous media. *Int. J. Numer. Methods Engrg.*, 2013. Submitted.

[65] P. J. Phillips and M. F. Wheeler. A coupling of mixed and continuous Galerkin finite element methods for poroelasticity I : the continuous in time case. *Comput. Geosci.*, 11 :131–144, 2007.

[66] P. J. Phillips and M. F. Wheeler. A coupling of mixed and continuous Galerkin finite element methods for poroelasticity II : the discrete-in-time case. *Comput. Geosci.*, 11 :145–158, 2007.

[67] P. J. Phillips and M. F. Wheeler. A coupling of mixed and discontinuous Galerkin finite-element methods for poroelasticity. *Comput. Geosci.*, 12 :417–435, 2008.

[68] P. J. Phillips and M. F. Wheeler. Overcoming the problem of locking in linear elasticity and poroelasticity : an heuristic approach. *Comput. Geosci.*, 13 :5–12, 2009.

[69] D. K. Ponting. Corner point geometry in reservoir simulation. In *Proc. ECMOR I*, pages 45–65, Cambridge, 1989. In Clarendon Press, editor.

[70] A. Quarteroni and A. Valli. *Numerical approximation of partial differential equations*, volume 23 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1994.

[71] R. Rannacher and S. Turek. Simple nonconforming quadrilateral Stokes element. *Numer. Methods for Partial Differential Equations*, 8(2) :97–111, 1992.

[72] A. Settari and F. M. Mourits. Coupling of geomechanics and reservoir simulation models. In *Comput. Methods Adv. Geomech.*, pages 2151–2158, Rotterdam, The Netherlands, 1994. Balkema, Siriwardane & Zeman ed.

[73] G. Shaw and T. Stone. Finite volume methods for coupled stress/fluid flow in a commercial reservoir simulator. In *SPE Reservoir Simulation Symposium*, The Woodlands, Texas, 2005. SPE.

[74] R. E. Showalter. Diffusion in poro-elastic media. *J. Math. Anal. Appl.*, 251 :310–340, 2000.

[75] R. Stenberg. A family of mixed finite elements for the elasticity problem. *Numer. Math.*, 53 :513–538, 1988.

[76] G. Strang. Variational crimes in the finite element method. In A. Aziz, editor, *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*. Academic Press, New York, 1972.

[77] A. Ten Eyck and A. Lew. Discontinuous Galerkin methods for nonlinear elasticity. *Int. J. Numer. Methods Engrg.*, 67(9) :1204–1243, 2006.

[78] M. Vogelius. An analysis of the p-version of the finite element method for nearly incompressible materials. Uniformly valid, optimal error estimates. *Numer. Math.*, 41 :39–53, 1983.

[79] M. Vohralík and B. I. Wohlmuth. From face to element unknowns by local static condensation with application to nonconforming finite elements. *Comput. Methods Appl. Mech. Engrg.*, 253 :517–529, 2013.

[80] M. Vohralík and B. I. Wohlmuth. Mixed finite element methods : implementation with one unknown per element, local flux expressions, positivity, polygonal meshes, and relations to other methods. *Math. Mod. Meths. Appli. Sci. (M3AS)*, 23(5) :803–838, 2013.

[81] K. von Terzaghi. *Theoretical soil mechanics*. J. Wiley and Sons, New York, 1943.

[82] A. Ženíšek. The existence and uniqueness theorem in Biot's consolidation theory. *Aplik. Matem.*, 29(3) :194–211, 1984.

[83] J. Wan, L. J. Durlofsky, T. J. R. Hughes, and K. Aziz. Stabilized finite element methods for coupled geomechanics-reservoir flow simulations. In *SPE Reservoir Simulation Symposium*, Houston, Texas, 2003. SPE.

[84] M. F. Wheeler, G. Xue, and I. Yotov. Coupling multipoint flux mixed finite element methods with continuous Galerkin methods for poroelasticity. 2013. Submitted.

[85] J. P. Whiteley. Discontinuous Galerkin finite element methods for incompressible nonlinear elasticity. *Comput. Methods Appl. Mech. Engrg.*, 198(41–44) :3464–3478, 2009.

Annexe A

A generalized Raviart–Thomas space

Sommaire

This appendix is inspired from the article [28], written with Daniele A. Di Pietro and accepted for publication in Mathematics of Computation. In the spirit of Chapter III, we design a discrete space which can be seen as an extension to general polygonal or polyhedral meshes of the classical lowest-order Raviart–Thomas space. More precisely, this new space extends two classical properties of this latter, namely (i) the (full) continuity of normal components of discrete functions at interfaces ($\boldsymbol{H}(\mathrm{div};\Omega)$-conformity), and (ii) the existence of an interpolator which preserves the mean value of the divergence inside each element. Since the construction as well as the proofs are very similar to the ones presented in Chapter III, only the main points are detailed.

A.1 Construction

For a general mesh $\mathcal{K}_h$, belonging to an admissible mesh sequence in the sense of Definition III.3, we first introduce the broken polynomial space

$$\mathbb{RT}_d^0(\mathcal{K}_h) := \mathbb{P}_d^0(\mathcal{K}_h)^d + \boldsymbol{x}\mathbb{P}_d^0(\mathcal{K}_h).$$

For a matching simplicial mesh $\mathcal{T}_h$, the standard lowest-order Raviart–Thomas space is the subspace of $\boldsymbol{H}(\mathrm{div};\Omega)$ of functions belonging to $\mathbb{RT}_d^0(\mathcal{T}_h)$. To perform a similar construction on general polygonal or polyhedral meshes, we consider the following space of DOFs, composed of vector cell unknowns and scalar face unknowns associated to the normal component of the discrete vector field:

$$\mathbb{V}_h := \left\{ \mathrm{v}_h = \left((\boldsymbol{v}_K \in \mathbb{R}^d)_{K\in\mathcal{K}_h}, (v_F^n \in \mathbb{R})_{F\in\mathcal{F}_{\mathcal{K}_h}} \right) \right\}.$$

As it is the case for the extension of the Crouzeix–Raviart space discussed in Chapter III, cell unknowns are used to define a piecewise constant subgrid correction of the gradient on the (fictitious) pyramidal submesh. The main difference with respect to the construction of Section III.2 is that we now define an isotropic instead of a full gradient operator. More specifically, we introduce the operator $\mathfrak{G}_h : \mathbb{V}_h \to \mathbb{P}_d^0(\mathcal{P}_h)$ which realizes the mapping $\mathrm{v}_h \mapsto \mathfrak{G}_h(\mathrm{v}_h)$ with

$$\mathfrak{G}_h(\mathrm{v}_h)_{|K_F} := G_K(\mathrm{v}_h) + R_{K_F}(\mathrm{v}_h), \qquad \forall K \in \mathcal{K}_h,\, \forall F \in \mathcal{F}_K,$$

where

$$G_K(\mathrm{v}_h) := \frac{1}{d\,|K|}\sum_{F\in\mathcal{F}_K} |F|\, v_F^n \boldsymbol{n}_F\cdot\boldsymbol{n}_{K,F},\ R_{K_F}(\mathrm{v}_h) := \frac{\eta}{d_{K,F}}(v_F^n\boldsymbol{n}_F - \boldsymbol{v}_K - G_K(\mathrm{v}_h)(\overline{\boldsymbol{x}}_F - \boldsymbol{x}_K))\cdot\boldsymbol{n}_{K,F}, \tag{A.1}$$

and $\eta > 0$ is a user-dependent parameter.

We can now introduce the reconstruction operator $\mathfrak{R}_h : \mathbb{V}_h \to \mathbb{RT}_d^0(\mathcal{P}_h)$ which realizes the mapping $\mathrm{v}_h \mapsto \mathfrak{R}_h(\mathrm{v}_h)$ with

$$\mathfrak{R}_h(\mathrm{v}_h)_{|K_F}(\boldsymbol{x}) = \boldsymbol{v}_K + \mathfrak{G}_h(\mathrm{v}_h)_{|K_F}(\boldsymbol{x} - \boldsymbol{x}_K), \qquad \forall K_F \in \mathcal{P}_h,\, \forall \boldsymbol{x} \in K_F. \tag{A.2}$$

Unlike (III.10), there holds for all $K \in \mathcal{K}_h$, $\boldsymbol{v}_K = \mathfrak{R}_h(\mathrm{v}_h)(\boldsymbol{x}_K)$, i.e., the cell unknown can now be interpreted as the value of the reconstruction at the cell center. This is a consequence of selecting the cell center as a starting point in (A.2). Thus, we consider the discrete space

$$\mathfrak{RT}(\mathcal{K}_h) := \mathfrak{R}_h(\mathbb{V}_h).$$

A.2 Conformity and approximation properties

In the spirit of Section III.3, we investigate the properties of the discrete space we have just introduced.

Lemma A.1 ($\boldsymbol{H}(\mathrm{div};\Omega)$-conformity)**.** *Assume $\eta = 1$ in* (A.1)*. Then, for all $\mathrm{v}_h \in \mathfrak{RT}(\mathcal{K}_h)$ and all $F \in \mathcal{F}_{\mathcal{P}_h}^{\mathsf{i}}$, there holds for all $\boldsymbol{x} \in F$,*

$$[\![\boldsymbol{v}_h]\!]_F(\boldsymbol{x})\cdot\boldsymbol{n}_F = 0.$$

Proof. Let $\boldsymbol{v}_h \in \mathfrak{RT}(\mathcal{K}_h)$ with $\boldsymbol{v}_h = \mathfrak{R}_h(\mathbb{v}_h)$, $F \in \mathcal{F}^{\mathrm{i}}_{\mathcal{P}_h}$, and $\boldsymbol{x} \in F$. We distinguish two cases.
(i) $F \in \mathcal{F}^{\mathrm{i}}_{\mathcal{K}_h}$ is an interface of the primal mesh $\mathcal{K}_h$ such that $F \subset \partial K_1 \cap \partial K_2$. For $i \in \{1,2\}$, let for the sake of brevity $G_i := G_{K_i}(\mathbb{v}_h)$, $R_i := R_{K_iF}(\mathbb{v}_h)$, $d_i := d_{K_i,F}(\boldsymbol{n}_{K_i,F}\cdot\boldsymbol{n}_F)$, and

$$\alpha_i := R_i(\boldsymbol{x}-\boldsymbol{x}_{K_i})\cdot\boldsymbol{n}_F = R_i d_i = \eta\left(v_F^n\boldsymbol{n}_F - \boldsymbol{v}_{K_i} - G_i(\overline{\boldsymbol{x}}_F - \boldsymbol{x}_{K_i})\right)\cdot\boldsymbol{n}_F,$$

where we have used the fact that $\boldsymbol{x} \in F$ to infer $(\boldsymbol{x}-\boldsymbol{x}_{K_i})\cdot\boldsymbol{n}_F = d_i$, and the fact that $\boldsymbol{n}_F = \boldsymbol{n}_{K_1,F} = -\boldsymbol{n}_{K_2,F}$ to infer $(\boldsymbol{n}_{K_i,F}\cdot\boldsymbol{n}_F)\,\boldsymbol{n}_{K_i,F} = \boldsymbol{n}_F$. Algebraic manipulations yield

$$\alpha_1 - \alpha_2 = -\eta\left[(\boldsymbol{v}_{K_1}-\boldsymbol{v}_{K_2})\cdot\boldsymbol{n}_F + G_1d_1 - G_2d_2\right].$$

Using the previous relation in the definition of the jump at $\boldsymbol{x} \in F$ it is inferred

$$\begin{aligned}\llbracket\boldsymbol{v}_h\rrbracket_F(\boldsymbol{x})\cdot\boldsymbol{n}_F &= \boldsymbol{v}_{h|K_1F}(\boldsymbol{x})\cdot\boldsymbol{n}_F - \boldsymbol{v}_{h|K_2F}(\boldsymbol{x})\cdot\boldsymbol{n}_F\\ &= (\boldsymbol{v}_{K_1}-\boldsymbol{v}_{K_2})\cdot\boldsymbol{n}_F + G_1d_1 - G_2d_2 + \alpha_1 - \alpha_2\\ &= (1-\eta)\left[(\boldsymbol{v}_{K_1}-\boldsymbol{v}_{K_2})\cdot\boldsymbol{n}_F + G_1d_1 - G_2d_2\right].\end{aligned}$$

As a consequence, the jump vanishes provided $\eta = 1$.
(ii) $F \in \mathcal{F}^{\mathrm{i}}_{\mathcal{P}_h}\backslash\mathcal{F}^{\mathrm{i}}_{\mathcal{K}_h}$ is a lateral pyramidal face such that there exist a unique element $K \in \mathcal{K}_h$ and two faces F_1, $F_2 \in \mathcal{F}_K$ such that $F \subset \partial K_{F_1} \cap \partial K_{F_2}$ (cf. Figure III.2a). There holds, letting for the sake of brevity $R_i := R_{K_{F_i}}(\mathbb{v}_h)$, $i \in \{1,2\}$,

$$\llbracket\boldsymbol{v}_h\rrbracket_F(\boldsymbol{x})\cdot\boldsymbol{n}_F = \boldsymbol{v}_{h|K_{F_1}}(\boldsymbol{x})\cdot\boldsymbol{n}_F - \boldsymbol{v}_{h|K_{F_2}}(\boldsymbol{x})\cdot\boldsymbol{n}_F = (R_1-R_2)(\boldsymbol{x}-\boldsymbol{x}_K)\cdot\boldsymbol{n}_F = 0,$$

since $(\boldsymbol{x}-\boldsymbol{x}_K)$ and $\boldsymbol{n}_F$ are orthogonal by definition. This concludes the proof. □

We remark that choosing $\boldsymbol{v}_K$ as a starting point for the reconstruction enables to prove the continuity of the normal component on lateral pyramidal faces. Besides, unlike Lemma III.4, the parameter η is here used to enforce the continuity of the normal component across the interfaces of the primal mesh rather than across lateral pyramidal faces. For the sake of completeness, we give the expression of the isotropic gradient operator in the case $\eta = 1$: for all $\mathbb{v}_h \in \mathbb{V}_h$,

$$\mathfrak{G}_h(\mathbb{v}_h)_{|K_F} = \frac{1}{d_{K,F}}(v_F^n\boldsymbol{n}_F - \boldsymbol{v}_K)\cdot\boldsymbol{n}_{K,F}, \qquad \forall K \in \mathcal{K}_h,\ \forall F \in \mathcal{F}_K.$$

Let now introduce the interpolator $\mathcal{I}_h^{\mathfrak{RT}} : H^1(\Omega)^d \to \mathfrak{RT}(\mathcal{K}_h)$ such that, for all $\boldsymbol{v} \in H^1(\Omega)^d$, $\mathcal{I}_h^{\mathfrak{RT}}(\boldsymbol{v}) := \mathfrak{R}_h(\mathbb{v}_h)$ with

$$\mathbb{V}_h \ni \mathbb{v}_h = \left((\Pi_h^1(\boldsymbol{v})(\boldsymbol{x}_K))_{K\in\mathcal{K}_h},\ (\langle\boldsymbol{v}\rangle_F\cdot\boldsymbol{n}_F)_{F\in\mathcal{F}_{\mathcal{K}_h}}\right).$$

The following result summarizes the most relevant approximation properties of $\mathcal{I}_h^{\mathfrak{RT}}$. The proof is omitted as it closely resembles that of Lemma III.5 or Corollary III.1.

Lemma A.2 (Approximation in $\mathfrak{RT}(\mathcal{K}_h)$)**.** *For all $\eta > 0$ in* (A.1) *and all $\boldsymbol{v} \in H^1(\Omega)^d$, there holds with $\boldsymbol{v}_h := \mathcal{I}_h^{\mathfrak{RT}}(\boldsymbol{v}) \in \mathfrak{RT}(\mathcal{K}_h)$,*

$$D_h(\boldsymbol{v}_h) = D_h(\boldsymbol{v}),$$

where the operator D_h is defined in (III.18)*. Moreover, there exists a real $C > 0$ independent of the meshsize such that, for all $h \in \mathcal{H}$, all $K \in \mathcal{K}_h$, and all $\boldsymbol{v} \in H^1(\Omega)^d \cap \boldsymbol{H}^1(\mathrm{div};\mathcal{K}_h)$ (see Corollary III.1) with $\boldsymbol{v}_h := \mathcal{I}_h^{\mathfrak{RT}}(\boldsymbol{v})$, there holds*

$$\|\boldsymbol{v}-\boldsymbol{v}_h\|_{0,K} + \|\boldsymbol{\nabla}\cdot\boldsymbol{v} - D_h(\boldsymbol{v}_h)\|_{0,K} \leqslant Ch_K\left(|\boldsymbol{v}|_{1,K} + |\boldsymbol{\nabla}\cdot\boldsymbol{v}|_{1,K}\right).$$

Remark A.1 (The matching simplicial case)**.** *When considering a matching simplicial mesh $\mathcal{T}_h$, in the spirit of Proposition III.2, we can prove that the lowest-order Raviart–Thomas space is a subspace of $\mathfrak{RT}(\mathcal{T}_h)$. This can then be accounted for in the proof of Lemma A.2 as it is detailed in Section III.4. We emphasize that the assumption $\eta = 1$ in Lemma A.1 remains mandatory also in that case for the continuity of normal values at interfaces. This is a consequence of choosing the cell unknown as a starting point in the reconstruction.*

Annexe B

On the steady Stokes problem

Sommaire

This appendix is inspired from the article [28], written with Daniele A. Di Pietro and accepted for publication in Mathematics of Computation. We briefly discuss an inf-sup stable method for the steady Stokes problem on general polygonal or polyhedral meshes, with velocity components in $\mathfrak{CR}(\mathcal{K}_h)$ (see Chapter III) and piecewise constant pressures. Since the proofs are very classical we only sketch them. We also investigate as we did in Section IV.3 the links of the proposed method with classical finite volume or finite element methods. Finally, we tackle the problem of large irrotational forcing terms and pinpoint a general strategy for their discrete treatment that we apply to the proposed method.

We consider an incompressible and viscous Newtonian fluid of constant unit dynamic viscosity, whose motion is governed by the steady Stokes equations. The problem consists in finding a vector-valued velocity field $\boldsymbol{u} : \Omega \to \mathbb{R}^d$, and a scalar-valued pressure $p : \Omega \to \mathbb{R}$, such that

$$\begin{aligned} -\triangle \boldsymbol{u} + \boldsymbol{\nabla} p &= \boldsymbol{f} && \text{in } \Omega, \\ \boldsymbol{\nabla}\cdot\boldsymbol{u} &= 0 && \text{in } \Omega, \\ \boldsymbol{u} &= \boldsymbol{0} && \text{on } \partial\Omega, \\ \frac{1}{|\Omega|}\int_\Omega p(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= 0, \end{aligned} \tag{B.1}$$

where $\boldsymbol{f} : \Omega \to \mathbb{R}^d$ represents the body force per unit volume acting on the fluid. For the sake of simplicity we focus on homogeneous Dirichlet boundary conditions.

B.1 Discretization

Let $\boldsymbol{U} := H^1_0(\Omega)^d$, and $P := L^2_0(\Omega)$, where $L^2_0(\Omega)$ has been introduced in (II.12). For $\boldsymbol{f} \in L^2(\Omega)^d$, the weak formulation of problem (B.1) reads: Find $(\boldsymbol{u}, p) \in \boldsymbol{U} \times P$ such that

$$\begin{aligned} a(\boldsymbol{u}, \boldsymbol{v}) + b(\boldsymbol{v}, p) &= (\boldsymbol{f}, \boldsymbol{v})_{0,\Omega} \qquad && \forall \boldsymbol{v} \in \boldsymbol{U}, \\ b(\boldsymbol{u}, q) &= 0 && \forall q \in P, \end{aligned} \tag{B.2}$$

where $a(\boldsymbol{w}, \boldsymbol{v}) := (\boldsymbol{\nabla} \boldsymbol{w}, \boldsymbol{\nabla} \boldsymbol{v})_{0,\Omega}$, and $b(\boldsymbol{v}, q) := -(\boldsymbol{\nabla} \cdot \boldsymbol{v}, q)_{0,\Omega}$. To approximate (B.2), let $\mathcal{K}_h$ be a general polygonal or polyhedral mesh, belonging to an admissible mesh sequence in the sense of Definition III.3, and define the following discrete spaces:

$$\boldsymbol{U}_h := \mathfrak{CR}_0(\mathcal{K}_h)^d, \qquad P_h := \mathbb{P}^0_d(\mathcal{K}_h) \cap L^2_0(\Omega),$$

where $\mathfrak{CR}_0(\mathcal{K}_h)$ is defined in (III.20). We equip $\boldsymbol{U}_h$ with the norm $\|\boldsymbol{\nabla}_h \boldsymbol{v}\|_{0,\Omega}$, see Proposition III.3, and P_h with the norm $\|q\|_{0,\Omega}$. We assume in the following $\eta = d$ in (III.9), so that the continuity of mean (or, equivalently, barycentric) values stated in Lemma III.4 holds, and consider the following discrete problem: Find $(\boldsymbol{u}_h, p_h) \in \boldsymbol{U}_h \times P_h$ such that

$$\begin{aligned} a_h(\boldsymbol{u}_h, \boldsymbol{v}_h) + b_h(\boldsymbol{v}_h, p_h) &= (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega} \qquad && \forall \boldsymbol{v}_h \in \boldsymbol{U}_h, \\ b_h(\boldsymbol{u}_h, q_h) &= 0 && \forall q_h \in P_h, \end{aligned} \tag{B.3}$$

where $a_h(\boldsymbol{w}, \boldsymbol{v}) := (\boldsymbol{\nabla}_h \boldsymbol{w}, \boldsymbol{\nabla}_h \boldsymbol{v})_{0,\Omega}$, and $b_h(\boldsymbol{v}, q) := -(D_h(\boldsymbol{v}), q)_{0,\Omega}$, where D_h is the operator defined in (III.18). For all $(\boldsymbol{v}_h, q_h) \in \boldsymbol{U}_h \times P_h$, there holds $b_h(\boldsymbol{v}_h, q_h) = -(\boldsymbol{\nabla}_h \cdot \boldsymbol{v}_h, q_h)_{0,\Omega}$.

The link between locking-free approximations of quasi-incompressible linear elasticity and inf-sup stable approximations of the Stokes problem is well-known; cf., e.g., the discussion in [17, Section IV.3]. With regards to what we proved in Chapter IV on the locking-free aspect of a $\mathfrak{CR}(\mathcal{K}_h)$-based discretization of linear elasticity equations, it is not a surprise to have the following property.

Lemma B.1 (inf-sup stability for b_h)**.** *There exists $\beta > 0$, independent of the meshsize, such that, for all $q_h \in P_h$,*

$$\beta \|q_h\|_{0,\Omega} \leqslant \sup_{\boldsymbol{v}_h \in \boldsymbol{U}_h \setminus \{\boldsymbol{0}\}} \frac{b_h(\boldsymbol{v}_h, q_h)}{\|\boldsymbol{\nabla}_h \boldsymbol{v}_h\|_{0,\Omega}}.$$

Proof. This result is a consequence of (i) the fact that the interpolator $\mathcal{I}_h^{\mathfrak{CR}}$ (cf. Section III.3.2) can play the role of a Fortin operator when coupled with piecewise constant pressures (see Corollary III.1, and Lemma III.5 for the H^1-stability property), (ii) and Remark II.2. □

From a more general point of view, for inf-sup stable approximations of the Stokes problem with discontinuous pressures, one can obtain a locking-free primal method for elasticity by performing static condensation of pressures, that is equivalent to introducing a projection on the divergence operator (this strategy can e.g. be pursued for the method of [9, 10]).

The well-posedness of the discrete problem (B.3) follows from Lemma B.1 together with the coercivity of a_h (an immediate consequence of Proposition III.3). Using classical arguments, one can prove the convergence of the method (B.3), as well as an optimal error estimate. This estimate will be invoked in the discussion of Section B.3. As in Theorem IV.1, the continuity of mean values at interfaces (as a consequence of $\eta = d$) is used to bound the conformity error. Note that the use of the under-integrated divergence operator D_h in b_h also introduces a consistency error. Optimal error estimates for the L^2-error on the velocity can also be derived using the Aubin–Nitsche trick.

Theorem B.1 (Error estimate for (B.3))**.** *Assume that* $\boldsymbol{u} \in \boldsymbol{U} \cap H^2(\Omega)^d$ *and* $p \in P \cap H^1(\Omega)$*, where* $(\boldsymbol{u}, p)$ *denotes the unique solution to the weak formulation* (B.2)*. Then, there holds with* $C > 0$ *independent of the meshsize, of* $\boldsymbol{u}$*, and of* p*,*

$$\|\boldsymbol{\nabla} \boldsymbol{u} - \boldsymbol{\nabla}_h \boldsymbol{u}_h\|_{0,\Omega} + \|p - p_h\|_{0,\Omega} \leqslant Ch\, \mathcal{N}_{\mathrm{sto}}(\boldsymbol{u}, p),$$

where $(\boldsymbol{u}_h, p_h) \in \boldsymbol{U}_h \times P_h$ *denotes the unique solution to* (B.3)*, and* $\mathcal{N}_{\mathrm{sto}}(\boldsymbol{u}, p) := \|\boldsymbol{u}\|_{2,\Omega} + \|p\|_{1,\Omega}$*.*

B.2 Links with finite volume and finite element methods

As we did in Section IV.3, we investigate the links of our method with other related frameworks, depending on the treatment of the right-hand side.

B.2.1 Flux formulation and local conservation

Let consider the approximation (B.3) of the Stokes problem. We do not make any assumption on the value of $\eta > 1$. Let $(\boldsymbol{w}_h, r_h), (\boldsymbol{v}_h, q_h) \in \boldsymbol{U}_h \times P_h$ be two discrete functions, and denote by $(\mathbb{w}_h, \mathbb{r}_h), (\mathbb{v}_h, \mathbb{q}_h) \in \mathbb{U}_h \times \mathbb{P}_h$ the corresponding vectors of DOFs, where we have set $\mathbb{U}_h := \mathbb{V}_{h,0}^d$ (cf. (III.20)) and $\mathbb{P}_h := \{\mathbb{q}_h \in \mathbb{R}^{\mathcal{K}_h} \mid \sum_{K \in \mathcal{K}_h} |K|\, q_K = 0\}$. Then, proceeding as in Section IV.3.1, one can show that for the two families of fluxes $(\boldsymbol{\Phi}_{K,F}(\mathbb{w}_h, \mathbb{r}_h))_{K \in \mathcal{K}_h,\, F \in \mathcal{F}_K}$ with $\boldsymbol{\Phi}_{K,F}(\mathbb{w}_h, \mathbb{r}_h) = (\Phi_{K,F,i}(\mathbb{w}_h, \mathbb{r}_h))_{1 \leqslant i \leqslant d}$, and $(\phi_F(\mathbb{w}_h))_{F \in \mathcal{F}_{\mathcal{K}_h}}$ such that (the expression for the vectors $\boldsymbol{y}_{F',F}^K$ is provided in (IV.17), cf. Proposition IV.1)

$$\Phi_{K,F,i}(\mathbb{w}_h, \mathbb{r}_h) := \sum_{F' \in \mathcal{F}_K} |K_{F'}| \left(\boldsymbol{G}_{K_{F'}}(\mathbb{w}_{h,i}) - r_K \boldsymbol{e}_i\right) \cdot \boldsymbol{y}_{F',F}^K, \qquad \phi_F(\mathbb{w}_h) := |F|\, \boldsymbol{w}_F \cdot \boldsymbol{n}_F,$$

there holds,

$$\begin{aligned} a_h(\boldsymbol{w}_h, \boldsymbol{v}_h) + b_h(\boldsymbol{v}_h, r_h) &= \sum_{K \in \mathcal{K}_h} \sum_{F \in \mathcal{F}_K} \boldsymbol{\Phi}_{K,F}(\mathbb{w}_h, \mathbb{r}_h) \cdot (\boldsymbol{v}_F - \boldsymbol{v}_K), \\ -b_h(\boldsymbol{w}_h, q_h) &= \sum_{F \in \mathcal{F}_{\mathcal{K}_h}} \phi_F(\mathbb{w}_h) [\![\mathbb{q}_h]\!]_F, \end{aligned} \tag{B.4}$$

where a_h and b_h are defined as in Section B.1 and, with a slight abuse in notation, we have set for all $F \in \mathcal{F}_{\mathcal{K}_h}$, $[\![\mathbb{q}_h]\!]_F := [\![q_h]\!]_F$.

Here again, the main interest of this formulation is that it allows to prove a local conservation property similar to those encountered in standard finite volume methods. Proceeding as in Section IV.3.1, and approximating the right-hand side in (B.3) as $\sum_{K \in \mathcal{K}_h} |K|\, \boldsymbol{f}_K \cdot \boldsymbol{v}_K$ where we define $\boldsymbol{f}_K := \frac{1}{|K|} \int_K \boldsymbol{f} \mathrm{d}\boldsymbol{x}$, one can prove thanks to (B.4) that for every interface $F \in \mathcal{F}_{\mathcal{K}_h}^{\mathrm{i}}$ such that $F \subset \partial K_1 \cap \partial K_2$, there holds

$$\boldsymbol{\Phi}_{K_1,F}(\mathbb{u}_h, \mathbb{p}_h) = -\boldsymbol{\Phi}_{K_2,F}(\mathbb{u}_h, \mathbb{p}_h),$$

where $(\mathbb{u}_h, \mathbb{p}_h) \in \mathbb{U}_h \times \mathbb{P}_h$ are such that $\boldsymbol{u}_h := \mathfrak{R}_h(\mathbb{u}_h)$ and $p_{h|K} := p_K$ for all $K \in \mathcal{K}_h$, with $(\boldsymbol{u}_h, p_h) \in \boldsymbol{U}_h \times P_h$ unique solution to (B.3). Moreover, the mass flux $\phi_F(\mathbb{u}_h)$ is single-valued, and therefore conservative.

B.2.2 Link with the Crouzeix–Raviart solution

Let again consider the discretization (B.3) of the Stokes problem, let $\eta > 1$, and let consider a matching simplicial mesh that we denote $\mathcal{T}_h$. The classical Crouzeix–Raviart/$\mathbb{P}_d^0(\mathcal{T}_h)$ method

consists in finding $(\hat{\boldsymbol{u}}_h, \hat{p}_h) \in \hat{\boldsymbol{U}}_h \times P_h$ with $\hat{\boldsymbol{U}}_h := \mathbb{CR}_0(\mathcal{T}_h)^d$ (cf. (IV.21)) such that

$$\begin{aligned} a_h(\hat{\boldsymbol{u}}_h, \boldsymbol{v}_h) + b_h(\boldsymbol{v}_h, \hat{p}_h) &= (\boldsymbol{f}, \boldsymbol{v}_h)_{0,\Omega} \qquad && \forall \boldsymbol{v}_h \in \hat{\boldsymbol{U}}_h, \\ b_h(\hat{\boldsymbol{u}}_h, q_h) &= 0 && \forall q_h \in P_h. \end{aligned} \tag{B.5}$$

Proceeding as in Section IV.3.2, one can easily show that the solution to (B.5) can be recovered replacing the right-hand side of (B.3) by $(\boldsymbol{f}, \mathcal{I}_h^{\mathbb{CR}}(\boldsymbol{v}_h))_{0,\Omega}$. In other words, the system forces the subgrid corrections to vanish as soon as the right-hand side does not *see* the cell unknowns.

B.3 Large irrotational forcing terms

B.3.1 Position of the problem

We discuss here a general modification, applicable to any suitable discretization of the Stokes equations, that allows a proper discrete treatment of large irrotational forcing terms, and we apply it to the method (B.3). This modification necessitates the knowledge of a Helmholtz decomposition of the volumetric body force. It has to be noted that such a decomposition is not always easy to obtain, and that often in applications it is merely unknown.

We assume that the following Helmholtz decomposition of the volumetric body force in (B.1) is available:

$$\boldsymbol{f} = \boldsymbol{\Psi} - \boldsymbol{\nabla}\varphi, \tag{B.6}$$

where $\boldsymbol{\Psi} \in \boldsymbol{H}_0(\mathrm{div};\Omega) := \{\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div};\Omega) \mid \gamma_{\boldsymbol{n}}(\boldsymbol{v}) = 0\}$ ($\gamma_{\boldsymbol{n}}(\boldsymbol{v}) \in H^{-\frac{1}{2}}(\Gamma)$ is the normal trace $\boldsymbol{v}\cdot\boldsymbol{n}_{|\Gamma}$) is a solenoidal vector field such that $\boldsymbol{\nabla}\cdot\boldsymbol{\Psi} = 0$, and $\varphi \in H^1(\Omega) \cap L_0^2(\Omega)$ is a scalar potential ($\boldsymbol{\nabla}\varphi$ is called the irrotational part of the force). The weak formulation of problem (B.1) with right-hand side given by (B.6) reads: Find $(\boldsymbol{u}, p) \in \boldsymbol{U} \times P$ such that

$$\begin{aligned} a(\boldsymbol{u}, \boldsymbol{v}) + b(\boldsymbol{v}, p) &= l(\boldsymbol{v}) \qquad && \forall \boldsymbol{v} \in \boldsymbol{U}, \\ b(\boldsymbol{u}, q) &= 0 && \forall q \in P, \end{aligned} \tag{B.7}$$

with bilinear forms a and b defined as in Section B.1 and $l(\boldsymbol{v}) := (\boldsymbol{\Psi}, \boldsymbol{v})_{0,\Omega} - b(\boldsymbol{v}, \varphi)$. Denoting by $(\boldsymbol{u}_{\boldsymbol{\Psi}}, p_{\boldsymbol{\Psi}})$ the solution to (B.7) with $\varphi \equiv 0$ (no irrotational part), there holds

$$\boldsymbol{u} = \boldsymbol{u}_{\boldsymbol{\Psi}}, \qquad p = p_{\boldsymbol{\Psi}} - \varphi. \tag{B.8}$$

As pointed out in [47], mimicking or approaching property (B.8) at the discrete level is a key ingredient to obtain an accurate approximation of the velocity field for large values of $|\varphi|_{1,\Omega}$. As a matter of fact, yet stable mixed finite element methods usually do not give satisfactory results since spurious oscillations appear on the velocity field as $|\varphi|_{1,\Omega}$ grows. This phenomenon can be partially handled by considering either pointwise divergence-free (by opposition to discretely divergence-free) approximations, either by using grad-div stabilizations. In the case when the Helmholtz decomposition is available, we propose to handle that problem by an appropriate treatment of the right-hand side. We hence consider the following approximation to (B.7): Find $(\boldsymbol{u}_h, p_h) \in \boldsymbol{U}_h \times P_h$ such that

$$\begin{aligned} a_h(\boldsymbol{u}_h, \boldsymbol{v}_h) + b_h(\boldsymbol{v}_h, p_h) &= l_h(\boldsymbol{v}_h) \qquad && \forall \boldsymbol{v}_h \in \boldsymbol{U}_h, \\ b_h(\boldsymbol{u}_h, q_h) &= 0 && \forall q_h \in P_h, \end{aligned} \tag{B.9}$$

with bilinear forms a_h and b_h defined as in Section B.1 and $l_h(\boldsymbol{v}_h) := (\boldsymbol{\Psi}, \boldsymbol{v}_h)_{0,\Omega} - b_h(\boldsymbol{v}_h, \Pi_h^0(\varphi))$, where Π_h^0 classically denotes the L^2-orthogonal projector onto $\mathbb{P}_d^0(\mathcal{K}_h)$. Note that for $\varphi \in L_0^2(\Omega)$, $\Pi_h^0(\varphi) \in L_0^2(\Omega)$. The sole difference with respect to (B.3) lies in the treatment of the source term, which is designed so that the following property holds true.

Proposition B.1 (Discrete counterpart of property (B.8))**.** *Denote by* $(\boldsymbol{u}_{\boldsymbol{\Psi},h}, p_{\boldsymbol{\Psi},h})$ *the solution to problem* (B.9) *with* $\varphi \equiv 0$*. There holds*

$$\boldsymbol{u}_h = \boldsymbol{u}_{\boldsymbol{\Psi},h}, \qquad p_h = p_{\boldsymbol{\Psi},h} - \Pi_h^0(\varphi).$$

The following result now shows that the velocity approximation is unaffected by the irrotational part of the source term.

Theorem B.2 (Error estimate for (B.9))**.** *Assume* $\boldsymbol{u} \in \boldsymbol{U} \cap H^2(\Omega)^d$ *and* $p \in P \cap H^1(\Omega)$*. Then, there holds with real numbers* $C_1 > 0$ *and* $C_2 > 0$ *independent of the meshsize, of* $\boldsymbol{u}$*, and of* p*, but depending on the mesh regularity parameters and on* Ω*,*

$$\|\boldsymbol{\nabla}\boldsymbol{u} - \boldsymbol{\nabla}_h\boldsymbol{u}_h\|_{0,\Omega} \leqslant C_1 h \mathcal{N}_{\mathrm{sto}}(\boldsymbol{u}_{\boldsymbol{\Psi}}, p_{\boldsymbol{\Psi}}), \qquad \|p - p_h\|_{0,\Omega} \leqslant C_2 h \left(\mathcal{N}_{\mathrm{sto}}(\boldsymbol{u}_{\boldsymbol{\Psi}}, p_{\boldsymbol{\Psi}}) + |\varphi|_{1,\Omega}\right),$$

where $\mathcal{N}_{\mathrm{sto}}(\cdot,\cdot)$ *is defined in Theorem B.1.*

Proof. Using Theorem B.1 for the solution to problem (B.9) with $\varphi \equiv 0$, we infer

$$\|\boldsymbol{\nabla}\boldsymbol{u}_{\boldsymbol{\Psi}} - \boldsymbol{\nabla}_h\boldsymbol{u}_{\boldsymbol{\Psi},h}\|_{0,\Omega} + \|p_{\boldsymbol{\Psi}} - p_{\boldsymbol{\Psi},h}\|_{0,\Omega} \leqslant Ch\mathcal{N}_{\mathrm{sto}}(\boldsymbol{u}_{\boldsymbol{\Psi}}, p_{\boldsymbol{\Psi}}),$$

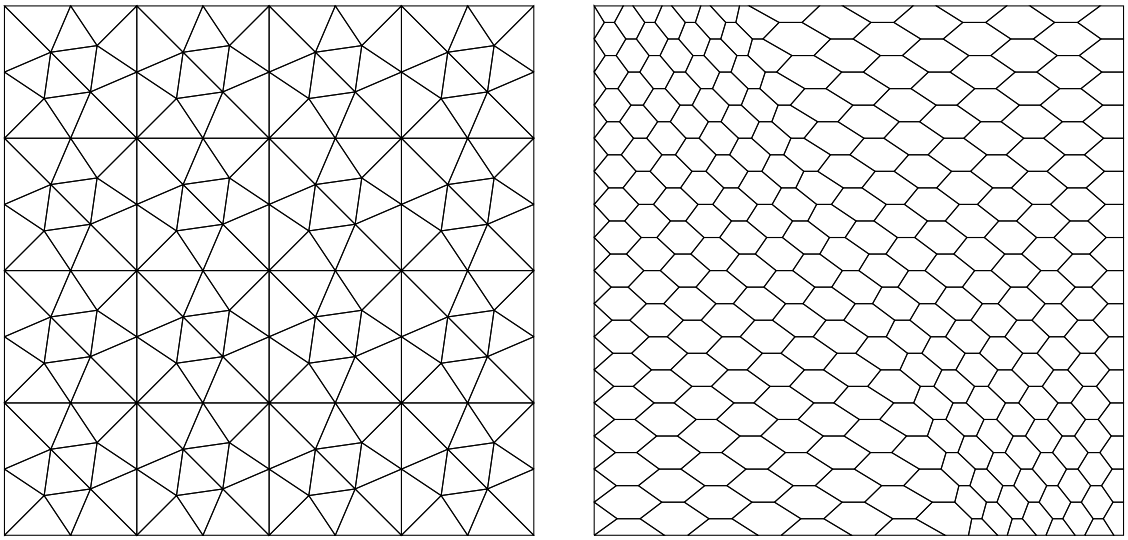
where $C > 0$ has the same dependencies as C_1 and C_2. The estimate for $\|\boldsymbol{\nabla}\boldsymbol{u} - \boldsymbol{\nabla}_h\boldsymbol{u}_h\|_{0,\Omega}$ is an immediate consequence of (B.8) and Proposition B.1. To estimate $\|p - p_h\|_{0,\Omega}$, we invoke again (B.8) and Proposition B.1 to infer $\|p - p_h\|_{0,\Omega} \leqslant \|p_{\boldsymbol{\Psi}} - p_{\boldsymbol{\Psi},h}\|_{0,\Omega} + \|\varphi - \Pi_h^0(\varphi)\|_{0,\Omega}$, and conclude using the above estimate for $\|p_{\boldsymbol{\Psi}} - p_{\boldsymbol{\Psi},h}\|_{0,\Omega}$ and the approximation properties of the L^2-orthogonal projector. □

B.3.2 Application

To check the theoretical results, we consider a 2D numerical example based on the following manufactured solution on the unit square domain $\Omega := (0,1)^2$:

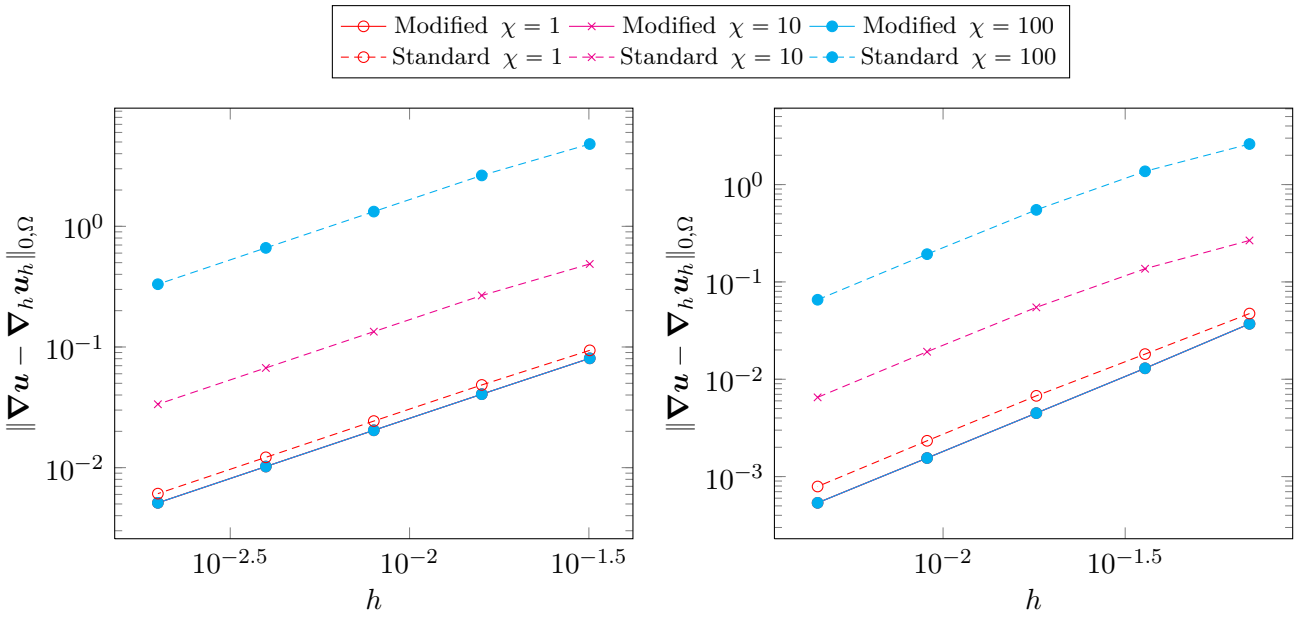
$$u_x = -e^x(y\cos(y) + \sin(y)), \qquad u_y = e^x y \sin(y), \qquad p_{\boldsymbol{\Psi}} = 2\exp(x)\sin(y) - C_{p_{\boldsymbol{\Psi}}},$$

with $C_{p_{\boldsymbol{\Psi}}}$ such that $p_{\boldsymbol{\Psi}}$ has zero-mean on Ω. The right-hand side is such that $\boldsymbol{\Psi} \equiv \boldsymbol{0}$, and the potential is chosen such that $\varphi = -\chi\sin(2\pi x)\sin(2\pi y)$, where χ is a positive parameter that allows to adjust its intensity. Note that this solution satisfies the regularity assumptions of Theorem B.2. In Figure B.2, we compare the numerical results obtained with the modified right-hand side (B.9) to those obtained with a standard treatment (B.3) for the matching triangular and hexagonal-dominant mesh families depicted in Figure B.1. The results confirm that a standard treatment of the right-hand side does not yield satisfactory results (the error on the velocity increases with χ), whereas the treatment proposed in this section under the assumption that a Helmholtz decomposition of the source term is available yields the robustness of the velocity approximation with respect to the potential intensity. Note that in practical implementations, one can solve the problem with $\varphi \equiv 0$ and then post-process the pressure approximation according to Proposition B.1.



(a) Matching triangular (b) Hexagonal-dominant

Figure B.1: Members of the 2D mesh families for the numerical test of Section B.3.2.



(a) Matching triangular mesh family (b) Hexagonal-dominant mesh family

Figure B.2: Effect of the treatment of the right-hand side (B.9) (Modified, solid lines) vs. (B.3) (Standard, dashed lines) when large irrotational volumetric forces are present.

Annexe C

A coercive finite volume discretization of linear elasticity equations

Sommaire

This appendix takes inspiration from the work [25], realized in collaboration with Daniele A. Di Pietro, Robert Eymard and Roland Masson, and presented to the Sixth International Symposium on Finite Volumes for Complex Applications (FVCA6) held in Prague in June 2011. We introduce a Hybrid Finite Volume discretization (cf. [40]) of the linear elasticity model with nonhomogeneous (possibly mixed-type) boundary conditions (C.1). The coercivity issue is treated by adding rigidity to the system. More precisely, we interpolate the tangential component(s) of the displacement on mesh faces by using normal unknowns belonging to a stencil of neighboring faces. The method is proved to converge in two and three space dimensions on a benchmark test-case. However, up to now, no theoretical result confirming these practical observations is available.

Under the very same assumptions as in Section II.1.1, we consider the following linear elasticity problem in a homogeneous medium $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, with boundary Γ such that $\Gamma = \Gamma_{\mathrm{D}} \cup \Gamma_{\mathrm{N}}$ (Γ_{D} has nonzero measure and $\Gamma_{\mathrm{D}} \cap \Gamma_{\mathrm{N}} = \varnothing$) and unit outward normal $\boldsymbol{n}$: Find a vector-valued displacement field $\boldsymbol{u} : \Omega \to \mathbb{R}^d$ such that

$$\begin{aligned} -\boldsymbol{\nabla}\cdot\underline{\underline{\sigma}}(\boldsymbol{u}) &= \boldsymbol{f} && \text{in } \Omega, \\ \boldsymbol{u} &= \boldsymbol{u}_{\mathrm{D}} && \text{on } \Gamma_{\mathrm{D}}, \\ \underline{\underline{\sigma}}(\boldsymbol{u})\boldsymbol{n} &= \boldsymbol{g} && \text{on } \Gamma_{\mathrm{N}}, \end{aligned} \tag{C.1}$$

where $\boldsymbol{f} : \Omega \to \mathbb{R}^d$ denotes the vector-valued body force per unit volume, and where the only difference with respect to system (II.1) lies in the introduction of the nonhomogeneous boundary terms $\boldsymbol{u}_{\mathrm{D}} : \Gamma_{\mathrm{D}} \to \mathbb{R}^d$ and $\boldsymbol{g} : \Gamma_{\mathrm{N}} \to \mathbb{R}^d$.

C.1 The Hybrid Finite Volume setting

We here adopt the notation introduced in Chapter V since the Hybrid Finite Volume (HFV) discretization enters the framework of Gradient discretizations, cf. [33, Section 5.3] for the proofs. We briefly recall in that section the main features of the HFV setting.

Let $\mathcal{D}$ be a (vector-valued) Hybrid Finite Volume discretization of the displacement field for problem (C.1). Following [33], let denote $\mathcal{M}$ the associated mesh, the subscript $\mathcal{D}$ being ignored for the sake of simplicity. Thus, $\mathcal{M}$ is a finite family of nonempty open (disjoint) polygonal or polyhedral control volumes K, such that $\overline{\Omega} = \bigcup_{K\in\mathcal{M}} \overline{K}$. The meshsize $h_\mathcal{D}$ is defined as $h_\mathcal{D} = \max_{K\in\mathcal{M}} h_K$ where h_K denotes the diameter of K, and is such that $h_{\mathcal{D}_m} \to 0$ as $m \to +\infty$ for any sequence $(\mathcal{D}_m)_{m\in\mathbb{N}}$ of displacement Hybrid Finite Volume discretizations. The set of faces of the mesh is denoted by $\mathcal{E}$ (the subscript $\mathcal{D}$ is here again ignored) and splits into boundary faces $\mathcal{E}^{\rm ext}$ and inner interfaces $\mathcal{E}^{\rm int}$. Among boundary faces, we denote by $\mathcal{E}_{\rm D}^{\rm ext} \neq \varnothing$ and $\mathcal{E}_{\rm N}^{\rm ext}$ the subsets of boundary faces respectively satisfying Dirichlet and Neumann conditions, which are such that $\mathcal{E}_{\rm D}^{\rm ext} \cap \mathcal{E}_{\rm N}^{\rm ext} = \varnothing$ and $\mathcal{E}^{\rm ext} = \mathcal{E}_{\rm D}^{\rm ext} \cup \mathcal{E}_{\rm N}^{\rm ext}$. The generic element of $\mathcal{E}$ is denoted σ and its barycenter $\overline{\boldsymbol{x}}_\sigma$. The set of faces of each cell $K \in \mathcal{M}$ is denoted $\mathcal{E}_K$, and we assume that every $K \in \mathcal{M}$ admits a cell center denoted $\boldsymbol{x}_K$, cf. Definition III.2. Finally, $d_{K,\sigma}$ stands for the orthogonal distance between $\boldsymbol{x}_K$ and the face σ, and the open pyramid of apex $\boldsymbol{x}_K$ and base $\sigma \in \mathcal{E}_K$ is denoted K_σ, in such a way that $\overline{K} = \bigcup_{\sigma\in\mathcal{E}_K} \overline{K_\sigma}$. The pyramidal submesh of $\mathcal{M}$ thus engendered is denoted $\mathcal{P}$, which is not a standard notation.

Taking into account these notations, the discrete setting is the one introduced in Sections III.1.1 and III.1.2. The notion of admissible mesh sequence is in particular given by Definition III.3. Before going further, we introduce the following set of discrete unknowns

$$\boldsymbol{X}_\mathcal{D} := \left\{ \boldsymbol{v}_\mathcal{D} = \left((\boldsymbol{v}_K \in \mathbb{R}^d)_{K\in\mathcal{M}}, (\boldsymbol{v}_\sigma \in \mathbb{R}^d)_{\sigma\in\mathcal{E}} \right) \right\}, \tag{C.2}$$

which is an equivalent of (III.7) for vector-valued elements.

Definition C.1 (Displacement Hybrid Finite Volume discretization)**.** The Hybrid Finite Volume discretization of the displacement is defined by $\mathcal{D} := (\boldsymbol{X}_{\mathcal{D},\rm D}, \Pi_\mathcal{D}, \boldsymbol{\nabla}_\mathcal{D}, T_\mathcal{D})$, where

(i) the set of discrete unknowns $\boldsymbol{X}_{\mathcal{D},\rm D}$ is defined as

$$\boldsymbol{X}_{\mathcal{D},\rm D} := \{\boldsymbol{v}_\mathcal{D} \in \boldsymbol{X}_\mathcal{D} \mid \boldsymbol{v}_\sigma = \boldsymbol{0},\ \forall \sigma \in \mathcal{E}_{\rm D}^{\rm ext}\}, \tag{C.3}$$

where $\boldsymbol{X}_\mathcal{D}$ is given by (C.2), and (C.3) is an equivalent of (III.19) for vector-valued elements;

(ii) the reconstruction of the approximate function $\Pi_\mathcal{D} : \boldsymbol{X}_{\mathcal{D},\rm D} \to \mathbb{P}_d^0(\mathcal{M})^d$ is given by

$$\forall \boldsymbol{v}_\mathcal{D} \in \boldsymbol{X}_{\mathcal{D},\rm D}, \quad \forall K \in \mathcal{M}, \quad \Pi_\mathcal{D} \boldsymbol{v}_{\mathcal{D}|K} = \boldsymbol{v}_K;$$

(iii) the discrete gradient operator $\boldsymbol{\nabla}_\mathcal{D} : \boldsymbol{X}_{\mathcal{D},\rm D} \to \mathbb{P}_d^0(\mathcal{P})^{d,d}$ is defined as

$$\forall \boldsymbol{v}_\mathcal{D} \in \boldsymbol{X}_{\mathcal{D},\rm D}, \quad \forall K \in \mathcal{M}, \quad \forall \sigma \in \mathcal{E}_K, \quad \boldsymbol{\nabla}_\mathcal{D} \boldsymbol{v}_{\mathcal{D}|K_\sigma} = \boldsymbol{\nabla}_{K_\sigma} \boldsymbol{v}_\mathcal{D} := \boldsymbol{\nabla}_K \boldsymbol{v}_\mathcal{D} + \boldsymbol{R}_{K_\sigma} \boldsymbol{v}_\mathcal{D},$$

where, for $\eta > 1$ user-dependent parameter,

$$\boldsymbol{\nabla}_K \boldsymbol{v}_\mathcal{D} = \frac{1}{|K|} \sum_{\sigma\in\mathcal{E}_K} |\sigma|\, \boldsymbol{v}_\sigma \otimes \boldsymbol{n}_{K,\sigma}, \quad \boldsymbol{R}_{K_\sigma} \boldsymbol{v}_\mathcal{D} = \frac{\eta}{d_{K,\sigma}} \left(\boldsymbol{v}_\sigma - \boldsymbol{v}_K - \boldsymbol{\nabla}_K \boldsymbol{v}_\mathcal{D} \left(\overline{\boldsymbol{x}}_\sigma - \boldsymbol{x}_K \right) \right) \otimes \boldsymbol{n}_{K,\sigma}, \tag{C.4}$$

and furthermore $\|\boldsymbol{\nabla}_\mathcal{D} \boldsymbol{v}_\mathcal{D}\|_{0,\Omega}$ is a norm on $\boldsymbol{X}_{\mathcal{D},\rm D}$;

(iv) the reconstruction of the approximate trace $T_{\mathcal{D}} : \boldsymbol{X}_{\mathcal{D},\mathrm{D}} \to \mathbb{P}_d^0\left(\mathcal{E}_{\mathrm{N}}^{\mathrm{ext}}\right)^d$ is given by

$$\forall \boldsymbol{v}_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},\mathrm{D}}, \quad \forall \sigma \in \mathcal{E}_{\mathrm{N}}^{\mathrm{ext}}, \quad T_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}|\sigma} = \boldsymbol{v}_\sigma.$$

The choice of $\eta > 1$ has already been discussed in Remark III.1. In the numerical experiments of Section C.4, we will consider $\eta = d^{1/2}$, which is more in the finite volume spirit as it allows to recover the two-point finite volume scheme on superadmissible meshes, cf. [40]. It also differs from the numerical experiments of Chapters IV and V, where the value $\eta = d$ (which is optimal in the finite element sense) was adopted. Note that in the classical definition of the HFV method, meaning the one of [40], the parameter η has not been thought to be tuned and is merely taken equal to $d^{1/2}$.

We see from Definition C.1 that the only difference between the Hybrid Finite Volume discretization and the generalized Crouzeix–Raviart space introduced in Chapter III is the definition of the reconstruction mappings (for the approximate function and trace). Here, the reconstructed approximate function is not gradient-based, but only linked to the gradient operator through a discrete Friedrichs' inequality and a limit-conformity property, see Chapter V.

As it enters the framework of Gradient discretizations, note that any sequence of displacement Hybrid Finite Volume discretizations is coercive, enjoys optimal approximation properties (termed as *consistency* in [33]), is limit-conforming, and in addition admits a sequence of Fortin operators (this is a consequence of Lemma III.5, Corollary III.1, and of the fact that a HFV discretization and its related generalized Crouzeix–Raviart space share the same gradient operator).

Remark C.1 (Norm on $\boldsymbol{X}_{\mathcal{D},\mathrm{D}}$)**.** *In Section III.5, we prove that the L^2-norm of the gradient operator defines a norm on $\mathfrak{CR}_{\mathrm{D}}(\mathcal{K}_h)$ for all $\eta > 1$ by showing its equivalence with the usual dG norm on piecewise polynomial spaces. Here, to establish the same result with a piecewise constant reconstruction mapping, we use the equivalence of the L^2-norm of the gradient operator stated in [40, Lemma 4.1] (in the scalar-valued case) with the following discrete H^1_{D}-norm on $\boldsymbol{X}_{\mathcal{D},\mathrm{D}}$ (which is a seminorm on $\boldsymbol{X}_{\mathcal{D}}$)*

$$|\boldsymbol{v}_{\mathcal{D}}|^2_{\boldsymbol{X}_{\mathcal{D}}} := \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \frac{|\sigma|}{d_{K,\sigma}} |\boldsymbol{v}_\sigma - \boldsymbol{v}_K|^2.$$

For further use, we introduce the following reduced set of discrete unknowns

$$\mathfrak{X}_{\mathcal{D}} := \left\{ \mathfrak{v}_{\mathcal{D}} = \left((\mathfrak{v}_K \in \mathbb{R}^d)_{K\in\mathcal{M}}, (\mathfrak{v}_\sigma \in \mathbb{R}^d)_{\sigma\in\mathcal{E}_{\mathrm{D}}^{\mathrm{ext}}}, (\mathfrak{v}_\sigma^n \in \mathbb{R})_{\sigma\in\mathcal{E}^{\mathrm{int}}\cup\mathcal{E}_{\mathrm{N}}^{\mathrm{ext}}} \right) \right\},$$

and the projection operator $\mathfrak{P}_{\mathcal{D}} : \boldsymbol{X}_{\mathcal{D}} \to \mathfrak{X}_{\mathcal{D}}$ which maps any $\boldsymbol{v}_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D}}$ onto

$$\mathfrak{P}_{\mathcal{D}}(\boldsymbol{v}_{\mathcal{D}}) = \left((\boldsymbol{v}_K)_{K\in\mathcal{M}}, (\boldsymbol{v}_\sigma)_{\sigma\in\mathcal{E}_{\mathrm{D}}^{\mathrm{ext}}}, (\boldsymbol{v}_\sigma\cdot\boldsymbol{n}_\sigma)_{\sigma\in\mathcal{E}^{\mathrm{int}}\cup\mathcal{E}_{\mathrm{N}}^{\mathrm{ext}}} \right),$$

where $\boldsymbol{n}_\sigma$ is a unit vector normal to σ which orientation is fixed, cf. Section III.1.3. We also define $\mathfrak{X}_{\mathcal{D},\mathrm{D}}$ as $\mathfrak{X}_{\mathcal{D},\mathrm{D}} := \mathfrak{P}_{\mathcal{D}}(\boldsymbol{X}_{\mathcal{D},\mathrm{D}})$. We endow $\mathfrak{X}_{\mathcal{D},\mathrm{D}}$ with the following norm (which is a seminorm on $\mathfrak{X}_{\mathcal{D}}$):

$$|\mathfrak{v}_{\mathcal{D}}|_{\mathfrak{X}_{\mathcal{D}}} := \inf_{\boldsymbol{v}_{\mathcal{D}}\in\boldsymbol{X}_{\mathcal{D}} | \mathfrak{P}_{\mathcal{D}}(\boldsymbol{v}_{\mathcal{D}})=\mathfrak{v}_{\mathcal{D}}} |\boldsymbol{v}_{\mathcal{D}}|_{\boldsymbol{X}_{\mathcal{D}}}. \tag{C.5}$$

C.2 Interpolation of the displacement tangential component(s) on faces

The main novelty of the discretization proposed in the next section lies in the definition of a linear interpolation operator $\mathfrak{I}_{\mathcal{D}} : \mathfrak{X}_{\mathcal{D}} \to \boldsymbol{X}_{\mathcal{D}}$. This linear interpolation operator is designed

to be second order accurate in order to preserve the order of approximation of the scheme. It is also exact for normal unknowns in the sense that $\mathfrak{P}_{\mathcal{D}}(\mathfrak{I}_{\mathcal{D}}(\mathfrak{v}_{\mathcal{D}})) = \mathfrak{v}_{\mathcal{D}}$ for all $\mathfrak{v}_{\mathcal{D}} \in \mathfrak{X}_{\mathcal{D}}$. Finally, it is local in the sense that it computes the displacement field $\boldsymbol{v}_\sigma$ at a given face $\sigma \in \mathcal{E}^{\rm int} \cup \mathcal{E}_{\rm N}^{\rm ext}$ in terms of a given number of normal components $\boldsymbol{v}_{\sigma'}\cdot\boldsymbol{n}_{\sigma'}$ taken in a stencil $\mathcal{S}_\sigma \subset \mathcal{E}$ of neighboring faces σ' of σ (imposing that $\sigma \in \mathcal{S}_\sigma$). An example of construction of such an interpolator is given below. Another example can be found in [58] in the context of large-eddy simulation (LES).

Given a face $\sigma \in \mathcal{E}^{\rm int} \cup \mathcal{E}_{\rm N}^{\rm ext}$, for each component $i \in [\![1,d]\!]$ of the displacement field $\boldsymbol{v}_\sigma$, we look for a linear interpolation of the form

$$\overline{v}_\sigma^i(\boldsymbol{x}) = \sum_{j=1}^d \alpha_\sigma^{ij} x_j + \beta_\sigma^i.$$

In order to determine the $d(d+1)$ coefficients $(\alpha_\sigma^{ij})_{i,j\in[\![1,d]\!]}$, $(\beta_\sigma^i)_{i\in[\![1,d]\!]}$ as linear combinations of normal components $\boldsymbol{v}_{\sigma'}\cdot\boldsymbol{n}_{\sigma'}$, $\sigma' \in \mathcal{S}_\sigma$, we hence look for a set $\mathcal{S}_\sigma$ of $d(d+1)$ neighboring faces σ' of the face σ (imposing that $\sigma \in \mathcal{S}_\sigma$) such that the system of equations $\overline{\boldsymbol{v}}_\sigma(\overline{\boldsymbol{x}}_{\sigma'})\cdot\boldsymbol{n}_{\sigma'} = \boldsymbol{v}_{\sigma'}\cdot\boldsymbol{n}_{\sigma'}$ is nonsingular. The set $\mathcal{S}_\sigma$ is built using the following greedy algorithm:

1. initialization: for a given number $l > d(d+1)$, we select the l closest neighboring faces of the face σ which are sorted from the closest to the furthest using the distance between their barycenter and $\overline{\boldsymbol{x}}_\sigma$: $\sigma_0 = \sigma, \sigma_1, \ldots, \sigma_{l-1}$. We set $\mathcal{S}_\sigma = \{\sigma\}$ and $q = 1$, $k = 0$;
2. while $q < d(d+1)$ and $k < l-1$:
   (a) $k \leftarrow k+1$;
   (b) if the equation $\overline{\boldsymbol{v}}_\sigma(\overline{\boldsymbol{x}}_{\sigma_k})\cdot\boldsymbol{n}_{\sigma_k} = \boldsymbol{v}_{\sigma_k}\cdot\boldsymbol{n}_{\sigma_k}$ is linearly independent from the set of equations $\overline{\boldsymbol{v}}_\sigma(\overline{\boldsymbol{x}}_{\sigma'})\cdot\boldsymbol{n}_{\sigma'} = \boldsymbol{v}_{\sigma'}\cdot\boldsymbol{n}_{\sigma'}$ for all $\sigma' \in \mathcal{S}_\sigma$, then $\mathcal{S}_\sigma \leftarrow \mathcal{S}_\sigma \cup \{\sigma_k\}$; $q \leftarrow q+1$;
3. if $q < d(d+1)$, the algorithm is rerun with a larger value for l.

Note that imposing $\sigma \in \mathcal{S}_\sigma$ guarantees as required the property $\mathfrak{P}_{\mathcal{D}}(\mathfrak{I}_{\mathcal{D}}(\mathfrak{v}_{\mathcal{D}})) = \mathfrak{v}_{\mathcal{D}}$ for all $\mathfrak{v}_{\mathcal{D}} \in \mathfrak{X}_{\mathcal{D}}$. At the end of the process, the tangential component(s) of the displacement on a face $\sigma \in \mathcal{E}^{\rm int} \cup \mathcal{E}_{\rm N}^{\rm ext}$ is (are) obtained as $\overline{\boldsymbol{v}}_\sigma(\overline{\boldsymbol{x}}_\sigma)\cdot\boldsymbol{t}_\sigma^i$, where $(\boldsymbol{t}_\sigma^i)_{i\in[\![1,d-1]\!]}$ defines an orthonormal basis of the face σ.

The use of the interpolation operator will bring two improvements to the discretization: first a reduction of the number of unknowns and secondly a stabilization of the discretization (rigidity adding).

C.3 Discrete variational formulation

Recalling the notation $H^1_{\rm D}(\Omega) := \{v \in H^1(\Omega) \mid v_{|\Gamma_{\rm D}} = 0\}$, we first write the weak formulation of the continuous problem (C.1). We assume that $\boldsymbol{f} \in L^2(\Omega)^d$, $\boldsymbol{u}_{\rm D} \in H^{\frac12}(\Gamma_{\rm D})^d$, and $\boldsymbol{g} \in H^{\frac12}(\Gamma_{\rm N})^d$. The weak problem reads: Find $\boldsymbol{u} \in H^1(\Omega)^d$ such that $\boldsymbol{u} = \boldsymbol{u}_{\rm D}$ on $\Gamma_{\rm D}$ and such that

$$a(\boldsymbol{u},\boldsymbol{v}) = (\boldsymbol{f},\boldsymbol{v})_{0,\Omega} + (\boldsymbol{g},\boldsymbol{v})_{0,\Gamma_{\rm N}} \qquad \forall \boldsymbol{v} \in H^1_{\rm D}(\Omega)^d, \tag{C.6}$$

where $a(\boldsymbol{w},\boldsymbol{v}) := 2\mu(\underline{\underline{\varepsilon}}(\boldsymbol{w}),\underline{\underline{\varepsilon}}(\boldsymbol{v}))_{0,\Omega} + \lambda(\boldsymbol{\nabla}\cdot\boldsymbol{w},\boldsymbol{\nabla}\cdot\boldsymbol{v})_{0,\Omega}$. The weak formulation (C.6) could be exclusively written in $H^1_{\rm D}(\Omega)^d$ thanks to the introduction of a lifting operator for $\boldsymbol{u}_{\rm D} \in H^{\frac12}(\Gamma_{\rm D})^d$.

We introduce the following discrete operators $\boldsymbol{\nabla}_{\mathcal{D}}\cdot : \boldsymbol{X}_{\mathcal{D},{\rm D}} \to \mathbb{P}^0_d(\mathcal{P})$, $\underline{\underline{\varepsilon_{\mathcal{D}}}} : \boldsymbol{X}_{\mathcal{D},{\rm D}} \to \mathbb{P}^0_d(\mathcal{P})^{d,d}$ and $\underline{\underline{\sigma_{\mathcal{D}}}} : \boldsymbol{X}_{\mathcal{D},{\rm D}} \to \mathbb{P}^0_d(\mathcal{P})^{d,d}$ such that, for all $\boldsymbol{v}_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},{\rm D}}$,

$$\boldsymbol{\nabla}_{\mathcal{D}}\cdot\boldsymbol{v}_{\mathcal{D}} := {\rm tr}(\boldsymbol{\nabla}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}}), \quad \underline{\underline{\varepsilon_{\mathcal{D}}}}(\boldsymbol{v}_{\mathcal{D}}) := \frac12(\boldsymbol{\nabla}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}} + \boldsymbol{\nabla}_{\mathcal{D}}\boldsymbol{v}_{\mathcal{D}}{}^T), \quad \underline{\underline{\sigma_{\mathcal{D}}}}(\boldsymbol{v}_{\mathcal{D}}) := 2\mu\,\underline{\underline{\varepsilon_{\mathcal{D}}}}(\boldsymbol{v}_{\mathcal{D}}) + \lambda\boldsymbol{\nabla}_{\mathcal{D}}\cdot\boldsymbol{v}_{\mathcal{D}}\underline{\underline{I_d}},$$

where we make use of the Definition C.1 of the gradient operator.

Then, the discrete variational formulation reads: Find $\mathbf{u}_\mathcal{D} \in \mathfrak{X}_\mathcal{D}$ such that $\mathbf{u}_\sigma = \boldsymbol{u}_\mathrm{D}^\sigma$ for all $\sigma \in \mathcal{E}_\mathrm{D}^\mathrm{ext}$ and such that

$$a_\mathcal{D}(\mathbf{u}_\mathcal{D}, \mathfrak{v}_\mathcal{D}) = (\boldsymbol{f}, \Pi_\mathcal{D} \circ \mathfrak{I}_\mathcal{D}(\mathfrak{v}_\mathcal{D}))_{0,\Omega} + (\boldsymbol{g}, T_\mathrm{D} \circ \mathfrak{I}_\mathcal{D}(\mathfrak{v}_\mathcal{D}))_{0,\Gamma_\mathrm{N}} \qquad \forall \mathfrak{v}_\mathcal{D} \in \mathfrak{X}_{\mathcal{D},\mathrm{D}}, \tag{C.7}$$

where $a_\mathcal{D}(\mathfrak{w}_\mathcal{D}, \mathfrak{v}_\mathcal{D}) := (\underline{\underline{\sigma_\mathcal{D}}} \circ \mathfrak{I}_\mathcal{D}(\mathfrak{w}_\mathcal{D}), \underline{\underline{\varepsilon_\mathcal{D}}} \circ \mathfrak{I}_\mathcal{D}(\mathfrak{v}_\mathcal{D}))_{0,\Omega}$, and where $\boldsymbol{u}_\mathrm{D}^\sigma := \int_\sigma \boldsymbol{u}_\mathrm{D} / |\sigma|$ is an average value.

It is important to keep in mind that, as numerical experiments tend to confirm, searching for a discrete solution in $\boldsymbol{X}_\mathcal{D}$ using the bilinear form $a_\mathcal{D}$ without interpolation leads to an unstable scheme with vanishing eigenvalues on triangular or tetrahedral meshes, especially for mixed-type boundary conditions. As we will observe in the next section, the introduction of the interpolation operator seems to provide a stabilization of the discrete formulation. In a way, adding rigidity to the system makes more likely the fact that a discrete Korn's inequality may hold on $\mathfrak{X}_{\mathcal{D},\mathrm{D}}$. Note however that, up to now, no theoretical justification is available. This stabilization technique is an alternative to a least-square penalization of discrete functions jumps as presented in Chapter IV. This alternative does not require to define a gradient-based piecewise affine reconstruction and allows to remain in the finite volume spirit.

As far as computational costs are concerned, the interpolation of the tangential component(s) of the displacement on mesh faces leads to a drastic reduction of the number of unknowns. Note also that cell unknowns $(\mathbf{u}_K)_{K\in\mathcal{M}}$ can be easily eliminated without any fill in since their are only related in each cell to the face unknowns of the cell. This further reduces the number of degrees of freedom to the faces normal component of the displacement only. However, a substantial drawback linked to this technique is the important stencil of neighboring faces that is needed for the construction. This increases the calculation (owing to the resolution of local problems) and assembling times, and deteriorates the matrix conditioning.

Remark C.2 (Quasi-incompressible materials)**.** *When considering a quasi-incompressible material ($\lambda \to +\infty$), the discrete bilinear form of problem* (C.7) *can be modified, using the notation introduced in Definition C.1, as*

$$a_\mathcal{D}(\mathfrak{w}_\mathcal{D}, \mathfrak{v}_\mathcal{D}) := 2\mu(\underline{\underline{\varepsilon_\mathcal{D}}} \circ \mathfrak{I}_\mathcal{D}(\mathfrak{w}_\mathcal{D}), \underline{\underline{\varepsilon_\mathcal{D}}} \circ \mathfrak{I}_\mathcal{D}(\mathfrak{v}_\mathcal{D}))_{0,\Omega} + \lambda \sum_{K\in\mathcal{M}} |K| \boldsymbol{\nabla}_K \cdot \mathfrak{I}_\mathcal{D}(\mathfrak{w}_\mathcal{D}) \boldsymbol{\nabla}_K \cdot \mathfrak{I}_\mathcal{D}(\mathfrak{v}_\mathcal{D}),$$

which enables to take advantage of the existence of a Fortin operator, thus guaranteeing the locking-free aspect of the discretization. To define the Fortin operator, for $\boldsymbol{v} \in H^1(\Omega)^d$, we first introduce $\boldsymbol{v}_\mathcal{D} \in \boldsymbol{X}_\mathcal{D}$ such that

$$\boldsymbol{X}_\mathcal{D} \ni \boldsymbol{v}_\mathcal{D} := \left(\left(\int_K \boldsymbol{v} / |K| \right)_{K\in\mathcal{M}}, \left(\int_\sigma \boldsymbol{v} / |\sigma| \right)_{\sigma\in\mathcal{E}} \right).$$

Then, the Fortin operator is given by $\mathfrak{X}_\mathcal{D} \ni \mathfrak{v}_\mathcal{D} := \mathfrak{P}_\mathcal{D}(\boldsymbol{v}_\mathcal{D})$. The cell-wise conservation property of divergence for the above interpolator can be proved as in Lemma III.5 and Corollary III.1, accounting for the fact that normal face unknowns are preserved by the projection operator $\mathfrak{P}_\mathcal{D}$ and the interpolation operator $\mathfrak{I}_\mathcal{D}$. As far as the H^1-stability property of the Fortin operator is concerned, we both use the fact that $|\mathfrak{v}_\mathcal{D}|_{\mathfrak{X}_\mathcal{D}} \leqslant |\boldsymbol{v}_\mathcal{D}|_{\boldsymbol{X}_\mathcal{D}}$ according to (C.5)*, and the fact that $|\boldsymbol{v}_\mathcal{D}|_{\boldsymbol{X}_\mathcal{D}} \leqslant C_\mathrm{S} \|\boldsymbol{\nabla} \boldsymbol{v}\|_{0,\Omega}$ with $C_\mathrm{S} > 0$ independent of $\mathcal{D}$, which is a consequence of [38, (3.38)],* (III.1) *and Lemma III.1.*

Remark C.3 (Coupling with piecewise constant pressures)**.** *Let define the following bilinear form on $\mathfrak{X}_{\mathcal{D},\mathrm{D}} \times P_\mathcal{D}$, where $P_\mathcal{D} := \mathbb{P}_d^0(\mathcal{M})$:*

$$b_\mathcal{D}(\mathfrak{v}_\mathcal{D}, q_\mathcal{D}) := -\sum_{K\in\mathcal{M}} q_K \sum_{\sigma\in\mathcal{E}_K\setminus\mathcal{E}_\mathrm{D}^\mathrm{ext}} |\sigma| \, \mathfrak{v}_\sigma^n (\boldsymbol{n}_\sigma \cdot \boldsymbol{n}_{K,\sigma}).$$

This bilinear form may appear in poroelastic mechanics-flow couplings when considering mixed-type mechanical boundary conditions (then the use of $\mathfrak{X}_{\mathcal{D},\mathrm{D}}$ enables to stabilize the linear elasticity model without penalizing jumps). Owing to the existence of a Fortin operator on $\mathfrak{X}_{\mathcal{D},\mathrm{D}}$, cf. Remark C.2, this bilinear form satisfies a (uniform) inf-sup condition.

C.4 Numerical experiments

Let $\Omega := (0,1)^d$. The convergence of the scheme (C.7) is assessed in two and three space dimensions for an exact solution such that, for $i \in \{1,\dots,d\}$,

$$u_i = e^{\cos\left(\sum_{j=1}^d \alpha_{ij} x_j\right)},$$

where $\underline{\underline{\alpha}}$ is a $d \times d$ tensor to be precised. The medium is homogeneous with (constant) Lamé parameters such that $\lambda = \mu = 1$. The right-hand side $\boldsymbol{f}$ is obtained as the divergence of the stress tensor and Dirichlet boundary conditions are imposed on the whole boundary. The stabilization parameter is chosen in a finite volume spirit such that $\eta = d^{1/2}$ in (C.4), and the H^1 and L^2 relative errors on the displacement are computed as

$$\frac{\|\boldsymbol{\nabla} \boldsymbol{u} - \boldsymbol{\nabla}_{\mathcal{D}} \boldsymbol{u}_{\mathcal{D}}\|_{0,\Omega}}{\|\boldsymbol{\nabla} \boldsymbol{u}\|_{0,\Omega}} \approx \frac{\left(\sum_{K\in\mathcal{K}_h} |K| \, |(\boldsymbol{\nabla} \boldsymbol{u})(\boldsymbol{x}_K) - \boldsymbol{\nabla}_K \boldsymbol{u}_{\mathcal{D}}|^2\right)^{1/2}}{\left(\sum_{K\in\mathcal{K}_h} |K| \, |(\boldsymbol{\nabla} \boldsymbol{u})(\boldsymbol{x}_K)|^2\right)^{1/2}},$$

$$\frac{\|\boldsymbol{u} - \Pi_{\mathcal{D}} \boldsymbol{u}_{\mathcal{D}}\|_{0,\Omega}}{\|\boldsymbol{u}\|_{0,\Omega}} \approx \frac{\left(\sum_{K\in\mathcal{K}_h} |K| \, |\boldsymbol{u}(\boldsymbol{x}_K) - \boldsymbol{u}_K|^2\right)^{1/2}}{\left(\sum_{K\in\mathcal{K}_h} |K| \, |\boldsymbol{u}(\boldsymbol{x}_K)|^2\right)^{1/2}},$$

where $\boldsymbol{u}_{\mathcal{D}} := \mathfrak{I}_{\mathcal{D}}(\mathfrak{u}_{\mathcal{D}})$ with $\mathfrak{u}_{\mathcal{D}} \in \mathfrak{X}_{\mathcal{D}}$ solution to (C.7). For the sake of simplicity, these relative errors are referred to as $\|\boldsymbol{\nabla} \boldsymbol{u} - \boldsymbol{\nabla}_{\mathcal{D}} \boldsymbol{u}_{\mathcal{D}}\|_{0,\Omega}$ and $\|\boldsymbol{u} - \Pi_{\mathcal{D}} \boldsymbol{u}_{\mathcal{D}}\|_{0,\Omega}$ in the plots axes. The implementation for the two-dimensional case is based on the same `C++` prototype as in Sections IV.4 and V.4, while the results in 3D have been obtained using a Fortran prototype developed by Roland Masson.

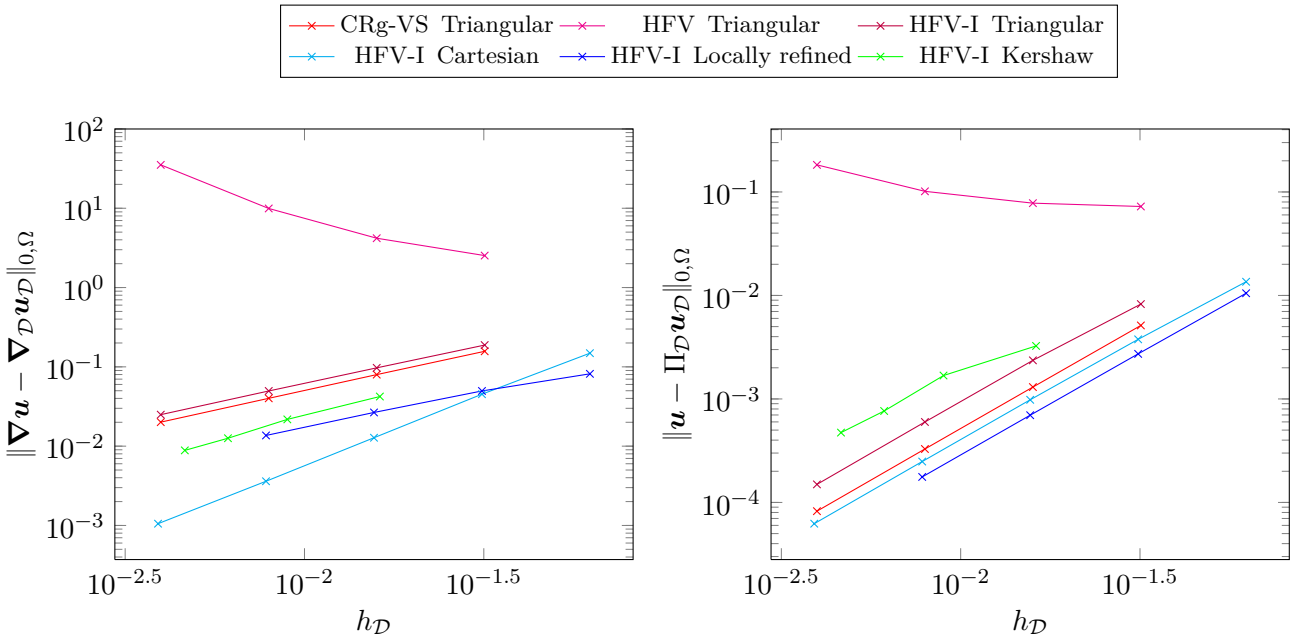
C.4.1 A two-dimensional test-case

Let set

$$\underline{\underline{\alpha}} := \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}.$$

We consider the matching triangular (cf. Figure IV.1a), Cartesian (cf. Figure IV.1b), locally refined Cartesian (cf. Figure IV.1c), and Kershaw-type (cf. Figure IV.1d) mesh sequences of the FVCA5 benchmark. We solve problem (C.7) and plot on Figure C.1 the H^1 and L^2 relative errors for the different mesh sequences. The method is referred to as HFV-I (for Interpolation). For comparison, we include the results for the CRg-VS method (IV.5)–(IV.6) (with $\eta = d$) and for the HFV method without interpolation (referred to as HFV) on the matching triangular mesh sequence.

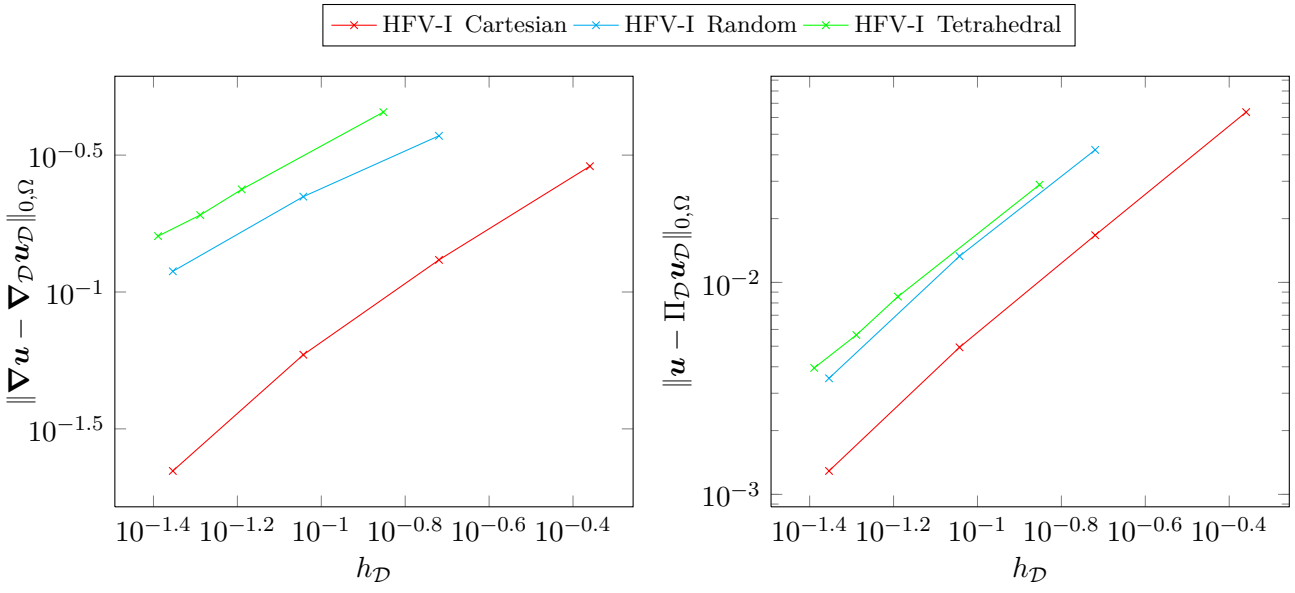
We first observe that the HFV method (without interpolation) does not converge on triangular mesh sequences, even for pure Dirichlet boundary conditions. This result is not surprising in the light of Section II.1.2.1 and Proposition III.2. Note that the HFV method (without interpolation) practically converges on Cartesian mesh sequences. The method HFV-I (with interpolation) defines a convergent scheme for any mesh sequence tested here. The expected



(a) H^1 relative error (b) L^2 relative error

Figure C.1: Convergence results for the HFV-I method on the two-dimensional test-case of Section C.4.1.

convergence orders (one in the H^1 seminorm and two in the L^2 norm) are obtained for all kinds of sequences with a supra-convergent behavior on Cartesian meshes. In addition, the errors compare to those obtained with the CRg-VS method on triangular meshes.



(a) H^1 relative error (b) L^2 relative error

Figure C.2: Convergence results for the HFV-I method on the three-dimensional test-case of Section C.4.2.

C.4.2 A three-dimensional test-case

Let set

$$\underline{\underline{\alpha}} := \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ -1 & 1 & 2 \end{pmatrix}.$$

Let consider the Cartesian (mesh family A), randomly distorted (mesh family AA), and matching tetrahedral (mesh family B) mesh sequences from the FVCA6 3D benchmark (cf. the link to get some illustrations). The randomly distorted sequence has the particularity to possess nonplanar faces. We solve problem (C.7) and plot on Figure C.2 the H^1 and L^2 relative errors for the different mesh sequences.

The results of Figure C.2 exhibit a first order convergence in the H^1 seminorm and second order convergence in the L^2 norm for the HFV-I method on the three different mesh sequences.

Table des figures

Résumé. Cette thèse s'intéresse à la conception de méthodes de discrétisation non-conforme pour un modèle de poromécanique. Le but de ce travail est de simplifier les couplages liant la géomécanique d'un milieu poreux à l'écoulement polyphasique compositionnel ayant cours en son sein tels qu'ils sont réalisés actuellement dans l'industrie pétrolière, en discrétisant sur un même maillage, typiquement non-conforme car à l'image de la lithologie, la mécanique et l'écoulement. La nouveauté consiste donc à traiter la mécanique par une méthode d'approximation non-conforme sur maillages généraux. Dans cette thèse, nous nous concentrons sur un modèle d'élasticité linéaire. Les difficultés inhérentes à son approximation non-conforme sont son manque de coercivité (se traduisant par la nécessité de satisfaire une inégalité de Korn sur un espace discret discontinu), ainsi que le phénomène de verrouillage numérique lorsque le matériau tend à devenir incompressible. Dans une première partie, nous construisons un espace d'approximation sur maillages généraux, s'apparentant à une extension de l'espace de Crouzeix–Raviart. Nous explicitons ses propriétés d'approximation et de conformité, et montrons que ce dernier est adapté à une discrétisation primale coercive et robuste au *locking* du modèle d'élasticité sur maillages généraux. La méthode proposée est moins coûteuse que son équivalent éléments finis (en termes de propriétés) $\mathbb{P}_2^d$. Nous nous intéressons dans une deuxième partie à l'approximation non-conforme d'un modèle couplé de poroélasticité. Nous étudions la convergence d'une famille de schémas numériques dont la discrétisation en espace utilise le formalisme des schémas Gradient, auquel appartient la méthode développée pour la mécanique. Nous prouvons la convergence de telles approximations vers la solution de régularité minimale du problème continu, indépendamment des paramètres physiques du système.

Mots-clés : poroélasticité, méthodes non-conformes, maillages généraux, volumes finis, verrouillage numérique, problèmes de point-selle

Abstract. This manuscript focuses on the conception of nonconforming discretization methods for a poromechanical model. The aim of this work is to ease the coupling between the geomechanics and the multiphase compositional Darcy flow in porous media by discretizing mechanics and flow on the same mesh, typically nonconforming as it represents the lithology. Hence, the novelty hinges on a nonconforming treatment of mechanics on general meshes. In this work, we focus on a linear elasticity model. The nonconforming approximation of such a model is not straightforward owing to its lack of coercivity (meaning that a discrete Korn's inequality must hold on a discontinuous discrete space) and to the numerical locking phenomenon occurring as the material becomes incompressible. In a first part, we design an approximation space on general meshes, which can be viewed as an extension of the so-called Crouzeix–Raviart space. We study its approximation and conformity properties, and prove that this latter is well-adapted to the design of a primal, coercive, and locking-free discretization of the elasticity model on general meshes. The proposed method is less costly than its finite element equivalent (in terms of properties) $\mathbb{P}_2^d$. In a second part, we tackle the nonconforming approximation of a coupled poroelasticity model. We study the convergence of a family of numerical schemes whose space discretization relies on the Gradient schemes framework, to which belongs the method developed for mechanics. We prove the convergence of such approximations toward the minimal regularity solution of the continuous problem, and independently of the choice of physical parameters.

Keywords: poroelasticity, nonconforming methods, general meshes, finite volumes, numerical locking, saddle-point problems

Laboratoire d'accueil : UMR 8050 LAMA Université Paris-Est Marne-la-Vallée