Finite volume methods on general meshes for nonlinear evolution systems

Konstantin Brenner

N° d'ordre: 1234

THÈSE

Présentée pour obtenir

LE GRADE DE DOCTEUR EN SCIENCES
DE L'UNIVERSITÉ PARIS-SUD XI

Spécialité: Mathématiques

par

Konstantin BRENNER

Méthodes de volumes finis sur maillages quelconques pour des systèmes d'évolution non linéaires

Soutenue le 8 Novembre 2011 devant la Commission d'examen:

M.	Alexandre	ERN
M.	Robert	EYMARD
Mme.	Danielle	HILHORST
M.	Pascal	OMNES
M.	Roland	MASSON
M.	Mazen	SAAD
M.	Marc	BONNET

Thèse préparée au
Département de Mathématiques d'Orsay
Laboratoire de Mathématiques (UMR 8628), Bât. 425
Université Paris-Sud 11
91 405 Orsay CEDEX

Méthodes de volumes finis sur maillages quelconques pour des systèmes d'évolution non linéaires

Résumé

Les travaux de cette thèse portent sur des méthodes de volumes finis sur maillages quelconque pour la discrétisation de problèmes d'évolution non linéaires modélisant le transport de contaminants en milieu poreux et les écoulements diphasiques.

Au Chapitre 1, nous étudions une famille de schémas numériques pour la discrétisation d'une équation parabolique dégénérée de convection-reaction-diffusion modélisant le transport de contaminants dans un milieu poreux qui peut être hétérogène et anisotrope. La discrétisation du terme de diffusion est basée sur une famille de méthodes qui regroupe les schémas de volumes finis hybrides, de différences finies mimétiques et de volumes finis mixtes. Le terme de convection est traité à l'aide d'une famille de méthodes qui s'appuient sur les inconnues hybrides associées aux interfaces du maillage. Cette famille contient à la fois les schémas centré et amont. Les schémas que nous étudions permettent une discrétisation localement conservative des termes d'ordre un et d'ordre deux sur des maillages arbitraires en dimensions d'espace deux et trois. Nous démontrons qu'il existe une solution unique du problème discret qui converge vers la solution du problème continu et nous présentons des résultats numériques en dimensions d'espace deux et trois, en nous appuyant sur des maillages adaptatifs.

Au Chapitre 2, nous proposons un schéma de volumes finis hybrides pour la discrétisation d'un problème d'écoulement diphasique incompressible et immiscible en milieu poreux. On suppose que ce problème a la forme d'une équation parabolique dégénérée de convection-diffusion en saturation couplée à une équation uniformément elliptique en pression. On considère un schéma implicite en temps, où les flux diffusifs sont discrétisés par la méthode des volumes finis hybride, ce qui permet de pouvoir traiter le cas d'un tenseur de perméabilité anisotrope et hétérogène sur un maillage très général, et l'on s'appuie sur un schéma de Godunov pour la discrétisation des flux convectifs, qui peuvent être non monotones et discontinus par rapport aux variables spatiales. On démontre l'existence d'une solution discrète, dont une sous-suite converge vers une solution faible du problème continu. On présente finalement des cas test bidimensionnels.

Le Chapitre 3 porte sur un problème d'écoulement diphasique, dans lequel la courbe de pression capillaire admet des discontinuité spatiales. Plus précisément on suppose que l'écoulement prend place dans deux régions du sol aux propriétés très différentes, et l'on suppose que la loi de pression capillaire est discontinue en espace à la frontière entre les deux régions, si bien que la saturation de l'huile et la pression globale sont discontinues à travers cette frontière avec des conditions de raccord non linéaires à l'interface. On discrétise le problème à l'aide d'un schéma, qui coïncide avec un schéma de volumes finis standard dans chacune des deux régions, et on démontre la convergence d'une solution approchée vers une solution faible du problème continu. Les test numériques présentés à la fin du chapitre montrent que le schéma permet de reproduire le phénomène de piégeage de la phase huile.

Mots-clefs : Diffusion hétérogène et anisotrope, maillages non conformes, schémas de volumes finis, équations paraboliques dégénérées de convection–réaction–diffusion, écoulements diphasiques, pression capillaire discontinue.

Finite volume methods on general meshes for nonlinear evolution systems

Abstract

In Chapter 1 we study a family of finite volume schemes for the numerical solution of degenerate parabolic convection-reaction-diffusion equations modeling contaminant transport in porous media. The discretization of possibly anisotropic and heterogeneous diffusion terms is based upon a family of numerical schemes, which include the hybrid finite volume scheme, the mimetic finite difference scheme and the mixed finite volume scheme. One discretizes the convection term by means of a family of schemes which makes use of the discrete unknowns associated to the mesh interfaces, and contains as special cases an upwind scheme and a centered scheme. The numerical schemes which we study are locally conservative and allow computations on general multi-dimensional meshes. We prove that the unique discrete solution converges to the unique weak solution of the continuous problem. We also investigate the solvability of the linearized problem obtained during Newton iterations. Finally we present a number of numerical results in space dimensions two and three using nonconforming adaptive meshes and show experimental orders of convergence for upwind and centered discretizations of the convection term.

In Chapter 2 we propose a finite volume method on general meshes for the numerical simulation of an incompressible and immiscible two-phase flow in porous media. We consider the case that it can be written as a coupled system involving a degenerate parabolic convection-diffusion equation for the saturation together with a uniformly elliptic equation for the global pressure. The numerical scheme, which is implicit in time, allows computations in the case of a heterogeneous and anisotropic permeability tensor. The convective fluxes, which are non monotone with respect to the unknown saturation and discontinuous with respect to the space variables, are discretized by means of a special Godunov scheme. We prove the existence of a discrete solution which converges, along a subsequence, to a solution of the continuous problem. We present a number of numerical results in space dimension two, which confirm the efficiency of the numerical method.

Chapter 3 is devoted to the study of a two-phase flow problem in the case that the capillary pressure curve is discontinuous with respect to the space variable. More precisely we assume that the porous medium is composed of two different rocks, so that the capillary pressure is discontinuous across the interface between the rocks. As a consequence the oil saturation and the global pressure are discontinuous across the interface with nonlinear transmission conditions. We discretize the problem by means of a numerical scheme which reduces to a standard finite volume scheme in each sub-domain and prove the convergence of a sequence of approximate solutions towards a weak solution of the continuous problem. The numerical tests show that the scheme can reproduce the oil trapping phenomenon.

Keywords : Heterogeneous anisotropic diffusion, nonconforming grids, finite volume schemes, degenerate parabolic convection–reaction–diffusion equation, contaminant transport with adsorption, flow in porous media, two-phase flow, convergence of approximate solutions, discontinuous capillarity.

Remerciements

Mes premiers remerciements vont naturellement à Madame Danielle Hilhorst, qui a dirigé ma thèse, et qui m'a également fait découvrir l'analyse numérique et a guidé avec une constante sollicitude mes premiers pas d'enseignant-chercheur.

Ses remarquables connaissances scientifiques, ses qualités pédagogiques rares ansi que la profonde humanité dont elle a fait preuve, tant pour mon accueil en France que lors de l'élaboration de ma thèse, font de madame Danielle Hilhorst une directrice de thèse exceptionnelle.

Je souhaiterais ensuite remercier chaleureusement, d'une part Monsieur Robert Eymard et Monsieur Mazen Saad qui m'ont fait l'honneur d'accepter d'être les rapporteurs de ma thèse et d'autre part Monsieur Alexandre Ern, Monsieur Roland Masson et Monsieur Pascal Omnes qui ont accepté de participer au jury. Leur investissement et l'intérêt qu'ils portent à mes travaux ainsi que la richesse des échanges que nous avons eus lors de nos rencontres à l'occasion de conférences m'encouragent dans le domaine de recherche que je me suis fixé et font de ces mathématiciens un exemple très inspirant.

Je voudrais aussi exprimer toute ma reconnaissance à Monsieur Marc Bonnet, membre de jury invité de ma thèse, pour l'intense et fructueuse collaboration de ces trois dernières années. Son expertise ainsi que ses conseils avisés m'ont apporté une véritable initiation à l'hydrogéologie, tout en me donnant l'opportunité de mettre en pratique les méthodes numériques étudiées dans cette thèse.

Le dernier chapitre de cette thèse m'a permis de collaborer avec Monsieur Clément Cancès que je tiens à remercier : ses qualités scientifiques et humaines ont fait de ce travail un moment remarquable.

Mes remerciements vont également à toute l'équipe d'Analyse Numérique et d'Equation aux Dérivées Partielles de l'Université Paris-Sud pour son accueil chaleureux lors de ces trois dernières années.

Enfin mes pensées vont à ma famille : à mon père Mikhael Brenner, à ma mère Elena Koustova et à ma sœur Emilia Koustova, qui sont et ont toujours été une source infinie d'encouragements, d'inspiration et d'amour. Je remercie tous mes proches qui m'ont soutenu lors de ces trois années, surtout Christine et Leonid. Sans leur soutien et leur patience cette thèse n'aurait jamais vu le jour.

A mes parents

Table des matières

Introduction

Contexte général

Les problèmes d'écoulement et de transport en milieu poreux représentent un domaine de recherche à la fois d'un grand intérêt scientifique et d'une importance essentielle pour la pratique, de part leur vaste domaine d'applications. Un exemple est donné par l'écoulement de l'eau dans les nappes aquifères, qui sont essentielles à l'alimentation des villes en eau potable ; la protection et le développement des ressources en eau représentent un défi environnemental majeur, ce qui rend nécessaires une compréhension profonde des écoulements et du transport de contaminants dans les nappes aquifères, ainsi que le développement d'outils appropriés pour les simulations numériques.

D'autre part, le développement intensif de l'industrie pétrolière dans les dernières décennies a créé une demande d'analyse mathématique et numérique d'écoulements multiphasiques. Il s'agit d'appréhender la coexistence de plusieurs fluides, comme par exemple l'huile et l'eau, en milieu poreux. Les simulations numériques peuvent servir à améliorer le taux de récupération des hydrocarbures lors de l'exploitation d'un réservoir pétrolier ou de prédire leur mouvement naturel à l'échelle d'un bassin.

On remarque que, dans les applications pratiques, le milieu poreux est souvent hétérogène et anisotrope, et qu'il peut avoir de plus une géométrie multi-dimensionnelle complexe. Une telle complexité doit être prise en compte par les méthodes numériques utilisées, ce qui est l'une des motivations principales de cette thèse. On étudie ici deux problèmes issus de la géologie, le transport réactif avec adsorption en milieu poreux et un problème d'écoulement diphasique immiscible et incompressible. Dans les deux cas il est nécessaire de résoudre une équation non linéaire parabolique dégénérée, qui est de plus couplée avec une équation elliptique dans le cas d'un écoulement diphasique.

Notre étude s'appuie sur la méthode des volumes finis [65]. On peut évidemment envisager d'appliquer le schéma de volumes finis le plus simple, appelé encore "schéma à deux points" [65], pour la discrétisation d'equations de convection-réaction-diffusion et celle d'écoulements diphasiques [90], [67]. Ce schéma possède des propriétés mathématiques très élégantes, comme en particulier un principe de maximum discret dans de nombreux cas. De plus il est peu coûteux, puisqu'il n'utilise qu'une seule inconnue par maille et qu'il possède un "stencil" compact (5 points sur les maillages rectangulaires bidimensionnels et 9 points sur les maillages tridimensionnelles hexahédriques). En revanche ce schéma ne s'applique que sur des maillages satisfaisant une condition d'orthogonalité et qu'à des problèmes elliptiques ou paraboliques où le tenseur de diffusion se réduit à un scalaire ; de façon plus générale, le maillage doit être aligné avec les directions principales du tenseur de diffusion. Si l'on souhaite s'affranchir de la condition d'orthogonalité, ou pouvoir traiter également le cas de tenseurs de diffusion hétérogènes et anisotropes, une possibilité est de combiner l'application d'une méthode de type éléments finis pour la discrétisation du terme de diffusion avec un schéma de volumes finis pour la discrétisation de tous les autres termes, en construisant également un maillage dual. C'est cette solution qui a été adoptée par [72] et [73]. Notons que dans [73] le maillage dual d'éléments finis triangulaires ou tétrahédriques est construit à partir d'un maillage de volumes finis assez général, qui peut être non conforme.

Une autre possibilité est de s'appuyer sur une famille de schémas de volumes finis développée par Aavatsmark et ses collaborateurs, [1], [2], [3]. Ces schémas sont connus sous l'abréviation MPFA correspondant à l'appellation "multi-point flux approximation". Cette classe de schémas permet la discrétisation de termes de diffusion anisotropes et hétérogènes sur des maillages très généraux. Toutefois, ils ne conduisent pas toujours à des formes bilinéaires coercives.

Plusieurs variantes de méthodes de volumes finis sur maillages quelconques ont été développées au cours des dernières années, comme en témoigne le benchmark sur la diffusion anisotrope et hétérogène organisé à l'occasion des conferences FVCA 5 [79] et FVCA 6 [70]. La méthode de volumes finis hybrides (méthode SUSHI) a été proposée par [66] et généralisée par [58]. Il a de plus été démontré [58] que si l'on applique un tel schéma à un problème elliptique, on obtient une formulation équivalente à une version généralisée de schémas de différences finies mimétiques, [23], [30], [32], [28], [102], [89], et à des schémas de volumes finis mixtes [57], [88], [43], ce qui rend possible d'envisager une méthode hybrid mimetic mixed (HMM) qui réunisse les trois. Ces schémas permettent tous la discrétisation de termes de diffusion anisotropes et très hétérogènes sur des maillages arbitraires. En général, ils nécessitent d'avantage d'inconnues discrètes que le schéma de volumes finis standard ou que les schémas MPFA, comme par exemple une inconnue par maille et une par interface. Notons qu'il est souvent possible de diminuer le nombre d'inconnues, soit de façon algébrique en éliminant les inconnues des mailles, soit de manière approchée [66], [87] en éliminant les inconnues d'interfaces. Dans des travaux récents [101] et [56], la méthode (HMM) a été appliquée à la discrétisation d'une équation stationnaire de convection-diffusion.

Cette thèse est composée de trois chapitres indépendants. Au Chapitre 1, nous étudions une famille des schémas numériques pour une équation parabolique dégénérée de convection-reaction-diffusion modélisant le transport de contaminants en milieu poreux avec adsorption. La discrétisation est basée sur une famille de méthodes de volumes finis analysées par [101]. Le terme de convection est discrétisé à l'aide de méthodes utilisant les inconnues hybrides, associées aux interfaces de maillage. Cette famille contient à la fois les schémas centré et amont. Un avantage est qu'elle permet une discrétisation localement conservative des termes d'ordre un et d'ordre deux sur un seul maillage et qu'elle peut être facilement implémentée en dimensions d'espace deux et trois. Au Chapitre 2, nous nous appuyons sur des résultats obtenus au Chapitre 1 et étudions un schéma de type volumes finis sur maillage quelconque pour un problème d'écoulement diphasique en milieu poreux, qui a la forme d'une équation parabolique dégénérée de convection-diffusion couplée à une équation uniformément elliptique. Le Chapitre 3 porte sur un problème d'écoulement diphasique, dans lequel la courbe de pression capillaire admet des discontinuités spatiales.

Chapitre 1 : Méthode de volumes finis sur maillages quelconques pour une équation de convection-réaction-diffusion

Au Chapitre 1 nous étudions une famille de méthodes de volumes finis sur maillages quelconques pour l'équation parabolique dégénérée de convection-réaction-diffusion

$$\frac{\partial \beta(u)}{\partial t} - \nabla \cdot (\mathbf{\Lambda} \nabla u) + \nabla \cdot (\mathbf{V} u) + f(u) = q \text{ dans } \Omega \times (0, T). \tag{0.1}$$

Cette équation joue un rôle important dans différents domaines d'applications comme par exemple le transport de contaminants en milieux poreux avec adsorption [86], [22]. La vitesse de l'écoulement V, qui est supposée connue dans le contexte de ce chapitre, est souvent calculée à l'aide de la loi de Darcy. Dans le cadre du transport de contaminants, la fonction inconnue u représente la concentration d'un contaminant dissout, qui, en contraste avec la partie adsorbée, est mobile et transporté par le flux souterrain. La matrice pleine $\mathbf{\Lambda}$ est un tenseur de diffusion-dispersion, la fonction f représente les réactions chimiques, et q est un terme source. Dans notre étude, nous prenons aussi en compte le phénomène d'adsorption du contaminant par la matrice poreuse en supposant que ce processus est très rapide par rapport au déplacement du contaminant. Plus précisément on suppose que la partie dissoute du contaminant est en équilibre avec celle qui est adsorbée. Dans le cas d'un écoulement saturé, les pores sont remplis de fluide et la fonction β peut s'écrire sous la forme

$$\beta(\mathbf{x}, u) = \phi(\mathbf{x}) u + \rho(\mathbf{x}) a(u),$$

où $\phi(\mathbf{x})$ est la porosité, $\rho(\mathbf{x})$ la densité volumique du milieu et la fonction $a(u)$ représente la concentration d'une phase adsorbée et immobile. Dans la pratique la fonction $a(u)$ peut être ou bien donnée par l'isotherme de Langmuir

$$a(u) = \frac{k_1 u}{1 + k_2 u}, \quad k_1, k_2 > 0,$$

ou bien par l'isotherme de Freundlich

$$a(u) = \mu_1 u^{\mu_2}, \quad \mu_1 > 0, 0 < \mu_2 < 1.$$

On remarque que dans ce dernier cas on a $a'(0) = +\infty$, d'où $\beta_u(\mathbf{x}, 0) = +\infty$, si bien que l'équation (0.1) est parabolique dégénérée.

Dans les applications pratiques les coefficients du tenseur de diffusion-dispersion $\mathbf{\Lambda} = \mathbf{\Lambda}(\mathbf{x}, \mathbf{V})$ sont typiquement donnés par

$$\begin{aligned} \mathbf{\Lambda}_{ii} &= \phi(\mathbf{x}) \left(\alpha_t |\mathbf{V}| + (\alpha_l - \alpha_t) \tfrac{\mathbf{V}_i^2}{|\mathbf{V}|} + \alpha_m \right) \\ \mathbf{\Lambda}_{ij} &= \phi(\mathbf{x}) (\alpha_l - \alpha_t) \tfrac{\mathbf{V}_i \mathbf{V}_j}{|\mathbf{V}|}. \end{aligned}$$

Les schémas hybride, mixte et mimétique (applelés schémas HMM [58]) ont été initialement proposés pour les problèmes elliptiques. L'idée essentielle de ces trois schémas est d'introduire les analogues discrets de notions continues (le flux de gradient, la divergence, le gradient discret, les produits scalaires, la formule de Stokes) et de prescrire des relations consistantes entre eux. Le problème est ensuite considéré sous une forme variationnelle mixte. Le schéma résultant peut être considéré comme un schéma de volumes finis car il utilise les flux discrets à travers les interfaces du maillage et on impose que ces flux soient continus. D'autre part ce schéma s'apparente aux méthodes de type éléments finis puisqu'il s'appuie sur une formulation faible du problème discret.

Dans la formulation hybride les inconnues discrètes représentent une approximation de la solution dans les éléments de volumes et aux interfaces du maillage, et la forme bilinéaire associée au problème variationnel discret est coercive et symétrique. Ceci permet d'établir une borne supérieure de la norme dans H^1-discret de la solution approchée et d'obtenir des estimations de différences de translatées un espace uniformes par rapport à la taille de la discrétisation. Plus précisément, nous nous appuyons sur [58] pour la discrétisation d'équations paraboliques dégénérées. L'analyse est basée sur les mêmes outils que dans [58] ou [101] en ce que concerne la discrétisation spatiale. On démontre tout d'abord que la solution discrète est uniformément bornée dans l'espace $L^\infty(0, T; L^2(\Omega))$ puis dans en espace analogue à $L^2(0, T; H^1(\Omega))$. Les estimation a priori permettent de déduire l'existence et l'unicité de la solution approchée. On démontre ensuite des estimations uniformes sur les différences de translatées en espace et en temps. Ces estimations permettent de déduire du théorème de Fréchet–Kolmogorov la compacité relative dans $L^2(\Omega \times (0, T))$ d'une famille des solutions approchées et d'établir ensuite la convergence forte d'une sous-suite dans $L^2(\Omega \times (0, T))$ vers une solution faible du problème continu. Après avoir démontré la convergence du schéma, nous effectuons des tests numériques en dimensions deux et trois d'espace. En particulier nous comparons numériquement les applications du schéma amont et du schéma centré pour la discrétisation du terme de convection. On démontre également que l'utilisation des maillage adaptatifs (non conformes) peut considérablement réduire le coût de calcul. A la fin du Chapitre 1, on discute de la stratégie à aborder pour la résolution du système discret non linéaire. Nous étudions l'inversibilité du problème linéarisé provenant de la méthode de Newton, ce qui n'est pas trivial à cause du caractère parabolique dégénéré de l'équation (0.1). Dans le cas d'une discrétisation monotone du terme de convection, on montre qu'un tel problème est inversible.

Chapitre 2 : Méthode de volumes finis généralisée pour un problème d'écoulement diphasique

Au Chapitre 2, nous étudions un schéma de volumes finis hybrides pour la résolution d'un problème d'écoulement diphasique immiscible et incompressible en milieu poreux. Le problème a la forme d'une loi de conservation pour chaque phase où nous avons substitué la loi de Darcy diphasique

$$\omega \partial_t s_n - \nabla \cdot \left(\mathbf{K} \frac{kr_n(s_n)}{\mu_n} (\nabla p_n - \rho_n \mathbf{g}) \right) = k_n \text{ dans } \Omega \times (0, T), \tag{0.2}$$

$$\omega \partial_t s_w - \nabla \cdot \left(\mathbf{K} \frac{kr_w(s_w)}{\mu_w} (\nabla p_w - \rho_w \mathbf{g}) \right) = k_w \text{ dans } \Omega \times (0, T) \tag{0.3}$$

et où s_w et s_n correspondent aux saturations des phases mouillante et non mouillante. On remarque que dans le cas d'un écoulement huile - eau, c'est la phase mouillante qui correspond à l'eau et la phase non mouillante aux hydrocarbures. Les termes de perméabilité relative kr_n et kr_w dans la loi de Darcy diphasique modélisent la coexistence des deux fluides dans le milieu poreux. On a de plus

$$s_n + s_w = 1, \tag{0.4}$$

et l'on suppose que les pressions des phases sont liées par la loi de pression capillaire

$$p_n + p_w = \pi(s). \tag{0.5}$$

On introduit finalement la notion de pression globale p (cf. [14] et [44])

$$p = p_w + \int_0^s \frac{\lambda_n(a)}{\lambda_n(a) + \lambda_w(a)} \pi'(a) da$$

où $s = s_n$ est la saturation de la phase non mouillante. Si l'on définit de plus le flux fractionnel f, la mobilité totale λ et la diffusion capillaire φ

$$\lambda(s) = \frac{kr_n(s)}{\mu_n} + \frac{kr_w(1-s)}{\mu_n}, \quad f(s) = \frac{kr_n(s)}{kr_n(s) + \frac{\mu_n}{\mu_w} kr_w(1-s)},$$

$$\varphi(s) = \int_0^s \frac{kr_n(a) kr_w(a)}{\mu_w kr_n(a) + \mu_o kr_w(a)} \pi'(a) da,$$

on déduit que le système (0.2)-(0.5) est équivalent au système

$$-\nabla \cdot \mathbf{q} = k_w + k_n, \quad \mathbf{q} = -\mathbf{K} \left(\lambda(s) \nabla p - \xi(s) \mathbf{g}\right) \text{ dans } \Omega \times (0,T), \tag{0.6}$$

$$\omega \frac{\partial s}{\partial t} + \nabla \cdot (\mathbf{q} f(s) + \gamma(s) \mathbf{K} \mathbf{g}) - \nabla \cdot (\mathbf{K} \nabla \varphi(s)) = k_n \text{ dans } \Omega \times (0,T), \tag{0.7}$$

où

$$\gamma(s) = (\rho_o - \rho_w) \frac{\lambda_n(s) \lambda_w(s)}{\lambda_n(s) + \lambda_w(s)} \qquad \text{and} \qquad \xi(s) = (\lambda_n(s) \rho_o + \lambda_w(s) \rho_w).$$

On suppose que $\lambda(s) = \dfrac{kr_n(s)}{\mu_n} + \dfrac{kr_w(1-s)}{\mu_n} \geq \underline{\lambda} > 0$, ce que implique que les deux premières équations sont équivalentes à une équation uniformément elliptique en la pression globale p, tandis que la troisième équation est parabolique dégénérée en la saturation s.

Une grande variété de méthodes numériques ont été appliquées au problème diphasique ; citons en particulier la méthode des éléments finis (cf. [48], [50], [51], [16], [44], [64] et [63]), la méthode de Galerkin discontinu (cf. e.g. [19], [61]) et la méthode des volumes finis standard (cf. e.g. [90], [67], [5], [12], [91]). Des résultats de convergence du schéma numérique ont été démontrés par [16], [46] et [51] pour la méthode des éléments finis, et par [8], [90] et [71] dans le cas de la méthode des volumes finis. De plus on remarque que dans [71] le problème à été considéré dans sa formulation initiale (0.2)-(0.5), sans faire intervenir la notion de pression globale. Citons également des résultats de convergence obtenus pour le problème de déplacement miscible [94], [43], [97].

Nous étudions ici un schéma de type volumes finis hybrides implicite en temps pour le système (0.6)-(0.7). La discrétisation des termes d'ordre deux est basée sur la méthode des volumes finis sur maillage quelconque [66], si bien que le tenseur de perméabilité absolu peut être anisotrope et hétérogène. Comme le tenseur $\mathbf{K}$ est discontinu en espace aux interfaces entre deux milieux, les termes $\mathbf{K}(\mathbf{x})\mathbf{g}\xi(\cdot)$ et $\mathbf{q}f(\cdot) + \mathbf{K}(\mathbf{x})\mathbf{g}\gamma(\cdot)$ le sont également, si bien qu'ils nécessitent un traitement adapté (cf e.g. [100], [99], [75]). On utilise ici le schéma de Godunov proposé par [82], en s'appuyant sur les inconnues d'interfaces.

Le système continu (0.6)-(0.7) possède une structure mathématique remarquable, qui se transmet au schéma discret et nous permet d'obtenir des estimations dans un espace discret analogue à $L^\infty(0,T;H^1(\Omega))$ pour la pression globale approchée $p_{\mathcal{D},\delta t}$, ainsi que dans des espaces discrets analogues aux espaces $L^2(0,T;H^1(\Omega))$ et $L^\infty(0,T;H^1(\Omega))$ pour la saturation approchée $s_{\mathcal{D},\delta t}$. Nous démontrons ensuite l'existence d'une solution approchée $(s_{\mathcal{D},\delta t}, p_{\mathcal{D},\delta t})$ ainsi que des estimations sur les différences de translatées en temps et en espace de la saturation discrète $s_{\mathcal{D},\delta t}$; on en déduit sa compacité relative par le théorème de Fréchet–Kolmogorov. Si l'on note $\{s_{\mathcal{D},\delta t}, p_{\mathcal{D},\delta t}\}$ une famille de solutions du problème discret, nous démontrons qu'il existe une sous-suite qui converge vers une solution faible du problème continu quand le diamètre de la discrétisation spatiale et le pas de temps tendent vers zero. Il s'agit d'une convergence $L^2(\Omega \times (0,T))$ forte en saturation et faible en pression globale. Nous présentons finalement des résultats de simulations numériques.

Chapitre 3 : Écoulement diphasique avec pression capillaire discontinue

On considère de nouveau le système (0.2)-(0.5), dans le cas où l'écoulement prend place dans deux régions du sol Ω_1 et Ω_2 aux propriétés très différentes, et l'on suppose que la loi de pression capillaire $\pi = \pi(x,s)$ est discontinue en espace à la frontière entre les deux régions. Cela peut conduire à un phénomène de piégeage de la phase non-mouillante, cf. [59], [24], [40], ce qui s'exprime par le fait que l'huile ne peut pas traverser la frontère entre les deux régions du sol tandis que c'est possible pour l'eau. De plus la saturation s et le pression globale p sont discontinues à travers cette frontière avec des conditions de raccord non linéaires à l'interface. On néglige ici l'anisotropie du tenseur de perméabilité $\mathbf{K}$. En s'appuyant sur un formalisme de fonctions multivoques proposé par [39], on obtient le système d'équations

$$\begin{cases} \nabla \cdot \mathbf{q}_i = 0, \quad \mathbf{q}_i = -(\lambda_i(s_i)\nabla p_i - \mathbf{g}_i\xi(s_i)) & \text{dans } \Omega_i, \\ \omega\dfrac{\partial s_i}{\partial t} + \nabla \cdot (\mathbf{q}_i f_i(s) + \mathbf{g}\gamma_i(s_i)) - \nabla \cdot (\nabla\varphi_i(s_i)) = 0 & \text{dans } \Omega_i, \\ \displaystyle\sum_{i\in\{1,2\}} \mathbf{K}_i \cdot \mathbf{n}_{\Gamma,i} = 0 & \text{sur } \Gamma, \\ \displaystyle\sum_{i\in\{1,2\}} (\mathbf{K}_i f_i(s) + \mathbf{g}\gamma_i(s_i) - \nabla\varphi_i(s_i)) \cdot \mathbf{n}_{\Gamma,i} = 0 & \text{sur } \Gamma, \\ \tilde{\pi} \in \tilde{\pi}_1(s_1) \cap \tilde{\pi}_2(s_2) & \text{sur } \Gamma, \\ p_1 - Z_1(\pi) = p_2 - Z_2(\pi) & \text{sur } \Gamma, \end{cases}$$

où $\tilde{\pi}_i(s), i = 1,2$ sont des graphes monotones, représentant des courbes de pression capillaire discontinues au travers de l'interface $\Gamma = \overline{\Omega}_1 \cap \overline{\Omega}_2$. On remarque que dans chaque sous-domaine Ω_i, il s'agit de la résolution d'un système de la forme (0.6)-(0.7). Plusieurs méthodes de simulation numérique ([62], [60], [40], [61]) ont été proposées pour approcher ce problème. La convergence d'un schéma numérique a été démontrée dans [40] dans le cas d'un domaine unidimensionnel où la gravité n'est pas prise en compte ; il est alors possible d'éliminer le pression globale et de remplacer le système (0.2)-(0.5) par une équation parabolique dégénérée scalaire. A notre connaissance le résultat de convergence n'avait été jamais établi pour le problème complet en dimension d'espace supérieure. On applique le schéma de volumes finis "flux à deux points" [65] pour la discrétisation spatiale.

Après avoir proposé un schéma numérique, nous démontrons des estimations a priori dans un espace discret analogue à l'espace $L^2(0,T;H^1(\Omega_i))$ sur la pression globale $P_{\mathcal{D},i}$ et la diffusion capillaire $\varphi(s_{\mathcal{D},i}), i=1,2$. (les notations diffèrent légèrement de celles du Chapitre 2). De plus on peut montrer que la saturation $s_{\mathcal{D},i}$ est uniformément bornée, ou plus précisément que $0 \leq s_{\mathcal{D},i} \leq 1, i=1,2$. On démontre ensuite l'existence d'une solution discrète et la convergence d'une sous-suite vers une solution faible du problème continu en s'appuyant sur des arguments liés à la compacité.

Première partie

A finite volume method on general meshes for a degenerate parabolic convection-reaction-diffusion equation

Abstract We propose a finite volume method on general meshes for the discretization of a degenerate parabolic convection-reaction-diffusion equation. Equations of this type arise in many contexts, such as for example the modeling of contaminant transport in porous media. The diffusion term, which can be anisotropic and heterogeneous, is discretized using a recently developed hybrid mimetic mixed framework. We construct a family of discretizations for the convection term, which uses the hybrid interface unknowns. We consider a wide range of unstructured possibly nonmatching polyhedral meshes in arbitrary space dimension. The scheme is fully implicit in time, it is locally conservative and robust with respect to the Péclet number. We obtain a convergence result based upon a priori estimates and the Fréchet–Kolmogorov compactness theorem. We implement the scheme both in two and three space dimensions and compare the numerical results obtained with the upwind and the centered discretizations of the convection term numerically.

This work was supported by the GdR MoMaS (PACEN/CNRS, ANDRA, BRGM, CEA, EdF, IRSN), France.

1 Introduction

In this paper we study a finite volume method on general meshes for degenerate parabolic convection-reaction-diffusion equations of the form

$$\frac{\partial \beta(u)}{\partial t} - \nabla \cdot (\mathbf{\Lambda} \nabla u) + \nabla \cdot (\mathbf{V} u) + f(u) = q. \tag{1.1}$$

Equations of this type arise for instance in the modeling of contaminant transport in porous media. The unknown function u represents the concentration of the dissolved species, which diffuses and is transported by the groundwater. An essential element in our study is the process of adsorption by a porous skeleton, which is supposed to be very fast. More particularly we suppose that the dissolved and the absorbed parts of the species are in equilibrium ; this is modeled by the function β, where β' may be infinite in several points. The matrix $\mathbf{\Lambda}$ is a possibly anisotropic and heterogeneous diffusion-dispersion tensor, $\mathbf{V}$ is the velocity field, the function f stands for the chemical reactions, and q is the source term. We suppose that the mesh is quite general, and possibly nonmatching. Therefore, also in view of the anisotropy in the diffusion term, we can not apply the standard finite volume method [65].

Finite volume schemes have often been applied to the equation (1.1), see e.g. [5], [20], [67], [43]. The upwind discretization of the convection term allows finite volume schemes to be stable in convection dominated case, however standard finite volume schemes do not permit to handle anisotropic diffusion on general meshes. On the other hand finite element method allows a very simple discretization of full diffusion tensors, they were used a lot for the discretization of equation (1.1), see e.g. [16], [53], [54]. A possible solution is to split equation (1.1) into a hyperbolic part and a parabolic part, by means of an operator splitting method ; one can find such an analysis in [84], [85], where the advection term was treated by the method of characteristics. The other quite intuitive idea is to take "best from both worlds" [72], which leads to combined finite volume-finite element schemes ; we refer to [72] for this approach. In order to solve this class of equations, Eymard, Hilhorst and Vohralík [72] discretize the diffusion term by means of piecewise linear nonconforming (Crouzeix–Raviart) finite elements over a triangularization of the space domain, or using the stiffness matrix of the hybridization of the lowest order Raviart–Thomas mixed finite element method. The other terms are discretized by means of a finite volume scheme on a dual mesh, where the dual volumes are constructed around the sides of the original triangularization. In the second paper of Eymard et al. [73] the time evolution, convection, reaction, and sources terms are discretized on a given grid, which can be nonmatching and can contain nonconvex elements, by means of a cell-centered finite volume method. In order to discretize the diffusion term, they construct a conforming simplicial mesh with vertices given by the original grid and use the finite element method. In this way, the scheme is fully consistent and the discrete solution is naturally continuous across the interfaces between the subdomains with nonmatching grids, without introducing supplementary equations and unknowns nor interpolating the discrete solutions at the interfaces.

The finite volume methods for the discretization of anisotropic diffusion on general meshes is a subject of wide interest (see for instance the results of the benchmarks organized at the FVCA 5 conference [79]). The hybrid finite volume (HFV) scheme on general meshes (also known as SUSHI) was first proposed in [66] and then generalized in [58]. It has also been shown in [58], that the HFV method applied to the stationary diffusion problem is equivalent to some generalized mimetic finite difference (MFD), [23], [30], [32], [28], [102], [89], or a mixed finite volume (MFV), [57], [88], [43] methods, so that one can think of a global hybrid mimetic mixed (HMM) framework. In the recent publications [101] and [56] the HMM method was applied to the stationary convection-diffusion equation. A family of discretization for a convection term was proposed and analyzed both theoretically and numerically. In this paper we study the family of the HHM methods applied to a nonlinear degenerate parabolic equation. In general we exploit the HFV point of view; however since the diffusion and convection operators in (1.1) are linear with respect to u, the numerical scheme which we consider can also be seen as a MFD or a MFV scheme. We discretize a convection term using a family of scheme involving the hybrid unknowns. This family contains the hybrid equivalent of the classical upwind and central schemes. In Section 2 we present the nonlinear parabolic problem together with its geological context. In Section 3 we describe the numerical scheme. We prove a priori estimates for the discrete solution in $L^\infty(0,T;L^2(\Omega))$ and in a discrete space analogous to the space $L^2(0,T;H^1(\Omega))$ in Sections 4 and we show the existence and uniqueness of the solution of the discrete scheme in Section 5. In Section 6, we prove an estimate on differences of time and space translates. These estimates imply a relative compactness property of sequences of approximate solutions by the Fréchet–Kolmogorov theorem. We deduce the strong convergence in L^2 of the approximate solutions to the unique solution of the continuous problem in Section 7. For the proofs, we apply methods inspired upon those of [65] and [66]. In Section 8, we finally present results of numerical tests, which confirm the validity of the numerical method, both in space dimensions two and three. Finally in Section 9 we investigate the solvability of the linearized problem obtained during Newton iterations.

2 Parabolic degenerate problem

We consider the parabolic degenerate convection-diffusion-reaction problem

$$(\mathcal{P}b)\begin{cases} \dfrac{\partial \beta(u)}{\partial t} - \nabla\cdot(\mathbf{\Lambda}(\mathbf{x})\nabla u) + \nabla\cdot(\mathbf{V}(\mathbf{x})u) + f(u) = q(\mathbf{x},t), \quad (\mathbf{x},t)\in Q_T, \\ u(\mathbf{x},t) = 0 \quad \mathbf{x}\in\partial\Omega, \quad t\in(0,T), \\ u(\mathbf{x},0) = u_0(\mathbf{x}), \quad \mathbf{x}\in\Omega, \end{cases}$$

where Ω is a bounded open connected polyhedral subset of $\mathbb{R}^d$, $d \in \mathbb{N}^*$, $T > 0$ and $Q_T = \Omega \times (0,T)$.

The parabolic equation in Problem $(\mathcal{P}b)$ models contaminant transport in a saturated porous medium ; in this case the velocity of the underlying flow $\mathbf{V}$ satisfies the Darcy's problem. In practical applications one has to solve both flow and transport problems (see for example [43] or [25]). However, for the sake of simplicity we assume that the velocity field $\mathbf{V}$ is given.

The coefficients of the diffusion-dispersion tensor $\mathbf{\Lambda} = \mathbf{\Lambda}(\mathbf{x},\mathbf{V})$ are usually given by

$$\begin{aligned} \mathbf{\Lambda}_{ii} &= \phi(\mathbf{x})\left(\alpha_t|\mathbf{V}| + (\alpha_l - \alpha_t)\tfrac{\mathbf{V}_i^2}{|\mathbf{V}|} + \alpha_m\right) \\ \mathbf{\Lambda}_{ij} &= \phi(\mathbf{x})(\alpha_l - \alpha_t)\tfrac{\mathbf{V}_i\mathbf{V}_j}{|\mathbf{V}|}, \end{aligned}$$

where α_l, α_t are the longitudinal and transverse dispersively, α_m is a molecular diffusion, and ϕ is the porosity [22]. In order to take into account the adsorption of a contaminant by the porous medium we define β as

$$\beta(\mathbf{x},u) = \phi(\mathbf{x})u + \rho(\mathbf{x})a(u),$$

where $\rho(\mathbf{x})$ is a bulk density of the porous medium and $a(u)$ stands for the concentration of the adsorbed phase (cf. [86] and [22]). We assume that the process of adsorption is fast compare to the displacement of the contaminant, so that the mobile phase u and the immobile phases $a(u)$ are in equilibrium. The function $a(u)$ can for example be given by Langmuir isotherm

$$a(u) = \frac{k_1 u}{1 + k_2 u}, \quad k_1, k_2 > 0,$$

or by Freundlich isotherm

$$a(u) = \mu_1 u^{\mu_2}, \quad \mu_1 > 0, 0 < \mu_2 < 1.$$

Remark that in the last case $a'(0) = +\infty$, so that $\beta_u(\mathbf{x},0) = +\infty$. The term $f(u)$ represents the chemical reactions ; in the case that the contaminant is a radioactive species, $f(u)$ take the form

$$f(u) = \lambda\beta(u),$$

where λ is a positive decay constant (cf. [22]).

In general we suppose that the datum satisfies the following hypotheses :

$(\mathcal{H}_1)$ $\beta \in C(\mathbb{R}), \beta(0) = 0$ is a strictly increasing function, which satisfies the growth condition $|\beta(a)-\beta(b)| \geq \underline{\beta}|a-b|$, $\underline{\beta} > 0$ for all $a, b \in \mathbb{R}$; moreover there exist $P > 0$ and C_β, such that $|\beta(u)| \leq C_\beta$ for $|u| \leq P$ and β is a Lipschitz continuous with a constant $\overline{\beta}$ for $|u| \geq P$;
$(\mathcal{H}_2)$ $\mathbf{\Lambda} \in (L^\infty(\Omega))^{2d}$ is such that for a.e. $\mathbf{x} \in \Omega$, $\mathbf{\Lambda}(\mathbf{x})$ is a symmetric positive definite matrix with the set of eigenvalues included in $[\underline{\lambda}, \overline{\lambda}]$;
$(\mathcal{H}_3)$ $\mathbf{V} \in H(\mathrm{div}, \Omega) \bigcap \mathrm{L}^\infty(\Omega)$ is such that $\nabla \cdot \mathbf{V} \geq 0$ a.e. in Ω ;
$(\mathcal{H}_4)$ $f \in C(\mathbb{R})$, $f(0) = 0$ and there exists $M > 0$ such that $uf(u) > 0$ and $f(u)$ is Lipschitz continuous function with constant L_f for all $u < 0$ or $u > M$; moreover we suppose that f does not decrease too fast i.e. there exists $\underline{f} > 0$ such that $(f(u) - f(v))(u - v) \geq -\underline{f}(u - v)^2$ for all $u, v \in \mathbb{R}$;
$(\mathcal{H}_5)$ $q \in L^2(Q_T)$.
$(\mathcal{H}_6)$ $u_0 \in L^\infty(\Omega)$;
We now present a definition of a weak solution of Problem $(\mathcal{P}b)$.

Definition 2.1. *We say that a function u is a weak solution of Problem $(\mathcal{P}b)$ if*
(i) $u \in L^2(0, T; H_0^1(\Omega))$;
(ii) $\beta(u) \in L^\infty(0, T; L^2(\Omega))$;
(iii) u satisfies the integral equality

$$-\int_0^T \int_\Omega \beta(u)\psi_t \, d\mathbf{x}dt - \int_\Omega \beta(u_0)\psi(\cdot, 0) \, d\mathbf{x} + \int_0^T \int_\Omega \mathbf{\Lambda}\nabla u \cdot \nabla\psi \, d\mathbf{x}dt$$
$$-\int_0^T \int_\Omega u\mathbf{V} \cdot \nabla\psi \, d\mathbf{x}dt + \int_0^T \int_\Omega f(u)\psi \, d\mathbf{x}dt = \int_0^T \int_\Omega q\psi \, d\mathbf{x}dt$$

for all $\psi \in L^2(0, T; H_0^1(\Omega))$ with $\psi_t \in L^\infty(Q_T)$, $\psi(\cdot, T) = 0$.

Remark 2.1. *In the case that the reaction function f is nondecreasing, the uniqueness of the weak solution of Problem $(\mathcal{P}b)$ follows from [86].*

3 The numerical scheme

3.1 The main definitions

In order to describe the numerical scheme we introduce below some notations related to the space and time discretizations.

Definition 3.1 (Discretization of Ω)**.** *Let Ω be a polyhedral open bounded connected subset of $\mathbb{R}^d$, with $d \in \mathbb{N}^*$, and $\partial\Omega = \overline{\Omega}\backslash\Omega$ its boundary. A discretization of Ω, denoted by $\mathcal{D}$, is defined as the triplet $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$, where :*

1. $\mathcal{M}$ is a finite family of non empty connex open disjoint subsets of Ω (the "control volumes") such that $\overline{\Omega} = \bigcup_{K\in\mathcal{M}} \overline{K}$. For any $K \in \mathcal{M}$, let $\partial K = \overline{K}\backslash K$ be the boundary of K ; we define $m(K) > 0$ as the measure of K and h_K as the diameter of K.

2. $\mathcal{E}$ is a finite family of disjoint subsets of $\overline{\Omega}$ (the "edges" of the mesh), such that, for all $\sigma \in \mathcal{E}$, σ is a non empty open subset of a hyperplane of $\mathbb{R}^d$, whose $(d-1)$-dimensional measure $m(\sigma)$ is strictly positive. We also assume that, for all $K \in \mathcal{M}$, there exists a *subset $\mathcal{E}_K$ of $\mathcal{E}$ such that $\partial K = \bigcup_{\sigma\in\mathcal{E}_K}\overline{\sigma}$. For each $\sigma \in \mathcal{E}$, we set $\mathcal{M}_\sigma = \{K \in \mathcal{M} | \sigma \in \mathcal{E}_K\}$. We then assume that, for all $\sigma \in \mathcal{E}$, either $\mathcal{M}_\sigma$ has exactly one element and then $\sigma \in \partial\Omega$ (the set of these interfaces called boundary interfaces, is denoted by $\mathcal{E}_{ext}$) or $\mathcal{M}_\sigma$ has exactly two elements (the set of these interfaces called interior interfaces, is denoted by $\mathcal{E}_{int}$). For all $\sigma \in \mathcal{E}$, we denote by $\mathbf{x}_\sigma$ the barycenter of σ. For all $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$, we denote by $\mathbf{n}_{K,\sigma}$ the outward normal unit vector.*

3. $\mathcal{P}$ is a family of points of Ω indexed by $\mathcal{M}$, denoted by $\mathcal{P} = (\mathbf{x}_K)_{K\in\mathcal{M}}$, such that for all $K \in \mathcal{M}$, $\mathbf{x}_K \in K$; moreover K is assumed to be $\mathbf{x}_K$-star-shaped, which means that for all $\mathbf{x} \in K$, there holds $[\mathbf{x}_K, \mathbf{x}] \in K$. Denoting by $d_{K,\sigma}$ the Euclidean distance between $\mathbf{x}_K$ and the hyperplane containing σ, one assumes that $d_{K,\sigma} > 0$. We denote by $D_{K,\sigma}$ the cone of vertex $\mathbf{x}_K$ and basis σ.

Next we introduce some extra notations related to the mesh. The size of the discretization $\mathcal{D}$ is defined by

$$h_{\mathcal{D}} = \sup_{K\in\mathcal{M}} diam(K); \tag{3.1}$$

moreover we define

$$\theta_{\mathcal{D}} = \max\Big(\max_{\sigma\in\mathcal{E}_{int},\{K,L\}=\mathcal{M}_\sigma} \frac{d_{K,\sigma}}{d_{L,\sigma}}, \max_{K\in\mathcal{M}_\sigma,\sigma\in\mathcal{E}_K} \frac{h_K}{d_{K,\sigma}}\Big). \tag{3.2}$$

Thus imposing a uniform bound on $\theta_{\mathcal{D}}$ forces the meshes to be sufficiently regular. As it was done in [66] we associate with the mesh the following spaces of discrete unknowns

$$\begin{aligned} X_{\mathcal{D}} &= \{((v_K)_{K\in\mathcal{M}}, (v_\sigma)_{\sigma\in\mathcal{E}}), v_K \in \mathbb{R}, v_\sigma \in \mathbb{R}\}, \\ X_{\mathcal{D},0} &= \{v \in X_{\mathcal{D}} \text{ such that } (v_\sigma)_{\sigma\in\mathcal{E}_{ext}} = 0\}. \end{aligned} \tag{3.3}$$

Moreover, for each function $\psi = \psi(\mathbf{x})$ smooth enough we define $P_{\mathcal{D}}\psi \in X_D$ in following way

$$\begin{aligned} (P_{\mathcal{D}}\psi)_K &= \psi(\mathbf{x}_K) \quad \text{for all} \quad K \in \mathcal{M}, \\ (P_{\mathcal{D}}\psi)_\sigma &= \psi(\mathbf{x}_\sigma) \quad \text{for all} \quad \sigma \in \mathcal{E}. \end{aligned}$$

Definition 3.2 (Time discretization). *We divide the time interval $(0, T)$ into N equal time steps of length $\delta t = T/N$, where δt is the uniform time step defined by $\delta t = t_n - t_{n-1}$.*

Taking into account the time discretization leads us to define of the following discrete spaces

$$X_{\mathcal{D},\delta t} = X_{\mathcal{D}}^N = \{(v^n)_{n\in\{1,\dots,N\}}, v^n \in X_{\mathcal{D}}\}$$

and

$$X_{\mathcal{D},0,\delta t} = X_{\mathcal{D},0}^N = \{(v^n)_{n\in\{1,\dots,N\}}, v^n \in X_{\mathcal{D},0}\}.$$

Remark 3.1. *For the sake of simplicity, we restrict our study to the case of constant time steps. Nevertheless all results presented below can be easily extended to the case of a non uniform time discretization.*

3.2 Finite volume scheme

After formally integrating the first equation of $(\mathcal{P}b)$ on the domain $K \times (t_{n-1}, t_n)$ for each $K \in \mathcal{M}$ and $n = 1, \ldots, N$, we obtain

$$\int_K \beta(u(\mathbf{x}, t_n)) - \beta(u(\mathbf{x}, t^{n-1}))\, d\mathbf{x} + \sum_{\sigma \in \mathcal{E}_K} \int_{t_{n-1}}^{t_n} \int_\sigma (-\mathbf{\Lambda} \nabla u + \mathbf{V} u) \cdot \mathbf{n}_{K,\sigma}\, d\gamma dt$$
$$+ \int_{t_{n-1}}^{t_n} \int_K f(u)\, d\mathbf{x} dt = \int_{t_{n-1}}^{t_n} \int_K q\, d\mathbf{x} dt.$$

For all $K \in \mathcal{M}$ and all $\sigma \in \mathcal{E}_K$ we define

$$V_{K,\sigma} = \int_\sigma \mathbf{V} \cdot \mathbf{n}_{K,\sigma} d\gamma \text{ and } q_K^n = \frac{1}{\delta t\ m(K)} \int_{t_{n-1}}^{t_n} \int_K q\, d\mathbf{x} dt.$$

Next, let $F_{K,\sigma}^D(u^n) \approx -\int_\sigma \mathbf{\Lambda} \nabla u \cdot \mathbf{n}_{K,\sigma} d\gamma$ and $F_{K,\sigma}^C(u^n) \approx \int_\sigma V u \cdot \mathbf{n}_{K,\sigma} d\gamma$ be an approximation of the diffusive and convective fluxes respectively, which are defined below by (3.16) and (3.18). The time implicit finite volume scheme corresponding to Problem $(\mathcal{P}b)$ is given by :

The initial condition

$$u_K^0 = \frac{1}{m(K)} \int_K u_0(\mathbf{x})\, d\mathbf{x} \tag{3.4}$$

for all $K \in \mathcal{M}$.

The discrete equations

$$\begin{aligned} m(K)(\beta(u_K^n) - \beta(u_K^{n-1})) + \delta t \sum_{\sigma \in \mathcal{E}_K} F_{K,\sigma}^D(u^n) + \delta t \sum_{\sigma \in \mathcal{E}_K} F_{K,\sigma}^C(u^n) \\ + \delta t\ m(K) f(u_K^n) = \delta t\ m(K) q_K^n \end{aligned} \tag{3.5}$$

for all $K \in \mathcal{M}$. We remark that for each time step the number of equations is $card(\mathcal{M})$, whereas the number of discrete unknowns is equal to $card(\mathcal{M}) + card(\mathcal{E})$. Therefore we need to introduce $card(\mathcal{E})$ additional equations corresponding to the interface values. For boundary faces these equations are obtained by writing the discrete analog of the Dirichlet boundary condition

$$u_\sigma^n = 0 \quad \text{for all} \quad \sigma \in \mathcal{E}_{ext}. \tag{3.6}$$

For interior faces, we follow the main idea of the finite volume method by imposing the local conservation of the discrete fluxes (e.g. [58], [101])

$$(F_{K,\sigma}^D(u^n) + F_{K,\sigma}^C(u^n)) + (F_{L,\sigma}^D(u^n) + F_{L,\sigma}^C(u^n)) = 0 \tag{3.7}$$

for all $\sigma \in \mathcal{E}_{int}$ with $\mathcal{M}_\sigma = \{K, L\}$.

3.3 The discrete weak formulation

We will define below $F_{K,\sigma}^D$ and $F_{K,\sigma}^C$ in more details, but we first give an alternative variational formulation of the discrete scheme (3.4)-(3.7). Let $(v^n)_{n \in \{1,\ldots,N\}}$ be an arbitrary sequence of elements of $X_{\mathcal{D},0}$; multiplying equation (3.5) by v_K^n and summing on all control volumes $K \in \mathcal{M}$ leads to :

$$\sum_{K\in\mathcal{M}} m(K)v_K^n \frac{\beta(u_K^n)-\beta(u_K^{n-1})}{\delta t} + \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}(v_K^n F_{K,\sigma}^D(u^n) + v_K^n F_{K,\sigma}^C(u^n)) \\ + \sum_{K\in\mathcal{M}} m(K)v_K^n f(u_K^n) = \sum_{K\in\mathcal{M}} m(K)v_K^n q_K^n.$$

Using (3.7), we obtain that

$$\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} v_\sigma^n(F_{K,\sigma}^D(u^n) + F_{K,\sigma}^C(u^n)) = 0 \text{ for all } v^n \in X_{\mathcal{D},0}, \tag{3.8}$$

which yields the following discrete weak formulation :

Let u_K^0 be defined by :

$$u_K^0 = \frac{1}{m(K)}\int_K u_0(\mathbf{x})\, d\mathbf{x} \quad \textit{for all } K \in \mathcal{M} \tag{3.9}$$

For each $n \in \{1,\dots,N\}$ find $u^n \in X_{\mathcal{D},0}$ such that for all $v^n \in X_{\mathcal{D},0}$:

$$\sum_{K\in\mathcal{M}} m(K)v_K^n \frac{\beta(u_K^n)-\beta(u_K^{n-1})}{\delta t} + < v^n, u^n >_D + < v^n, u^n >_C \\ + \sum_{K\in\mathcal{M}} m(K)v_K^n f(u_K^n) = \sum_{K\in\mathcal{M}} m(K)v_K^n q_K^n, \tag{3.10}$$

with

$$< v, u >_D = \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}(v_K - v_\sigma)F_{K,\sigma}^D(u^n) \tag{3.11}$$

and

$$< v, u >_C = \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}(v_K - v_\sigma)F_{K,\sigma}^C(u^n). \tag{3.12}$$

Remark 3.2. *Remark that the problem* (3.4)-(3.7) *is equivalent to (3.10)-(3.12). Indeed, let δ_{ij} be the Kroneker symbol, by setting $v_\sigma^n = 0$ for all $\sigma \in \mathcal{E}$, and $v'_K = \delta_{KK'}$ for all $K' \in \mathcal{M}$ and for a given K one recover* (3.5)*, and setting $v_K = 0$ for all $K \in \mathcal{M}$ and $v_{\sigma'} = \delta_{\sigma\sigma'}$ for all $\sigma' \in \mathcal{E}$ yields* (3.7)*. The homogeneous Dirichlet boundary condition* (3.6) *follows from the fact that $u^n \in X_{\mathcal{D},0}$.*

3.4 Diffusion term

In order to complete the numerical scheme we still have to express the discrete fluxes $F_{K,\sigma}^D$ and $F_{K,\sigma}^C$ in terms of the discrete unknowns. We briefly recall below the definition of the hybrid finite volume scheme [58]. First we define the discrete gradient

$$\nabla_K u = \frac{1}{m(K)}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)(u_\sigma - u_K)\mathbf{n}_{K,\sigma} \quad \forall K \in \mathcal{M}, \quad \forall u \in X_{\mathcal{D}}. \tag{3.13}$$

Remark that the geometrical relation

$$\sum_{\sigma\in\mathcal{E}_K} m(\sigma)\mathbf{n}_{K,\sigma}(\mathbf{x}_\sigma - \mathbf{x}_K)^T = m(K)Id \tag{3.14}$$

holds for each K. Let $\psi(\mathbf{x})$ be a function, piecewise linear on the control volumes of the mesh ; then $\nabla_K P_{\mathcal{D}}(\psi) = \nabla\psi(\mathbf{x})|_{x\in K}$. Let $R_{K,\sigma}(u)$ denote

$$R_{K,\sigma}(u) = u_\sigma - u_K - \nabla_K u \cdot (\mathbf{x}_\sigma - \mathbf{x}_K), \tag{3.15}$$

which is the second order error term, vanishing for piecewise linear functions, let $\mathbb{B}_K = \left(\mathbb{B}_K^{\sigma,\sigma'}\right)_{\sigma,\sigma'\in\mathcal{E}_K}$ be a positive definite matrix. Then the numerical fluxes $\left(F_{K,\sigma}^D(u)\right)_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$ are uniquely defined through the following discrete integration by parts rule

$$\sum_{\sigma\in\mathcal{E}_K}(v_K - v_\sigma)F_{K,\sigma}^D(u) = \mathbf{\Lambda}_K\nabla_K v\cdot\nabla_K u + \sum_{\sigma,\sigma'\in\mathcal{E}_K}\mathbb{B}_K^{\sigma,\sigma'}R_{K,\sigma}(v)R_{K,\sigma'}(u) \text{ for all } v\in X_{\mathcal{D}}, \tag{3.16}$$

where $\mathbf{\Lambda}_K = \int_K \mathbf{\Lambda}(\mathbf{x})\ d\mathbf{x}$. The last term in (3.16) is a stabilization term which is introduced in order to insure the bilinear form $<\cdot,\cdot>_D$ is coercive. The choice of $\mathbb{B}_K$ is not arbitrary, but as it was pointed out in [58] and [101] it must fulfill the following condition :
There exist two positive constants $\underline{s}$ and $\overline{S}$ such that for all $u \in X_{\mathcal{D}}$

$$\underline{s}\sum_{\sigma\in\mathcal{E}_K}\frac{m(\sigma)}{d_{K,\sigma}}(R_{K,\sigma}(u))^2 \leq \sum_{\sigma,\sigma'\in\mathcal{E}_K}\mathbb{B}_K^{\sigma,\sigma'}R_{K,\sigma}(u)R_{K,\sigma'}(u) \leq \overline{S}\sum_{\sigma\in\mathcal{E}_K}\frac{m(\sigma)}{d_{K,\sigma}}(R_{K,\sigma}(u))^2.$$

Remark that in view of (3.16), (3.13) and (3.15) there exists a family of positive definite matrix $(\mathbb{D}_K)_{K\in\mathcal{M}}$ such that

$$\sum_{\sigma,\sigma'\in\mathcal{E}_K}(v_K - v_\sigma)F_{K,\sigma}^D(u) = \sum_{\sigma,\sigma'\in\mathcal{E}_K}\mathbb{D}_K^{\sigma,\sigma'}(v_K - v_\sigma)(u_K - u_{\sigma'}) \text{ for all } K\in\mathcal{M}. \tag{3.17}$$

If $\mathbb{B}_K$ is diagonal with $\mathbb{B}_K^{\sigma,\sigma} = b_{K,\sigma}\dfrac{m(\sigma)}{d_{K,\sigma}}$, then we obtain the SUSHI scheme [66]. It is worth noting that if we choose $b_{K,\sigma} = \sqrt{d}$ for all $K\in\mathcal{M}$ and $\sigma\in\mathcal{E}_K$ in the simple case that Λ is a scalar and that the mesh satisfies the orthogonality property $\mathbf{n}_{K,\sigma} = \dfrac{\mathbf{x}_\sigma - \mathbf{x}_K}{d_{K,\sigma}}$, the expression for $F_{K,\sigma}^D(u)$ simplifies (see. [66, Lemma 2.1]) ; moreover if convection is absent we obtain the usual two point scheme. Nevertheless, it may be useful to optimize the choice of $\alpha_{K,\sigma}$ as it is done in [11].

Lemma 3.1. *Let $\mathcal{D}$ be a discretization of Ω in sense of Definition 3.1, moreover let $\theta \geq \theta_{\mathcal{D}}$ be given. Then for all $\psi \in C^2(\overline{\Omega})$, there exists a positive constant C only depending on d, θ and ψ such that*

$$\|\nabla_K P_{\mathcal{D}}\psi - \nabla\psi\|_{(L^\infty(\Omega))^d} \leq Ch_{\mathcal{D}}.$$

The proof of this Lemma is given in [66].

3.5 Convection term

Our discretization of the convective term is based upon a convective flux which involves the hybrid unknowns u_σ. Following [101] for all $u\in X_{\mathcal{D}}$ we define

$$F_{K,\sigma}^C(u) = A(V_{K,\sigma})u_K + B(V_{K,\sigma})u_\sigma, \tag{3.18}$$

where A and B satisfy the assumptions :

$(\mathcal{H}_{C,1})$ A, B are Lipschitz functions such that $A(0) = B(0) = 0$;
$(\mathcal{H}_{C,2})$ $A(s) + B(s) = s$ for all $s \in \mathbb{R}$;
$(\mathcal{H}_{C,3})$ There exists $C_{AB} \geq 0$ such that $A(s) - B(s) \geq -C_{AB}|s|$ for all $s \in \mathbb{R}$.

Different choices of A and B lead to different schemes, for example setting $A(s) = \frac{1}{2}(s + |s|)$ and $B(s) = \frac{1}{2}(s - |s|)$ one obtains an upwind scheme i.e.

$$F^C_{K,\sigma}(u) = \begin{cases} V_{K,\sigma} u_K, & \text{if } V_{K,\sigma} \geq 0 \\ V_{K,\sigma} u_\sigma, & \text{if } V_{K,\sigma} < 0. \end{cases} \tag{3.19}$$

On the other hand setting $A(s) = 0$ and $B(s) = s$ leads to a sort of a centered scheme

$$F^C_{K,\sigma}(u) = V_{K,\sigma} u_\sigma. \tag{3.20}$$

In Section 8 we compare the schemes (3.19) and (3.20) numerically. It is shown that the centered scheme (3.20) is more accurate, while the upwind scheme (3.19) preserves the approximate solution from unphysical oscillations in the convection dominated case.

3.6 Basic properties of the scheme

The space $X_\mathcal{D}$ defined in (3.3) is equipped with the following semi-norm.

Definition 3.3. *Let $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ be a discretization of Ω ; then for all $v \in X_\mathcal{D}$ we define*

$$|v|^2_{X_\mathcal{D}} = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \frac{m(\sigma)}{d_{K,\sigma}} (v_\sigma - v_K)^2. \tag{3.21}$$

Moreover, let $N \in \mathbb{N}^*$ and $\delta t = T/N$, for all $v \in X_{\mathcal{D},\delta t}$ we define

$$|v|^2_{X_{\mathcal{D},\delta t}} = \sum_{n=1}^{N} \delta t |v|^2_{X_\mathcal{D}}. \tag{3.22}$$

The semi-norm $|\cdot|_{X_\mathcal{D}}$ (or $|\cdot|_{X_{\mathcal{D},\delta t}}$) is a norm on the space $X_{\mathcal{D},0}$ (on $X_{\mathcal{D},0,\delta t}$ respectively). Let us also define a discrete analog of the $\|\cdot\|_{1,p}$ norm.

Definition 3.4 (The discrete space $H_\mathcal{M}(\Omega)$)**.** *Let $1 \leq p < \infty$ and let $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ be a discretization of Ω. Let $H_\mathcal{M}(\Omega) \subset L^2(\Omega)$ be the set of piecewise constant functions on the control volumes of the mesh $\mathcal{M}$ for each $v \in H_\mathcal{M}(\Omega)$ we define $v_K = v(\mathbf{x})|_{\mathbf{x}\in K}$. For all $v \in H_\mathcal{M}(\Omega)$ and for all $\sigma \in \mathcal{E}_{int}$ with $\mathcal{M}_\sigma = \{K, L\}$ we define $D_\sigma v = |v_K - v_L|$ and $d_\sigma = d_{K,\sigma} + d_{L,\sigma}$, and for all $\sigma \in \mathcal{E}_{ext}$ with $\mathcal{M}_\sigma = \{K\}$, we set $D_\sigma v = |v_K|$ and $d_\sigma = d_{K,\sigma}$. We then define the following family of norms*

$$\|v\|^p_{1,p,\mathcal{M}} = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma) d_{K,\sigma} \left(\frac{D_\sigma v}{d_\sigma}\right)^p ; \tag{3.23}$$

so that in particular

$$\|v\|^2_{1,2,\mathcal{M}} = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma) d_{K,\sigma} \left(\frac{D_\sigma v}{d_\sigma}\right)^2 .$$

Definition 3.5. *Let $\mathcal{D}$ be a discretization of Ω , we define the projection operators $\Pi_{\mathcal{D}} : X_{\mathcal{D}} \to L^2(\Omega)$ and $\nabla_{\mathcal{D}} : X_{\mathcal{D}} \to (L^2(\Omega))^d$ by*

$$\Pi_{\mathcal{D}} v(\mathbf{x}) = v_K \text{ for all } (\mathbf{x}, t) \in K$$

and

$$\nabla_{\mathcal{D}} v(\mathbf{x}) = \nabla_K v \text{ for all } (\mathbf{x}, t) \in K$$

respectively. Let also $N \in \mathbb{N}^*$ and $\delta t = T/N > 0$; we define $\Pi_{\mathcal{D},\delta t} : X_{\mathcal{D},\delta t} \to L^2(Q_T)$ and $\nabla_{\mathcal{D},\delta t} : X_{\mathcal{D},\delta t} \to (L^2(Q_T))^d$ by

$$\Pi_{\mathcal{D},\delta t} v(\mathbf{x}, t) = v_K^n \text{ for all } (\mathbf{x}, t) \in K \times (t_{n-1}, t_n]$$

and

$$\nabla_{\mathcal{D},\delta t} v(\mathbf{x}, t) = \nabla_K v^n \text{ for all } (\mathbf{x}, t) \in K \times (t_{n-1}, t_n].$$

Next we recall two results from [66] which we will use below.

Lemma 3.2. *Let $\mathcal{D}$ be a discretization of Ω , and let $\theta \geq \theta_{\mathcal{D}}$ be given. Then there exists a positive constant C only depending on θ and d such that*

$$\|\nabla_{\mathcal{D}} v\|_{L^2(\Omega)} \leq C |v|_{X_{\mathcal{D}}} \text{ for all } v \in X_{\mathcal{D}}.$$

Lemma 3.3. *Let $\mathcal{D}$ be a discretization of Ω , then there holds*

$$\|v\|_{1,2,\mathcal{M}} \leq |v|_{X_{\mathcal{D}}} \quad \text{for all } v \in X_{\mathcal{D},0}.$$

We show below that the bilinear forms defined in (3.11) and (3.12) satisfy continuity and coercivity properties.

Lemma 3.4. *Let $\mathcal{D}$ be a discretization of Ω , and let $\theta \geq \theta_{\mathcal{D}}$ be given, then :*

(i) There exist positive constants C_1 and α which do not depend on h such that

$$| < u, v >_D | \leq C_1 |u|_{X_{\mathcal{D}}} |v|_{X_{\mathcal{D}}} \tag{3.24}$$

and

$$< u, u >_D \geq \alpha |u|^2_{X_{\mathcal{D}}} \tag{3.25}$$

for all $u, v \in X_{\mathcal{D}}$.

(ii) There exist two positive constant C_2 and C_3 which does not depend on h such that

$$| < u, v >_C | \leq C_2 |u|_{X_{\mathcal{D}}} |v|_{X_{\mathcal{D}}} \tag{3.26}$$

and

$$< u, u >_C \geq -C_3 h_{\mathcal{D}} |u|^2_{X_{\mathcal{D}}} \tag{3.27}$$

for all $u, v \in X_{\mathcal{D},0}$. Moreover $C_3 = 0$ if $C_{AB} = 0$, which is the case for the upwind scheme (3.19).

Proof : (i) We refer to [66, Lemma 4.4] for a proof.

(ii) The proof of the second statement of (ii) is given in Remark 3.4 of [101]. It remains to show (3.26). By the definition (3.12), (3.19) and in view of $(\mathcal{H}_{C,2})$ we have that

$$\begin{aligned} < u,v >_C &= \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}(v_K - v_\sigma)\left(A(V_{K,\sigma})u_K + B(V_{K,\sigma})u_\sigma\right) \\ &= \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} V_{K,\sigma}(v_K - v_\sigma)u_K - \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} B(V_{K,\sigma})(v_K - v_\sigma)(u_K - u_\sigma) \end{aligned}$$

which implies using $(\mathcal{H}_{C,1})$

$$< u,v >_C \leq \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} V_{K,\sigma}(v_K - v_\sigma)u_K + C\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} |V_{K,\sigma}|(v_K - v_\sigma)(u_K - u_\sigma).$$

Using the Cauchy-Schwarz inequality and the bound $d_{K,\sigma} \leq h_{\mathcal{D}}$ we have that

$$\begin{aligned} | < u,v >_C | &\leq \sqrt{d}\|\mathbf{V}\|_{L^\infty(\Omega)}\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)\frac{(v_K - v_\sigma)^2}{d_{K,\sigma}}\right)^{\frac{1}{2}}\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\frac{m(\sigma)d_{K,\sigma}}{d}u_K^2\right)^{\frac{1}{2}} \\ &+h_{\mathcal{D}}C\|\mathbf{V}\|_{L^\infty(\Omega)}\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)\frac{(v_K - v_\sigma)^2}{d_{K,\sigma}}\right)^{\frac{1}{2}}\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)\frac{(u_K - u_\sigma)^2}{d_{K,\sigma}}\right)^{\frac{1}{2}}. \end{aligned}$$

and

$$| < u,v >_C | \leq \|\mathbf{V}\|_{L^\infty(\Omega)}\left(\sqrt{d}\cdot|v|_{X_{\mathcal{D}}}\|\Pi_{\mathcal{D}}u\|_{L^2(\Omega)} + h_{\mathcal{D}}C\cdot|v|_{X_{\mathcal{D}}}|u|_{X_{\mathcal{D}}}\right)$$

since $\sum_{\sigma\in\mathcal{E}_K} d_{K,\sigma}m(\sigma) = m(K)d$. In view of Lemma 3.3 and the discrete Poincaré inequality implied by Lemma 6.3 below we conclude that

$$| < u,v >_C | \leq C_2|u|_{X_{\mathcal{D}}}|v|_{X_{\mathcal{D}}}.$$

□

Next we recall a technical lemma presented in [72], namely Lemma 8.2, which will be useful for the a priori estimates of the next section

Lemma 3.5. *Let $B(s)$, $s\in\mathbb{R}$ be defined by*

$$B(s) = \beta(s)s - \int_0^s \beta(\tau)d\tau,$$

with β satisfying hypothesis $(\mathcal{H}_1)$. Then $B(s) \geq \frac{1}{2}s^2\underline{\beta}$.

4 A priori estimates

We define below an approximate solution of Problem $(\mathcal{P}b)$.

Definition 4.1 (Approximate solution). *Let $\mathcal{D}$ be a discretization of Ω, $N \in \mathbb{N}*$ and $\delta t = T/N > 0$. We say that the $u_{\mathcal{D},\delta t} = (u^n)_{n\in\{1,\ldots,N\}} \in X^N_{\mathcal{D},\delta t,0}$, is an approximate solution of Problem $(\mathcal{P}b)$ if for all $n \in \{1,\ldots,N\}$, u^n satisfies (3.9)-(3.12). We also denote $u_{\mathcal{D},\delta t}$ the function defined by*

$$u_{\mathcal{D},\delta t}(\mathbf{x},0) = u^0_K \textit{ for all } \mathbf{x} \in K,$$
$$u_{\mathcal{D},\delta t}(\mathbf{x},t) = u^n_K \textit{ for all } (\mathbf{x},t) \in K \times (t_{n-1},t_n]$$

and

$$u_{\mathcal{D},\delta t}(\mathbf{x},t) = u^n_\sigma \textit{ for all } (\mathbf{x},t) \in \sigma \times (t_{n-1},t_n].$$

Lemma 4.1 (A priori estimate). *Let $u_{\mathcal{D},\delta t}$ be an approximate solution of Problem $(\mathcal{P}b)$. There exists $C \in \mathbb{R}^+$ and also $C_{h_\mathcal{D}} \in \overline{\mathbb{R}^+}$ only depending on α and C_3 (defined in Lemma 3.4), such that if $h_\mathcal{D} \le C_{h_\mathcal{D}}$, then $u_{\mathcal{D},\delta t}$ satisfies*

$$\|u_{\mathcal{D},\delta t}\|_{L^\infty(0,T;L^2(\Omega))} + |u_{\mathcal{D},\delta t}|_{X_{\mathcal{D},\delta t}} \le C; \tag{4.1}$$

moreover

$$\|\beta(u_{\mathcal{D},\delta t})\|_{L^\infty(0,T;L^2(\Omega))} \le C. \tag{4.2}$$

Proof : Let $m \in [1,N]$ be an arbitrary integer. Summing on $n \in \{1,\ldots,m\}$ the equation (3.10) with $v^n = u^n$ for each n we obtain

$$\sum_{K\in\mathcal{M}} m(K)\sum_{n=1}^m u^n_K(\beta(u^n_K) - \beta(u^{n-1}_K)) + \sum_{n=1}^m \delta t(< u^n, u^n >_D + < u^n, u^n >_C)$$
$$+\sum_{n=1}^m \delta t \sum_{K\in\mathcal{M}} m(K)u^n_K f(u^n_K) = \sum_{n=1}^m\sum_{K\in\mathcal{M}} \delta t\; m(K)u^n_K q^n_K.$$

Next, we consider the function B from Lemma 3.5 defined by

$$B(u) = \beta(u)u - \int_0^u \beta(\tau)d\tau.$$

One can see that the following relation holds

$$B(u^n_K) - B(u^{n-1}_K) = u^n_K(\beta(u^n_K) - \beta(u^{n-1}_K)) - \int_{u^{n-1}_K}^{u^n_K}(\beta(\tau) - \beta(u^{n-1}_K))d\tau$$

and since β is nondecreasing we have that

$$\int_{u^{n-1}_K}^{u^n_K}(\beta(\tau) - \beta(u^{n-1}_K))d\tau \ge 0,$$

which implies

$$\begin{aligned}\sum_{K\in\mathcal{M}} m(K)(B(u^m_K) - B(u^0_K)) &= \sum_{K\in\mathcal{M}} m(K)\sum_{n=1}^m (B(u^n_K) - B(u^{n-1}_K))\\ &\le \sum_{K\in\mathcal{M}} m(K)\sum_{n=1}^m u^n_K(\beta(u^n_K) - \beta(u^{n-1}_K)).\end{aligned}$$

In view of Lemma 3.5 we have that

$$\frac{1}{2}\underline{\beta}u^2 \leq B(u) \leq u\beta(u) \leq \frac{(\beta(u))^2}{\underline{\beta}},$$

which yields

$$\frac{1}{2}\underline{\beta}\|u_{\mathcal{D},\delta t}(\cdot,t_m)\|^2_{L^2(\Omega)} - \frac{1}{\underline{\beta}}\|\beta(u_0)\|^2_{L^2(\Omega)} \leq \sum_{K\in\mathcal{M}} m(K)\sum_{n=1}^{m} u_K^n(\beta(u_K^n) - \beta(u_K^{n-1})).$$

We remark that in view of the hypothesis $(\mathcal{H}_4)$ one has

$$uf(u) \geq M \min_{0\leq u\leq M} f(u),$$

since $\min_{0\leq u\leq M} f(u) \leq 0$. In view of Lemma 3.4, the equation (3.11) and using $(\mathcal{H}_2)$ we finally conclude that

$$\begin{aligned}&\frac{1}{2}\underline{\beta}\|u_{\mathcal{D},\delta t}(\cdot,t_m)\|^2_{L^2(\Omega)} + \alpha|u_{\mathcal{D},\delta t}|^2_{X_{\mathcal{D},\delta t}}\\ &\leq C + C_3 h_{\mathcal{D}}|u_{\mathcal{D},\delta t}|^2_{X_{\mathcal{D},\delta t}} + \sum_{n=1}^{N}\sum_{K\in\mathcal{M}} \delta t m(K)u_K^n q_K^n,\end{aligned} \tag{4.3}$$

where α and C_3 are the constants form Lemma 3.4. Applying Cauchy-Schwarz and Young's inequality to the last term in (4.3) leads to

$$\begin{aligned}\frac{1}{2}\underline{\beta}\|u_{\mathcal{D},\delta t}(\cdot,t_m)\|^2_{L^2(\Omega)} &+ \alpha|u_{\mathcal{D},\delta t}|^2_{X_{\mathcal{D},\delta t}}\\ &\leq C + C_3 h_{\mathcal{D}}|u_{\mathcal{D},\delta t}|^2_{X_{\mathcal{D},\delta t}} + \|u_{\mathcal{D},\delta t}\|_{L^2(Q_T)}\|q\|_{L^2(Q_T)}\\ &\leq C + C_3 h_{\mathcal{D}}|u_{\mathcal{D},\delta t}|^2_{X_{\mathcal{D},\delta t}} + \frac{\varepsilon}{2}\|u_{\mathcal{D},\delta t}\|^2_{L^2(Q_T)} + \frac{1}{2\varepsilon}\|q\|^2_{L^2(Q_T)}.\end{aligned}$$

Then, in view of $(\mathcal{H}_3)$ we obtain

$$\begin{aligned}\frac{1}{2}\underline{\beta}\|u_{\mathcal{D},\delta t}\|^2_{L^\infty(0,T;L^2(\Omega))} &\leq C + C_3 h_{\mathcal{D}}|u_{\mathcal{D},\delta t}|^2_{X_{\mathcal{D},\delta t}}\\ &+ \frac{\varepsilon}{2}T\|u_{\mathcal{D},\delta t}\|^2_{L^\infty(0,T;L^2(\Omega))} + \frac{1}{2\varepsilon}\|q\|^2_{L^2(Q_T)}\end{aligned}$$

and

$$\begin{aligned}\alpha|u_{\mathcal{D},\delta t}|^2_{X_{\mathcal{D},\delta t}} &\leq C + C_3 h_{\mathcal{D}}|u_{\mathcal{D},\delta t}|^2_{X_{\mathcal{D},\delta t}}\\ &+ \frac{\varepsilon}{2}T\|u_{\mathcal{D},\delta t}\|^2_{L^\infty(0,T;L^2(\Omega))} + \frac{1}{2\varepsilon}\|q\|^2_{L^2(Q_T)}.\end{aligned}$$

We now choose $\varepsilon = \underline{\beta}/(2T)$, which gives

$$\frac{1}{4}\underline{\beta}\|u_{\mathcal{D},\delta t}\|^2_{L^\infty(0,T;L^2(\Omega))} \leq C + C_3 h_{\mathcal{D}}|u_{\mathcal{D},\delta t}|^2_{X_{\mathcal{D},\delta t}}$$

and

$$\frac{\alpha}{2}|u_{\mathcal{D},\delta t}|^2_{X_{\mathcal{D},\delta t}} \leq C + C_3 h_{\mathcal{D}}|u_{\mathcal{D},\delta t}|^2_{X_{\mathcal{D},\delta t}}.$$

If

$$h_{\mathcal{D}} \leq C_{h_{\mathcal{D}}} = \frac{\alpha}{4C_3}, \tag{4.4}$$

then (4.1) follows. The assumption $(\mathcal{H}_1)$ implies that there exists a positive C, such that

$$|\beta(s)| \leq C(1+|s|) \text{ for all } s\in\mathbb{R}, \tag{4.5}$$

which in turn implies the estimate (4.2). □

Remark 4.1. *We assume from now on that the size of the mesh $h_{\mathcal{D}}$ is always chosen in a suitable way (by (4.4) for example), so that the a priori estimates (4.1) and (4.2) hold.*

5 Existence and uniqueness of the discrete solution

Remark 5.1 (Extended discrete problem)**.** *Let $\nu > 0$ and $w \in H_{\mathcal{M}}(\Omega)$ (cf. Definition 3.4), we consider the following extended one step problem. Find $u^{nu} \in X_{\mathcal{D},0}$ such that for all $v \in X_{\mathcal{D},0}$:*

$$\begin{aligned}\nu \sum_{K\in\mathcal{M}} m(K) v_K \frac{\beta(u_K^\nu) - \beta(w_K)}{\delta t} + < v, u^\nu >_D + \nu < v, u^\nu >_C \\ +\nu \sum_{K\in\mathcal{M}} m(K) v_K f(u_K^\nu) = \nu \sum_{K\in\mathcal{M}} m(K) v_K q_K.\end{aligned} \tag{5.1}$$

It can be shown that, under an appropriate condition on $h_{\mathcal{D}}$ (see Remark 4.1), the solution of the extended problem (5.1) satisfies

$$\delta t |u^\nu|_{X_{\mathcal{D}}}^2 \le \nu C. \tag{5.2}$$

Theorem 5.1 (Existence of a discrete solution)**.** *The problem (3.9)-(3.12) possesses at least one solution.*

Proof : Let us consider the extended problem (5.1), which can be written as the abstract system of nonlinear equations

$$H(u^\nu, u^{n-1}, \nu) = 0, \tag{5.3}$$

where H is a continuous mapping from $H_{\mathcal{M}(\Omega)} \times X_{\mathcal{D},0} \times [0,1]$ to $X_{\mathcal{D},0}$. It is worth noting that in view of Remark 3.2 (see also (3.5) and (3.7)), we may write the system (5.3) in the form

$$H_K\left(\beta(u_K^\nu), \beta(w_K), u_K^\nu, (u_\sigma^\nu)_{\sigma\in\mathcal{E}_K}, \nu\right) = 0 \text{ for all } K \in \mathcal{M}$$

and

$$H_\sigma\left((u_K^\nu)_{K\in\mathcal{M}_\sigma}, \left((u_\sigma^\nu)_{\sigma\in\mathcal{E}_K}\right)_{K\in\mathcal{M}_\sigma}, \nu\right) = 0 \text{ for all } \sigma \in \mathcal{E}_{int},$$

where $u^\nu \in X_{\mathcal{D},0}$. In view of Remark 5.1 and the estimate (4.2), there exists $C > 0$ which does not depend on ν such that

$$\delta t |u^\nu|_{X_{\mathcal{D}}}^2 \le C$$

for any u^ν satisfying $H(u^\nu, u^{n-1}, \nu)$. Setting $R = \sqrt{C/\delta t + 1}$ we deduce that

$$|u^\nu|_{X_{\mathcal{D}}} < R \text{ for all } (\nu, u^\nu) \in [0,1] \times X_{\mathcal{D},0} \text{ such that } H(u^\nu, u^{n-1}, \nu) = 0.$$

Therefore the system (5.3) has no solutions on the boundary of the ball B_R of radius R for $\nu \in [0,1]$. Next, we denote by $d(H(\cdot, u^{n-1}, \nu) = 0, B_R, 0)$ the topological degree of the application $H(\cdot, u^{n-1}, \nu)$ with respect to the ball B_R and right-hand side 0. For

$\nu = 0$ the system $H(u^\nu, u^{n-1}, \nu) = 0$ reduces to a linear system with a positive definite matrix. Thus, in view of the homotopy invariance of the topological degree we have that

$$d(H(\cdot, u^{n-1}, \nu), B_R, 0) = d(H(\cdot, u^{n-1}, 0), B_R, 0) = 1 \text{ for all } \nu \in [0,1],$$

where we have applied [55, Theorem 3.1 (d1) and (d3)]. Thus, by [55, Theorem 3.1 (d4)]. Then, there exists u^n such that $H(u^n, u^{n-1}, 1) = 0$, so that u^n is a solution of (3.9)-(3.12). □

Theorem 5.2 (Uniqueness of the discrete solution)**.** *Let $\mathcal{D}$ be a discretization of Ω, then if the time step satisfies*

$$\delta t < \underline{\beta}/\underline{f} \tag{5.4}$$

and the mesh size $h_{\mathcal{D}}$ is small enough, then the solution of the problem (3.9)-(3.12) is unique.

Proof : We give a proof by contradiction. Let $u_{\mathcal{D},\delta t}$ and $\widetilde{u}_{h,k}$ be two different solutions of (3.9)-(3.12), such that $u^m = \widetilde{u}^m$ for all $m = 1, \ldots, n-1$, but $u^n \neq \widetilde{u}^n$. We define $r^n = u^n - \widetilde{u}^n$. In view of (3.10) with $v = r^n$ we have that

$$\sum_{K \in \mathcal{M}} m(K) r_K^n \frac{\beta(u_K^n) - \beta(\widetilde{u}_K^n)}{\delta t} + < r^n, r^n >_D + < r^n, r^n >_C$$

$$+ \sum_{K \in \mathcal{M}} m(K) r_K^n (f(u_K^n) - f(\widetilde{u}_K^n)) = 0.$$

We apply Lemma 3.4 as well as the assumptions $(\mathcal{H}_1)$ and $(\mathcal{H}_4)$ in order to estimate each term in the equation above. We obtain

$$(\underline{\beta}/\delta t - \underline{f}) \sum_{K \in \mathcal{M}} m(K) (r_K^n)^2 + \alpha |r^n|^2_{X_{\mathcal{D}}} \leq C_3 h_{\mathcal{D}} |r^n|^2_{X_{\mathcal{D}}},$$

Finally, in view of (5.4) we deduce that

$$|r^n|_{X_{\mathcal{D}}} = 0$$

for $h_{\mathcal{D}}$ sufficiently small. □

6 Estimates on space and time translates

6.1 Estimates on time translates

To begin with we state without proof two technical lemmas which will be useful for proving the estimate on time translates.

Lemma 6.1. *Let $T > 0$, $\tau \in (0, T)$, $N \in \mathbb{N}^*$, $\delta t = T/N$ be given and $(a^n)_{n \in \mathbb{N}^*}$ be a family of non negative real values. Let $\lceil s \rceil$ denotes the smallest integer larger or equal to s. Then*

$$\int_0^{T-\tau} \sum_{\lceil t/\delta t \rceil + 1 \leq n \leq \lceil (t+\tau)/\delta t \rceil} a^n dt \leq \tau \sum_{n=1}^{N} a^n.$$

Lemma 6.2. *Let $T>0$, $\tau\in(0,T)$, $N\in\mathbb{N}^*$, $\delta t=T/N$, $\zeta\in[0,\tau]$ be given and $(a^n)_{n\in\mathbb{N}^*}$ be a family of nonnegative real values. Let $\lceil s\rceil$ denotes the smallest integer larger or equal to s. Then*

$$\int_0^{T-\tau}\sum_{\lceil t/\delta t\rceil+1\le n\le\lceil(t+\tau)/\delta t\rceil} a^{\lceil(t+\zeta)/\delta t\rceil}dt\le\tau\sum_{n=1}^{N}a^n.$$

Theorem 6.1. *Let $\mathcal{D}$ be a discretization of Ω and let $\{u_{\mathcal{D},\delta t}\}$ be a solution of the discrete problem in sense of Definition 4.1. Let also $\theta\ge\theta_{\mathcal{D}}$ be given. Then there exists a positive constant C only depending on θ such that*

$$\int_0^{T-\tau}\int_\Omega(u_{\mathcal{D},\delta t}(x,t+\tau)-u_{\mathcal{D},\delta t}(x,t))^2dxdt\le C\tau,\tag{6.1}$$

for all $\tau\in(0,T)$.

Proof : To begin with we use the hypothesis $(\mathcal{H}_1)$ to obtain

$$\begin{aligned}
&\underline{\beta}\int_0^{T-\tau}\int_\Omega(u_{\mathcal{D},\delta t}(x,t+\tau)-u_{\mathcal{D},\delta t}(x,t))^2\ d\mathbf{x}dt\\
&=\underline{\beta}\int_0^{T-\tau}\sum_{K\in\mathcal{M}}m(K)(u_K^{\lceil(t+\tau)/\delta t\rceil}-u_K^{\lceil t/\delta t\rceil})^2dt\\
&\le\int_0^{T-\tau}\sum_{K\in\mathcal{M}}m(K)(u_K^{\lceil(t+\tau)/\delta t\rceil}-u_K^{\lceil t/\delta t\rceil})(\beta(u_K^{\lceil(t+\tau)/\delta t\rceil})-\beta(u_K^{\lceil t/\delta t\rceil}))dt\\
&=\int_0^{T-\tau}\sum_{K\in\mathcal{M}}m(K)(u_K^{\lceil(t+\tau)/\delta t\rceil}-u_K^{\lceil t/\delta t\rceil})\sum_{\lceil t/\delta t\rceil+1\le n\le\lceil(t+\tau)/\delta t\rceil}m(K)(\beta(u_K^n)-\beta(u_K^{n-1}))dt.
\end{aligned}$$

For a given δt and for all real t and τ we define the following set

$$n(t,\tau)=\{n\in\mathbb{N},\ \lceil t/\delta t\rceil+1\le n\le\lceil(t+\tau)/\delta t\rceil\},$$

which can be empty. Then, the discrete equation (3.5) implies

$$\begin{aligned}
\underline{\beta}\int_0^{T-\tau}\int_\Omega(u_{\mathcal{D},\delta t}(x,t+\tau)-u_{\mathcal{D},\delta t}(x,t))^2\ d\mathbf{x}dt\le\int_0^{T-\tau}\sum_{K\in\mathcal{M}}(u_K^{\lceil(t+\tau)/\delta t\rceil}-u_K^{\lceil t/\delta t\rceil})\\
\cdot\sum_{n\in n(t,\tau)}\delta t\left(m(K)q_K^n-\sum_{\sigma\in\mathcal{E}_K}\left(F_{K,\sigma}^D(u^n)+F_{K,\sigma}^C(u^n)\right)-m(K)f(u_K^n)\right)dt.
\end{aligned}$$

Let us define the expressions $A_{D,C}$, A_R and A_S by

$$\begin{aligned}
A_{D,C}&=\int_0^{T-\tau}\sum_{n\in n(t,\tau)}\delta t\sum_{K\in\mathcal{M}}(u_K^{\lceil(t+\tau)/\delta t\rceil}-u_K^{\lceil t/\delta t\rceil})\sum_{\sigma\in\mathcal{E}_K}\left(F_{K,\sigma}^D(u^n)+F_{K,\sigma}^C(u^n)\right)dt,\\
A_R&=\int_0^{T-\tau}\sum_{n\in n(t,\tau)}\delta t\sum_{K\in\mathcal{M}}m(K)(u_K^{\lceil(t+\tau)/\delta t\rceil}-u_K^{\lceil t/\delta t\rceil})f(u_K^n)dt,\\
A_S&=\int_0^{T-\tau}\sum_{n\in n(t,\tau)}\delta t\sum_{K\in\mathcal{M}}m(K)(u_K^{\lceil(t+\tau)/\delta t\rceil}-u_K^{\lceil t/\delta t\rceil})q_K^ndt,
\end{aligned}$$

which we will estimate below. In view of (3.8), (3.11) and (3.12) we obtain

$$A_{D,C} = \int_0^{T-\tau} \sum_{n\in n(t,\tau)} \delta t(< u^{\lceil (t+\tau)/\delta t\rceil} - u^{\lceil t/\delta t\rceil}, u^n >_F + < u^{\lceil (t+\tau)/\delta t\rceil} - u^{\lceil t/\delta t\rceil}, u^n >_T)dt$$

In view of Lemma 3.4 we have that $| < u, v >_F | + | < u, v >_T | \le C|u|_{X_\mathcal{D}}|v|_{X_\mathcal{D}}$ for all $u, v \in X_{\mathcal{D},0}$ and since $2ab \le a^2 + b^2$ one has

$$|A_{D,C}| \le C \int_0^{T-\tau} \sum_{n\in n(t,\tau)} \delta t(|u^{\lceil (t+\tau)/\delta t\rceil}|_{X_\mathcal{D}} + |u^{\lceil t/\delta t\rceil}|_{X_\mathcal{D}})|u^n|_{X_\mathcal{D}} dt$$

$$\le C(\int_0^{T-\tau} \sum_{n\in n(t,\tau)} \delta t|u^{\lceil t/\delta t\rceil}|^2_{X_\mathcal{D}} + \int_0^{T-\tau} \sum_{n\in n(t,\tau)} \delta t|u^{\lceil (t+\tau)/\delta t\rceil}|^2_{X_\mathcal{D}} + \int_0^{T-\tau} \sum_{n\in n(t,\tau)} \delta t|u^n|^2_{X_\mathcal{D}} dt).$$

It follows from the estimate (4.1) and the Lemmas 6.1 and 6.2 that

$$|A_{D,C}| \le \tau C \sum_{n=1}^{N} \delta t|u^n|^2_{X_\mathcal{D}} \le C\tau.$$

Next, we estimate the term A_R; remark that the inequality (4.5) also holds for the function f instead of β. Hence,

$$\sum_{K\in\mathcal{M}} m(K)v_K f(u_K) \le \frac{1}{2}\|v\|^2_{L^2(\Omega)} + C\left(m(\Omega) + \|u\|^2_{L^2(\Omega)}\right) \tag{6.2}$$

for all $v, u \in H_\mathcal{M}$, which in turn implies that

$$\begin{aligned} |A_R| \;&\le \frac{1}{2}\int_0^{T-\tau} \sum_{n\in n(t,\tau)} \delta t\left(\|u_{\mathcal{D},\delta t}(\cdot, \lceil t/\delta t\rceil)\|^2_{L^2(\Omega)} + \|u_{\mathcal{D},\delta t}(\cdot, \lceil (t+\tau)/\delta t\rceil)\|^2_{L^2(\Omega)}\right) dt \\ &+C\int_0^{T-\tau} \sum_{n\in n(t,\tau)} \delta t\left(m(\Omega) + \|u_{\mathcal{D},\delta t}(\cdot, t_n)\|^2_{L^2(\Omega)}\right) dt. \end{aligned}$$

One more time it follows from the estimate (4.1) and the Lemmas 6.1 and 6.2 that

$$|A_R| \le \tau C \sum_{n=1}^{N} \delta t\left(m(\Omega) + \|u_{\mathcal{D},\delta t}(\cdot, t_n)\|^2_{L^2(\Omega)}\right) \le \tau C.$$

In the same way we proceed for the term $|A_S|$, one has that

$$|A_S| \le \tau\left(\|u_{\mathcal{D},\delta t}\|^2_{L^2(Q_T)} + \|q\|^2_{L^2(Q_T)}\right).$$

Finally we use an a priori estimate (4.1) and the hypothesis $(\mathcal{H}_5)$ to complete the proof.
□

6.2 Estimates on space translates

In this section we prove an estimate on the L^2-norm of differences of space translates of the discrete solution. We state without proof two results from [66], which are useful in our study.

Lemma 6.3. *Let $d \geq 1$, $1 \leq p < \infty$ and Ω be an open bounded connected subset of $\mathbb{R}^d$. Let $\mathcal{D}$ be a mesh of Ω . Let $\eta > 0$ be such that $\eta \leq d_{K,\sigma}/d_{L,\sigma} \leq 1/\eta$ for all $\sigma \in \mathcal{M}_\sigma = \{K, L\}$. Then, there exists $q > p$ only depending on p and there exists a positive constant C, only depending on d, Ω, p and η such that :*

$$\|u\|_{L^q(\Omega)} \leq C\|u\|_{1,p,\mathcal{M}} \tag{6.3}$$

for all $u \in H_{\mathcal{M}}(\Omega)$. We recall that $H_{\mathcal{M}}(\Omega) \subset L^2(\Omega)$ is the set of piecewise constant functions on the control volumes of the mesh.

Lemma 6.4. *Let $d \geq 1$ and Ω be a polyhedral open bounded connected subset of $\mathbb{R}^d$. Let $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ be a discretization of Ω and let $u \in H_{\mathcal{M}}(\Omega)$. Then, with notation of Definition 3.4 :*

$$\|u(\cdot + \mathbf{y}) - u\|_{L^1(\mathbb{R}^d)} \leq |\mathbf{y}|\sqrt{d}\|u\|_{1,1,\mathcal{M}}, \tag{6.4}$$

where u is defined on the whole $\mathbb{R}^d$, taking $u = 0$ outside Ω.

Next we show that a similar inequality holds in every L^p-norm.

Lemma 6.5. *Let $d \geq 1$, $1 \leq p < \infty$ and Ω be an open bounded connected subset of $\mathbb{R}^d$ and $T > 0$. Let $\mathcal{D}$ be a discretization of Ω . Let $\eta > 0$ such that $\eta \leq d_{K,\sigma}/d_{L,\sigma} \leq 1/\eta$ for all $\sigma \in \mathcal{M}_\sigma = \{K, L\}$. There exist $C > 0$ and $\rho > 0$, only depending on d, p, Ω and η such that*

$$\|u(\cdot + \mathbf{y}) - u\|_{L^p(\mathbb{R}^d)} \leq C|\mathbf{y}|^\rho\|u\|_{1,p,\mathcal{M}},$$

where u is defined on $\mathbb{R}^d$, taking $u = 0$ outside Ω.

Proof : In view of Lemma 6.3, there exist $q > p$ and a positive constant C such that

$$\|u\|_{L^q(\mathbb{R}^d)} \leq C\|u\|_{1,p,\mathcal{M}}. \tag{6.5}$$

We apply the Interpolation Inequality [4, Theorem 2.11, p.27]

$$\|u(\cdot + \mathbf{y}) - u\|_{L^p(\mathbb{R}^d)} \leq \|u(\cdot + \mathbf{y}) - u\|^\rho_{L^1(\mathbb{R}^d)}\|u(\cdot + \mathbf{y}) - u\|^{1-\rho}_{L^q(\mathbb{R}^d)}, \tag{6.6}$$

where

$$\rho = \frac{1}{p} \cdot \frac{q-p}{q-1}.$$

Moreover (6.5) implies that

$$\|u(\cdot + \mathbf{y}) - u\|_{L^q(\mathbb{R}^d)} \leq 2\|u\|_{L^q(\mathbb{R}^d)} \leq C\|u\|_{1,p,\mathcal{M}},$$

so that by (6.4) and (6.6) implies that

$$\|u(\cdot + \mathbf{y}) - u\|_{L^p(\mathbb{R}^d)} \leq C|\mathbf{y}|^\rho(\|u\|_{1,1,\mathcal{M}})^\rho(\|u\|_{1,p,\mathcal{M}})^{1-\rho}.$$

Applying Hölder inequality we obtain that

$$\|u\|_{1,1,\mathcal{M}} \leq C\|u\|_{1,p,\mathcal{M}}$$

for some positive constant C. Then

$$\|u(\cdot + \mathbf{y}) - u\|_{L^p(\mathbb{R}^d)} \leq C|\mathbf{y}|^\rho\|u\|_{1,p,\mathcal{M}}.$$

□

Theorem 6.2. *Let $\mathfrak{D}$ be a sequence of discretizations of Ω and such that there exists a positive constant θ satisfying $\theta_{\mathcal{D}} \leq \theta$ for all $\mathcal{D} \in \mathfrak{D}$. Let $\delta\mathfrak{t}$ be a sequence of real positif numbers, such that $T/\delta t \in \mathbb{N}$ for all $\delta t \in \delta\mathfrak{t}$. Let $u_{\mathfrak{D},\delta\mathfrak{t}} = (u_{\mathcal{D},\delta t})_{\mathcal{D}\in\mathfrak{D},\delta t\in\delta\mathfrak{t}}$ be the sequence of approximate solutions corresponding to $\mathfrak{D}$ and $\delta\mathfrak{t}$. Then $u_{\mathfrak{D},\delta\mathfrak{t}}$ is relatively compact in $L^2(Q_T)$.*

Proof : To begin with, we extend $u_{\mathcal{D},\delta t}$ by zero outside of Q_T. Applying the Lemma 6.5 with $p = 2$ yields

$$\|u_{\mathcal{D},\delta t}(\cdot + \mathbf{y}, t) - u_{\mathcal{D},\delta t}(\cdot, t)\|_{L^2(\mathbb{R}^d)} \leq C|\mathbf{y}|^\rho \|u_{\mathcal{D},\delta t}(\cdot, t)\|_{1,2,\mathcal{M}}$$

for some positive constants $\rho > 0$ and $C > 0$. Integrating on $(0, T)$ we obtain

$$\|u_{\mathcal{D},\delta t}(\cdot + \mathbf{y}, \cdot) - u_{\mathcal{D},\delta t}\|^2_{L^2(\mathbb{R}^d\times(0,T))} \leq C|\mathbf{y}|^{2\rho} \sum_{n=1}^{N} \delta t \|u_{\mathcal{D},\delta t}(\cdot, t_n)\|^2_{1,2,\mathcal{M}}.$$

Then in view of Lemma 3.3 and the estimate (4.1) we obtain the bound

$$\|u_{\mathcal{D},\delta t}(\cdot + \mathbf{y}) - u_{\mathcal{D},\delta t}\|_{L^2(\mathbb{R}^d\times(0,T))} \leq C|\mathbf{y}|^\rho,$$

which, combined with (6.1) gives

$$\begin{aligned} &\|u_{\mathcal{D},\delta t}(\cdot + \mathbf{y}, \cdot + \tau) - u_{\mathcal{D},\delta t}\|_{L^2(\mathbb{R}^d\times(0,T))} \\ &\leq \|u_{\mathcal{D},\delta t}(\cdot + \mathbf{y}, \cdot + \tau) - u_{\mathcal{D},\delta t}(\cdot + \mathbf{y}, \cdot)\|_{L^2(\mathbb{R}^d\times(0,T))} + \|u_{\mathcal{D},\delta t}(\cdot + \mathbf{y}, \cdot) - u_{\mathcal{D},\delta t}\|_{L^2(\mathbb{R}^d\times(0,T))} \\ &\leq C(\sqrt{\tau} + |\mathbf{y}|^\rho). \end{aligned}$$

Then the Fréchet-Kolmogorov Compactness Theorem implies that the family $u_{\mathfrak{D},\delta\mathfrak{t}}$ is relatively compact in $L^2(\mathbb{R}^d \times (0, T))$ and thus in $L^2(Q_T)$. □

7 Convergence result

We begin with a technical lemma which is useful for passing to the limit in the nonlinear terms of (1.1).

Lemma 7.1. *Let (a, b) be an interval of $\mathbb{R}$ and let the function $g \in C(\mathbb{R})$ be Lipschitz continuous on $\mathbb{R} \setminus (a, b)$. Let $1 \leq p < \infty$ and let $\{u_m\}$ be a sequence in $L^2(Q_T)$ and $u \in L^2(Q_T)$, such that*

$$\lim_{m\to\infty} \|u_m - u\|_{L^2(Q_T)} = 0.$$

Then there exists a subsequence of $\{u_m\}$, which we still denote by $\{u_m\}$ such that

$$\lim_{m\to\infty} \|g(u_m) - g(u)\|_{L^2(Q_T)} = 0.$$

Proof : The proof of this Lemma can be found in Section 6.2.1 of [72], it relies on splitting of g into a sum g_1 and g_2, where g_1 is bounded and g_2 is Lipschitz on $\mathbb{R}$. □

Theorem 7.1. *Let $\mathfrak{D}$ be a sequence of discretizations of Ω, such that $h_{\mathcal{D}}$ tends to zero along $\mathfrak{D}$ and let θ be a positive constant such that $\theta_{\mathcal{D}} \le \theta$ for all $\mathcal{D} \in \mathfrak{D}$. Let $\delta\mathfrak{t}$ be a sequence of real positif numbers, such that $T/\delta t \in \mathbb{N}$ for all $\delta t \in \delta\mathfrak{t}$ and such that δt tends to zero along $\delta\mathfrak{t}$. Let $u_{\mathfrak{D},\delta\mathfrak{t}} = (u_{\mathcal{D},\delta t})_{\mathcal{D}\in\mathfrak{D},\delta t\in\delta\mathfrak{t}}$ be the sequence of approximate solutions corresponding to $\mathfrak{D}$ and $\delta\mathfrak{t}$. Then there exists a subsequence of $u_{\mathfrak{D},\delta\mathfrak{t}}$, which we denote again by $u_{\mathfrak{D},\delta\mathfrak{t}}$, such that $u_{\mathcal{D},\delta t} \to u$ strongly in $L^2(Q_T)$ as $h_{\mathcal{D}}, \delta t \to 0$, where u is a weak solution of Problem $(\mathcal{P}b)$. Moreover $u \in L^2(0,T;H^1_0(\Omega))$ and $\nabla_{\mathcal{D},\delta t}u_{\mathcal{D},\delta t}$ weakly converge in $L^2(Q_T)^d$ to ∇u. In the case that F is nondecreasing, the whole sequence $u_{\mathfrak{D},\delta\mathfrak{t}}$ converges to the unique weak solution u of Problem $(\mathcal{P}b)$.*

Proof : By Theorem 6.2 there exists a subsequence of $u_{\mathfrak{D},\delta\mathfrak{t}}$ that we still denote by $u_{\mathfrak{D},\delta\mathfrak{t}}$ and a function $u \in L^2(Q_T)$ such that $u_{\mathcal{D},\delta t} \to u$ strongly in $L^2(Q_T)$ as $h_{\mathcal{D}}, \delta t \to 0$ (and also in $L^2(\mathbb{R}^d \times (0,T))$ taking $u_{\mathcal{D},\delta t} = 0$ outside of Q_T). Let $\nabla_{\mathcal{D},\delta t}u_{\mathcal{D},\delta t}$ be prolonged by zero outside of Q_T, in view of (4.1) and Lemma 3.2 there exists a function $\mathbf{G} \in L^2(\mathbb{R} \times (0,T))^d$ such that $\nabla_{\mathcal{D},\delta t}u_{\mathcal{D},\delta t}$ weakly converge in $L^2(\mathbb{R} \times (0,T))^d$ to $\mathbf{G}$ along a subsequence as $h_{\mathcal{D}}, \delta t \to 0$. The proof of the fact that $\mathbf{G} = \nabla u$ is a straightforward adaptation of [66, Lemma 4.2] to the time dependent problem. Thus

$$\nabla_{\mathcal{D},\delta t}u_{\mathcal{D},\delta t} \to \nabla u \text{ weakly in } L^2(Q_T), \tag{7.1}$$

which in particular implies that $u \in L^2(0,T;H^1_0(\Omega))$ since $u = 0$ outside of Q_T.

Next we show that u is a weak solution of Problem $(\mathcal{P}b)$. For this purpose, we introduce the function space

$$\Psi = \{\psi \in C^{2,1}(\overline{\Omega} \times [0,T]), \quad \psi = 0 \text{ on } \partial\Omega \times [0,T], \quad \psi(\cdot,T) = 0\}.$$

Taking an arbitrary $\psi \in \Psi$, we define the sequence of elements of $X_{\mathcal{D},0}$

$$\psi^n = P_{\mathcal{D}}\psi(\cdot,t_n) \text{ for all } n \in \{1,\dots,N\}$$

so that $\psi^n_K = \psi(\mathbf{x}_K,t_n)$ and $\psi^n_\sigma = \psi(\mathbf{x}_\sigma,t_n)$. Next setting

$$v^n = \psi^{n-1} \text{ for all } n \in \{1,\dots,N\}$$

in (3.10), we obtain, also in view of (3.11) and (3.12), that

$$T_T + T_D + T_C + T_R = T_S,$$

where

$$T_T = \sum_{n=1}^{N} \sum_{K\in\mathcal{M}} m(K)(\beta(u^n_K) - \beta(u^{n-1}_K))\psi^{n-1}_K,$$

$$T_D = \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\psi^{n-1}_K - \psi^{n-1}_\sigma)F^D_{K,\sigma}(u^n),$$

$$T_C = \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\psi^{n-1}_K - \psi^{n-1}_\sigma)F^C_{K,\sigma}(u^n),$$

$$T_R = \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} m(K)\psi^{n-1}_K f(u^n_K)$$

and

$$T_S = \sum_{n=1}^{N} \delta t \sum_{K \in \mathcal{M}} m(K) \psi_K^{n-1} q_K^n.$$

We successively search for the limit of each of these terms as $h_\mathcal{D}$ and δt tend to zero.

Convection and diffusion terms. We refer to Theorem 3.7 of [101] (where a convergence result is shown for a stationary convection-diffusion problem) for a proof of the limits

$$T_C \to -\int_0^T \int_\Omega u(\mathbf{x},t)\mathbf{V}(\mathbf{x}) \cdot \nabla\psi(\mathbf{x},t)\ d\mathbf{x}dt \tag{7.2}$$

and

$$T_D \to \int_0^T \int_\Omega \nabla\psi(\mathbf{x},t) \cdot \mathbf{\Lambda}(\mathbf{x})\nabla u(\mathbf{x},t)\ d\mathbf{x}dt \tag{7.3}$$

as $h_\mathcal{D}, \delta t \to 0$.

Time evolution term. The proof of the fact that

$$T_T \to -\int_0^T \int_\Omega \beta(u)\psi_t\ d\mathbf{x}dt - \int_\Omega \beta(u_0)\psi(\cdot,0)\ d\mathbf{x}$$

along a subsequence of $u_{\mathfrak{D},\delta\mathfrak{t}}$ is classical. It is based upon the chain rule, the regularity of ψ, the strong convergence of $u_{\mathcal{D},\delta t}$ and Lemma 7.1 applied to the function β. The complete proof can be found in [72].

Reaction term. We deduce from the regularity of ψ and from Lemma 7.1 that the term T_R converges (up to a subsequence) to

$$\int_0^T \int_\Omega f(u(\mathbf{x},t))\psi(\mathbf{x},t)\ d\mathbf{x}dt$$

as $h_\mathcal{D}, \delta t \to 0$.

Source term. Similarly one can show that

$$T_S = \sum_{n=1}^{N} \sum_{K\in\mathcal{M}} \int_{t_{n-1}}^{t_n} \int_K \psi(\mathbf{x}_K, t_{n-1}) q(\mathbf{x},t)\ d\mathbf{x}dt \to \int_0^T \int_\Omega \psi(\mathbf{x},t) q(\mathbf{x},t)\ d\mathbf{x}dt$$

as $h_\mathcal{D}, \delta t \to 0$.

It follows from (4.2) that $\beta(u) \in L^\infty(0,T;L^2(\Omega))$. Moreover we deduce from the density of the set Φ in the set $\{\psi \in L^2(0,T;H_0^1(\Omega)), \psi_t \in L^\infty(Q_T), \psi(\cdot,T) = 0\}$ that u is a weak solution of the continuous problem $(\mathcal{P}b)$ in the sense of Definition 2.1. In the case that f is nondecreasing so that the solution of Problem $(\mathcal{P}b)$ is unique (cf. Remark 2.1) we conclude that the whole family $u_{\mathfrak{D},\delta\mathfrak{t}}$ converges to u as $h_\mathcal{D}, \delta t \to 0$. □

8 Numerical simulations

In this section we present the results of numerical simulations. The purpose is to test our scheme in the case of problems with a known analytical solution.

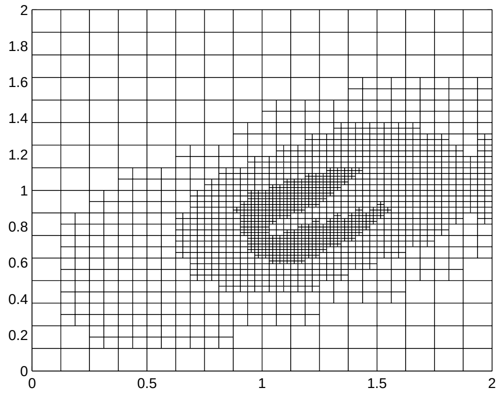

FIGURE 1 – The locally refined grid at $t = 1$

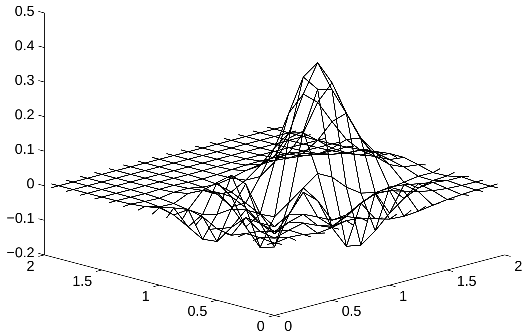

FIGURE 2 – The oscillating solution at time $t = 1$ using the centered scheme with $h_{\mathcal{D}} = 0.1768$

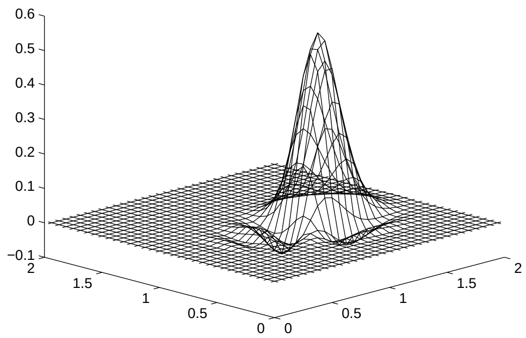

FIGURE 3 – The oscillating are reduced with the mesh refinement, $h_{\mathcal{D}} = 0.0884$

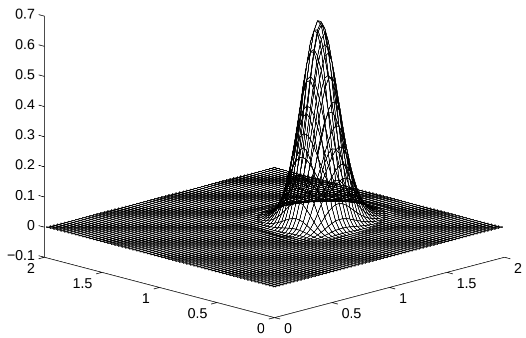

FIGURE 4 – The solution is not oscillating any more, $h_{\mathcal{D}} = 0.0442$

8.1 Numerical test I

First we consider the linear parabolic equation

$$\frac{\partial u}{\partial t} - \nabla \cdot (\delta \nabla u) + \nabla \cdot (\mathbf{V}(\mathbf{x})u) = 0 \text{ in } \Omega \times (0, T).$$

It possesses the exact solution

$$u_{ext}(\mathbf{x}, t) = \frac{1}{200\delta \cdot t + 1} e^{-50\frac{|\mathbf{x}-\mathbf{x}_0-\mathbf{V}t|^2}{200\delta \cdot t+1}},$$

which we numerically compute with prescribing the corresponding boundary and initial conditions. This solution is a Gaussian peak initially centered at the point $\mathbf{x}_0$, which is transported by the field $\mathbf{V}$, and diffusing. We set $\Omega = (0, 2)^2$, $T = 1$, $\mathbf{V} = (0.8, 0.4)$ and $\mathbf{x}_0 = (0.5, 0.5)$ so that during the time interval $[0, T]$ the peak of the exact solution u_{ext} moves from the point $(0.5, 0.5)$ to the point $(1.3, 0.9)$. In this test case we consider two values of the diffusion coefficient δ. The values $\delta = 0.1$ and $\delta = 0.001$ correspond, respectively, to diffusion-dominated and convection-dominated regimes of the problem. Our purpose is to study the behavior of the upwind (3.19) and the centered (3.20) schemes for the discretization of the convection term. First we consider the sequence of uniform space and time discretizations. Then for both schemes we perform numerical simulations on the locally refined (and nonconforming) adaptive mesh (see Figure 1). We use simple heuristic considerations to decrease the computational error. At each time step n the mesh is refined in each volume element K where the local norm $\sum_{\sigma\in\mathcal{E}_K}\frac{m(\sigma)}{d_{K,\sigma}}(u_\sigma^n-u_K^n)^2$ is lager then some tolerance parameter. Conversely, the mesh is unrefined in the regions where the approximate solution is flat.

N	h	# of elements	Err upwind	eoc	Err centered	eoc
8	0.1768	256	$1.33 \cdot 10^{-1}$	–	$1.27 \cdot 10^{-1}$	–
32	0.0884	1024	$1.12 \cdot 10^{-1}$	0.27	$7.00 \cdot 10^{-2}$	1.03
128	0.0442	4096	$8.57 \cdot 10^{-2}$	0.44	$3.33 \cdot 10^{-2}$	1.25
512	0.0221	16384	$5.74 \cdot 10^{-2}$	0.66	$1.66 \cdot 10^{-2}$	1.70
128	–	1951	$6.58 \cdot 10^{-2}$		–	
128	–	1753	–		$3.57 \cdot 10^{-2}$	

TABLE 1 – Number of time steps N, mesh diameter $h_{\mathcal{D}}$, number of elements, the L^2 error for upwind and centered schemes and the experimental order of convergence *eoc* for $\delta = 0.001$

In Table 1 we present simulation results with various mesh sizes $h_{\mathcal{D}}$ and time steps δt in the convection-dominated case. The first three lines of the table indicate the results of the simulations preformed on uniform grids. First we remark that the precision of the upwind scheme is rather poor, which is due to numerical diffusion. On the other hand, if the mesh size $h_{\mathcal{D}}$ is not small enough, then the centered scheme can lead to oscillating solutions, which is the case for $h_{\mathcal{D}} = 0.1768$ and $h_{\mathcal{D}} = 0.0884$ (see Figures 2 and 3). In the last two lines of the table we provide results obtained on adaptive grids. Remark that for both schemes we can significantly reduce the number of unknowns by a suitable concentration of the computational efforts; moreover since we do not impose many constraints on the mesh (in particular it can be nonconforming), the grid adaptation is not an expensive issue. In the diffusion-dominated case (Table 2) both upwind and centered scheme seem to be more accurate. In this case, since the solution is not concentrated in space the mesh adaptation is less efficient.

N	h	# of elements	Err upwind	eoc	Err centered	eoc
8	0.1768	256	$5.91 \cdot 10^{-3}$	–	$3.17 \cdot 10^{-3}$	–
32	0.0884	1024	$3.06 \cdot 10^{-3}$	1.32	$1.37 \cdot 10^{-3}$	1.75
128	0.0442	4096	$1.55 \cdot 10^{-3}$	1.29	$6.42 \cdot 10^{-4}$	1.92
512	0.0221	16384	$7.83 \cdot 10^{-4}$	1.16	$3.10 \cdot 10^{-4}$	1.98
128	-	3589	$1.37 \cdot 10^{-3}$		–	
128	-	3586	–		$6.96 \cdot 10^{-4}$	

TABLE 2 – Number of time steps N, mesh diameter $h_{\mathcal{D}}$, number of elements, the L^2 error for upwind and centered schemes and the experimental order of convergence *eoc* for $\delta = 0.1$

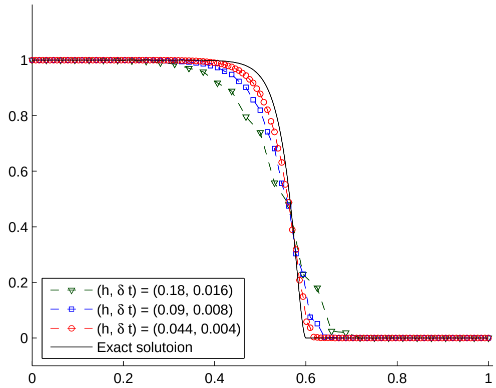

FIGURE 5 – Numerical solution for $\delta = 0.01, t = 0.5$, upwind scheme.

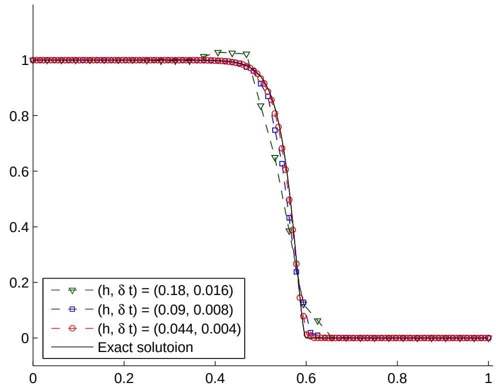

FIGURE 6 – Numerical solution for $\delta = 0.01, t = 0.5$, centered scheme.

8.2 Numerical test II

Next we consider a degenerate parabolic equation which possesses a traveling wave solution, namely

$$\frac{\partial(u^{\frac{1}{2}})}{\partial t} - \nabla \cdot (\delta \nabla u) + \nabla \cdot ((v, 0, 0)u) = 0$$

in the domain

$$\Omega = (0, 1)^2 \text{ and } T = 1.$$

This equation admits the following 1-dimensional exact solution

$$u(x, y, t) = (1 - e^{\frac{v}{2\delta}(x - vt - p)})^2 \text{ for } x \leq vt + p,$$

$$u(x, y, t) = 0 \text{ for } x > vt + p,$$

where p, v, δ are parameters to be defined. We set $p = 0.2$, $v = 0.8$, and consider two values of δ, namely $\delta = 0.01$, $\delta = 0.0001$. The initial state is given by the exact solution at the time $t = 0$ and we prescribe corresponding Dirichlet boundary conditions on the sides $x = 0$ and $x = 1$. The null flux boundary condition is imposed on the remaining part of the boundary.

First remark that both schemes leads to a solution with compact support ; moreover the speed of propagation is in good agrement with the theoretical one. As it can expected, in the diffusion dominated case ($\delta = 0.01$, Figures 5 and 6) the second order centered scheme is clearly more accurate then the upwind scheme. However it produces small oscillations if the mesh size $h_{\mathcal{D}}$ is too large. In the case that ($\delta = 0.0001$, Figures 7 and 8), even though it is diffusive, the upwind scheme is more adapted. The centered scheme does not resolve the sharp interface better then the upwind one ; moreover it is oscillatory even on a fine mesh.

8.3 Numerical test III

Let us consider the equation

$$\frac{\partial(u + u^{\frac{1}{2}})}{\partial t} - \nabla \cdot (\mathbf{\Lambda}(\mathbf{x}) \nabla u) + \nabla \cdot (\mathbf{V}(\mathbf{x}) u) + \frac{1}{2} u^{\frac{1}{2}} = 0$$

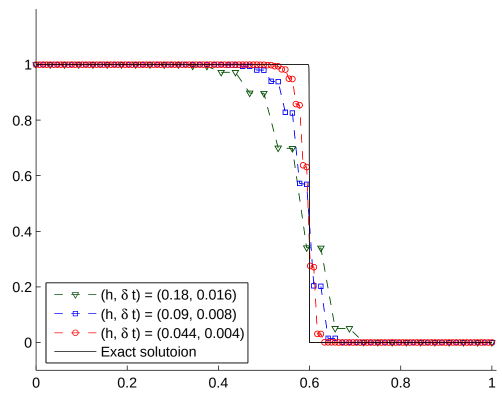

FIGURE 7 – Numerical solution for $\delta = 0.0001, t = 0.5$, upwind scheme.

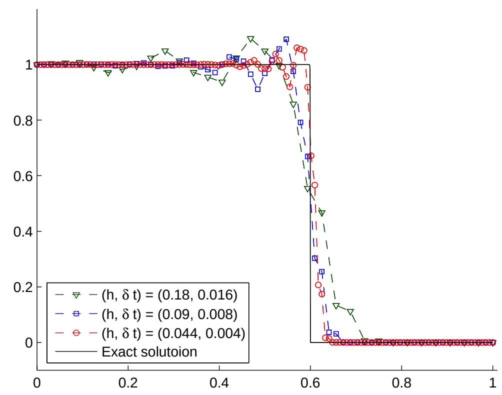

FIGURE 8 – Numerical solution for $\delta = 0.0001, t = 0.5$, centered scheme.

in the 3-dimensional space domain $\Omega = (0,2)\times(0,1)\times(0,1)$. We define the discontinuous $\mathbf{\Lambda}$ and $\mathbf{V}$ fields as follows :
For all $x_1 \leq 1$ we set

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{V} = (4,0,0);$$

for all $x_1 > 1$ we set

$$\mathbf{\Lambda} = \begin{pmatrix} 8 & -5 & -2 \\ -5 & 20 & -7 \\ -2 & -7 & 19 \end{pmatrix} \quad \text{and} \quad \mathbf{V} = (4,8,8).$$

The initial and the Dirichlet boundary conditions are given by the exact solution

$$u(\mathbf{x},t) = e^{x_1+x_2+x_3-t-3}.$$

We remark that the velocity field $\mathbf{V}$ and the total flux $\mathbf{\Lambda}(\mathbf{x})\nabla u + \mathbf{V}(\mathbf{x})u$ have a continuous normal trace across the discontinuity $x = 1$. We perform the simulations on 3-dimensional hexahedral meshes with random refinement (see Figure 9), so that the mesh is nonmatching. In Table 3 we present simulation results with various mesh sizes $h_{\mathcal{D}}$ and time steps δt ; we denote by Err the maximum relative error in L^2-norm, namely

$$Err = \max_{n\in\{1,\ldots,N\}} \frac{\|u_{h,t}(\cdot,t_n) - \Pi_{h_{\mathcal{D}},\delta t} u(\cdot,t_n)\|_{L^2(\Omega)}}{\|\Pi_{h_{\mathcal{D}},\delta t} u(\cdot,t_n)\|_{L^2(\Omega)}}.$$

9 Remarks on the implementation of the numerical scheme

In this section we assume that in addition to $(\mathcal{H}_1)$ and $(\mathcal{H}_4)$ the following hypothesis is satisfied.

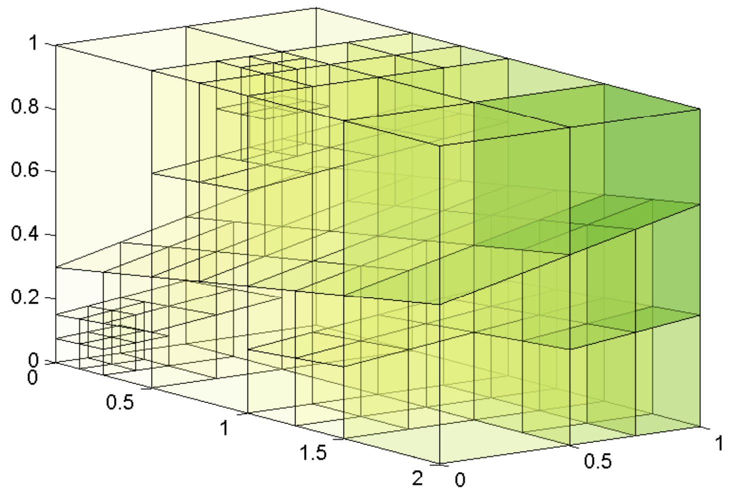

FIGURE 9 – Approximate solution on the nonmatching hexahedral mesh at $t = 1$

N	h	# of elements	Err upwind	Err centered
50	0.99	58	$4.35 \cdot 10^{-2}$	$2.69 \cdot 10^{-2}$
100	0.52	436	$1.65 \cdot 10^{-2}$	$1.08 \cdot 10^{-2}$
200	0.27	1570	$7.16 \cdot 10^{-3}$	$5.28 \cdot 10^{-3}$

TABLE 3 – Number of time steps N, mesh diameter $h_{\mathcal{D}}$, number of elements, number of faces and the relative error for nonmatching hexahedral meshes

$(\mathcal{H}_{\beta,f})$ The functions $\varphi = \beta^{-1}$ and $g = f \circ \varphi$ are piecewise continuously differentiable.

Let the initial data be given; we recall that at each time step one has to solve the nonlinear problem (3.10)-(3.12), which in view of (3.17) and (3.18) may be written as

For each $n \in \{1, \dots, N\}$ find $u^n \in X_{\mathcal{D},0}$ such that for all $v^n \in X_{\mathcal{D},0}$

$$\sum_{K\in\mathcal{M}} \frac{m(K)}{\delta t} v_K^n(\beta(u_K^n) - \beta(u_K^{n-1})) + \sum_{K\in\mathcal{M}} \sum_{\sigma,\sigma'\in\mathcal{E}_K} \mathbb{D}_K^{\sigma,\sigma'}(v_K^n - v_\sigma^n)(u_K^n - u_{\sigma'}^n)$$
$$+ \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (v_K^n - v_\sigma^n)(A(V_{K,\sigma})u_K^n + B(V_{K,\sigma})u_\sigma^n) + \sum_{K\in\mathcal{M}} m(K) v_K^n f(u_K^n) \tag{9.1}$$
$$= \sum_{K\in\mathcal{M}} m(K) v_K^n q_K^n.$$

Since $\beta'(0) = +\infty$ the Newton method can not be directly applied. In order to overcome this difficulty, the first idea would be to introduce new discrete unknowns

$$w^n = \beta(u^n), \text{ and thus } u^n = \varphi(w^n), \text{ where } \varphi = \beta^{-1},$$

so that the equation (9.1) becomes

$$\sum_{K\in\mathcal{M}} \frac{m(K)}{\delta t} v_K^n (w_K^n - w_K^{n-1}) + \sum_{K\in\mathcal{M}} \sum_{\sigma,\sigma'\in\mathcal{E}_K} \mathbb{D}_K^{\sigma,\sigma'} (v_K^n - v_\sigma^n)\,(\varphi(w_K^n) - \varphi(w_{\sigma'}^n))$$

$$+ \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (v_K^n - v_\sigma^n)\,(A(V_{K,\sigma})\varphi(w_K^n) + B(V_{K,\sigma})\varphi(w_\sigma^n)) + \sum_{K\in\mathcal{M}} m(K) v_K^n g(w_K^n) \tag{9.2}$$

$$= \sum_{K\in\mathcal{M}} m(K) v_K^n q_K^n.$$

However we remark that the linearized system corresponding to the problem (9.2) may not be solvable. This situation may occur, if at some Newton iteration m the corresponding value $w_\sigma^{m,n}$ (cf. (9.4) below) is equal to 0, which than implies that $\varphi'(w_\sigma^{m,n}) = 0$. Noting that the system (9.2) depends on $(w_\sigma^n)_{\sigma\in\mathcal{E}}$ only through the terms $(\varphi(w_\sigma^n))_{\sigma\in\mathcal{E}}$, this lead us to choose a new set of discrete unknowns

$$z_K^n = \beta(u_K^n) \text{ for all } K \in \mathcal{M} \text{ and } z_\sigma^n = u_\sigma^n \text{ for all } \sigma \in \mathcal{E},$$

so that the system (9.2) takes the form

$$\sum_{K\in\mathcal{M}} \frac{m(K)}{\delta t} v_K^n (z_K^n - z_K^{n-1}) + \sum_{K\in\mathcal{M}} \sum_{\sigma,\sigma'\in\mathcal{E}_K} \mathbb{D}_K^{\sigma,\sigma'} (v_K^n - v_\sigma^n)\,(\varphi(z_K^n) - z_{\sigma'}^n)$$

$$+ \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (v_K^n - v_\sigma^n)\,(A(V_{K,\sigma})\varphi(z_K^n) + B(V_{K,\sigma}) z_\sigma^n) + \sum_{K\in\mathcal{M}} m(K) v_K^n g(z_K^n) \tag{9.3}$$

$$= \sum_{K\in\mathcal{M}} m(K) v_K^n q_K^n.$$

The solvability of the linearized system corresponding to (9.3) is still not obvious. However we show in Lemma 9.1 below that it is invertible for certain forms of the discrete convective flux (including the upwind scheme (3.19)) under a suitable condition on the time step. Applying Newton's method leads to the following sequence of M linearized problems to be solved on each time step n. *For each $m \in \{1, \ldots, M\}$ find $\delta z^m \in X_{\mathcal{D},0}$ such that for all $v^m \in X_{\mathcal{D},0}$*

$$\begin{aligned}
&\sum_{K\in\mathcal{M}} \frac{m(K)}{\delta t} v_K^m \delta z_K^m + \sum_{K\in\mathcal{M}} \sum_{\sigma,\sigma'\in\mathcal{E}_K} (v_K^m - v_\sigma^m) \mathbb{D}_K^{\sigma,\sigma'} (\varphi'(z_K^{n,m-1})\delta z_K^m - \delta z_{\sigma'}^m) \\
&\qquad + \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (v_K^m - v_\sigma^m)(A(V_{K,\sigma})\varphi'(z_K^{n,m-1})\delta z_K^m + B(V_{K,\sigma})\delta z_\sigma^m) \\
&\qquad + \sum_{K\in\mathcal{M}} m(K) v_K^n g'(z_K^{n,m-1}) \delta z_K^m \\
&\qquad = \sum_{K\in\mathcal{M}} v_K^m b_K(z_K^{n-1}, z^{n,m-1}, q_K^n) + \sum_{\sigma\in\mathcal{E}} v_\sigma^m b_\sigma(z^{n,m-1}),
\end{aligned} \tag{9.4}$$

where $z^{n,0} = z^{n-1}, z^{n,m+1} = z^{n,m} + \delta z^{m+1}$ (we recall that they correspond both to values indexed by K and by σ), and where the terms b_K and b_σ on the right-hand side correspond to the accumulation of all the data evaluated on previous time and linearization steps. Taking advantage of the variational formulation we will show below that the system (9.4) is invertible.

Remark 9.1. *Note that the functions φ and g are piecewise C^1 functions, so that the corresponding derivatives φ' and g' are finite but not uniquely defined ; in the points of discontinuity a certain value of φ' and g' has to be fixed. Also remark that since the numerical scheme does not preserve the maximum principle, the functions φ and g have to be defined on whole $\mathbb{R}$ (i.e. for negative concentrations), so that the choice of extensions may affect their smoothness. In practice it may occur that φ, g are smooth for physically relevant concentrations $s \in [0,1]$, but not on $\mathbb{R}$.*

Next, taking advantage of the variational formulation we show that the system (9.4) is invertible.

Lemma 9.1. *Let $(\alpha_K)_{K\in\mathcal{M}}, (g'_K)_{K\in\mathcal{M}}$ and $(\varphi'_K)_{K\in\mathcal{M}} \geq 0$ be three sequences of real values, for all $v, w \in X_\mathcal{D}$ we define the following bilinear forms*

$$< v, u >_{D,\alpha} = \sum_{K\in\mathcal{M}} \sum_{\sigma,\sigma'\in\mathcal{E}_K} \mathbb{D}_K^{\sigma,\sigma'} (\alpha_K v_K - v_\sigma)(\varphi'_K u_K - u_{\sigma'}), \tag{9.5}$$

$$< v, u >_{C,\alpha} = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\alpha_K v_K - v_\sigma) \left(A(V_{K,\sigma}) \varphi'_K u_K + B(V_{K,\sigma}) u_\sigma \right), \tag{9.6}$$

and

$$< v, u >_\alpha = \sum_{K\in\mathcal{M}} \alpha_K m(K) \left(\frac{1}{\delta t} + g'_K \right) v_K u_K + < v, u >_{D,\alpha} + < v, u >_{C,\alpha} . \tag{9.7}$$

Assume that the discrete flux $F^C_{K,\sigma}(\cdot)$ is monotone, namely that $A(V_{K,\sigma}) \geq 0$ and $B(V_{K,\sigma}) \leq 0$ as it is the case for the upwind scheme (3.19) ; also assume that $1/\delta t + \min_{K\in\mathcal{M}} g'_K > 0$; then there exist $(\alpha_K)_{K\in\mathcal{M}}$ such that $< \cdot, \cdot >_\alpha$ is coercive.

Proof : We begin with the form $< \cdot, \cdot >_{D,\alpha}$ by noting that in view of symmetry of $\mathbb{D}_K^{\sigma,\sigma'}$ one has

$$\sum_{\sigma,\sigma'\in\mathcal{E}_K} \mathbb{D}_K^{\sigma,\sigma'} u_{\sigma'} = \sum_{\sigma,\sigma'\in\mathcal{E}_K} \mathbb{D}_K^{\sigma,\sigma'} u_\sigma = \frac{1}{2} \sum_{\sigma,\sigma'\in\mathcal{E}_K} \mathbb{D}_K^{\sigma,\sigma'} (u_\sigma + u_{\sigma'}).$$

Thus, the definition (9.5) implies

$$\begin{aligned} < u, u >_{D,\alpha} &= \sum_{K\in\mathcal{M}} \sum_{\sigma,\sigma'\in\mathcal{E}_K} \mathbb{D}_K^{\sigma,\sigma'} \left(\alpha_K \varphi'_K (u_K)^2 - \alpha_K u_K u_{\sigma'} - \varphi'_K u_K u_\sigma + u_\sigma u_{\sigma'} \right) \\ &= \sum_{K\in\mathcal{M}} \sum_{\sigma,\sigma'\in\mathcal{E}_K} \mathbb{D}_K^{\sigma,\sigma'} \left(\alpha_K \varphi'_K (u_K)^2 - \frac{1}{2} (\alpha_K + \varphi'_K) u_K (u_{\sigma'} + u_\sigma) + u_\sigma u_{\sigma'} \right) . \end{aligned} \tag{9.8}$$

The equality $4ab = (a+b)^2 - (a-b)^2$ applied to the first term on the right-hand side of (9.8) implies that

$$\begin{aligned} < u, u >_{D,\alpha} = \sum_{K\in\mathcal{M}} \sum_{\sigma,\sigma'\in\mathcal{E}_K} \mathbb{D}_K^{\sigma,\sigma'} \left(\frac{1}{2} (\alpha_K + \varphi'_K) u_K - u_{\sigma'} \right) \left(\frac{1}{2} (\alpha_K + \varphi'_K) u_K - u_\sigma \right) \\ - \frac{1}{4} \sum_{K\in\mathcal{M}} D_K \left(\alpha_K - \varphi'_K \right)^2 (u_K)^2 , \end{aligned}$$

where $D_K = \sum_{\sigma,\sigma' \in \mathcal{E}_K} \mathbb{D}_K^{\sigma,\sigma'}$, which is a positive value since $\mathbb{D}_K$ is positive definite for all $K \in \mathcal{M}$. Thus

$$< u, u >_{D,\alpha} \geq < l_\alpha(u), l_\alpha(u) >_D - \frac{1}{4} \sum_{K \in \mathcal{M}} D_K \left(\alpha_K - \varphi'_K\right)^2 \left(u_K\right)^2, \tag{9.9}$$

where l_α is the linear application defined by

$$(l_\alpha(u))_K = \frac{1}{2}(\alpha_K + \varphi'_K) u_K \text{ and } (l_\alpha(u))_\sigma = u_\sigma. \tag{9.10}$$

Next we consider the bilinear form $< \cdot, \cdot >_{C,\alpha}$. By (9.6) we have

$$\begin{aligned} < u, u >_{C,\alpha} &= \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} A(V_{K,\sigma}) \alpha_K \varphi'_K (u_K)^2 \\ &+ \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} B(V_{K,\sigma}) \alpha_K u_K u_\sigma - A(V_{K,\sigma}) \varphi'_K u_K u_\sigma - B(V_{K,\sigma})(u_\sigma)^2 \end{aligned}$$

Recall that by the assumption $A(V_{K,\sigma}) \geq 0$ and $B(V_{K,\sigma}) \leq 0$ so that

$$-A(V_{K,\sigma}) \varphi'_K u_K u_\sigma \geq -\frac{A(V_{K,\sigma})}{2} (\varphi'_K)^2 (u_K)^2 - \frac{A(V_{K,\sigma})}{2} (u_\sigma)^2$$

and

$$B(V_{K,\sigma}) \alpha_K u_K u_\sigma \geq \frac{B(V_{K,\sigma})}{2} \alpha_K^2 (u_K)^2 + \frac{B(V_{K,\sigma})}{2} (u_\sigma)^2,$$

which implies

$$\begin{aligned} < u, u >_{C,\alpha} &\geq \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} \left(\frac{1}{2} B(V_{K,\sigma}) \alpha_K^2 + A(V_{K,\sigma}) \alpha_K \varphi'_K - \frac{1}{2} A(V_{K,\sigma}) (\varphi'_K)^2 \right) (u_K)^2 \\ &- \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} \frac{1}{2} \left(A(V_{K,\sigma}) + B(V_{K,\sigma}) \right) (u_\sigma)^2. \end{aligned} \tag{9.11}$$

Note that by the assumption $(\mathcal{H}_{C,2})$, $A(V_{K,\sigma}) + B(V_{K,\sigma}) = V_{K,\sigma}$. Thus, in view of Dirichlet boundary condition the last term in the inequality (9.11) vanishes. Adding and subtracting the term $\frac{1}{2} A(V_{K,\sigma}) \alpha_K^2 u_K^2$ and using again $(\mathcal{H}_{C,2})$ we obtain

$$< u, u >_{C,\alpha} \geq -\frac{1}{2} \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} A(V_{K,\sigma}) (\alpha_K - \varphi'_K)^2 (u_K)^2 + \frac{1}{2} \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} V_{K,\sigma} \alpha_K^2 u_K^2. \tag{9.12}$$

The assumptions $(\mathcal{H}_3)$ implies

$$< u, u >_{C,\alpha} \geq -\frac{1}{2} \sum_{K \in \mathcal{M}} A_K (\alpha_K - \varphi'_K)^2 (u_K)^2, \tag{9.13}$$

where $A_K = \sum_{\sigma \in \mathcal{E}_K} A(V_{K,\sigma})$. Gathering (9.7), (9.9) and (9.13) we obtain

$$\begin{aligned} < u, u >_\alpha &\geq < l_\alpha(u), l_\alpha(u) >_D \\ &+ \sum_{K \in \mathcal{M}} \left(\alpha_K m(K) \left(\frac{1}{\delta t} + g'_K \right) - \frac{1}{2} \left(\frac{1}{2} D_K + A_K \right) (\alpha_K - \varphi'_K)^2 \right) (u_K)^2. \end{aligned}$$

For all $K \in \mathcal{M}$ it is always possible to find $\alpha_K > \varphi'_K \geq 0$ such that

$$\alpha_K m(K) \left(\frac{1}{\delta t} + g'_K \right) \geq \frac{1}{2} \left(\frac{1}{2} D_K + A_K \right) (\alpha_K - \varphi'_K)^2. \tag{9.14}$$

Therefor in view of Lemma 3.4 there exist a positive C such that

$$< u, u >_{D,\alpha} \geq < l_\alpha(u), l_\alpha(u) >_D \geq C |l_\alpha(u)|^2_{X_\mathcal{D}}.$$

The fact that $\alpha_K + \varphi'_K > 0$ implies $Ker(l) = \{0\}$, so that the application $|l_\alpha(\cdot)|_{X_\mathcal{D}}$ define a norm on $X_{\mathcal{D},0}$. Since for a fixed discretization $\mathcal{D}$ all the norms on finite dimensional space $X_{\mathcal{D},0}$ are equivalent, we finally obtain that

$$< u, u >_{D,\alpha} \geq C |l_\alpha(u)|^2_{X_\mathcal{D}} \geq \widetilde{C} |u|^2_{X_\mathcal{D}}$$

for some positive $\widetilde{C}$. □

Remark that in view of the definition of g (see $(\mathcal{H}_{\beta,f})$) the condition (5.4) implies $1/\delta t + g'(w) > 0$ for all $w \in \mathbb{R}$.

Deuxième partie

Hybrid finite volume scheme for two-phase flow in porous media

Abstract We apply a finite volume method on general meshes for the discretization of an incompressible and immiscible two-phase flow in porous media. The problem is considered in the global pressure formulation. Mathematically, it amounts to solve an elliptic equation for the global pressure, with an anisotropic and heterogeneous permeability tensor coupled to a parabolic degenerate convection-diffusion equation for a saturation, again with the same permeability tensor. Extending ideas which we had previously developed for the numerical solution of a degenerate parabolic convection-reaction-diffusion equation we discretize the diffusion terms by means of a hybrid finite volume scheme, while we use a Godunov scheme for the non monotone convection flux. We prove the convergence of the numerical scheme in arbitrary space dimension and we present results of a number of numerical tests in space dimension two.

This work was supported by the GdR MoMaS (PACEN/CNRS, ANDRA, BRGM, CEA, EdF, IRSN), France.

10 Introduction

The two-phase flow in porous media is an important problem arising in many engineering and scientific context as e.g the secondary oil recovery, the basin modeling and the contaminated sites remediation [21], [17], [44]. In this paper we consider the simplified *dead-oil* model, that is to say, we assume that there are only two incompressible and immiscible fluids. The problem is described by the mass conservation of each phase together with Darcy-Muskat law

$$\omega \partial_t s_o - \nabla \cdot \left(\mathbf{K} \frac{kr_o(s_o)}{\mu_o} (\nabla p_o - \rho_o \mathbf{g}) \right) = k_o \tag{10.1}$$

$$\omega \partial_t s_w - \nabla \cdot \left(\mathbf{K} \frac{kr_w(s_w)}{\mu_w} (\nabla p_w - \rho_w \mathbf{g}) \right) = k_w \tag{10.2}$$

where ω is the porosity, $\mathbf{K}$ absolute permeability, s_w stands for the wetting phase (water) saturation and s_o for the non-wetting phase (oil) saturation. The relative permeabilities $kr_o(s_o)$ and $kr_w(s_w)$ model the effects of coexistence of both phases in porous medium; ρ_α and μ_a denote the densities and the viscosity of phase $\alpha \in \{o, w\}$, $\mathbf{g}$ is the gravity vector. It is also assumed that the porous medium is saturated and that two phases are immiscible

$$s_o + s_w = 1. \tag{10.3}$$

The phase pressures are related by a capillary pressure law

$$p_o + p_w = \pi(s). \tag{10.4}$$

In view of (10.3) we can take as unknown only one saturation, for instance $s = s_o$. We denote by $\lambda_o(s) = \dfrac{kr_o(s)}{\mu_o}$ and $\lambda_w(s) = \dfrac{kr_w(1-s)}{\mu_w}$ the relative mobilities, which are such that $\lambda_o(0) = \lambda_w(1) = 0$. In this paper we transform the system (10.1)-(10.4) into the so-called global pressure or fractional flow formulation. The global pressure p which is defined by

$$p = p_w + \int_0^s \frac{\lambda_o(a)}{\lambda_o(a) + \lambda_w(a)} \pi(a) da$$

was first introduced in [14] and [44]. We also define the fractional flow f, total mobility λ and capillary diffusion φ

$$\lambda(s) = \lambda_o(s) + \lambda_w(s), \quad f(s) = \frac{\lambda_o(s)}{\lambda(s)}, \quad \varphi(s) = \int_0^s \lambda_w(\tau) f(\tau) \pi'(\tau) d\tau.$$

The system (10.1)-(10.4) is equivalent to

$$-\nabla \cdot \mathbf{K} \left(\lambda(s) \nabla p - \xi(s) \mathbf{g} \right) = k_w + k_o \ \text{ in } \ \Omega \times (0, T), \tag{10.5}$$

$$\mathbf{q} = -\mathbf{K} \left(\lambda(s) \nabla p - \xi(s) \mathbf{g} \right) \ \text{ in } \ \Omega \times (0, T), \tag{10.6}$$

$$\omega \frac{\partial s}{\partial t} + \nabla \cdot (\mathbf{q} f(s) + \gamma(s) \mathbf{K} \mathbf{g}) - \nabla \cdot (\mathbf{K} \nabla \varphi(s)) = k_o \ \text{ in } \ \Omega \times (0, T), \tag{10.7}$$

where

$$\gamma(s) = (\rho_o - \rho_w) \frac{\lambda_o(s) \lambda_w(s)}{\lambda_o(s) + \lambda_w(s)} \qquad \text{and} \qquad \xi(s) = \left(\lambda_o(s) \rho_o + \lambda_w(s) \rho_w \right).$$

The usual assumption $\lambda_o + \lambda_w \geq \underline{\lambda} > 0$ implies that the first equation is uniformly elliptic in p, whereas the second one is parabolic degenerate with respect to the saturation s. The system (10.5)-(10.7) has a remarkable mathematical structure, which permits to obtain a number of energy estimates. Many numerical methods were proposed for solving two-phase problem, such as finite element methods (see e.g. [64], [63], [44], [16], [46], [48], [50], [51]), discontinuous Galerkin (see e.g. [19], [61]) and finite volume methods (see e.g. [5], [67], [90], [71], [8], [91], [12]). The convergence results for the different finite elements schemes were obtained in [16], [46] and [51]. For finite volume schemes convergence was shown in [5], [90] and [71], where the initial formulation (10.1)-(10.4) was considered. We also would like to mention some interesting convergence results which were obtained in the case of miscible displacement [94], [97], [43].

The heterogeneity and anisotropy of porous media is a numerical challenge even when studying the elliptic problem derived from Darcy's law for one-phase problem. Many schemes was proposed and analyzed in last decades for its discretization. See [79] and [70] for more references and for the detailed description and comparison of those numerical methods. In this paper we propose an implicit fully coupled Hybrid finite volume scheme for the system (10.5)-(10.7). The discretization of the diffusion terms is based upon a hybrid finite volume method [66], which allows the tensor $\mathbf{K}$ to be anisotropic and highly variable in space. Remark that heterogeneity also affects the first order terms $\mathbf{K}(\mathbf{x})\mathbf{g}\xi(\cdot)$ and $\mathbf{q}f(\cdot) + \mathbf{K}(\mathbf{x})\mathbf{g}\gamma(\cdot)$, which may be discontinuous with respect to the space variable $\mathbf{x}$; in that case they require a suitable treatment see e.g. [100], [99], [75]. We apply the Godunov scheme proposed in [82], which seems to be a natural choice, since the hybrid (interface) unknowns are used.

Assumptions on the data : $(\mathcal{H}_1)$ $\varphi \in C(\mathbb{R}), \varphi(0) = 0$, is a strictly increasing piecewise continuously differentiable Lipschitz continuous function with a Lipschitz constant L_φ. We assume that the function φ^{-1} is Hölder continuous, namely that there exists $H_\varphi > 0$ and $\alpha \in (0, 1]$ such that $|s_1 - s_2| \leq H_\varphi |\varphi(s_1) - \varphi(s_2)|^\alpha$. It is also such that

$\varphi(s) = s$ for all $s < 0$ and $\varphi(s) - \varphi(1) = s - 1$ for all $s > 1$;
$(\mathcal{H}_2)$ The functions $\lambda, \xi, \gamma, f \in C([0,1])$ are Lipschitz continuous ; for $\mathfrak{f} = \lambda, \xi, \gamma, f$ we denote by $L_{\mathfrak{f}}$ the corresponding Lipschitz constant.
$(\mathcal{H}_{2a})$ λ is such that $0 < \underline{\lambda} \leq \lambda(s)$;
$(\mathcal{H}_{2b})$ ξ and γ are convex functions, moreover γ is such that $\gamma(0) = \gamma(1) = 0$; $\mathbf{g}$ is a constant vector from $\mathbb{R}^d$;
$(\mathcal{H}_{2c})$ f is a nondecreasing function and it satisfies $f(0) = 0, f(1) = 1$;
$(\mathcal{H}_3)$ The functions λ, ξ, γ, f are constant outside of $(0, 1)$, i.e. for $\mathfrak{f} = \lambda, \xi, \gamma, f$ we assume that $\mathfrak{f}(s) = \mathfrak{f}(0)$ for all $s < 0$ and $\mathfrak{f}(s) = \mathfrak{f}(1)$ for all $s > 1$.
$(\mathcal{H}_5)$ $s_0 \in L^\infty(\Omega)$; $\omega \in L^\infty(\Omega)$ and such that $0 < \underline{\omega} \leq \omega(\mathbf{x}) \leq \overline{\omega}$ for a.e. in $\mathbf{x} \in \Omega$;
$(\mathcal{H}_6)$ $k_o, k_w \in L^\infty(0, T; L^2(\Omega))$ and such that $k_o + k_w \geq 0$ a.e. in Q_T.

Remark 10.1. *The physical range of values for the saturation is* $[0, 1]$*, however we extend the definition of all nonlinear functions in* (10.5)-(10.7) *outside of* $[0, 1]$*, since the numerical scheme which we study does not preserve neither the maximum principle, nor the positivity of* s *(nor the bound* $s \leq 1$*). It is worth noting that one can replace* $(\mathcal{H}_3)$ *by the assumption that the functions* λ, ξ, γ, f *are in* $L^\infty(\mathbb{R})$*. In turn the assumption* $(\mathcal{H}_1)$ *can also weakened, namely the capillary diffusion* φ *can be extended by an arbitrary strictly increasing function, such that* φ^{-1} *is Lipschitz continuous on* $\mathbb{R} \setminus (0, 1)$*.*

Assumptions on the geometry :
$(\mathcal{H}_{4a})$ Ω is a polyhedral open bounded connected subset of $\mathbb{R}^d$, with $d \in \mathbb{N}^*$, and $\partial\Omega = \overline{\Omega} \backslash \Omega$ its boundary.
$(\mathcal{H}_{4b})$ $\mathbf{K}$ is a piecewise constant function from Ω to $\mathcal{M}_d(\mathbb{R})$, where $\mathcal{M}_d(\mathbb{R})$ denotes the set of real $d \times d$ matrices. More precisely we assume that there exist a finite family $(\Omega_i)_{i \in \{1,\ldots,I\}}$ of open connected polyhedral in $\mathbb{R}^d$, such that $\overline{\Omega} = \bigcup_{i \in \{1,\ldots,I\}} \overline{\Omega_i}$, $\Omega_i \bigcap \Omega_j = \emptyset$ if $i \neq j$ and such that $\mathbf{K}(\mathbf{x})|_{\Omega_i} = \mathbf{K}_i \in \mathcal{M}_d(\mathbb{R})$. By $\Gamma_{i,j}$ we denote the interface between the sub-domains i, j, $\Gamma_{i,j} = \overline{\Omega_i} \bigcap \overline{\Omega_j}$. We suppose that there exist two positive constants $\underline{\mathrm{K}}$ and $\overline{\mathrm{K}}$ such that for all $i \in \{1, \ldots, I\}$ the eigenvalues of the symmetric positive definite $\mathbf{K}_i$ are included in $[\underline{\mathrm{K}}, \overline{\mathrm{K}}]$.

Remark 10.2. *For the sake of simplicity we have assumed that the heterogeneity of the medium is only expressed through the* $\mathbf{x}$*-dependence of the absolute permeability tensor* $\mathbf{K} = \mathbf{K}(\mathbf{x})$*. However it is very simple to extend the analysis to the case where* λ, ξ, γ *and* f *depend on the rock type. On the other hand if we suppose that* φ *is discontinuous in space, this may lead to significant difficulties. The analysis of the case where the capillary pressure field (and also* φ*) is discontinuous was carried out in [38] and [26]. It is also worth noting that the partitioning of* Ω *introduced in* $\mathcal{H}_{4b}$ *is only used in order to provide a control on the gravity terms (see Remark 12.2) below). In the case that the gravity effects are neglected, one can consider a fully heterogenous permeability field* $\mathbf{K}$*.*

Let $T > 0$, we consider the system (10.5)-(10.7) in the domain $Q_T = \Omega \times (0, T)$ together with the initial condition

$$s(\cdot, 0) = s_0 \text{ in } \Omega \tag{10.8}$$

and homogeneous Dirichlet boundary conditions

$$p = 0 \text{ and } s = 0 \text{ on } \partial\Omega \times (0, T). \tag{10.9}$$

We now present the definition of a weak solution of the problem (10.5)-(10.9).

Definition 10.1 (Weak solution). *A function pair (s,p) is a weak solution of the problem* (10.5)-(10.6) *if*
(i) $s \in L^\infty(0,T;L^2(\Omega))$;
(ii) $\varphi(s) \in L^2(0,T;H_0^1(\Omega))$;
(iii) $p \in L^\infty(0,T;H_0^1(\Omega))$;
(iv) for all $\psi, \chi \in L^2(0,T;H_0^1(\Omega))$ *with* $\varphi_t \in L^\infty(Q_T)$, $\varphi(\cdot,T)=0$, *s and p satisfy the integral equalities*

$$-\int_0^T\int_\Omega s\psi_t\ d\mathbf{x}dt - \int_\Omega s_0\psi(\cdot,0)\ d\mathbf{x} - \int_0^T\int_\Omega \mathbf{K}(f(s)\mathbf{q}+\gamma(s)\mathbf{g})\cdot\nabla\psi\ d\mathbf{x}dt + \int_0^T\int_\Omega \mathbf{K}\nabla\varphi(s)\cdot\nabla\psi\ d\mathbf{x}dt = \int_0^T\int_\Omega k_o\psi\ d\mathbf{x}dt$$

and

$$-\int_0^T\int_\Omega \mathbf{q}\cdot\nabla\chi\ d\mathbf{x}dt = \int_0^T\int_\Omega (k_w+k_o)\chi\ d\mathbf{x}dt,$$

where $\mathbf{q}$ *is given by* (10.6).

In Section 11 we present the finite volume scheme and some technical lemmas. In Section 12 we provide the a priori estimates and we prove an existence of a discrete solution. We prove the estimates on space and time translates of a discrete solution in Section 13, those estimates allow to establish a strong convergence property for a subsequence of discrete saturation. The convergence result is shown in Section 14. Finally in Section 15 we present a number of numerical results obtained on different two-dimensional meshes.

11 The finite volume scheme

11.1 The main definitions

In order to describe the numerical scheme we introduce below some notations related to the space and time discretization, which follows [66].

Definition 11.1 (Discretization of Ω). *Let Ω be a polyhedral open bounded connected subset of $\mathbb{R}^d$, with $d \in \mathbb{N}^*$, $\partial\Omega = \overline{\Omega}\backslash\Omega$ its boundary, and $(\Omega_i)_{i\in\{1,\dots,I\}}$ it's partition in the sense of $(\mathcal{H}_4)$. A discretization of Ω, denoted by $\mathcal{D}$, is defined as the triplet $\mathcal{D}=(\mathcal{M},\mathcal{E},\mathcal{P})$, where :*

1. *$\mathcal{M}$ is a finite family of non empty connex open disjoint subsets of Ω (the "control volumes") such that $\overline{\Omega}=\bigcup_{K\in\mathcal{M}}\overline{K}$. For any $K\in\mathcal{M}$, let $\partial K=\overline{K}\backslash K$ be the boundary of K ; we define $m(K)>0$ as the measure of K and h_K as the diameter of K. We also assume that the mesh resolve the structure of the medium, i.e. for all $K\in\mathcal{M}$ there exist $i\in\{1,\dots,I\}$ such that $K\subset\Omega_i$.*

2. *$\mathcal{E}$ is a finite family of disjoint subsets of $\overline{\Omega}$ (the "edges" of the mesh), such that, for all $\sigma\in\mathcal{E}$, σ is a non empty open subset of a hyperplane of $\mathbb{R}^d$, whose $(d-1)$-dimensional measure $m(\sigma)$ is strictly positive. We also assume that, for all $K\in\mathcal{M}$, there exists a subset $\mathcal{E}_K$ of $\mathcal{E}$ such that $\partial K=\bigcup_{\sigma\in\mathcal{E}_K}\overline{\sigma}$. For each $\sigma\in\mathcal{E}$, we set $\mathcal{M}_\sigma=\{K\in\mathcal{M}|\sigma\in\mathcal{E}_K\}$. We then assume that, for all $\sigma\in\mathcal{E}$, either $\mathcal{M}_\sigma$ has exactly one element and then*

$\sigma \in \partial\Omega$ (the set of these interfaces called boundary interfaces, is denoted by $\mathcal{E}_{ext}$) or $\mathcal{M}_\sigma$ has exactly two elements (the set of these interfaces called interior interfaces, is denoted by $\mathcal{E}_{int}$). For all $\sigma \in \mathcal{E}$, we denote by $\mathbf{x}_\sigma$ the barycenter of σ. For all $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$, we denote by $\mathbf{n}_{K,\sigma}$ the outward normal unit vector.

3. $\mathcal{P}$ is a family of points of Ω indexed by $\mathcal{M}$, denoted by $\mathcal{P} = (\mathbf{x}_K)_{K\in\mathcal{M}}$, such that for all $K \in \mathcal{M}$, $\mathbf{x}_K \in K$; moreover K is assumed to be $\mathbf{x}_K$-star-shaped, which means that for all $\mathbf{x} \in K$, there holds $[\mathbf{x}_K, \mathbf{x}] \in K$. Denoting by $d_{K,\sigma}$ the Euclidean distance between $\mathbf{x}_K$ and the hyperplane containing σ, one assumes that $d_{K,\sigma} > 0$. We denote by $D_{K,\sigma}$ the cone of vertex $\mathbf{x}_K$ and basis σ.

Next we introduce some extra notations related to the mesh. The size of the discretization $\mathcal{D}$ is defined by

$$h_{\mathcal{D}} = \sup_{K\in\mathcal{M}} diam(K); \tag{11.1}$$

moreover we define

$$\theta_{\mathcal{D}} = \max\Big(\max_{\sigma\in\mathcal{E}_{int},\{K,L\}=\mathcal{M}_\sigma} \frac{d_{K,\sigma}}{d_{L,\sigma}}, \max_{K\in\mathcal{M}_\sigma,\sigma\in\mathcal{E}_K} \frac{h_K}{d_{K,\sigma}}\Big). \tag{11.2}$$

Imposing a uniform bound on $\theta_{\mathcal{D}}$ forces the mesh to be sufficiently regular. Next, we define several discrete spaces, which are going to be used in the sequel.

Definition 11.2 (The hybrid space $X_{\mathcal{D}}(\Omega)$). *Let $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ be a discretization of Ω. We define*

$$\begin{aligned} X_{\mathcal{D}} &= \{((v_K)_{K\in\mathcal{M}}, (v_\sigma)_{\sigma\in\mathcal{E}}), v_K \in \mathbb{R}, v_\sigma \in \mathbb{R}\}, \\ X_{\mathcal{D},0} &= \{v \in X_{\mathcal{D}} \text{ such that } (v_\sigma)_{\sigma\in\mathcal{E}_{ext}} = 0\}. \end{aligned} \tag{11.3}$$

The space X_D is equipped with the semi-norm $|\cdot|_{X_{\mathcal{D}}}$ defined by

$$|v|^2_{X_{\mathcal{D}}} = \sum_{K\in\mathcal{M}} |v|^2_{X_{\mathcal{D},K}}, \text{ where } |v|^2_{X_{\mathcal{D},K}} = \sum_{\sigma\in\mathcal{E}_K} \frac{m(\sigma)}{d_{K,\sigma}}(v_\sigma - v_K)^2 \text{ for all } v \in X_{\mathcal{D}}. \tag{11.4}$$

Note that $|\cdot|_{X_{\mathcal{D}}}$ is a norm on the space $X_{\mathcal{D},0}$.

Moreover, for each function $\psi = \psi(\mathbf{x})$ regular enough we define its projection $P_{\mathcal{D}}\psi \in X_D$ on the space $X_{\mathcal{D}}$ in following way

$$\begin{aligned} (P_{\mathcal{D}}\psi)_K &= \psi(\mathbf{x}_K) \quad \text{for all} \quad K \in \mathcal{M}, \\ (P_{\mathcal{D}}\psi)_\sigma &= \psi(\mathbf{x}_\sigma) \quad \text{for all} \quad \sigma \in \mathcal{E}. \end{aligned}$$

Definition 11.3 (The discrete flux space $\mathcal{Q}_{\mathcal{D}}(\Omega)$). *Let $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ be a discretization of Ω. We define*

$$\mathcal{Q}_{\mathcal{D}} = \{(q_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}, q_{K,\sigma} \in \mathbb{R}\}. \tag{11.5}$$

Next we introduce the time discretization.

Definition 11.4 (Time discretization). *We divide the time interval $(0, T)$ into N equal time steps of length $\delta t = T/N$, where δt is the uniform time step defined by $\delta t = t_n - t_{n-1}$.*

Taking into account the time discretization leads us to define of the following discrete spaces

$$X_{\mathcal{D},\delta t} = X_{\mathcal{D}}^N = \{(v^n)_{n\in\{1,\dots,N\}}, v^n \in X_{\mathcal{D}}\}$$

and

$$X_{\mathcal{D},\delta t,0} = X_{\mathcal{D},0}^N = \{(v^n)_{n\in\{1,\dots,N\}}, v^n \in X_{\mathcal{D},0}\};$$

moreover we define the following semi-norm on $X_{\mathcal{D},\delta t}$

$$|v|^2_{X_{\mathcal{D},\delta t}} = \sum_{n=1}^{N} \delta t |v|^2_{X_{\mathcal{D}}}. \tag{11.6}$$

11.2 The numerical scheme

11.2.1 The discrete problem

In this section we present the fully implicit finite volume scheme for the problem (10.5)-(10.9). Let us introduce the discrete saturation $\left((s_K^n)_{K\in\mathcal{M}}, (s_\sigma^n)_{\sigma\in\mathcal{E}}\right)_{n\in\{1,\dots,N\}} \in X_{\mathcal{D},\delta t}$ and the discrete global pressure $\left((p_K^n)_{K\in\mathcal{M}}, (p_\sigma^n)_{\sigma\in\mathcal{E}}\right)_{n\in\{1,\dots,N\}} \in X_{\mathcal{D},\delta t}$, which are the main discrete unknowns. Moreover let $\mathfrak{f}$ denote λ, ξ, γ or f we introduce the following notation $\mathfrak{f}_K^n = \mathfrak{f}(s_K^n)$ and $\mathfrak{f}_\sigma^n = \mathfrak{f}(s_\sigma^n)$ for all $K \in \mathcal{M}, \sigma \in \mathcal{E}$ and $n \in \{1,\dots,N\}$. Let $k_{i,K}^n$ denote the mean value of the source term $k_i(\mathbf{x},t)$ over a cell $K \times (t_{n-1}, t_n)$, i.e.

$$k_{i,K}^n = \frac{1}{m(K)\delta t}\int_{t_{n-1}}^{t_n}\int_K k_i(\mathbf{x},t)\, d\mathbf{x}dt \quad \text{with} \quad i \in \{w,n\}. \tag{11.7}$$

We denote the porous volume of the element K by $\omega(K)$,

$$\omega(K) = \int_K \omega(\mathbf{x})\, d\mathbf{x}.$$

Next, let $Q_{K,\sigma}^n$ be an approximation of the total flux through the interface σ

$$Q_{K,\sigma}^n \approx \frac{1}{\delta t}\int_{t_{n-1}}^{t_n}\int_\sigma \mathbf{K}(\lambda(s)\nabla p - \xi(s)\mathbf{g}) \cdot \mathbf{n}_{K,\sigma}\mathbf{d}\nu dt \tag{11.8}$$

and let $F_{K,\sigma}^n$ be an approximation of the non-wetting phase flux

$$F_{K,\sigma}^n \approx \frac{1}{\delta t}\int_{t_{n-1}}^{t_n}\int_\sigma (\mathbf{q}f(s) + \gamma(s)\mathbf{K}\mathbf{g} - \mathbf{K}\nabla\varphi(s)) \cdot \mathbf{n}_{K,\sigma}\mathbf{d}\nu dt. \tag{11.9}$$

The numerical fluxes $Q_{K,\sigma}^n$ and $F_{K,\sigma}^n$ have to be constructed as functions of the discrete unknowns. Using the notations (11.7) and (11.8) we discretize the equation (10.5) by

$$\sum_{\sigma\in\mathcal{E}_K} Q_{K,\sigma}^n = m(K)(k_{w,K}^n + k_{o,K}^n) \text{ for all } K \in \mathcal{M}. \tag{11.10}$$

We also prescribe the continuity of the fluxes

$$Q_{K,\sigma}^n + Q_{L,\sigma}^n = 0 \text{ for all } \sigma \in \mathcal{E}_{int} \text{ with } \{K,L\} = \mathcal{M}_\sigma. \tag{11.11}$$

Similarly, the equation (10.7) is discretized by

$$\omega(K)\frac{s_K^n - s_K^{n-1}}{\delta t} + \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^n = m(K)k_{o,K}^n \text{ for all } K \in \mathcal{M}, \tag{11.12}$$

and

$$F_{K,\sigma}^n + F_{L,\sigma}^n = 0 \text{ for all } \sigma \in \mathcal{E}_{int} \text{ with } \{K, L\} = \mathcal{M}_\sigma. \tag{11.13}$$

The discrete equations (11.16)-(11.13) have to be prescribed at each time step $n \in \{1, \dots, N\}$. We prescribe the initial and the boundary conditions for the numerical scheme by setting

$$s_K^0 = \frac{1}{m(K)} \int_K s_0(\mathbf{x})\, d\mathbf{x} \text{ for all } K \in \mathcal{M} \tag{11.14}$$

and

$$s_\sigma^n = p_\sigma^n = 0 \text{ for all } \sigma \in \mathcal{E}_{ext}. \tag{11.15}$$

Remark that opposite to the classical two-point flux approximation, the discrete fluxes $Q_{K,\sigma}^n$ and $F_{K,\sigma}^n$ (which still remain to be constructed) are not *a priori* continuous across the element's interfaces, so that the continuity is prescribed in the scheme by (11.11) and (11.13).

11.2.2 The discrete weak formulation

Following the ideas of [66] we write the scheme in the variational form.
For each $n \in \{1, \dots, N\}$ find $s^n \in X_{\mathcal{D},0}$ and $p^n \in X_{\mathcal{D},0}$ such that for all $v^n, w^n \in X_{\mathcal{D},0}$:

$$\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (v_K^n - v_\sigma^n) Q_{K,\sigma}^n = \sum_{K\in\mathcal{M}} m(K) v_K^n (k_{w,K}^n + k_{o,K}^n), \tag{11.16}$$

$$\sum_{K\in\mathcal{M}} \omega(K) w_K^n \frac{s_K^n - s_K^{n-1}}{\delta t} + \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (w_K^n - w_\sigma^n) F_{K,\sigma}^n = \sum_{K\in\mathcal{M}} m(K) w_K^n k_{o,K}^n, \tag{11.17}$$

$$s_K^0 = \frac{1}{m(K)} \int_K s_0(\mathbf{x})\, d\mathbf{x}. \tag{11.18}$$

In order to complete the scheme we have to define the numerical fluxes $Q_{K,\sigma}^n$ and $F_{K,\sigma}^n$. Let $\mathbf{K}_K$ denote the mean value of $\mathbf{K}(\mathbf{x})$ over a cell K,

$$\mathbf{K}_K = \frac{1}{m(K)} \int_K \mathbf{K}(\mathbf{x})\, d\mathbf{x} \tag{11.19}$$

and let

$$g_{K,\sigma} = m(\sigma) \mathbf{K}_K \mathbf{g} \cdot \mathbf{n}_{K,\sigma}. \tag{11.20}$$

Note that $g_{K,\sigma}$ satisfies

$$\sum_{\sigma\in\mathcal{E}_K} g_{K,\sigma} = 0 \text{ for all } K \in \mathcal{M}, \tag{11.21}$$

but not necessarily

$$g_{K,\sigma} + g_{L,\sigma} = 0 \text{ with } \{K, L\} = \mathcal{M}_\sigma.$$

The above equality remains true for the interfaces which are "interior" with respect to some sub-domain Ω_i, that is to say

$$g_{K,\sigma} + g_{L,\sigma} = 0 \text{ with } \{K, L\} = \mathcal{M}_\sigma \text{ for all } \sigma \notin \Gamma_{i,j}, \tag{11.22}$$

for any i, j. Next, we define $Q^n_{K,\sigma}$ and $F^n_{K,\sigma}$ by

$$Q^n_{K,\sigma} = \lambda^n_K \mathcal{F}_{K,\sigma}(p^n) + \mathcal{G}(\xi(\cdot)g_{K,\sigma}; s^n_K, s^n_\sigma), \tag{11.23}$$

$$F^n_{K,\sigma} = \mathcal{G}(Q^n_{K,\sigma} f(\cdot) + \gamma(\cdot)g_{K,\sigma}; s^n_K, s^n_\sigma) + \mathcal{F}_{K,\sigma}(\varphi(s^n)), \tag{11.24}$$

where $\lambda^n_K = \lambda^n_K$ for all $K \in \mathcal{M}$ and $n \in \{1, \ldots, N\}$, in general. The terms $\mathcal{F}_{K,\sigma}(\cdot)$ correspond to the diffusive fluxes which are discretized using the SUSHI scheme (Section 11.2.4). The terms $\mathcal{G}(\cdot)$ stand for the discretization of the convective fluxes, using the Godunov scheme (Section 11.2.3 below, see also [82, p. 3]).

11.2.3 The Godunov scheme and the convection term

Let $a, b \in \mathbb{R}$ and $f \in L^\infty(\mathbb{R})$ we define the Godunov flux by

$$\mathcal{G}(f; a, b) = \begin{cases} \min_{s\in[a,b]} f(s) & \text{if } a \le b, \\ \max_{s\in[b,a]} f(s) & \text{if } b \le a. \end{cases} \tag{11.25}$$

Since it is often useful to write the discrete flux in more explicit form we define

$$\mathcal{S}(f; a, b) = \begin{cases} \text{argmin}_{s\in[a,b]} f(s) & \text{if } a \le b, \\ \text{argmax}_{s\in[b,a]} f(s) & \text{if } b \le a, \end{cases}$$

and we introduce the following notations

$$\begin{aligned} \xi^n_{K,\sigma} &= \xi(\mathcal{S}(\xi(\cdot)g_{K,\sigma}; s^n_K, s^n_\sigma)), \\ f^n_{K,\sigma} &= f(\mathcal{S}(Q^n_{K,\sigma} f(\cdot) + \gamma(\cdot)g_{K,\sigma}; s^n_K, s^n_\sigma)), \\ \gamma^n_{K,\sigma} &= \gamma(\mathcal{S}(Q^n_{K,\sigma} f(\cdot) + \gamma(\cdot)g_{K,\sigma}; s^n_K, s^n_\sigma)). \end{aligned} \tag{11.26}$$

Using the notations (11.26) we can write the discrete fluxes in the form

$$Q^n_{K,\sigma} = \lambda^n_K \mathcal{F}_{K,\sigma}(p^n) + \xi^n_{K,\sigma} g_{K,\sigma} \tag{11.27}$$

and

$$F^n_{K,\sigma} = Q^n_{K,\sigma} f^n_{K,\sigma} + \gamma^n_{K,\sigma} g_{K,\sigma} + \mathcal{F}_{K,\sigma}(\varphi(s^n)). \tag{11.28}$$

11.2.4 The discrete gradient and the diffusion term

In this section we recall a construction of the discrete gradient and of the numerical flux $\mathcal{F}_{K,\sigma}(\cdot)$ proposed in [66]. Let $u \in X_\mathcal{D}$, for all $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$ we define

$$\nabla_{K,\sigma} u = \nabla_K u + R_{K,\sigma} u \cdot \mathbf{n}_{K,\sigma}, \tag{11.29}$$

where

$$\nabla_K u = \frac{1}{m(K)} \sum_{\sigma\in\mathcal{E}_K} m(\sigma)(u_\sigma - u_K)\mathbf{n}_{K,\sigma} \tag{11.30}$$

and

$$R_{K,\sigma} u = \frac{\sqrt{d}}{d_{K,\sigma}}(u_\sigma - u_K - \nabla_K u \cdot (\mathbf{x}_\sigma - \mathbf{x}_K)). \tag{11.31}$$

Note that the stabilizing term $R_{K,\sigma}$ is a second order error term, which vanishes for piecewise linear functions. We define the discrete gradient $\nabla_{\mathcal{D}} u$ as the piecewise constant function equal to $\nabla_{K,\sigma} u$ in the cone $D_{K,\sigma}$ with vertex $\mathbf{x}_K$ and basis σ

$$\nabla_{\mathcal{D}} u(\mathbf{x})|_{\mathbf{x}\in D_{K,\sigma}} = \nabla_{K,\sigma} u.$$

Let $u = (u^n)_{n\in\{1,\ldots,N\}} \in X_{\mathcal{D},\delta t}$, taking into account the time discretization, we define the discret gradient $\nabla_{\mathcal{D},\delta t} u(\mathbf{x},t)$ by

$$\nabla_{\mathcal{D},\delta t} u(\mathbf{x},t)|_{t\in(t_{n-1},t_n]} = \nabla_{\mathcal{D}} u^n(\mathbf{x}), \tag{11.32}$$

for all $\mathbf{x} \in \Omega$ and all $n \in \{1,\ldots,N\}$. For an arbitrary $u \in X_D$ the numerical flux $\mathcal{F}_{K,\sigma}(u)$ can be defined through the following discrete integration by parts formula

$$\sum_{\sigma\in\mathcal{E}_K} (v_K - v_\sigma)\mathcal{F}_{K,\sigma}(u) = \sum_{\sigma\in\mathcal{E}_K} m(D_{K,\sigma})\mathbf{K}_K \nabla_{K,\sigma} u \cdot \nabla_{K,\sigma} v \text{ for all } v \in X_{\mathcal{D}}, \tag{11.33}$$

which in particular implies that

$$\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (v_K - v_\sigma)\mathcal{F}_{K,\sigma}(u) = \int_\Omega \mathbf{K}\nabla_D u \cdot \nabla_{\mathcal{D}} v \, d\mathbf{x} \text{ for all } v \in X_{\mathcal{D}}. \tag{11.34}$$

The explicit form of $\mathcal{F}_{K,\sigma}$ can be obtained by setting $v_K - v_\sigma = 1$ and $v_K - v_{\sigma'} = 0$ for all $\sigma' \neq \sigma$. We refer to [66] for more details on construction of $\mathcal{F}_{K,\sigma}$ and its practical implementation. Next we state without proof three results from [66].

Lemma 11.1 (Strong consistency). *Let $\mathcal{D}$ be a discretization of Ω in sense of Definition 11.1, moreover let $\theta \geq \theta_{\mathcal{D}}$ be given. Then for all $\psi \in C^2(\overline{\Omega})$, there exist a positive constant C only depending on d, θ and φ such that*

$$|\nabla_K P_{\mathcal{D}}\psi - \nabla\psi(\mathbf{x})| \leq Ch \text{ for all } \mathbf{x} \in K \tag{11.35}$$

and also

$$\|\nabla_{\mathcal{D}} P_{\mathcal{D}}\psi - \nabla\psi\|_{(L^\infty(\Omega))^d} \leq Ch. \tag{11.36}$$

The following lemma, which is the slightly modified version of [66, Lemma 4.1] shows the equivalence between the semi-norm in $X_{\mathcal{D}}$ and the L^2-norm of the discrete gradient.

Lemma 11.2. *Let $\mathcal{D}$ be a discretization of Ω and let $\theta \geqslant \theta_{\mathcal{D}}$ be given. Then there exists $m > 0$ and $M > 0$ only depending on θ and d such that*

$$m|v|_{X_{\mathcal{D},K}} \leq \|\nabla_{\mathcal{D}} v\|_{L^2(K)} \leq M|v|_{X_{\mathcal{D},K}} \quad \text{for all } K \in \mathcal{M} \text{ for all } v \in X_{\mathcal{D}}.$$

Lemma 11.3 (Discrete Poincaré inequality). *There exists a positive constant C, independent of the mesh size $h_{\mathcal{D}}$ such that*

$$\|\Pi_{\mathcal{D},\delta t} u\|_{L^2(\Omega)} \leq C|u|_{X_{\mathcal{D}}} \text{ for all } u \in X_{\mathcal{D}}. \tag{11.37}$$

Proof : The result follows from Lemma 5.3 of [66]. □
The direct consequence of (11.33) and Lemma 11.2 is the lemma below.

Lemma 11.4. *Let u, v be an arbitrary couple from $X_{\mathcal{D}}$, then there exist two positive constants C_1, C_2 independent of the mesh size such that*

$$\left| \sum_{\sigma\in\mathcal{E}_K} (v_K - v_\sigma)\mathcal{F}_{K,\sigma}(u) \right| \le C_1 |u|_{X_{\mathcal{D},K}} |v|_{X_{\mathcal{D},K}} \tag{11.38}$$

and

$$\sum_{\sigma\in\mathcal{E}_K} (u_K - u_\sigma)\mathcal{F}_{K,\sigma}(u) \ge C_2 |u|^2_{X_{\mathcal{D},K}} \tag{11.39}$$

Lemma 11.4 implies the following useful technical result.

Lemma 11.5. *Let $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ be a discretization of Ω, let $q \in \mathcal{Q}_{\mathcal{D}}$ and $u \in X_{\mathcal{D}}$. Then*

$$\left| \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} q_{K,\sigma}\mathcal{F}_{K,\sigma}(u) \right| \le C_1 |u|_{X_{\mathcal{D}}} \left(\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \frac{m(\sigma)}{d_{K,\sigma}} q^2_{K,\sigma} \right)^{\frac{1}{2}}. \tag{11.40}$$

Proof : Let K be an element of $\mathcal{M}$, setting $v_K = 0$ and $v_\sigma = f_{K,\sigma}$ in (11.38) we obtain

$$\left| \sum_{\sigma\in\mathcal{E}_K} q_{K,\sigma}\mathcal{F}_{K,\sigma}(u) \right| \le |u|_{X_{\mathcal{D},K}} \left(\sum_{\sigma\in\mathcal{E}_K} \frac{m(\sigma)}{d_{K,\sigma}} q^2_{K,\sigma} \right)^{\frac{1}{2}}.$$

Proceeding the same way for all $K \in \mathcal{M}$ we get

$$\begin{aligned} \left| \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} q_{K,\sigma}\mathcal{F}_{K,\sigma}(u) \right| &\le \sum_{K\in\mathcal{M}} \left| \sum_{\sigma\in\mathcal{E}_K} q_{K,\sigma}\mathcal{F}_{K,\sigma}(u) \right| \\ &\le C_1 \sum_{K\in\mathcal{M}} |u|_{X_{\mathcal{D},K}} \left(\sum_{\sigma\in\mathcal{E}_K} \frac{m(\sigma)}{d_{K,\sigma}} q^2_{K,\sigma} \right)^{\frac{1}{2}}. \end{aligned}$$

We use the Cauchy-Schwarz inequality to complete the proof. □

Lemma 11.6. *Let $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ be a discretization of Ω, let $v \in X_{\mathcal{D}}$ then there exists a positive constant C independent of the mesh size such that for all $n \in \{1 \dots N\}$*

$$\left| \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (v_K - v_\sigma)\mathcal{G}(Q^n_{K,\sigma} f(\cdot) + \gamma(\cdot) g_{K,\sigma}; s^n_K, s^n_\sigma) \right| \le C |v|_{X_{\mathcal{D}}} (|p|_{X_{\mathcal{D}}} + 1).$$

Proof : In view of (11.24), (11.27) and (11.28) we have that

$$\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (v_K - v_\sigma)\mathcal{G}(Q^n_{K,\sigma} f(\cdot) + \gamma(\cdot) g_{K,\sigma}; s^n_K, s^n_\sigma) = T_1 + T_2 + T_3,$$

where

$$\begin{aligned} T_1 &= \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (v_K - v_\sigma)\gamma^n_{K,\sigma} g_{K,\sigma} \\ T_2 &= \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (v_K - v_\sigma)\xi^n_{K,\sigma} f^n_{K,\sigma} g_{K,\sigma} \\ T_3 &= \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (v_K - v_\sigma)\lambda^n_K \mathcal{F}_{K,\sigma}(p^n) f^n_{K,\sigma} \end{aligned}$$

In view of (11.20)

$$T_1 = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma)(v_K - v_\sigma)\gamma^n_{K,\sigma}\mathbf{K}_K\mathbf{g}\cdot\mathbf{n}_{K,\sigma}$$

Remark that for all $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$ one has $m(\sigma)d_{K,\sigma} = m(D_{K,\sigma})d$, where d is the space dimension. Using Cauchy-Schwarz inequality

$$\begin{aligned}(T_1)^2 &\leq \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \frac{m(\sigma)}{d_{K,\sigma}}(v_K - v_\sigma)^2 \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}(\gamma^n_{K,\sigma})^2|\mathbf{g}|^2\overline{\mathrm{K}}^2 \\ &\leq \overline{\mathrm{K}}^{\ 2}|\mathbf{g}|^2 m(\Omega)d\|\gamma\|^2_{L^\infty((0,1))}|v|^2_{X_\mathcal{D}}.\end{aligned} \tag{11.41}$$

In the same way

$$(T_2)^2 \leq \overline{\mathrm{K}}^{\ 2}|\mathbf{g}|^2 m(\Omega)d\|\xi\|^2_{L^\infty((0,1))}\|f\|^2_{L^\infty((0,1))}|v|^2_{X_\mathcal{D}}. \tag{11.42}$$

Applying Lemma 11.5 to the term T_3 we obtain

$$\begin{aligned}(T_3)^2 &\leq |p|^2_{X_\mathcal{D}} \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \frac{m(\sigma)}{d_{K,\sigma}}\left(\lambda^n_K(v_K - v_\sigma)f^n_{K,\sigma}\right)^2 \\ &\leq \|\lambda\|^2_{L^\infty((0,1))}\|f\|^2_{L^\infty((0,1))}|p|^2_{X_\mathcal{D}}|v|^2_{X_\mathcal{D}}.\end{aligned} \tag{11.43}$$

Gathering (11.41)-(11.43) we complete the proof. □

Finally we present the technical lemma which is used in the proof of the a priori estimates.

Lemma 11.7. *Let $\varphi(s)$ satisfying the hypothesis $(\mathcal{H}_1)$ and let the function Φ be defined by*

$$\Phi(s) = \int_0^s \varphi(\tau)d\tau. \tag{11.44}$$

Then,

$$\frac{1}{2L_\varphi}(\varphi(s))^2 \leq \Phi(s) \leq \frac{L_\varphi}{2}s^2. \tag{11.45}$$

Proof : The function φ is invertible, then setting $\xi = \varphi(\tau)$ in (11.44) gives

$$\Phi(s) = \int_0^{\varphi(s)} \frac{\xi d\xi}{\varphi'(\varphi^{-1}(\xi))} \geq \frac{1}{L_\varphi}\int_0^{\varphi(s)} \xi d\xi = \frac{1}{2L_\varphi}(\varphi(s))^2.$$

On the other hand

$$\Phi(s) \leq L_\varphi \int_0^s \tau d\tau = \frac{L_\varphi}{2}s^2.$$

□

12 A priory estimates and existence of discrete solution

12.1 A priori estimates

Definition 12.1 (Approximate solution)**.** *Let $\mathcal{D}$ be a discretization of Ω, $N \in \mathbb{N}^*$ and $\delta t = T/N > 0$. We say that the sequence $(s_{\mathcal{D},\delta t}, p_{\mathcal{D},\delta t}) = (s^n, p^n)_{n\in\{1,\ldots,N\}} \in (X_{\mathcal{D},\delta t,0})$ is* *an approximate solution of the problem* (10.5)-(10.9) *if for all* $n \in \{1, \dots, N\}$, (s^n, p^n) *satisfies* (11.16)-(11.18). *We also denote by* $s_{\mathcal{D},\delta t}$ *and* $p_{\mathcal{D},\delta t}$ *the function pair defined by*

$$s_{\mathcal{D},\delta t}(\mathbf{x},0) = s_K^0 \textit{ for all } \mathbf{x} \in K, \tag{12.1}$$
$$s_{\mathcal{D},\delta t}(\mathbf{x},t) = s_K^n \textit{ for all } (\mathbf{x},t) \in K \times (t_{n-1}, t_n], \tag{12.2}$$
$$s_{\mathcal{D},\delta t}(\mathbf{x},t) = s_\sigma^n \textit{ for all } (\mathbf{x},t) \in \sigma \times (t_{n-1}, t_n] \tag{12.3}$$

and

$$p_{\mathcal{D},\delta t}(\mathbf{x},t) = p_K^n \textit{ for all } (\mathbf{x},t) \in K \times (t_{n-1}, t_n], \tag{12.4}$$
$$p_{\mathcal{D},\delta t}(\mathbf{x},t) = p_\sigma^n \textit{ for all } (\mathbf{x},t) \in \sigma \times (t_{n-1}, t_n]. \tag{12.5}$$

Remark 12.1. *In general the capillary diffusion* $\varphi(s_{\mathcal{D},\delta t})$ *is more regular that* $s_{\mathcal{D},\delta t}$ *and it is easer to work with the function* $\varphi_{\mathcal{D},\delta t}$*, thus the element* $\varphi_{\mathcal{D},\delta t} = \varphi(s_{\mathcal{D},\delta t})$ *of* $X_{\mathcal{D},\delta t,0}$ *can be considered as a set of primary discret unknowns. We also define the function* $\varphi_{\mathcal{D},\delta t}(\mathbf{x},t)$

$$\varphi_{\mathcal{D},\delta t}(\mathbf{x},t) = \varphi(s_{\mathcal{D},\delta t}(\mathbf{x},t)) \textit{ for all } (\mathbf{x},t) \in Q_T. \tag{12.6}$$

Theorem 12.1 (A priori estimate)**.** *Let* $(s_{\mathcal{D},\delta t}, p_{\mathcal{D},\delta t})$ *be a solution of the discrete problem* (11.16)-(11.18). *Then there exists a positive constant* C *independent of* $h_{\mathcal{D}}$ *and* δt *such that*

$$\|\nabla_{\mathcal{D},\delta t} p_{\mathcal{D},\delta t}\|_{L^\infty(0,T;L^2(\Omega))} + \|\varphi_{\mathcal{D},\delta t}\|_{L^\infty(0,T;L^2(\Omega))} + \|\nabla_{\mathcal{D},\delta t}\varphi_{\mathcal{D},\delta t}\|_{L^2(Q_T)} \le C. \tag{12.7}$$

Proof : **Pressure equation.** In order to obtain the estimate on the first term in (12.7) we use p^n as a test element in the pressure equation (11.16), which implies

$$\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (p_K^n - p_\sigma^n)(\lambda_K^n \mathcal{F}_{K,\sigma}(p^n) + \xi_{K,\sigma}^n g_{K,\sigma}) = \sum_{K\in\mathcal{M}} m(K) p_K^n (k_{w,K}^n + k_{o,K}^n). \tag{12.8}$$

Let us first estimate the term $T_\xi = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (p_K^n - p_\sigma^n)\xi_{K,\sigma}^n g_{K,\sigma}$. In view of (11.20) we deduce from Cauchy-Schwarz inequality that

$$\begin{aligned}(T_\xi)^2 &= \Big(\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma)(p_K^n - p_\sigma^n)\mathbf{n}_{K,\sigma} \cdot \xi_{K,\sigma}^n \mathbf{K}_K \mathbf{g}\Big)^2 \\ &\le \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \frac{m(\sigma)}{d_{K,\sigma}} (p_K^n - p_\sigma^n)^2 \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma) d_{K,\sigma} (|\mathbf{K}_K \mathbf{g}| \xi_{K,\sigma}^n)^2.\end{aligned}$$

Then, in view of the assumptions $(\mathcal{H}_2)$, $(\mathcal{H}_3)$ and $(\mathcal{H}_{4b})$, and also thanks to (11.4) and Lemma 11.2 we deduce that

$$|T_\xi| \le d m(\Omega) \overline{K} |\mathbf{g}| \|\xi\|_{L^\infty((0,1))} \|\nabla_{\mathcal{D},\delta t} p(\cdot, t_n)\|_{L^2(\Omega)}. \tag{12.9}$$

In view of (11.7) and (12.4) the right-hand side of (12.8) can be written as

$$\sum_{K\in\mathcal{M}} m(K) p_K^n (k_{w,K}^n + k_{o,K}^n) = \sum_{K\in\mathcal{M}} \frac{1}{\delta t} \int_{t_{n-1}}^{t_n} \int_\Omega p_{\mathcal{D},\delta t}(\mathbf{x},t)(k_w(\mathbf{x},t) + k_o(\mathbf{x},t)) \, d\mathbf{x} dt$$

Applying Cauchy-Schwarz inequality and the discrete Poincaré inequality (Lemma 11.3) we obtain

$$\begin{aligned}\sum_{K\in\mathcal{M}} m(K)p_K^n(k_{w,K}^n + k_{o,K}^n) &\le \|p_{\mathcal{D},\delta t}(\cdot,t_n)\|_{L^2(\Omega)}\|k_w + k_o\|_{L^\infty(0,T;L^2(\Omega))}\\ &\le C\|\nabla_{\mathcal{D},\delta t}p_{\mathcal{D},\delta t}(\cdot,t_n)\|_{L^2(\Omega)}\|k_w + k_o\|_{L^\infty(0,T;L^2(\Omega))}.\end{aligned}$$

so that in view of $(\mathcal{H}_6)$,

$$\sum_{K\in\mathcal{M}} m(K)p_K^n(k_{w,K}^n + k_{o,K}^n) \le C\|\nabla_{\mathcal{D},\delta t}p_{\mathcal{D},\delta t}(\cdot,t_n)\|_{L^2(\Omega)}. \tag{12.10}$$

Gathering (11.39), (12.8), (12.9), (12.10) we obtain

$$\underline{\lambda}\|\nabla_{\mathcal{D},\delta t}p_{\mathcal{D},\delta t}(\cdot,t_n)\|^2_{L^2(\Omega)} \le \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}(p_K^n - p_\sigma^n)\lambda_K^n\mathcal{F}_{K,\sigma}(p^n) \le C\|\nabla_{\mathcal{D},\delta t}p_{\mathcal{D},\delta t}(\cdot,t_n)\|_{L^2(\Omega)},$$

with some positive C; the proof of the estimate on the first term of (12.7) is complete. The estimate on the other terms of (12.7) can be obtained by setting $w^n = \varphi(s^n)$ in the saturation equation (11.17) and summing over $n \in \{1,\ldots,m\}$, which yields

$$\begin{aligned}\sum_{n=1}^m\sum_{K\in\mathcal{M}}\omega(K)\varphi_K^n(s_K^n - s_K^{n-1}) + \delta t\sum_{n=1}^m\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}(\varphi_K^n - \varphi_\sigma^n)F_{K,\sigma}(\varphi^n)\\ = \sum_{n=1}^m\sum_{K\in\mathcal{M}}\delta t m(K)\varphi_K^n k_{o,K}^n.\end{aligned}$$

We define the terms

$$T_t = \sum_{n=1}^m\sum_{K\in\mathcal{M}}\omega(K)\varphi_K^n(s_K^n - s_K^{n-1}),$$

$$T_C = \sum_{n=1}^m\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\delta t(\varphi_K^n - \varphi_\sigma^n)\mathcal{G}(Q_{K,\sigma}^n f(\cdot) + \gamma(\cdot)g_{K,\sigma}; s_K^n, s_\sigma^n).$$

and

$$T_D = \sum_{n=1}^m\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\delta t(\varphi_K^n - \varphi_\sigma^n)\mathcal{F}_{K,\sigma}(\varphi(s^n)).$$

Accumulation term. Remark that using the notations of Lemma 11.7

$$\Phi(s_K^n) - \Phi(s_K^{n-1}) = \varphi_K^n(s_K^n - s_K^{n-1}) + \int_{s_K^{n-1}}^{s_K^n}(\varphi(\tau) - \varphi_K^n)d\tau. \tag{12.11}$$

Since the function φ is increasing, the second term in the right-hand side of (12.11) is negative, which implies that

$$T_t \ge \sum_{n=1}^m\sum_{K\in\mathcal{M}}\omega(K)(\Phi(s_K^n) - \Phi(s_K^{n-1})) = \sum_{K\in\mathcal{M}}\omega(K)\Phi(s_K^m) - \sum_{K\in\mathcal{M}}\omega(K)\Phi(s_K^0)).$$

It follows from (11.45) and the assumption $(\mathcal{H}_5)$ that

$$\begin{aligned}T_t &\ge \frac{1}{2L_\varphi}\sum_{K\in\mathcal{M}}\omega(K)(\varphi_K^m)^2 - \frac{L_\varphi}{2}\sum_{K\in\mathcal{M}}\omega(K)(s_K^0)^2\\ &\ge \frac{\underline{\omega}}{2L_\varphi}\|\varphi_{\mathcal{D},\delta t}(\cdot,t_m)\|^2_{L^2(\Omega)} - \frac{\overline{\omega}L_\varphi}{2}\|s_{\mathcal{D},\delta t}(\cdot,0)\|^2_{L^2(\Omega)}.\end{aligned}$$

Convection term. In this subsection we use the simplified notation

$$\mathcal{G}^n_{K,\sigma}(a,b) = \mathcal{G}\left(Q^n_{K,\sigma} f(\cdot) + \gamma(\cdot) g_{K,\sigma}; a, b\right) \text{ for all } a,b \in \mathbb{R}. \tag{12.12}$$

For all $K \in \mathcal{M}, \sigma \in \mathcal{E}_K$ and $s \in \mathbb{R}$ we define the function

$$G^n_{K,\sigma}(s) = \int_0^s \mathcal{G}^n_{K,\sigma}(\tau,\tau)\varphi'(\tau)d\tau, \tag{12.13}$$

which is such that

$$\begin{aligned} G^n_{K,\sigma}(s^n_K) - G^n_{K,\sigma}(s^n_\sigma) &= (\varphi_K - \varphi_\sigma)\mathcal{G}^n_{K,\sigma}(s_K, s_\sigma) \\ &+ \int_{s_\sigma}^{s_K} \left(\mathcal{G}^n_{K,\sigma}(\tau,\tau) - \mathcal{G}^n_{K,\sigma}(s_K,s_\sigma)\right)\varphi'(\tau)d\tau. \end{aligned} \tag{12.14}$$

In view of (11.25) the second term on the right-hand side of (12.14) is negative, which implies the following estimate

$$T_C \geq \sum_{n=1}^m \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \delta t \left(G^n_{K,\sigma}(s^n_K) - G^n_{K,\sigma}(s^n_\sigma)\right).$$

In view of (12.12), (12.13) and (11.25) we have that

$$G^n_{K,\sigma}(s^n_K) - G^n_{K,\sigma}(s^n_\sigma) = Q^n_{K,\sigma}\int_{s^n_\sigma}^{s^n_K} f(\tau)\varphi'(\tau)d\tau + g_{K,\sigma}\int_{s^n_\sigma}^{s^n_K} \gamma(\tau)\varphi'(\tau)d\tau.$$

It follows from (11.16) and the hypotheses $(\mathcal{H}_1), (\mathcal{H}_{2c})$ that

$$\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} Q^n_{K,\sigma}\int_{s^n_\sigma}^{s^n_K} f(\tau)\varphi'(\tau)d\tau = \sum_{K\in\mathcal{M}} m(K)(k^n_{w,K} + k^n_{o,K})\int_0^{s^n_K} f(\tau)\varphi'(\tau)d\tau \geq 0,$$

where we have set

$$v^n_K = \int_0^{s^n_K} f(\tau)\varphi'(\tau)d\tau \quad \text{and} \quad v^n_\sigma = \int_0^{s^n_\sigma} f(\tau)\varphi'(\tau)d\tau.$$

Therefore we have that

$$\begin{aligned} T_C &\geq \sum_{n=1}^N \delta t \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} g_{K,\sigma}\int_{s^n_\sigma}^{s^n_K} \gamma(\tau)\varphi'(\tau)d\tau \\ &\geq \sum_{n=1}^N \delta t \sum_{K\in\mathcal{M}} \int_0^{s^n_K} \gamma(\tau)\varphi'(\tau)d\tau \sum_{\sigma\in\mathcal{E}_K} g_{K,\sigma} \quad - \sum_{n=1}^N \delta t \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} g_{K,\sigma}\int_0^{s^n_\sigma} \gamma(\tau)\varphi'(\tau)d\tau \end{aligned}$$

In view of (11.21), (11.22) and using the homogeneous Dirichlet boundary condition we obtain

$$T_C \geq -\sum_{n=1}^N \delta t \sum_{K\in\mathcal{M}} \sum_{0\leq i\leq j\leq J} \sum_{\sigma\in\mathcal{E}_K\cap\Gamma_{i,j}} g_{K,\sigma}\int_0^{s_\sigma} \gamma(\tau)\varphi'(\tau)d\tau.$$

Therefore

$$T_C \geq -2T \sum_{0\leq i\leq j\leq J} m(\Gamma_{i,j})\overline{\mathrm{K}}|\mathbf{g}|L_\varphi\|\gamma\|_{L^1((0,1))}. \tag{12.15}$$

Diffusion and source terms. It follows from the equality (11.33)

$$T_D \geq \underline{\mathrm{K}} \|\nabla_{\mathcal{D},\delta t}\varphi_{\mathcal{D},\delta t}\|^2_{L^2(Q_T)}.$$

In view of the estimates on T_t and T_D we deduce that

$$\begin{aligned}\frac{\underline{\omega}}{2L_\varphi}\|\varphi_{\mathcal{D},\delta t}\|^2_{L^\infty(0,T;L^2(\Omega))} &+ \underline{\mathrm{K}}\|\nabla_{\mathcal{D},\delta t}\varphi_{\mathcal{D},\delta t}\|^2_{L^2(Q_T)} \\ &\leq \frac{\overline{\omega}L_\varphi}{2}\|s_{\mathcal{D},\delta t}(\cdot,0)\|^2_{L^2(\Omega)} + \sum_{n=1}^{N}\sum_{K\in\mathcal{M}} \delta t m(K)\varphi_K^n k_{o,K}^n - T_C. \end{aligned} \tag{12.16}$$

Applying Cauchy-Schwarz and Young's inequality to the last term in (12.16) leads to

$$\begin{aligned}\frac{\underline{\omega}}{2L_\varphi}\|\varphi_{\mathcal{D},\delta t}\|^2_{L^\infty(0,T;L^2(\Omega))} &+ \underline{\mathrm{K}}\|\nabla_{\mathcal{D},\delta t}\varphi_{\mathcal{D},\delta t}\|^2_{L^2(Q_T)} \\ &\leq \frac{\varepsilon}{2}\|\varphi_{\mathcal{D},\delta t}\|^2_{L^2(Q_T)} + \frac{1}{2\varepsilon}\|k_o\|^2_{L^2(Q_T)} + \frac{\overline{\omega}L_\varphi}{2}\|s_{\mathcal{D},\delta t}(\cdot,0)\|^2_{L^2(\Omega)} 0 T_C \\ &\leq \frac{\varepsilon T}{2}\|\varphi_{\mathcal{D},\delta t}\|^2_{L^\infty(0,T;L^2(\Omega))} + \frac{1}{2\varepsilon}\|k_o\|^2_{L^2(Q_T)} + \frac{\overline{\omega}L_\varphi}{2}\|s_{\mathcal{D},\delta t}(\cdot,0)\|^2_{L^2(\Omega)} 0 T_C.\end{aligned}$$

Setting $\varepsilon = \underline{\omega}/(2TL_\varphi)$ we complete the proof. □

Remark 12.2. *Remark that the partitioning of Ω in to the set of homogeneous media $(\Omega_i)_{i\in\{1,\ldots,I\}}$ serves to control the term T_C in* (12.15)*. Therefore if the gravity effects are neglected the assumption* $(\mathcal{H}_{4b})$ *can be weakened.*

12.2 Existence of a discrete solution

In this subsection we shows that there exists a sequence $((p^n, s^n))_{n\in\{1,\ldots,N\}} \in X^2_{\mathcal{D},\delta t,0}$ satisfying the variational discrete problem. The existence result makes use of the topological degree argumentation [[55], Theorem 3.1].

Theorem 12.2. *The problem* (11.16)-(11.18) *has at least one solution.*

Proof : In view of Remark 12.1 the discrete problem (11.16)-(11.18) can be written as a system of $2N(card(\mathcal{M})+card(\mathcal{E}))$ nonlinear equations $H_{\mathcal{D},\delta t}(\varphi, p) = 0$. For $\nu \in [0,1]$ we define

- $f^\nu(s) = \nu f(s)$,
- $\gamma^\nu(s) = \nu\gamma(s)$,
- $\xi^\nu(s) = \nu\xi(s)$,
- $\lambda^\nu(s) = \nu\lambda(s) + (1-\nu)\underline{\lambda}$

and we consider the following extended problem

$$-\nabla\cdot\mathbf{K}\left(\lambda(s^\nu)\nabla p^\nu - \xi^\nu(s^\nu)\mathbf{g}\right) = k_w + k_o \ \text{ in } \ \Omega\times(0,T), \tag{12.17}$$

$$\mathbf{q}^\nu = -\mathbf{K}\left(\lambda^\nu(s^\nu)\nabla p^\nu - \xi^\nu(s^\nu)\mathbf{g}\right) \ \text{ in } \ \Omega\times(0,T), \tag{12.18}$$

$$\omega\frac{\partial s^\nu}{\partial t} + \nabla\cdot\left(\mathbf{u}^\nu f^\nu(s^\nu) + \gamma^\nu(s^\nu)\mathbf{K}\mathbf{g}\right) - \nabla\cdot\left(\mathbf{K}\nabla\varphi(s^\nu)\right) = k_o \ \text{ in } \ \Omega\times(0,T), \tag{12.19}$$

together with initial and boundary conditions (10.8), (10.9). We denote by $H^{\nu}_{\mathcal{D},\delta t}(\phi,p)=0$ the corresponding discrete problem and $(\varphi^{\nu}_{\mathcal{D},\delta t},p^{\nu}_{\mathcal{D},\delta t})$ its solution, which satisfies the a priori estimates (12.7) instead of $(\varphi_{\mathcal{D},\delta t},p_{\mathcal{D},\delta t})$. Then, there exists $R>0$ such that

$$(\varphi^{\nu}_{\mathcal{D},\delta t},p^{\nu}_{\mathcal{D},\delta t})\in B_R=\{(u,v)\in X^2_{\mathcal{D},\delta t,0},\|\nabla_{\mathcal{D},\delta t}u\|^2_{L^2(Q_T)}+\|\nabla_{\mathcal{D},\delta t}v\|^2_{L^2(Q_T)}<R\}$$

for all $\nu\in[0,1]$. Therefore for any $\nu\in[0,1]$, $(\varphi^{\nu}_{\mathcal{D},\delta t},p^{\nu}_{\mathcal{D},\delta t})$ does not belongs to ∂B_R, which implies that the topological degree of $H^{\nu}_{\mathcal{D},\delta t}(\phi,p)$ with respect to B_R and right hand side 0 is constant. Remark that for $\nu=0$ the system (12.17)-(12.19), becomes uncoupled

$$-\nabla\cdot\mathbf{K}\underline{\lambda}\nabla p^0=k_w+k_o \quad\text{in}\quad \Omega\times(0,T), \tag{12.20}$$

$$\omega\frac{\partial s^0}{\partial t}-\nabla\cdot(\mathbf{K}\nabla\varphi(s^0))=k_o \quad\text{in}\quad \Omega\times(0,T). \tag{12.21}$$

We first remark that since the equation (12.20) does not contain s. It follows from [66] that the discrete problem corresponding to (12.20) and (10.9) admits the unique solution $p^0_{\mathcal{D},\delta t}$; moreover the existence and uniqueness of $\varphi^0_{\mathcal{D},\delta t}$ satisfying the discrete version of the problem (12.21), (10.8) and (10.9) has been shown in [10] where a family of schemes, containing in particular the SUSHI scheme was applied to a scalar parabolic degenerate convection-reaction-diffusion equation. Consequently there exists the unique solution $(p^0_{\mathcal{D},\delta t},\varphi^0_{\mathcal{D},\delta t})$ to the problem $H^0_{\mathcal{D},\delta t}(\phi,p)$, and the corresponding topological degree is equal to ±1. From the homotopy invariance of the topological degree we deduce that here exists at least one solution to $H^1_{\mathcal{D},\delta t}(\phi,p)=0$. □

13 Estimates on space and time translates

13.1 Estimates on space translates

We state here an estimate on space translates, which is a consequence of the a priori estimate (given by Theorem 12.1), Lemma 11.2 and [10, Lemma 3.4, Lemma 6.5] (see also the basic article [66]).

Lemma 13.1. *Let $\mathcal{D}$ be a discretization of Ω and let $\delta t=T/N>0$ with some $N\in\mathbb{N}^*$; let $(s_{\mathcal{D},\delta t},p_{\mathcal{D},\delta t})\in X^2_{\mathcal{D},\delta t,0}$ be a solution of the discrete problem* (11.16)-(11.18). *We recall that the function $\varphi_{\mathcal{D},\delta t}$ is defined by $\varphi_{\mathcal{D},\delta t}(\mathbf{x},t)=\varphi(s_{\mathcal{D},\delta t}(\mathbf{x},t))$. Let also $\theta\geq\theta_{\mathcal{D}}$ be given, then there exist $C>0$ and $\rho>0$, which depend on θ, but does not depend on mesh and time step size, such that*

$$\|\varphi_{\mathcal{D},\delta t}(\cdot+\mathbf{y},\cdot)-\varphi_{\mathcal{D},\delta t}\|_{L^2(\mathbb{R}^d\times(0,T))}\leq C|\mathbf{y}|^{\rho},$$

taking $\varphi_{\mathcal{D},\delta t}=0$ outside of Q_T.

13.2 Estimates on time translates

To begin with we state without proof two technical lemmas which will be useful for proving the estimate on time translates.

Lemma 13.2. *Let $T > 0$, $\tau \in (0,T)$, $N \in \mathbb{N}^*$, $\delta t = T/N$ be given and $(a^n)_{n\in\mathbb{N}^*}$ be a family of non negative real values. Let $\lceil s \rceil$ denotes the smallest integer larger or equal to s. Then*

$$\int_0^{T-\tau} \sum_{\lceil t/\delta t\rceil+1\le n\le\lceil(t+\tau)/\delta t\rceil} a^n dt \le \tau \sum_{n=1}^{N} a^n.$$

Lemma 13.3. *Let $T > 0$, $\tau \in (0,T)$, $N \in \mathbb{N}^*$, $\delta t = T/N$, $\zeta \in [0,\tau]$ be given and $(a^n)_{n\in\mathbb{N}^*}$ be a family of nonnegative real values. Let $\lceil s \rceil$ denotes the smallest integer larger or equal to s. Then*

$$\int_0^{T-\tau} \sum_{\lceil t/\delta t\rceil+1\le n\le\lceil(t+\tau)/\delta t\rceil} a^{\lceil(t+\zeta)/\delta t\rceil} dt \le \tau \sum_{n=1}^{N} a^n.$$

Next we prove the estimate on differences of time translates.

Lemma 13.4. *Let $\mathcal{D}$ be a discretization of Ω and let $\delta t = T/N > 0$ with some $N \in \mathbb{N}^*$; let $(s_{\mathcal{D},\delta t}, p_{\mathcal{D},\delta t}) \in X^2_{\mathcal{D},\delta t,0}$ be a solution of the discrete problem (11.16)-(11.18). Let also $\theta \ge \theta_{\mathcal{D}}$ be given, then there exists a positive constant C only depending on θ such that*

$$\int_{\Omega\times(0,T-\tau)} (\varphi_{\mathcal{D},\delta t}(\mathbf{x}, t+\tau) - \varphi_{\mathcal{D},\delta t}(\mathbf{x}, t))^2 \, d\mathbf{x}dt \le C\tau \tag{13.1}$$

for all $\tau \in (0,T)$.

Proof : The discrete saturation equation (11.17) yields

$$\begin{aligned}\sum_{K\in\mathcal{M}} \omega(K) w_K (s_K^n - s_K^{n-1}) = \\ &- \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (w_K - w_\sigma)\mathcal{G}(Q^n_{K,\sigma} f(\cdot) + \gamma(\cdot) g_{K,\sigma}; s_K^n, s_\sigma^n) \\ &- \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (w_K - w_\sigma)\mathcal{F}_{K,\sigma}(\varphi(s^n)) + \delta t \sum_{K\in\mathcal{M}} m(K) w_K k^n_{o,K}.\end{aligned} \tag{13.2}$$

for all $w \in X_{\mathcal{D},0}$. In view of Lemma 11.6 and Lemma 11.4 one has

$$\begin{aligned}&|\sum_{K\in\mathcal{M}} \omega(K) w_K (s_K^n - s_K^{n-1})| \\ &\qquad \le C\delta t |w|_{X_{\mathcal{D}}} (|p^n|_{X_{\mathcal{D}}} + |\varphi^n|_{X_{\mathcal{D}}} + 1) + \delta t \sum_{K\in\mathcal{M}} m(K) w_K k^n_{o,K}.\end{aligned} \tag{13.3}$$

Next, applying the Cauchy-Schwarz and discrete Poincaré inequality (cf. inequality (11.37)) to the last term of (13.3)

$$\begin{aligned}&|\sum_{K\in\mathcal{M}} \omega(K) w_K (s_K^n - s_K^{n-1})| \\ &\qquad \le C\delta t |w|_{X_{\mathcal{D}}} (|p^n|_{X_{\mathcal{D}}} + |\varphi^n|_{X_{\mathcal{D}}} + \frac{1}{\delta t}\|k_o\|_{L^2(\Omega)\times(t_{n-1},t_n)} + 1).\end{aligned} \tag{13.4}$$

In view of hypothesis $(\mathcal{H}_1)$ we obtain

$$\begin{aligned}
&\frac{1}{L_\varphi}\int_0^{T-\tau}\int_\Omega (\varphi_{\mathcal{D},\delta t}(x,t+\tau)-\varphi_{\mathcal{D},\delta t}(x,t))^2 \; d\mathbf{x}dt \\
&\quad = \frac{1}{L_\varphi}\int_0^{T-\tau}\sum_{K\in\mathcal{M}}\omega(K)\left(\varphi_K^{\lceil (t+\tau)/\delta t\rceil}-\varphi_K^{\lceil t/\delta t\rceil}\right)^2 dt \\
&\quad \le \int_0^{T-\tau}\sum_{K\in\mathcal{M}}\omega(K)\left(\varphi_K^{\lceil (t+\tau)/\delta t\rceil}-\varphi_K^{\lceil t/\delta t\rceil}\right)\left(s_K^{\lceil (t+\tau)/\delta t\rceil}-s_K^{\lceil t/\delta t\rceil}\right) dt \\
&\quad = \int_0^{T-\tau}\sum_{\lceil t/\delta t\rceil+1\le n\le\lceil (t+\tau)/\delta t\rceil}\;\sum_{K\in\mathcal{M}}\omega(K)\left(\varphi_K^{\lceil (t+\tau)/\delta t\rceil}-\varphi_K^{\lceil t/\delta t\rceil}\right)\left(s_K^n-s_K^{n-1}\right) dt.
\end{aligned}$$

For a given δt and for all real t and τ we define the following set

$$n(t,\tau)=\{n\in\mathbb{N},\ \lceil t/\delta t\rceil+1\le n\le\lceil (t+\tau)/\delta t\rceil\},$$

which can be empty. Then, in view of (13.4), Lemma 13.2 and Lemma 13.3 we obtain

$$\begin{aligned}
&\frac{1}{L_\varphi}\int_0^{T-\tau}\int_\Omega (\varphi_{\mathcal{D},\delta t}(x,t+\tau)-\varphi_{\mathcal{D},\delta t}(x,t))^2 \; d\mathbf{x}dt \\
&\quad \le C\int_0^{T-\tau}\sum_{n\in n(t,\tau)}\delta t\left(|\varphi^{\lceil (t+\tau)/\delta t\rceil}|_{X_\mathcal{D}}+|\varphi^{\lceil t/\delta t\rceil}|_{X_\mathcal{D}}\right) \\
&\qquad\qquad \left(|p^n|_{X_\mathcal{D}}+|\varphi^n|_{X_\mathcal{D}}+\frac{1}{\delta t}\|k_o\|_{L^2(\Omega\times(t_{n-1},t_n))}+1\right) \\
&\quad \le C\int_0^{T-\tau}\sum_{n\in n(t,\tau)}\delta t(|\varphi^{\lceil (t+\tau)/\delta t\rceil}|^2_{X_\mathcal{D}} \\
&\qquad\qquad +|\varphi^{\lceil t/\delta t\rceil}|^2_{X_\mathcal{D}}+|p^n|^2_{X_\mathcal{D}}+|\varphi^n|^2_{X_\mathcal{D}}+\frac{1}{\delta t}\|k_o\|^2_{L^2(\Omega\times(t_{n-1},t_n))}+1) \\
&\quad \le C\tau\sum_{n=1}^N\delta t\left(|\varphi^n|^2_{X_\mathcal{D}}+|p^n|^2_{X_\mathcal{D}}+\frac{1}{\delta t}\|k_o\|^2_{L^2(\Omega\times(t_{n-1},t_n))}+1\right).
\end{aligned}$$

Finally we use Theorem 12.1 and the hypothesis $(\mathcal{H}_6)$ to complete the proof. $\square$

14 Convergence result

This section is devoted to the proof of the following theorem.

Theorem 14.1. *Let $\mathfrak{D}$ be a sequence of discretizations of Ω and such that there exists a positive constant θ satisfying $\theta_\mathcal{D}\le\theta$ for all $\mathcal{D}\in\mathfrak{D}$. Let $\delta\mathfrak{t}$ be a sequence of real positif numbers, such that $T/\delta t\in\mathbb{N}$ for all $\delta t\in\delta\mathfrak{t}$ and such that δt tends to zero along $\delta\mathfrak{t}$. Let $(s_{\mathfrak{D},\delta\mathfrak{t}},p_{\mathfrak{D},\delta\mathfrak{t}})=(s_{\mathcal{D},\delta t},p_{\mathcal{D},\delta t})_{\mathcal{D}\in\mathfrak{D},\delta t\in\delta\mathfrak{t}}$ be a sequence of approximate solutions corresponding to $\mathfrak{D}$ and $\delta\mathfrak{t}$, and let $\varphi_{\mathfrak{D},\delta\mathfrak{t}}=\varphi(s_{\mathfrak{D},\delta\mathfrak{t}})$. Then there exists a subsequence of $(s_{\mathfrak{D},\delta\mathfrak{t}},p_{\mathfrak{D},\delta\mathfrak{t}})$, which we denote again by $(s_{\mathfrak{D},\delta\mathfrak{t}},p_{\mathfrak{D},\delta\mathfrak{t}})$, such that $s_{\mathcal{D},\delta t}\to s$ strongly in $L^2(Q_T)$ and $p_{\mathcal{D},\delta t}\to p$ weakly in $L^2(Q_T)$ as $h_\mathcal{D},\delta t\to 0$, where (s,p) is a weak solution of the problem $(10.5)-(10.7)$.*

14.1 Relative compactness of $\varphi_{\mathfrak{D},\delta t}$

The relative compactness of $\varphi_{\mathfrak{D},\delta t}$ follows from Theorem 12.1 and the estimates on time and space translates. Indeed, let us prolonge $\varphi_{\mathcal{D},\delta t}$ by zero outside of Q_T and remark that we may prolonge as well $\nabla_{\mathcal{D},\delta t}\varphi_{\mathcal{D},\delta t}$ by 0 outside of Q_T. The Fréchet-Kolmogorov Compactness Theorem implies that the family $\varphi_{\mathfrak{D},\delta t}$ is relatively compact in $L^2(Q_T)$, which yields in turn that $\varphi_{\mathfrak{D},\delta t}$ contains a subsequence strongly converging in $L^2(Q_T)$ (and also in $L^2(\mathbb{R}^d \times (0,T))$) to some $\phi \in L^\infty(0,T;L^2(\mathbb{R}^d))$. In view of (12.7), we deduce that $(\nabla_{\mathcal{D},\delta t}\varphi_{\mathcal{D},\delta t})_{\mathcal{D}\in\mathfrak{D},\delta t\in\delta t}$ is weakly relatively compact in $L^2(\mathbb{R}^d \times (0,T))$ and it can be shown (see [10, Theorem 7.1] and [66, Lemma 4.2]) that $\nabla_{\mathcal{D},\delta t}\varphi_{\mathcal{D},\delta t} \rightharpoonup \nabla\phi$ weakly in $L^2(\mathbb{R}^d \times (0,T))$. Thus $\phi \in L^2(0,T;H^1(\mathbb{R}^d))$; since $\phi = 0$ for all $\mathbf{x} \in \mathbb{R}^d \setminus \Omega$, it follows that $\phi \in L^2(0,T;H_0^1(\Omega))$.

14.2 Weak relative compactness of $p_{\mathfrak{D},\delta t}$

Since the pressure equation does not involves any time derivatives on p, we can not obtain any estimate time translates. Nevertheless, Theorem 12.1 and discrete Poincaré inequality implies that $p_{\mathfrak{D},\delta t}$ is bounded in $L^\infty(0,T;L^2(\Omega))$. Then up to a sequence $p_{\mathcal{D},\delta t}$ converges weakly in $L^2(Q_T)$ (and once again in $L^2(\mathbb{R}^d \times (0,T))$ taking $p_{\mathcal{D},\delta t} = 0$ outside of Q_T) to some $p \in L^\infty(0,T;L^2(\mathbb{R}^d))$; moreover using same arguments as for $\varphi_{\mathcal{D},\delta t}$, we deduce that $\nabla_{\mathcal{D},\delta t}p_{\mathcal{D},\delta t} \rightharpoonup \nabla p$ weakly in $L^2(\mathbb{R}^d \times (0,T))$ and $p \in L^2(0,T;H_0^1(\Omega))$.

14.3 Saturation equation

The assumption $(\mathcal{H}_1)$ implies that

$$\|s_{\mathcal{D},\delta t}\|_{L^\infty(0,T;L^2(\Omega))} \le m(\Omega) + \|\varphi_{\mathcal{D},\delta t}\|_{L^\infty(0,T;L^2(\Omega))}.$$

Therefore in view of Theorem 12.1 and we have that $\{s_{\mathcal{D},\delta t}\}$ contains a subsequence (denoted again by $\{s_{\mathcal{D},\delta t}\}$) such that $s_{\mathcal{D},\delta t}$ converges to some $s \in L^\infty(0,T;L^2(\Omega))$ strongly in $L^2(Q_T)$ as $h_{\mathcal{D}}, \delta t \to 0$. It remains to show that (s,p) satisfies the integral equalities (iv) of Definition 10.1, for this purpose we introduce the function space

$$\Psi = \{\psi \in C^{2,1}(\overline{\Omega} \times [0,T]), \quad \psi = 0 \text{ on } \partial\Omega \times [0,T], \quad \psi(\cdot,T) = 0\}.$$

Taking an arbitrary $\psi \in \Psi$, we define the sequence of elements of $X_{\mathcal{D},0}$

$$\psi^n = P_{\mathcal{D}}\psi(\cdot,t_n) \text{ for all } n \in \{1,\dots,N\},$$

which implies $\psi_K^n = \psi(\mathbf{x}_K,t_n)$ and $\psi_\sigma^n = \psi(\mathbf{x}_\sigma,t_n)$. Setting $w^n = \psi^{n-1}$ in the discrete saturation equation we multiply it by the time step δt and we sum over $n \in \{1,\dots,N\}$ to obtain

$$S_T + S_C + S_D = S_S,$$

where

$$T_T = \sum_{n=1}^{N} \sum_{K\in\mathcal{M}} \omega(K)\psi_K^{n-1}(s_K^n - s_K^{n-1}),$$

$$T_C = \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\psi_K^{n-1} - \psi_\sigma^{n-1})\mathcal{G}(Q_{K,\sigma}^n f(\cdot) + \gamma(\cdot) g_{K,\sigma}; s_K^n, s_\sigma^n)$$

$$T_D = \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\psi_K^{n-1} - \psi_\sigma^{n-1}) F_{K,\sigma}(\varphi(s^n))$$

and

$$T_S = \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} m(K)\psi_K^{n-1} k_{o,K}^n.$$

14.3.1 Time evolution term

Using discrete integration by parts and the fact that $\varphi(\mathbf{x}, T) = 0$ we obtain

$$T_T = -\sum_{n=1}^{N} \sum_{K\in\mathcal{M}} \omega(K)(\psi_K^n - \psi_K^{n-1}) s_K^n - \sum_{K\in\mathcal{M}} \omega(K)\psi_K^0.s_K^0.$$

Clearly

$$\sum_{K\in\mathcal{M}} \omega(K)\psi_K^0 s_K^0 \to \int_\Omega \omega(\mathbf{x})\psi(\mathbf{x}, 0) s_0(\mathbf{x})\; d\mathbf{x};$$

next we define

$$T_T^1 = \sum_{n=1}^{N} \sum_{K\in\mathcal{M}} \omega(K)(\psi_K^n - \psi_K^{n-1}) s_K^n - \int_0^T \int_\Omega s(\mathbf{x}, t)\psi_t(\mathbf{x}, t)\; d\mathbf{x}dt,$$

and we add and subtract $\displaystyle\int_{t_{n-1}}^{t_n} \int_K s_K^n \psi_t(\mathbf{x}, t)\; d\mathbf{x}dt$ in each term to obtain

$$\begin{aligned} T_T^1 &= \sum_{n=1}^{N} \sum_{K\in\mathcal{M}} m(K) s_K^n \int_{t_{n-1}}^{t_n} (\psi_t(\mathbf{x}_K, t) - \psi_t(\mathbf{x}, t))\; dt \\ &+ \int_0^T \int_\Omega (s_{\mathcal{D},\delta t}(\mathbf{x}, t) - s(\mathbf{x}, t))\psi_t(\mathbf{x}, t)\; d\mathbf{x}dt. \end{aligned} \tag{14.1}$$

In view of the regularity of ψ we have that for all $x \in K$ and all $K \in \mathcal{M}$ it holds

$$|\psi_t(\mathbf{x}_K, t) - \psi_t(\mathbf{x}, t)| \le C h_{\mathcal{D}}.$$

Since $s_{\mathcal{D},\delta t}$ is bounded in $L^\infty(0, T; L^2(\Omega))$ the first term in (14.1) tends to zero as $h_{\mathcal{D}}, \delta t \to 0$. Further, since $|\psi_t(\mathbf{x}, t)| \le C_\psi$, the second term in (14.1) also tends to zero in view of the strong convergence of $s_{\mathcal{D},\delta t}$.

14.3.2 Convection term

Using the notation of (11.28) we split T_C into $T_C = T_C^1 + T_C^2$ with

$$T_C^1 = \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\psi_K^{n-1} - \psi_\sigma^{n-1}) Q_{K,\sigma}^n f_{K,\sigma}^n,$$

$$T_C^2 = \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\psi_K^{n-1} - \psi_\sigma^{n-1}) g_{K,\sigma} \gamma_{K,\sigma}^n.$$

First we study the limit of the term T_C^2, which can be written as $T_C^2 = T_C^{21} + T_C^{22}$,

$$T_C^{21} = \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\psi_K^{n-1} - \psi_\sigma^{n-1}) g_{K,\sigma} \gamma_K^n,$$
$$T_C^{22} = \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\psi_K^{n-1} - \psi_\sigma^{n-1}) g_{K,\sigma} (\gamma_{K,\sigma}^n - \gamma_K^n).$$

Using (11.20) and (11.30) one has

$$\begin{aligned} T_C^{21} &= \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \gamma_K^n \mathbf{K}_K \mathbf{g} \cdot \left(\sum_{\sigma\in\mathcal{E}_K} m(\sigma)(\psi_K^{n-1} - \psi_\sigma^{n-1}) \mathbf{n}_{K,\sigma} \right) \\ &= - \sum_{n=1}^{N} \sum_{K\in\mathcal{M}} \int_{t_{n-1}}^{t_n} \int_K \gamma(s_{\mathcal{D},\delta t}) \mathbf{K} \mathbf{g} \cdot \nabla_K \psi^{n-1}. \end{aligned}$$

We deduce from the regularity of ψ and Lemma 11.1 that

$$|\nabla_K \psi^n - \nabla\psi(\mathbf{x}, t)| \le C(h_{\mathcal{D}} + \delta t)$$

for all $(\mathbf{x}, t) \in K \times (t_{n-1}, t_n)$, which implies in view of the strong convergence of $s_{\mathcal{D},\delta t}$ to s and the Lipschitz continuity of γ that

$$T_C^{21} \to - \int_0^T \int_\Omega \gamma(s) \mathbf{K} \mathbf{g} \cdot \nabla\psi \; d\mathbf{x} dt$$

as δt and $h_{\mathcal{D}}$ tend to zero. Next we show that $\lim_{h_{\mathcal{D}},\delta t\to 0} T_C^{22} = 0$. Thanks to the regularity of ψ we have

$$|\psi_K^n - \psi_\sigma^n| \le C_\psi |\mathbf{x}_K - \mathbf{x}_\sigma| \le C_\psi \theta_{\mathcal{D}} d_{K,\sigma}, \tag{14.2}$$

therefore in view of $(\mathcal{H}_{2b})$ and $(\mathcal{H}_{4b})$ there exists a positive constant C such that

$$\begin{aligned} |T_C^{22}| &= \left| \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma)(\psi_K^{n-1} - \psi_\sigma^{n-1}) \mathbf{K}_K \mathbf{g} (\gamma_{K,\sigma}^n - \gamma_K^n) \right| \\ &\le C \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma) d_{K,\sigma} |\gamma_{K,\sigma}^n - \gamma_K^n|. \end{aligned}$$

Remark that $\mathcal{S}\left(Q_{K,\sigma}^n f(\cdot) + \gamma(\cdot) g_{K,\sigma}; s_K^n, s_\sigma^n\right) \in [\min(s_K^n, s_\sigma^n), \max(s_K^n, s_\sigma^n)]$ therefore in view of (11.26) and the Lipschitz continuity of γ we obtain

$$|T_C^{22}| \le C \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma) d_{K,\sigma} |s_K^n - s_\sigma^n|.$$

The Hölder continuity of function φ^{-1} (the assumption $(\mathcal{H}_1)$) implies that

$$\begin{aligned} |T_C^{22}| &\le C \sum_{n=1}^{N} \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \delta t m(\sigma) d_{K,\sigma} |\varphi_K^n - \varphi_\sigma^n|^\alpha \\ &\le C \left(\sum_{n=1}^{N} \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \delta t m(\sigma) d_{K,\sigma} |\frac{\varphi_K^n - \varphi_\sigma^n}{d_{K,\sigma}}|^\alpha \right) h_{\mathcal{D}}^\alpha. \end{aligned}$$

Applying Hölder inequality we obtain

$$
\begin{aligned}
|T_C^{22}| \; &\leq C \left(\sum_{n=1}^{N} \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \delta t m(\sigma) d_{K,\sigma} |\frac{\varphi_K^n - \varphi_\sigma^n}{d_{K,\sigma}}|^2 \right)^{\frac{\alpha}{2}} \\
&\cdot \left(\sum_{n=1}^{N} \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\delta t m(\sigma) d_{K,\sigma})^{\frac{1-\alpha/2}{\beta}} \right)^{\beta} h_{\mathcal{D}}^{\alpha},
\end{aligned}
$$

where $\frac{\alpha}{2} + \beta = 1$. Using the a priori estimates we finally obtain that

$$
|T_C^{22}| \leq C \left(dm(\Omega) T \right)^{1-\frac{\alpha}{2}} \|\varphi\|^{\alpha}_{X_{\mathcal{D},\delta t}} h_{\mathcal{D}}^{\alpha},
$$

where $C = C_\psi \theta_{\mathcal{D}} L_\gamma H_\varphi \overline{\mathrm{K}} |\mathbf{g}|$. Thus, $\lim_{h_{\mathcal{D}}\to 0} T_C^{22} = 0$. Next we show that

$$
T_C^1 \to \int_0^T \int_\Omega f(s) \mathbf{K} \left(\lambda(s) \nabla p - \xi(s) \mathbf{g} \right) \cdot \nabla \psi \; d\mathbf{x} dt
$$

as $h_{\mathcal{D}}, \delta t \to 0$. In view of (11.27) we can write T_C^1 as $T_C^1 = T_C^{11} + T_C^{12} + T_C^{13}$, where

$$
\begin{aligned}
T_C^{11} &= \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\psi_K^{n-1} - \psi_\sigma^{n-1}) \xi_{K,\sigma}^n f_{K,\sigma}^n g_{K,\sigma}, \\
T_C^{12} &= \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\psi_K^{n-1} - \psi_\sigma^{n-1}) \lambda_K^n \mathcal{F}_{K,\sigma}(p^n) f_K^n, \\
T_C^{13} &= \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} (\psi_K^{n-1} - \psi_\sigma^{n-1}) \lambda_K^n \mathcal{F}_{K,\sigma}(p^n) (f_{K,\sigma}^n - f_K^n).
\end{aligned}
$$

Using the same arguments as for the term T_C^2 one can show that

$$
T_C^{11} \to -\int_0^T \int_\Omega f(s) \xi(s) \mathbf{K} \mathbf{g} \cdot \nabla \psi \; d\mathbf{x} dt.
$$

It follows from (11.34) and (11.32) that the term T_C^{12} can be written as

$$
T_C^{12} = \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} \int_K \lambda(s_{\mathcal{D},\delta t}) f(s_{\mathcal{D},\delta t}) \mathbf{K} \nabla_{\mathcal{D},\delta t} p_{\mathcal{D},\delta t} \cdot \nabla_{\mathcal{D}} \psi^{n-1} \; d\mathbf{x} dt
$$

The regularity of ψ combined with Lemma 11.1 implies that

$$
\|\nabla_{\mathcal{D}} \psi^n(\mathbf{x}) - \nabla \psi(\mathbf{x}, t)\|_{(L^\infty(\Omega))^d} \leq C(h_{\mathcal{D}} + \delta t) \tag{14.3}
$$

for all $(\mathbf{x}, t) \in K \times (t_{n-1}, t_n)$. In view of the Lipschitz continuity of the functions λ and f, also in view of the weak convergence of $\nabla_{\mathcal{D},\delta t} p_{\mathcal{D},\delta t}$ to ∇p we conclude that

$$
T_C^{12} \to \int_0^T \int_\Omega \lambda(s) f(s) \mathbf{K} \nabla p \cdot \nabla \psi \; d\mathbf{x} dt.
$$

It remains to show that T_C^3 tends to zero as $h_{\mathcal{D}}, \delta t \to 0$. Thanks to Lemma 11.5 the term T_C^{13} can be estimated by

$$
(T_C^{13})^2 \leq C_1^2 \|\nabla_{\mathcal{D},\delta t} p\|^2_{L^2(Q_T)} \sum_{n=1}^{N} \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \frac{m(\sigma)}{d_{K,\sigma}} \lambda^2(s_K^n) (\psi_K^{n-1} - \psi_\sigma^{n-1})^2 (f_{K,\sigma}^n - f_K^n)^2
$$

which implies

$$(T_C^{13})^2 \le C_1^2 C_\psi^2 \theta_{\mathcal{D}}^2 \|\lambda\|^2_{L^\infty((0,1))} \|\nabla_{\mathcal{D},\delta t} p\|^2_{L^2(Q_T)} \sum_{n=1}^N \delta t \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma) d_{K,\sigma} (f^n_{K,\sigma} - f^n_K)^2.$$

Using the same arguments that for the term T_C^{22} one shows that $|T_C^{13}| \le Ch^\alpha \to 0$ as $h_{\mathcal{D}} \to 0$.

14.3.3 Diffusion term

In view of (11.36) and (11.33) one has

$$T_D = \sum_{n=1}^N \int_{t_{n-1}}^{t_n} \int_\Omega \mathbf{K} \nabla_{\mathcal{D}} \varphi(s^n) \cdot \nabla_{\mathcal{D}} \psi^{n-1} \; d\mathbf{x}dt.$$

In view of (14.3) and since $\nabla_{\mathcal{D},\delta t} \varphi_{\mathcal{D},\delta t} \to \nabla\varphi(s)$ weakly in $L^2(Q_T)$ we deduce that the term T_D tends to $\int_0^T \int_\Omega \mathbf{K}\nabla\varphi(s) \cdot \nabla\psi \; d\mathbf{x}dt$ as $h_{\mathcal{D}}, \delta t$ tend to zero.

14.3.4 Source term

In view of regularity of ψ and thanks to the definition (11.7) of $k^n_{o,K}$ one has

$$\sum_{n=1}^N \sum_{K\in\mathcal{M}} \delta t m(K) \psi_K^{n-1} k^n_{o,K} \to \int_0^T \int_\Omega \psi k_o \; d\mathbf{x}dt$$

as $h_{\mathcal{D}}, \delta t \to 0$.

14.4 Pressure equation

Setting $v^n = \psi^n$ in the discrete pressure equation (11.16) we multiply it by the time step δt and we sum over $n \in \{1 \dots N\}$, we obtain

$$\sum_{n=1}^N \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \delta t (\psi^n_K - \psi^n_\sigma) Q^n_{K,\sigma} = \sum_{n=1}^N \sum_{K\in\mathcal{M}} \delta t m(K) \psi^n_K (k^n_{w,K} + k^n_{o,K}).$$

Replacing $f^n_{K,\sigma}$ by 1 in the definition of T_C^1 allows to conclude that

$$\sum_{n=1}^N \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \delta t (\psi^n_K - \psi^n_\sigma) Q^n_{K,\sigma} \to \int_0^T \int_\Omega (\lambda(s)\mathbf{K}\nabla p - \xi(s)\mathbf{K}\mathbf{g}) \cdot \nabla\psi \; d\mathbf{x}dt.$$

It is also clear that

$$\sum_{n=1}^N \sum_{K\in\mathcal{M}} \delta t m(K) \psi^n_K k^n_{o,K} \to \int_0^T \int_\Omega \psi k_o \; d\mathbf{x}dt$$

as $h_{\mathcal{D}}, \delta t \to 0$. We deduce from the density of the set Ψ in the set $\{\psi \in L^2(0,T; H^1_0(\Omega)), \psi_t \in L^\infty(Q_T), \psi(\cdot,T) = 0\}$ that (s,p) is a weak solution of the continuous problem (10.5)-(10.7) in the sense of Definition 10.1.

15 Numerical experiments

In this section we present some numerical results obtained using the scheme.

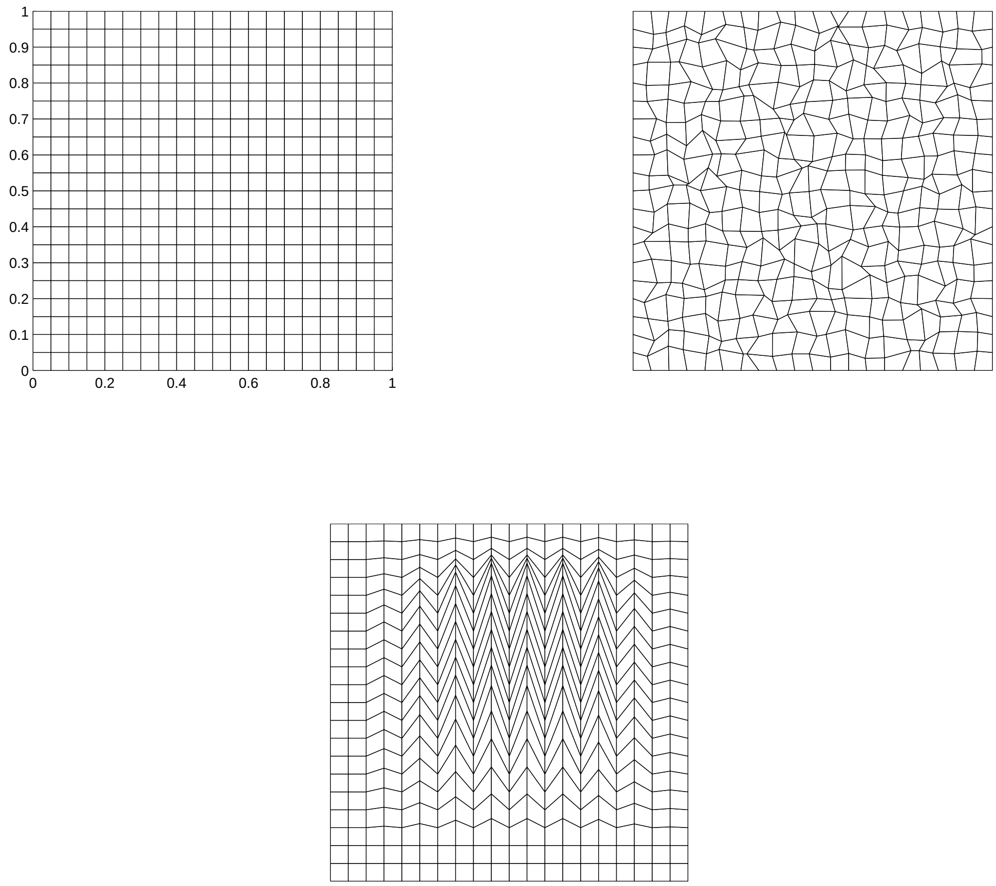

FIGURE 10 – 20×20 orthogonal, randomly perturbed and Kershaw meshes

15.1 Five spot problem I

We consider the quarter five spot problem, which is the model problem in oil recovery ; more precisely we consider a two dimensional horizontal (gravity is neglected) domain $\Omega = (0,1)^2$, which is initially saturated in oil ($s_0 = 1$ in Ω). The oil-phase is then displaced by water, which is injected at the lower left corner ; the production well is placed at the upper right corner. Assume that the oil-phase and water-phase mobilities and capillary pressure are respectively defined by

$$\lambda_o(s) = \frac{s^3}{2}, \quad \lambda_w(s) = \frac{(1-s)^3}{2} \text{ and } \pi(s) = \frac{1}{2}\sqrt{\frac{s}{1-s}}.$$

Recall that the nonlinear functions in (10.5)-(10.7) are defined by

$$\lambda(s) = \lambda_o(s) + \lambda_w(s), \quad f(s) = \frac{\lambda_o(s)}{\lambda(s)}, \quad \varphi(s) = \int_0^s \lambda_w(\tau) f(\tau) \pi'(\tau) d\tau.$$

The effects of wells are modeled by terms on the right-hand side of (10.5) and (10.7)

$$k_o(s) = f(\overline{s})q^+ + f(s)q^-$$

and

$$k_w(s) = (1 - f(\overline{s}))\, q^+ + (1 - f(s))\, q^-,$$

where $q^+ = q^+(\mathbf{x}, t)$ and $q^- = q^-(\mathbf{x}, t)$ denote the production and the injection rates. The value $\overline{s}$ is the oil saturation of an injected fluid, which is set to $\overline{s} = 0$. We assume that q^+ and q^- are Dirac functions

$$q^+ = \delta(\mathbf{x}) \text{ and } q^- = \delta(\mathbf{x} - (1, 1)).$$

We recall that due to the fact that $\varphi'(0) = 0$ the water front propagates with a finite speed. We consider three types of meshes, namely the square mesh, the randomly perturbed quadrangular mesh and the Kershaw mesh (see Figure 10). For each type of mesh we also consider three different resolutions 20×20, 40×40 and 80×80 elements ; the meshes are constructed in such a way that the regularity parameter $\theta_{\mathcal{D}}$ remains almost constant for each class namely $\theta_{\mathcal{D}} = 2\sqrt{2}$ for orthogonal meshes, $\theta_{\mathcal{D}} \approx 30$ for random and Kershow meshes.

Figure 11 represents the water-phase saturation profile along the diagonal $((0, 0), (0, 1))$ at time $t = 0.6$ for the family of orthogonal grids. Note that in this case the hybrid discretization of the diffusion operator is equivalent to the classical two-point flux approximation scheme. Since the exact solution is not available for this problem we compare the results obtained using "bad" (randomly perturbed and Kershaw) meshes with the results corresponding to the orthogonal grids (see Figures 11-12). We first remark that the solution obtained on the square and randomly perturbed meshes are in very good agreement. However the Kershaw mesh seems to be more challenging. The corresponding numerical solution has a non negligible deviation from the reference one, which we believe is due to the mesh orientations effects breaking the symmetry of the solution (see Figure 13). A positive point is that for a sufficiently small mesh size the speed of the water front propagation is correct.

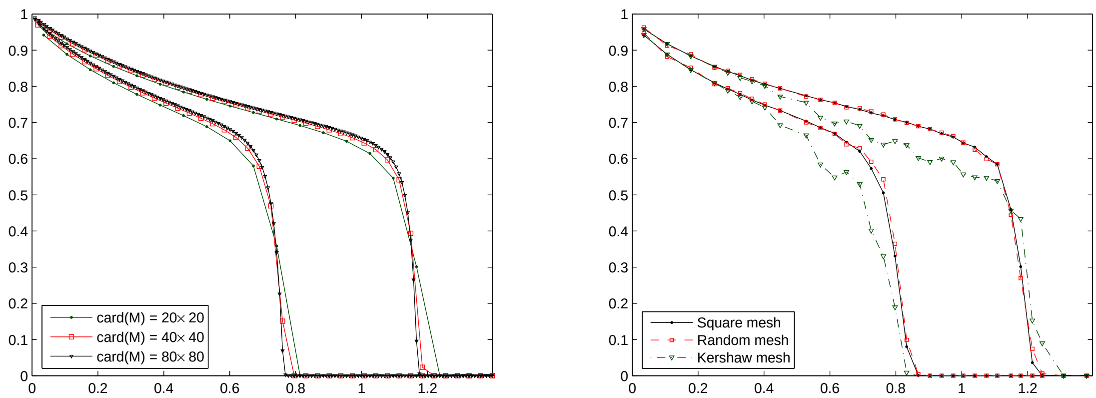

FIGURE 11 – Water profile at time $t = 0.3$ and $t = 0.6$ for a family of orthogonal meshes (left) and the same profile for three type of meshes with 20×20 elements (right)

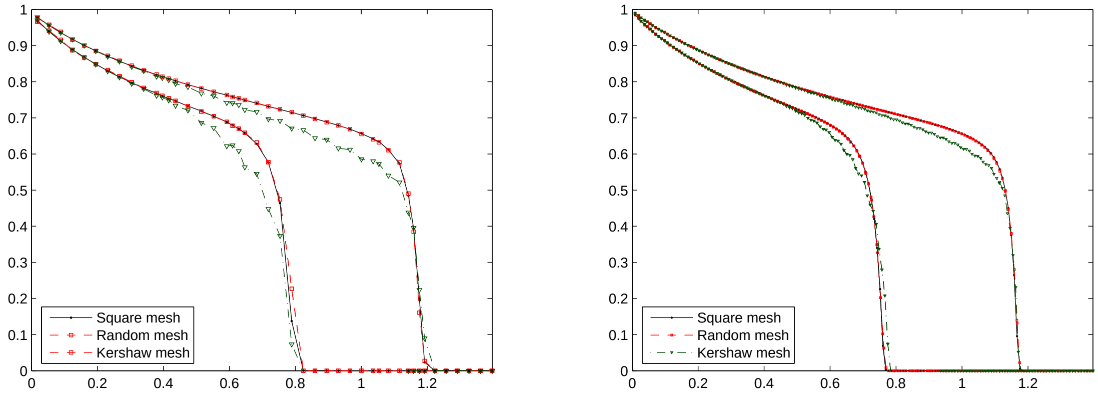

FIGURE 12 – Water profile at time $t = 0.3$ and $t = 0.6$ for three type of meshes with 40×40 elements (left), 80×80 elements (right)

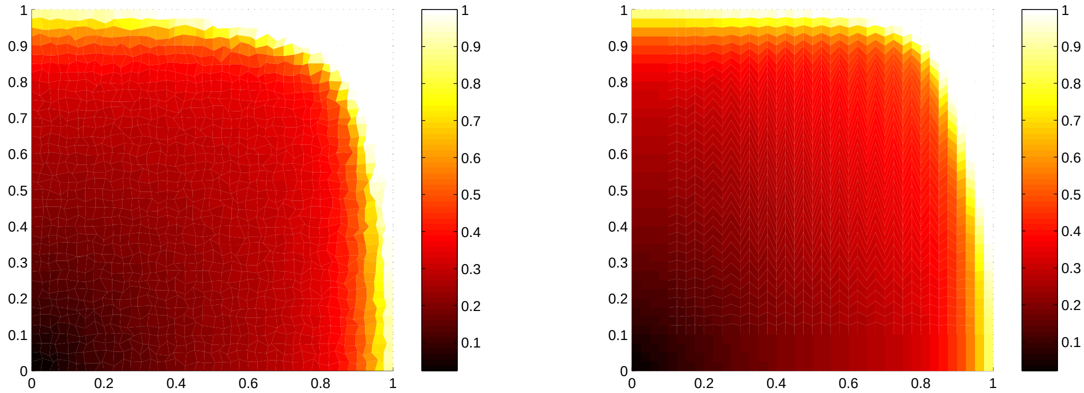

FIGURE 13 – Oil saturation at time $t = 0.6$ using 40×40 randomly perturbed (left) and Kershaw (right) meshes

15.2 Five spot problem II (Heterogeneous and anisotropic medium)

Next we consider two test cases proposed in [91], which allow to investigate the qualitative behavior of the scheme in the case of anisotropic and heterogeneous absolute permeability tensor $\mathbf{K}$.

First we assume that the porous medium is homogeneous. The absolute permeability tensor $\mathbf{K}$ is defined by $\mathbf{K} = \mathbf{R}_\theta \mathbf{D} \mathbf{R}'_\theta$, where

$$\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 10^{-3} \end{pmatrix} \text{ with } \mathbf{R}_\theta = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix},$$

for all real θ. First we set $\theta = \pi/4$ and we present on Figure 14 the approximate water-phase saturation and the global pressure fields at time $t = 0.02$. Note that due to the anisotropy the solution form a relatively fine strip propagating along the direction $(cos\theta, sin\theta)$, which is an eigenvector of $\mathbf{K}$ corresponding to the value 1.

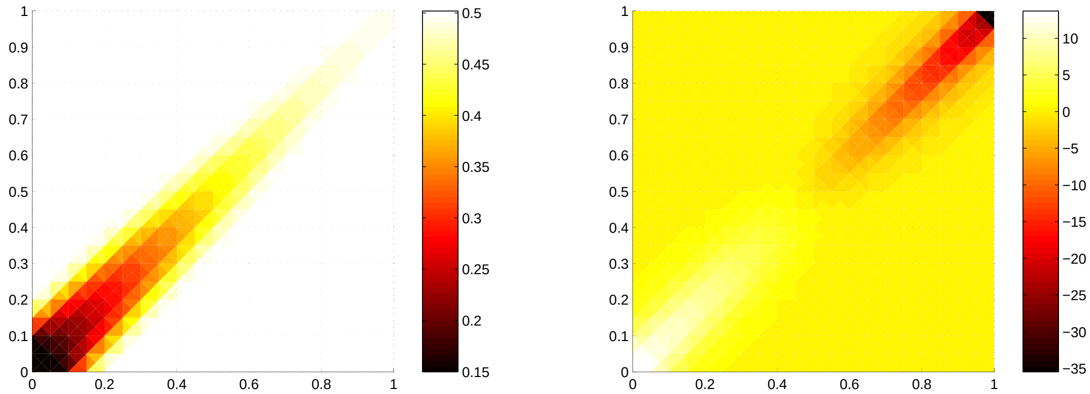

FIGURE 14 – Oil saturation (left) and global pressure field (right) at time $t = 0.02$ (homogeneous case)

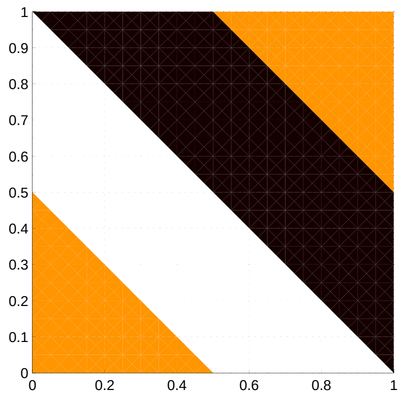

FIGURE 15 – Absolute permeability field

Next let us consider a more complex geometry. We assume that the computational domain $\Omega = (0,1)^2$ is composed of four layers

$$\begin{array}{ll} \Omega_1 = \Omega \cap \{|x_1| + |x_2| < 0.5\}, & \Omega_2 = \Omega \cap \{0.5 < |x_1| + |x_2| < 1\}, \\ \Omega_3 = \Omega \cap \{1 < |x_1| + |x_2| < 1.5\}, & \Omega_4 = \Omega \cap \{|x_1| + |x_2| > 1.5\}, \end{array}$$

which we represent on Figure 15 by different colors. The piecewise constant absolute permeability field is defined by $\mathbf{K}(\mathbf{x})|_{\mathbf{x}\in\Omega_i} = \mathbf{K}_i$, with $i \in \{1,\dots,4\}$ and

$$\mathbf{K}_1 = \mathbf{K}_4 = \mathbf{R}_{\pi/4}\mathbf{D}\mathbf{R}'_{\pi/4}, \quad \mathbf{K}_2 = \mathbf{D}, \quad \mathbf{K}_3 = \mathbf{R}_{\pi/2}\mathbf{D}\mathbf{R}'_{\pi/2}.$$

On Figures 16 we present the approximate water-phase saturation and global pressure field at time $t = 0.07$. As in the previous case water propagates in the most permeable direction, which changes at the interface between two different layers.

Remark 15.1. *In conclusion to this section we can state that the scheme seems to be perform well on the distorted meshes and it gives a fairly correct in the case of a highly anisotropic and heterogeneous absolute permeability fields. In future it would be interesting to compare our scheme with others methods both in the cases with and without gravity.*

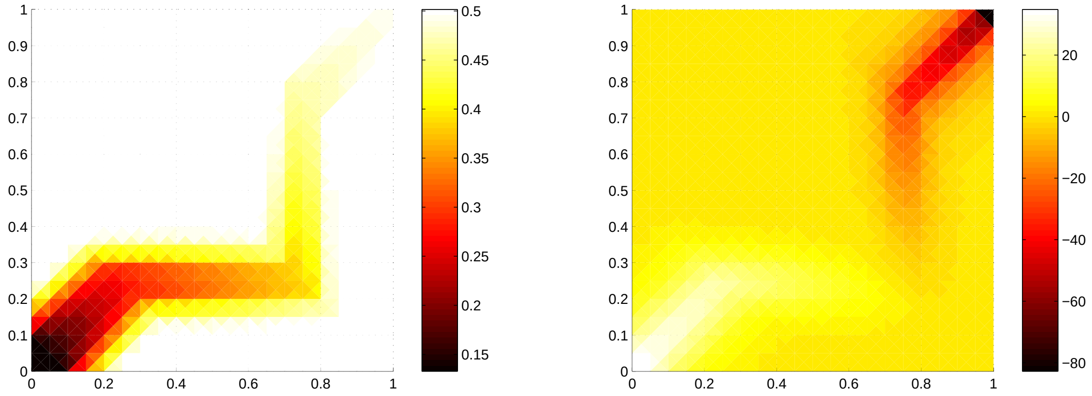

FIGURE 16 – Oil saturation (left) and global pressure field (right) at time $t = 0.07$ (heterogeneous case)

Troisième partie

Finite volume approximation for an immiscible two-phase flow in porous media with discontinuous capillary pressure

Abstract We consider a immiscible incompressible two-phase flow within a porous medium made of two different rocks. The flow is governed by the Darcy-Muskat law and a capillary pressure law, that depends on space in a discontinuous way, so that the capillary pressure field can be discontinuous at the interface. Using the concept of multivalued phase pressures, we introduce a notion of weak solution for the flow, and we establish the convergence of a finite volume approximation towards a weak solution.

This work was supported by the GdR MoMaS (PACEN/CNRS, ANDRA, BRGM, CEA, EdF, IRSN), France.

16 Introduction

16.1 Multivalued phase pressures

Models of immiscible two-phase flows are widely used in oil engineering to predict the motion of oil in the subsoil. They have been widely studied from a mathematical point of view (see e.g. [6], [7], [14], [15], [45]) and from a numerical point of view (see e.g. [44], [50], [51], [48], [71], [90]). In this model, sometimes called *dead-oil* approximation, it is assumed that there are only two phases oil and water, and that each phase is only composed of a single component.

The governing equations are derived by substituting the Darcy-Muskat (or diphasic Darcy) law into the conservation equations for both phases, that is that for each phase $\alpha \in \{o, w\}$ (o corresponds to the oil phase, while w corresponds to the water phase) :

$$\phi \partial_t s_\alpha - \operatorname{div}\left(K \frac{kr_\alpha(s_\alpha)}{\mu_\alpha} \left(\nabla p_\alpha - \rho_\alpha \mathbf{g} \right) \right) = 0, \tag{16.1}$$

where $\phi = \phi(\mathbf{x})$ is the porosity of the rock ($\phi \in (0,1)$ in the domain Ω), s_α is the saturation of the phase α, the permeability of the porous medium K is supposed to be a positive scalar function, the relative permeability kr_α of the phase α is a increasing function of the saturation s_α, satisfying $kr_\alpha(0) = 0$ and $kr_\alpha(1) = 1$, μ_α, p_α and ρ_α denote respectively the viscosity, the pressure and the density of the phase α, and $\mathbf{g}$ is the gravity vector. Assuming that the porous medium is saturated, one has

$$s_o + s_w = 1. \tag{16.2}$$

This relation allows to eliminate the water saturation. We note $s := s_o$, so that $s_w = 1 - s$.

Classically (see [95], [77], [78]), it is assumed that the phase pressures are connected by the equality

$$p_o - p_w = \pi(s_o), \tag{16.3}$$

where π is the capillary pressure function, which is supposed to be strictly increasing on $(0, 1)$.

As it has been stressed in [6], the natural topology for the phase pressures in such a flow is prescribed by the quantity

$$\sum_{\alpha \in \{o,w\}} \int_0^T \int_\Omega K \frac{kr_\alpha(s_\alpha)}{\mu_\alpha} \left(\nabla p_\alpha \right)^2 \, d\mathbf{x} dt. \tag{16.4}$$

Note that when the phase α vanishes, i.e. $s_\alpha = 0$, then (16.4) provides no control on the pressure p_α —it is indeed difficult to define the pressure of a missing phase—. As a consequence, if $s_\alpha = 0$ and p_β is known ($\beta \neq \alpha$), then p_α is not defined in a unique way, but it is multivalued, i.e. it can take any value lower than a threshold value, for which the phase α would appear. Practically,

$$p_o \leq p_w + \pi(0) \quad \text{if } s_o = 0 \tag{16.5}$$

and

$$p_w \leq p_o - \pi(1) \quad \text{if } s_o = 1. \tag{16.6}$$

We will take advantage of the multivalued formalism in order to deal with the case where the porous medium is composed of several rock types, and where the functions describing the porous medium depend in a discontinuous way of space. As it was stressed in [59], [24], [35], [36], the discontinuities of the capillary pressure function (16.3) have a large influence on the behavior of the flow, leading to *oil-trapping*. Several numerical methods ([62], [60], [40], [61]) have been proposed to approximate this problem, but, to our knowledge, no convergence proof has been provided for the full problem in several space dimensions. Indeed, the convergence result stated in [40] holds in the one-dimensional case, where the problem (16.1)-(16.3) becomes a single degenerate parabolic equation. In [60], simplifying assumptions are made so that the problem also reduces to a single degenerate parabolic equation. The scope of this paper is to deal with the two balance equations (16.1). In the case where the porous medium has smooth variations, it is possible to show the continuity of p_α, in the sense that on each Lipschitz continuous hypersurface of Ω, p_α admits strong traces on both sides $p_{\alpha,1}$, $p_{\alpha,2}$, and that these traces coincide :

$$p_{\alpha,1} = p_{\alpha,2}, \qquad \text{for } \alpha \in \{o, w\}. \tag{16.7}$$

If at the level of such a hypersurface Γ, the characteristic of the medium change in a discontinuous way, then, unless we have a compatibility condition on the capillary pressure functions (see [38]), the continuity of the pressure can not be prescribed by (16.7), but one should use the graph formalism :

$$p_{\alpha,1} \cap p_{\alpha,2} \neq \emptyset, \qquad \text{for } \alpha \in \{o, w\}. \tag{16.8}$$

Note that if the phase pressures are single-valued, then the conditions (16.7) and (16.8) are equivalent. Moreover, because of the conservation of mass, one has

$$\sum_{i\in\{1,2\}} K_i \frac{kr_{\alpha,i}(s_i)}{\mu_\alpha} \left(\nabla p_{\alpha,i} - \rho_\alpha \mathbf{g}\right) \cdot \mathbf{n}_i = 0, \tag{16.9}$$

where K_i and $kr_{\alpha,i}(s_i)$ denote the traces of K and $kr_\alpha(s)$ on each side of the interface, and $\mathbf{n}_i$ denotes the outward normal to Γ with respect to the side i of the interface.

The fact that the phase pressures p_α are multivalued when $s_\alpha = 0$ implies that the capillary pressure functions $s \mapsto \pi_i(s)$ must also be multivalued for $s = 0$ and $s = 1$. Thus we introduce the monotonous graphs $\tilde{\pi}_i$ defined by

$$\tilde{\pi}_i(s) = \begin{cases} [-\infty, \pi_i(0)] & \text{if } s = 0, \\ \pi_i(s) & \text{if } s \in (0,1), \\ [\pi_i(1), +\infty] & \text{if } s = 1, \end{cases}$$

for $i = 1, 2$. The capillary pressure graphs admit a continuous inverse functions, denoted by $\tilde{\pi}_i^{-1}$, which is defined on $\mathbb{R}$ by

$$\tilde{\pi}_i^{-1}(p) := \begin{cases} 0 & \text{if } p \leq \pi_i(0), \\ \pi_i^{-1}(p) & \text{if } p \in (\pi_i(0), \pi_i(1)), \\ 1 & \text{if } p \geq \pi_i(1). \end{cases}$$

Requiring (16.8) implies that, at an interface where the porous medium is discontinuous, one has

$$\tilde{\pi}_1(s_1) \cap \tilde{\pi}_2(s_2) \neq \emptyset, \tag{16.10}$$

where $\pi_1(s_1)$ and $\pi_2(s_2)$ denote the traces of $\pi(s)$ on both sides of the interface. It has been shown in [41] and [33] that this interface condition is natural, at least in the one-dimensional case. But this single relation is not sufficient to deal with the case of two equations such as in (16.1), and further information has to be given along the discontinuity lines.

16.2 The model problem and assumptions on the data

We assume that the porous medium Ω is a connex open bounded polygonal subset of $\mathbb{R}^d$, and is made of two disjoint homogeneous rocks Ω_i, $i \in \{1, 2\}$, which are both open polygonal subsets of $\mathbb{R}^d$. We assume moreover, for the sake of simplicity, that $\partial\Omega_i \cap \partial\Omega$ contains a nonempty open subset of $\partial\Omega$ for $i \in \{1, 2\}$. We denote by Γ the interface between Ω_1 and Ω_2, i.e.

$$\overline{\Gamma} = \partial\Omega_1 \cap \partial\Omega_2.$$

For all functions a depending on the physical characteristics of the rock, we use the notation $a_i = a(\cdot, x)$ if $x \in \Omega_i$.

Also remark that as it has been stressed in [37], the gravity plays a crucial role in the so called *oil-trapping* phenomenon.

We assume that at the initial time $t = 0$, the composition of the fluid is known

$$s_{|t=0} = s_0 \in L^\infty(\Omega; [0, 1]). \tag{16.11}$$

We also assume that both phase fluxes are equal to 0 through the boundary $\partial\Omega \times [0, T]$ where $T > 0$ is a positive fixed (but arbitrary) final time :

$$K_i \frac{kr_{\alpha,i}(s_i)}{\mu_\alpha} (\nabla p_{\alpha,i} - \rho_\alpha \mathbf{g}) \cdot \mathbf{n}_i = 0, \quad \text{on } (\partial\Omega \cap \partial\Omega_i) \times (0, T). \tag{16.12}$$

Nevertheless, it should be possible to deal with other types of boundary conditions, such as Dirichlet conditions on a part of the boundary and Neumann conditions on the remaining part.

We denote by Q_T and $Q_{i,T}$ the cylinders

$$Q_T := \Omega \times (0, T), \qquad Q_{i,T} := \Omega_i \times (0, T).$$

In order to control the energy of the flow, we have to make the following natural assumption on the capillary pressure functions.

Assumption 1. *The function π_i belongs to $\mathcal{C}^1([0,1])$.*

Remark 16.1. *As a direct consequence of Assumption 1,*

$$\tilde{\pi}_i^{-1} \in L^1(\mathbb{R}_-) \qquad \text{and} \qquad (1-\tilde{\pi}_i^{-1}) \in L^1(\mathbb{R}_+), \qquad i \in \{1,2\}.$$

Remark 16.2. *For the sake of simplicity we have assumed that the function π_i is finite at $s=0$ and $s=1$. However it is not difficult to take into account an unbounded capillary pressure curves, namely $\pi_i \in \mathcal{C}^1((0,1)) \cap L^1((0,1))$. In order to do so one has to consider the family of functions $\pi_i^\varepsilon \in \mathcal{C}^1([0,1])$, such that $\pi_i^\varepsilon \to \pi_i$ in $L^1((0,1))$ (see [38]). It can be shown that the analysis presented below remains true for the limit curves $\pi_i \in \mathcal{C}^1((0,1))$.*

16.3 Global pressure formulation of the problem

The lack of control on the phase pressures, described in Section 16.1, leads to important mathematical difficulties. A classical mathematical tool consists in introducing the so-called *global pressure* P to circumvent this difficulty. We define, for $x \in \Omega_i$

$$P = p_w + \int_0^{\pi_i(s)} \frac{kr_{o,i}(\pi_i^{-1}(a))}{kr_{o,i}(\pi_i^{-1}(a)) + \frac{\mu_o}{\mu_w} kr_{w,i}(\pi_i^{-1}(a))} da, \tag{16.13}$$

$$= p_o - \int_0^{\pi_i(s)} \frac{kr_{w,i}(\pi_i^{-1}(a))}{kr_{w,i}(\pi_i^{-1}(a)) + \frac{\mu_w}{\mu_o} kr_o(\pi_i^{-1}(a))} da. \tag{16.14}$$

The advantage of this fictive pressure relies on the fact that if the domain Ω is homogeneous ([44]), or if Ω varies smoothly ([15], [45]), then the global pressure belongs to the space $L^\infty((0,T); H^1(\Omega))$. This regularity result does not remain true, as it will be shown in the sequel, in the case of a discontinuous capillary pressure.

We define the fractional function $f_i(s) = \dfrac{kr_{o,i}(s)}{kr_{o,i}(s) + \frac{\mu_o}{\mu_w} kr_{w,i}(s)}$ and we introduce the Kirchhoff transform

$$\varphi_i(s) = \int_0^s K_i \frac{kr_{o,i}(a) kr_{w,i}(a)}{\mu_w kr_{o,i}(a) + \mu_o kr_{w,i}(a)} \pi_i'(a) da, \qquad \forall s \in (0,1), \tag{16.15}$$

that we extend in a continuous way by constants outside of $(0,1)$. We make moreover the following assumption on the functions φ_i.

Assumption 2. *For $i \in \{1,2\}$, the functions φ_i are Lipschitz continuous and increasing on $[0,1]$.*

It is well known that the system (16.1)–(16.3) can be rewritten in $Q_{i,T}$ under the form

$$\begin{cases} \phi_i \partial_t s + \operatorname{div}\left(f_i(s)\mathbf{q}_i + \gamma_i(s)\mathbf{g} - \nabla\varphi_i(s)\right) = 0, \\ \operatorname{div}\mathbf{q}_i = 0, \\ \mathbf{q}_i = -M_i(s)\nabla P + \zeta_i(s)\mathbf{g}, \end{cases} \tag{16.16}$$

where

$$\gamma_i(s) = K_i(\rho_o - \rho_w) \frac{kr_{o,i}(s) kr_{w,i}(s)}{\mu_w kr_{o,i}(s) + \mu_o kr_{w,i}(s)} \quad \text{and} \quad \zeta_i(s) = K_i \left(\frac{kr_{o,i}(s)}{\mu_o}\rho_o + \frac{kr_{w,i}(s)}{\mu_w}\rho_w \right). \tag{16.17}$$

The total mobility is defined by $M_i(s) = K_i \left(\frac{kr_{o,i}(s)}{\mu_o} + \frac{kr_{w,i}(s)}{\mu_w} \right)$.

Assumption 3. *For $i \in \{1,2\}$, the functions M is such that $M(s) \geq \alpha_M > 0$.*

The boundary conditions on the phase fluxes (16.12) are given by

$$\mathbf{q}_i \cdot \mathbf{n}_i = 0, \qquad (f_i(s)\mathbf{q}_i + \gamma_i(s)\mathbf{g} - \nabla\varphi_i(s)) \cdot \mathbf{n}_i = 0, \qquad \text{on } (\partial\Omega \cap \partial\Omega_i) \times (0,T). \tag{16.18}$$

Concerning the transmission conditions on the interface Γ, prescribing the relation (16.10) is not sufficient. It has to be replaced by : there exists π such that

$$\pi \in \tilde{\pi}_1(s_1) \cap \tilde{\pi}_2(s_2), \tag{16.19}$$

$$P_1 - Z_1(\pi) = P_2 - Z_2(\pi), \tag{16.20}$$

where $Z_i(p) = \int_0^p f_i \circ \pi_i^{-1}(u)du$. Note that (16.19) ensures that (16.10) holds. Equation (16.20) consists in requiring the continuity of the water pressure in some weak sense implying (16.8) for $\alpha = w$. On the other hand, adding π on both side of (16.20) leads to the continuity in the same weak sense of the oil pressure.

The conservation of the total mass and of the oil mass give

$$\sum_{i\in\{1,2\}} \mathbf{q}_i \cdot \mathbf{n}_i = 0 \quad \text{on } \Gamma, \tag{16.21}$$

$$\sum_{i\in\{1,2\}} (f_i(s)\mathbf{q}_i + \gamma_i(s)\mathbf{g} - \nabla\varphi_i(s)) \cdot \mathbf{n}_i = 0 \quad \text{on } \Gamma, \tag{16.22}$$

where $\mathbf{n}_i$ denotes the outward normal to Γ with respect to Ω_i.

Since the global pressure P is defined up to a constant, we have to impose a condition to select a solution. More precisely we impose that

$$m_{\Omega_1}(P)(t) = 0, \text{ where } m_{\Omega_i}(P)(t) := \frac{1}{m(\Omega_i)} \int_{\Omega_i} P(x,t)\, d\mathbf{x} \text{ for } i \in \{1,2\}. \tag{16.23}$$

We now define a weak solution of Problem (16.16)-(16.23).

Definition 16.1. *We say that a function pair (s,P) is a weak solution of Problem (16.16)-(16.23) if :*

1. $s \in L^\infty(Q_T;[0,1])$ *;*
2. $\varphi_i(s), P \in L^2(0,T;H^1(\Omega_i))$*, with* $m_{\Omega_1}(P)(t) = 0$ *for almost every* $t \in (0,T)$ *;*
3. *there exists a measurable function π on $\Gamma \times (0,T)$ such that, for a.e. $(x,t) \in \Gamma \times (0,T)$,* (16.19)–(16.20) *hold;*
4. *for all $\psi \in \mathcal{C}_c^\infty\left(\overline{\Omega} \times [0,T)\right)$, the following integral equalities hold :*

$$\int_0^T \sum_{i\in\{1,2\}} \int_{\Omega_i} \mathbf{q}_i \cdot \nabla\psi \; d\mathbf{x}dt = 0, \tag{16.24}$$

$$\begin{aligned}\int_0^T \int_\Omega \phi s \partial_t \psi \ d\mathbf{x}dt + \int_\Omega \phi s_0 \psi(\cdot, 0) \ d\mathbf{x} \\ = \int_0^T \sum_{i\in\{1,2\}} \int_{\Omega_i} (f_i(s)\mathbf{q}_i + \gamma_i(s)\mathbf{g} + \nabla\varphi_i(s)) \cdot \nabla\psi \ d\mathbf{x}dt,\end{aligned} \tag{16.25}$$

where

$$\mathbf{q}_i = -M_i(s)\nabla P + \zeta_i(s)\mathbf{g}.$$

We will use several time the following lemma, which ensures that the global pressure jump $P_1 - P_2$ at the interface belongs to $L^\infty(\Gamma \times (0,T))$.

Lemma 16.1. *The function $p \mapsto Z_1(p) - Z_2(p)$ belongs to $\mathcal{C}^1(\mathbb{R};\mathbb{R})$, is uniformly bounded on $\mathbb{R}$ and admits finite limits as $p \to \pm\infty$.*

Proof : Define

$$\Upsilon_i(p) = \begin{cases} \displaystyle\int_0^p \left(f_i \circ \tilde\pi_i^{-1}(p) - 1\right) dp & \text{if } p \ge 0, \\ \displaystyle\int_0^p f_i \circ \tilde\pi_i^{-1}(p) \ dp & \text{if } p < 0, \end{cases}$$

then $Z_1(p) - Z_2(p) = \Upsilon_1(p) - \Upsilon_2(p)$. Hence, we deduce that if $\Upsilon_1(p), \Upsilon_2(p)$ have finite limits for $p \to \pm\infty$, then $Z_1 - Z_2$ also does, since $f_i(1) = 1$. Since Υ_1, Υ_2 are nonincreasing functions, it only remains to check that they are bounded. Let $p \ge 0$, then

$$0 \ge \Upsilon_i(p) \ge -\int_0^p |f_i \circ \pi_i^{-1}(p) - f_i(1)| dp \ge -L_{f_i} \int_0^p |\pi_i^{-1}(p) - 1| dp \ge -L_{f_i} \|\pi_i^{-1} - 1\|_{L^1(\mathbb{R}_+)}.$$

Similarly, for $p < 0$, one has

$$0 \le \Upsilon_i(p) \le L_{f_i} \|\pi_i^{-1}\|_{L^1(\mathbb{R}_-)}.$$

We conclude the proof of Lemma 16.1 by applying Remark 16.2. □

17 The finite volume approximation

17.1 Discretization of $\mathbf{Q_T}$

Definition 17.1. *(**Admissible mesh of** Ω) An admissible mesh of Ω is given by a set $\mathcal{T}$ of open bounded convex subsets of Ω called control volumes, a family $\mathcal{E}$ of subsets of $\overline{\Omega}$ contained in hyperplanes of $\mathbb{R}^d$ with strictly positive measure, and a family of points $(x_K)_{K\in\mathcal{T}}$ (the "centers" of control volumes) satisfying the following properties :*

1. *there exists $i \in \{1,2\}$ such that $K \subset \Omega_i$. We note $\mathcal{T}_i = \{K \in \mathcal{T}, K \subset \Omega_i\}$;*
2. *$\overline{\bigcup_{K\in\mathcal{T}_i} K} = \overline{\Omega_i}$. Thus, $\overline{\bigcup_{K\in\mathcal{T}} K} = \overline{\Omega}$;*
3. *for any $K \in \mathcal{T}$, there exists a subset $\mathcal{E}_K$ of $\mathcal{E}$ such that $\partial K = \bigcup_{\sigma\in\mathcal{E}_K} \overline{\sigma}$. Furthermore, $\mathcal{E} = \bigcup_{K\in\mathcal{T}} \mathcal{E}_K$;*
4. *for any $(K,L) \in \mathcal{T}^2$ with $K \neq L$, either the "length"(i.e. the $(d-1)$ Lebesgue measure) of $\overline{K} \cap \overline{L}$ is 0 or $\overline{K} \cap \overline{L} = \overline{\sigma}$ for some $\sigma \in \mathcal{E}$. In the latter case, we write $\sigma = K|L$, and*

- $\mathcal{E}_i = \{\sigma \in \mathcal{E},\ \exists (K,L) \in \mathcal{T}_i^2,\ \sigma = K|L\}$, $\mathcal{E}_{\rm int} = \mathcal{E}_1 \cup \mathcal{E}_2$, $\mathcal{E}_{K,{\rm int}} = \mathcal{E}_K \cap \mathcal{E}_{\rm int}$,
- $\mathcal{E}_{\rm ext} = \{\sigma \in \mathcal{E},\ \sigma \subset \partial\Omega\}$, $\mathcal{E}_{K,{\rm ext}} = \mathcal{E}_K \cap \mathcal{E}_{\rm ext}$,
- $\mathcal{E}_\Gamma = \{\sigma \in \mathcal{E},\ \exists (K,L) \in \mathcal{T}_1 \times \mathcal{T}_2,\ \sigma = K|L\}$, $\mathcal{E}_{K,\Gamma} = \mathcal{E}_K \cap \mathcal{E}_\Gamma$;

5. *The family of points $(x_K)_{K\in\mathcal{T}}$ is such that $x_K \in K$ (for all $K \in \mathcal{T}$) and, if $\sigma = K|L$, it is assumed that the straight line (x_K, x_L) is orthogonal to σ.*

For a control volume $K \in \mathcal{T}_i$, we denote by $\mathcal{N}_K = \{L \in \mathcal{T}_i, \sigma = K|L \in \mathcal{E}_{K,i}\}$ the set of the neighbors and by $m(K)$ its measure. For all $\sigma \in \mathcal{E}$,we denote by $m(\sigma)$ the $(d-1)$-Lebesgue measure of σ. If $\sigma \in \mathcal{E}_K$, we note $d_{K,\sigma} = d(x_K, \sigma)$, and we denote by $\tau_{K,\sigma}$ the transmissibility of K through σ, defined by $\tau_{K,\sigma} = \frac{m(\sigma)}{d_{K,\sigma}}$. If $\sigma = K|L$, we note $d_{K,L} = d(x_K, x_L)$ and $\tau_{KL} = \frac{m(\sigma)}{d_{K,L}}$. The size of the mesh is defined by :

$$size(\mathcal{T}) = \max_{K\in\mathcal{T}} diam(K),$$

and a geometrical factor, connected with the regularity of the mesh, is defined by

$$reg(\mathcal{T}) = \max_{K\in\mathcal{T}} \left(\sum_{\sigma = K|L \in \mathcal{E}_{K,{\rm int}}} \frac{m(\sigma) d_{K,L}}{m(K)} \right).$$

Definition 17.2. (Uniform time discretization of $(0,T)$**)** *A uniform time discretization of $(0,T)$ is given by an integer value N and a sequence of real values $(t^n)_{n\in\{0,\dots,N+1\}}$. We define $\delta t = \frac{T}{N+1}$ and, $\forall n \in \{0,\dots,N\}$, $t^n = n\delta t$. Thus we have $t^0 = 0$ and $t^{N+1} = T$.*

Remark 17.1. *We can easily prove all the results of this paper for a general time discretization, but for the sake of simplicity, we choose to only consider uniform time discretizations.*

Definition 17.3. (Space-time discretization of Q_T**)** *A finite volume discretization $\mathcal{D}$ of Q_T is a family*

$$\mathcal{D} = (\mathcal{T}, \mathcal{E}, (x_K)_{K\in\mathcal{T}}, N, (t^n)_{n\in\{0,\dots,N\}}),$$

where $(\mathcal{T}, \mathcal{E}, (x_K)_{K\in\mathcal{T}})$ is an admissible mesh of Ω in the sense of definition 17.1 and $(N, (t^n)_{n\in\{0,\dots,N\}})$ is a discretization of $(0,T)$ in the sense of definition 17.2. For a given mesh $\mathcal{D}$, one defines :

$$size(\mathcal{D}) = max(size(\mathcal{T}), \delta t), \qquad and\ reg(\mathcal{D}) = reg(\mathcal{T}).$$

17.2 Definition of the scheme and main result

For $K \in \mathcal{T}_i$, we denote by $g_K(s) = g_i(s)$ for all function g whose definition depends on the subdomain Ω_i, as for example $\phi_i, \varphi_i, M_i, f_i, Z_i, \dots$.

The total flux balance equation is discretized by

$$\sum_{\sigma\in\mathcal{E}_K} m(\sigma) Q_{K,\sigma}^{n+1} = 0, \qquad \forall n \in \{0,\dots,N\}, \forall K \in \mathcal{T}, \tag{17.1}$$

with

$$Q_{K,\sigma}^n = \begin{cases} \frac{M_{K,L}(s_K^n, s_L^n)}{d_{K,L}} (P_K^n - P_L^n) + \mathcal{R}(\zeta_{K,\sigma}; s_K^n, s_L^n) & \text{if } \sigma = K|L \in \mathcal{E}_{K,i}, \\ \frac{M_K(s_K^n)}{d_{K,\sigma}} \left(P_K^n - P_{K,\sigma}^n\right) + \mathcal{R}\left(\zeta_{K,\sigma}; s_K^n, s_{K,\sigma}^n\right) & \text{if } \sigma \in \mathcal{E}_{K,\Gamma}, \\ 0 & \text{if } \sigma \in \mathcal{E}_{K,{\rm ext}}, \end{cases} \tag{17.2}$$

where $M_{K,L}(s_K^{n+1}, s_L^{n+1}) = M_{L,K}(s_L^{n+1}, s_K^{n+1})$ is a mean value between $M_K(s_K^{n+1})$ and $M_L(s_L^{n+1})$. For example, we can consider, as in [90], the harmonic mean

$$M_{K,L}(s_K^{n+1}, s_L^{n+1}) = \frac{M_K(s_K^{n+1}) M_K(s_L^{n+1}) d_{K,L}}{d_{L,\sigma} M_K(s_K^{n+1}) + d_{K,\sigma} M_K(s_L^{n+1})}.$$

The function $\zeta_{K,\sigma}$ is defined by $\zeta_{K,\sigma}(s) = \zeta_K(s)\mathbf{g}\cdot\mathbf{n}_{K,\sigma}$, where $\mathbf{n}_{K,\sigma}$ denotes the outward normal to σ with respect to K. For a function f, we denote by $\mathcal{R}(f; a, b)$ the Riemann solver

$$\mathcal{R}(f; a, b) = \begin{cases} \min_{c\in[a,b]} f(c) & \text{if } \ a \le b, \\ \max_{c\in[b,a]} f(c) & \text{if } \ b \le a. \end{cases} \tag{17.3}$$

The oil-flux balance equation is discretized as follows :

$$\phi_K \frac{s_K^{n+1} - s_K^n}{\delta t} m(K) + \sum_{\sigma\in\mathcal{E}_K} m(\sigma) F_{K,\sigma}^{n+1} = 0, \tag{17.4}$$

with

$$F_{K,\sigma}^n = \begin{cases} Q_{K,\sigma}^n \ f_K(\overline{s}_{K,\sigma}^n) + \mathcal{R}(\gamma_{K,\sigma}; s_K^n, s_L^n) + \dfrac{\varphi_K(s_K^n) - \varphi_K(s_L^n)}{d_{K,L}} \text{ if } \ \sigma = K|L \in \mathcal{E}_{K,i}, \\ Q_{K,\sigma}^n \ f_K(\overline{s}_{K,\sigma}^n) + \mathcal{R}(\gamma_{K,\sigma}; s_K^n, s_{K,\sigma}^n) + \dfrac{\varphi_K(s_K^n) - \varphi_K(s_{K,\sigma}^n)}{d_{K,\sigma}} \text{ if } \ \sigma \in \mathcal{E}_{K,\Gamma}, \\ 0 \text{ if } \ \sigma \in \mathcal{E}_{K,\text{ext}}, \end{cases} \tag{17.5}$$

where $\gamma_{K,\sigma}(s) = \gamma_K(s)\mathbf{g}\cdot\mathbf{n}_{K,\sigma}$ and $\overline{s}_{K,\sigma}^{n+1}$ is the upwind value defined by

$$\overline{s}_{K,\sigma}^{n+1} = \begin{cases} s_K^{n+1} & \text{if } \ Q_{K,\sigma}^{n+1} \ge 0, \\ s_L^{n+1} & \text{if } \ Q_{K,\sigma}^{n+1} < 0 \text{ and } \sigma = K|L \in \mathcal{E}_{K,i}, \\ s_{K,\sigma}^{n+1} & \text{if } \ Q_{K,\sigma}^{n+1} < 0 \text{ and } \sigma \in \mathcal{E}_{K,\Gamma}. \end{cases} \tag{17.6}$$

The interface values $\left(s_{K,\sigma}^{n+1}, s_{L,\sigma}^{n+1}, P_{K,\sigma}^{n+1}, P_{L,\sigma}^{n+1}\right)$ for $\sigma = K|L \in \mathcal{E}_\Gamma$ are defined by the following nonlinear system. For all $\sigma = K|L \in \mathcal{E}_\Gamma$, for all $n \in \{0, \ldots, N\}$, there exists $\pi_\sigma^{n+1} \in \mathbb{R}$ such that

$$\pi_\sigma^{n+1} \in \tilde{\pi}_K(s_{K,\sigma}^{n+1}) \cap \tilde{\pi}_L(s_{L,\sigma}^{n+1}), \tag{17.7}$$

$$P_{K,\sigma}^{n+1} - Z_K\left(\pi_\sigma^{n+1}\right) = P_{L,\sigma}^{n+1} - Z_L\left(\pi_\sigma^{n+1}\right), \tag{17.8}$$

$$Q_{K,\sigma}^{n+1} + Q_{L,\sigma}^{n+1} = 0, \tag{17.9}$$

$$F_{K,\sigma}^{n+1} + F_{L,\sigma}^{n+1} = 0. \tag{17.10}$$

Moreover, we impose the discrete counterpart of the equation (16.23), that is, for all $n \in \{0, \ldots, N\}$,

$$\sum_{K\in\mathcal{T}_1} m(K) P_K^{n+1} = 0. \tag{17.11}$$

We will show below in Section 17.3 that the system (17.7)-(17.10) possesses a solution. We denote by $\mathcal{X}(\mathcal{D}, i)$ the finite dimensional space of piecewise constant functions $u_\mathcal{D}$ defined almost everywhere in $Q_{i,T}$ having a trace on the interface Γ, i.e.

$$\begin{aligned} \mathcal{X}(\mathcal{D}, i) \ := \ & \big\{u_{\mathcal{D},i} \in L^\infty(Q_{i,T}) \text{ and for all } (K, \sigma, n) \in \mathcal{T} \times \mathcal{E}_\Gamma \times \{0, \ldots, N\}, \\ & \quad u_{\mathcal{D},i} \text{ is constant on } K \times (t^n, t^{n+1}], u_{\mathcal{D},i} \text{ is constant on } \sigma \times (t^n; t^{n+1})\big\}, \end{aligned}$$

and by $\mathcal{X}(\mathcal{D})$ the space of the functions $u_\mathcal{D}$ whose restriction $(u_\mathcal{D})_{|_{\overline{Q}_{i,T}}}$ belongs to $\mathcal{X}(\mathcal{D}, i)$. We define the solution $(s_\mathcal{D}, P_\mathcal{D}) \in \mathcal{X}(\mathcal{D})^2$ of the scheme by

$$s_\mathcal{D}(x,t) = s_K^{n+1}, \quad P_\mathcal{D}(x,t) = P_K^{n+1} \quad \text{if } (x,t) \in K \times (t^n, t^{n+1}],$$

and, for $x \in \sigma = K|L \subset \Gamma$ for some $K \in \mathcal{T}_1$, $L \in \mathcal{T}_2$, for $t \in (t^n, t^{n+1})$, the traces

$$s_{\mathcal{D}_{|\Gamma,1}}(x,t) = s_{K,\sigma}^{n+1}, \qquad s_{\mathcal{D}_{|\Gamma,2}}(x,t) = s_{L,\sigma}^{n+1}.$$

In this paper we prove the following convergence result.

Theorem 1. *Assume that Assumptions 1 and 2 hold. Let $(\mathcal{D}_m)_m$ be a sequence of admissible discretizations of Q_T in the sense of Definition 17.3, then for all $m \in \mathbb{N}$, there exists a discrete solution $(s_{\mathcal{D}_m}, P_{\mathcal{D}_m}) \in \mathcal{X}(\mathcal{D}_m)^2$ to the scheme. Moreover, if* $\lim_{m\to\infty}$ *$size(\mathcal{D}_m) = 0$, and if there exists $\zeta > 0$ such that, for all m, $reg(\mathcal{D}_m) \le \zeta$, then up to a subsequence, $s_{\mathcal{D}_m}$ converges, towards $s \in L^\infty(Q_T; [0,1])$ in the $L^p(Q_T)$ topology for all $p \in [1,\infty)$, $P_{\mathcal{D}_m}$ converges to P weakly in $L^2(Q_T)$, where (s,P) is a weak solution of Problem* (16.16)-(16.23) *in the sense of Definition 16.1.*

17.3 The interface condition system

Define, for all $\sigma = K|L \in \mathcal{E}_\Gamma$, for all $n \in \{0, \dots, N\}$,

$$P_\sigma^{n+1} := P_{K,\sigma}^{n+1} - Z_K\left(\pi_\sigma^{n+1}\right) = P_{L,\sigma}^{n+1} - Z_L(\pi_\sigma^{n+1}), \tag{17.12}$$

and

$$Q_{K,\sigma}^{n+1}(\pi_\sigma^{n+1}) := \alpha_K^{n+1}\left(P_K^{n+1} - P_\sigma^{n+1} - Z_K(\pi_\sigma^{n+1})\right) + \mathcal{R}(\zeta_{K,\sigma}; s_K^{n+1}, \tilde{\pi}_K^{-1}(\pi_\sigma^{n+1})), \tag{17.13}$$

where $\alpha_K^{n+1} = \frac{M_K(s_K^{n+1})}{d_{K,\sigma}}$. Then, the balance of the fluxes on the interface (17.9)–(17.10) can be rewritten as

$$Q_{K,\sigma}^{n+1}(\pi_\sigma^{n+1}) + Q_{L,\sigma}^{n+1}(\pi_\sigma^{n+1}) = 0 \tag{17.14}$$

$$\begin{aligned} &Q_{K,\sigma}^{n+1}(\pi_\sigma^{n+1}) f_K\left(\overline{s}_{K,\sigma}^{n+1}(\pi_\sigma^{n+1})\right) + \mathcal{R}(\gamma_{K,\sigma}; s_K^{n+1}, \tilde{\pi}_K^{-1}(\pi_\sigma^{n+1})) + \frac{\varphi_K(s_K^{n+1}) - \varphi_K \circ \tilde{\pi}_K^{-1}(\pi_\sigma^{n+1})}{d_{K,\sigma}} \\ &\quad + Q_{L,\sigma}^{n+1}(\pi_\sigma^{n+1}) f_L\left(\overline{s}_{L,\sigma}^{n+1}(\pi_\sigma^{n+1})\right) + \mathcal{R}(\gamma_{L,\sigma}; s_L^{n+1}, \tilde{\pi}_L^{-1}(\pi_\sigma^{n+1})) + \frac{\varphi_L(s_L^{n+1}) - \varphi_L \circ \tilde{\pi}_L^{-1}(\pi_\sigma^{n+1})}{d_{L,\sigma}} = 0, \end{aligned} \tag{17.15}$$

where

$$\overline{s}_{K,\sigma}^{n+1}(p) = \begin{cases} s_K^{n+1} & \text{if} \quad Q_{K,\sigma}^{n+1}(p) \ge 0, \\ \pi_K^{-1}(p) & \text{if} \quad Q_{K,\sigma}^{n+1}(p) < 0. \end{cases} \tag{17.16}$$

We deduce from (17.14) that

$$\begin{aligned} P_\sigma^{n+1} &= \frac{\alpha_K^{n+1}(P_K^{n+1} - Z_K(\pi_\sigma^{n+1})) + \alpha_L^{n+1}(P_L^{n+1} - Z_L(\pi_\sigma^{n+1}))}{\alpha_K^{n+1} + \alpha_L^{n+1}} \\ &\quad + \frac{\mathcal{R}(\zeta_{K,\sigma}; s_K^{n+1}, \tilde{\pi}_K^{-1}(\pi_\sigma^{n+1})) + \mathcal{R}(\zeta_{L,\sigma}; s_L^n, \tilde{\pi}_L^{-1}(\pi_\sigma^{n+1}))}{\alpha_K^{n+1} + \alpha_L^{n+1}} \end{aligned} \tag{17.17}$$

and thus that

$$
\begin{aligned}
Q_{K,\sigma}^{n+1}(\pi_\sigma^{n+1}) &= \frac{\alpha_K^{n+1}\alpha_L^{n+1}}{\alpha_K^{n+1}+\alpha_L^{n+1}}\left(P_K^{n+1}-P_L^{n+1}-Z_K(\pi_\sigma^{n+1})+Z_L(\pi_\sigma^{n+1})\right)\\
&\quad+\frac{\alpha_L^{n+1}\mathcal{R}(\zeta_{K,\sigma};s_K^{n+1},\tilde\pi_K^{-1}(\pi_\sigma^{n+1}))-\alpha_K^{n+1}\mathcal{R}(\zeta_{L,\sigma};s_L^{n},\tilde\pi_L^{-1}(\pi_\sigma^{n+1}))}{\alpha_K^{n+1}+\alpha_L^{n+1}}.
\end{aligned}
\tag{17.18}
$$

As a direct consequence of Lemma 16.1, $Q_{K,\sigma}^{n+1}$ belong to $\mathcal{C}^1(\mathbb{R};\mathbb{R})$ and admits finite limits as $p\to\pm\infty$.

Denote by

$$
\begin{aligned}
\Psi_\sigma^{n+1}(p) &:= Q_{K,\sigma}^{n+1}(p)\left(f_K(\overline{s}_{K,\sigma}(p))-f_L(\overline{s}_{L,\sigma}(p))\right)\\
&+\mathcal{R}(\gamma_{K,\sigma};s_K^{n+1},\tilde\pi_K^{-1}(p))+\mathcal{R}(\gamma_{L,\sigma};s_L^{n+1},\tilde\pi_L^{-1}(p))\\
&+\frac{\varphi_K(s_K^{n+1})-\varphi_K\circ\pi_K^{-1}(p)}{d_{K,\sigma}}+\frac{\varphi_L(s_L^{n+1})-\varphi_L\circ\pi_L^{-1}(p)}{d_{L,\sigma}},
\end{aligned}
$$

then Ψ_σ is continuous on $\mathbb{R}$.

Lemma 17.1. *Let $(s_K^{n+1},s_L^{n+1})\in[0,1]^2$, there exists $\pi_\sigma^{n+1}\in[\min_i\pi_i(0),\max_i\pi_i(1)]$ such that $\Psi_\sigma^{n+1}(\pi_\sigma^{n+1})=0$.*

Proof : From the definition (17.16) of $\overline{s}_{K,\sigma}^{n+1}(p)$, since $\lim_{p\to\min_i\pi_i(0)}\pi_K^{-1}(p)=0$, and since $Q_{K,\sigma}^{n+1}(p)$ admits a limit as $p\to\min_i\pi_i(0)$, one has

$$
\lim_{p\to\min_i\pi_i(0)} Q_{K,\sigma}^{n+1}(p)\left(f_K(\overline{s}_{K,\sigma}^{n+1}(p))-f_L(\overline{s}_{L,\sigma}^{n+1}(p))\right)\geq 0
$$

and also

$$
\lim_{p\to\min_i\pi_i(0)}\mathcal{R}(\gamma_{M,\sigma};s_M^{n+1},\tilde\pi_M^{-1}(p))=\max_{s\in[0,s_M]}\gamma_{M,\sigma}(s)\geq 0,\qquad\text{with}\qquad M\in\{K,L\}.
$$

This yields that

$$
\lim_{p\to\min_i\pi_i(0)}\Psi_\sigma^{n+1}(p)\geq\frac{\varphi_K(s_K^{n+1})}{d_{K,\sigma}}+\frac{\varphi_L(s_L^{n+1})}{d_{L,\sigma}}\geq 0.
$$

One obtains similarly that $\lim_{p\to\max_i\pi_i(1)}\Psi_\sigma^{n+1}(p)\leq 0$. One conclude thanks to the continuity of Ψ_σ^{n+1}. □

Proposition 17.2. *Let $\sigma=K|L\in\mathcal{E}_\Gamma$, and let $\left(s_K^{n+1},s_L^{n+1},P_K^{n+1},P_L^{n+1}\right)\in\mathbb{R}^4$, then there exists a solution $\left(\pi_\sigma^{n+1},s_{K,\sigma}^{n+1},s_{L,\sigma}^{n+1},P_{K,\sigma}^{n+1},P_{L,\sigma}^{n+1}\right)\in[\min_i\pi_i(0),\max_i\pi_i(1)]\times[0,1]^2\times\mathbb{R}^2$ to the nonlinear system* (17.7)–(17.10).

Proof : Let $\pi_\sigma^{n+1}\in\mathbb{R}$ be a solution of the equation $\Psi_\sigma^{n+1}(\pi_\sigma^{n+1})=0$, whose existence has been claimed in Lemma 17.1. Firstly, defining $s_{K,\sigma}^{n+1}:=\pi_K^{-1}(\pi_\sigma^{n+1})$ and $s_{L,\sigma}^{n+1}:=\pi_L^{-1}(\pi_\sigma^{n+1})$, one has directly that

$$
\pi_\sigma^{n+1}\in\tilde\pi_K(s_{K,\sigma}^{n+1})\cap\tilde\pi_L(s_{L,\sigma}^{n+1}).
$$

As it was noticed in Lemma 16.1, the function $p \mapsto Z_K(p) - Z_L(p)$ is uniformly bounded. Hence, the values

$$P_{K,\sigma}^{n+1} := \frac{\alpha_K^{n+1} P_K^{n+1} + \alpha_L^{n+1} P_L^{n+1} + \alpha_L^{n+1} \left(Z_K(\pi_\sigma^{n+1}) - Z_L(\pi_\sigma^{n+1}) \right)}{\alpha_K^{n+1} + \alpha_L^{n+1}} + \frac{\mathcal{R}\left(\zeta_{K,\sigma}; s_K^{n+1}, \tilde{\pi}_K^{-1}(\pi_\sigma^{n+1})\right) + \mathcal{R}\left(\zeta_{L,\sigma}; s_L^n, \tilde{\pi}_L^{-1}(\pi_\sigma^{n+1})\right)}{\alpha_K^{n+1} + \alpha_L^{n+1}}$$

and

$$P_{L,\sigma}^{n+1} := \frac{\alpha_K^{n+1} P_K^{n+1} + \alpha_L^{n+1} P_L^{n+1} + \alpha_K^{n+1} \left(Z_L(\pi_\sigma^{n+1}) - Z_K(\pi_\sigma^{n+1}) \right)}{\alpha_K^{n+1} + \alpha_L^{n+1}} + \frac{\mathcal{R}\left(\zeta_{K,\sigma}; s_K^{n+1}, \tilde{\pi}_K^{-1}(\pi_\sigma^{n+1})\right) + \mathcal{R}\left(\zeta_{L,\sigma}; s_L^n, \tilde{\pi}_L^{-1}(\pi_\sigma^{n+1})\right)}{\alpha_K^{n+1} + \alpha_L^{n+1}}$$

are finite. It is now easy to check that $\left(\pi_\sigma^{n+1}, s_{K,\sigma}^{n+1}, s_{L,\sigma}^{n+1}, P_{K,\sigma}^{n+1}, P_{L,\sigma}^{n+1}\right)$ is a solution to the system (17.7)–(17.10) thanks to the analysis carried out above. □

18 *A priori* estimates and existence of a discrete solution

18.1 $\mathbf{L}^\infty(\mathbf{Q_T})$ estimate on the saturation

Proposition 18.1. *Let* $(s_\mathcal{D}, P_\mathcal{D})$ *be a solution to the scheme* (17.1)–(17.11), *then*

$$0 \le s_\mathcal{D} \le 1 \qquad a.e. \text{ in } Q_T. \tag{18.1}$$

Proof : We will prove that for all $K \in \mathcal{T}$, for all $n \in \{0, \dots, N\}$,

$$s_K^{n+1} \le 1.$$

The proof for obtaining $s_K^{n+1} \ge 0$ is similar.

Using the definition (17.5) of $F_{K,\sigma}^{n+1}$, one can rewrite (17.4) under the form

$$H_K\left(s_K^{n+1}, s_K^n, \left(s_L^{n+1}\right)_{L\in\mathcal{N}_K}, \left(s_{K,\sigma}^{n+1}\right)_{\sigma\in\mathcal{E}_{K,\Gamma}}, \left(Q_{K,\sigma}^{n+1}\right)_{\sigma\in\mathcal{E}_K}\right) = 0, \tag{18.2}$$

where H_K is non increasing with respect to s_K^n, $\left(s_L^{n+1}\right)_{L\in\mathcal{N}_K}$, $\left(s_{K,\sigma}^{n+1}\right)_{\sigma\in\mathcal{E}_{K,\Gamma}}$. Making use of the notations $a\top b = \max(a,b)$, we obtain that

$$H_K\left(s_K^{n+1}, s_K^n\top 1, \left(s_L^{n+1}\top 1\right)_{L\in\mathcal{N}_K}, \left(s_{K,\sigma}^{n+1}\top 1\right)_{\sigma\in\mathcal{E}_{K,\Gamma}}, \left(Q_{K,\sigma}^{n+1}\right)_{\sigma\in\mathcal{E}_K}\right) \le 0.$$

We remark that for all $K \in \mathcal{T}$ and for all $s \in [0,1]$ one has

$$\sum_{\sigma\in\mathcal{E}_K} m(\sigma)\gamma_{K,\sigma}(s) = 0. \tag{18.3}$$

Combining (18.3) and (17.1) we have

$$H_K\left(1, 1, (1)_{L\in\mathcal{N}_K}, (1)_{\sigma\in\mathcal{E}_{K,i}}, \left(Q_{K,\sigma}^{n+1}\right)_{\sigma\in\mathcal{E}_K}\right) = 0.$$

Hence, using once again the monotonicity of H_K, one obtains

$$H_K\left(1, s_K^n \top 1, \left(s_L^{n+1}\top 1\right)_{L\in\mathcal{N}_K}, \left(s_{K,\sigma}^{n+1}\top 1\right)_{\sigma\in\mathcal{E}_{K,\Gamma}}, \left(Q_{K,\sigma}^{n+1}\right)_{\sigma\in\mathcal{E}_K}\right) \le 0.$$

Since $a\top b$ is either equal to a or to b, one has

$$H_K\left(s_K^{n+1}\top 1, s_K^n \top 1, \left(s_L^{n+1}\top 1\right)_{L\in\mathcal{N}_K}, \left(s_{K,\sigma}^{n+1}\top 1\right)_{\sigma\in\mathcal{E}_{K,\Gamma}}, \left(Q_{K,\sigma}^{n+1}\right)_{\sigma\in\mathcal{E}_K}\right) \le 0. \tag{18.4}$$

Next we remark that for any $\sigma = K|L \in \mathcal{E}_\Gamma$ the the equation (17.10) can be written as

$$H_\sigma\left(s_K^{n+1}, s_L^{n+1}, \left(s_M^{n+1}\right)_{M\in\mathcal{N}_K\cup\mathcal{N}_L}, \left(s_{M,\sigma}^{n+1}\right)_{M\in\mathcal{N}_K\cup\mathcal{N}_L}, \left(Q_{M,\sigma}^{n+1}\right)_{M\in\mathcal{N}_K\cup\mathcal{N}_L}\right) = 0,$$

where H_σ is non increasing with respect to $\left(s_M^{n+1}\right)_{M\in\mathcal{N}_K\cup\mathcal{N}_L}, \left(s_{M,\sigma}^{n+1}\right)_{M\in\mathcal{N}_K\cup\mathcal{N}_L}$. Thanks to (17.9) and using $\gamma_i(1) = 0$ for $i \in \{1, 2\}$ we obtain

$$H_\sigma\left(1, 1, (1)_{M\in\mathcal{N}_K\cup\mathcal{N}_L}, (1)_{M\in\mathcal{N}_K\cup\mathcal{N}_L}, \left(Q_{M,\sigma}^{n+1}\right)_{M\in\mathcal{N}_K\cup\mathcal{N}_L}\right) = 0.$$

Using the same arguments as for (18.4) one has that

$$H_\sigma\left(s_K^{n+1}\top 1, s_L^{n+1}\top 1, \left(s_M^{n+1}\top 1\right)_{M\in\mathcal{N}_K\cup\mathcal{N}_L}, \left(s_{M,\sigma}^{n+1}\top 1\right)_{M\in\mathcal{N}_K\cup\mathcal{N}_L}, \left(Q_{M,\sigma}^{n+1}\right)_{M\in\mathcal{N}_K\cup\mathcal{N}_L}\right) \le 0. \tag{18.5}$$

Multiplying (18.4) by δt and summing over $K \in \mathcal{T}$ provides, using (18.5) and the conservativity of the scheme,

$$\sum_{K\in\mathcal{T}} \phi_K (s_K^{n+1} - 1)^+ m(K) \le \sum_{K\in\mathcal{T}} \phi_K (s_K^n - 1)^+ m(K).$$

Since $s_0 \in L^\infty(Q_T; [0,1])$, $s_K^0 \in [0,1]$ for all $K \in \mathcal{T}$. A straightforward induction allows us to conclude. $\square$

18.2 Energy estimate

Definition 18.1. *We define the discrete $L^2(0,T;H^1(\Omega_i))$ semi-norm of an element $u_\mathcal{D} \in \mathcal{X}(\mathcal{D}, i)$ by*

$$|u_\mathcal{D}|_{\mathcal{D},i}^2 := \sum_n \delta t \sum_{\sigma=K|L\in\mathcal{E}_i} \tau_{KL}\left(u_K^{n+1} - u_L^{n+1}\right)^2 + \sum_n \delta t \sum_{K\in\mathcal{T}_i} \sum_{\sigma\in\mathcal{E}_{K,\Gamma}} \tau_{K\sigma}\left(u_K^{n+1} - u_{K,\sigma}^{n+1}\right)^2.$$

Lemma 18.2. *The following inequalities hold :*

– *for all $\sigma = K|L \in \mathcal{E}_{\text{int}}$,*

$$Q_{K,\sigma}^{n+1} f_K\left(\overline{s}_{K,\sigma}^{n+1}\right)\left(\pi_K(s_K^{n+1}) - \pi_K(s_L^{n+1})\right) \ge Q_{K,\sigma}^{n+1}\left(Z_K(\pi_K(s_K^{n+1})) - Z_K(\pi_K(s_L^{n+1}))\right); \tag{18.6}$$

– *for all $\sigma \in \mathcal{E}_{K,\Gamma}$,*

$$Q_{K,\sigma}^{n+1} f_K\left(\overline{s}_{K,\sigma}^{n+1}\right)\left(\pi_K(s_K^{n+1}) - \pi_\sigma^{n+1})\right) \ge Q_{K,\sigma}^{n+1}\left(Z_K(\pi_K(s_K^{n+1})) - Z_K(\pi_\sigma^{n+1})\right). \tag{18.7}$$

Proof : Since $f_K \circ \pi_K^{-1}$ is a non decreasing function, then function $Z_K : p \mapsto \int_0^p f_K \circ \pi_K^{-1}(a)da$ is convex, so that for all $(a, b) \in \mathbb{R}^2$,

$$f_K \circ \pi_K^{-1}(a)\,(b-a) \le Z_K(b) - Z_K(a) \le f_K \circ \pi_K^{-1}(b)\,(b-a)\,.$$

The inequalities (18.6) and (18.7) follow from the definition (17.6) of $\overline{s}_{K,\sigma}^{n+1}$. □

Lemma 18.3. *Let us define*

$$\mathcal{G}_{K,\sigma}(p) := \int_0^p \gamma_{K,\sigma}\left(\tilde{\pi}_K^{-1}(\tau)\right) d\tau \tag{18.8}$$

for all $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_K$. Then, the following estimates hold :
– *for all $\sigma = K|L \in \mathcal{E}_{\text{int}}$,*

$$\mathcal{R}(\gamma_{K,\sigma}; s_K^{n+1}, s_L^{n+1})\left(\pi_K(s_K^{n+1}) - \pi_K(s_L^{n+1})\right) \ge \mathcal{G}_{K,\sigma}(\pi_K(s_K^{n+1})) - \mathcal{G}_{K,\sigma}(\pi_K(s_L^{n+1})) \tag{18.9}$$

– *for all $\sigma \in \mathcal{E}_{K,\Gamma}$,*

$$\mathcal{R}(\gamma_{K,\sigma}; s_K^{n+1}, s_{K,\sigma}^{n+1})\left(\pi_K(s_K^{n+1}) - \pi_\sigma^{n+1}\right) \ge \mathcal{G}_{K,\sigma}(\pi_K(s_K^{n+1})) - \mathcal{G}_{K,\sigma}(\pi_\sigma^{n+1}). \tag{18.10}$$

Proof : For any $a, b \in \mathbb{R}$ one has

$$\begin{aligned}\mathcal{R}(\gamma_{K,\sigma}; \tilde{\pi}_K^{-1}(a), \tilde{\pi}_K^{-1}(b))\,(a-b)) &= \int_b^a \gamma_{K,\sigma}\left(\tilde{\pi}_K^{-1}(p)\right) dp \\ &+ \int_b^a \mathcal{R}\left(\gamma_{K,\sigma}; \tilde{\pi}_K^{-1}(a), \tilde{\pi}_K^{-1}(b)\right) - \gamma_{K,\sigma}\left(\tilde{\pi}_K^{-1}(p)\right) dp.\end{aligned} \tag{18.11}$$

We only have to remark that in view of (17.3) the last term in the right hand side of (18.11) is positive. □

Lemma 18.4. *For all $K \in \mathcal{T}$, for all $n \in \{0, \ldots, N\}$ and for all $\sigma \in \mathcal{E}_{K,\Gamma}$, one has*

$$\begin{aligned}&\left(\varphi_K(s_K^{n+1}) - \varphi_K(s_{K,\sigma}^{n+1})\right)\left(\pi_K(s_K^{n+1}) - \pi_\sigma^{n+1}\right) \\ &\qquad \ge \left(\varphi_K(s_K^{n+1}) - \varphi_K(s_{K,\sigma}^{n+1})\right)\left(\pi_K(s_K^{n+1}) - \pi_K(s_{K,\sigma}^{n+1})\right).\end{aligned} \tag{18.12}$$

Proof : Assume that $s_{K,\sigma}^{n+1} \in (0,1)$, then $\tilde{\pi}_K(s_{K,\sigma}^{n+1}) = \{\pi_K(s_{K,\sigma}^{n+1})\}$, thus the inequality (18.12) is in fact an equality. Assume now that $s_{K,\sigma}^{n+1} = 0$, then $\pi_\sigma^{n+1} \le \pi_K(s_{K,\sigma}^{n+1}) \le \pi_K(s_K^{n+1})$, and $\varphi_K(s_{K,\sigma}^{n+1}) \le \varphi_K(s_K^{n+1})$. The inequality (18.12) follows. Similarly, if $s_{K,\sigma}^{n+1} = 1$, then $\pi_\sigma^{n+1} \ge \pi_K(s_{K,\sigma}^{n+1}) \ge \pi_K(s_K^{n+1})$, and $\varphi_K(s_{K,\sigma}^{n+1}) \ge \varphi_K(s_K^{n+1})$, leading also to (18.12). □

Proposition 18.5. *There exists C_1, depending only on α_M, $\min_i K_i$, μ_o, μ_w , $\max_i \|\pi_i\|_{L^1((0,1))}$ and Ω, such that*

$$\sum_{i\in\{1,2\}} \left(|P_\mathcal{D}|_{\mathcal{D},i}^2 + |\varphi(s_\mathcal{D})|_{\mathcal{D},i}^2\right) \le C_1. \tag{18.13}$$

Proof : Multiplying the equation (17.4) by $\delta t \pi_K(s_K^{n+1})$ and summing over $K \in \mathcal{T}$ and $n \in \{0, \ldots, N\}$ yields, after reorganizing the sum,

$$A + B = 0, \tag{18.14}$$

where

$$\begin{aligned}
A &= \sum_{n=0}^{N} \sum_{K\in\mathcal{T}} \phi_K \pi_K(s_K^{n+1}) \left(s_K^{n+1} - s_K^n\right) m(K),\\
B &= \sum_{n=0}^{N} \delta t \sum_{\sigma=K|L\in\mathcal{E}_{\rm int}} m(\sigma) F_{K,\sigma}^{n+1} \left(\pi_K(s_K^{n+1}) - \pi_K(s_L^{n+1})\right)\\
&\quad + \sum_{n=0}^{N} \delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_{K,\Gamma}} m(\sigma) F_{K,\sigma}^{n+1} \left(\pi_K(s_K^{n+1}) - \pi_\sigma^{n+1}\right),
\end{aligned}$$

where we have used (17.10). The definition (17.5) of $F_{K,\sigma}^{n+1}$ gives

$$B = B_1 + B_2 + B_3, \tag{18.15}$$

where

$$\begin{aligned}
B_1 &= \sum_{n=0}^{N} \delta t \sum_{\sigma=K|L\in\mathcal{E}_{\rm int}} m(\sigma) Q_{K,\sigma}^{n+1} f_K(\overline{s}_{K,\sigma}^{n+1}) \left(\pi_K(s_K^{n+1}) - \pi_K(s_L^{n+1})\right)\\
&\quad + \sum_{n=0}^{N} \delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_{K,\Gamma}} m(\sigma) Q_{K,\sigma}^{n+1} f_K(\overline{s}_{K,\sigma}^{n+1}) \left(\pi_K(s_K^{n+1}) - \pi_\sigma^{n+1}\right),\\
B_2 &= \sum_{n=0}^{N} \delta t \sum_{\sigma=K|L\in\mathcal{E}_{\rm int}} m(\sigma) \mathcal{R}(\gamma_{K,\sigma}; s_K^{n+1}, s_L^{n+1}) \left(\pi_K(s_K^{n+1}) - \pi_K(s_L^{n+1})\right)\\
&\quad + \sum_{n=0}^{N} \delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_{K,\Gamma}} m(\sigma) \mathcal{R}(\gamma_{K,\sigma}; s_K^{n+1}, s_{K,\sigma}^{n+1}) \left(\pi_K(s_K^{n+1}) - \pi_\sigma^{n+1}\right),\\
B_3 &= \sum_{n=0}^{N} \delta t \sum_{\sigma=K|L\in\mathcal{E}_{\rm int}} \tau_{KL} \left(\varphi_K(s_K^{n+1}) - \varphi_K(s_L^{n+1})\right) \left(\pi_K(s_K^{n+1}) - \pi_K(s_L^{n+1})\right)\\
&\quad + \sum_{n=0}^{N} \delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_{K,\Gamma}} \tau_{K\sigma} \left(\varphi_K(s_K^{n+1}) - \varphi_K(s_{K,\sigma}^{n+1})\right) \left(\pi_K(s_K^{n+1}) - \pi_\sigma^{n+1})\right).
\end{aligned}$$

It follows from Lemma 18.2 that

$$\begin{aligned}
B_1 &\geq \sum_{n=0}^{N} \delta t \sum_{\sigma=K|L\in\mathcal{E}_{\rm int}} m(\sigma) Q_{K,\sigma}^{n+1} \left(Z_K(\pi_K(s_K^{n+1})) - Z_K(\pi_K(s_L^{n+1}))\right)\\
&\quad + \sum_{n=0}^{N} \delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_{K,\Gamma}} m(\sigma) Q_{K,\sigma}^{n+1} \left(Z_K(\pi_K(s_K^{n+1})) - Z_K(\pi_\sigma^{n+1})\right).
\end{aligned}$$

Multiplying the equation (17.1) by $\delta t \left(P_K^{n+1} - Z_K(\pi_K(s_K^{n+1}))\right)$ and summing over $K \in \mathcal{T}$

and $n \in \{0, \dots, N\}$ yields, after reorganizing the sum and using (17.8) and (17.9),

$$\sum_{n=0}^{N} \delta t \left(\sum_{\sigma=K|L \in \mathcal{E}_{\text{int}}} m(\sigma) Q_{K,\sigma}^{n+1} (P_K^{n+1} - P_L^{n+1}) + \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_{K,\Gamma}} m(\sigma) Q_{K,\sigma}^{n+1} (P_K^{n+1} - P_{K,\sigma}^{n+1}) \right)$$
$$= \sum_{n=0}^{N} \delta t \sum_{\sigma=K|L \in \mathcal{E}_{\text{int}}} m(\sigma) Q_{K,\sigma}^{n+1} \left(Z_K(\pi_K((s_K^{n+1})) - Z_K(\pi_K(s_L^{n+1}))\right)$$
$$+ \sum_{n=0}^{N} \delta t \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_{K,\Gamma}} m(\sigma) Q_{K,\sigma}^{n+1} \left(Z_K(\pi_K(s_{K,}^{n+1})) - Z_K(\pi_\sigma^{n+1})\right).$$

Therefore, using the definition (17.2) of $Q_{K,\sigma}^n$, we deduce that

$$B_1 \geq B_4 + B_5, \tag{18.16}$$

where

$$B_4 = \sum_{n=0}^{N} \delta t \left(\sum_{\sigma=K|L \in \mathcal{E}_{\text{int}}} \frac{m(\sigma) M_{K,L}}{d_{K,L}} (P_K^{n+1} - P_L^{n+1})^2 + \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_{K,\Gamma}} \frac{m(\sigma) M_K}{d_{K,\sigma}} (P_K^{n+1} - P_{K,\sigma}^{n+1})^2 \right)$$

and

$$\begin{aligned} B_5 \; &= \sum_{n=0}^{N} \delta t \sum_{\sigma=K|L \in \mathcal{E}_{\text{int}}} m(\sigma) \mathcal{R}(\zeta_{K,\sigma}; s_K^{n+1}, s_L^{n+1})(P_K^{n+1} - P_L^{n+1}) \\ &+ \sum_{n=0}^{N} \delta t \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_{K,\Gamma}} m(\sigma) \mathcal{R}(\zeta_{K,\sigma}; s_K^{n+1}, s_{K,\sigma}^{n+1})(P_K^{n+1} - P_{K,\sigma}^{n+1}). \end{aligned} \tag{18.17}$$

Using the fact that for all $s \in \mathbb{R}$, $M_i(s) \geq \alpha_M > 0$ we obtain

$$B_4 \geq \alpha_M \sum_{i \in \{1,2\}} |P_{\mathcal{D}}|_{\mathcal{D},i}^2 . \tag{18.18}$$

The Cauchy–Schwarz inequality applied to the right hand side of (18.17) implies

$$\begin{aligned} |B_5| \; \leq & \left(\sum_{n=0}^{N} \delta t \sum_{\sigma=K|L \in \mathcal{E}_{\text{int}}} m(\sigma) d_{K,L} \mathcal{R}(\zeta_{K,\sigma}; s_K^{n+1}, s_L^{n+1})^2 \right)^{\frac{1}{2}} \\ & \qquad \cdot \left(\sum_{n=0}^{N} \delta t \sum_{\sigma=K|L \in \mathcal{E}_{\text{int}}} \frac{m(\sigma)}{d_{K,L}} (P_K^{n+1} - P_L^{n+1})^2 \right)^{\frac{1}{2}} \\ & + \left(\sum_{n=0}^{N} \delta t \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_{K,\Gamma}} m(\sigma) d_{K,\sigma} \mathcal{R}(\zeta_{K,\sigma}; s_K^{n+1}, s_{K,\sigma}^{n+1})^2 \right)^{\frac{1}{2}} \\ & \qquad \cdot \left(\sum_{n=0}^{N} \delta t \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_{K,\Gamma}} \frac{m(\sigma)}{d_{K,\sigma}} (P_K^{n+1} - P_{K,\sigma}^{n+1})^2 \right)^{\frac{1}{2}} . \end{aligned}$$

Hence,

$$B_5^2 \leq \frac{3}{2} T|\mathbf{g}|^2 d \sum_{i\in\{1,2\}} m(\Omega_i)\|\zeta_i\|^2_{L^\infty((0,1))} \sum_{i\in\{1,2\}} |P_{\mathcal{D}}|^2_{\mathcal{D},i}\,. \tag{18.19}$$

We recall that d stands for the spatial dimension of Ω. Combining (18.16), (18.18) and (18.19) one has

$$B_1 \geq \alpha_M \sum_{i\in\{1,2\}} |P_{\mathcal{D}}|^2_{\mathcal{D},i} - \left(\frac{3}{2} T|\mathbf{g}|^2 d \sum_{i\in\{1,2\}} m(\Omega_i)\|\zeta_i\|^2_{L^\infty((0,1))}\right)^{\frac{1}{2}} \left(\sum_{i\in\{1,2\}} |P_{\mathcal{D}}|^2_{\mathcal{D},i}\right)^{\frac{1}{2}}. \tag{18.20}$$

We now will show the estimates on the term B_2. Using Lemma 18.3 we have

$$\begin{aligned} B_2 \;&\geq \sum_{n=0}^{N} \delta t \sum_{\sigma=K|L\in\mathcal{E}_{\rm int}} m(\sigma)\left(\mathcal{G}_{K,\sigma}\left(\pi_K(s_K^{n+1})\right) - \mathcal{G}_{K,\sigma}\left(\pi_K(s_L^{n+1})\right)\right) \\ &+\sum_{n=0}^{N} \delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_{K,\Gamma}} m(\sigma)\left(\mathcal{G}_{K,\sigma}\left(\pi_K(s_K^{n+1})\right) - \mathcal{G}_{K,\sigma}\left(\pi_\sigma^{n+1}\right)\right) \end{aligned} \tag{18.21}$$

Recombining terms we obtain

$$\begin{aligned} B_2 \;&\geq \sum_{n=0}^{N} \delta t \sum_{K\in\mathcal{T}} \sum_{\mathcal{E}_{K,\rm int}} m(\sigma)\mathcal{G}_{K,\sigma}\left(\pi_K(s_K^{n+1})\right) \\ &+\sum_{n=0}^{N} \delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_{K,\Gamma}} m(\sigma)\left(\mathcal{G}_{K,\sigma}\left(\pi_K(s_K^{n+1})\right) - \mathcal{G}_{K,\sigma}\left(\pi_\sigma^{n+1}\right)\right) \end{aligned}$$

which in view of (18.8) and (18.3) implies

$$B_2 \geq -\sum_{n=0}^{N} \delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_{K,\Gamma}} m(\sigma)\mathcal{G}_{K,\sigma}\left(\pi_\sigma^{n+1}\right).$$

Remark that if $\sigma = K|L \in \mathcal{E}_\Gamma$ then the function $\mathcal{G}_{K,\sigma}(p)+\mathcal{G}_{L,\sigma}(p)$ in general is not equal to zero. However we can write an lower bound for the term B_2. Indeed, comparing the definition (16.15) of φ_i with the definition (16.17) of γ_i, and using the fact that $\gamma_i(0) = 0$ and $\gamma_i(1) = 0$ one has

$$\int_0^{\pi_\sigma^n} \gamma_K \circ \pi_K^{-1}(p)dp = \int_0^{s_{K,\sigma}^n} \gamma_K(a)\pi_K'(a)da = (\rho_o - \rho_w)\varphi_K(s_{K,\sigma}^n)$$

and thus, in view of Proposition 18.1

$$B_2 \geq -|\rho_o - \rho_w||\mathbf{g}| \max_{i\in\{1,2\}} \varphi_i(1)\; m(\Gamma)T.$$

Because of the definition (16.15) of the function φ_i, then, for all $(a,b) \in [0,1]^2$,

$$(\varphi_i(a) - \varphi_i(b))(\pi_i(a) - \pi_i(b)) \geq \frac{\max(\mu_o,\mu_w)}{K_i}\left(\varphi_i(a) - \varphi_i(b)\right)^2. \tag{18.22}$$

Then it follows from Lemma 18.4 and for inequality (18.22) that

$$B_3 \geq \frac{\max(\mu_o, \mu_w)}{\min_{i\in\{1,2\}} K_i} \sum_{i\in\{1,2\}} |\varphi_i(s_{\mathcal{D}})|^2_{\mathcal{D},i} \,. \tag{18.23}$$

We define $\Pi_i(s) = \int_0^s \pi_i(a)da$, then Π_i is a continuous convex function. As a consequence, for all $(a,b) \in [0,1]^2$,

$$\pi_i(b)(b-a) \geq \Pi_i(b) - \Pi_i(a).$$

Therefore,

$$A \geq \sum_{n=0}^{N} \sum_{K\in\mathcal{T}} \phi_K \left(\Pi_K(s_K^{n+1}) - \Pi_K(s_K^n)\right) m(K) = \sum_{K\in\mathcal{T}} \phi_K \left(\Pi_K(s_K^{N+1}) - \Pi_K(s_K^0)\right) m(K).$$

Using the fact that, for all $(a,b) \in [0,1]^2$, one has

$$\Pi_i(b) - \Pi_i(a) = \int_a^b \pi_i(u)du \geq -\int_0^1 |\pi_i(u)|du,$$

it follows from Proposition 18.1 that

$$A \geq -\sum_{i\in\{1,2\}} \phi_i m(\Omega_i) \|\pi_i\|_{L^1((0,1))}. \tag{18.24}$$

Taking (18.20), (18.23), (18.23) and (18.24) into account in (18.14) we have.

$$\begin{aligned} \alpha_M \sum_{i\in\{1,2\}} |P_{\mathcal{D}}|^2_{\mathcal{D},i} \; &- \left(\frac{3}{2}\frac{T|\mathbf{g}|^2}{d} \sum_{i\in\{1,2\}} m(\Omega_i)\|\zeta_i\|^2_{L^\infty((0,1))}\right)^{\frac{1}{2}} \left(\sum_{i\in\{1,2\}} |P_{\mathcal{D}}|^2_{\mathcal{D},i}\right)^{\frac{1}{2}} \\ &+ \frac{\max(\mu_o, \mu_w)}{\min_{i\in\{1,2\}} K_i} \sum_{i\in\{1,2\}} |\varphi_i(s_{\mathcal{D}})|^2_{\mathcal{D},i} \leq C, \end{aligned} \tag{18.25}$$

where Applying Young's inequality to (18.21) we complete the proof of Proposition 18.5. Indeed,

$$\frac{\alpha_M}{2} \sum_{i\in\{1,2\}} |P_{\mathcal{D}}|^2_{\mathcal{D},i} + \frac{\max(\mu_o, \mu_w)}{\min_{i\in\{1,2\}} K_i} \sum_{i\in\{1,2\}} |\varphi_i(s_{\mathcal{D}})|^2_{\mathcal{D},i} \leq C,$$

□

Proposition 18.6. *There exists C only depending on Ω_i, C_1 and $\|Z_1 - Z_2\|_\infty$ such that*

$$\|P_{\mathcal{D}}\|_{L^2(Q_T)} \leq C.$$

Proof : In view of the discrete Poincaré-Wirtinger inequality [76] and Proposition 18.5, there exists C depending only on Ω_i and C_1 such that

$$\iint_{Q_{i,T}} \left(P_{\mathcal{D}} - m_{\Omega_i}(P_{\mathcal{D}})\right)^2 \, d\mathbf{x}dt \leq C.$$

In order to conclude the proof, it only remains to check that $m_{\Omega_2}(P_{\mathcal{D}})$ is uniformly bounded.

□

18.3 Existence of a discrete solution

Proposition 18.7. *There exists (at least) a solution to the scheme* (17.4)-(17.11).

Proof : The proof is based on a topological degree argument (see for example [55]). For $\nu \in [0,1]$, we introduce the functions

- $f_i^\nu(s) = \nu f_i(s) + (1-\nu)s$,
- $\zeta_i^\nu(s) = \nu\zeta_i(s), \quad \gamma_i^\nu(s) = \nu\gamma_i(s)$
- $M_i^\nu(s) = \nu M_i(s) + (1-\nu)\alpha_M$,
- $\lambda_i^\nu(s) = \nu\lambda_i(s) + (1-\nu)\alpha_M s(1-s)$,
- $\pi_i^\nu(s) = \nu\pi_i(s) + (1-\nu)\pi_1(s)$,
- $\varphi_i^\nu(s) = \displaystyle\int_0^s \lambda_i^\nu(a)\,(\pi_i^\nu)'(a)da$,
- $Z_i^\nu(s) = \displaystyle\int_{s^\star}^s f_i^\nu(a)\,(\pi_i^\nu)'(a)da$.

We denote by $(s_{\mathcal{D}}^\nu, P_{\mathcal{D}}^\nu)$ the solution to the modified scheme. For $\nu = 0$, the problem becomes homogeneous, corresponding to the equations

$$\begin{cases} \partial_t s^0 - \mathrm{div}\,(s^0\nabla P^0 - \nabla\varphi^0(s^0)) = 0, \\ -\alpha_M \Delta P^0 = 0. \end{cases} \tag{18.26}$$

The pressure equation provides a classical linear finite volume scheme which is completely uncoupled from the saturation equation. The transmission conditions (17.9),(17.8) turn to

$$P_{K,\sigma}^{n+1,0} = P_{L,\sigma}^{n+1,0} = \frac{\tau_{K\sigma}P_K^{n+1,0} + \tau_{L\sigma}P_L^{n+1,0}}{\tau_{K\sigma} + \tau_{L\sigma}},$$

and thus

$$Q_{K,\sigma}^{n+1,0} = \tau_{KL}\left(P_K^{n+1,0} - P_L^{n+1,0}\right).$$

Note that the *a priori* estimates (18.1) and (18.13) still hold for $(s_{\mathcal{D}}^\nu, P_{\mathcal{D}}^\nu)$ instead of $(s_{\mathcal{D}}, P_{\mathcal{D}})$. We introduce now a new parameter $\eta \in [0,1]$, and we approximate the problem

$$\begin{cases} \partial_t s^{0,\eta} - \eta\,\mathrm{div}\,(s^{0,\eta}\nabla P^0 - \nabla\varphi^0(s^{0,\eta})) = 0, \\ -\alpha_M \Delta P^0 = 0. \end{cases}$$

The corresponding discrete solution $s_{\mathcal{D}}^{0,\eta}$ satisfies

$$0 \le s_{\mathcal{D}}^{0,\eta} \le 1, \quad \forall \eta \in [0,1]. \tag{18.27}$$

We introduce the compact set

$$\mathcal{K} = \left\{(u_{\mathcal{D}}, v_{\mathcal{D}}) \in (\mathcal{X}(\mathcal{D}))^2 \;\middle|\; \|u_{\mathcal{D}}\|_\infty \le 2 \text{ and } |v_{\mathcal{D}}|_{\mathcal{D}} \le 2C_1\right\},$$

where C_1 is the quantity introduced in Proposition 18.5. Since, for $\nu = \eta = 0$, the problem turns to an invertible linear problem, we can claim that the corresponding topological degree is equal to ± 1 (following the sign of the determinant of the underlying matrix). One can let first η go to 1, and thanks to (18.13),(18.27), $\left(s_{\mathcal{D}}^{0,\eta}, P_{\mathcal{D}}^0\right)$ never belongs to the boundary $\partial\mathcal{K}$ of $\mathcal{K}$. Hence, the topological degree is constant for $\eta \in [0,1]$, and, for $\eta = 1$, the discrete counterpart of (18.26) admits at least a solution. Letting then ν tend to 1 provides thanks to similar arguments the existence of a solution to the scheme (17.4)-(17.7). $\square$

19 Compactness properties of the discrete solution

In order to prove the convergence of the scheme, we will use the method presented in [65] to derive the relative compactness of the sequencies $(s_{\mathcal{D}_m})_{m\in\mathbb{N}}$ and $(P_{\mathcal{D}_m})_{m\in\mathbb{N}}$, where $(\mathcal{D}_m)_{m\in\mathbb{N}}$ is a sequence of admissible discretizations of $\Omega\times(0,T)$ in the sense of Definition 17.3, for which the discretization parameter $h_m := \text{size}(\mathcal{D}_m)$ tends to 0 as $m\to\infty$, while the regularity parameter $\text{reg}(\mathcal{D}_m)$ remains bounded.

Firstly, since $0\le s_{\mathcal{D}_m}\le 1$ almost everywhere in Q_T, we can claim that there exists $s\in L^\infty(Q_T;[0,1])$, such that, up to a subsequence,

$$s_{\mathcal{D}_m} \rightharpoonup s \text{ in the } L^\infty(Q_T) \text{ weak-}\star\text{sense as } m\to\infty.$$

This is of course not sufficient to pass to the limit, so that we seek for additional compactness on the family of approximate solutions $(s_{\mathcal{D}_m}, P_{\mathcal{D}_m})_m$.

19.1 Estimates on space and time translates

We recall here the lemmas 4.2 and 4.3 of [65].

Lemma 19.1. *Let $u_\mathcal{D}$ be an element of $\mathcal{X}(\mathcal{D})$, then for all $\xi\in\mathbb{R}^d$,*

$$\int_0^T\int_{\Omega_{i,\xi}} \left(u_\mathcal{D}(x+\xi,t)-u_\mathcal{D}(x,t)\right)^2\, d\mathbf{x}dt \le |u_\mathcal{D}|^2_{\mathcal{D},i}|\xi|\left(|\xi|+2\text{size}(\mathcal{D})\right),$$

where $\Omega_{i,\xi}=\{x\in\Omega_i \mid [x,x+\xi]\subset\Omega_i\}$.

Lemma 19.2. *Let $u_\mathcal{D}$ be an element of $\mathcal{X}(\mathcal{D})$, and let $T_i(u_\mathcal{D})$ the function of $L^2(\mathbb{R}^{d+1})$ defined by*

$$T_i(u_\mathcal{D})(x,t) = \begin{cases} u_\mathcal{D}(x,t) & \text{if } (x,t)\in\Omega_i\times(0,T),\\ 0 & \text{otherwise,}\end{cases}$$

then for all $\xi\in\mathbb{R}^d$,

$$\int_0^T\int_{\mathbb{R}^d} \left(T_i(u_\mathcal{D})(x+\xi,t)-T_i(u_\mathcal{D})(x,t)\right)^2\, d\mathbf{x}dt \le |u_\mathcal{D}|^2_{\mathcal{D},i}|\xi|\left(|\xi|+2\text{size}(\mathcal{D})+2m(\partial\Omega_i)\,\|u_\mathcal{D}\|_\infty\right),$$

where $\Omega_{i,\xi}=\{x\in\Omega_i \mid [x,x+\xi]\subset\Omega_i\}$.

Lemma 19.3. *There exists C_3, which does not depend on size($\mathcal{T}$), δt nor on τ such that for all $\tau\in(0,T)$,*

$$\int_0^{T-\tau}\sum_{i\in\{1,2\}}\int_{\Omega_i} \left(\varphi_i(s_\mathcal{D})(x,t+\tau)-\varphi_i(s_\mathcal{D})(\mathbf{x},t)\right)^2\, d\mathbf{x}dt \le C_3\tau. \tag{19.1}$$

Lemma 19.3 is an extension of Lemma 4.6 of [65] (see also Proposition 5.1 in [71]).

Proposition 19.4. *The sequence $(\varphi_i(s_{\mathcal{D}_m}))_m$ converges strongly in $L^2(Q_{i,T})$, up to a subsequence, towards the function $\varphi_i(s)\in L^2(0,T;H^1(\Omega_i))$.*

Proof : First recall that by Proposition 18.1 $(\varphi_i(s_{\mathcal{D}_m}))_m$ is bounded in $L^\infty(Q_{i,T})$ for $i \in \{1,2\}$ and that by Proposition 18.5 the sequence $(|\varphi_i(s_{\mathcal{D}_m})|_{\mathcal{D}_m,i})_m$ is bounded. Thanks to the lemmas 19.2 and 19.3 and the Kolmogorov compactness criterion (see e.g. [27] or [65, Theorem 3.9]), it follows that $(T_i(\varphi_i(s_{\mathcal{D}_m})))_m$ is relatively compact in $L^2(\mathbb{R}^{d+1})$ for $i \in \{1,2\}$. Thus we can extract a subsequence, which we denote again denoted by $(T_i(\varphi_i(s_{\mathcal{D}_m})))_m$ such that both $T_1(\phi_1(s_{\mathcal{D}_m}))$ and $T_2(\phi_2(s_{\mathcal{D}_m}))$ converge to their limit strongly in $L^2(Q_{1,T})$ and $L^2(Q_{2,T})$ respectively. As a direct consequence, $(\varphi_i(s_{\mathcal{D}_m}))_m$ converges in $L^2(Q_{i,T})$ for $i \in \{1,2\}$ towards a function ϕ, which satisfies, thanks to Lemma 19.1,

$$\int_0^T \int_{\Omega_{i,\xi}} (\phi(x+\xi,t) - \phi(x,t))^2 \, d\mathbf{x}dt \le C|\xi|^2, \qquad \forall \xi \in \mathbb{R}^d.$$

This implies (see [27]) that $\phi \in L^2(0,T;H^1(\Omega_i))$. It remains to identify ϕ as $\varphi_i(s)$, $i \in \{1,2\}$. This can be done using Minty's lemma (see e.g. [68, Theorem 4.1]). □

Corollary 19.5. *Up to a subsequence, $(s_{\mathcal{D}_m})_m$ converges towards s strongly in $L^p(Q_T)$ for all $p \in [1,\infty)$.*

Proof : Since $(\varphi_i(s_{\mathcal{D}_m}))_m$ converges in $L^2(Q_T)$ towards $\varphi_i(s)$, it converges (up to a new subsequence) almost everywhere in Q_T. Since φ_i^{-1} is continuous, $s_{\mathcal{D}_m}$ tends to s almost everywhere. The result then follows from the uniform bound on $(s_{\mathcal{D}_m})_m$ stated in Proposition 18.1. □

Lemma 19.6. *There exists $\mathfrak{P} \in L^2(0,T;H^1(\Omega_i))$ such that, up to a subsequence,*

$$P_{\mathcal{D}_m} - m_{\Omega_i}(P_{\mathcal{D}_m}) \rightharpoonup \mathfrak{P} \text{ weakly in } L^2(Q_{i,T}) \text{ as } m \to \infty.$$

Proof : In view of the discrete Poincaré-Wirtinger inequality [76], there exists C depending only on Ω_i and on the quantity C_1 introduced in Proposition 18.5 such that

$$\|P_{\mathcal{D}_m} - m_{\Omega_i}(P_{\mathcal{D}_m})\|_{L^2(Q_{i,T})} \le C, \text{ for } i \in \{1,2\}.$$

Hence the sequence $(P_{\mathcal{D}_m} - m_{\Omega_i}(P_{\mathcal{D}_m}))_m$ converges weakly in $L^2(Q_{i,T})$ towards a function $\mathfrak{P}$. Therefore, for all $\xi \in \mathbb{R}^d$,

$$P_{\mathcal{D}_m}(\cdot+\xi,\cdot) - P_{\mathcal{D}_m} \rightharpoonup \mathfrak{P}(\cdot+\xi,\cdot) - \mathfrak{P} \quad \text{weakly in } \; L^2(\Omega_{i,\xi} \times (0,T)) \text{ as } m \to \infty.$$

The lower semi-continuity of the norm for the weak L^2 topology implies that

$$\int_0^T \int_{\Omega_{i,\xi}} (\mathfrak{P}(x+\xi,t) - \mathfrak{P}(x,t))^2 \, d\mathbf{x}dt \le \liminf_{m\to\infty} \int_0^T \int_{\Omega_{i,\xi}} (P_{\mathcal{D}_m}(x+\xi,t) - P_{\mathcal{D}_m}(x,t))^2 \, d\mathbf{x}dt.$$

We deduce from Proposition 18.5 and Lemma 19.1 that

$$\int_0^T \int_{\Omega_{i,\xi}} (\mathfrak{P}(x+\xi,t) - \mathfrak{P}(x,t))^2 \, d\mathbf{x}dt \le C_1|\xi|^2,$$

ensuring that $\mathfrak{P}$ belongs to $L^2(0,T;H^1(\Omega_i))$. □

19.2 Convergence of the traces

We denote by $s_{\mathcal{D}_{|\Gamma,i}}$ (resp. $P_{\mathcal{D}_{|\Gamma,i}}$) the trace of $s_{\mathcal{D}}$ (resp. $P_{\mathcal{D}}$) on Γ from the side of Ω_i, defined by

$$s_{\mathcal{D}_{|\Gamma,i}}(x,t) = s_{K,\sigma}^{n+1}, \quad P_{\mathcal{D}_{|\Gamma,i}}(x,t) = P_{K,\sigma}^{n+1}, \qquad \forall (x,t) \in \sigma \times (t^n, t^{n+1}],$$

where $\sigma \in \mathcal{E}_{K,\Gamma}$, $K \subset \Omega_i$.

It has been proven in Proposition 19.4 that $\varphi_i(s_{\mathcal{D}_m})$ converges strongly in $L^2(Q_{i,T})$ towards $\varphi_i(s) \in L^2(0,T;H^1(\Omega_i))$. Hence, $\varphi_1(s)$ and $\varphi_2(s)$ admits a traces in the sense of $L^2(\Gamma \times (0,T))$. Since φ_i^{-1} is continuous, s also admits a traces on the interface, denoted by s_1 and s_2. We claim in Corollary 19.10 below that $s_{\mathcal{D}_{m|\Gamma,i}}$ converges *strongly* in $L^p(\Gamma \times (0,T))$ towards s_i for all $p \in [1,\infty)$.

We now introduce another definition of the trace, denoted by $\tilde{u}_{|\Gamma,i}$. For a function u of $\mathcal{X}(\mathcal{D})$ we define

$$\tilde{u}_{|\Gamma,i}(x,t) := u_K^{n+1} \text{ if } (x,t) \in \sigma \times (t^n, t^{n+1}], \sigma \subset \Gamma \cap \partial K, K \subset \Omega_i.$$

Lemma 19.7. *Let $u \in \mathcal{X}(\mathcal{D})$, then*

$$\int_0^T \int_\Gamma \left| u_{|\Gamma,i} - \tilde{u}_{|\Gamma,i} \right| \, d\mathbf{x}dt \le |u|_{\mathcal{D}} \left(T m(\Gamma) \mathrm{size}(\mathcal{D}) \right)^{1/2}.$$

Proof : From the definitions of the traces of u,

$$\int_0^T \int_\Gamma \left| u_{|\Gamma,i} - \tilde{u}_{|\Gamma,i} \right| \, d\mathbf{x}dt = \sum_{n=0}^{N} \delta t \sum_{K \in \mathcal{T}_i} \sum_{\sigma \in \mathcal{E}_{K,\Gamma}} m(\sigma) |u_{K,\sigma}^{n+1} - u_K^{n+1}|.$$

Cauchy-Schwarz inequality yields that

$$\begin{aligned} \int_0^T \int_\Gamma \left| u_{|\Gamma,i} - \tilde{u}_{|\Gamma,i} \right| \, d\mathbf{x}dt \le \; & \left(\sum_{n=0}^{N} \delta t \sum_{K \in \mathcal{T}_i} \sum_{\sigma \in \mathcal{E}_{K,\Gamma}} \tau_{K,\sigma} (u_{K,\sigma}^{n+1} - u_K^{n+1})^2 \right)^{1/2} \\ & \cdot \left(\sum_{n=0}^{N} \delta t \sum_{K \in \mathcal{T}_i} \sum_{\sigma \in \mathcal{E}_{K,\Gamma}} m(\sigma) d_{K,\sigma} \right)^{1/2}. \end{aligned}$$

The result follows. □

Since Ω_i is supposed to be polygonal, Γ is made of a finite number of faces $(\Gamma_j)_{1 \le j \le J}$ contained in affine hyperplanes of $\mathbb{R}^d$. We denote by $\mathbf{n}_{i,j}$ the outward normal to Γ_j with respect to Ω_i. For $\varepsilon > 0$ and $j \in \{1, \dots, J\}$, we define the open subset $\omega_{i,j,\varepsilon}$ of Ω_i as the largest cylinder of width ε generate by Γ_j and $n_{i,j}$ included in Ω_i, that is

$$\omega_{i,j,\varepsilon} := \left\{ x - h\mathbf{n}_{i,j} \in Q_{i,T} \; | x \in \Gamma_j, 0 < h < \varepsilon \text{ and } [x, x - \varepsilon \mathbf{n}_{i,j}] \subset \overline{\Omega}_i \right\}. \tag{19.2}$$

We also define the subset $\Gamma_{i,j,\varepsilon} = \partial\omega_{i,j,\varepsilon} \cap \Gamma_j$ of Γ_j, that satisfies

$$m(\Gamma_j \setminus \Gamma_{i,j,\varepsilon}) \le C\varepsilon, \tag{19.3}$$

where C only depends on Ω.

Lemma 19.8. *Let $u \in \mathcal{X}(\mathcal{D})$, then for all $j \in \{1, \dots, J\}$,*

$$\int_0^T \frac{1}{\varepsilon} \int_{\Gamma_{i,j,\varepsilon}} \int_0^\varepsilon \left(\tilde{u}_{|\Gamma,i}(x,t) - u(x - h\mathbf{n}_{i,j}, t)\right)^2 dh \; d\mathbf{x}dt \leq |u|^2_{\mathcal{D}} \left(\varepsilon + \mathrm{size}(\mathcal{D})\right).$$

Proof : For all $\sigma \in \mathcal{E}_{\mathrm{int}}$, we denote by

$$\chi_\sigma(x,y) := \begin{cases} 1 & \text{if } (x,y) \cap \sigma \text{ is reduced to a single point,} \\ 0 & \text{otherwise.} \end{cases}$$

We also introduce, for almost all $x \in \Gamma_{i,j,\varepsilon}$, for almost $h \in (0,\varepsilon)$ and for all $t \in (0,T)$, the quantity

$$\begin{aligned} T_{\mathcal{D}}(x,h,t) &:= \left|\tilde{u}_{|\Gamma,i}(x,t) - u_{\mathcal{D}}(x - h\mathbf{n}_{i,j}, t)\right| \\ &\leq \sum_{\sigma = K|L \in \mathcal{E}_i} \chi_\sigma(x, x - h\mathbf{n}_{i,j}) \left|u_K^{n+1} - u_L^{n+1}\right| \end{aligned}$$

if $t \in (t^n, t^{n+1}]$. It follows from the Cauchy-Schwarz inequality that, for $t \in (t^n, t^{n+1}]$,

$$\begin{aligned} (T_{\mathcal{D}}(x,h,t))^2 \leq & \left(\sum_{\sigma = K|L \in \mathcal{E}_i} \chi_\sigma(x, x - h\mathbf{n}_{i,j}) \frac{\left(u_K^{n+1} - u_L^{n+1}\right)^2}{d_{KL}|\mathbf{n}_{i,j} \cdot \mathbf{n}_{KL}|} \right) \\ & \times \left(\sum_{\sigma = K|L \in \mathcal{E}_i} \chi_\sigma(x, x - h\mathbf{n}_{i,j}) d_{KL} |\mathbf{n}_{i,j} \cdot \mathbf{n}_{KL}| \right). \end{aligned}$$

For almost all $x \in \Gamma_{i,j,\varepsilon}$, there exists a unique $K_1 \in \mathcal{T}_i$ such that $x \in \partial K_1$. Moreover, for almost all $h \in (0,\varepsilon)$, there exists a unique $K_2 \in \mathcal{T}_i$ such that $x - hn_i$ belongs to K_2 (possibly K_2 coincides with K_1). Hence,

$$\begin{aligned} \sum_{\sigma = K|L \in \mathcal{E}_{\mathrm{int}}} \chi_\sigma(x, x - h\mathbf{n}_{i,j}) d_{KL} \quad |\mathbf{n}_{i,j} \cdot \mathbf{n}_{KL}| &= (x_{K_1} - x_{K_2}) \cdot \mathbf{n}_{i,j} \\ &\leq (x_{K_1} - x) \cdot \mathbf{n}_{i,j} + h + |(x_{K_2} - (x - h\mathbf{n}_{i,j})) \cdot \mathbf{n}_{i,j}|. \end{aligned} \tag{19.4}$$

Since $x - h\mathbf{n}_{i,j}$ belongs to K_2, we have

$$|(x_{K_2} - (x - h\mathbf{n}_{i,j})) \cdot \mathbf{n}_{i,j}| \leq \mathrm{size}(\mathcal{D}),$$

and since x belongs to Γ_i, $(x_{K_1} - x) \cdot \mathbf{n}_{i,j} \leq 0$. Then we obtain

$$\sum_{\sigma = K|L \in \mathcal{E}_{\mathrm{int}}} \chi_\sigma(x, x - h\mathbf{n}_{i,j}) d_{KL} |\mathbf{n}_{i,j} \cdot \mathbf{n}_{KL}| \leq \varepsilon + \mathrm{size}(\mathcal{D}). \tag{19.5}$$

For all $\sigma \in \mathcal{E}_{\mathrm{int}}$ with $\sigma \cap \omega_{i,j,\varepsilon} = \emptyset$ and all $h \in (0,\varepsilon)$, one has

$$\int_{\Gamma_i^\varepsilon} \chi_\sigma(x, x - hn_i) \; d\mathbf{x} = 0.$$

For all $\sigma \in \mathcal{E}_{i,j,\varepsilon} := \{\sigma \in \mathcal{E}_i \mid \sigma \cap \omega_{i,j,\varepsilon} \neq \emptyset\}$, one has

$$\forall h \in (0,\varepsilon), \qquad \int_{\Gamma_{i,j,\varepsilon}} \chi_\sigma(x, x - h\mathbf{n}_{i,j}) \; d\mathbf{x} \leq m(\sigma) |\mathbf{n}_{i,j} \cdot \mathbf{n}_{KL}|. \tag{19.6}$$

We obtain from (19.5) and (19.6) that for all $t \in (t^n, t^{n+1}]$, for all $h \in (0, \varepsilon)$,

$$\int_{\Gamma_{i,j,\varepsilon}} (T_{\mathcal{D}}(x,h,t))^2 \; d\mathbf{x} \le (\varepsilon + \text{size}(\mathcal{D})) \sum_{\sigma=K|L\in\mathcal{E}_{i,j,\varepsilon}} \tau_{KL} \left(u_K^{n+1} - u_L^{n+1}\right)^2,$$

which complete the proof. □

Proposition 19.9. *The sequence* $\left(\varphi_i(s_{\mathcal{D}_m|_{\Gamma,i}})\right)_m$ *converges towards* $\varphi_i(s_i)$ *strongly in* $L^1(\Gamma \times (0,T))$ *as* $m \to \infty$.

Proof : For notation convenience, we remove the subscripts m in the proof. Denote by

$$A_{i,j,\mathcal{D}} := \int_0^T \int_{\Gamma_j} \left|\varphi_i(s_{\mathcal{D}|_{\Gamma,i}}) - \varphi_i(s_i)\right| \; d\mathbf{x}dt, \tag{19.7}$$

then in view of Lemma 19.7 and Proposition 18.5, there exists C not depending on $\mathcal{D}$ such that

$$A_{i,j,\mathcal{D}} = \int_0^T \int_{\Gamma_j} \left|\varphi_i(\tilde{s}_{\mathcal{D}|_{\Gamma,i}}) - \varphi_i(s_i)\right| \; d\mathbf{x}dt + C\text{size}(\mathcal{D})^{1/2}. \tag{19.8}$$

By (19.3), one has

$$\int_0^T \int_{\Gamma_j} \left|\varphi_i(\tilde{s}_{\mathcal{D}|_{\Gamma,i}}) - \varphi_i(s_i)\right| \; d\mathbf{x}dt \le \int_0^T \int_{\Gamma_{i,j,\varepsilon}} \left|\varphi_i(\tilde{s}_{\mathcal{D}|_{\Gamma,i}}) - \varphi_i(s_i)\right| \; d\mathbf{x}dt + \varphi_i(1)C\varepsilon. \tag{19.9}$$

Next we apply the triangle inequality to deduce that

$$\int_0^T \int_{\Gamma_{i,j,\varepsilon}} \left|\varphi_i(\tilde{s}_{\mathcal{D}|_{\Gamma,i}}) - \varphi_i(s_i)\right| \; d\mathbf{x}dt \le B_{1,\mathcal{D},\varepsilon} + B_{2,\mathcal{D},\varepsilon} + B_{3,\varepsilon}, \tag{19.10}$$

where

$$\begin{aligned}
B_{1,\mathcal{D},\varepsilon} &= \frac{1}{\varepsilon} \int_0^T \int_{\Gamma_{i,j,\varepsilon}} \int_0^\varepsilon \left|\varphi_i(\tilde{s}_{\mathcal{D}|_{\Gamma,i}})(x,t) - \varphi_i(s_{\mathcal{D}})(x - h\mathbf{n}_{i,j}, t)\right| dh \; d\mathbf{x}dt, \\
B_{2,\mathcal{D},\varepsilon} &= \frac{1}{\varepsilon} \int_0^T \int_{\omega_{i,j,\varepsilon}} |\varphi_i(s_{\mathcal{D}}) - \varphi_i(s)| \; d\mathbf{x}dt, \\
B_{3,\varepsilon} &= \frac{1}{\varepsilon} \int_0^T \int_{\Gamma_{i,j,\varepsilon}} \int_0^\varepsilon |\varphi_i(s_i)(x,t) - \varphi_i(s)(x - h\mathbf{n}_{i,j}, t)| \, dh \; d\mathbf{x}dt,
\end{aligned}$$

where we have used (19.2). From Cauchy-Schwarz inequality, one has

$$(B_{1,\mathcal{D},\varepsilon})^2 \le m(\Gamma_{i,j,\varepsilon})T \int_0^T \int_{\Gamma_{i,j,\varepsilon}} \frac{1}{\varepsilon} \int_0^\varepsilon \Big(\varphi_i(\tilde{s}_{\mathcal{D}|_{\Gamma,i}})(x,t) - \varphi_i(s_{\mathcal{D}})(x - h\mathbf{n}_{i,j}, t)\Big)^2 dh \; d\mathbf{x}dt,$$

and then, from Proposition 18.5 and Lemma 19.8, one has

$$|B_{1,\mathcal{D},\varepsilon}| \le \left(C_1(\text{size}(\mathcal{D}) + \varepsilon) m(\Gamma_i) T\right)^{1/2}. \tag{19.11}$$

We can now let size($\mathcal{D}$) tend to 0 in (19.10). Thanks to Proposition 19.4, we can claim that

$$\lim_{\text{size}(\mathcal{D})\to 0} B_{2,\mathcal{D},\varepsilon} = 0.$$

Then it follows from (19.9) and (19.11) that

$$\limsup_{\text{size}(\mathcal{D})\to 0} \int_0^T \int_{\Gamma_j} \left|\varphi_i(\tilde{s}_{\mathcal{D}_{|\Gamma,i}}) - \varphi_i(s_i)\right| \, d\mathbf{x}dt \le C(\varepsilon + \sqrt{\varepsilon}) + B_{3,\varepsilon}. \tag{19.12}$$

Since $\varphi_i(s_i)$ is the trace of $\varphi_i(s)$ on Γ, $\lim_{\varepsilon\to 0} B_{3,\varepsilon} = 0$. Therefore, letting ε tend to 0 in (19.12) implies that

$$\lim_{\text{size}(\mathcal{D})\to 0} \int_0^T \int_{\Gamma_j} \left|\varphi_i(\tilde{s}_{\mathcal{D}_{|\Gamma,i}}) - \varphi_i(s_i)\right| \, d\mathbf{x}dt = 0.$$

Then the result follows from (19.7) and (19.2). □

Corollary 19.10. *The sequence* $\left(s_{\mathcal{D}_{m|\Gamma,i}}\right)_m$ *converges towards* s_i *strongly in* $L^p(\Gamma \times (0,T))$ *for all* $p \in [1,\infty)$.

Proof : This corollary is just a consequence from the fact that $\varphi_i(s_{\mathcal{D}_{m|\Gamma,i}})$ converges, up to a subsequence, almost everywhere on $\Gamma\times(0,T)$, from the fact that φ_i^{-1} is continuous and from the fact that $s_{\mathcal{D}_{m|\Gamma,i}}$ is essentially uniformly bounded between 0 and 1. □

Lemma 19.11. *The sequence* $\left((P_{\mathcal{D}_m})_{|\Gamma,i} - m_{\Omega_i}(P_{\mathcal{D}})\right)_m$ *converges towards* $\mathfrak{P}_i$ *weakly in* $L^2(\Gamma \times (0,T))$.

Proof : Let $\psi \in \mathcal{D}(\Gamma_i \times (0,T))$, then, there exists $\varepsilon_\star$ such that, for all $\varepsilon \in (0,\varepsilon_\star)$, $\text{supp}(\psi) \subset \Gamma_{i,j,\varepsilon} \times (0,T)$. We aim to prove that

$$\lim_{\text{size}(\mathcal{D})\to 0} \int_0^T \int_{\Gamma_j} \left(P_{\mathcal{D}_{|\Gamma,i}} - m_{\Omega_i}(P_{\mathcal{D}}) - \mathfrak{P}_i\right) \psi \, d\mathbf{x}dt = 0. \tag{19.13}$$

Thanks to Lemma 19.7 and to Proposition 18.5, it is sufficient to show that

$$\lim_{\text{size}(\mathcal{D})\to 0} \int_0^T \int_{\Gamma_j} \left(\tilde{P}_{\mathcal{D}_{|\Gamma,i}} - m_{\Omega_i}(P_{\mathcal{D}}) - \mathfrak{P}_i\right) \psi \, d\mathbf{x}dt = 0.$$

Let $\varepsilon \in (0,\varepsilon_\star)$, then one has

$$\int_0^T \int_{\Gamma_j} \left(\tilde{P}_{\mathcal{D}_{|\Gamma,i}} - m_{\Omega_i}(P_{\mathcal{D}_m}) - \mathfrak{P}_i\right) \psi \, d\mathbf{x}dt = E_{1,\mathcal{D},\varepsilon} + E_{2,\mathcal{D},\varepsilon} + E_{3,\varepsilon},$$

where

$$\begin{aligned}
E_{1,\mathcal{D},\varepsilon} &= \int_0^T \frac{1}{\varepsilon} \int_{\Gamma_{i,j,\varepsilon}} \int_0^\varepsilon \left(\tilde{P}_{\mathcal{D}_{|\Gamma,i}}(x,t) - P_{\mathcal{D}}(x - h\mathbf{n}_{i,j},t)\right) \psi(x,t) dh \, d\mathbf{x}dt,\\
E_{2,\mathcal{D},\varepsilon} &= \int_0^T \frac{1}{\varepsilon} \int_{\Gamma_{i,j,\varepsilon}} \int_0^\varepsilon \left(P_{\mathcal{D}}(x - h\mathbf{n}_{i,j},t) - m_{\Omega_i}(P_{\mathcal{D}}) - \mathfrak{P}(x - h\mathbf{n}_{i,j},t)\right) \psi(x,t) dh \, d\mathbf{x}dt,\\
E_{3,\varepsilon} &= \int_0^T \frac{1}{\varepsilon} \int_{\Gamma_{i,j,\varepsilon}} \int_0^\varepsilon \left(\mathfrak{P}(x - h\mathbf{n}_{i,j},t) - \mathfrak{P}_i\right) \psi(x,t) dh \, d\mathbf{x}dt.
\end{aligned}$$

The Cauchy-Schwarz inequality gives that

$$(E_{1,\mathcal{D},\varepsilon})^2 \le \int_0^T \frac{1}{\varepsilon} \int_{\Gamma_{i,j,\varepsilon}} \int_0^\varepsilon \left(\tilde{P}_{\mathcal{D}_{|\Gamma,i}}(x,t) - P_{\mathcal{D}}(x - h\mathbf{n}_{i,j},t)\right)^2 dh\, d\mathbf{x}dt \times \int_0^T \int_{\Gamma_j} (\psi(x,t))^2\, d\mathbf{x}dt.$$

Using Proposition 18.5 and Lemma 19.8 yields

$$|E_{1,\mathcal{D},\varepsilon}| \le \|\psi\|_{L^2(\Gamma_j\times(0,T))} \left(C_1(\varepsilon + \mathrm{size}(\mathcal{D}))\right)^{1/2}.$$

It has been stated in Lemma 19.6 that $P_{\mathcal{D}} - m_{\Omega_i}(P_{\mathcal{D}})$ tends to $\mathfrak{P}$ weakly in $L^2(Q_{i,T})$ as size$(\mathcal{D})$ tends to 0, then

$$\lim_{\mathrm{size}(\mathcal{D})\to 0} E_{2,\mathcal{D},\varepsilon} = 0.$$

Therefore,

$$\limsup_{\mathrm{size}(\mathcal{D})\to 0} \left| \int_0^T \int_{\Gamma_j} \left(\tilde{P}_{\mathcal{D}_{|\Gamma,i}} - P_i\right) \psi\, d\mathbf{x}dt \right| \le C_\psi \sqrt{\varepsilon} + |E_{3,\varepsilon}|.$$

Since P_i is the trace on Γ of P from the side of Ω_i, one has

$$\lim_{\varepsilon\to 0} E_{3,\varepsilon} = 0.$$

Thus, letting $\varepsilon \to 0$, one obtains that for all $\psi \in \mathcal{D}(\Gamma_j \times (0,T))$,

$$\lim_{\mathrm{size}(\mathcal{D})\to 0} \int_0^T \int_{\Gamma_j} \left(\tilde{P}_{\mathcal{D}_{|\Gamma,i}} - m_{\Omega_i}(P_{\mathcal{D}}) - \mathfrak{P}_i\right) \psi\, d\mathbf{x}dt = 0. \tag{19.14}$$

A straightforward generalization of [65, Lemma 3.10] allows us to claim, using Proposition 18.5 and the discrete Poincaré-Wirtinger inequality [76], that $\left(\tilde{P}_{\mathcal{D}_{|\Gamma,i}} - m_{\Omega_i}(P_{\mathcal{D}})\right)_{\mathcal{D}}$ is uniformly bounded in $L^2(\Gamma \times (0,T))$. Then, we conclude, using a classical density argument, that (19.14) holds for all $\psi \in L^2(\Gamma_j \times (0,T))$. □

Proposition 19.12. *There exists $P \in L^2(0,T;H^1(\Omega_i))$ such that $P_{\mathcal{D}_m}$ tends to P weakly in $L^2(Q_T)$ as $m \to \infty$, and such that $\left(P_{\mathcal{D}_{m_{|\Gamma,i}}}\right)_m$ converges weakly in $L^2(\Gamma \times (0,T))$ towards P_i.*

Proof : Firstly, since we have enforced $m_{\Omega_1}(P_{\mathcal{D}_m}) = 0$, we can set $P := \mathfrak{P}$ in $Q_{1,T}$. Next we search for a uniform bound on $\|P_{\mathcal{D}_m}\|_{L^2(Q_{2,T})}$. In view of the discrete Poincaré-Wirtinger inequality

$$\|P_{\mathcal{D}_m}\|^2_{L^2(Q_{2,T})} \le (m_{\Omega_2}(P_{\mathcal{D}_m}))^2 + C, \tag{19.15}$$

it only remains to check that $m_{\Omega_2}(P_{\mathcal{D}_m})$ is uniformly bounded w.r.t. m. This is a consequence of the fact that, almost everywhere on $\Gamma \times (0,T)$, one has

$$m_{\Omega_2}(P_{\mathcal{D}_m}) = P_{\mathcal{D}_{m_{|\Gamma,1}}} - \left(P_{\mathcal{D}_{m_{|\Gamma,2}}} - m_{\Omega_2}(P_{\mathcal{D}_m})\right) - \left(Z_1(\pi_{\mathcal{D}_m}) - Z_2(\pi_{\mathcal{D}_m})\right).$$

Then, integrating on $\Gamma \times (0,T)$ and using Lemma 16.1 provides

$$|m_{\Omega_2}(P_{\mathcal{D}_m})| \le \frac{1}{m(\Gamma)T} \sum_{i\in\{1,2\}} \left\| P_{\mathcal{D}_{m_{|\Gamma,i}}} - m_{\Omega_i}(P_{\mathcal{D}_m}) \right\|_{L^1(\Gamma\times(0,T))} + \|Z_1 - Z_2\|_\infty.$$

The quantities $\left\| P_{\mathcal{D}_{m_{|\Gamma,i}}} - m_{\Omega_i}(P_{\mathcal{D}_m}) \right\|_{L^1(\Gamma\times(0,T))}$ are bounded by the proof of Lemma 19.11. Hence, in view of (19.15), $(P_{\mathcal{D}_m})_m$ converges towards some function P weakly in $L^2(Q_{i,T})$. From the analysis performed in the proof of Lemma 19.6, we deduce that $P \in L^2(0,T;H^1(\Omega_i))$, and from the analysis of Lemma 19.11, we deduce the weak convergence of the traces. □

Lemma 19.13. *Let $s_1, s_2 \in L^\infty(\Gamma\times(0,T))$ be the respective limits of $\left(s_{\mathcal{D}_{m_{|\Gamma,1}}}\right)_m$ and $\left(s_{\mathcal{D}_{m_{|\Gamma,2}}}\right)_m$, then,*

$$\tilde{\pi}_1(s_1) \cap \tilde{\pi}_2(s_2) \neq \emptyset \qquad a.e.\ on\ \Gamma\times(0,T). \tag{19.16}$$

Proof : For all $m \in \mathcal{N}$, one has

$$\tilde{\pi}_1(s_{\mathcal{D}_{m_{|\Gamma,1}}}) \cap \tilde{\pi}_2(s_{\mathcal{D}_{m_{|\Gamma,2}}}) \neq \emptyset.$$

Since the set $F = \{(a,b) \in [0,1]^2 \mid \tilde{\pi}_1(a) \cap \tilde{\pi}_2(b) \neq \emptyset\}$ is closed in $[0,1]^2$, we conclude that (19.16) holds. □

In the sequel, we denote by

$$T_{[A,B]}(s) = \begin{cases} s & \text{if } s \in [A,B], \\ A & \text{if } s \le A, \\ B & \text{if } s \ge B, \end{cases}$$

and by

$$\mathcal{U} = \{(x,t) \in \Gamma\times(0,T) \mid \{s_1,s_2\} = \{0,1\}\}, \qquad \mathcal{V} = \mathcal{U}^c.$$

Note that, thanks to Lemma 19.13, the set $\mathcal{U}$ is empty if $\min_i \pi_i(1) > \max_i \pi_i(0)$.

Lemma 19.14. *There exists a measurable function π defined on $\mathcal{V}$ with values in $\overline{\mathbb{R}}$, such that, up to a subsequence,*

$$\pi_{\mathcal{D}_m} \to \pi \qquad a.e.\ in\ \mathcal{V}.$$

Proof : We define the functions

$$\tilde{\varphi}_i : p \mapsto \int_{\pi_i(0)}^{p} K_i \frac{kr_{o,i}(\pi_i^{-1}(a))kr_{w,i}(\tilde{\pi}_i^{-1}(a))}{\mu_w kr_{o,i}(\pi_i^{-1}(a)) + \mu_o kr_{w,i}(\pi_i^{-1}(a))} da,$$

that satisfy the properties

$$\pi \in \tilde{\pi}_i(s) \implies \tilde{\varphi}_i(\pi) = \tilde{\varphi}_i(\pi_i(s)) = \varphi_i(s), \tag{19.17}$$

and

$$\text{its restriction } (\tilde{\varphi}_i)_{|[\pi_i(0),\pi_i(1)]} \text{ admits a continuous inverse function.} \tag{19.18}$$

Thanks to Proposition 19.9 and to (19.17), we can claim that, up to a subsequence, $\tilde{\varphi}_i(\pi_{\mathcal{D}_m})$ converges almost everywhere on $\Gamma\times(0,T)$ towards $\tilde{\varphi}_i(\pi_i(s_i))$. For a.e. $(x,t) \in \mathcal{V}$, the set $\tilde{\pi}_1(s_1) \cap \tilde{\pi}_2(s_2)$ is reduced to the singleton $\{\pi_{i_0}(s_{i_0})\}$ for some $i_0 \in \{1,2\}$. Thanks to (19.18), we can identify the limit π of $\pi_{\mathcal{D}_m}$ as $\pi_{i_0}(s_{i_0})$. □

Lemma 19.15. *Assume that* $[\min_i \pi_i(1), \max_i \pi_i(0)] \neq \emptyset$*, then there exists* $\pi \in L^\infty(\mathcal{U}; [\min_i \pi_i(1), \max \pi_i(0)])$ *such that, for all bounded interval* $\mathcal{J} \subset \mathbb{R}$ *such that* $[\min_i \pi_i(1), \max_i \pi_i(0)] \subset \overset{\circ}{\mathcal{J}}$*,*

$$T_{\mathcal{J}}(\pi_{\mathcal{D}_m}) \to \pi \quad \text{in the } L^\infty(\mathcal{U}) \text{ weak-}\star \text{ sense.}$$

Proof : For the sake of simplicity, we assume, without loss of generality, that $\pi_1(1) \leq \pi_2(0)$, then thanks to Lemma 19.13, almost everywhere in $\mathcal{U}$, $s_1 = 1$ and $s_2 = 0$.

The sequence $(T_{\mathcal{J}}(\pi_{\mathcal{D}_m}))_m$ is bounded in $L^\infty(\mathcal{U})$, thus, up to a subsequence, it converges towards a function $\pi_{\mathcal{J}}$ in the $L^\infty(\mathcal{U})$ weak-$\star$ sense. Let us now show that $\pi_{\mathcal{J}}$ does not depend on the choice of the bounded interval $\mathcal{J}$. Because of Lemma 19.13, one has, for a.e. $(x,t) \in \mathcal{U}$,

$$\liminf_m \pi_{\mathcal{D}_m} \geq \pi_1(1), \qquad \limsup_m \pi_{\mathcal{D}_m} \leq \pi_2(0). \tag{19.19}$$

Let $\mathcal{J}_1$ and $\mathcal{J}_2$ be two bounded intervals such that $[\pi_1(1), \pi_2(0)] \subset \overset{\circ}{\mathcal{J}}_k$ ($k \in \{1,2\}$). Then, it follows from (19.19) that, for a.e. $(x,t) \in \mathcal{U}$, for m large enough (depending on (x,t)),

$$T_{\mathcal{J}_1}(\pi_{\mathcal{D}_m}(x,t)) - T_{\mathcal{J}_2}(\pi_{\mathcal{D}_m}(x,t)) = 0.$$

As a consequence, the sequence $(T_{\mathcal{J}_1}(\pi_{\mathcal{D}_m}) - T_{\mathcal{J}_2}(\pi_{\mathcal{D}_m}))_m$ converges almost everywhere to 0 on $\mathcal{U}$, and is uniformly bounded in $L^\infty(\mathcal{U})$. The dominated convergence theorem yields that for all $\psi \in L^1(\mathcal{U})$,

$$\iint_{\mathcal{U}} (T_{\mathcal{J}_1}(\pi_{\mathcal{D}_m}) - T_{\mathcal{J}_2}(\pi_{\mathcal{D}_m}))\, \psi \; d\mathbf{x}dt \to 0 = \iint_{\mathcal{U}} (\pi_{\mathcal{J}_1} - \pi_{\mathcal{J}_2})\, \psi \; d\mathbf{x}dt.$$

Choosing $\psi = (\pi_{\mathcal{J}_1} - \pi_{\mathcal{J}_2})$ provides that $\pi_{\mathcal{J}_1} = \pi_{\mathcal{J}_2} = \pi$ almost everywhere in $\mathcal{U}$. $\square$

Lemma 19.16. *Assume that* $[\min_i \pi_i(1), \max_i \pi_i(0)] \neq \emptyset$*, then there exists* $\pi \in L^\infty(\mathcal{U})$ *such that, for all bounded interval* $\mathcal{J} \subset \mathbb{R}$ *such that* $[\min_i \pi_i(1), \max_i \pi_i(0)] \subset \overset{\circ}{\mathcal{J}}$*, the sequence* $(Z_i(T_{\mathcal{J}}(\pi_{\mathcal{D}_m})))_m$ *converges towards* $Z_i(\pi)$ *in the* $L^\infty(\mathcal{U})$ *weak-*$\star$ *sense.*

Proof : We suppose, without loss of generality, that $\pi_1(1) \leq \pi_2(0)$. Then on $\mathcal{U}$, $s_2 = 0$ and $s_1 = 1$. One has

$$Z_2(T_{\mathcal{J}}(\pi_{\mathcal{D}_m})) \;=\; \int_0^{\pi_2(0)} f_2 \circ \pi_2^{-1}(p)dp + \int_{\pi_2(0)}^{\pi_{\mathcal{D}_m}} f_2 \circ \pi_2^{-1}(p)dp.$$

Since for almost every $(x,t) \in \mathcal{U}$,

$$\limsup_m \pi_{\mathcal{D}_m}(x,t) \leq \pi_2(0),$$

and since $f_2 \circ \pi_2^{-1}(p) = 0$ for all $p \leq \pi_1(0)$, then for almost every $(x,t) \in \mathcal{U}$,

$$\int_{\pi_2(0)}^{\pi_{\mathcal{D}_m}(x,t)} f_2 \circ \pi_2^{-1}(p)dp \to 0 \text{ as } m \to \infty.$$

Since the function $Z_2 \circ T_{\mathcal{J}}$ is uniformly bounded on $\mathbb{R}$, the dominated convergence theorem yields that, for all $\psi \in L^1(\mathcal{U})$,

$$\lim_{m\to\infty} \int_{\mathcal{U}} Z_2(T_{\mathcal{J}}(\pi_{\mathcal{D}_m}))\psi \; d\mathbf{x}dt \to \iint_{\mathcal{U}} Z_2(\pi_2(0))\psi \; d\mathbf{x}dt = \iint_{\mathcal{U}} Z_2(\pi)\psi \; d\mathbf{x}dt.$$

Similarly, we obtain that

$$\iint_{\mathcal{U}} \left(Z_1(T_{\mathcal{J}}(\pi_{\mathcal{D}_m})) - T_{\mathcal{J}}(\pi_{\mathcal{D}_m})\right) \psi \; d\mathbf{x}dt \to \iint_{\mathcal{U}} \left(Z_1(\pi_1(1)) - \pi_1(1)\right) \psi \; d\mathbf{x}dt.$$

Since, thanks to Lemma 19.15, $T_{\mathcal{J}}(\pi_{\mathcal{D}_m})$ tends to π in the $L^\infty(\mathcal{U})$ weak-$\star$ sense, one has

$$\lim_{m\to\infty} \iint_{\mathcal{U}} Z_1(T_{\mathcal{J}}(\pi_{\mathcal{D}_m}))\psi \; d\mathbf{x}dt = \iint_{\mathcal{U}} \left(Z_1(\pi_1(1)) + \pi - \pi_1(1)\right) \psi \; d\mathbf{x}dt = \iint_{\mathcal{U}} Z_1(\pi)\psi \; d\mathbf{x}dt.$$

□

Proposition 19.17. *There exists a measurable function π on $\Gamma \times (0,T)$, with $\pi \in \tilde{\pi}_1(s_1) \cap \tilde{\pi}_2(s_2)$ a.e. on $\Gamma \times (0,T)$, with value in $[\min_i(\pi_i(0)), \max_i(\pi_i(1))]$ such that,*

$$Z_1(\pi_{\mathcal{D}_m}) - Z_2(\pi_{\mathcal{D}_m}) \to Z_1(\pi) - Z_2(\pi) \textit{ in the } L^\infty(\Gamma \times (0,T)) \textit{ weak-}\star \textit{ sense as } n \to \infty.$$

Proof : We know, from Lemma 16.1, that $Z_1(p) - Z_2(p)$ is uniformly bounded on $[\min_i \pi_i(0), \max_i \pi_i(1)]$. Hence, the sequence $(Z_1(\pi_{\mathcal{D}_m}) - Z_2(\pi_{\mathcal{D}_m}))_m$ converges in the $L^\infty(\Gamma \times (0,T))$ weak-$\star$ sense towards a function $\mathfrak{Z}$. Let $\psi \in L^1(\Gamma \times (0,T))$, then

$$\begin{aligned}\int_0^T \int_\Gamma \left(Z_1(\pi_{\mathcal{D}_m}) - Z_2(\pi_{\mathcal{D}_m})\right) \psi \; d\mathbf{x}dt &= \iint_{\mathcal{U}} \left(Z_1(\pi_{\mathcal{D}_m}) - Z_2(\pi_{\mathcal{D}_m})\right) \psi \; d\mathbf{x}dt \\ &+ \iint_{\mathcal{V}} \left(Z_1(\pi_{\mathcal{D}_m}) - Z_2(\pi_{\mathcal{D}_m})\right) \psi \; d\mathbf{x}dt.\end{aligned}$$

Thanks to Lemma 19.14, $\pi_{\mathcal{D}_m}$ tends almost everywhere to π on $\mathcal{V}$, then for almost every $(x,t) \in \mathcal{V}$, we can identify $\mathfrak{Z}(x,t)$ as $Z_1(\pi(x,t)) - Z_2(\pi(x,t))$. Thus

$$\lim_{m\to\infty} \iint_{\mathcal{V}} \left(Z_1(\pi_{\mathcal{D}_m}) - Z_2(\pi_{\mathcal{D}_m})\right) \psi \; d\mathbf{x}dt = \iint_{\mathcal{V}} \left(Z_1(\pi) - Z_2(\pi)\right) \psi \; d\mathbf{x}dt.$$

We denote by

$$\begin{aligned}A_m &= \iint_{\mathcal{U}} \left(Z_1(\pi_{\mathcal{D}_m}) - Z_2(\pi_{\mathcal{D}_m}) - Z_1(\pi) + Z_2(\pi)\right) \psi \; d\mathbf{x}dt, \\ &= \iint_{\mathcal{U}} \left(\Upsilon_1(\pi_{\mathcal{D}_m}) - \Upsilon_1(\pi)\right) \psi \; d\mathbf{x}dt + \iint_{\mathcal{U}} \left(\Upsilon_2(\pi_{\mathcal{D}_m}) - \Upsilon_2(\pi)\right) \psi \; d\mathbf{x}dt.\end{aligned}$$

Let $R \in \mathbb{R}$ such that $[\min_i \pi_i(0), \max_i \pi_i(1)] \subset [-R, R]$, then

$$A_m = B_{1,m}(R) - B_{2,m}(R) + C_m(R),$$

where

$$B_{i,m}(R) = \iint_{\mathcal{U}} \left(\Upsilon_i(\pi_{\mathcal{D}_m}) - \Upsilon_i(T_{[-R,R]}(\pi_{\mathcal{D}_m}))\right) \psi \; d\mathbf{x}dt$$

and

$$C_m(R) = \iint_{\mathcal{U}} \left(Z_1(T_{[-R,R]}(\pi_{\mathcal{D}_m})) - Z_2(T_{[-R,R]}(\pi_{\mathcal{D}_m})) - Z_1(\pi) + Z_2(\pi) \right) \psi \; d\mathbf{x}dt.$$

Let $\varepsilon > 0$, then since Υ_i admits finite limits as $p \to \min_i \pi_i(0)$ and $p \to \max_i \pi_i(1)$, there exists $R_0(\varepsilon) > 0$ such that

$$R > R_0(\varepsilon) \implies \|\Upsilon_i - \Upsilon_i \circ T_{[-R,R]}\|_\infty \leq \varepsilon.$$

Thus, for $R > R_0(\varepsilon)$ fixed,

$$|B_{i,m}(R)| \leq Tm(\Gamma)\varepsilon.$$

Thanks to Lemma 19.16,

$$\lim_{m\to\infty} C_m(R) = 0,$$

then, for all $\varepsilon > 0$,

$$\limsup_{m\to\infty} |A_m| \leq 2Tm(\Gamma)\varepsilon.$$

As a consequence, since the above estimate holds for all $\varepsilon > 0$, A_m tends to 0, concluding the proof of Proposition 19.17. □

20 End of the proof of Theorem 1

We have proven in the section 19 that, up to a subsequence, the sequence of approximate solutions $(s_{\mathcal{D}_m}, P_{\mathcal{D}_m})_m$ converge towards (s, P) as $m \to \infty$. Moreover, it as been stated in Lemmata 19.14 and 19.15 that $(\pi_{\mathcal{D}_m})_m$ converges in some sense on $\Gamma \times (0, T)$ towards a measurable function π. In order to conclude the proof of Theorem 1, it remains to check that (s, P) satisfy the weak formulations (16.24) and (16.25), and that the transmission conditions (16.19) and (16.20) are fulfilled. Let us begin by this latter point.

It follows from the construction of the function π carried out in Lemmata 19.14 and 19.15 that, for almost every $(x, t) \in \Gamma \times (0, T)$,

$$\pi(x, t) \in \tilde{\pi}_1(s_1(x, t)) \cap \tilde{\pi}_2(s_2(x, t)). \tag{20.1}$$

Let $\psi \in L^2(\Gamma \times (0, T))$, then thanks to (17.8), one has, for all $\psi \in L^2(\Gamma \times (0, T))$,

$$\int_0^T \int_\Gamma \left(P_{\mathcal{D}_{m_{|\Gamma,1}}} - P_{\mathcal{D}_{m_{|\Gamma,2}}} \right) \psi \; d\mathbf{x}dt = \int_0^T \int_\Gamma (Z_1(\pi_{\mathcal{D}_m}) - Z_2(\pi_{\mathcal{D}_m}))\, \psi \; d\mathbf{x}dt.$$

Letting m tend to ∞ provides, thanks to Propositions 19.12 and 19.17, that

$$\int_0^T \int_\Gamma (P_1 - P_2)\, \psi \; d\mathbf{x}dt = \int_0^T \int_\Gamma (Z_1(\pi) - Z_2(\pi))\, \psi \; d\mathbf{x}dt.$$

Hence,

$$P_1 - Z_1(\pi) = P_2 - Z_2(\pi) \quad \text{a.e. on } \Gamma \times (0, T). \tag{20.2}$$

In order to recover the weak formulations (16.24) and (16.25), we can apply to our case the analysis carried out in the proof of Theorem 5.1 in [90].

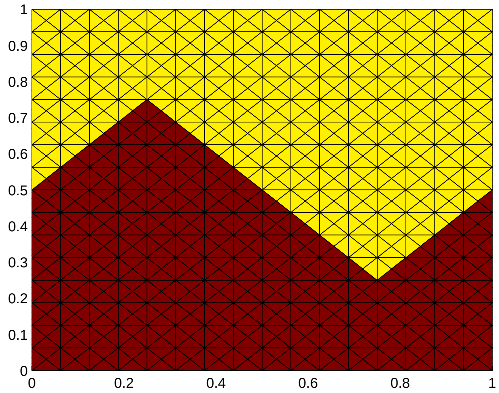

FIGURE 17 – The model porus medium $\Omega_1 \bigcup \Omega_2$

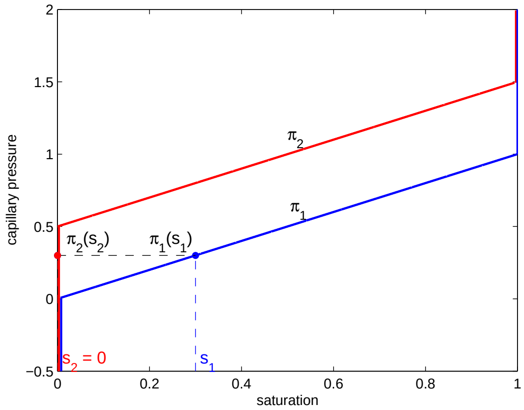

FIGURE 18 – Capillary pressure connection at time t = 0

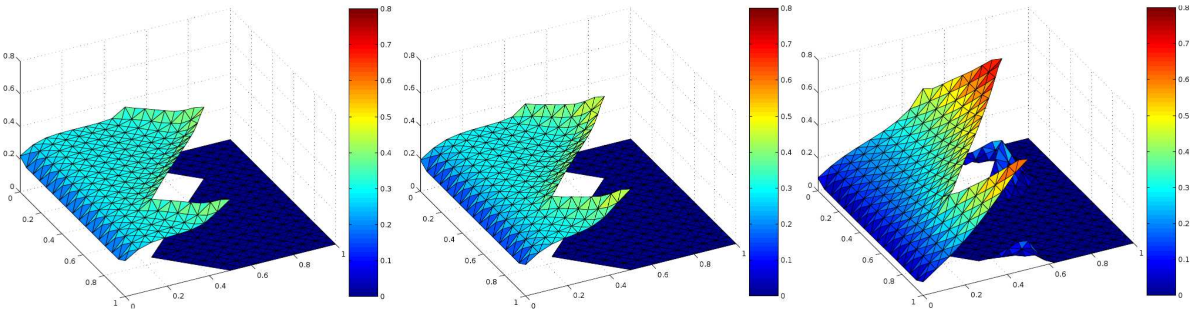

FIGURE 19 – Saturation for $t = 0.06$, $t = 0.11$ and $t = 0.6$

21 Numerical results

In this Section we consider a model porous medium $\Omega = (0,1)^2$ composed of two layers Ω_1 and Ω_2, which are separated by an "S-shaped" interface Γ (see Fig. 17), and which have different capillary pressure laws. The oil and water densities are given by $\rho_o = 0.81$, $\rho_w = 1$ respectively, and $\mathbf{g} = -9.81\mathbf{e}_y$. We suppose that the porosity is such that $\phi_i = 1, i \in \{1,2\}$, and we define the oil and water mobilities by

$$\eta_{o,i}(s) = 0.5s^2 \quad \text{and} \quad \eta_{w,i} = (1-s)^2, \quad i \in \{1,2\}.$$

Moreover we suppose that the capillary pressure curves have the form

$$\pi_1(s) = s \quad \text{and} \quad \pi_2(s) = 0.5 + s.$$

In the first test case we suppose that the layer Ω_1 contains some quantity of oil and it is situated below Ω_2, which is saturated with water, that is to say $\Omega_1 = \{(x,y) \in \Omega \mid y < \Gamma(x)\}$ and $\Omega_2 = \{(x,y) \in \Omega \mid y > \Gamma(x)\}$. The initial saturation is given by

$$s_0(x) = \begin{cases} 0.3 & \text{if } x \in \Omega_1, \\ 0 & \text{otherwise.} \end{cases}$$

The flow is driven by buoyancy, making the oil move along $\mathbf{e}_y$ until it reaches the interface Γ. As one can see on the figures 19 and 20, for $t \leq 0.11$, oil can not access

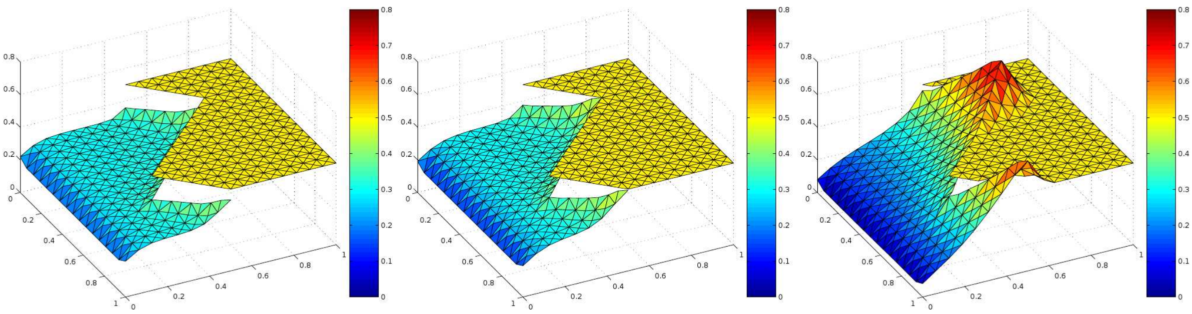

FIGURE 20 – Capillary pressure for $t = 0.06$, $t = 0.11$ and $t = 0.6$

the domain Ω_2, since the capillary pressure $\pi_1(s_1)$ is lower than the threshold value $\pi_2(0) = 0.5$, which is called *the entry pressure* (see Fig. 18). Hence the saturation below the interface s_1 increases, as well as the capillary pressure $\pi_1(s_1)$. As soon as the capillary pressure $\pi_1(s_1)$ reaches the entry pressure $\pi_2(0)$, oil starts to penetrate in the domain Ω_2. Nevertheless, as pointed out in [24, 40], a finite quantity of oil remains trapped under the rock discontinuity. This phenomenon is called *oil trapping*. It is worth noting that the solution at $t = 0$ satisfies (16.19), thus in the absence of gravity the initial distribution of oil-phase would be a steady state solution $s(\mathbf{x}, t) = s_0(\mathbf{x})$.

In the second test case we assume that the oil is initially situated in the rock with a higher *entry pressure* pressure i.e.

$$s_0(x) = \begin{cases} 0.3 & \text{if } x \in \Omega_2, \\ 0 & \text{otherwise.} \end{cases}$$

where this time $\Omega_1 = \{(x, y) \in \Omega \mid y > \Gamma(x)\}$ and $\Omega_2 = \{(x, y) \in \Omega \mid y < \Gamma(x)\}$. This time the flow is driven not only by a buoyancy, but also by a difference in the capillary pressure potential (the solution at $t = 0$ does not fulfill (16.19)). As a result the oil-phase can immediately penetrate the domain Ω_1. The figure 21 shows that the oil propagates in the domain Ω_1 with a finite speed.

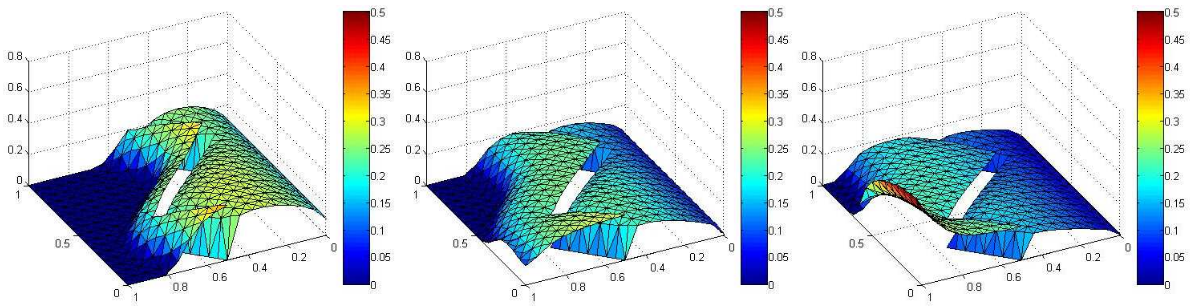

FIGURE 21 – Saturation for $t = 0.3$, $t = 1$ and $t = 1.7$

Références

[1] I. Aavatsmark. An introduction to multipoint flux approximations for quadrilateral grids. *Computational Geosciences*, 6 :405–432, 2002.

[2] I. Aavatsmark, T. Barkve, . Bøe, and T. Mannseth. Discretization on unstructured grids for inhomogeneous, anisotropic media. part i : Derivation of the methods. *SIAM Journal on Scientific Computing*, 19 :1700–1716, 1998.

[3] I. Aavatsmark, T. Barkve, . Bøe, and T. Mannseth. Discretization on unstructured grids for inhomogeneous, anisotropic media. part ii : Discussion and numerical results. *SIAM Journal on Scientific Computing*, 19 :1717–1736, 1998.

[4] R. A. Adams and J. Fournier. *Sobolev Spaces, Pure and Applied Mathematics*, volume 140. Academic Press, 2003.

[5] M. Afif and B. Amaziane. On convergence of finite volume schemes for a degenerate convection-diffusion equation arising in flow in porous media. *Comput. Methods Appl. Mech. Engrg.*, 191 :5265–5286, 2002.

[6] H. W. Alt and E. DiBenedetto. Nonsteady flow of water and oil through inhomogeneous porous media. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 12(3) :335–392, 1985.

[7] H. W. Alt, S. Luckhaus, and A. Visintin. On nonstationary flow through porous media. *Ann. Mat. Pura Appl. (4)*, 136 :303–316, 1984.

[8] B. Amaziane and M. Jurak. A new formulation of immiscible compressible two-phase flow in porous media. *Comptes Rendus Mécanique*, 336(7) :600–605, 2008.

[9] B. Andreianov, M. Bendahmane, and K. Karlsen. Discrete duality finite volume schemes for doubly nonlinear degenerate hyperbolic-parabolic equations. *J. Hyperbolic Diff. Equ. 7*, 1 :1–67, 2010.

[10] O. Angelini, K. Brenner, and D. Hilhorst. A finite volume method on general meshes for a degenerate parabolic convection-reaction-diffusion equation. 2011. to appear.

[11] O. Angelini, C. Chavant, E. Chénier, and R. Eymard. A finite volume scheme for diffusion problems on general meshes applying monotony constraints. *SIAM J. Numer. Anal.*, 47 :4193–4213, 2010.

[12] O. Angelini, C. Chavant, E. Chénier, R. Eymard, and S. Granet. Finite volume approximation of a diffusion-dissolution model and application to nuclear waste storage. *J. Math. and Comp. in Sim.*, 81(10), 2011.

[13] P. Angot, V. Dolejsi, M. Feistauer, and J. Felcman. Analysis of a combined barycentric finite volume - nonconforming finite element method for nonlinear convection - diffusion problem. *Applications of Mathematics*, 43 :263–310, 1998.

[14] S. N. Antontsev, A. V. Kazhikhov, and V. N. Monakhov. Boundary value problems in mechanics of nonhomogeneous fluids. *Studies in Mathematics and its Applications*, 22, 1990. Translated from the Russian.

[15] T. Arbogast. The existence of weak solutions to single porosity and simple dual porosity models of twophase incompressible flow. *Nonlinear Anal. Theory Methods Appl.*, 19 :1009–1031, 1992.

[16] T. Arbogast, M. F. Wheeler, and N. Zhang. A nonlinear mixed finite element method for a degenerate parabolic equation arising in flow in porous media. *SIAM J. Numer. Anal.*, 33 :1669–1687, 1996.

[17] K. Aziz and A. Settari. *Petroleum Reservoir Simulation.* Applied Science Publishers, 1979.

[18] G. E. K. B. Cockburn and C.-W. Shu. *Discontinuous Galerkin methods. Theory, Computation and Applications.* 2000.

[19] P. Bastian and B. Rivière. Discontinuous galerkin methods for two-phase flow in porous media. 2004. Technical report 2004-28, IWR (SFB 359).

[20] L. Baughman and N. Walkington. Co-volume methods for degenerate parabolic problems. *Numer. Math*, 64 :45–67, 1993.

[21] J. Bear. *Dynamics of fluids in porous media.* Elsevier, 1972.

[22] J. Bear and A. Verruijt. *Modeling groundwater flow and pollution.* Reidel, 1987.

[23] M. Berndt, K. Lipnikov, J. D. Moulton, and M. Shashkov. Convergence of mimetic finite difference discretizations of the diffusion equation. *EastWest J. Numer. Math.*, 9 :253–284, 2001.

[24] M. Bertsch, R. D. Passo, and C. J. van Duijn. Analysis of oil trapping in porous media flow. *SIAM J. Math. Anal.*, 35(1) :245–267, 2003.

[25] K. Brenner. Hybrid finite volume scheme for a two-phase flow problem. To appear.

[26] K. Brenner, C. Cancès, and D. Hilhorst. A convergent finite volume scheme for two-phase flows in porous media with discontinuous capillary pressure field. *Finite Volumes for Complex App. VI*, 1 :185–194, 2011.

[27] H. Brézis. *Analyse Fonctionnelle : Théorie et applications.* Masson, Paris, 1983.

[28] F. Brezzi, A. Buffa, and K. Lipnikov. Mimetic finite differences for elliptic problems. *Math. Mod. Numer. Anal.*, 43 :277–295, 2009.

[29] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods.* Springer-Verlag, 1991.

[30] 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.

[31] 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 :3682–3692, 2007.

[32] F. Brezzi, K. Lipnikov, and V. Simoncini. A family of mimetic finite difference methods on polygonal and polyhedral meshes. *Math. Models Methods Appl. Sci.*, 15(10) :1533–1551, 2005.

[33] F. Buzzi, M. Lenzinger, and B. Schweizer. Interface conditions for degenerate two-phase flow equations in one space dimension. *Analysis*, 29(3) :299–316, 2009.

[34] C. Cancès. Nonlinear parabolic equations with spatial discontinuities. *NoDEA Nonlinear Differential Equations Appl.*, 15(4-5) :427–456, 2008.

[35] C. Cancès. Asymptotic behavior of two-phase flows in heterogeneous porous media for capillarity depending only on space. i. convergence to the optimal entropy solution. *SIAM J. Math. Anal.*, 42(2) :946–971, 2010.

[36] C. Cancès. Asymptotic behavior of two-phase flows in heterogeneous porous media for capillarity depending only on space. ii. non-classical shocks to model oil-trapping. *SIAM J. Math. Anal.*, 42(2) :972–995, 2010.

[37] C. Cancès. On the effects of discontinuous capillarities for immiscible two-phase flows in porous media made of several rock-types. *Paper of the conference NTMC09, to appear in Networks and heterogeneous media*, 2010.

[38] C. Cancès and M. Pierre. An existence result for multidimensional immiscible two-phase flows with discontinuous capillary pressure field. 2010. In preparation.

[39] C. Cancés. Écoulement diphasiques en milieux poreux hétérogènes : modélisation et analyse. *Thèse de doctorat*, 2008.

[40] C. Cancès. Finite volume scheme for two-phase flow in heterogeneous porous media involving capillary pressure discontinuities. *M2AN*, 43 :973–1001, 2009.

[41] C. Cancès, T. Gallouët, and A. Porretta. Two-phase flows involving capillary barriers in heterogeneous porous media. *Interfaces and Free Bound.*, 2009 :239–258, 11.

[42] A. Cangiani, G. Manzini, and A. Russo. Convergence analysis of a mimetic finite difference method for general secondorder elliptic problems. *SIAM J. Numer. Anal.*, 47(4) :2612–2637, 2009.

[43] C. Chainais-Hillairet and J. Droniou. Convergence analysis of a mixed finite volume scheme for an ellipticparabolic system modeling miscible fluid flows in porous media. *SIAM J. Numer. Anal.*, 45(5) :2228–2258, 2007.

[44] G. Chavent and J. Jaffré. Mathematical models and finite elements for reservoir simulation. *NorthHolland*, 1986.

[45] Z. Chen. Degenerate two-phase incompressible flow. I. Existence, uniqueness and regularity of a weak solution. *J. Differential Equations*, 171(2) :203–232, 2001.

[46] Z. Chen and R. Ewing. Fullydiscrete finite element analysis of multiphase flow in groundwater hydrology. *SIAM J. Num . Anal.*, 34 :2228–2253, 1997.

[47] Z. Chen, R. Ewing, E. Jiang, and A. Spagnuolo. Error analysis for characteristics-based methods for degenerate parabolic problems. *SIAM J. Numer. Anal.*, 40 :1491–1515, 2002.

[48] Z. Chen and R. E. Ewing. Degenerate two-phase incompressible flow. III. Sharp error estimates. *Numer. Math.*, 90(2) :215–240, 2001.

[49] Z. Chen, G. Huan, and Y. Ma. *Computational methods for multiphase flows in porous media*. SIAM, 2006.

[50] Z. Chen and N. L. Khlopina. Degenerate two-phase incompressible flow problems. I. Regularization and numerical results. *Commun. Appl. Anal.*, 5(3) :319–334, 2001.

[51] Z. Chen and N. L. Khlopina. Degenerate two-phase incompressible flow problems. II. Error estimates. *Commun. Appl. Anal.*, 5(4) :503–521, 2001.

[52] Y. Coudière, J.-P. Vila, and P. Villedieu. Convergence rate of a finite volume scheme for a two dimensional diffusion convection problem. *M2AN Math. Model. Numer. Anal.*, 33 :493–516, 1999.

[53] C. Dawson. Analysis of an upwind-mixed finite element method for nonlinear contaminant transport equations. *SIAM J. Numer. Anal.*, 35 :1709–1724, 1998.

[54] C. Dawson and V. Aizinger. Upwind mixed methods for transport equations. *Comput. Geosci*, 3 :93–110, 1999.

[55] K. Deimling. *Nonlinear Functional Analysis*. Springer-Verlag, 1985.

[56] J. Droniou. Remarks on discretizations of convection terms in hybrid mimetic mixed methods. *Network and Heterogeneous Media 5*, 3, 2010.

[57] J. Droniou and R. Eymard. Study of the mixed finite volume method for stokes and navierstokes equations. *Numer. Meth. P. D. E.*, 25(1) :137–171, 2009.

[58] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin. A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods. *M3AS*. to appear.

[59] C. V. Duijn, J. Molenaar, and M. J. D. Neef. The effect of capillary forces on immiscible two-phase flow in heterogeneous porous media. *Transport in Porous Media*, 21 :71–93, 1995.

[60] G. Enchéry, R. Eymard, , and A. Michel. Numerical approximation of a two-phase flow in a porous medium with discontinuous capillary forces. *SIAM J. Numer. Anal.*, 43(6) :2402–2422, 2006.

[61] A. Ern, I. Mozolevski, , and L. Schuh. Discontinuous galerkin approximation of two-phase flows in heterogeneous porous media with discontinuous capillary pressures. *CERMICS Research Report 398*, 2009. submitted.

[62] B. G. Ersland, M. S. Espedal, and R. Nybø. Numerical methods for flow in a porous medium with internal boundaries. *Comput. Geosci.*, 2(3) :217–240, 1998.

[63] R. Ewing and R. F. Heinmann. Mixed finite element approximation of phase velocities in compositional reservoir simulation. *Comput. Methods Appl. Mech. Engrg. (R.E. Ewing, ed.)*, 47 :161–176, 1984.

[64] R. Ewing and M. F. Wheeler. Galerkin methods for miscible displacement problems with point sources and sinks–unit mobility ratio. *Mathematical Methods in Energy Research*, pages 40–59, 1983.

[65] R. Eymard, T. Gallouët, and R. Herbin. *Finite Volume Methods, Handbook of Numerical Analysis*, volume 7. P.G. Ciarlet and J.L. Lions eds, Elsevier Science B.V., 2000.

[66] 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. of Num. Anal.*, 30(4) :1009–1043, 2010.

[67] R. Eymard, T. Gallouët, R. Herbin, and A. Michel. Convergence of a finite volume scheme for nonlinear degenerate parabolic equations. *Numer. Math.*, 92 :41–82, 2002.

[68] R. Eymard, T. Gallouët, D. Hilhorst, and Y. N. Slimane. Finite volumes and nonlinear diffusion equations. *M2AN Math. Model. Numer. Anal.*, 32 :747–761, 1998.

[69] R. Eymard, M. Gutnic, and D. Hilhorst. The finite volume method for the richards equation. *Comput. Geosci.*, 3 :259–294, 2000.

[70] R. Eymard, G. Henry, R. Herbin, F. Hubert, R. Klofkorn, and G. Manzini. 3d benchmark on discretization schemes for anisotropic diffusion problems on general grids. *Finite Volumes for Complex App. VI*, 2 :95–130, 2011.

[71] R. Eymard, R. Herbin, and A. Michel. Mathematical study of a petroleum-engineering scheme. *Math. Mod. Numer. Anal.*, 37(6) :937–972, 2003.

[72] R. Eymard, D. Hilhorst, and M. Vohralík. A combined finite volume–nonconforming/mixed-hybrid finite element scheme for degenerate parabolic problems. *Numer. Math.*, 105 :73–131, 2006.

[73] R. Eymard, D. Hilhorst, and M. Vohralík. A combined finite volume-finite element scheme for the discretization of strongly nonlinear convection-diffusion-reaction problems on nonmatching grids. *Numer. Methods for Partial Differ. Equations*, 26 :612–646, 2009.

[74] M. Feistauer, J. Felcman, and M. Lukacova-Medvidova. On the convergence of a combined finite volume-finite element method for nonlinear convection-diffusion problems. *Numer. Methods for Partial Differ. Equations*, 13 :163–190, 1997.

[75] T. Gallouët. Hyperbolic equations and systems with discontinuous coefficients or source terms. *Proceedings of Equadiff-11*, Bratislava, 2005.

[76] A. Glitzky and J. Griepentrog. Discrete sobolev-poincaré inequalities for voronoi finite volume approximations. *SIAM J. Numer. Anal.*, 48 :2489–2513, 2010.

[77] S. Hassanizadeh and W. Gray. Mechanics and thermodynamics of multiphase flow in porous media including interphase boundaries. *Adv. Water Resour.*, 13 :169–186, 1990.

[78] S. Hassanizadeh and W. Gray. Thermodynamic basis of capillary pressure in porous media. *Water Resources Res.*, 29 :3389–3405, 1993.

[79] R. Herbin and F. Hubert. Benchmark on discretization schemes for anisotropic diffusion problems on general grids for anisotropic heterogeneous diffusion problems. *Finite Volumes for Complex App. V*, pages 659–692, 2008.

[80] H. Holden, K. H. Karlsen, and K.-A. Lie. Operator splitting methods for degenerate convection-diffusion equations ii : Numerical examples with emphasis on reservoir simulation and sedimentation. Technical report, Computational Geosciences, 2000.

[81] J. Hyman, J. Morel, M. Shashkov, and S. Steinberg. Mimetic finite difference methods for the diffusion equation. *Comput. Geosci.*, 6 :333–352, 2002.

[82] J. Jaffré. Numerical calculation of the flux across an interface between two rock types of a porous medium for a two-phase flow. *Hyperbolic Problems : Theory, Numerics, Applications, J. Glimm, M. Graham, J. Grove, and B. Plohr, eds.*, pages 165–177, 1996.

[83] K. H. Karlsen, N. H. Risebro, and J. D. Towers. Upwind difference approximations for degenerate parabolic convection-diffusion equations with a discontinuous coefficient. *IMA J. Numer. Anal.*, 22 :623–664, 2002.

[84] J. Kačur. Solution of degenerate convection-diffusion problems by the method of characteristics. *SIAM J. Numer. Anal.*, 39 :858–879, 2001.

[85] J. Kačur and R. van Keer. Solution of contaminant transport with adsorption in porous media by the method of characteristics. *ESAIM : M2AN Math. Model. Numer. Anal.*, 35 :981–1006, 2001.

[86] P. Knabner and F. Otto. Solute transport in porous media with equilibrium and nonequilibrium multiple-site adsorption : uniqueness of weak solutions. *Nonlinear Anal.*, 42 :381–403, 2000.

[87] R. H. L. Agelas, R. Eymard. A nine-point finite volume scheme for the simulation of diffusion in heterogeneous media. *Comptes Rendus Mathématique*, 347(11–12) :673–676, 2009.

[88] K. Lipnikov, M. Shashkov, and I. Yotov. Local flux mimetic finite difference methods. *Numer. Math.*, 112 :115–152, 2009.

[89] A. C. G. Manzini. Flux reconstruction and pressure postprocessing in mimetic finite difference methods. *Comput. Methods Appl. Mech. Engrg.*, 197/9-12 :933–945, 2008.

[90] A. Michel. A finite volume scheme for two-phase immiscible flow in porous media. *SIAM J. Numer. Anal.*, 41(4) :1301–1317, 2003. (electronic).

[91] K. Nikitin. Nonlinear finite volume method for two phase flow in porous media. *Math. Model.*, 20 :11 :131–147, 2010.

[92] P. Nochetto and C. Verdi. Approximation of degenerate parabolic problems using numerical integration. *SIAM J. Numer. Anal.*, 25 :784–814, 1988.

[93] R. H. Nochetto, A. Schmidt, and C. Verdi. A posteriori error estimation and adaptivity for degenerate parabolic problems. *Math. Comput.*, 69 :1–24, 2000.

[94] M. Ohlberger. Convergence of a mixed finite element - finite volume method for the two phase flow in porous media. *East-West J. Numer. Math.*, 5(3) :183–210, 1997.

[95] D. Pavone. Macroscopic equations derived from space averaging for immiscible two-phase flow in porous media. *Revue de l'Institut français du pétrole*, 44(1) :29–41, janvier-février 1989.

[96] D. W. Peaceman. *Fundamentals of Numerical Reservoir Simulation.* Elsevier Scientific Publishing Co., 1977.

[97] B. Riviere and N. Walkington. Convergence of a discontinuous galerkin method for the miscible displacement under low regularity. *SIAM J. Numer. Anal.* to appear.

[98] J. Rulla and N. J. Walkington. Optimal rates of convergence for degenerate parabolic problems in two dimensions. *SIAM J. Numer. Anal.*, 33 :56–67, 1996.

[99] N. Seguin and J. Vovelle. Analysis and approximation of a scalar conservation law with a flux function with discontinuous coefficient. *M3AS*, 13(2) :221–257, 2003.

[100] J. D. Towers. Convergence of a difference scheme for conservation laws with a discontinuous flux. *SIAM J. Numer. Anal.*, 38(2) :681–698, 2000. (electronic).

[101] L. B. D. Veiga, J. Droniou, and G. Manzini. A unified approach for handling convection terms in finite volumes and mimetic discretization methods for elliptic problems. *IMA J. Numer. Anal.*, DOI : 10.1093/imanum/drq018 :1–45, 2010.

[102] L. B. D. Veiga, K. Lipnikov, and G. Manzini. Convergence analysis of the high-order mimetic finite difference method. *Numer. Math.*, 113(3) :325–356, 2009.

[103] L. B. D. Veiga and G. Manzini. An a posteriori error estimator for the mimetic finite difference approximation of elliptic problems. *Int. J. for Numer. Meth. in Engnr.*, 76(11) :1696–1723, 2008.

[104] L. B. D. Veiga and G. Manzini. A higherorder formulation of the mimetic finite difference method. *SIAM J. Sci. Comput.*, 31(1) :732–760, 2008.