Méthode de Galerkin discontinue pour un modèle stratigraphique

Abdelaziz Taakili

ACADÉMIE DE BORDEAUX |_|_|_|_|_|_|_|_|_|_|

THÈSE

présenté à

l'Université de Pau et des Pays de l'Adour

École Doctorale des Sciences exactes et leurs applications

par

Abdelaziz TAAKILI

pour obtenir le grade de

Docteur

Discipline : **Mathématiques**

MÉTHODE DE GALERKIN DISCONTINUE POUR UN MODÈLE STRATIGRAPHIQUE

Soutenue le 2 juillet 2008 devant le jury composé de

M.	Roland	BECKER	Directeur
M.	Abderrahmane	BENDALI	Président, Examinateur
M.	Robert	EYMARD	Rapporteur
M.	Gérard	GAGNEUX	Examinateur
M.	Peter	HANSBO	Rapporteur
M.	Guy	VALLET	Co-directeur

Après avis des rapporteurs : M. Robert EYMARD
M. Peter HANSBO

-2008-

À mes parents

À mes soeurs et frères

Remerciement

Je voudrais tout d'abord remercier mes directeurs de thèse, Monsieur Roland Becker et Monsieur Guy Vallet. Je leurs suis très reconnaissant d'avoir été des directeurs de thèse très responsable. Leur expérience, leur disponibilité et leur confiance m'ont permis de mener à bien ce travail.

Je suis très sensible à l'honneur que m'ont fait Monsieur Robert Eymard et Monsieur Peter Hansbo en acceptant d'être les rapporteurs de cette thèse et de participer au jury. J'apprécie sincèrement leur intérêt pour mes travaux ainsi que leurs commentaires avisés. Je voudrais aussi exprimer ma gratitude profonde à Monsieur Abderrahmane Bendali pour avoir accepté d'être le président du jury et à Monsieur Gérard Gagneux pour avoir accepté d'être examinateur et membres du jury.

Je tiens aussi à exprimer ma reconnaissance au Directeur du Laboratoire de Mathématiques Appliquées de l'Université de Pau, Monsieur Mohamed Amara ainsi qu'au nouveau Directeur, Monsieur Laurent Bordes qui m'ont permis de bénificier de bonnes conditions de ce travail.

Je voudrais exprimer mes remerciements profonds à mes parents et à toute ma famille pour leur soutien permanent, leurs encouragements et leur amour.

Mes remerciement vont à mes chers amis de Pau : A. Mokrani, R. Bouchouirbat, A. El abdellaoui, M. El ossmani, M. Jamil, Rachid et A. Ezziani ainsi qu'à mes amies: Nour el-houda, N. Khadra et H. Karim. Ils m'ont accompagné durant mon chemin et ils m'ont toujours aidé lorsque j'ai eu besoin d'eux.

Enfin, je tiens à remercier très chaleureusement tous mes collègues doctorants et ATERs du Laboratoire de Mathématiques Appliquées. Je leurs souhaite une bonne continuation professionnelle.

Méthode de Galerkin discontinue pour un modèle stratigraphique

Résumé

Dans cette thèse, on considère un modèle stratigraphique issu de la modélisation géologique de la formation d'un bassin sédimentaire. Ce modèle original a été initialement développé par l'Institut Français du Pétrole (IFP). Il décrit la formation de bassins sédimentaires à une seule lithologie ; il décrit également le transport et l'accumulation de sédiments et prend en compte les phénomènes d'érosion, de sédimentation et les apports de sédiments aux frontières du bassin. L'aspect mathématique original de ce problème réside dans l'imposition d'une contrainte sur le taux d'érosion qui doit rester inférieur à une fonction donnée E. Ce qui nous amène à considérer une loi de conservation de type dégénéré.
Ce travail est organisé comme suit: dans le **Chapitre 1** le DgFem (Discontinuous Galerkin Finite Element Method) pour un problème elliptique est présenté avec une estimation *a priori* de l'erreur. Nous considérons le problème suivant : comment choisir le paramètre γ apparaissant dans le terme de stabilisation. On remarque par ailleurs que la solution DgFem converge vers la solution éléments finis conformes lorsque γ tend vers l'infini. Ce résultat théorique est confirmé par des tests numériques. Nous terminons ce chapitre par des techniques standards "red-green refinement" pour le raffinement local du maillage. Dans le **Chapitre 2**, on présente le modèle mathématique. On établit l'existence d'une solution au problème par le biais d'une discrétisation implicite en temps, en établissant des estimations *a priori* appropriées. Le **chapitre 3** concerne la discrétisation en temps avec une méthode de Galerkin discontinue pour une version linéarisée simplifiée du problème. Nous introduisons tout d'abord la discrétisation en temps, nous établissons une estimation *a priori* de l'erreur. À la fin du chapitre, nous proposons quelques résultats numériques.
Le **chapitre 4** traite de la discrétisation du problème stratigraphique : nous utilisons un schéma Dg(0) implicite en temps et un schéma DgFem(p), $p \geq 0$ pour la variable d'espace. En ce qui concerne la discrétisation de l'espace, le choix des flux à l'interface entre deux éléments du maillage est très important, d'autant plus que la diffusion introduit un terme non-linéaire et non-négatif. La moyenne pondérée est préférée à la moyenne arithmétique classiquement utilisée dans les méthodes SIPG. Ce choix est justifié par la présence d'une frontière libre provenant de la contrainte. Dans le cas $p = 0$, nous prouvons que le schéma vérifie implicitement la contrainte, ce qui rend cette méthode compatible avec la modélisation continue. L'existence et l'unicité d'une solution discrète sont prouvées. Des simulations numériques sont présentées. Nous terminons le chapitre avec un algorithme adaptatif combinant les schéma DgFem(0) et DgFem(p), $p \geq 1$.

A discontinuous Galerkin method for a model from stratigraphy

Abstract

In this thesis, we consider a stratigraphic model arising from the modelling of geological basin formation. This model has been initially developed by the Institut Français du Pétrole (see [30]). It takes into account sedimentation, transport and accumulation, erosion phenomena, and others. The original mathematical aspect of this model is the imposition of a constraint on the time-derivative of the unknown u. This leads to the consideration of a class of conservation laws of degenerate type.

This work is organized as follows: in **Chapter 1**, some notation used in this thesis are collected. Then, the DgFem for an elliptic problem is presented with some *a priori* error estimates. We consider the problem: how to choose the parameter γ involved in the stabilization term. We remark that the DgFem solution converges to the conforming finite element solution when the parameter γ tends to infinity. This theoretical result is confirmed by numerical tests. This chapter ends with some standard techniques, used along this thesis for local mesh refinement based on red-green refinement. In **Chapter 2**, the mathematical model of stratigraphy is presented. An existence result of a solution of problem is proved by means of an implicit time-discretization, establishing appropriate *a priori* estimations.

Chapter 3 is concerned with the time discontinuous Galerkin discretization of the Sobolev equation which is a simple case of the stratigraphic model. We first introduce the time discretization, then we give an *a priori* error analysis. At the end of the chapter, we propose some numerical results.

Chapter 4 deals with the discretization of the stratigraphic problem: an implicit Dg(0) scheme in time and a DgFem(p), $p \geq 0$ scheme in space. Concerning the space discretization, the choice of the flux at the interface between two elements is very important, especially since the diffusion term is a nonlinear and nonnegative term. A weighted average is preferred to the arithmetic one, classically used in SIPG methods. This choice is justified by the presence of the free set discriminated by the constraint. In the case of lowest order case, we prove that the scheme satisfies implicitly the constraint, which makes this method compatible with the continuous model. We prove that a discrete solution exists and is unique as soon as τ is greater than a positive threshold τ^*. Some numerical simulations are presented. We finish the chapter with a p-adaptive algorithm combining the DgFem(0) and DgFem(p), $p \geq 1$ schemes.

Contents

Introduction

The modelling of sedimentary basins allows to recall the history of hydrocarbons since their genesis. It takes into account many physical and geological phenomena such as: erosion, compaction, sedimentation, formation of faults, plate tectonics play, chemical reactions, etc. This modelling is of primary importance in petroleum engineering because it allows, in particular, an effective selection of the sites of drillings. Other fields are also covered by this modelling, like the study of pollution risks in the ground by one or more contaminants, of research miners, and others. The computation of sedimentation and erosion processes lead to a better knowledge of the geometry of the layers, and of their lithological nature (see, for example [25]). In this thesis, we consider a stratigraphic model arising from the modelling of geological basin formation. This model has been initially developed by the Institut Français du Pétrole (see [30]). It takes into account sedimentation, transport and accumulation, erosion phenomena, and others. The original mathematical aspect of this model is the imposition of a constraint on the time-derivative of the unknown u. This leads to the consideration of a class of conservation laws of degenerate type.
For more information on the physical descriptions and numerical aspects of the monolithological and multilithological cases of these models see R. Eymard *et al.* [26], [24], and V. Gervais et al. [29].

Concerning the theoretical aspects of the monolithological case, S. N. Antontsev *et al.* (see [6]) propose a new conservative formulation, which is equivalent to the model proposed by R. Eymard *et al.* (see [26], [24]):

$$\partial_t u - \operatorname{div}[\lambda K(x)\nabla u] = 0,\ \lambda \in H(\partial_t u + E), \tag{1}$$

where, H denotes the maximal monotone graph of the Heaviside function, u denotes the thickness of sediment and E denotes the admissible erosion rate. The advantage of this formulation is that it implicitly contains the constraint

$$\partial_t u + E \geq 0. \tag{2}$$

We propose in Chapter 2 a description of the model.
The existence and the uniqueness of a solution to this problem is still an open problem.

Indeed, on the one hand, these "new unilateral problems" proposed by J.-L. Lions in [35] p. 420 where the constraint is imposed on the time-derivative of the solution and not on the solution itself are not much developed in the literature. On the other hand, a nonlinear function of $\partial_t u$ as a viscosity term is a delicate problems, with local hyperbolic behaviour and hysteretic effects (see [4]).

The mathematical analysis of an implicit time-discretization of the problem and in particular an explicit study of the one-dimensional sedimentation has been considered by S. N. Antontsev, G. Gagneux, R. Luce and G. Vallet (see [3]).

In the framework of a perfect physical equilibrium (see D. Granjeon [31]), the flux $\vec{q}$ is considered to be proportional to the slope by writing that $\vec{q} \propto -\nabla u$. Taking into account a balance of the slope, and, according to a Darcy-Barenblatt law (see [11]), one considers that $\vec{q} \propto -\nabla(u + \tau \partial_t u)$, where $\tau > 0$ is a time-scale. Then, we get the following problem:

$$\partial_t u - \mathrm{div}[\lambda K(x) \nabla(u + \tau \partial_t u)] = 0, \quad \lambda \in H(\partial_t u + E). \tag{3}$$

Existence and uniqueness results of the solution to the above differential inclusion are still open problems. A modified model where H is replaced by a Lipschitz continuous function a, for example the Yosida approximation of H, is analyzed by S. N. Antontsev *et al.* [4]. An existence result of a solution to this nonlinear degenerate pseudoparabolic[1] equation by means of an adapted compactness argument is considered when $K = 1$. Local hyperbolic behaviour is proved too. Thanks to the regularity of the solutions (*i.e.* u and $\partial_t u$ in $H^1(\Omega)$), it can be shown that the problem is equivalent to the following one:

$$\partial_t u - \mathrm{div}[\lambda K \nabla u + \tau K \nabla \partial_t u] = 0, \quad \lambda \in H(\partial_t u + E). \tag{4}$$

Recently, the results on existence and uniqueness are generalized by S. N. Antontsev *et al.* [5] to nonlinear diffusion coefficients K, time-dependent E and $\lambda = a(\partial_t u + E)$.

Equations of pseudoparabolic type are used in the modelling of many different physical phenomena such as:

- the theory of porous media (see G. I. Barenblatt [11], C. Cuesta *et al.* [17], R. E. Ewing [23] and J. Garcia-Azorero *et al.* [28]),
- in aggregation of populations (V. Padron [36]),
- in solvent uptake in polymeric solids (W. P. Düll [19]),
- singular perturbation problems (G. I. Barenblatt *et al.* [12], R. E. Ewing [22] and P. I. Plotnikov [37]).

[1] A problem with the time-derivative of the solution in the second order operator.

In this thesis, we extend the results of the above cited articles to more general data. In particular, we consider a source term f and a space-time dependent function E.
The main contribution of this thesis is the development of a robust and efficient numerical scheme for the computation of the approximate solution of the stratigraphic problem (4), with $\lambda = a(\partial_t u + E)$.

Our approach is based on the discontinuous Galerkin finite element method (DgFem). The choice of this method is motivated by its flexibility and robustness to treat equations with varying and degenerate coefficients. In the lowest order case ($p = 0$), it coincides with the finite volume discretization used before to treat the stratigraphic equation.
The piecewise polynomial trial functions are discontinuous. Approximate continuity is imposed by the use of appropriate penalty terms involving jumps of the function values of inter-element edges. These terms have to take into account the character of the continuous operators. Therefore, special care has to be taken in order to discretize the novel aspects of the problem: the time-dependent and degenerate diffusion coefficient, as well as the pseudoparabolic regularization.
The primary motivation for DgFem is its robustness and the possibility to obtain higher-order discretizations. Moreover, the local nature of the trial spaces enables us to locally adapt the approximation order (hp-methods), which seems to be of interest for the problem (4) and more generally for parabolic equations with dominant transport terms where solutions vary rapidly on small parts of the domain. Another advantages of the DgFem method is that it leads to a natural flux function, defined on each edge of the triangulation of the domain, and which satisfies an element-wise conservation law, a property desired in many applications. Finally, the structure of the mass matrices (block diagonal) is an attractive feature in the context of time-dependent problems, especially if explicit time discretizations are used. For the above mentioned reasons, there has been an increased interest in DgFem and different variants have been developed. In the last ten years, a unified analysis of the different DgFem methods has been developed, for an overview see the article of Arnold, Brezzi, Cockburn and Marini [8].

Among the DgFem variant, we are interested in this work by the SIPG (Symmetric Interior Penalty Galerkin) method, established following the works of Baker [10], Douglas and Dupont [18], Wheeler [9] and Arnold [7]. The bilinear form is symmetric, and the jumps of the approximate solution, as well as the Dirichlet boundary conditions, are penalized. The stabilization parameter has to be large enough. The NIPG method of Rivière, Wheeler and Girault [38] is very similar, except the sign of one term. In this case, the positivity of the bilinear form is provided for $p \geq 2$, without any stabilization terms. A disadvantage of the NIPG method, unlike the SIPG one, lies in the fact that the convergence of the error in L^2 norm is in general not increasing, even under the elliptic regularity hypothesis.

In this thesis, we first consider the simple case of the stratigraphic problem when the con-

straint is inactive ($\partial_t u + E > 0$). Then, the problem is reduced to a linear pseudoparabolic problem, which is also known as the Sobolev equation.
The nature of such problems is transient and, therefore, an appropriate time stepping scheme has to be applied in numerical simulations in order to obtain an approximate solution. A flexible and robust time discretization method is the Dg(r) (discontinuous Galerkin) which is based on a variational formulation of the initial value problems and approximation by piecewise polynomial of degree r in time. We develop a generalization of this well-known scheme for the heat equation to the Sobolev equation. Then, error estimates, explicit in the polynomial degrees of order r and time step k, are derived. Some numerical experiments illustrate the theoretical results.

For the stratigraphic problem with constraint, the implicit Dg(0) scheme is used in time. Concerning the space discretization, a variant of the SIPG method is proposed. A difference with the above-cited works, dealing with constant coefficients, appears in the choice of the consistency terms where the arithmetic average usually used in SIPG is replaced by the weighted averages involving the nonlinear diffusion coefficient a, a similar discretization has been used by A. Ern *et al.* [21] for a stationary convection-diffusion problem with discontinuous diffusion coefficients. This leads us to consider the harmonic average of the function a and the arithmetic average of the normal derivative of the unknown, in order to impose the continuity of the flux at the interface between two mesh elements. This choice is motivated by the nature of the problem since it may degenerate in order to take into account the constraint. By allowing the solution to be discontinuous at the interface, the penalization term is weighted by the harmonic average of the function a. This choice leads in the lowest-order case, DgFem(0), to discrete solution which implicitly satisfy the constraint. Existence of the discrete solution for general p is also established if $\tau > 0$ and its uniqueness is shown when τ is bigger than a given positive threshold τ^*.

It is well known that higher order discontinuous Galerkin methods do not respect the maximum principle. Our aim is then to propose an adaptive algorithm that combines DgFem(0) scheme and DgFem(p) $p \geq 1$ scheme. This is done with the objective to obtain higher-order accuracy, while still verifying the constraint. In this work, a p-adaptive algorithm for a stratigraphic problem is presented. The idea of the algorithm is to first solve the problem using higher-order DgFem, and then, using an interface indicator, to reduce the approximation order of the selected elements to zero.
Let us now present the detailed plan of each chapter.

This work is organized as follows: in **Chapter 1**, some notations used in this thesis are collected. Then, the DgFem for an elliptic problem is presented with some *a priori* error estimates. We consider the problem: how to choose the parameter γ involved in the stabilization term. It depends on the solution and on the inverse estimate constant, which itself depends on the approximation order. Then, we remark that the discontinuous

Galerkin solution converges to the conforming finite element solution when the parameter γ tends to infinity. This theoretical result is confirmed by numerical tests.
This chapter ends with some standard techniques, used along this thesis, as local mesh refinement based on red-green refinement and triangular element with arbitrary order (see R. Verfürth [41]).
In **Chapter 2**, the mathematical model of stratigraphy is presented. An existence result of a solution of problem (4) is proved by means of an implicit time-discretization, establishing appropriate *a priori* estimations.
Chapter 3 is concerned with the time discontinuous Galerkin discretization of the Sobolev equation which is a simple case of the stratigraphic model. We first introduce the time discretization, then we give an *a priori* error analysis. With a particular choice of the projection operator, we prove that the $\mathrm{Dg}(r)$ time error is controlled by the interpolation error. At the end of the chapter, we propose some numerical results.
Chapter 4 deals with the discretization of the stratigraphic problem: an implicit Dg(0) scheme in time and a DgFem(p) $p \geq 0$ scheme in space. Concerning the space discretization, the choice of the flux at the interface between two elements is very important, especially since the diffusion term is a nonlinear and nonnegative term. A weighted average is preferred to the arithmetic one, classically used in SIPG methods. This choice is justified by the presence of the free set discriminated by the constraint. In the case of lowest order DgFem(0) discretization, as in the continuous case, we prove that the scheme satisfies implicitly the constraint, which makes this method compatible with the continuous model. We prove that a discrete solution exists and is unique as soon as τ is greater than a positive threshold τ^*. Some numerical simulations using DgFem(p) scheme, $p \geq 0$, are presented. We finish the chapter with a p-adaptive algorithm combining the DgFem(0) and DgFem(p) $p \geq 1$ schemes. The aim of this algorithm is to impose $p = 0$ at any element incompatible with the constraint. Some numerical simulations are presented. The numerical tests have been realized with a C^{++} code based on the library **Concha**.

Chapter 1

Notation and preliminary results

1.1 Introduction

In the following section, notations used along this work are collected. In section 3.2, we first present the DgFem for elliptic problem then an *a priori* error analysis is developed. We remark that the DgFem solution coincides with the Fem solution when the penalty parameter γ tends to infinity. This is confirmed by some numerical results. In section 1.5, a h-adaptive algorithm is presented, based on residual *a posteriori* error estimates (see [33] for the Laplacian case and [20] for the advection-reaction-diffusion equations with anisotropic and discontinuous diffusivity). Implementation issues concerning the red-green techniques [41] for the local mesh refinement are described. Some numerical examples are shown.

1.2 Notation and preliminaries

We suppose that $\Omega \subset \mathbb{R}^2$ is a bounded polygonal domain, and that h is a triangular mesh in a family of shape-uniform meshes [16].
We denote by $\mathcal{K}_h$ the set of triangles and by $\mathcal{S}_h$ the set of edges; divided into interior edges $\mathcal{S}_h^{int}$ and boundary edges $\mathcal{S}_h^{\partial}$. An interior edges $S \in \mathcal{S}_h^{int}$ is shared by two triangles. We arbitrarily chose a normal n_S pointing from K^+ to K^-, see Figure 1.1. In $\mathcal{K}_h$, we consider the space $H^m(\mathcal{K}_h)$, $m \geq 1$, defined by

$$H^m(\mathcal{K}_h) = \{v \in L^2(\Omega) \text{ such that } v|_K \in H^m(K) \text{ for all } K \in \mathcal{K}_h\}. \tag{1.1}$$

For $p \in \mathbb{N}$, we define the discontinuous finite element space :

$$V_h^p = \left\{v_h \in L^2(\Omega) : \ v_h|_K \in P^{p_K} \text{ for all } K \in \mathcal{K}_h\right\}, \tag{1.2}$$

where, for any integer k, P^k is the space of polynomial functions of maximal degree k.

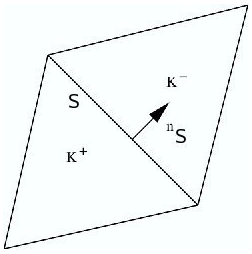


Figure 1.1: Two adjacent triangles sharing edge S.

Due to the discontinuity of the approximation space, the weak formulation reveals jumps terms through the cell interfaces. We use the standard notation concerning the jumps and averages for $v_h \in V_h^p$, $S \in \mathcal{S}_h^{int}$, and $x \in S$:

$$v_h^{\pm}(x) \;=\; \lim_{\varepsilon \to 0^+} v_h(x \mp \varepsilon n_S), \quad [v_h]_S = v_h^+ - v_h^-.$$

For a boundary edge we set $[v_h]_S := v_h^-$.
In addition, let κ be a bounded positive piecewise continuous function with respect to h. We define the weighted average of $v_h \in V_h$ on an interior edge S by

$$\{\frac{\partial v_h}{\partial n_S}\}_{S,\kappa} := \frac{\kappa^-}{\kappa^+ + \kappa^-}\kappa^+ \frac{\partial v_h^+}{\partial n_S}|_S + \frac{\kappa^+}{\kappa^+ + \kappa^-}\kappa^- \frac{\partial v_h^-}{\partial n_S}|_S, \tag{1.3}$$

we observe that:

$$\{\frac{\partial v_h}{\partial n_S}\}_{S,\kappa} = \frac{\kappa^-\kappa^+}{\kappa^+ + \kappa^-}(\frac{\partial v_h^+}{\partial n_S}|_S + \frac{\partial v_h^-}{\partial n_S}|_S). \tag{1.4}$$

For convenience, we extend the above definition to the boundary edges as follow: for $S \in \mathcal{S}_h^{\partial}$, we set $\{\frac{\partial v_h}{\partial n}\}_{S,\kappa} = \kappa \frac{\partial v_h}{\partial n_S}|_S$.
For a constant function κ, we get the standard arithmetic average on an interior edge S as

$$\{\frac{\partial v_h}{\partial n_S}\}_{S,\kappa} := \frac{\kappa}{2}\frac{\partial v_h^+}{\partial n_S}|_S + \frac{\kappa}{2}\frac{\partial v_h^-}{\partial n_S}|_S. \tag{1.5}$$

In addition, we introduce the following notations:

- x_K is the orthocenter of $K \in \mathcal{K}_h$,
- x_S is the barycenter of edge $S \in \mathcal{S}_h$,
- $h_S = |x_{K^+} - x_{K^-}|$ for $S \in \mathcal{S}_h^{int}$,
- $h_S = |x_{K^-} - x_S|$ for $S \in \mathcal{S}_h^{\partial}$,

Let us recall this well-known variant of Poincaré's inequality (see [15]) :

Lemma 1.1 *For all $v_h \in V_h^p$, we have*

$$\|v_h\|_{0,\Omega}^2 \leq C \left\{ \sum_{S \in \mathcal{S}_h^{int}} \frac{1}{h_S} \|[v_h]\|_S^2 + \sum_{S \in \mathcal{S}_h^{\partial}} \frac{1}{h_S} \|v_h\|_S^2 + \sum_{K \in \mathcal{K}_h} \|\nabla v_h\|_K^2 \right\}, \tag{1.6}$$

where C is a constant depending on Ω.

The following inverse estimate holds (see [40]): For all $K \in \mathcal{K}_h$, and for all $v_h \in V_h^p$, there is a constant C depends only on p and the minimum angle such that

$$\|v_h\|_{0,\partial K} \leq C\, h_K^{-1/2} \|v_h\|_{0,K}. \tag{1.7}$$

1.3 DgFem discretization

1.3.1 Model problem

We consider the following elliptic problem with homogeneous Dirichlet boundary conditions:

$$\begin{aligned} -\operatorname{div}(\kappa \nabla u) + \alpha u &= f \text{ in } \Omega, \\ u &= 0 \text{ on } \partial\Omega. \end{aligned} \tag{1.8}$$

We assume that the constant α satisfies $\alpha \geq 0$, the diffusivity $\kappa \in L^\infty(\Omega)$ is a scalar function such that $0 < \kappa_{min} \leq \kappa \leq \kappa_{max}$ and $f \in L^2(\Omega)$. The weak formulation of (1.8) is: find $u \in H_0^1(\Omega)$ such that

$$\int_\Omega \kappa\, \nabla u \cdot \nabla v\, dx + \alpha \int_\Omega u\, v\, dx = \int_\Omega f\, v\, dx \quad \forall v \in H_0^1(\Omega). \tag{1.9}$$

The existence and the uniqueness of u is given by the Lax-Milgram theorem.

1.3.2 Discrete formulation

For the discretization of (1.9), we introduce the following bilinear form for given u_h and v_h in V_h^p:

(1.10)

$$\begin{aligned} A(u_h, v_h) := \sum_{K \in \mathcal{K}_h} \int_K \kappa\, \nabla u_h \cdot \nabla v_h\, dx + \alpha \sum_{K \in \mathcal{K}_h} \int_K u_h\, v_h\, dx - \sum_{S \in \mathcal{S}_h} \int_S \{\frac{\partial u_h}{\partial n_S}\}_{S,\kappa} [v_h]_S\, ds \\ - \sum_{S \in \mathcal{S}_h} \int_S [u_h]_S \{\frac{\partial v_h}{\partial n_S}\}_{S,\kappa}\, ds + \sum_{S \in \mathcal{S}_h} \frac{\gamma}{h_S} \int_S \gamma_S [u_h][v_h]\, ds, \end{aligned}$$

where $2\gamma_S$ is the harmonic average of κ and the penalty parameter γ is assumed to be large enough. Defining the linear form:

$$L(v_h) := \int_\Omega f\, v_h\, dx, \tag{1.11}$$

the discrete problem consists on finding u_h in V_h^p such that:

$$A(u_h, v_h) = L(v_h), \quad \forall\, v_h \in V_h^p. \tag{1.12}$$

We equip the finite element space V_h^p with the following norms:

$$\|v\|_{1,h}^2 := \|v\|_{0,\Omega}^2 + \|\kappa^{1/2}\nabla v\|_{0,\Omega}^2 + |h_S^{-1/2}[v]_S|_{\gamma_S}^2,$$

and

$$\|v\|_{h,A}^2 := \|v\|_{1,h}^2 + \sum_{S\in\mathcal{S}_h} h_S \|\frac{\partial v}{\partial n_S}\|_{0,S}^2,$$

where $|v|_{\gamma_S}$ is given by

$$|v|_{\gamma_S}^2 = \sum_{S\in\mathcal{S}_h} \int_S \gamma_S\, v^2\, ds.$$

As consequence of the inverse estimate (1.7), the norms $\|.\|_{1,h}$ and $\|.\|_{h,A}$ are equivalent on the subspace V_h^p. The stability of the method can now be proved.

Lemma 1.2 *(Coercivity) There exists a constant $C > 0$ such that for all v_h in V_h^p*

$$A(v_h, v_h) \geq C\|v_h\|_{h,A}^2. \tag{1.13}$$

Proof. Let $v_h \in V_h^p$. We have

$$A(v_h, v_h) = \|\kappa^{1/2}\nabla v_h\|_{0,\Omega}^2 + \alpha\|v_h\|_{0,\Omega}^2 - 2\sum_{S\in\mathcal{S}_h}\int_S \{\frac{\partial v_h}{\partial n_S}\}_{S,\kappa}[v_h]_S\, ds + \gamma|h_S^{-1/2}[v_h]|_{\gamma_S}^2. \tag{1.14}$$

Consider now the third term in the right hand side. Let $S \in \mathcal{S}_h$. Using Young's inequality, we get

(1.15)

$$\begin{aligned} -2\int_S \{\frac{\partial v_h}{\partial n_S}\}_{S,\kappa}[v_h]_S\, ds &\geq -h_S\varepsilon \int_S \frac{\kappa^+ + \kappa^-}{\kappa^+\kappa^-}|\{\frac{\partial v_h}{\partial n_S}\}_{S,\kappa}|^2\, ds - \frac{1}{\varepsilon\, h_S}\int_S \frac{\kappa^+\kappa^-}{\kappa^+ + \kappa^-}[v_h]_S^2\, ds \\ &\geq -\varepsilon C\|\kappa^{1/2}\nabla v_h\|_{0,K^+\cup K^-}^2 - \frac{1}{\varepsilon}|h_S^{-1/2}[v_h]_S|_{\gamma_S}^2, \end{aligned}$$

where in the last step, we have used the inequality (1.7). Choosing ε such that $1 - \varepsilon\, C = c$, and taking $\gamma_0 = c + \varepsilon^{-1}$, we obtain:

$$A(v_h, v_h) \geq c\|v_h\|_{1,h}^2 + (\gamma - \gamma_0)|h_S^{-1/2}[v_h]_S|_{\gamma_S}^2, \tag{1.16}$$

the assertion follows by chosen γ large enough. □

Lemma 1.3 *(Consistence) Let u be the solution of (1.9) and u_h the solution of (1.12). Assume that $u \in H^2(\mathcal{K}_h)$. Then, for all v_h in V_h^p*

$$A(u - u_h, v_h) = 0. \tag{1.17}$$

Proof. Let $v_h \in V_h^p$. Since $u \in H_0^1(\Omega) \cap H^2(\mathcal{K}_h)$, we have

$$\begin{aligned} A(u, v_h) := \sum_{K \in \mathcal{K}_h} \int_K \kappa \, \nabla u \cdot \nabla v_h \, dx + \alpha \sum_{K \in \mathcal{K}_h} \int_K u \, v_h \, dx \\ - \sum_{S \in \mathcal{S}_h} \int_S \{\frac{\partial u}{\partial n_S}\}_{S,\kappa} [v_h]_S \, ds. \end{aligned} \tag{1.18}$$

Using the fact that $\kappa \frac{\partial u}{\partial n_S}$ is "continuous" on the interior edges $\mathcal{S}_h^{int}$ yields

$$\{\frac{\partial u}{\partial n_S}\}_{S,\kappa} = (\frac{\kappa^+}{\kappa^+ + \kappa^-} + \frac{\kappa^-}{\kappa^+ + \kappa^-})(\kappa \, \frac{\partial u}{\partial n_S}) = \kappa \, \frac{\partial u}{\partial n_S}.$$

Integrating by parts leads to

$$\sum_{K \in \mathcal{K}_h} \int_K \kappa \, \nabla u \cdot \nabla v_h \, dx - \sum_{S \in \mathcal{S}_h} \int_S \{\frac{\partial u}{\partial n_S}\}_{S,\kappa} [v_h]_S \, ds = - \sum_{K \in \mathcal{K}_h} \int_K \operatorname{div}(\kappa \, \nabla u) v_h \, dx.$$

We get

$$A(u, v_h) = \sum_{K \in \mathcal{K}_h} \int_K (- \operatorname{div}(\kappa \, \nabla u) + \alpha \, u) v_h \, dx = \int_\Omega f \, v_h \, dx = A(u_h, v_h),$$

yielding the assertion. □

For the continuity property of the bilinear form A, we can prove the following result

$$A(u, v) \leq C(\|\kappa\|_\infty, \|\alpha\|_\infty) \|u\|_{h,A} \|v\|_{h,A} \quad u, \, v \in V_h^p. \tag{1.19}$$

1.3.3 *A priori* error analysis

In this section, we establish an *a priori* error estimate for DgFem in energy norm. The analysis is performed by using the continuity, the coercivity and the consistence properties established previously.

Theorem 1.1 *Let u be the solution of the problem (1.8) and u_h the discrete solution of (1.12). Assume that $u \in H^{p+1}(\mathcal{K}_h)$, then, there is a constant C independent of h such that*

$$\|u - u_h\|_{h,A} \leq C \, h^p \|u\|_{H^{p+1}(\mathcal{K}_h)}. \tag{1.20}$$

Proof. Let $v_h \in V_h^p$, owing to lemma 1.2, 1.3

$$\begin{aligned} \|u_h - v_h\|_{h,A}^2 &\leq C\, A(u_h - v_h, u_h - v_h) = C\, A(u - v_h, u_h - v_h) \\ &\leq C\, \|u - v_h\|_{h,A} \|u_h - v_h\|_{h,A}, \end{aligned} \tag{1.21}$$

yields

$$\|u_h - v_h\|_{h,A} \leq C\, \|u - v_h\|_{h,A}.$$

Taking $v_h = \Pi_h u \in V_h^p$ the L^2-projection of u

$$\|u_h - \Pi_h u\|_{h,A} \leq C\, \|u - \Pi_h u\|_{h,A}.$$

Now, using the triangle inequality, we get

$$\|u - u_h\|_{h,A} \leq C\, \|u - \Pi_h u\|_{h,A}. \tag{1.22}$$

The assertion follows by using the standard approximation properties of the L^2-orthogonal projector Π_h. □

The error estimate in L^2 norm can be improved by using the duality argument, as shown in the next theorem.

Theorem 1.2 *Under the assumptions of Theorem 1.1, there is a constant C independent of h such that*

$$\|u - u_h\|_{0,\Omega} \leq C\, h^{p+1} \|u\|_{H^{p+1}(\mathcal{K}_h)}. \tag{1.23}$$

1.3.4 Dependence on γ

Thanks to (1.16), one gets that

$$\int_\Omega f u_h^\gamma \, dx = A(u_h^\gamma, u_h^\gamma) \geq c \|u_h^\gamma\|_{1,h}^2 + (\gamma - \gamma_0) |h_S^{-1/2} [u_h^\gamma]_S|_{\gamma_S}^2. \tag{1.24}$$

Then, inequality (1.24) yields that $(u_h^\gamma)_\gamma$ is a bounded sequence for the norm $\|.\|_{1,h}$ and that $|h_S^{-1/2}[u_h^\gamma]_S|_{\gamma_S}$ tends to 0 when γ tends to infinity.
For any accumulation point w_h of sequence $(u_h^\gamma)_\gamma$, since at the limit $|h_S^{-1/2}[w_h]_S|_{\gamma_S} = 0$, it is an element of the conforming finite element space.
Moreover, passing to the limits with respect to γ ($\gamma \to +\infty$) in the discrete formulation with some test-functions in the continuous FE-space, proves that $w_h = u_h$. Then, one gets the convergence of all the sequence u_h^γ to u_h.
This proves that the DgFem solution converges to the conforming finite element solution when γ tend to infinity.

To clarify the order of convergence, let us recall, from Larson [34], the following result:

Proposition 1.1 *Let u_h^γ the DgFem solution of the problem (1.8), u_h the Fem solution of (1.8) and u the solution to the continuous problem. If u is assumed $H^{p+1}(\Omega)$, then, for all $\gamma \geq \gamma_0$ one has*

$$\|u_h^\gamma - u_h\|_{h,A} \leq \frac{C}{\gamma - \gamma_0} h^p \|u\|_{p+1}, \tag{1.25}$$

where C is a positive constants independent of h and γ.

1.4 Numerical results

In this section, we present the results of some numerical experiments. In the first and the second test, we study the behaviour of the error with respect to h, the mesh parameter, and p the approximation order. In the third test, we are interested in the behaviour of u_h^γ with respect to γ (when it tends to ∞). In order to illustrate this, we consider the following elliptic problem with $\kappa = 1$ and $\alpha = 0$;

$$-\Delta u = f \ \ in \ \ \Omega =]0,1[^2, \tag{1.26}$$

where f is chosen such that the exact solution is given by

$$u(x,y) = \sin(\pi(x-y)).$$

In this case, the DgFem is the classical SIPG method. We start by studying the convergence with respect to h, for $p = 1, 2, 3, 4$.

1.4.1 h-DgFem

In this section, we illustrate the behaviour of the error with respect to the discretization parameter h, for a fixed approximation order p in $\{1, 2, 3, 4\}$. In the figure below, we represent the energy and L^2 norms of the error with respect to h;

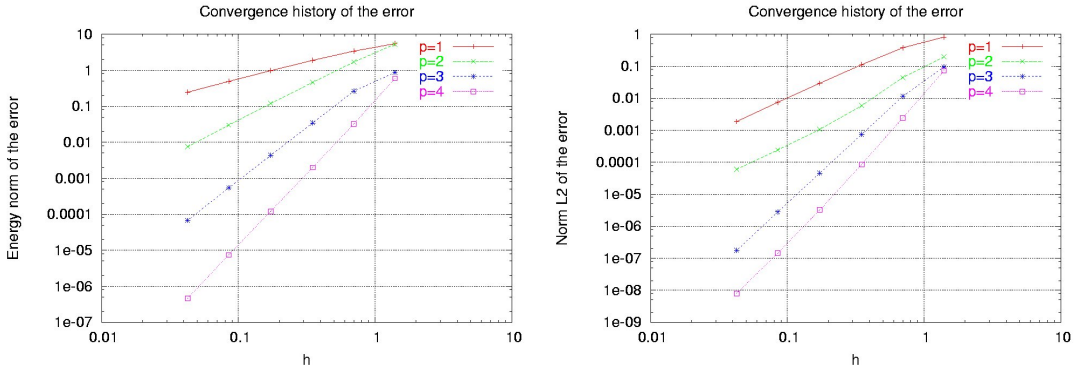


Figure 1.2: $\|u - u_h\|_{1,h}$ and $\|u - u_h\|_{0,\Omega}$ as a function of h in log-log scale.

These figures show the convergence of the error with respect to h and that the convergence is of order $p+1$ in L^2 norm and of order p in energy norm which confirm the theoretical results.

1.4.2 Exponential convergence of p-DgFem

In this example, problem (1.26) is considered, with a fixed discretization parameter h. The numerical illustration concerns the behaviour of the convergence with respect to the approximation order p. The result is presented in Figure 1.3

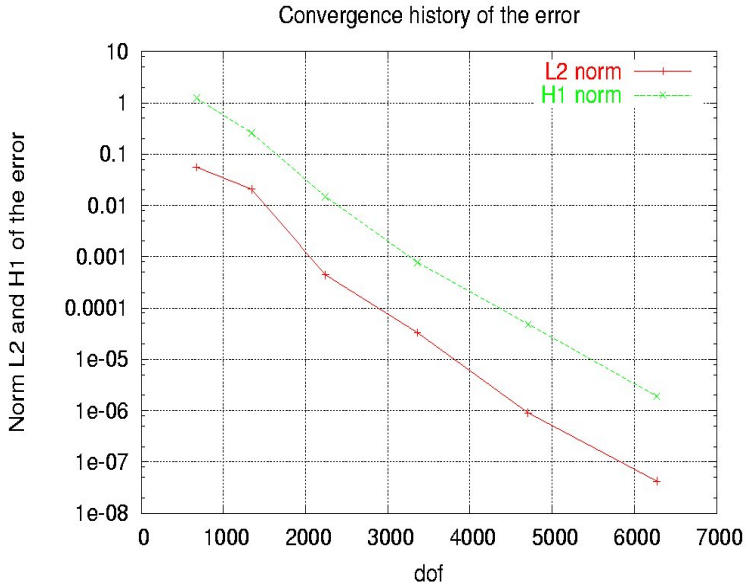


Figure 1.3: $\|u - u_h\|_{1,h}$ and $\|u - u_h\|_{0,\Omega}$ as a function of the approximation order p.

The figure illustrates that, the refinement p-uniform delivers exponential convergence with respect to the number of degree of freedom because of the regularity of the exact solution. This is demonstrated by using a linear scale for the number of degree of freedom and the logarithmic scale for the error.

1.4.3 Convergence study (variable approximation order)

In this section, we study the convergence of the error in L^2 and energy norms with respect to the discretization parameter h with uniform and variable approximation order. The result is represented in the figure below :

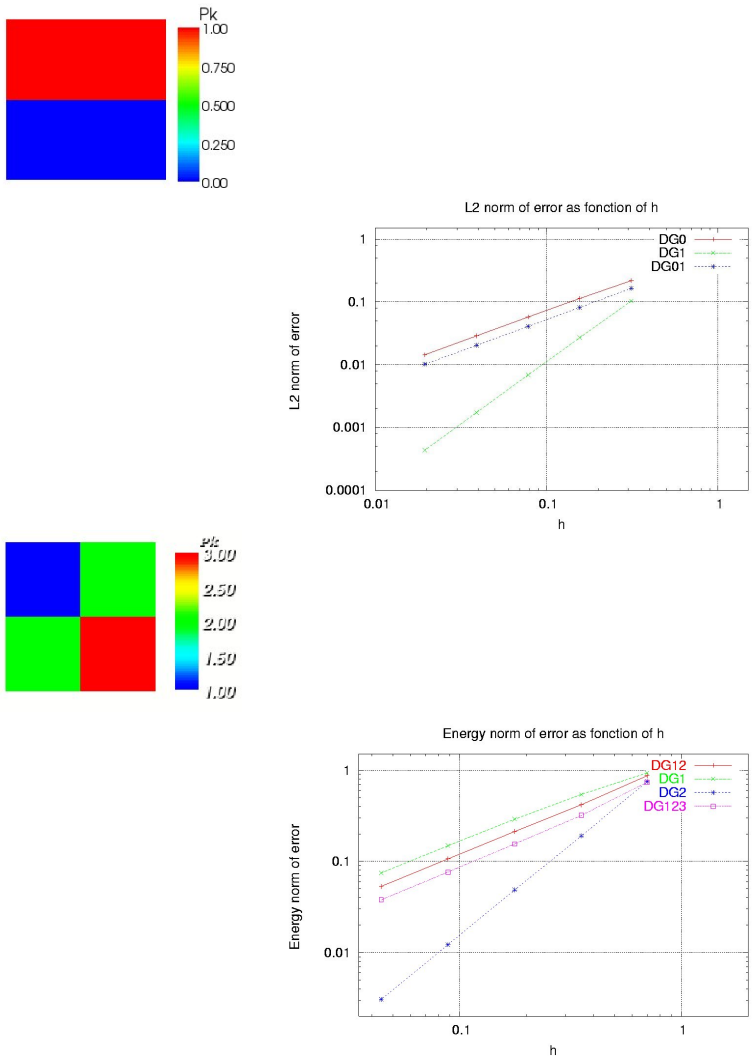


Figure 1.4: $\|u - u_h\|_{0,\Omega}$ (top) and $\|u - u_h\|_{1,h}$ (bottom) with respect to h in log-log scale.

This numerical tests have been realized in the objective to validate a code written in C^{++}

with approximation order depends on the mesh element. A convergence rate equal to $\min(p_K) + 1$ in L^2 norm is observed, because the exact solution is regular. Increasing uniformly p gives a good approximation in this case.

1.4.4 Dependence on γ

In this section, a fixed mesh is considered. Then, we are interested in the behaviour of the difference between u_h^γ: the discrete solution given by the DgFem; and u_h the discrete solution given by the standard Galerkin method.
The figure below illustrates the convergence with respect to γ. Three different meshes are considered (with a log-log scale) and confirm the theoretical result proved in section 1.3.4.

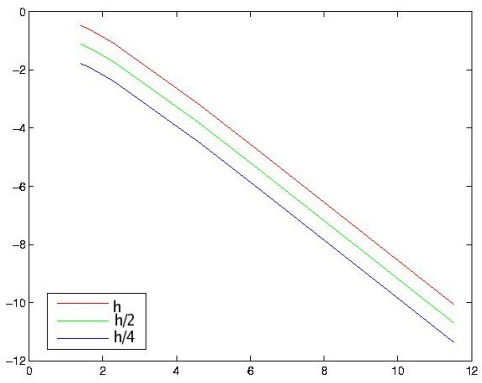


Figure 1.5: $\|u_h^\gamma - u_h\|_{H^1(\mathcal{K}_h)}$ as a function of γ in log-log scale.

The figure shows that, for a large γ, the DgFem gives the same error as the CFE. Indeed, if we choose γ in an optimal way, the DgFem error is less than the CFE error. In figure 1.7, we present the L^2 norm of the error with respect to γ for $p = 1$ and optimal gamma with respect to p;

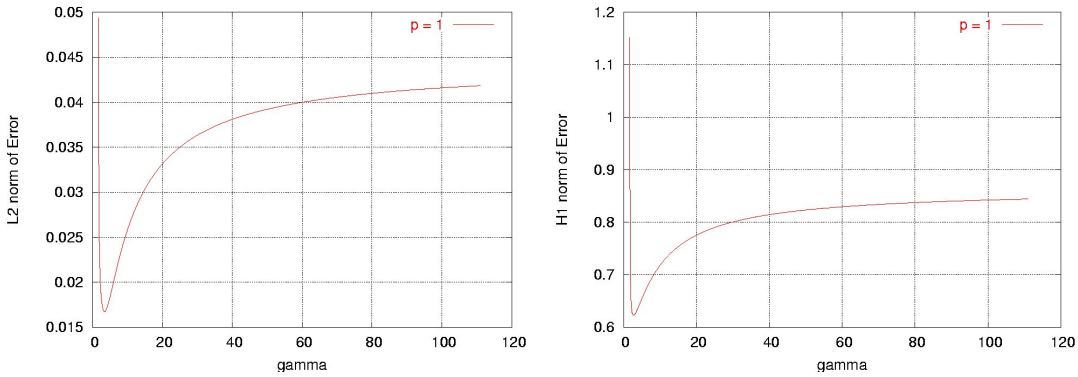


Figure 1.6: $\|u - u_h\|_{0,\Omega}$ as a function of γ (left) and $\|u - u_h\|_{1,h}$ as a function of γ (right).

This figure illustrates the existence of an optimal value γ_0 of the parameter γ. It depends *a priori* on the approximation order p. Since no theoretical results prove this, we propose in the Figure 1.7 some numerical indications of the optimal value γ_0 as a function of p.

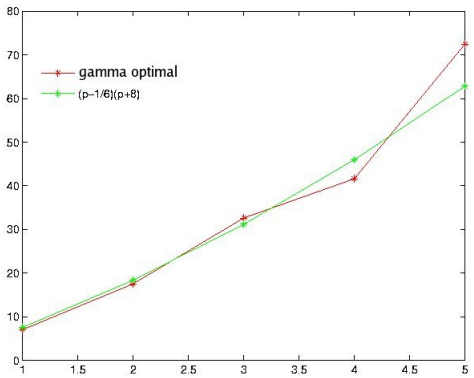


Figure 1.7: gamma optimal as a function of p.

1.5 h-adaptive DgFem discretization

1.5.1 Red-green mesh refinement

One standard method of local mesh refinement in 2-D is the red-green refinement (see R. Verfürth [41]), especially for Delaunay meshes. A triangle marked for refinement is split into four smaller sons, so called red triangles, as shown in Figure 1.8. At the intersections

of the triangle edges, new points are inserted. Thereby, four triangles are generated, whose edges are connected to these points. A pleasant property of this method is that all new triangles are geometrically similar to the original one, and thereby, the element quality of the new triangles is similar to the original one.

Figure 1.8: A red-refined triangle

Using this method, any neighboring triangles can be red-refined (see Figure 1.9) and any of the red triangles can be red-refined itself. But red-refined triangles will produce extra points along an edge of neighboring triangles that are not refined to the same level. These triangles must be refined irregularly, which are labeled as green triangles and are of lower quality (shown in Figure 1.10). Green-refinement is only performed if a single point is inserted. If more points are inserted or higher refinement of this triangle is wanted, the green triangles are removed and the parent triangle is red-refined, too.

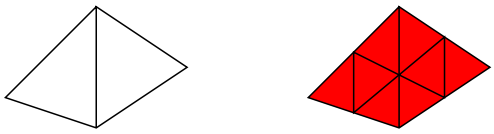


Figure 1.9: A red-refined triangle with red-refined neighbor

By continuation of these two strategies, a local mesh refinement is produced. Red triangles can be surrounded by red triangles as well as green triangles, and red triangles can also be further red-refined, which may also induce the generation of surrounding green triangles. With a hierarchical memory representation, refinement and also coarsening of already refined triangles can be handled easily.

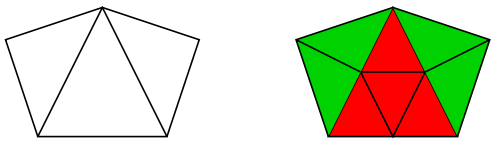


Figure 1.10: A red-refined triangle with green-refined neighbor

If we regard Ω as a pseudo-element whose "sons" are the elements of $\mathcal{T}_0$, the process of creating an admissible mesh can be viewed as the creation of an element tree, in which the root is Ω. The level l_i of T_i then corresponds to the distance from T_i to Ω.

1-irregular rule

In general, it is advantageous to regularize an admissible mesh by restricting the number of irregular vertex on each edge. There are number of ways to accomplish this regularization, we shall mainly consider the following *1-irregular rule*.

1-irregular rule: Refine any unrefined element for which any of the sides contains more then one irregular vertex.

We shall also use the following two-neighbor rule:

two-neighbor rule: Refine any element with two neighbors that have been regularly refined.

The properties of the 1-irregular rule when applied to triangular meshes differ somewhat from the case of quadrilateral element. Unlike quadrilaterals, refinement of triangular elements does not always generate new regular vertex. Let the $l+1$ level mesh $\mathcal{T}_l$ be generated by successive refinement of the center element. $\mathcal{T}_3$ is illustrated in Figure 1.11. All the $\mathcal{T}_l$ are 1-irregular. However, the only regular nodes for $l > 2$ are the six boundary nodes.

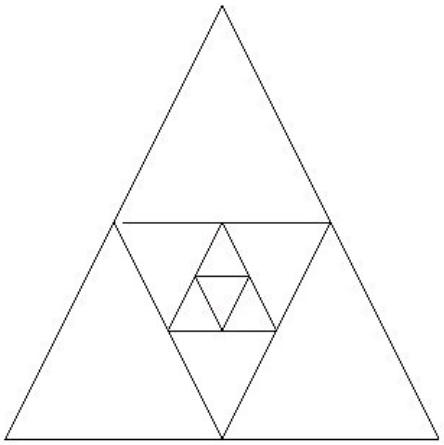


Figure 1.11: $\mathcal{T}_3$

Given a 1-irregular mesh $\mathcal{T}^{'}$, we can generate a more refined mesh $\mathcal{T}^{"}$ by applying wherever possible, the following green rule:

Green refinement rule

With as few elements as possible, triangulate any unrefined element with an irregular vertex on one or more of it sides. The three situations in which the green rule can occur are shown in Figure 1.12.

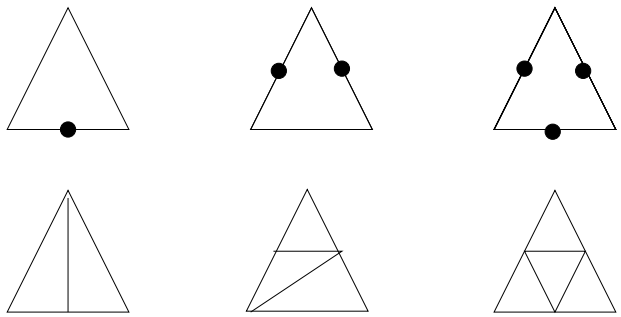


Figure 1.12: The green refinement rule

Algorithm of local mesh refinement

In this section, we present an algorithm for the generation of local mesh refinement described previously. It implements the 1-irregular and two-neighbor rules.
In this algorithm, we assume that a logical-valued function **IsRefined** is available which indicates, for a given element in the mesh, whether the element to be refined. An element in the mesh may be refined either because **IsRefined** indicates that it should be refined, or if it satisfies the two-neighbor rule or the 1-irregular rule. The algorithm of refinement is as follows:

- **Step 1** : Delete the green triangle.
- **Step 2** : Refine red triangle.
- **Step 3** : Perform the 1-irregular rule.
- **Step 4** : Perform the two-neighbor rule.
- **Step 5** : Refine the green triangle using green rule.

1.5.2 *A posteriori* error estimate

An *a posteriori* error estimate is a rigorous bound to a prescribed error quantity in terms of available information. Here, we consider control of the energy error $\|u - u_h\|_{h,A}$.

$$\|u - u_h\|_{h,A}^2 \leq \sum_K \varepsilon_K^2 \tag{1.27}$$

where ε_K is a computable quantity depending on u_h and the data. The following theorem gives an upper bound of the error in energy norm. The prove of the theorem is given in [33] for an elliptic problem in the case of $\kappa = 1$ and $\alpha = 0$. An adaptive algorithm has been considered and a result of the convergence of the algorithm is given. Recently, an *a posteriori* error analysis for the advection-reaction-diffusion equations with anisotropic

and discontinuous diffusivity is considered in [20]. Let us recall from [33] the following theorem.

Theorem 1.3 *Let u be the solution of (1.9) and u_h be the solution of (1.12) with $\kappa = 1$ and $\alpha = 0$. There holds*

$$\sum_{K\in\mathcal{K}_h} \|u-u_h\|^2_{0,K} \leq C\left(\sum_{K\in\mathcal{K}_h}\eta_K(f,u_h)+\sum_{S\in\mathcal{S}_h^{int}}\eta_S(u_h)\right), \tag{1.28}$$

where C is a constant independent of h and γ, η_K and η_S are given by

$$\eta_K = h_K^2\|f+\Delta u_h\|^2_{0,K},\ \ \eta_S = h_S\,|[\partial_{n_S}u_h]_S|^2_S + \frac{\gamma^2}{h_S}|[u_h]_S|^2_S$$

In addition the following estimates hold

- *(i) Suppose that f is a piecewise polynomial on $\mathcal{K}_h$. Then for each $K\in\mathcal{K}_h$,*

$$\eta_K \leq c\,\|u-u_h\|^2_{0,K}.$$

- *(ii) For $S\in\partial K^+\cap\partial K^-$,*

$$h_S\,|[\partial_{n_S}u_h]_S|^2_S \leq c\,(\|u-u_h\|^2_{0,K^+}+\|u-u_h\|^2_{0,K^-}).$$

- *(iii) There exists γ_1 depending only on p and the minimal angle such that for all $\gamma\geq\gamma_1$*

$$\sum_{S\in\mathcal{S}_h^{int}}\frac{\gamma^2}{h_S}|[u_h]_S|^2_S \leq c\sum_{K\in\mathcal{K}_h}\|u-u_h\|^2_{0,K},$$

c is a constant independent of h and γ.

1.5.3 *h*-adaptive algorithm

In this section, we describe the outline of the h-adaptive algorithm. We repeat the algorithm that automatically constructs an adapted solution in the approximation space by h-refinement.

- Step 1 (initialization): Assume an initial coarse mesh $\mathcal{K}_h$ consisting of piecewise-polynomial ($p=2$). User input $\theta\in[0,1]$ and $Tol>0$ for the energy norm of the approximation error function η.

- Step 2: Compute coarse mesh approximation $u_h\in V_h^p$ on $\mathcal{K}_h$.

- Step 3 (Error estimator): Compute the indicator ε_K on every element K_i, $i = 1, ..., N_K$ in the mesh, calculate the global error estimate η, if $\eta < Tol$, stop computation.

- Step 4 (Marking element): Let the number ε_K be given as in steps 3. We find $\mathcal{A}^{(j)} \subset K^{(j)}$ be the solution of the following optimization problem:

$$\text{Minimize } Cardinal\mathcal{A}^{(j)}, \tag{1.29}$$

under the constraint

$$\sum_{K \in \mathcal{A}^{(j)}} \varepsilon_K^2 \geq \theta^2 \sum_{K \in K^{(j)}} \varepsilon_K^2. \tag{1.30}$$

- Step 5 (h-refinement): Refine $K \in \mathcal{A}^{(j)}$, then we establish a new finite element space, and we continue with step 3.

1.5.4 Numerical examples

In this section, we present the results of some numerical experiments.

L-shaped domain

In order to illustrate the h-adaptive algorithm, we consider the following elliptic problem with non homogeneous Dirichlet boundary condition;

$$\begin{aligned} -\Delta u &= 0 \text{ in } \Omega \\ u &= g \text{ on } \partial\Omega, \end{aligned} \tag{1.31}$$

where r and θ are the polar coordinates and g is given by

$$g(x, y) = r(x, y)^{2/3} \sin(2/3(\theta(x, y) + 2\pi)), \ (x, y) \in \partial\Omega. \tag{1.32}$$

The geometry of the computational domain is displayed in figure 1.13

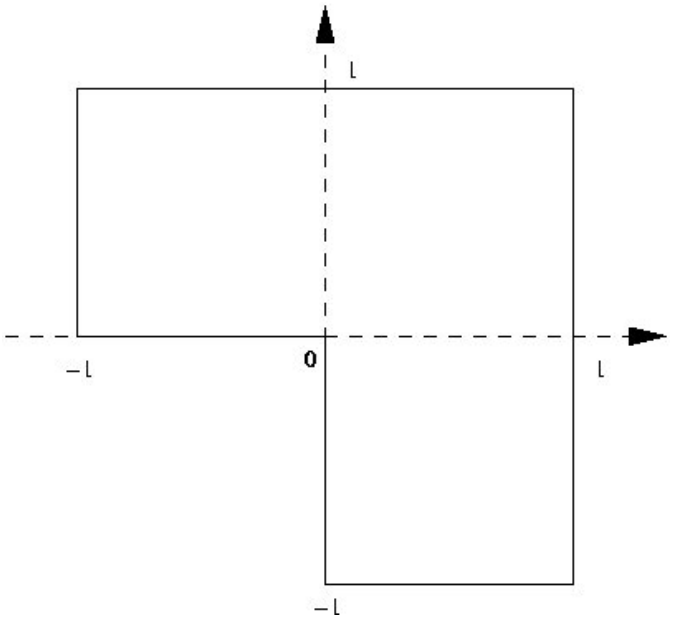


Figure 1.13: The L-shaped domain Ω

The exact solution is given by:

$$u(x,y) = r(x,y)^{2/3}\sin(2/3(\theta(x,y)+2\pi)), \text{ for all } x,y \in \Omega. \tag{1.33}$$

This problem has a re-entrant corner located at the origin, thereby producing a singularity on the solution (non-convexity of the domain). We present in the figures below, some iterations of adaptive mesh refinement and the numerical solution obtained in the final mesh;

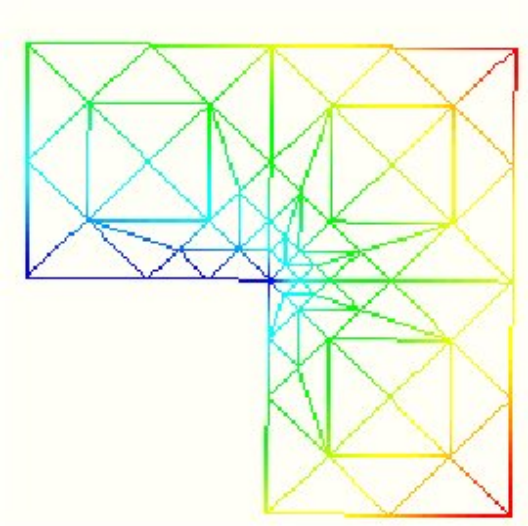


Figure 1.14: Mesh refinement iteration 2

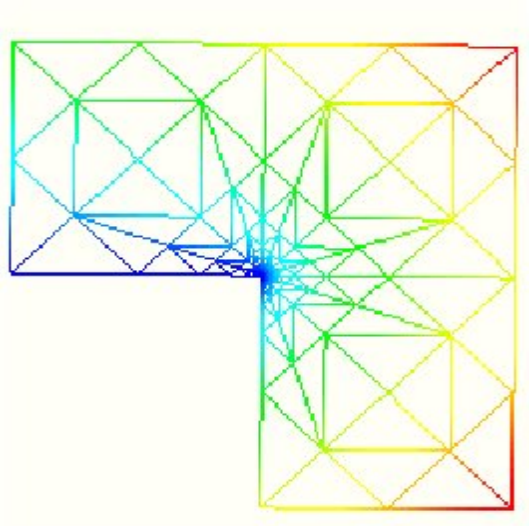


Figure 1.15: Mesh refinement iteration 4

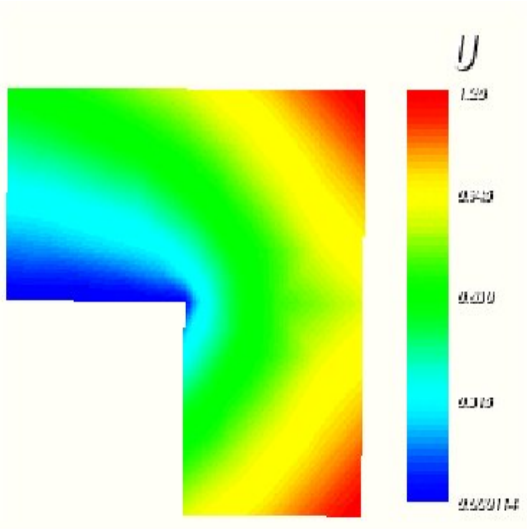


Figure 1.16: Numerical solution

The figure 1.17 displays the convergence history on $H^1(\mathcal{K}_h)$ norm of the error with respect to the number of degree of freedom in log-log scale;

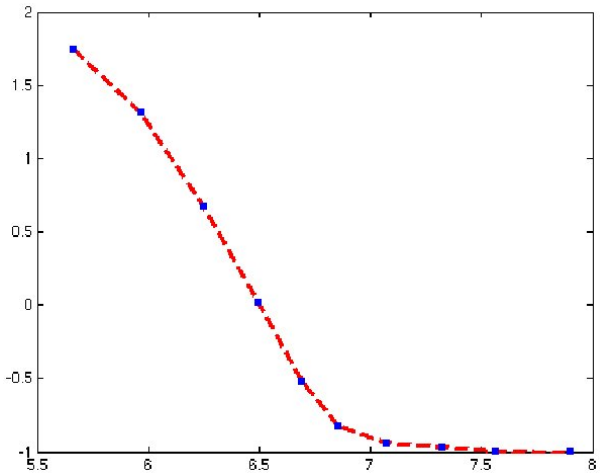


Figure 1.17: Convergence history on $H^1(\mathcal{K}_h)$ norm of the error h-adapt($p = 2$)

Internal layer

In this example, we consider the following elliptic problem with non-homogeneous Dirichlet boundary condition:

$$-\Delta u = f \text{ in }]0,1[^2, \tag{1.34}$$

where the right hand side f is chosen such that the exact solution is given by

$$u(x,y) = \arctan(\alpha(r(x,y) - \pi/3)), \quad \alpha = 60, \tag{1.35}$$

with $r(x,y) = \sqrt{x^2 + y^2}$.
The exact solution u is regular with high gradient. We present in the figure below some iterations of the algorithm and the corresponding numerical solution;

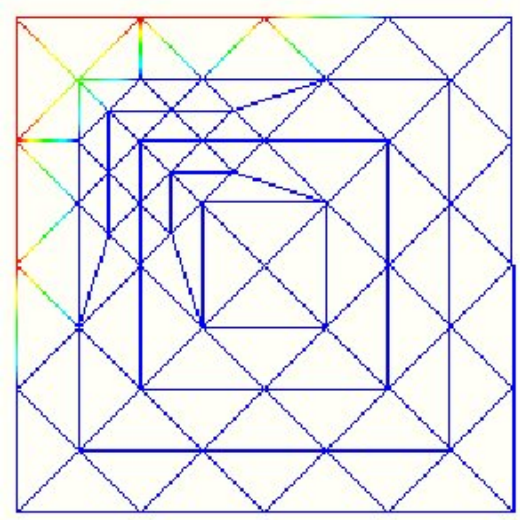


Figure 1.18: Mesh refinement, first iteration

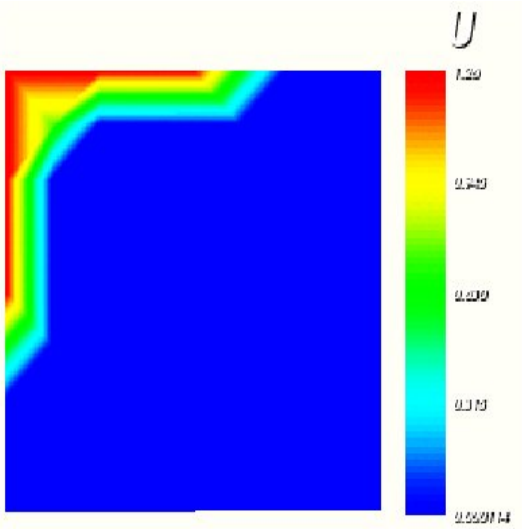


Figure 1.19: Numerical solution

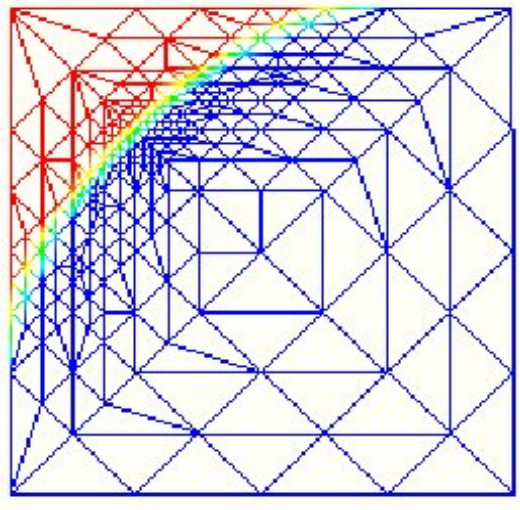


Figure 1.20: Mesh refinement, iteration 4

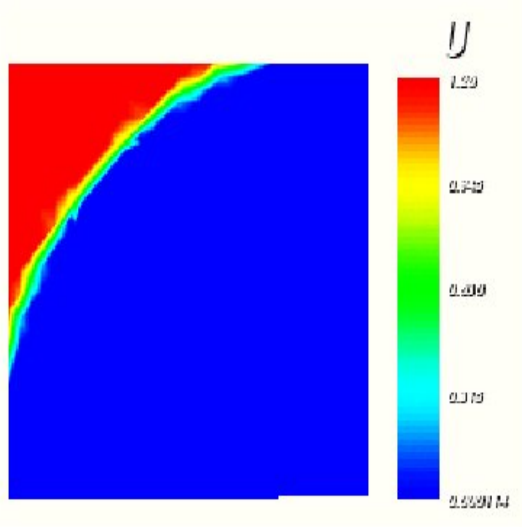


Figure 1.21: Numerical solution

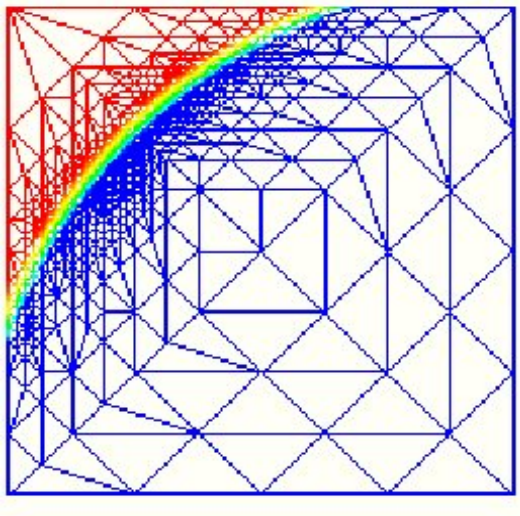


Figure 1.22: Mesh refinement iteration 8

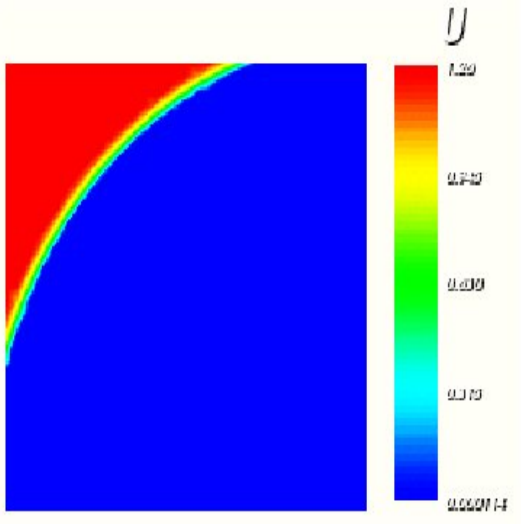


Figure 1.23: Numerical solution

The figure 1.24 displays the convergence history of the $H^1(\mathcal{K}_h)$ norm of error with respect to the number of degree of freedom in logarithmic scale;

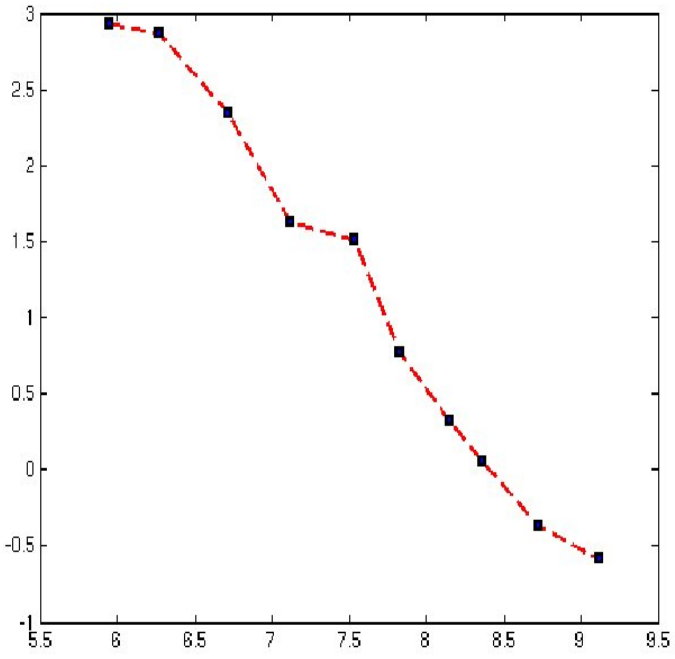


Figure 1.24: Convergence history on $H^1(\mathcal{K}_h)$ norm of error, h-adapt$(p = 2)$ in log-log scale

Chapter 2

Presentation of the model and study of the existence

2.1 Introduction

In this chapter, we present and study a mathematical problem arising from the modelling of maximal erosion rate in geological stratigraphic phenomena. The equation of such problems is nonlinear; the diffusion coefficient depends of the time-derivative of the unknown u and degenerates in order to take implicitly into account a global constraint on the time-derivative of u :

$$\partial_t u + E \geq 0, \tag{2.1}$$

where E is a prescribed maximum erosion rate of the sediments.
The outlines of this chapter is organized as follows: in the following section, we present the mathematical model. In section 2.3, a mathematical formulation that contains implicitly the constraint (2.1) is presented. In section 2.4, we give a result of existence of a solution to the model problem.

2.2 Mathematical model

We consider a sedimentary basin Ξ with basis $\Omega \subset \mathbb{R}^N$ $(N = 1, 2)$ (see Figure 2.1), assumed to be a bounded domain with a Lipschitz boundary. It is determined by a known vertical position $H(x, t)$ for each instant t and each position x. The position of this base is due to vertical displacements of tectonics and the variation of the sea level. It provides a description of the transport laws of sediments and their coupling, as well as the flux of sediments at the edges of the basin.
For a positive number T, we denote by $Q =]0, T[\times\Omega$ and u the sediments thickness; the

topography is then given by $u + H$. The mathematical model is obtained by writing the mass conservation equation, the Darcy-Barenblatt's law and by taking into account a constraint of obstacle (limited erosion rate).

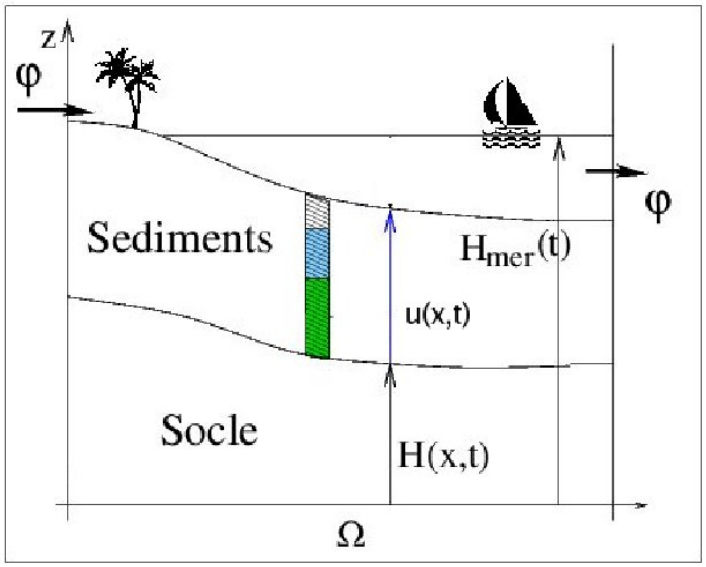


Figure 2.1: a sedimentary basin in 2D

2.2.1 Equation of conservation

We consider that the sediments thickness u satisfies the mass conservation equation:

$$\partial_t u + \mathrm{div}(\vec{q}) = f \ \text{ in } \ Q, \tag{2.2}$$

where $\vec{q}$ is the flux of the sediments.
According to Darcy-Barenblatt's law (see [11]), the flux of sediments $\vec{q}$ is given by the relation:

$$\vec{q} = -K\nabla(u + \tau\partial_t u), \tag{2.3}$$

with K is the viscosity rate, $\tau > 0$.
In a sedimentary basin formation process, sediments must first be produced *in situ* by weathering effects prior to be transported by surfacing erosion. Thus, R. Eymard *et al.* [26] introduce a maximum erosion rate

$$\partial_t u + E \geq 0, \quad a.e.\ in\ Q, \tag{2.4}$$

where $E(x,t) \geq 0$ denotes the admissible maximum erosion rate.

In order to reconcile the constraint with a conservative formulation, a limiter λ is introduced to correct the theoretical diffusive flow into a real diffusive flow:

$$\vec{q} = -\lambda K \nabla(u + \tau \partial_t u), \tag{2.5}$$

where λ is a unknown function with values *a priori* in $[0, 1]$ which will be defined later. Moreover, $K = 1$ would be consider in the sequel.

2.2.2 Boundary conditions

In the real physical framework, one has to consider a partition of the boundary: $\Gamma = \Gamma_e \cup \Gamma_s$. An input flux φ_e is imposed on Γ_e and an unilateral conditions $\partial_t u + E \geq 0$, $\vec{q}.\vec{n} + \varphi_s \geq 0$ and $(\partial_t u + E)(\vec{q}.\vec{n} + \varphi_s) = 0$ on Γ_s (see S. N. Antontsev *et al.* [2]).
In our academic work, Dirichlet boundary conditions would be considered.

2.2.3 Mathematical modelling of λ

In [26], the authors propose to consider a measurable flux limiter λ that satisfies

$$0 \leq \lambda \leq 1, \quad \partial_t u + E \geq 0, \quad (1 - \lambda)(\partial_t u + E) = 0, \text{ a.e. in } \Omega \times (0, T). \tag{2.6}$$

S.N. Antontsev, G. Gagneux and G. Vallet [6] propose a new conservative formulation which contains implicitly (2.6): for all $u \in L^2(0, T; H_0^1(\Omega))$, with $\partial_t u \in L^2(0, T; H_0^1(\Omega))$, (2.2,2.5,2.6) is equivalent to the following formulation:

$$\partial_t u - \operatorname{div}[\lambda \nabla(u + \tau \partial_t u)] = f \text{ in } \Omega \times (0, T), \quad \lambda \in H(\partial_t u + E) \cap L^\infty(Q). \tag{2.7}$$

Here, for the sake of simplicity, homogeneous Dirichlet boundary conditions on u and $\partial_t u$ are considered. $u(0, .) = u_0 \in H_0^1(\Omega)$, $E \in L^\infty(0, T; H^1(\Omega))$, $f \in L^\infty(0, T; L^2(\Omega))$ and H denotes the maximal monotone graph of the Heaviside function.

The advantage of this writing is that, as soon as $f + E \geq 0$ is assumed, the constraint (2) is implicitly satisfied.
Indeed, on the one hand, obviously (2.6) implies that $\lambda \in H(\partial_t u + E)$.
On the other hand, if $\lambda \in H(\partial_t u + E)$, one just has to prove that $\partial_t u + E \geq 0$. Since the solution is understood in a weak sense: $\forall\, v \in H_0^1(\Omega)$,

$$\int_\Omega [\partial_t u v + \lambda \nabla(u - \tau E) \nabla v + \tau \nabla(\partial_t u + E) \nabla v] dx = \int_\Omega f v dx.$$

Then, for the available test function $(\partial_t u + E)^-$, we get that

$$-\int_\Omega (E + f)(\partial_t u + E)^- dx = \int_\Omega \left|(\partial_t u + E)^-\right|^2 dx + \int_\Omega \lambda \nabla(u + \tau \partial_t u) \nabla(\partial_t u + E)^- dx. \tag{2.8}$$

That leads to $\partial_t u + E \geq 0$ a.e. in $\Omega \times (0,T)$ since $\lambda \mathbf{1}_{\{\partial_t u + E < 0\}} = 0$ a.e.

Now, we are interested in deriving a new model, equivalent to the previous one. Since $\lambda \in H(\partial_t u + E)$, $\partial_t u + E \geq 0$ and $\partial_t u + E \in H_0^1(\Omega)$, thanks to the lemma of Saks, we have

$$\begin{aligned} \lambda \nabla(u + \tau \partial_t u) &= \lambda \nabla(u + \tau(\partial_t u + E) - \tau E) \\ &= \lambda \nabla u + \tau \lambda \nabla(\partial_t u + E) - \tau \lambda \nabla E \\ &= \lambda \nabla u + \tau \nabla(\partial_t u + E)^+ - \tau \lambda \nabla E, \ \lambda \in H \\ &= \lambda \nabla(u - \tau E) + \tau \nabla(\partial_t u + E). \end{aligned} \tag{2.9}$$

Thus, the problem (2.7) is equivalent to the following one:

$$\partial_t u - \operatorname{div}[\lambda \nabla(u - \tau E)] - \tau \Delta(\partial_t u + E) = f, \ \ \lambda \in H(\partial_t u + E). \tag{2.10}$$

Results of existence and uniqueness of a solution to the above differential inclusion are still open problems. Our purpose is to analyze a modified model where H is replaced by a, a regular approximation of H, vanishing on $\mathbb{R}^-$ in order the constraint to be respected.

2.3 Mathematical formulation

In this section, we are interested in the following pseudoparabolic problem **(P)**:

$$\partial_t u - \operatorname{div}[a(\partial_t u + E) \nabla(u - \tau E)] - \tau \Delta(\partial_t u + E) = f \ \ in \ \ \Omega \times (0,T). \tag{2.11}$$

The initial height is given by : $u(0,.) = u_0$ in Ω, where $u_0 \in H_0^1(\Omega)$, homogeneous Dirichlet condition are assumed for u and $\partial_t u$ and the following assumptions are made concerning the data.

$$\mathbf{(H)} : \begin{cases} \tau > 0, \ E \in L^\infty(0,T; H^1(\Omega)) \text{ is a nonnegative function,} \\ f \in L^\infty(0,T; L^2(\Omega)), \ f + E \geq 0 \text{ in } Q, \text{ and} \\ a : \ \mathbb{R} \to [0,1] \text{ is a Hölder continuous function,} \\ a \in C^{0,\theta}(\mathbb{R}), \text{ with } \theta \geq 1/2, \text{ and vanishing on }]-\infty, 0]. \end{cases} \tag{2.12}$$

Then, we would say that

Definition 2.1 *A solution to problem (**P**) is any u in $H^1(0,T; H_0^1(\Omega))$, such that for any v in $H_0^1(\Omega)$, and for t a.e. in $(0,T)$,*

$$\begin{cases} u(0,.) = u_0 \ \ in \ \ \Omega \ \ and \ \ \forall \, v \in H_0^1(\Omega), \\ \displaystyle\int_\Omega [\partial_t u v + a(\partial_t u + E) \nabla(u - \tau E) \nabla v + \tau \nabla(\partial_t u + E) \nabla v] dx = \int_\Omega f v dx. \end{cases} \tag{2.13}$$

Lemma 2.1 *If u is a solution to (2.13). Then the constraint $\partial_t u + E \geq 0$ a.e. in Q is implicitly satisfied.*

Proof. One just to adapt the demonstration given by (2.8) since $a(\partial_t u + E) \mathbf{1}_{\{\partial_t u + E < 0\}} = 0$. □

2.4 Existence of a solution

In this section, we study the existence of a solution to problem (2.13). To prove the existence of a solution of this problem, we proceed as follow: we initially propose an existence and uniqueness lemma of the solution to an additional stationary problem. Then we prove the existence of a sequence of solutions to an implicit time-discretized problem. And, via some *a priori* estimations, we prove that the problem (2.13) has a solution by passing to the limits with respect to the time step.

Lemma 2.2 *Let us consider a bounded function $b \in C^{0,\theta}(\mathbb{R})$ with $\theta \geq \frac{1}{2}$, κ in $H^1(\Omega)$, f in $L^2(\Omega)$, E in $H^1(\Omega)$ a nonnegative function. Then, a unique solution w exists in $H_0^1(\Omega)$ such that, for all v in $H_0^1(\Omega)$*

$$\int_\Omega [wv + b(w+E)\nabla(\kappa - \tau E)\nabla v + \tau\nabla(w+E)\nabla v]dx = \int_\Omega fv dx. \tag{2.14}$$

Proof. The result of existence of a solution is based on the theorem of fixed point of Schauder.
Concerning the uniqueness, we consider two admissible solutions w_1 and w_2. Let $\xi = w_1 - w_2$ and for $\mu > 0$, we define p_μ by

$$p_\mu(t) = \min\left[1, \ln[\max(1, \frac{e\ t}{\mu})]\right]. \tag{2.15}$$

p_μ is a non-increasing Lipschitz-continuous function with $p_\mu(0) = 0$. The Lipschitz-constant is equal to $\frac{e}{\mu}$. Therefore, $v = p_\mu(\xi)$ is a suitable test function and we have that

$$0 = \int_\Omega \xi p_\mu(\xi)dx + \int_\Omega [b(w_1+E) - b(w_2+E)]\nabla(\kappa - \tau E)\nabla p_\mu(\xi)dx + \tau\int_\Omega \nabla\xi\nabla p_\mu(\xi)dx.$$

Thus,

$$0 = \int_\Omega \xi p_\mu(\xi)dx + \int_{\Omega\cap\{\frac{\mu}{e}<\xi<\mu\}} \frac{1}{\xi}[b(w_1+E) - b(w_2+E)]\nabla(\kappa - \tau E)\nabla\xi dx + \tau\int_{\Omega\cap\{\frac{\mu}{e}<\xi<\mu\}} \frac{1}{\xi}|\nabla\xi|^2 dx.$$

So, it comes that :

$$\begin{aligned}\int_\Omega \xi p_\mu(\xi)dx \ &+ \ \tau\int_{\Omega\cap\{\frac{\mu}{e}<\xi<\mu\}} \frac{1}{\xi}|\nabla\xi|^2 dx \leq c\int_{\Omega\cap\{\frac{\mu}{e}<\xi<\mu\}} \xi^{\theta-1}|\nabla(\kappa - \tau E)\nabla\xi|dx \\ &\leq \ \frac{c^2}{2\tau}\int_{\Omega\cap\{\frac{\mu}{e}<\xi<\mu\}} \xi^{2\theta-1}|\nabla(\kappa - \tau E)|^2 dx + \int_{\Omega\cap\{\frac{\mu}{e}<\xi<\mu\}} \frac{\tau}{2\xi}|\nabla\xi|^2 dx.\end{aligned}$$

In particular, we have

$$\int_\Omega \xi p_\mu(\xi)dx + \tau\int_{\Omega\cap\{\frac{\mu}{e}<\xi<\mu\}} \frac{1}{2\xi}|\nabla\xi|^2 dx \leq \frac{c^2\mu^{2\theta-1}}{2\tau}\int_{\Omega\cap\{\frac{\mu}{e}<\xi<\mu\}} |\nabla(\kappa - \tau E)|^2 dx. \tag{2.16}$$

Since the function $\mu \to 1_{]\frac{\mu}{e},\mu[}$ converge simply to 0 as $\mu \to 0^+$, the theorem of Lebesgue yields

$$\int_{\Omega \cap \{\frac{\mu}{e}<\xi<\mu\}} |\nabla(\kappa - \tau E)|^2 dx \to 0, \ \ as\ \mu \to 0^+.$$

Moreover,

$$\lim_{\mu \to 0^+} p_\mu(\xi) = sign_0^+(\xi) \text{ in } L^2(\Omega).$$

Thus

$$|(w_1 - w_2)^+|_{L^1(\Omega)} \leq 0, \text{ and } w_1 \leq w_2.$$

Note that w_1 and w_2 play the same role, to deduce that $w_1 = w_2$. □

2.4.1 The semi-discretized problem

We consider a uniform partition of $(0,T)$ into N subintervals $[t_k, t_{k+1}[$, $k = 0, ..., N-1$. We denote by $\Delta t = t_{k+1} - t_k$ the time step. Assume that $E^{k+1} \in H^1(\Omega)$ with $E^{k+1} \geq 0$, $f^{k+1} \in L^2(\Omega)$ with $f^{k+1} + E^{k+1} \geq 0$. Denote by $\|.\|_0$ the L^2 norm and by $\|.\|_1$ the H^1 norm. The semi-discretized problem is written: find $u^{k+1} \in H_0^1(\Omega)$ for $k = 0, ..., N-1$ such that

$$\begin{aligned} \frac{u^{k+1} - u^k}{\Delta t} - \text{div}[a(\frac{u^{k+1} - u^k}{\Delta t} + E^{k+1})\nabla(u^{k+1} - \tau E^{k+1})] \\ -\tau\Delta(\frac{u^{k+1} - u^k}{\Delta t} + E^{k+1}) = f^{n+1}, \\ u^0 = u_0 \text{ in } \Omega. \end{aligned} \tag{2.17}$$

Since we are interested in variational solutions, the problem is: find $u^{k+1} \in H_0^1(\Omega)$ for $k = 0, ..., N-1$, such that for all v in $H_0^1(\Omega)$

$$\begin{cases} \frac{1}{\Delta t}\int_\Omega u^{k+1} v\, dx + \int_\Omega a(\frac{w - u^k}{\Delta t} + E^{k+1})\nabla(u^{k+1} - \tau E^{k+1}) \cdot \nabla v\, dx \\ \qquad + \frac{\tau}{\Delta t}\int_\Omega \nabla u^{k+1} \cdot \nabla v\, dx = \int_\Omega f^{k+1} v\, dx \\ \quad + \frac{1}{\Delta t}\int_\Omega u^k v\, dx + \frac{\tau}{\Delta t}\int_\Omega \nabla u^k \cdot \nabla v\, dx - \tau\int_\Omega \nabla E^{k+1} \cdot \nabla v\, dx, \\ \qquad u^0 = u_0 \ \text{ in } \Omega. \end{cases} \tag{2.18}$$

Note that we have just to prove the result for the first iteration. Denote by $u = u^1$, $E = E^1$, $f^1 = f$ and $w := \frac{u - u_0}{\Delta t}$. The problem becomes:

$$\begin{aligned} \int_\Omega wv\, dx + \Delta t\int_\Omega (a(w+E) + \frac{\tau}{\Delta t})\nabla w \cdot \nabla v\, dx + \int_\Omega a(w+E)\nabla(u_0 - \tau E) \cdot \nabla v\, dx \\ = \int_\Omega fv\, dx - \tau\int_\Omega \nabla E \cdot \nabla v\, dx. \end{aligned} \tag{2.19}$$

Lemma 2.3 *Assume hypothesis (H) for data a and τ. For a given $u_0 \in H_0^1(\Omega)$, a unique solution w exists to the problem (2.19).*
The discrete version of the constraint is implicitly satisfied: $\frac{u-u_0}{\Delta t} + E \geq 0$ a.e. in Ω.

Proof. The existence of w is based on the fixed point theorem of Schauder-Tychonov. Let $\psi : H_0^1(\Omega) \to H_0^1(\Omega), S \mapsto w = \psi(S)$ the unique solution of the following linear problem: Find $w \in H_0^1(\Omega)$ such that for all $v \in H_0^1(\Omega)$:

$$\begin{aligned} \int_\Omega wv\,dx + \Delta t \int_\Omega (a(S+E) + \frac{\tau}{\Delta t})\nabla w \cdot \nabla v\,dx &= \int_\Omega fv\,dx \\ - \int_\Omega a(S+E)\nabla(u_0 - \tau E)\nabla v\,dx - \tau \int_\Omega \nabla E \nabla v dx. \end{aligned} \tag{2.20}$$

The existence of a solution to this linear problem is guaranteed by the Lax-Milgram lemma since $\Delta t\, a(S+E) + \tau \geq \tau > 0$. Using $v = w$ in the weak formulation, we have

$$\begin{aligned} \|w\|_0^2 + \int_\Omega (\Delta t\, a(S+E) + \tau)|\nabla w|^2 dx &= \int_\Omega fw dx - \int_\Omega a(S+E)\nabla(u_0 - \tau E)\nabla w dx \\ &- \tau \int_\Omega \nabla E \nabla w dx. \end{aligned}$$

Using the assumptions, Cauchy-Schwarz and Young's inequalities, we obtain

$$\frac{1}{2}\|w\|_0^2 + \frac{\tau}{2}\|\nabla w\|_0^2 \leq \frac{1}{2}\|f\|_0^2 + 4\tau\|\nabla E\|_0^2 + \frac{1}{\tau}\|\nabla u_0\|_0^2.$$

We conclude that

$$\|w\|_1 \leq C(\tau, \|f\|_0, \|\nabla u_0\|_0, \|\nabla E\|_0).$$

The set $K = \{v \in H_0^1(\Omega);\ \|v\|_1 \leq C\}$ is not empty, bounded, convex and strongly closed. Thus weakly compact in $H_0^1(\Omega)$. Moreover, $\psi(K) \subset K$.
Consider $S_n \rightharpoonup S$ in $H_0^1(\Omega)$. Thus $S_{n'} \to S$ in $L^2(\Omega)$ and a.e. in Ω for a sub-sequence, since the injection of $H_0^1(\Omega) \subset L^2(\Omega)$ is compact. Since a is continuous and bounded, for any v in $H_0^1(\Omega)$,

$$a(S_{n'} + E)\nabla v \to a(S+E)\nabla v \quad \text{a.e. in } \Omega \text{ and in } \left(L^2(\Omega)\right)^d.$$

Let $w_{S_{n'}} := \psi(S_{n'})$. The boundedness of $(w_{S_{n'}})$ in $H_0^1(\Omega)$ allows us to extract a sub-sequence $(w_{S_{n''}})$ such that

$$w_{S_{n''}} \rightharpoonup \xi \text{ in } H_0^1(\Omega).$$

By passing to the limits ($n'' \to \infty$), we conclude that ξ satisfies: $\xi \in H_0^1(\Omega)$ and for all $v \in H_0^1(\Omega)$

$$\begin{aligned} \int_\Omega \xi v\,dx + \Delta t \int_\Omega (a(S+E) + \frac{\tau}{\Delta t})\nabla \xi \cdot \nabla v\,dx \\ = \int_\Omega fv\,dx - \int_\Omega a(S+E)\nabla(u_0 - \tau E)\cdot \nabla v\,dx - \tau \int_\Omega \nabla E \cdot \nabla v\,dx. \end{aligned} \tag{2.21}$$

Thus $\xi = w_S$ (the uniqueness of the solution) and all the sequence $(w_{S_n})_n$ converges weakly to w_S. ψ is sequentially continuous with respect to the weak convergence. The existence of a fixed point w_s then results from the Schauder fixed point theorem.
For the discrete constraint, the assertion follows by taking in (2.19) $v = (\frac{u-u_0}{\Delta t} + E)^-$. □

Lemma 2.4 *Assume hypothesis (H) for data a and τ, $f^{k+1} \in L^2(\Omega)$, $E^{k+1} \in H^1(\Omega)$ such that $f^{k+1} + E^{k+1} \geq 0$ and u_0 in $H_0^1(\Omega)$. Then a sequence (u^{k+1}) exists in $H_0^1(\Omega)$, solution to the problem (2.18). As expected, the discrete version of the constraint is satisfied.*

$$\frac{u^{k+1} - u^k}{\Delta t} + E^{k+1} \geq 0 \text{ a.e. in } \Omega.$$

Proof. The existence of the sequence, as well as the constraint condition, is an induction of the previous result. □

2.4.2 *A priori* estimates

For all sequences $(u^k)_{0 \leq k \leq N-1}$, with $u^0 = u_0$, we define in Q the functions

$$u_{\Delta t}(t) = \sum_{k=0}^{N-1} u^k \, \mathbf{1}_k(t),$$

where $\mathbf{1}_k$ is the characteristic function in the interval $[k\Delta t, (k+1)\Delta t[$, and

$$\tilde{u}_{\Delta t}(t) = (t - k\Delta t)\frac{u^{k+1} - u^k}{\Delta t} + u^k, \text{ if } t \in [k\Delta t, (k+1)\Delta t[, \ 0 \leq k \leq N-1. \tag{2.22}$$

We note that, the derivative $\partial_t \tilde{u}_{\Delta t}$, calculated in the distribution sense is, t a.e.

$$\partial_t \tilde{u}_{\Delta t}(t) = \sum_{k=0}^{N-1} \frac{u^{k+1} - u^k}{\Delta t} \, \mathbf{1}_k(t).$$

For a.e. t in $(0,T)$ and for all v in $H_0^1(\Omega)$, one has that

$$\int_\Omega [\partial_t \tilde{u}_{\Delta t} v + a(\partial_t \tilde{u}_{\Delta t} + E_{\Delta t})\nabla(u_{\Delta t} - \tau E_{\Delta t})\nabla v + \tau \nabla(\partial_t \tilde{u}_{\Delta t} + E_{\Delta t})\nabla v]dx = \int_\Omega f_{\Delta t} v dx, \tag{2.23}$$

where $E_{\Delta t}$ and $f_{\Delta t}$ are given by

$$E_{\Delta t} = \sum_{k=0}^{N-1} \frac{1}{\Delta t} \int_{k\Delta t}^{(k+1)\Delta t} E(.,s)ds \mathbf{1}_k, \quad f_{\Delta t} = \sum_{k=0}^{N-1} \frac{1}{\Delta t} \int_{k\Delta t}^{(k+1)\Delta t} f(.,s)ds \mathbf{1}_k.$$

Thanks to those notations, we are able to say that

Lemma 2.5 *For small Δt*

- *The sequences $(u_{\Delta t})$ and $(\tilde{u}_{\Delta t})$ are bounded in $L^\infty(0,T;H_0^1(\Omega))$,*
- *The sequence $(\partial_t \tilde{u}_{\Delta t})$ is bounded in $L^\infty(0,T;H_0^1(\Omega))$.*

Proof. Let us consider $v = \frac{u^{k+1}-u^k}{\Delta t}$ in (2.18), in order to get

$$\begin{aligned}\|\frac{u^{k+1}-u^k}{\Delta t}\|_0^2 + \tau\|\nabla(\frac{u^{k+1}-u^k}{\Delta t})\|_0^2 &\leq \|\nabla(u^{k+1}-\tau E^{k+1})\|_0\|\nabla(\frac{u^{k+1}-u^k}{\Delta t})\|_0 \\ &+ \tau\|\nabla E^{k+1}\|_0\|\nabla(\frac{u^{k+1}-u^k}{\Delta t})\|_0 + \|f^{k+1}\|_0\|\frac{u^{k+1}-u^k}{\Delta t}\|_0.\end{aligned}$$

Thus, for any positive ϵ_1, ϵ_2, one has that

$$\begin{aligned}\|\frac{u^{k+1}-u^k}{\Delta t}\|_0^2 + \tau\|\nabla(\frac{u^{k+1}-u^k}{\Delta t})\|_0^2 &\leq \frac{\epsilon_1+\epsilon_2}{2}\|\nabla(\frac{u^{k+1}-u^k}{\Delta t})\|_0^2 + \frac{1}{2\epsilon_1}\|\nabla(u^{k+1}-\tau E^{k+1})\|_0^2 \\ &+ \frac{\tau^2}{2\epsilon_2}\|\nabla E^{k+1}\|_0^2 + \frac{1}{2}\|f^{k+1}\|_0^2 + \frac{1}{2}\|\frac{u^{k+1}-u^k}{\Delta t}\|_0^2.\end{aligned}$$

With $\epsilon_1 = \epsilon_2 = \frac{\tau}{2}$, we have

$$\begin{aligned}\frac{1}{2}\|\frac{u^{k+1}-u^k}{\Delta t}\|_0^2 + \frac{\tau}{2}\|\nabla(\frac{u^{k+1}-u^k}{\Delta t})\|_0^2 &\leq \frac{1}{\tau}\|\nabla(u^{k+1}-\tau E^{k+1})\|_0^2 + \tau\|\nabla E^{k+1}\|_0^2 + \frac{1}{2}\|f^{k+1}\|_0^2 \\ &\leq \frac{2}{\tau}\|\nabla u^{k+1}\|_0^2 + 3\tau\|\nabla E^{k+1}\|_0^2 + \frac{1}{2}\|f^{k+1}\|_0^2,\end{aligned}$$

and

$$\begin{aligned}\|\frac{u^{k+1}-u^k}{\Delta t}\|_0^2 + \tau\|\nabla(\frac{u^{k+1}-u^k}{\Delta t})\|_0^2 &\leq \frac{8\Delta t^2}{\tau}\|\sum_{m=0}^{k}\nabla(\frac{u^{m+1}-u^m}{\Delta t})\|_0^2 + \frac{8}{\tau}\|\nabla u_0\|_0^2 \\ &+ 6\tau\|\nabla E^{k+1}\|_0^2 + \|f^{k+1}\|_0^2 \\ &\leq \frac{8\Delta t^2}{\tau}(k+1)\sum_{m=0}^{k}\|\nabla(\frac{u^{m+1}-u^m}{\Delta t})\|_0^2 + M \\ &\leq \frac{8T\Delta t}{\tau}\sum_{m=0}^{k}\|\nabla(\frac{u^{m+1}-u^m}{\Delta t})\|_0^2 + M,\end{aligned}$$

where M is given by

$$M = \frac{8}{\tau}\|\nabla u_0\|_0^2 + 6\tau\|\nabla E^{k+1}\|_0^2 + \|f^{k+1}\|_0^2.$$

Denote by $x_k = \|\nabla(\frac{u^{k+1}-u^k}{\Delta t})\|_0^2$, using hypothesis (H) for data E and f, there is a constant C independent of Δt such that

$$x_k \leq \frac{8T\Delta t}{\tau^2}\sum_{m=0}^{k} x_m + C.$$

If $\Delta t \leq \frac{\tau^2}{16T}$, we get

$$x_k \leq \frac{16T\Delta t}{\tau^2}\sum_{m=0}^{k-1} x_m + 2C.$$

On the one hand, by using the discrete version of the lemma of Gronwall, one obtains

$$x_k \leq 2C\exp(\frac{16T}{\tau^2}k\Delta t) \leq C\exp(\frac{16T^2}{\tau^2}).$$

Thus

$$\|\nabla(\frac{u^{k+1}-u^k}{\Delta t})\|_0^2 \leq C, \ and \ \ \|\partial_t \tilde{u}_{\Delta t}\|_{L^\infty(0,T;H_0^1(\Omega))} \leq C. \tag{2.24}$$

On the other hand, we have

$$\begin{aligned}
\|\nabla u^k\|_0 &= \|\sum_{m=0}^{k-1}\nabla(\frac{u^{m+1}-u^m}{\Delta t}).\Delta t + \nabla u_0\|_0 \\
&\leq \sum_{m=0}^{k-1}\Delta t\|\nabla(\frac{u^{m+1}-u^m}{\Delta t})\|_0 + \|\nabla u_0\|_0 \\
&\leq (\sum_{m=0}^{k-1}\Delta t)^{1/2}(\sum_{m=0}^{k-1}\frac{\|\nabla(u^{m+1}-u^m)\|_0^2}{\Delta t})^{1/2} + \|\nabla u_0\|_0 \\
&\leq \sqrt{T}C + \|\nabla u_0\|_0.
\end{aligned}$$

Thus, we get

$$\|u_{\Delta t}\|_{L^\infty(0,T;H_0^1(\Omega))} \leq C. \tag{2.25}$$

From (2.24) and (2.25) we deduce that

$$\|\tilde{u}_{\Delta t}\|_{L^\infty(0,T;H_0^1(\Omega))} \leq C, \tag{2.26}$$

and the assertions in Lemma (2.5) are proved. □

2.4.3 Existence result

Let us prove now the existence of a solution u to the problem (P).

Proposition 2.1 *There exists at least an element u in $H^1(0,T;H_0^1(\Omega))$ such that, t a.e. in $(0,T)$: $\forall\, v \in H_0^1(\Omega)$,*

$$\int_\Omega [\partial_t uv + a(\partial_t u + E)\nabla(u - \tau E)\nabla v + \tau\nabla(\partial_t u + E)\nabla v]dx = \int_\Omega fvdx. \tag{2.27}$$

Moreover, the constraint $\partial_t u + E \geq 0$ is implicitly satisfied.

Proof. The proof is based on *a priori* estimations on the sequence (u^n).
From Lemma (2.5), we conclude that, the sequence $(\tilde{u}_{\Delta t})$ is bounded in $H^1(0,T;H_0^1(\Omega))$. Thus u in $H^1(0,T;H_0^1(\Omega))$ exists and a sub-sequence, still denoted by $(\tilde{u}_{\Delta t})$ may be extracted such that, $\tilde{u}_{\Delta t}$ converges weakly to u in $H^1(0,T;H_0^1(\Omega))$ and, for all $t \in (0,T)$,

$$\tilde{u}_{\Delta t}(t) \rightharpoonup u(t) \text{ in } H_0^1(\Omega).$$

In particular, we have $u(0,.) = u_0$ a.e. in Ω. Moreover, we have $\forall\, t \in [k\Delta t, (k+1)\Delta t[$,

$$\begin{aligned} \|u_{\Delta t}(t) - \tilde{u}_{\Delta t}(t)\|_{H_0^1(\Omega))} = \|\tilde{u}_{\Delta t}(k\Delta t) - \tilde{u}_{\Delta t}(t)\|_{H_0^1(\Omega))} &\leq \int_{k\Delta t}^{(k+1)\Delta t} \|\partial_t \tilde{u}_{\Delta t}(s)\|_{H_0^1(\Omega))} ds, \\ &\leq C\Delta t. \end{aligned}$$

Then, for all $t \in (0,T)$

$$u_{\Delta t}(t) \rightharpoonup u(t) \text{ in } H_0^1(\Omega).$$

Moreover, note that by construction, for a.e. t in $]0,T[$, $E_{\Delta t}(t)$ converges to $E(t)$ in $H^1(\Omega)$ and $f_{\Delta t}(t)$ to $f(t)$ in $L^2(\Omega)$.
From Lemma (2.5), we have that the sequence $(\partial_t \tilde{u}_{\Delta t})$ is bounded in $L^\infty(0,T;H_0^1(\Omega))$. Then, there is Z, subset of $(0,T)$, with $\mathcal{L}((0,T),Z) = 0$, such that, in addition to the above convergences, for all $t \in Z$ the sequence $(\partial_t \tilde{u}_{\Delta t})(t)$ is in a fixed bounded subset of $H_0^1(\Omega)$. We may extract a sub-sequence denoted by $(\partial_t \tilde{u}_{\Delta t_t})$, such that

$$\partial_t \tilde{u}_{\Delta t_t}(t) \rightharpoonup \xi(t) \text{ in } H_0^1(\Omega).$$

The embedding of $H_0^1(\Omega)$ in $L^2(\Omega)$ is compact, thus

$$\partial_t \tilde{u}_{\Delta t_t}(t) \to \xi(t) \text{ in } L^2(\Omega) \text{ a.e. in } \Omega \text{ by a sub-sequence denoted by the same way.} \tag{2.28}$$

Thus, since a is continuous and bounded, for all v in $H_0^1(\Omega)$, we have

$$a(\partial_t \tilde{u}_{\Delta t_t}(t) + E_{\Delta t}(t))\nabla v \to a(\xi(t) + E(t))\nabla v \text{ in } (L^2(\Omega))^d. \tag{2.29}$$

The derivative operators $\frac{\partial}{\partial x_i}$ are linear and continuous from $H_0^1(\Omega)$ into $L^2(\Omega)$, thus we get that

$$\begin{aligned} \forall\, t \in Z, \qquad & \nabla(\tilde{u}_{\Delta t}(t) - \tau E_{\Delta t}(t)) \rightharpoonup \nabla(u(t) - \tau E(t)) \text{ in } (L^2(\Omega))^d, \\ & \nabla(\partial_t \tilde{u}_{\Delta t_t}(t) + E_{\Delta t}(t)) \rightharpoonup \nabla(\xi(t) + E(t)) \text{ in } (L^2(\Omega))^d. \end{aligned}$$

Passing to the limits on Δt ($\Delta t \to 0^+$) in (2.23), for all v in $H_0^1(\Omega)$, $\xi(t)$ satisfies the equation

$$\int_\Omega [\xi(t)v + a(\xi(t) + E(t))\nabla(u(t) - \tau E(t))\nabla v + \tau \nabla(\xi(t) + E(t))\nabla v]dx = \int_\Omega f(t)v dx.$$

Then, using lemma (2.2), with $b = a$ and $\kappa = u(t)$ and hypothesis (H), the solution $\xi(t)$ is unique in $H_0^1(\Omega)$. Then, all the sequence $(\partial_t \tilde{u}_k(t))_{\Delta t}$, and not only the sub-sequences $(\partial_t \tilde{u}_{\Delta t_t})(t)$ extract from $(\partial_t \tilde{u}_{\Delta t})(t)$, converges in $H_0^1(\Omega)$ toward $\xi(t)$. Thus, we have for all t in Z

$$\partial_t \tilde{u}_{\Delta t}(t) \rightharpoonup \xi(t) \text{ in } H_0^1(\Omega).$$

Therefore, the function $\xi :]0, T[\to H_0^1(\Omega)$ is weakly measurable (indeed, for any g in $H^{-1}(\Omega)$, $t \mapsto \langle g, \xi(t)\rangle$ is the limit of a sequence of measurable functions $t \mapsto \langle g, \partial_t \tilde{u}_{\Delta t}(t)\rangle$) and consequently ξ is a measurable function thanks to the theorem of Pettis [42], since $H_0^1(\Omega)$ is a separable set.
Moreover, for any v in $L^2(0, T; H_0^1(\Omega))$,

$$(\partial_t \tilde{u}_{\Delta t}(t), v(t)) \rightharpoonup (\xi(t), v(t)) \text{ a.e. in }]0, T[.$$

As the sequence $(\partial_t \tilde{u}_{\Delta t})$ is bounded in $L^\infty(0, T; H_0^1(\Omega))$, there exists a constant C such that

$$|(\partial_t \tilde{u}_{\Delta t}(t), v(t))| \leq C \|v(t)\|_1.$$

Thus, we conclude that $\partial_t \tilde{u}_{\Delta t} \rightharpoonup \xi$ in $L^2(0, T; H_0^1(\Omega))$.
Thus, we get that $\xi = \partial_t u$, then by passing to the limits ($\Delta t \to 0^+$), we have, for t a.e. in $(0, T)$, for all v in $L^2(0, T; H_0^1(\Omega))$

$$\int_\Omega [\partial_t u v + a(\partial_t u + E)\nabla(u - \tau E)\nabla v + \tau \nabla(\partial_t u + E)\nabla v]dx = \int_\Omega f v dx, \tag{2.30}$$

i.e. a solution to problem (P) exists. □

Chapter 3

Dg time discretization of the Sobolev equation

3.1 Introduction

The aim of this chapter is to discuss the time-approximation of the Sobolev equation; which is a linearized version of the stratigraphic model with parameter λ equal to one.
Thus, we get the following linear problem:

$$\partial_t u - \Delta u - \tau \Delta \partial_t u = f, \tag{3.1}$$

for $x \in \Omega$, $t \in J = (0, T]$, $\tau > 0$, where Ω is a bounded domain in $\mathbb{R}^d$ with Lipschitz boundary $\partial\Omega$.
We consider homogeneous Dirichlet boundary condition for u and for $\partial_t u$, and the following initial condition:

$$u(x, 0) = u_0(x), \quad x \in \Omega. \tag{3.2}$$

where $u_0 \in H_0^1(\Omega)$.

Equation (3.1) is of pseudoparabolic type, which means that the time-derivative appears in the highest order term of the operator. Such equations are called Sobolev equations. They arise in many applications (cf. chap. Introduction). The nature of such problems is transient and, therefore, an appropriate time stepping scheme has to be applied to obtain an approximative solution. A flexible and efficient time discretization method is the discontinuous Galerkin finite element one (Dg) which is based on variational formulation of initial value problems.
This type of equation was studied by Benjamin, Bona, and Mahony [13] to describe unidirectional long dispersive waves. Theoretical results on existence, uniqueness, regularity,

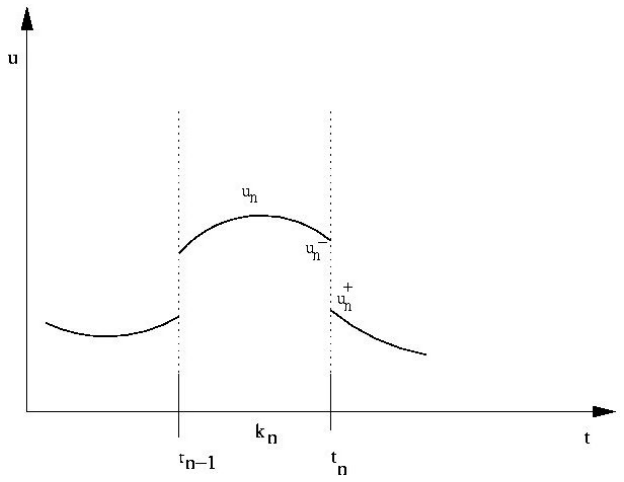
and decay at infinite time of solution for (3.1) are studied in Benjamin [13] and Tran [1], for example. Numerical approximations based on finite differences, finite elements, and spectral methods have been considered in Douglas, Bona, Ewing [32, 14, 23].

The outline of this chapter is as follows: In section 3.2, we introduce the discontinuous Galerkin finite element method (Dg) for the problems (3.1)-(3.2). In section 3.3, error estimates that are explicit in the polynomial order r and time step k are derived. The section 3.4 concerns the discretization in time and space of the problem. In section 3.5, we present some numerical results.

To describe the time-discretization, we use Bochner spaces of functions which map a time interval $I = (a, b)$ into a separable Hilbert space V: we denote by $L^p(I; V),\ 1 \leq p \leq \infty$ and $H^k(I; V),\ 0 \leq k \in \mathbb{R}$, the corresponding Lebesgue and Sobolev spaces. $C^k(\bar{I}; V)$ are the functions that are k times continuously differentiable. $\mathcal{P}^r(I; V)$ denotes the set of all the polynomials of degree less or equal to r, with coefficients in V, *i.e.* $p \in \mathcal{P}^r(I; V)$ if and only if $p(t) = \sum_{j=0}^{r} \mathcal{X}_j t^j$ for some $\mathcal{X}_j \in V$ and $t \in I$. P_r denotes the orthogonal projector of $L^2(I, V)$ on $\mathcal{P}^r(I, V)$, and we set $\mathbb{V} = L^2(I, V)$.
We emphasize here that the space V typically is infinite dimensional function space and, in this sense, the theoretical setting of this chapter is semi-discrete. In practice, the spatial operator might also have to be discretized. This will be addressed in section 3.4 below.

3.2 Dg time discretization

In order to discretize Problem (3.1) in time by Dg, let us start by introducing time meshes. Let $\mathcal{M}$ be a partition of $J =]0, T[$ into $N(\mathcal{M})$ subintervals $\{I_n\}_{n=1}^{N}$ given by $I_n =]t_{n-1}, t_n[$ with nodes $0 =: t_0 < t_1 < ... < t_{N-1} < t_N := T$. The length of time step I_n is given by $k_n := t_n - t_{n-1}$.
The idea of Dg is to approximate the exact solution u by a semi-discrete function u_k which, on each step I_n, consists in a polynomial in t of order r_n with coefficients in V.



Of course, u is not required to be continuous across the time nodes. This allows us to write the Dg as a time stepping scheme. In order to deal with the discontinuities across nodes t_n, we define the left- and right-handed limits of a function $u:\ J \to H := L^2(\Omega)$ (or $u:\ J \to V := H_0^1(\Omega)$) to be

$$u_n^+ = \lim_{s \to 0^+} u(t_n + s),\ 0 \le n \le N-1,\ \ u_n^- = \lim_{s \to 0^+} u(t_n - s),\ 1 \le n \le N,$$

when these limits exist.
Furthermore, the jump of u across t_n is defined as

$$[u]_n = u_n^+ - u_n^-,\ \ 0 \le n \le N-1.$$

The semi-discrete space in which we want to discretize (3.1)-(3.2) in time is

$$\mathcal{V}_k^r = \{u:\ J \to V:\ u|_{I_n} \in \mathcal{P}^r(I_n; V),\ 1 \le n \le N\}. \tag{3.3}$$

We consider the following discontinuous Galerkin approximation of (3.1)-(3.2): Find $u_k \in \mathcal{V}_k^r$ such that

$$\forall\ v_k \in\ \mathcal{V}_k^r, \quad a(u_k, v_k) = F(v_k). \tag{3.4}$$

The bilinear form $a(.,.)$ and the form F are given by

$$\begin{aligned} a(u,v) &= \int_0^{t_N} \big((u',v) + (\nabla u, \nabla v) + \tau(\nabla u', \nabla v)\big)dt + \sum_{n=1}^{N-1}([u]_n, v_n^+) \\ &+ (u_0^+, v_0^+) + \tau \sum_{n=1}^{N-1}([\nabla u]_n, \nabla v_n^+) + \tau(\nabla u_0^+, \nabla v_0^+), \end{aligned} \tag{3.5}$$

and

$$F(v) = (u_0, v_0^+) + \tau(\nabla u_0, \nabla v_0^+) + \int_0^{t_N} (f, v)dt.$$

where $u' = \partial_t u$.

Remark 3.1 *Due to the discontinuity of the trial and test space, we may choose the test function v vanishing outside I_n. Thus, (3.4) can be interpreted as an implicit time marching scheme, where u_k is obtained by solving successively evolution problems on I_n (for $n = 1, ..., N$) with initial values $u_{k,n-1}^-$. More precisely: if u_k is already given on the time intervals I_k, $1 \le k \le n-1$, we determine u_k on I_n by solving:*
Find $u_k \in \mathcal{P}^r(I_n; V)$ such that

$$\begin{aligned} \int_{I_n} \{(u_k', v_k)_H &+ (\nabla(u_k + \tau u_k'), \nabla v_k)_H\}dt + (u_{k,n-1}^+, v_{k,n-1}^+)_H + \tau(\nabla u_{k,n-1}^+, \nabla v_{k,n-1}^+)_H \\ &= \int_{I_n} (f, v_k)_H dt + (u_{k,n-1}^-, v_{k,n-1}^+)_H + \tau(\nabla u_{k,n-1}^-, \nabla v_{k,n-1}^+)_H, \end{aligned} \tag{3.6}$$

for all $v_k \in \mathcal{P}^r(I_n; V)$. Here we set $u_{k,0}^- = u_0$.

Lemma 3.1 *For all $v_k, w_k \in \mathcal{V}_k^r$, one has that*

$$\begin{aligned} a(v_k, w_k) &= \sum_{n=1}^{N} \int_{I_n} \{-(v_k, w_k')_H + (\nabla v_k, \nabla(w_k - \tau w_k'))_H\}dt \\ &- \sum_{n=1}^{N-1} (v_{k,n}^-, [w_k]_n)_H + (v_{k,N}^-, w_{k,N}^-)_H \\ &- \tau \sum_{n=1}^{N-1} (\nabla v_{k,n}^-, [\nabla w_k]_n)_H + \tau(\nabla v_{k,N}^-, \nabla w_{k,N}^-)_H, \end{aligned} \tag{3.7}$$

$$\begin{aligned} a(v_k, v_k) &= \sum_{n=1}^{N} \int_{I_n} \|\nabla v_k\|_H^2 dt + \frac{1}{2}\|v_{k,0}^+\|_H^2 + \frac{\tau}{2}\|\nabla v_{k,0}^+\|_H^2 \\ &+ \frac{1}{2}\sum_{n=1}^{N-1} \|[v_k]_n\|_H^2 + \frac{\tau}{2}\sum_{n=1}^{N-1} \|[\nabla v_k]_n\|_H^2. \end{aligned} \tag{3.8}$$

Proof. Integration by parts in time and rearranging the nodal contributions give (3.7). To prove (3.8), note that

$$a(v_k, v_k) = \frac{1}{2}a(v_k, v_k) + \frac{1}{2}a(v_k, v_k) =: T_1 + T_2.$$

Evaluating T_1 with (3.5) and T_2 with (3.7) show the assertion. □

Notation:
Let us set in the sequel the following discrete energy norm: $\forall\, v_k \in \mathcal{V}_k^r$, $|||v_k|||^2 := a(v_k, v_k)$.

Proposition 3.1 *The Dg (3.4) has a unique solution $u_k \in \mathcal{V}_k^r$. Moreover, if $u \in H^1(0,T;V)$ is the solution to (3.1)-(3.2), one has the Galerkin orthogonality*

$$\forall\, v_k \in\ \mathcal{V}_k^r, \quad a(u-u_k, v_k) = 0.$$

Proof. Since the bilinear form a is continuous and coercive and the form linear F is continuous, the Proposition comes from the Lax-Milgram theorem. □

3.3 *A priori* error analysis

In this section, we derive error estimates for the Dg, explicit in time steps k_n and in the r-order polynomials. First, we introduce a projector and show that it is well defined. Then, we derive estimate for this projector and we give *a priori* error estimate for Dg.

3.3.1 Interpolation error

In this section we introduce and analyze a projector. Let $I = (-1,1)$ and denote by $\{L_i\}_{i\geq 0}$, $L_i \in \mathcal{P}^i(I)$, the Legendre polynomials on I. For $u, v \in H^1(\Omega)$, we define the scalar product

$$(u,v)_\tau = (u,v)_H + \tau(\nabla u, \nabla v)_H,$$

where $(.,.)$ denotes the L^2 scalar product.

Definition 3.1 *Let $I = (-1,1)$. For a function $u \in\ L^2(I;V)$ which is continuous at $t = 1$, we define $\Pi^r u \in\ \mathcal{P}^r(I;V)$, $r \in \mathbb{N}$, $r \geq 1$, via the $r+1$ conditions*

$$\forall\, q \in\ \mathcal{P}^{r-1}(I;V), \quad \int_I (\Pi^r u - u, q)_\tau dt = 0, \qquad \Pi^r u(1) = u(1) \text{ in } V. \tag{3.9}$$

Lemma 3.2 *Π^r is well defined.*

Proof. **Uniqueness**:
Assume that u_1 and u_2 are two polynomials in $\mathcal{P}^r(I;V)$ which satisfy (3.9), especially, we have $u_1(1) = u_2(1)$ in V.
Denote by L_i, $i \geq 0$, the Legendre polynomial of degree $i \leq r$. The difference $u_1 - u_2$ can be developed into the series:

$$u_1 - u_2 = \sum_{i=0}^{r} v_i L_i, \text{ with } v_i = \frac{2i+1}{2} \int_I (u_1 - u_2) L_i dt\ \in\ V.$$

Fix now $k \in \{0, ..., r-1\}$. From (3.9), it follows that

$$\forall\, v \in V, \quad \int_I (u_1 - u_2, vL_k)_\tau dt = 0.$$

Using the orthogonality properties of the Legendre polynomials, we get

$$(v_k, v)_\tau = 0 \;\; for\ all\ v \in V.$$

We conclude that $v_k = 0\ in\ V$, which gives

$$u_1 - u_2 = v_r L_r, \;\; v_r \in V.$$

Since $u_1(1) = u_2(1)$, and $L_r(1) = 1$, we have $v_r = 0$, which proves the uniqueness of a polynomial satisfying the conditions in Definition 3.1.
The existence follows similarly by setting

$$\Pi^r u = \sum_{i=0}^{r-1} u_i L_i + (u(1) - \sum_{i=0}^{r-1} u_i) L_r. \tag{3.10}$$

□

Definition 3.2 *On an arbitrary interval $I_n = (t_{n-1}, t_n)$, with $k_n := t_n - t_{n-1}$, we define $\Pi^r_{I_n}$ via the linear map $Q : (-1, 1) \to I_n$, $\hat{t} \mapsto t = \frac{1}{2}(t_{n-1} + t_n + \hat{t}k_n)$ as*

$$\Pi^r_{I_n} u = [\Pi^r (u \circ Q)] \circ Q^{-1}.$$

Lemma 3.3 *Let $u \in H^1(I; V)$ and let $u = \sum_{i=0}^{\infty} u_i L_i$ be the Legendre expansion of u with coefficient $u_i = \frac{2i+1}{2} \int_I u L_i(t) dt \in V$. For $r \in \mathbb{N}_0$, we denote by P^r the $L^2(I; V)$-projection onto $\mathcal{P}^r(I; V)$. There holds:*

$$\|u - \Pi^r u\|^2_{L^2(I;V)} = \|u - P^r u\|^2_{L^2(I;V)} + \frac{2}{2r+1} \|u(1) - (P^r u)(1)\|^2_V.$$

Proof. From the Definition of Π^r we have: $\Pi^r u = \sum_{i=0}^{r-1} u_i L_i + (u(1) - P^r u(1)) L_r$. Therefore,

$$u - \Pi^r u = (u - P^r u) - (u(1) - P^r u(1)) L_r.$$

The assertion follows by using the orthogonality properties of Legendre polynomials. □

We recall the following approximation result from [39]. There, the proof is presented for real-value functions, but the extension to the Bochner spaces considered here is straightforward.

Proposition 3.2 *Let $I = (-1, 1)$ and let $u \in H^{k+1}(I; V)$ for some integer $k \geq 1$. Then there exists $q \in \mathcal{P}^r(I; V)$, $r \geq 1$, such that*

$$\|u' - q'\|^2_{L^2(I;V)} \leq \frac{(r-s)!}{(r+s)!}\|u^{(s+1)}\|^2_{L^2(I;V)},$$

$$\|u - q\|^2_{L^2(I;V)} \leq \frac{1}{max(1, r^2)}\frac{(r-s)!}{(r+s)!}\|u^{(s+1)}\|^2_{L^2(I;V)},$$

for any $0 \leq s \leq min(r, k)$. Additionally, $q(\pm 1) = u(\pm 1)$ if $r \geq 1$.

Now, we give an estimation of the projector Π^r:

Theorem 3.1 *Let $I_n = (t_{n-1}, t_n)$, $k_n := t_n - t_{n-1}$, $u \in H^{r_n+1}(I_n; V)$, $r_n \geq 1$ the approximation order on I_n. Then, we have*

$$\|u - \Pi^{r_n}_{I_n} u\|_{L^2(I_n;V)} \leq \big(\frac{k_n}{2}\big)^{r_n+1}\frac{2\sqrt{2}}{r_n^2 r_n!}\|u^{(r_n+1)}\|_{L^2(I_n;V)}. \tag{3.11}$$

Proof. For the simplicity, we give the proof for the reference element $I = (-1, 1)$ with approximation order r. Then, we conclude by using the transformation Q.
Assume that $u \in H^{r+1}(I; V)$, using the Proposition 3.2, we obtain the following approximation estimate for P^r:

$$\|u - P^r u\|^2_{L^2(I;V)} \leq \frac{1}{max(1, r^2)}\frac{1}{(2r)!}\|u^{(r+1)}\|^2_{L^2(I;V)}, \tag{3.12}$$

For the second term in Lemma 3.3, we use the following Darboux-Christoffel formula:

$$(P^r u)(s) = \frac{r+1}{2}\int_{-1}^{1}\frac{L_{r+1}(s)L_r(t) - L_r(s)L_{r+1}(t)}{s-t}u(t)dt, \tag{3.13}$$

and

$$\frac{r+1}{2}\int_{-1}^{1}\frac{L_{r+1}(s)L_r(t) - L_r(s)L_{r+1}(t)}{s-t}dt = 1.$$

Then we get

$$(P^r u)(s) - u(s) = \frac{r+1}{2}\int_{-1}^{1}\frac{L_{r+1}(s)L_r(t) - L_r(s)L_{r+1}(t)}{s-t}(u(t) - u(s))dt.$$

In particular for $s = 1$,

$$(P^r u)(1) - u(1) = \frac{r+1}{2}\int_{-1}^{1}\frac{L_{r+1}(t) - L_r(t)}{t-1}(u(t) - u(1))dt.$$

By using the following Legendre polynomial properties, we come to

$$L_r = \frac{1}{2^r r!}\big(\frac{d}{dt}\big)^r(1-t^2)^r,$$

thus

$$L_{r+1} - L_r = \frac{1}{2^r r!}\big(\frac{d}{dt}\big)^r\big(-(1+t)(1-t^2)^r\big).$$

Now, assume that u in $H^{r+1}(I;V)$, with $r \geq 1$. By setting $\alpha_r = \frac{r+1}{2^{r+1}r!}$, we obtain that

$$\begin{aligned}
(P^r u)(1) - u(1) &= \alpha_r \int_I \big(\frac{d}{dt}\big)^r\big(-(1+t)(1-t^2)^r\big)\frac{u(t)-u(1)}{t-1}dt \\
&= -\alpha_r \int_I \big(\frac{d}{dt}\big)^r\big((1+t)(1-t^2)^r\big)\int_t^1 \frac{u'(s)}{1-t}dsdt \\
&= -\alpha_r \int_I \big(\frac{d}{dt}\big)^r\big((1+t)(1-t^2)^r\big)\int_0^1 u'\big((1-t)\sigma+t\big)d\sigma dt \\
&= (-1)^{r+1}\alpha_r \int_I \big((1+t)(1-t^2)^r\big)\int_0^1 (1-\sigma)^r u^{(r+1)}\big((1-t)\sigma+t\big)d\sigma dt \\
&= (-1)^{r+1}\alpha_r \int_I \big((1+t)(1-t^2)^r\big)\int_t^1 u^{(r+1)}(s)\frac{(1-s)^r}{(1-t)^{r+1}}dsdt. \\
\|(P_r u)(1) - u(1)\|_V &\leq \alpha_r \int_I \big|(1+t)(1-t^2)^r\big|\big(\frac{1}{1-t}\big)^{r+1}\int_t^1 (1-s)^r\|u^{(r+1)}(s)\|_V dsdt.
\end{aligned}$$

Using Cauchy-Schwarz inequality, we get

$$\begin{aligned}
\|(P^r u)(1) - u(1)\|_V &\leq \alpha_r \int_{-1}^1 (1+t)^{r+1}\big(\frac{1}{1-t}\big)\big(\int_t^1 \|u^{(r+1)}(s)\|_V^2 ds\big)^{\frac{1}{2}}\big(\int_t^1 (1-s)^{2r}ds\big)^{\frac{1}{2}}dt \\
&\leq \frac{\alpha_r}{\sqrt{2r+1}}\int_{-1}^1 (1+t)^{r+1}(1-t)^{r-\frac{1}{2}}\big(\int_{-1}^1 \|u^{(r+1)}(s)\|_V^2 ds\big)^{\frac{1}{2}}dt \\
&\leq \frac{\alpha_r}{\sqrt{2r+1}}\|u^{(r+1)}\|_{L^2(I;V)}\int_{-1}^1 (1+t)^{r+1}(1-t)^{r-\frac{1}{2}}dt.
\end{aligned}$$

Setting $s = \frac{1-t}{2}$, we get

$$\begin{aligned}
\|(P^r u) - u(1)\|_V &\leq 2^{2r+\frac{3}{2}}\frac{\alpha_r}{\sqrt{2r+1}}\|u^{(r+1)}\|_{L^2(I;V)}\int_0^1 (1-s)^{r+1}(s)^{r-\frac{1}{2}}ds \\
&\leq 2^{r+\frac{1}{2}}\frac{r+1}{r!\sqrt{2r+1}}\|u^{(k+1)}\|_{L^2(I;V)}\int_0^1 (1-s)^{r+1}(s)^{r-\frac{1}{2}}ds \\
&\leq 2^{r+\frac{1}{2}}\frac{(r+1)}{r!\sqrt{2r+1}}\|u^{(r+1)}\|_{L^2(I;V)}B(r+2,r+\frac{1}{2}), \qquad (3.14)
\end{aligned}$$

where B is the beta function and is defined by

$$\forall\ (p,q) \in]0,\infty[^2, B(p,q) = \int_0^1 (1-t)^{p-1}t^{q-1}dt,$$

which is related to the gamma function by

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

Then, we obtain

$$B(r+2, r+\frac{1}{2}) = \frac{\Gamma(r+2)\Gamma(r+\frac{1}{2})}{\Gamma(2r+\frac{5}{2})} = \frac{(r+1)!\Gamma(r+\frac{1}{2})}{\Gamma(2r+\frac{5}{2})}.$$

Replacing B in (3.14), we get

$$\begin{aligned}\|(P^r u)(1) - u(1)\|_V &\leq 2^{r+\frac{1}{2}} \frac{(r+1)(r+1)!\Gamma(r+\frac{1}{2})}{r!\sqrt{2r+1}\Gamma(2r+\frac{5}{2})} \|u^{(r+1)}\|_{L^2(I;V)} \\ &\leq 2^{r+\frac{1}{2}} \frac{(r+1)^2\Gamma(r+\frac{1}{2})}{\sqrt{2r+1}\Gamma(2r+\frac{5}{2})} \|u^{(r+1)}\|_{L^2(I;V)}.\end{aligned}$$

Now, we simplify the term in the right hand side to obtain

$$\frac{\Gamma(r+1/2)}{\Gamma(2r+5/2)} = \frac{(2r-1)\times(2r-3)\times ...\times 5\times 3\times 2^{2r+2}}{(4r+3)\times(4r+2)\times ...\times(2r+1)\times(2r-1)\times ...\times 5\times 3\times 2^r}.$$

Therefore,

$$\begin{aligned}\|(P^r u)(1) - u(1)\|_V &\leq 2^{2r+5/2} \frac{r+1}{(2r+1)^{3/2} r!} \prod_{i=1}^{r+1} \frac{i}{2(r+i)+1} \|u^{(r+1)}\|_{L^2(I;V)} \\ &\leq \frac{\sqrt{2}}{(2r+1)^{1/2} r!} \prod_{i=1}^{r+1} \frac{4i}{2(r+i)+1} \|u^{(r+1)}\|_{L^2(I;V)} \\ &\leq \frac{\sqrt{2}}{2^{\frac{2r}{3}+16}(2r+1)^{3/2} r!} \|u^{(r+1)}\|_{L^2(I;V)}.\end{aligned} \tag{3.15}$$

Now, using (3.15) and (3.12) in Lemma (3.3), we obtain the estimation for Π^r. □

3.3.2 *A priori* error estimate

Proposition 3.3 *Let u be the exact solution of (1.1) − (1.2) and u_k the semi-discrete solution of the Dg (3.4) in $\mathcal{V}_k^r$. Let $\mathcal{I}u \in \mathcal{V}_k^r$ be the interpolate of u which is defined on each time interval I_n by $\mathcal{I}u|_{I_n} = \Pi_{I_n}^r(u|_{I_n})$.*
Then there holds

$$\|u - u_k\|_{L^2(I;V)} \leq C\|u - \mathcal{I}u\|_{L^2(I;V)}.$$

The constant C is in particular independent of T.

Proof. : By Lemma 3.1 we have, for all $v_k,\ w_k \in \mathcal{V}_k^r$,

$$\begin{aligned}a(v_k - w_k, v_k - w_k) &= \int_J \|\nabla(v_k - w_k)\|_H^2 dt + \frac{1}{2}\|(v_k - w_k)_0^+\|_H^2 + \frac{\tau}{2}\|\nabla(v_k - w_k)_0^+\|_H^2 \\ &+ \frac{1}{2}\sum_{n=1}^{N-1} \|[v_k - w_k]_n\|_H^2 + \frac{\tau}{2}\sum_{n=1}^{N-1} \|[\nabla(v_k - w_k)]_n\|_H^2.\end{aligned} \tag{3.16}$$

Hence, we get

$$\int_J \|u_k - \mathcal{I}u\|_V^2 dt \leq a(u_k - \mathcal{I}u, u_k - \mathcal{I}u) \leq |a(u - \mathcal{I}u, u_k - \mathcal{I}u)|$$

where we have used, in the last step, Proposition 3.1. Writing Θ for $u_k - \mathcal{I}u$, we get with lemma 3.1 and the definition of $\mathcal{I}$ that

$$\begin{aligned}
\int_J \|u_k - \mathcal{I}u\|_V^2 dt &\leq \int_J \{-(u - \mathcal{I}u, \Theta')_H + (\nabla(u - \mathcal{I}u), \nabla(\Theta - \tau\Theta'))_H\} dt \\
&+ \sum_{n=1}^{N-1} ((u - \mathcal{I}u)_n^-, [\Theta]_n)_H + ((u - \mathcal{I}u)_N^-, w_{k,N}^-)_H \\
&+ \tau \sum_{n=1}^{N-1} (\nabla(u - \mathcal{I}u)_n^-, [\nabla\Theta]_n)_H + (\nabla(u - \mathcal{I}u)_N^-, \nabla\Theta_N^-)_H \qquad (3.17) \\
&= \int_J |(\nabla(u - \mathcal{I}u), \nabla\Theta)_H\}| dt \\
&\leq \int_J \|u - \mathcal{I}u\|_V \|\Theta\|_V dt. \qquad (3.18)
\end{aligned}$$

We conclude now with the Cauchy-Schwarz inequality that

$$\int_J \|u_k - \mathcal{I}u\|_V^2 dt \leq C \int_J \|u - \mathcal{I}u\|_V^2 dt.$$

Using the triangle inequality, we get

$$\|u - u_k\|_{L^2(I;V)} \leq 2\|u - \mathcal{I}u\|_{L^2(I;V)}.$$

□

Therefore, Proposition 3.3 and Theorem 3.1 give error estimates for the Dg (3.4) which are valid if the exact solution is at least in $H^1(I;V)$.

Theorem 3.2 *Let u be the exact solution of (3.1)-(3.2) and u_k the semi-discrete solution of the Dg (3.4). Assume that $u|_{I_n} \in H^{r_n+1}(I_n;V)$ for $0 \leq n \leq N$. Then we have*

$$\|u - u_k\|_{L^2(I_n;V)}^2 \leq \sum_{n=1}^{N} \left(\frac{k_n}{2}\right)^{2(r_n+1)} \frac{1}{r_n^4 (r_n)!^2} \|u^{(r_n+1)}\|_{L^2(I_n;V)}^2. \qquad (3.19)$$

3.4 Discretization in time and space

The Dg reduces the pseudoparabolic equation (3.1) in each time step I_n to a coupled elliptic system of $r+1$ equations. In order to obtain a fully discrete solution, this system has to be solved numerically. Note that for large r, this is very costly, the system has to be decoupled.

3.4.1 The spatial problems

On a generic time step $I = (t_0, t_1)$ with length $k = t_1 - t_0 > 0$ and approximation order r, the Dg semi-approximation u_k is found by solving the problem in (3.6). The right-hand side $f(t)$ and the initial condition u_{init} are the known data on the time step.
Let $\{\hat{\varphi}_i\}_{i=0}^r$ and $\{\hat{\psi}_i\}_{i=0}^r$ the two bases of the reference polynomial space $\mathcal{P}^r((-1,1))$, chosen as normalized Legendre polynomials. These bases define transported variants $\{\varphi_i\}_{i=0}^r$ and $\{\psi_i\}_{i=0}^r$ on $\mathcal{P}^r((t_0,t_1))$ given by $\varphi_i \circ F(\hat{t}) = \hat{\varphi}_i(\hat{t})$ and $\psi_i \circ F(\hat{t}) = \hat{\psi}_i(\hat{t})$, where F is the transformation $t = F(\hat{t}) = \frac{1}{2}(t_0 + t_1 + k\hat{t})$ from $(-1,1)$ onto (t_0,t_1).
Now if we set $u_k = \sum_{j=0}^r u_{k,j}\varphi_j$ and $v_k = \sum_{i=0}^r v_{k,i}\psi_i$ with coefficients $u_{k,j}, v_{k,i} \in V$, Problem (3.6) is then equivalent to the elliptic system, given in a variational sense:
Find $\{u_{k,j}\}_{j=0}^r \subset V$ such that for all $\{v_{k,i}\}_{i=0}^r \subset V$

$$
\begin{aligned}
(3.20) \qquad \sum_{i,j=0}^{r} \{ \big[\int_I \varphi_j' \psi_i dt + \varphi_j^+(t_0)\psi_i^+(t_0) \big] \big((u_{k,j}, v_{k,i}) + \tau(\nabla u_{k,j}, \nabla v_{k,i}) \big) \\
+ \big(\int_I \varphi_j \psi_i dt \big) (\nabla u_{k,j}, \nabla v_{k,i}) \} \\
= \sum_{i=0}^{r} \{ \big(\int_I f\psi_i dt, v_{k,i} \big) + \big((u_{init}, v_{k,i}) + \tau(\nabla u_{init}, \nabla v_{k,i}) \big) \psi_i^+(t_0) \}.
\end{aligned}
$$

We introduce the matrices

$$
\hat{A}_{ij} := \int_{-1}^{1} \hat{\varphi}_j' \hat{\psi}_i d\hat{t} + \hat{\varphi}_j^+(-1)\hat{\psi}_i^+(-1), \quad \hat{B}_{ij} := \int_{-1}^{1} \hat{\varphi}_j \hat{\psi}_i d\hat{t},
$$

which are expressed in terms of the bases $\{\hat{\varphi}_i\}$ and $\{\hat{\psi}_i\}$ on $(-1,1)$ and which are therefore independent of k. Then, (3.20) is to find $\{u_{k,j}\}_{j=0}^r \subset V$ such that for all $\{v_{k,i}\}_{i=0}^r \subset V$

(3.21)
$$
\sum_{i,j=0}^{r} \hat{A}_{ij} \big((u_{k,j}, v_{k,i}) + \tau(\nabla u_{k,j}, \nabla v_{k,i}) \big) + \frac{k}{2} \hat{B}_{ij} (\nabla u_{k,j}, \nabla v_{k,i}) = \sum_{i=0}^{r} \frac{k}{2} (\hat{f}_i^1, v_{k,i}) + (\hat{f}_i^2, v_{k,i}),
$$

where the right hand sides $\hat{f}_i^1$ and $\hat{f}_i^2$ are defined by

$$
\hat{f}_i^1(v) = \big(\int_{-1}^{1} [f \circ F] \hat{\psi}_i d\hat{t}, v \big), \quad \hat{f}_i^2(v) = \big((u_{init}, v) + \tau(\nabla u_{init}, \nabla v) \big) \hat{\psi}_i^+(-1).
$$

To obtain a fully discrete approximation of (3.1), (3.2) the system (3.21) has to be solved numerically by a Finite Element Method. if $\{u_{k,j}^h\}$ is a FE solution of (3.21) in $V_h \subset V$, then $u_k^h = \sum_{j=0}^r u_{k,j}^h \varphi_j$ approximates $u_k = \sum_{j=0}^r u_{k,j}\varphi_j$ on the time step I. We get for the

error

$$\|u_k - u_k^h\|^2_{L^2(I;V)} = \int_I \| \sum_{j=0}^{r} (u_{k,j} - u_{k,j}^h)\varphi_j \|^2_V dt$$

$$= \sum_{j=0}^{r} \|u_{k,j} - u_{k,j}^h\|^2_V (j+1/2) \int_I L_j^2 dt = \frac{k}{2} \sum_{j=0}^{r} \|u_{k,j} - u_{k,j}^h\|^2_V,$$

where we used the orthogonality properties of the Legendre polynomials.
Thus, we have the following proposition :

Proposition 3.4 *Let u be the exact solution of (3.1), (3.2) on $J = (0,T]$ and let u_k be the time discretization of u. On each interval I_n we develop $u_k|_{I_n}$ into $u_k|_{I_n} = \sum_{j=0}^{r} u_{k,j}^n \varphi_{n,j}$. Let $\{u_{k,j}^{h,n}\}$ be a Finite Element approximation of (3.21) and let u_k^h be the fully discrete solution. Then we have*

$$\|u - u_k^h\|^2_{L^2(J;V)} \leq C\|u - u_k\|^2_{L^2(J;V)} + C \sum_{n=1}^{N} k_n \sum_{j=0}^{r} \|u_{k,j}^n - u_{k,j}^{h,n}\|^2_V. \tag{3.22}$$

The first term in the error estimate (3.22) is the error of the time discretization. The second error contribution stems from the spatial discretization and will be discussed in more details.
For $r = 1$, the matrix $\hat{A}$ is given by

$$\hat{A} = \begin{pmatrix} \frac{1}{2} & \sqrt{\frac{3}{4}} \\ -\sqrt{\frac{3}{4}} & \frac{3}{2} \end{pmatrix}.$$

Let us denote by b the bilinear form defined by

$$b(\vec{u}, \vec{v}) = \sum_{i,j=0}^{r} \hat{A}_{ij}\big((u_j, v_i) + \tau(\nabla u_j, \nabla v_i)\big) + \frac{k}{2}\hat{B}_{ij}(\nabla u_j, \nabla v_i).$$

The bilinear form b is continue and V-coercive, and we have

$$b(\vec{u}, \vec{u}) = \sum_{i=0}^{r} \hat{A}_{ii}\big(\|u_i\|^2 + \tau\|\nabla u_i\|^2\big) + \frac{k}{2}\|\nabla u_i\|^2,$$

and we get

$$b(\vec{u}, \vec{u}) \geq \frac{\tau + k}{2} \sum_{i=0}^{r} \|\nabla u_i\|^2. \tag{3.23}$$

Now using (3.23) and the continuity of the bilinear form b, we get

$$\|\vec{u}_k^n - \vec{u}_k^{h,n}\|_V \leq \frac{C}{\tau + k_n} \inf_{\vec{v} \in V_h} \|\vec{u}_k^n - \vec{v}\|_V. \tag{3.24}$$

3.5 Numerical results

3.5.1 Convergence study

As a test problem, we choose the standard Sobolev equation with $d = 2$; the computational domain being the unit square $\Omega = (0,1)^2$ and the time interval $J = (0, 0.1)$. Here, we choose for the initial data: $u_0(x,y) = \sin(\pi x)\sin(\pi y)$, and for right hand side: $f = -\exp(-t)\sin(\pi x)\sin(\pi y)$, with $\tau = 1$.
u_0 is actually the first eigenfunction of the Laplacian and belongs to $H_0^1(\Omega)$.
The corresponding exact solution $u(t,x,y)$ is smooth in space and time and is given by :

$$u(t,x,y) = \exp(-t)\sin(\pi x)\sin(\pi y).$$

Since the solution is smooth, we would investigate the performance of the Dg by using a spatial discretization of known better order than the temporal one. In order to test the convergence on k, we choose a fixed time approximation order for all time steps of the partition $\mathcal{M}$. Convergence is then achieved by refining the time partition $\mathcal{M}$, *i.e* by increasing the number of time steps in the interval $J = (0, 0.1)$. We present in the figure below (see Figure 3.1) the convergence rate for the previous problem with $r = 0, 1$. In space discretization, we use the standard P_1 finite element method.

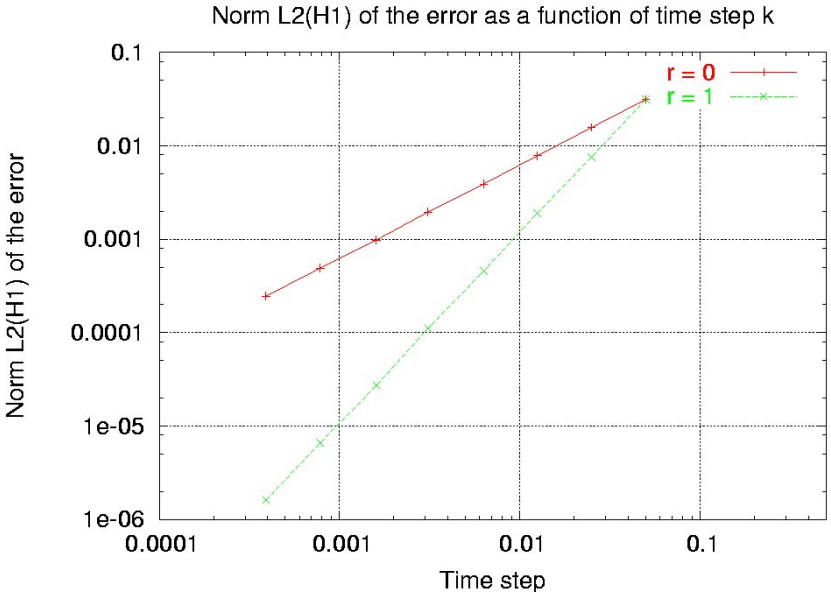


Figure 3.1: Convergence rate for previous problem, h-Dg

This figure shows the convergence of Dg scheme as a function of the time step. A convergence rate equal to $r + 1$ is observed when the approximation order with respect to time is r.

Chapter 4

Space-Time DgFem discretization for the stratigraphic model

4.1 Introduction

In this chapter, we are interested in the discretization of the stratigraphic problem studied in Chapter 2. We use the notations and the preliminary results given in Chapter 1. For the sake of simplicity, we shall assume in the sequel that E is a non-negative constant. Although in more realistic models E depends on other variables such as the bathymetry b, the difference between the sediment thickness and the sea level. The purpose of this chapter is to introduce a numerical scheme that implicitly contains the constraint (2). This is proved by using the lowest-order DgFem(0) with a particular choice of the flux at the interface between two meshes elements. It is well known that high order DgFem(p) do not respect the maximum principle, our aim is to propose a p-adaptive algorithm that combines DgFem(0) and DgFem(p), $p \geq 1$ in order to get more accuracy while still verifying the constraint.
The outline of this chapter is organized as follow: In the following section, the semi-discretized problem using Dg(0) scheme is presented. In section 4.3, we introduce the discontinuous Galerkin scheme for the semi-discretized problem, we prove the existence and the uniqueness of the discrete solution, then we prove that the lowest-order DgFem scheme satisfies implicitly the constraint. In section 4.4, the Newton algorithm for the numerical resolution of the system of nonlinear equations is presented. The section 4.5 concerns the numerical results. In section 4.6, a p-adaptive algorithm applied to the model problem is presented.

4.2 Time Dg discretization

We consider a uniform partition of $(0,T)$ into N subintervals $I_k = [t_k, t_{k+1}[$, $k = 0, ..., N-1$. We denote by $\Delta t = t_{k+1} - t_k$ the time step. To introduce the Dg time scheme, we rewrite the model problem in the following form: Find u and w in $H_0^1(\Omega)$ such that

$$\left\{\begin{array}{r} w - \partial_t u = E \text{ in } \Omega \times (0,T), \\ w - \mathrm{div}[a(w)\nabla u] - \tau \Delta w = f + E \text{ in } \Omega \times (0,T), \\ u(0,x) = u_0 \text{ in } \Omega. \end{array}\right. \tag{4.1}$$

The discontinuous Galerkin approximation for the problem (4.1) reads: Find $u_{\Delta t}$ and $w_{\Delta t}$ in $\mathcal{V}_k^r$ such that for all χ, ψ in $\mathcal{V}_k^r$

$$\left\{\begin{array}{l} \displaystyle\int_0^T \int_\Omega w_{\Delta t}\, \chi \, dx\, dt - \int_0^T \int_\Omega \partial_t u_{\Delta t}\, \chi \, dx\, dt - \sum_{k=1}^{N-1} \int_\Omega [u_{\Delta t}]_k \chi_k^+ \, dx - \int_\Omega u_0^+ \chi_0^+ \, dx \\ \qquad\qquad = \displaystyle\int_\Omega u_0 \chi_0^+ \, dx + \int_0^T \int_\Omega E_{\Delta t}\, \chi \, dx\, dt, \\ \displaystyle\int_0^T \int_\Omega w_{\Delta t}\, \psi \, dx\, dt + \int_0^T \int_\Omega a(w_{\Delta t}) \nabla u_{\Delta t} \cdot \nabla \psi \, dx\, dt + \tau \int_0^T \int_\Omega \nabla u_{\Delta t} \cdot \nabla \psi \, dx\, dt \\ \qquad\qquad = \displaystyle\int_0^T \int_\Omega (f_{\Delta t} + E_{\Delta t})\, \psi \, dx\, dt. \end{array}\right.$$

Since the functions χ and ψ are not required to be continuous at time t_k, one may assume that χ and ψ vanish outside I_k. Therefore, the system reduces to: for each I_k with $k \leq N-1$, determine $u_{\Delta t}$ and $w_{\Delta t}$ in $\mathcal{V}_k^r$ such that

$$\begin{array}{l} \displaystyle\int_{I_k} \int_\Omega w_{\Delta t}\, \chi \, dx\, dt - \int_{I_k} \int_\Omega \partial_t u_{\Delta t}\, \chi \, dx\, dt - \int_\Omega u_{\Delta t,k}^+ \chi_k^+ \, dx = - \int_\Omega u_{\Delta t,k}^- \chi_k^+ \, dx \\ \qquad\qquad + \displaystyle\int_{I_k} \int_\Omega E_{\Delta t}\, \chi \, dx\, dt, \\ \displaystyle\int_{I_k} \int_\Omega w_{\Delta t}\, \psi \, dx\, dt + \int_{I_k} \int_\Omega a(w_{\Delta t}) \nabla u_{\Delta t} \cdot \nabla \psi \, dx\, dt + \tau \int_{I_k} \int_\Omega \nabla u_{\Delta t} \cdot \nabla \psi \, dx\, dt \\ \qquad\qquad = \displaystyle\int_{I_k} \int_\Omega (f_{\Delta t} + E_{\Delta t})\, \psi \, dx\, dt. \end{array}$$

For the sake of simplicity, we consider in the sequel the case $r = 0$, *i.e.*, when the approximating functions are piecewise constant in time. Then $\partial_t u_{\Delta t} \equiv 0$ and $u_{\Delta t}(t) = u^{k+1} = u_{\Delta t,k}^+$, $w_{\Delta t}(t) = w^{k+1} = w_{\Delta t,k}^+$, and the method reduces to the implicit Euler method: Find u^{k+1}

and w^{k+1} in $H_0^1(\Omega)$ such that

$$\begin{aligned}\int_\Omega w^{k+1}\,\chi\,dx &= \frac{1}{\Delta t}\int_\Omega (u^{k+1}-u^k)\chi\,dx + \int_\Omega E^{k+1}\,\chi\,dx,\ \forall\,\chi\in L^2(\Omega),\\ \int_\Omega w^{k+1}\,\psi\,dx + \int_\Omega a(w^{k+1})\nabla u^{k+1}\cdot\nabla\psi\,dx &+ \tau\int_\Omega \nabla w^{k+1}\cdot\nabla\psi\,dx\\ &= \int_\Omega (f^{k+1}+E^{k+1})\,\psi\,dx,\ \forall\,\psi\in H_0^1(\Omega),\end{aligned}\tag{4.2}$$

where f^{k+1} and E^{k+1} are respectively the average of $f_{\Delta t}$ and $E_{\Delta t}$ over I_k. The first equation in (4.2) gives that $w^{k+1}=\frac{1}{\Delta t}(u^{k+1}-u^k)$. Replacing w^{k+1} by its values in the second equation of (4.2) and introducing the function $b(w):=a(\frac{w-u^k}{\Delta t}+E^{k+1})$, we obtain the equation to be solved in each time step: Find $u^{k+1}\in H_0^1(\Omega)$ such that for all v in $H_0^1(\Omega)$

$$\begin{cases}\dfrac{1}{\Delta t}\displaystyle\int_\Omega u^{k+1}v\,dx + \int_\Omega (b(u^{k+1})+\frac{\tau}{\Delta t})\nabla u^{k+1}\cdot\nabla v\,dx = \int_\Omega f^{k+1}v\,dx\\ \qquad\qquad +\dfrac{1}{\Delta t}\displaystyle\int_\Omega u^k v\,dx + \frac{\tau}{\Delta t}\int_\Omega \nabla u^k\cdot\nabla v\,dx,\\ \qquad\qquad u^0=u_0\ \text{ in }\Omega.\end{cases}\tag{4.3}$$

4.3 Space DgFem discretization

4.3.1 DgFem formulation

Our starting point is the time discrete problem (4.3). Let us introduce the following bilinear form for given $\rho_h\in V_h^p$:

(4.4)

$$\begin{aligned}A(\rho_h)(u_h,v_h) = \frac{1}{\Delta t}\int_\Omega u_h v_h\,dx + \sum_{K\in\mathcal{K}_h}\int_K (b(\rho_h)+\frac{\tau}{\Delta t})\nabla u_h\cdot\nabla v_h\,dx + \sum_{S\in\mathcal{S}_h}\frac{\gamma_0}{h_S}\int_S \gamma_S[u_h][v_h]\,ds\\ -\sum_{S\in\mathcal{S}_h}\int_S \{\frac{\partial u_h}{\partial n_S}\}_{S,b}[v_h]_S\,ds - \sum_{S\in\mathcal{S}_h}\int_S [u_h]_S\{\frac{\partial v_h}{\partial n_S}\}_{S,b}\,ds\\ -\sum_{S\in\mathcal{S}_h}\int_S \{\frac{\partial u_h}{\partial n_S}\}_{S,\tau/\Delta t}[v_h]_S\,ds - \sum_{S\in\mathcal{S}_h}\int_S [u_h]_S\{\frac{\partial v_h}{\partial n_S}\}_{S,\tau/\Delta t}\,ds,\end{aligned}$$

where $\gamma_S=\frac{\tau}{\Delta t}+\frac{2b^+b^-}{b^++b^-}$ and $\gamma_0>0$.
Defining the linear form

$$\begin{aligned}L^k(v_h) := \frac{1}{\Delta t}\int_\Omega u_h^k v_h\,dx + \sum_{K\in\mathcal{K}_h}\int_K \frac{\tau}{\Delta t}\nabla u_h^k\cdot\nabla v_h\,dx + \sum_{S\in\mathcal{S}_h}\frac{\gamma_0\tau}{h_S\Delta t}\int_S [u_h^k][v_h]\,ds\\ -\sum_{S\in\mathcal{S}_h}\int_S \{\frac{\partial u_h^k}{\partial n_S}\}_{S,\tau/\Delta t}[v_h]_S\,ds - \sum_{S\in\mathcal{S}_h}\int_S [u_h^k]_S\{\frac{\partial v_h}{\partial n_S}\}_{S,\tau/\Delta t}\,ds + \int_\Omega f^{k+1}v_h\,dx.\end{aligned}\tag{4.5}$$

The discrete approximation to (4.3) reads: Find $u_h^{k+1} \in V_h^p$ such that for all $v_h \in V_h^p$

$$A(u_h^{k+1})(u_h^{k+1}, v_h) = L^k(v_h). \tag{4.6}$$

A similar discretization has been used in [21] for a stationary convection-diffusion problem with discontinuous diffusion coefficients.
The stability of (4.6) is expressed by the following result.

Lemma 4.1 *Let for $\gamma_0 > 0$ and ρ a positive bounded piecewise continuous function*

$$|||u_h|||_{h,\rho} := \sqrt{\frac{1}{\Delta t}\|u_h\|^2 + \frac{\tau}{\Delta t}\sum_K \|\nabla u_h\|_K^2 + \frac{\tau}{\Delta t}\sum_{S\in\mathcal{S}_h}\frac{\gamma_0}{h_S}\|[u_h]_S\|_S^2}. \tag{4.7}$$

Then $|||\cdot|||_{h,\rho}$ is a norm on V_h^p and there exists $\gamma > 0$, independent of h, such that

$$A(\rho)(u_h, u_h) \geq \gamma\, |||u_h|||_{h,\rho}^2. \tag{4.8}$$

This result is responsible for the well-posedness of the discrete problem. As usual, it yields, in connection with consistency, to convergence.

4.3.2 Existence and uniqueness

On the discrete level, we obtain not only existence of a solution, but also uniqueness.

Proposition 4.1 *Suppose for $p > 0$ that γ_0 is sufficiently large. The problem (4.6) has at least one solution. In addition, if a is Lipschitz-continuous and τ sufficiently large, the solution is unique too.*

Proof. The proof of the existence of a solution to the problem (4.6) is based on the Brower's fixed point theorem. We consider an application $\psi : V_h^p \to V_h^p$ such that $\psi(\rho) = u_{h,\rho}^{k+1}$, where $u_{h,\rho}^{k+1}$ is the solution of the following linearized problem

$$\begin{cases} \text{Find } u_{h,\rho}^{k+1} \in V_h^p, \text{ such that} \\ \\ A(\rho)(u_{h,\rho}^{k+1}, v_h) = L^k(v_h) \ \ \forall\, v_h \in V_h^p. \end{cases} \tag{4.9}$$

Existence and uniqueness of $u_{h,\rho}^{k+1}$ follow from the Lax-Milgram theorem in conjunction with Lemma 4.1.
Let $\rho_n \to \rho$, choosing $v_h = u_{h,\rho}^{k+1}$, we show that the sequence $(u_{h,\rho_n})_n$ is bounded in V_h^p. We may extract a subsequence $u_{h,\rho_{n'}}$ such that $u_{h,\rho_{n'}} \to \xi$ as $n' \to \infty$. The sequence $(u_{h,\rho_{n'}})_{n'}$ satisfies the equation (4.9), thus by passing to the limits ($n' \to \infty$), we prove that ξ satisfies

the equation (4.9), we conclude that $\xi = \psi(\rho)$ (uniqueness of the solution). The whole sequence converges and ψ is continuous. Thus the assertion follows from the Brower fixed point theorem.

We now turn to the uniqueness proof. Assume that the problem (4.6) has two solutions u_h^1 and u_h^2. We set $w_h = u_h^1 - u_h^2$. We have for all $v_h \in V_h^p$

$$A(u_h^1)(w_h, v_h) = A(u_h^1)(u_h^2, v_h) - A(u_h^2)(u_h^2, v_h) =: R(v_h). \tag{4.10}$$

In particular for $v_h = w_h$, we have

$$C_1 |||w_h|||_{h,u_h^1}^2 \leq |A(u_h^1)(u_h^2, w_h) - A(u_h^2)(u_h^2, w_h)|.$$

The right hand-side is:

$$R(w_h) = \sum_{K\in\mathcal{K}_h} \int_K (b_1 - b_2)\, \nabla u_h^2 \cdot \nabla w_h \, dx + \sum_{S\in\mathcal{S}_h} \frac{\gamma_0}{h_S} \int_S \left(\gamma_S^1 - \gamma_S^2\right) [u_h^2][w_h] \, ds$$
$$- \sum_{S\in\mathcal{S}_h} \int_S \{\frac{\partial u_h^2}{\partial n_S}\}_{S,(b_1-b_2)} [w_h]_S \, ds - \sum_{S\in\mathcal{S}_h} \int_S [u_h^2]_S \{\frac{\partial w_h}{\partial n_S}\}_{S,(b_1-b_2)} \, ds,$$

where $b_i = b(u_h^i)$ and $\gamma_S^i = \frac{\tau}{\Delta t} + \frac{2b_i^+ b_i^-}{b_i^+ + b_i^-}$. Let L be the Lipschitz-constant of a. Then we have

$$\begin{aligned} |b(u_h^1) - b(u_h^2)| &= |a(\frac{u_h^1 - u^k}{\Delta t} + E^{k+1}) - a(\frac{u_h^2 - u^k}{\Delta t} + E^{k+1}) \\ &\leq L|\frac{u_h^1 - u^k}{\Delta t} + E^{k+1} - (\frac{u_h^2 - u^k}{\Delta t} + E^{k+1})| \leq \frac{L}{\Delta t}|u_h^1 - u_h^2| \\ &= \frac{L}{\Delta t}|w_h|. \end{aligned}$$

The standard inverse estimate $\|\nabla v_h\|_{K,\infty} \leq Ch^{-1/2}\|\nabla v_h\|_K$ gives $\|\nabla u_h^2\|_{\Omega,\infty} \leq Ch^{-1/2}|||u_h^2|||_{h,u_h^2} \leq Ch^{-1/2}$ with a data-dependent constant since u_h^2 solves (4.6). We therefore obtain with Lemma 1.24

$$\begin{aligned} \sum_{K\in\mathcal{K}_h} \int_K (b_1 - b_2)\, \nabla u_h^2 \cdot \nabla w_h \, dx &\leq \frac{L}{\Delta t} \sum_{K\in\mathcal{K}_h} \int_K |w_h| \nabla u_h^2 \cdot \nabla w_h \, dx \\ &\leq \frac{L\|\nabla u_h^2\|_{\infty,\Omega}}{\Delta t} \sum_{K\in\mathcal{K}_h} \int_K \|\nabla w_h\|_K^2 \\ &\leq \frac{CLh^{-1/2}}{\Delta t} \sum_{K\in\mathcal{K}_h} \int_K \|\nabla w_h\|_K^2 \\ &\leq \frac{C_1\tau}{2\Delta t} \sum_{K\in\mathcal{K}_h} \int_K \|\nabla w_h\|_K^2 \leq \frac{C_1}{2}|||w_h|||_{h,u_h^1}^2, \end{aligned}$$

provided $\tau \geq 2Ch^{-1/2}C_1^{-1}L\|\nabla u_h^2\|_{\infty,\Omega}$. Now the term can be absorbed by the left-hand side of (4.10).
The other terms are treated in similar way, and we find $w_h = 0$. □

4.3.3 DgFem(0)

In the lowest-order case, the DgFem scheme reduces to a cell-centered finite volume method, and the well-known theoretical results can be used, see for example[27].
In this case the bilinear form A reduces to

$$A(\rho_h)(u_h, v_h) = \frac{1}{\Delta t}\int_\Omega u_h v_h \, dx + \sum_{S\in\mathcal{S}_h} \frac{1}{h_S}\int_S \gamma_S [u_h][v_h] \, ds. \tag{4.11}$$

It is clear that the term $\frac{\gamma_S}{h_S}[u_h]$ is an approximation of the normal flux, and this implies that $\frac{\gamma_S}{h_S}$ has to be chosen as the distance of the orthocenters of the neighboring triangles of S, in case of an interior side. A similar argument is used for the boundary sides.
The error analysis of [27] assumes that the triangulation h satisfies the following condition: there exists $\eta > 0$ such that for any α angle of an element $K \in \mathcal{K}_h$ we have

$$\eta < \alpha < \frac{\pi}{2}. \tag{4.12}$$

4.3.4 Discrete maximum principle

In analogy to the continuous formulation, we prove that the DgFem(0) scheme with bilinear form (4.11) satisfies implicitly the constraint $\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1} \geq 0$ a.e. in Ω, for $k = 0, ..., N-1$.

Proposition 4.2 *Let u_h^{k+1} be the solution of the problem (4.6) with $p = 0$. Assume that f^{k+1} and E^{k+1} satisfy $f^{k+1} + E^{k+1} \geq 0$. Then, we have*

$$\frac{u_h^{k+1} - u_h^k}{\Delta t} + E^{k+1} \geq 0, \ \textit{a.e. in}\ \Omega, \ k \geq 0. \tag{4.13}$$

A similar result cannot be expected for $p \geq 1$ since the discretization of the elliptic operator does not lead to an M-matrix in this case. This will be illustrates by the numerical experiments below.
Proof. Let u_h^{k+1} be the solution of the problem (4.6), we have $\forall v_h \in V_h^0$

$$\begin{aligned}\int_\Omega (\frac{u_h^k - u_h^{k-1}}{\Delta t} + E^{k+1}) v_h \, dx + \sum_{S\in\mathcal{S}_h} \frac{1}{h_S}\int_S \frac{2b^+b^-}{b^+ + b^-}[u_h^{k+1}][v_h] \, ds \\ + \tau \sum_{S\in\mathcal{S}_h} \frac{1}{h_S}\int_S [\frac{u_h^{k+1} - u_h^k}{\Delta t} + E^{k+1}][v_h] \, ds \\ = \int_\Omega f^{k+1} v_h \, dx + \int_\Omega E^{k+1} v_h \, dx \ \forall\ v_h \in V_h^0.\end{aligned}$$

In particular for $v_h = (\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1})^- \in V_h^0$, we obtain

$$\begin{aligned}-\|(\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1})^-\|_0^2 + \sum_{S\in\mathcal{S}_h}\frac{1}{h_S}\int_S \frac{2b^+b^-}{b^++b^-}[u_h^{k+1}][(\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1})^-]\,ds \\ +\tau\sum_{S\in\mathcal{S}_h}\frac{1}{h_S}\int_S[\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1}][(\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1})^-]\,ds \\ = \int_\Omega f^{k+1}(\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1})^-\,dx + \int_\Omega E^{k+1}(\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1})^-\,dx.\end{aligned}$$

If on the one hand, $\frac{u_h^{k+1,+}-u_h^{k,+}}{\Delta t} + E^{k+1} \geq 0$ and $\frac{u_h^{k+1,-}-u_h^{k,-}}{\Delta t} + E^{k+1} \geq 0$, we have

$$[v_h] = (\frac{u_h^{k+1,+}-u_h^{k,+}}{\Delta t} + E^{k+1})^- - (\frac{u_h^{k+1,-}-u_h^{k,-}}{\Delta t} + E^{k+1})^- = 0.$$

On the other hand, $\frac{u_h^{k+1,+}-u_h^{k,+}}{\Delta t} + E^{k+1}$ or $\frac{u_h^{k+1,-}-u_h^{k,-}}{\Delta t} + E^{k+1}$ is non-positive and from the definition of the average of b, we obtain

$$\frac{2b^+b^-}{b^++b^-} = 0.$$

We conclude that

$$\sum_{S\in\mathcal{S}_h}\frac{1}{h_S}\int_S \frac{2b^+b^-}{b^++b^-}[u_h^{k+1}][(\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1})^-]\,ds = 0,$$

we obtain

$$\begin{aligned}-\|(\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1})^-\|_0^2 = \int_\Omega f^{k+1}(\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1})^-\,dx + \int_\Omega E^{k+1}(\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1})^-\,dx \\ +\tau\sum_{S\in\mathcal{S}_h}\frac{1}{h_S}\int_S |[(\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1})^-]|^2\,ds.\end{aligned}$$

Since the term in the right hand side is nonnegative

$$(\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1})^- = 0.$$

Thus $\frac{u_h^{k+1}-u_h^k}{\Delta t} + E^{k+1} \geq 0$. □

4.4 Nonlinear solver

The non-linear equations at each time-step are solved by a Newton-type algorithm which solves in each iteration l

$$B(u_h^{k+1,l})(u_h^{k+1,l+1} - u_h^{k+1,l}, v_h) = \alpha \left(L^k(v_h) - A(u_h^{k+1,l})(u_h^{k+1,l}, v_h) \right), \tag{4.14}$$

where

(4.15)

$$\begin{aligned} B(u_h)(w_h, v_h) = A(u_h)(w_h, v_h) + \sum_{K \in \mathcal{K}_h} \int_K a'_\varepsilon(u_h) \nabla w_h \cdot \nabla v_h \, dx + \sum_{S \in \mathcal{S}_h} \frac{\gamma_0}{h_S} \int_S \gamma'_S [w_h][v_h] \, ds \\ - \sum_{S \in \mathcal{S}_h} \int_S \{ \frac{\partial w_h}{\partial n_S} \}_{S, a'_\varepsilon} [v_h]_S \, ds - \sum_{S \in \mathcal{S}_h} \int_S [w_h]_S \{ \frac{\partial v_h}{\partial n_S} \}_{S, a'_\varepsilon} \, ds, \end{aligned}$$

with $\gamma'_S = \partial \gamma_S / \partial u_h$ and α is a positive number.
We denote by $F^k(u_h) = A(u_h)(u_h, v_h) - L^k(v_h)$, it is not necessary that $F^k(u_h^{k+1}) = 0$ has got a solution even if $F^k(u^{k+1}) = 0$ has. In this case the Newton iteration tends to be the minimizer of $\|F^k(u_h^{k+1})\|$.
It is well known that for sufficiently small α

$$\|F^k(u_h^{k+1,l+1})\| < \|F^k(u_h^{k+1,l})\|, \tag{4.16}$$

and

$$d_l = - \left(\frac{\partial F^k(u_h^{k+1,l})}{\partial u_h} \right)^{-1} F^k(u_h^{k+1,l}), \tag{4.17}$$

is a direction descent for $\|F^k(u_h^{k+1})\|$. The Newton iteration is

$$u_h^{k+1,l+1} = u_h^{k+1,l} + \alpha_l d_l, \tag{4.18}$$

where $0 < \alpha_l \leq 1$ is chosen as large as possible in order to allow quadratic convergence.
The Newton method is local, and convergence is assured only when $u_h^{k+1,0}$ is close enough to the solution. In general, the first guess may be outside the region of convergence. To improve convergence from bad initial guesses, a damping strategy is used for choosing α_l. It chooses the largest damping coefficient α out of the sequence $1, 1/2, 1/4, ...$, such that the following inequality holds:

$$\|F^k(u_h^{k+1,l+1})\| < \|F^k(u_h^{k+1,l})\|. \tag{4.19}$$

An important point of this strategy is that when $u_h^{k+1,l}$ approaches the solution, then $\alpha \to 1$ and thus the convergence rate increases.

Closely related to the above problem is the choice of the initial guess $u_h^{k+1,0}$. By default, the solver sets $u_h^{k+1,0}$ and then assembles the DgFem matrices A_k and b_k and computes

$$u_h^{k+1,1} = A_k^{-1} b_k,$$

and the Newton iteration is then started with $u_h^{k+1,1}$, which should be a better guess than U^0. Furthermore, if the equation is linear, then $u_h^{k+1,1}$ is the exact DgFem solution and the solver does not enter the Newton loop.
In general the exact Jacobian

$$J_l = \frac{\partial F^k(u_h^{k+1,l})}{\partial u_h},$$

is not available. A very simple approximation to J_l, which gives a fixed point iteration, is possible as follows. Essentially, for a given $u_h^{k+1,l}$, we compute the DgFem matrices A_k and b_k and we set

$$U^{k+1,l+1} = A_k^{-1} F. \tag{4.20}$$

This is equivalent to approximating the Jacobian J_l with the matrix A_k. Indeed, since $F^k(u_h^{k+1,l}) = A_k u_h^{k+1,l} - b_k$, putting $J_l = A_k$ we have

$$u_h^{k+1,l+1} = u_h^{k+1,l} - J_l^{-1} F^k(u_h^{k+1,l}) = u_h^{k+1,l} - A_k^{-1}(A_k u_h^{k+1,l} - b_k) = A_k^{-1} b_k,$$

In many cases the convergence rate is slow, but the cost of each iteration is cheap. The Newton algorithm is as follows

- Step 1 (Initialization): for a given u_0 initial condition, $Tol > 0$ for residual norm, Δt: time step and triangulation $\mathcal{K}_h$, we compute u_h^k the L^2-projection of u_0.
- Step 2 (Initial guess for Newton algorithm): construct the matrix A_k and the right hand side b_k, compute $u_h^{k,l} = A_k^{-1} b_k$ and $nr = \|F^k(u_h^{k,l})\|$.
- Step 3 (Newton algorithm)
  - Step 3.1: construct the Jacobian matrix J.
  - Step 3.2: solve $J_l * \delta u_h^k = -\alpha F^k(u_h^{k,l})$.
  - Step 3.3: compute $u_h^{k,l+1} = u_h^{k,l} + \delta u_h^k$ and $nrr = \|F^k(u_h^{k,l})\|$.
  - If $(nr < nrr)$, $\alpha = \alpha/2$ and continue with 3.2.
  - If $nrr < Tol$ continue with step 4.
- Step 4: Set $u_h^{k+1} = u_h^{k,l+1}$ and continue with step 2.

4.5 Numerical results

We consider problem (4.3) in the domain $\Omega =]-1,1[^2$ for $0 \leq t \leq T$ with homogeneous Dirichlet condition and a right hand side f equal to 0. The initial condition u_0 is given by

$$u_0(x,y) = -\sin(\pi x)\sin(\pi y),$$

and $a = a_\epsilon$ is given by

$$a_\epsilon(u) = \begin{cases} 0 \text{ if } u < 0, \\ \frac{3u^2}{\epsilon^2}(1 - \frac{2u}{3\epsilon}) \text{ if } 0 \leq u \leq \epsilon, \\ 1 \text{ if } u \geq \epsilon. \end{cases}$$

The meshes are obtained by uniform refined from a coarse mesh h_0, verifying the angle condition required for $p = 0$.
We fix ϵ at 0.1, the maximum erosion rate E is equal to 0.1 and the time step $\Delta t = 0.1$. We then compare the number of iterations needed for the convergence of the algorithm by using DgFem(p), $p = 0, 1, 2$. The algorithm stopped at a tolerance fixed at 10^{-12} (norm of residual) or if the maximum number of iteration fixed at 80 iterations is reached. The result is presented in the table below

τ	1			0.1			0.05		
t	0.1	0.2	0.4	0.1	0.2	0.4	0.1	0.2	0.4
$p = 0$	24	12	8	25	21	17	80	39	39
$p = 1$	22	9	9	38	52	48	46	55	52
$p = 2$	26	7	7	35	40	46	39	47	56

Table 4.1: Number of Newton iterations with respect to τ, t, and p.

The table shows that for τ equal to one, the convergence of Newton algorithm is achieved with a small number of iterations. As indicated by our uniqueness proof, for a small value of τ convergence is very slow.
Now, we fix the parameter τ at one and look for the dependence of the convergence of Newton algorithm with respect to the mesh size h. We represent in the table below the number of iterations as a function of the mesh size h at time $t = 0.2$.

N	Number of iterations		
	$p=0$	$p=1$	$p=2$
896	30	8	7
3584	12	7	7
14336	8	7	6

Table 4.2: Number of Newton iterations with respect to h and p for $\tau = 1$ at time $t = 0.2$.

4.5.1 Convergence study

In this section, we study the convergence of the error in norm L^2 for the problem (4.3) under uniform mesh refinement. We denote u_h^* the reference solution obtained by solving the problem (4.3) using $p = 2$ scheme in a fine mesh with 57344 element, then we compute the norm L^2 of the error as the norm of the difference between u_h and $u_{h_{\text{ref}}}$. We represent in the table below the norm L^2 of error as a function of h at time $t = 0.1$, the time step is fixed at $\Delta t = 0.1$.

Ne	$\|u_h^* - u_h\|_{L^2(\Omega)}$			rate		
	$p=0$	$p=1$	$p=2$	$p=0$	$p=1$	$p=2$
896	$1.51e-1$	$3.57e-2$	$1.48e-2$	–	–	–
3584	$6.95e-2$	$9.47e-3$	$1.48e-3$	1.11	1.91	2.90
14336	$3.32e-2$	$2.42e-3$	$2.25e-4$	1.06	1.96	2.70

Table 4.3: L^2 norm of error with respect to h and p for $\tau = 1$ at time $t = 0.1$.

The table (4.3) shows the convergence of the DgFem scheme with convergence rate approximatively equal to $p + 1$. As indicated by our proof of the uniqueness, the convergence is not guaranteed if the parameter τ is very small.

4.5.2 Numerical simulations

In this section, we present some numerical simulations. We solve the problem (4.3) using $p = 0, 1, 2$. We choose $\epsilon = 0.1$, the maximum erosion rate $E = 0.1$, the time step $\Delta t = 0.1$. Our aim is to test numerically if the discrete constraint $\frac{u_h^{n+1} - u_h^n}{\Delta t} + E^{n+1}$ is satisfied. If we consider the linearized problem, the numerical solution satisfied $\partial_t u_h \leq 0$ in part of domain where the initial condition is convex and $\partial_t u_h \geq 0$ in the rest of domain. Figure 4.1 shows the numerical solution and the constraint at different time using $p = 0, 1, 2$ schemes.

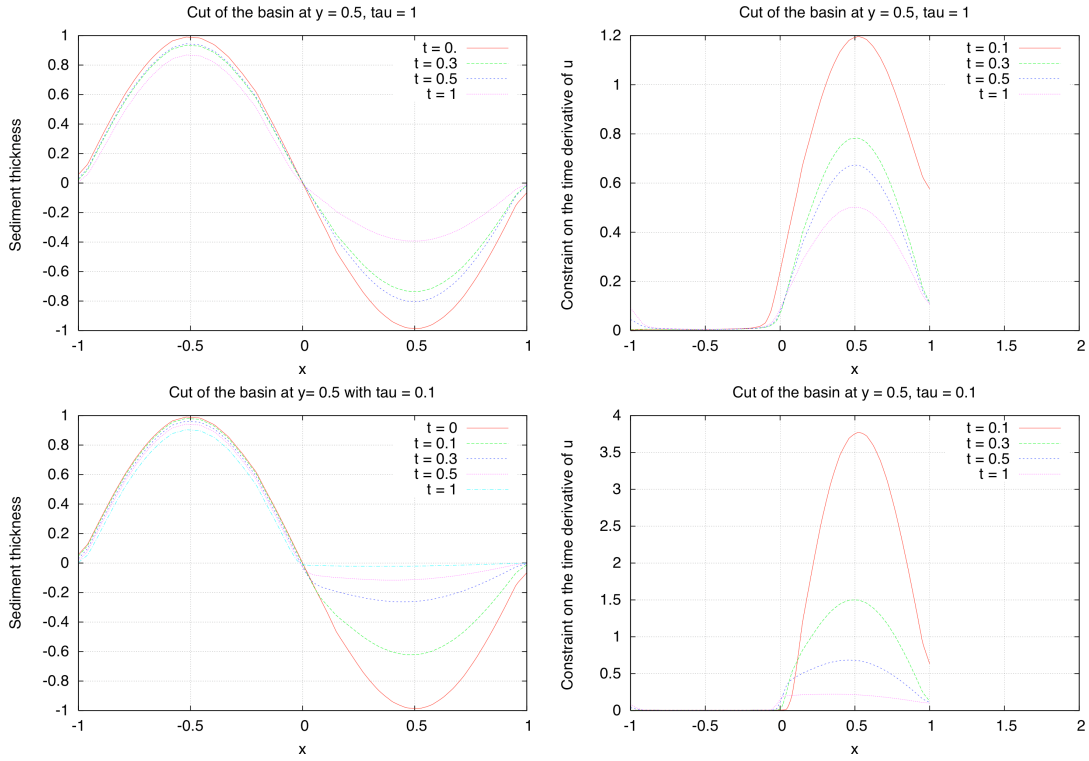


Figure 4.1: Vertical 1D cut at $y = 0.5$ of the numerical solution (left) and the constraint $\partial_t u_h + E$ (right) with $\tau = 1$ (top) and $\tau = 0.1$ (bottom), $p = 0$.

The figures show that the constraint is satisfied in all domain Ω and at each time step. When the constraint is active, we have $\partial_t u_h + E \simeq 0$. This confirm the theoretical results. In the sequel, we consider the same example with $p = 1$, we present in the figure below the numerical solution and the constraint at different time step with different values of τ

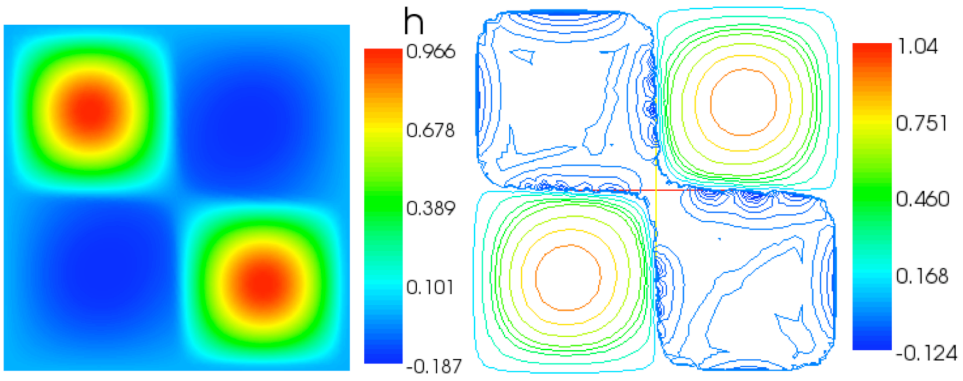


Figure 4.2: Numerical solution and the numerical approximation of the constraint at time t = 0.4, $\tau = 0.1$, $p = 1$.

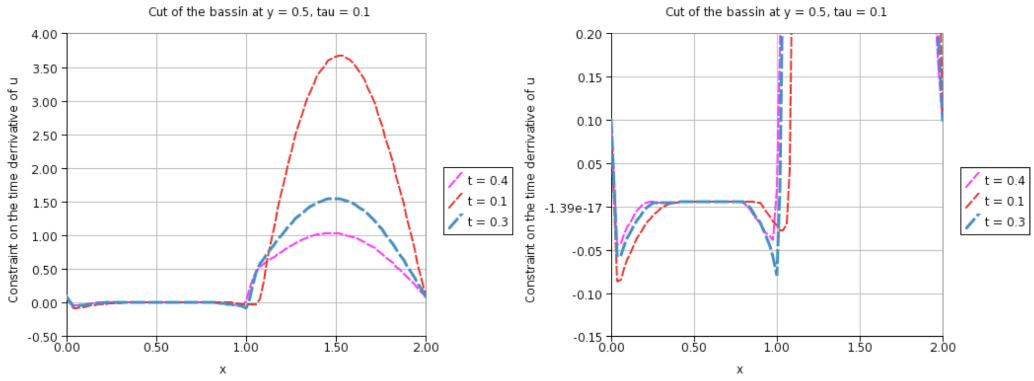


Figure 4.3: Vertical 1D cut at $y = 0.5$ of the constraint $\partial_t u_h + E$ with $\tau = 0.1$ and $p = 1$.

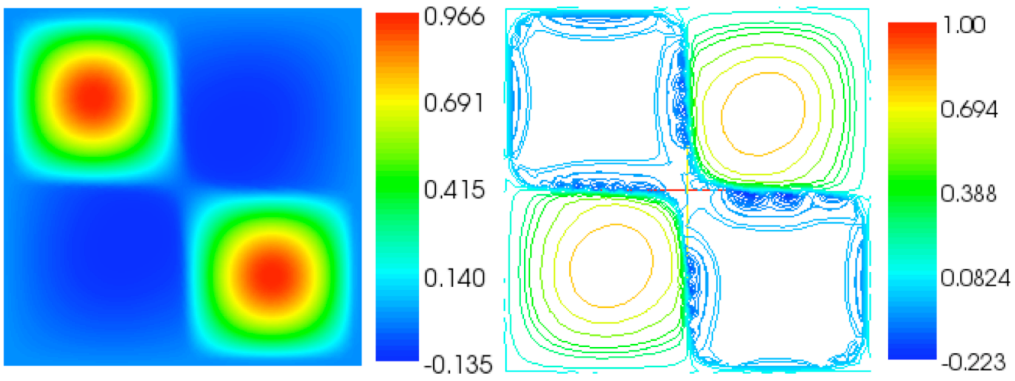


Figure 4.4: Numerical solution and the numerical approximation of constraint at time $t = 0.4$ with $\tau = 0.05$ and $p = 1$.

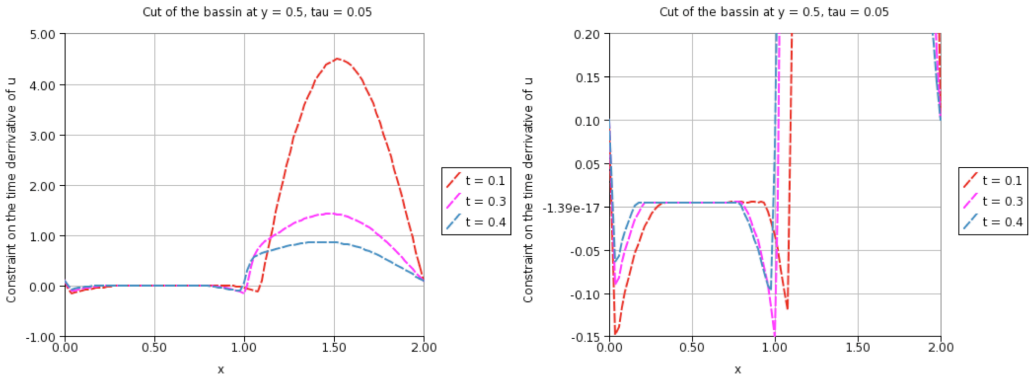


Figure 4.5: Vertical 1D cut at $y = 0.5$ of the constraint $\partial_t u_h + E$ with $\tau = 0.05$ and $p = 1$.

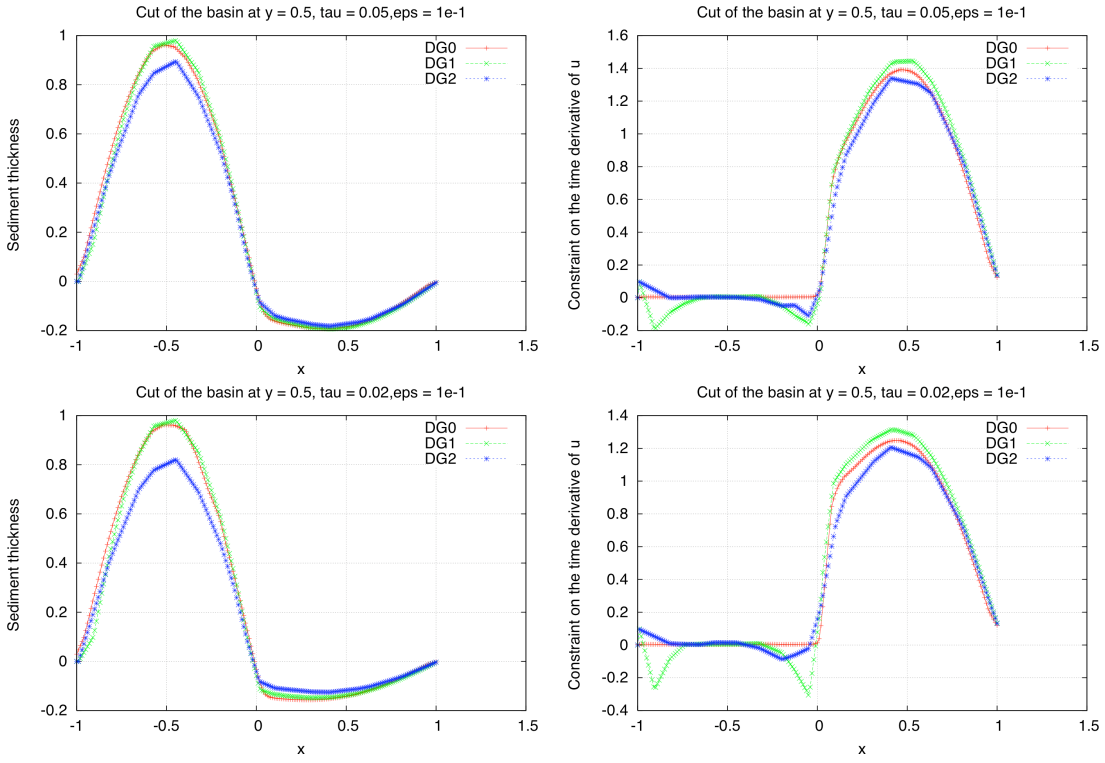


Figure 4.6: Vertical 1D cut at $y = 0.5$ of the numerical solution and the constraint $\partial_t u + E$ with $\tau = 0.05$ (top) and $\tau = 0.02$ (bottom) using DgFem(p) schemes, $p = 0, 1, 2$.

This numerical simulations show the influence of the parameter τ on the constraint, for a small values of τ, the constraint is not satisfied near the interface and the boundary if $p \geq 1$ is used.

In the following numerical examples, we fix the parameter τ to 0.05 and we look for the influence of the discretization parameters h and p on the constraint. The figure 4.7 presents the numerical constraint with respect to h and p at time $t = 0.1$.

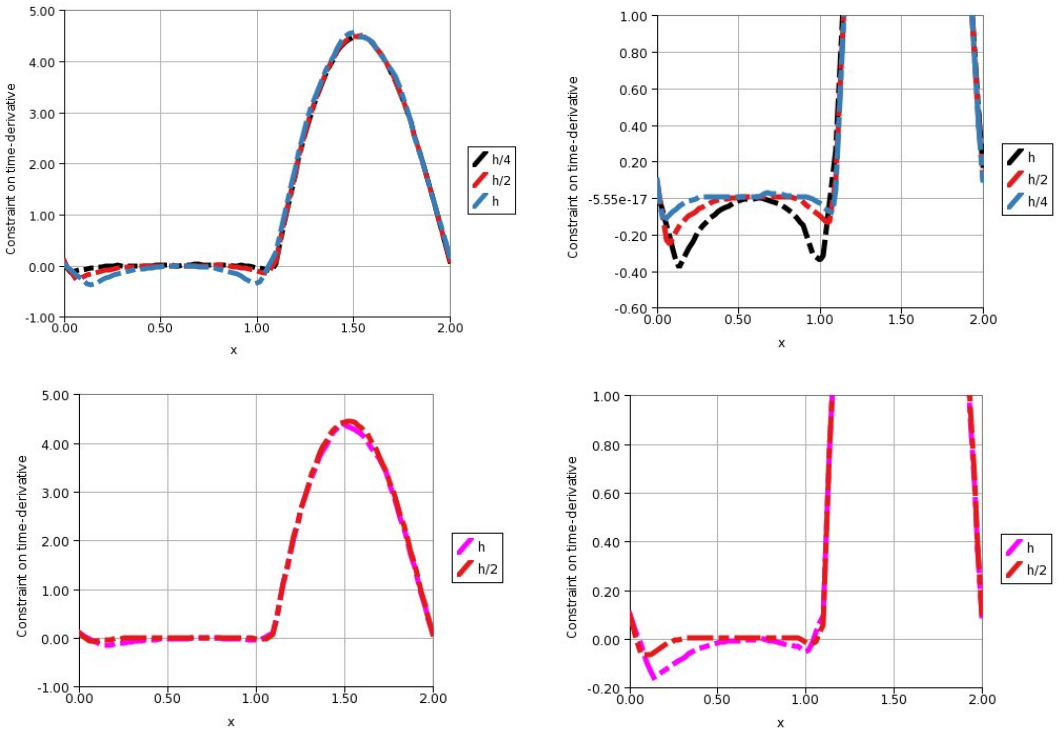


Figure 4.7: Vertical 1D cut at $y = 0.5$ of the numerical constraint $\partial_t u_h + E$ with $p = 1$ (top) and $p = 2$ (bottom), $\tau = 0.05$.

The figure illustrates the influence of the discretization parameters on the constraint. It tends to be satisfied, when we refine the mesh and increasing the approximation order. High order schemes and fine mesh means a big number of degree of freedom while a DgFem(0) gives a good approximation. Our aim is then to reduce to zero the approximation order in the zone where the constraint is not satisfied.

4.6 Adaptive algorithm

4.6.1 Introduction

The numerical results presented in the last section show that using DgFem(0) scheme, the constraint is implicitly satisfied. However, the constraint is not totally satisfies if we use a high order DgFem scheme, some negative values appear near the interface. With a small value of τ, this values became important. In this section an adaptive algorithm that combines the DgFem(0) scheme and high order DgFem scheme is given. The idea of the algorithm is to solve the problem by using a high order scheme, then thanks to an interface

indicator, we reduce to zero the approximation order of the selected triangle.

4.6.2 Adaptive algorithm

For $v_h \in V_h^{\mathbf{P}}$, we define its interpolate $\mathcal{I}_h$ by describing its values at the usual Lagrange interpolation nodes on each mesh element by taking the average of the values of v_h at the node,

$$\mathcal{I}_h^n(v_h)(n) = \frac{1}{|\mathcal{K}_n|} \sum_{K \in \mathcal{K}_n} v_h \backslash K(n), \tag{4.21}$$

where $\mathcal{K}_n$ is the set of mesh element that contains the node n and where $|\mathcal{K}_n|$ is the cardinal of the set. The p-adaptive algorithm is summarized as follow: we repeat the following algorithm until time $t = T$

- Step 1 (Initialization): for a given u_0 initial condition, $Tol > 0$ for residual norm, Δt: time step and triangulation $\mathcal{K}_h$, we compute u_{hp}^k the L^2-projection of u_0 in $V_h^{\mathbf{P}}$.
- Step 2 (Initial guess for Newton algorithm): construct the matrix A and the right hand side F, compute $u_{hp}^{k,l} = A_k^{-1} b_k$ in $V_h^{\mathbf{P}}$ and $nr = \|F^k(u_{hp}^{k,l})\|$.
- Step 3 (Newton algorithm)
  - Step 3.1: construct the Jacobian matrix J_l.
  - Step 3.2: solve $J_l * \delta u_{hp}^{k,l} = -\alpha F^k(u_{hp}^{n,k})$.
  - Step 3.3: compute $u_{hp}^{k,l+1} = u_{hp}^{k,l} + \delta u_{hp}^{k,l}$ in $V_h^{\mathbf{P}}$ and $nrr = \|F^k(u_{hp}^{k,l+1})\|$.
  - Step 3.4: If $(nr < nrr)$, $\alpha = \frac{\alpha}{2}$, continue with step 3.2.
  - Step 3.5: If $nrr < Tol$, continue with step 4.
- Step 4: for each element $K_i,\ i = 1, ..., m_h$, we compute

  $$\eta_i^j = \mathcal{I}_h(\frac{u_{hp}^{k,l+1} - u_{hp}^k}{\Delta t} + E_k)(n_j),\ j = 1, 2, 3.$$

  If $\eta_i^j \geq 0,\ i = 1, ..., m_h,\ j = 1, 2, 3$, set $u_{hp}^{k+1} = u_{hp}^{k,l+1}$ and continue with step 2
- Step 5 (Marking element): for a given $\eta_i^j\ i = 1, ..., m_h,\ j = 1, 2, 3$, we find a set $\mathcal{A}$ subset of $\mathcal{K}$, the set of mesh elements such that for all $i \in \mathcal{A}$:

  $$\eta_i^j \times \eta_i^k < 0,\ j = 1, 2, 3, \text{and } k \equiv (j + 1) \bmod 3.$$

- Step 6 (Refine): for $i \in \mathcal{A}$, we set $p_{K_i} = 0$, we then construct a new finite element space $V_h^{\tilde{\mathbf{P}}}$, we compute the L^2-projection of u_{hp}^k into $V_h^{\tilde{\mathbf{P}}}$ and we continue with step 2.
- Step 7: we set $u_{hp}^{k+1} = u_{hp}^{k,l+1}$, $u_{hp}^k = u_{hp}^{k+1}$ and continue with step 1.

4.6.3 Numerical simulations

In this section, we first consider a mesh with approximation order varying between zero and one see figure 4.8, then, we compare the result with uniform p, $p = 0, 1$. The figure 4.9 represents the numerical approximation of the time-derivative constraint $\partial_t u_h + E$;

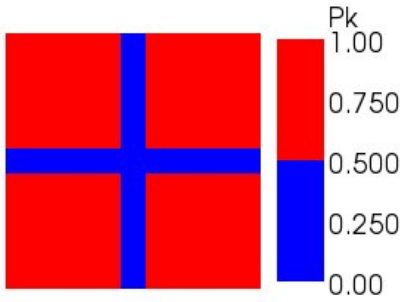


Figure 4.8: Mesh with approximation order varying between 0 and 1.

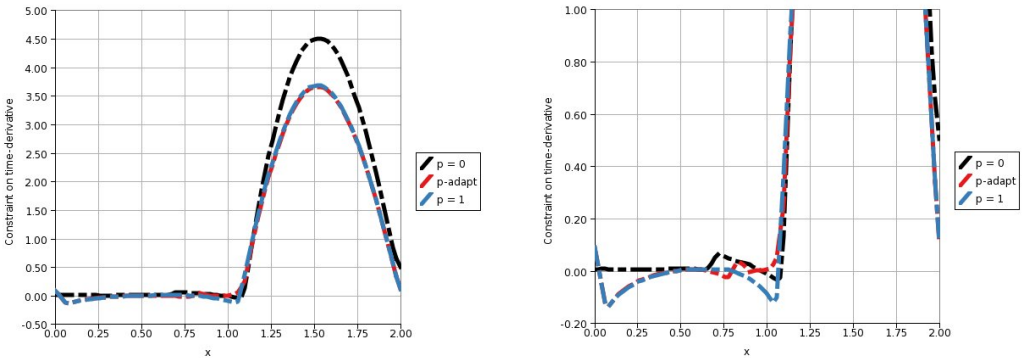


Figure 4.9: Vertical 1D cut at $y = 0.5$ of the numerical approximation of the time-derivative constraint at time $t = 0.1$, $\tau = 0.1$.

In the sequel, some numerical simulations using our p-adaptive algorithm are presented. The maximum erosion rate is fixed to $E = 5.$, the parameter $\tau = 1.$, $\varepsilon = 0.01$, the time step is fixed to $\Delta t = 0.1$, we then repeat the algorithm until the satisfaction of the constraint or a prescribed maximum iteration. The result is represented in the figure below

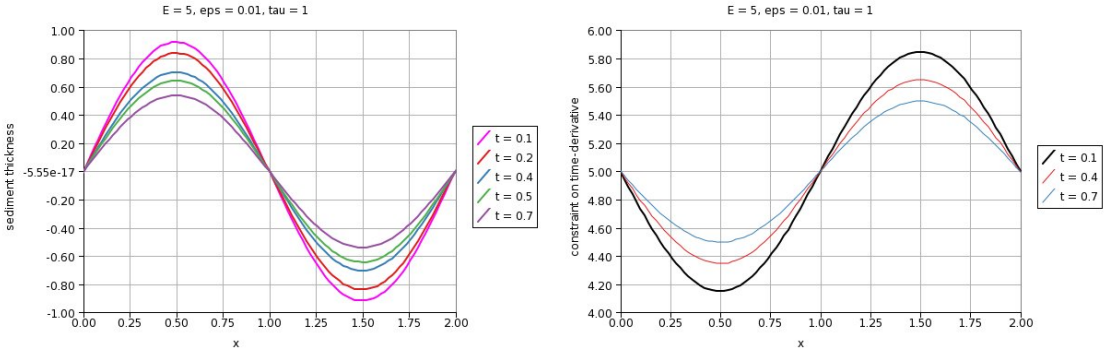


Figure 4.10: Vertical 1D cut at $y = 0.5$ of the numerical solution and the time-derivative constraint of u_h.

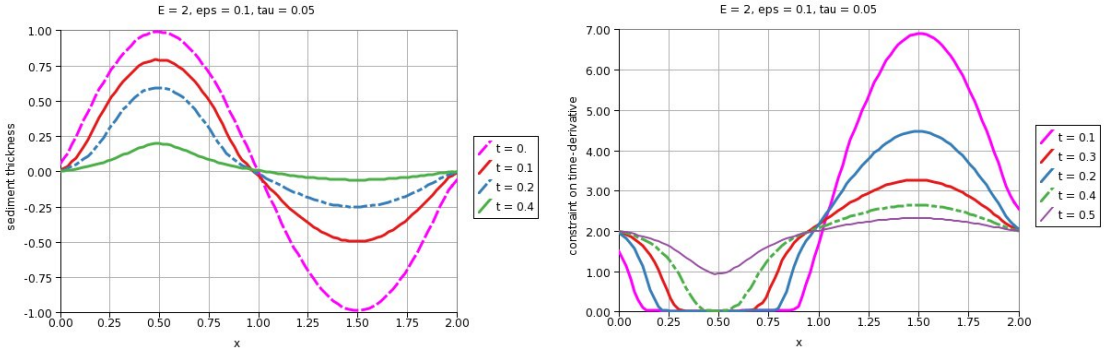


Figure 4.11: Vertical 1D cut at $y = 0.5$ of the numerical solution and the time-derivative constraint of u_h.

The figures (4.10,4.11) illustrate the influence of the maximum erosion rate parameter E on the sediment thickness. For large E, the erosion rate constraint is not active, in this case the algorithm converges in the first iteration. On the contrary case, for a small E, the erosion is constrained by the maximum erosion rate, for $\tau \geq 1$ the constraint is satisfied in the first iteration. Indeed, for $\tau < 1$ the constraint is not satisfied and the algorithm is repeated until the satisfaction of the constraint. We represent in the figure below the mesh and the numerical approximation of the time-derivative constraint of u_h at different iteration of the algorithm

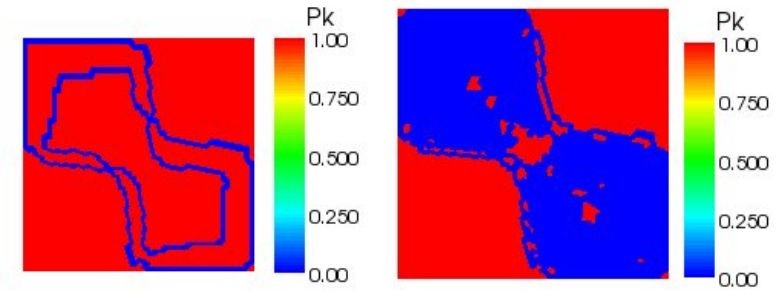


Figure 4.12: First and final adaptive mesh at time $t = 0.1$.

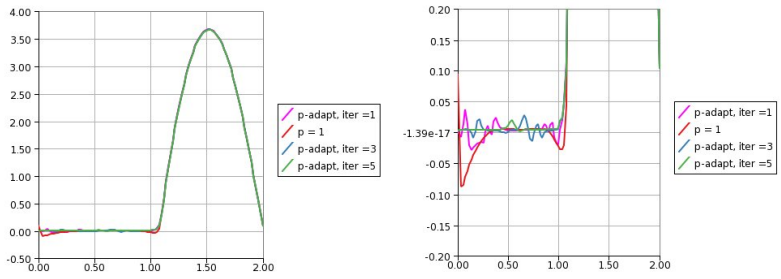


Figure 4.13: Vertical 1D cut of the numerical approximation of the time-derivative constraint at time $t = 0.1$, $\tau = 0.1$, $\varepsilon = 0.1$.

Some numerical oscillations between two mesh elements with different order appear in the first iterations of the algorithm. This oscillations disappeared in the last iteration which correspond to the final mesh see figure 4.12 where p is set to zero in the zone where the constraint is active.

4.7 Conclusions

We have presented a numerical scheme that implicitly takes into account the constraint on the time-derivative of the unknown. We have shown existence and uniqueness of the discrete solution under a condition on τ. Numerical and theoretical results show that the constraint (2) is implicitly satisfied if the $p = 0$ is used. Numerical experiments show that the constraint (2) is not satisfied when $p > 0$. An p-adaptive algorithm is proposed, some numerical simulations show that some numerical oscillations appear when a combined DgFem(0) and DgFem(p), $p > 0$ scheme is used, this oscillations disappeared after some iterations. Our aim in a future work is to propose an adaptive hp algorithm that combines h and p refinement in order to get more accuracy while still verifying the constraint.

Conclusions and perspectives

In this thesis, a mathematical model arising from stratigraphy under a constraint on the erosion rate is considered. The main feature of this model is that the constraint is implicitly satisfied in the equations.
In chapter 1, the discontinuous Galerkin finite element method applied to diffusion-advection problem has been presented. In particular, we have considered a way to choose the parameter γ that appears in the stabilization term. Some numerical simulations show the existence of optimal value of γ. An adaptive algorithm using red-green refinement for local mesh refinement has been detailed with some numerical examples.
In Chapter 2, the mathematical model is presented, we have proved the existence of a solution to the regularized problem. In the case of homogeneous problem and constant E, the uniqueness result is proved by S. Antontsev *et al.* in [5]. Our purpose in a future work is to generalize this approach to our case.
In Chapter 3, we have introduced and analyzed the discontinuous Galerkin time discretization for a linear pseudoparabolic problem which is a simplified linearized version of the model presented in Chapter 2. Then, *a priori* error estimates explicit in the time step Δt and the approximation order r has been established. The Chapter is concluded by a numerical example illustrating the theoretical result. As a complement of this Chapter, we propose to complete this result by estimating the space discretization error.
In Chapter 4, we have introduced the DgFem for the model problem. We have proved the monotonicity in the lowest-order case, this has been confirmed by some numerical simulations. We have established the existence of a discrete solution and its uniqueness under a condition on the parameter τ. Some numerical simulations show that, the constraint is not implicitly satisfied if a high order DgFem scheme is used. An adaptive algorithm combines DgFem(0) and high order DgFem scheme is proposed. Some numerical oscillations appear when a combined DgFem(0) and DgFem(p), $p > 0$ scheme is used, this will be ameliorated by refining in h each elements with $p = 0$. Our aim in a future work is on the one hand to propose an adaptive hp algorithm that combines h and p refinement in order to get accuracy and efficiency while still verifying the constraint. On the other hand, to do some numerical simulations for the physical problem by dealing with unilateral boundaries conditions.

Bibliography

[1] D. D. Ang and T. Tran. A nonlinear pseudoparabolic equations. Prog. Roy. Sec. Edinburgh, 114A:119–133, 1990.

[2] S. N. Antontsev, G. Gagneux, R. Luce, and G. Vallet. New unilateral problems in stratigraphy. M2AN Math. Model. Numer. Anal., 40(4):765–784, 2006.

[3] S. N. Antontsev, G. Gagneux, R. Luce, and G. Vallet. A non standard free boundary problem arising from stratigraphy. Anal. Appli., 4(3):209–236, 2006.

[4] S. N. Antontsev, G. Gagneux, R. Luce, and G. Vallet. On a pseudoparabolic problem with constraint. Differential and Integral Equations, 19(12):1391–1412, 2006.

[5] S. N. Antontsev, G. Gagneux, A. Mokrani, and G. Vallet. Stratigraphic modelling by the way of a pseudoparabolic problem with constraint. submitted, 2007.

[6] S. N. Antontsev, G. Gagneux, and G. Vallet. On some stratigraphic control problems. Journal of Applied Mechanics and Technical Physics, 44(6):821–828, 2003.

[7] D.N Arnold. An interior penalty finite element methods with discontinuous elements. SIAM J. Numer. Anal., 19(4):742–760, 1982.

[8] D.N. Arnold, F. Brezzi, B. Cockburn, and L.D. Marini. Unified analysis of discontinuous galerkin methods for elliptic problem. SIAM J. Numer. Anal., 39:1749–1779, 2002.

[9] I. Babuska and M. Zlamal. An elliptic collocation-finite element method with in- terior penalties. SIAM J. Numer. Anal., 15:152–161, 1978.

[10] G. A. Baker. Finite element methods for elliptic equations using nonconforming elements. Math. Comp., 31(137):45–59, 1977.

[11] G. I. Barenblatt. Similarity, self-similarity, and intermediate asymptotics. New York, London: Consultants Bureau. XVII, 1982.

[12] G. I. Barenblatt, M. Bertsch, R. Dal Passo, and M. Ughi. A degenerate pseudoparabolic regularization of a nonlinear forward-backward heat equation arising in the theory of heat and mass exchange in stably stratified turbulent shear flow. SIAM J. Math. Anal., 24(6):1414–1439, 1993.

[13] T. Benjamin, J. Bona, and J. Mahony. Model equations for long waves in nonlinear dispersive systems. Philos. Tran. Roy. Soc. London, 272:47–78, 1972.

[14] J. Bona, W. Pritchard, and L. Scott. Numerical schemes for a model for nonlinear dispersive waves. J. Comput. Phys., 60:167–186, 1985.

[15] S. C. Brenner. Poincaré-Friedrichs inequalities for piecewise H^1 functions. SIAM J. Numer. Anal., 41(1):306–324, 2003.

[16] P.G. Ciarlet. The finite element method for elliptic problems. Studies in Mathematics and its Applications. Vol. 4. Amsterdam - New York - Oxford: North-Holland Publishing Company., 1978.

[17] C. Cuesta and J. Hulshof. A model problem for groundwater flow with dynamic capillary pressure: stability of travelling waves. Nonlinear Anal., 52(4):1199–1218, 2003.

[18] J. Douglas and T. Dupont. Interior penalty procedures for elliptic and parabolic galerkin methods. Lecture Notes in Phys., Springer, Berlin, 58, 1976.

[19] W.-P. Düll. Some qualitative properties of solutions to a pseudoparabolic equation modeling solvent uptake in polymeric solids. Comm. Partial Differential Equations, 31(7-9):1117–1138, 2006.

[20] A. Ern, A. F. Stephansen, and M. Vohalík. Improved energy norm a posteriori error estimation based on flux reconstruction for discontinuous galerkin methods. SIAM J. Numer. Anal.(submitted.

[21] A. Ern, A. F. Stephansen, and P. Zunino. A discontinuous galerkin method with weighted averages for advection–diffusion equations with locally small and anisotropic diffusivity. IMA J. Numer. Anal., page doi: doi:10.1093/imanum/drm050, 2008.

[22] R. E. Ewing. The approximation of certain parabolic equations backward in time by Sobolev equations. SIAM J. Math. Anal., 6:283–294, 1975.

[23] R. E. Ewing. Time-stepping galerkin methods for nonlinear sobolev partial differential equations. SIAM J. Numer. Anal., 15:1125–1150, 1978.

[24] R. Eymard and T. Gallouët. Analytical and numerical study of a model of erosion and sedimentation. SIAM J. Numer. Anal., 43(6):2344–2370, 2006.

[25] R. Eymard, T. Gallouët, V. Gervais, and R. Masson. Convergence of numerical scheme for stratigraphic modeling. SIAM J. Numer. Anal., 2004.

[26] R. Eymard, T. Gallouët, D. Granjeon, R. Masson, and Q.H. Tran. Multilithology model under maximum erosion rate constraint. Internat. J. Numer. Methods Engrg., 60(2):527–548, 2004.

[27] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In Ciarlet, P. G. (ed.) et al., Handbook of numerical analysis. Vol. 7: Solution of equations in R^n (Part 3). Techniques of scientific computing (Part 3). Amsterdam: North-Holland/ Elsevier. 713-1020. 2000.

[28] J. Garcia-Azorero and A. de Pablo. Finite propagation for a pseudoparabolic equation: two-phase non-equilibrium flows in porous media. Nonlinear Anal., 33(6):551–573, 1998.

[29] V. Gervais and R. Masson. Mathematical and numerical analysis of a stratigraphic model. M2AN, 38(4):585–611, 2004.

[30] D. Granjeon. Modélisation stratigraphique déterministe; conception et applications d'un modèle diffusif 3d multilithologique. Ph.D Dissertation, Géosciences Rennes, Rennes, France, 1997.

[31] D. Granjeon. (Institut Français du Pétrole (IFP), personal communication, 2006.

[32] J. Douglas Jr, D. N. Arnold, and V. Thomée. Superconvergence of a finite element approximation to the solution of sobolev equation in a single space variable. Math. Comp., 36:53–63, 1981.

[33] O. Karakahian and F. Pascal. Convergence of adaptive discontinuous galerkin approximations of second-order elliptic problems. SIAM J. Numer. Anal., 45(2):641–665, 2007.

[34] M. G. Larson and A. J. Niklasson. Conservation properties for the continuous and discontinuous galerkin methods. Chalmers Finite Element Center, Goteborg Sweden, 2001.

[35] J.-L. Lions. Quelques méthodes de résolution des problèmes aux limites non linéaires. Dunod, 1969.

[36] V. Padrón. Effect of aggregation on population revovery modeled by a forward-backward pseudoparabolic equation. Trans. Amer. Math. Soc., 356(7):2739–2756 (electronic), 2004.

[37] P. I. Plotnikov. Passage to the limit with respect to viscosity in an equation with a variable direction of parabolicity. Differential Equations, 30(4):665–674, 734, 1994.

[38] B. Rivière, M. F. Wheeler, and V. Girault. Inproved energy estimates for interior penalty, constrained and discontinuous galerkin method for elleptic problem. I. Comput. Geosci., 3:337–360, 2000.

[39] C. Schwab. p- and hp-finite element methods. Oxford University Press, New York, 1998.

[40] V. Thomée. Galerkin finite element methods for parabolic problems. Springer Series in Computational Mathematics.

[41] R. Verfurth. A review of *a posteriori* error estimation and adaptive mesh-refinement techniques.

[42] K. Yosida. Functional analysis. Berlin-Göttingen-Heidelberg: Springer-Verlag, XI, 1965.

Résumé

Dans cette thèse, nous nous intéressons à un problème mathématique issu de la modélisation de taux d'érosion maximale dans la stratigraphie géologique. Une contrainte globale sur $\partial_t u$, la dérivée par rapport au temps de la solution, est la principale caractéristique de ce modèle. Ce qui nous amène à considérer une équation non linéaire pseudo-parabolique avec un coefficient de diffusion qui est une fonction non-linéaire de $\partial_t u$. En outre, le problème dégénère de telle sorte de tenir compte implicitement de la contrainte. Nous présentons un résultat de l'existence d'une solution au problème continu. Ensuite, une méthode DgFem (discontinuous Galerkin finite element method) pour son approximation numérique est développée. Notre objectif est d'utiliser les propriétés d'approximation constante par morceaux pour tenir compte implicitement de la contrainte.

Mots clés: Stratigraphie, méthode de Galerkin discontinue, contrainte, pseudo-parabolique.

Abstract

In this Thesis, we are interested in a mathematical problem arising from the modeling of maximal erosion rates in geological stratigraphy. A global constraint on $\partial_t u$, the time-derivative of the solution, is the main feature of this model. This leads to a non-linear pseudoparabolic equation with a diffusion coefficient which is a nonlinear function of $\partial_t u$. Moreover, the problem degenerates in order to take implicitly into account the constraint. We present a result of existence of a solution to the continuous problem. Then, a DgFem (discontinuous Galerkin finite element method) for its numerical approximation is developed. Our goal is to use the properties of piecewise constant approximation to keep implicitly the constraint.

Key words: Stratigraphy, discontinuous Galerkin method, constraint, pseudoparabolic.