UNIVERSITE SIDI MOHAMED BEN ABDELLAH
FACULTE DES SCIENCES ET TECHNIQUES - FES
Centre d'Etudes Doctorales : Sciences et Techniques de l'Ingénieur

N° d'ordre : 4/2017

THESE DE DOCTORAT

Présentée par

Mr: Mohammed Benzakour Amine

Spécialité : Mathématiques Appliquées

Sujet de la thèse : Schémas volumes finis implicites linéarisés pour des modèles de chimiotaxie.

Thèse présentée et soutenue le samedi 15 avril 2017 devant le jury composé de :

Nom Prénom	Titre	Etablissement	
Omar SIDKI	PES	Faculté des Sciences et Techniques de Fès	Président
Noura YOUSFI	PES	Faculté des Sciences Ben M'sik Casablanca	Rapporteur
Mustapha GHILANI	PES	ENSAM de Meknès	Rapporteur
Najib GUESSOUS	PES	Ecole Normale Supérieure de Fès	Rapporteur
Driss SEGHIR	PES	Faculté des Sciences de Meknès	Examinateur
Azeddine EL BARAKA	PES	Faculté des Sciences et Techniques de Fès	Examinateur
Mohammed AKHMOUCH	PES	Faculté des Sciences et Techniques de Fès	Directeur de thèse

Laboratoire d'accueil : Algèbre, Analyse Fonctionnelle et Applications

Etablissement : Faculté des Sciences et Techniques de Fès

بسم الله الرحمن الرحيم

وما بكم من نعمة فمن الله

Remerciements

Mes plus chaleureux remerciements s'adressent à mon directeur de thèse, le professeur Mohammed Akhmouch. Je le remercie pour m'avoir proposé un sujet de thèse très passionnant. Je tiens à lui exprimer ma profonde gratitude pour le temps qu'il m'a consacré, pour avoir été toujours disponible quand j'avais besoin d'aide et pour son soutient durant toutes ces années de thèse, même dans les moments les plus difficiles. Je tiens aussi à le remercier très sincèrement pour m'avoir fait aimer l'analyse numérique, et pour m'avoir initié à la recherche et cela dès le cycle licence.

J'exprime ma sincère reconnaissance au professeur Omar Sidki pour avoir présidé le Jury de la soutenance. Je le remercie également pour sa gentillesse et sa disponibilité.

Je remercie chaleureusement les professeurs Noura Yousfi, Mustapha Ghilani et Najib Guessous pour avoir accepté d'être les rapporteurs de cette thèse. Je les remercie également pour leurs remarques judicieuses et pour avoir accepté de discuter différents aspects de mon travail.

J'adresse mes sincères remerciements aux professeurs Driss Seghir et Azeddine El Baraka pour avoir accepté d'examiner ce travail et pour leurs différentes remarques et suggestions.

Je remercie aussi chaleureusement le professeur Najib Mahdou pour son aide et ses conseils pendant les années de cette thèse. J'adresse également un grand merci aux professeurs Mohammed Boutaleb, Ahmed Elhilali Alaoui et Lahcen Oukhtite, leur présence à la soutenance m'a fait le plus grand plaisir. Je souhaite aussi remercier très vivement tous les professeurs du département de mathématiques de la faculté des sciences et techniques de Fès.

Je tiens à remercier spécialement mes deux amis : Amine Allouhi et Noureddine Joudar qui m'ont apporté leur aide dans des phases déterminantes de mes études. Je remercie aussi chaleureusement tous mes collègues des différents laboratoires de recherche avec qui j'ai lié de forts liens d'amitié. Particulièrement, je tiens à remercier mes amis doctorants des deux laboratoires de mathématiques de la FST, et mes anciens collègues de master.

Un grand merci à mes anciens étudiants et étudiantes de licence et du master qui ont assisté à la soutenance.

Mes plus profonds remerciements vont à mes parents pour tout ce qu'ils ont fait pour moi, je ne saurai trouver les mots pour leur exprimer ma reconnaissance. C'est grâce à Allah puis grâce à vous que cette thèse a vu le jour. Je remercie aussi très chaleureusement mon frère Hamza et ma sœur Meryem pour leur soutien et leurs encouragements. J'adresse aussi un grand merci à tous les membres de ma famille.

Résumé

Les résultats présentés dans cette thèse sont dédiés à l'analyse et au développement de schémas numériques pour des modèles de chimiotaxie. Nous nous intéressons particulièrement à des systèmes d'équations aux dérivées partielles non-linéaires se basant sur le modèle de Keller-Segel et décrivant la formation de motifs soit dans le domaine de la bactériologie, ou dans le domaine du développement embryonnaire.

Pour tous ces modèles, nous présentons des approximations numériques adoptant une discrétisation temporelle implicite linéarisée. Cette approche permet d'obtenir des schémas linéaires découplés afin d'éviter le coût élevé de la résolution d'un système non-linéaire couplé. En ce qui concerne la discrétisation spatiale, nous adoptons des schémas volumes finis utilisant des approximations de type upwind, exponentiel ou hybride (upwind/centré), dans le but d'assurer la positivité et la stabilité de la solution numérique.

Ce travail a deux objectifs principaux : le premier objectif consiste à donner une analyse mathématique des schémas proposés. Ainsi, dans les différents chapitres de cette thèse, nous prouvons l'existence, l'unicité, la positivité et la convergence de la solution numérique. Les schémas implicites linéaires usuels ayant comme désavantage principal leur manque de précision, notre deuxième objectif est de développer de nouveaux schémas du même type qui soient plus précis et plus performants. Dans le chapitre 4, nous construisons ainsi un schéma semi-exponentiel en temps pour des modèles de chimiotaxie-croissance. Un schéma découplé corrigé pour une classe de modèles de chimiotaxie est aussi développé dans le chapitre 5. Plusieurs simulations et tests numériques sont présentés dans cette thèse afin de valider les résultats théoriques obtenus et pour démontrer l'efficience des méthodes développées.

Abstract

The results of this thesis are devoted to the analysis and development of numerical schemes for chemotaxis models. We are particularly interested in nonlinear systems of partial differential equations based on the Keller-Segel model, and describing the process of pattern formation which is performed by bacteria, or which appears during the embryonic development.

For all these models, we present numerical approximations which make use of a linearized implicit time discretization. This approach allows us to obtain linear schemes, and to avoid coupling between the equations to gain efficiency in term of computational cost. For the spatial discretization, we consider finite volume schemes, combined with upwind, exponential or hybrid (upwind/central) approximations to ensure the nonnegativity and the stability of the numerical solution.

This work has two main objectives. The first one is to give a mathematical study of the proposed schemes. Thus, in the different chapters of this thesis, the existence, uniqueness, nonnegativity and convergence of the numerical solution are proved. Since the main drawback of usual linearized implicit schemes is their lack of accuracy, our second objective is to develop new schemes of the same type which are more accurate and more efficient. In chapter 4, a time semi-exponentially fitted scheme for chemotaxis-growth models is thus constructed. A decoupled corrected scheme is also developed in chapter 5 for a class of chemotaxis models. Several simulations and numerical tests are performed in this thesis to validate our theoretical results and to demonstrate the efficiency of the developed methods.

Table des matières

Chapter 1

Introduction

Sommaire

1.1 Contexte général et motivation

1.1.1 Les mathématiques et les sciences du vivant

Malgré que les mathématiques soient à l'origine un ensemble de connaissances abstraites, elles jouent un rôle fondamental dans la compréhension du monde qui nous entoure. Cette "déraisonnable efficacité des mathématiques dans les sciences de la nature" comme l'a décrite Wigner [91] (prix Nobel de physique en 1963), est depuis longtemps une réalité universellement reconnue, tant elles ont prouvé au cours des années, leur capacité d'adaptation aux problèmes du monde réel. En sens inverse, les mathématiques ont toujours bénéficié de leur implication dans le développement des sciences. Les difficultés rencontrées dans celles-ci ont souvent stimulé des progrès considérables dans ses différentes branches.

La relation qu'entreprend les mathématiques avec la physique reste l'exemple le plus fascinant de cette complémentarité. En effet, il n'existe aucun domaine de la physique où elles ne sont omniprésentes. Notre façon de concevoir les phénomènes naturels de l'univers ne peut être précisément traduite que dans un langage mathématique. Les équations de Newton, d'Einstein, de Maxwell ou de Schrödinger, et beaucoup d'autres formulations des lois physiques montrent avec clairvoyance ce fait, ou bien peut être beaucoup plus, comme dit Feynman (prix Nobel de physique en 1965) : "seules les propriétés mathématiques du monde nous sont accessibles". D'autre part, l'intérêt que portent les mathématiciens pour les problèmes issus de la physique entrainent dans plusieurs cas, le développement de nouvelles méthodes, l'aboutissement à de nouveaux résultats ou la création de nouvelles branches mathématiques. C'est le cas par exemple du calcul différentiel et intégral qui a progressé grâce à des échanges avec la mécanique et la thermodynamique, de la théorie des équations aux dérivées partielles qui doit beaucoup aux travaux de d'Alembert sur la vibration des cordes, des séries de Fourier initialement développées pour la résolution de l'équation de la chaleur. L'influence de la physique sur la communauté mathématique se voit aussi clairement par l'attribution de la médaille Fields, plus prestigieuse récompense en mathématiques, au physicien théoricien Edward Witten en 1990, ainsi que par l'attribution de ce prix à plusieurs reprises pour des travaux en lien avec la physique.

Il reste que, ces dernières années, nous assistons à l'explosion de l'intérêt des mathématiciens pour un autre domaine scientifique fondamentalement différent : les sciences du vivant. La biologie mathématique, aussi appelée biomathématique, est un champ de recherche avec un vaste domaine d'application en sciences du vivant tels que : la biologie, l'écologie, la médecine, etc. Le lecteur intéressé pourra consulter l'excellent ouvrage de J.D. Murray [57], qui est une référence dans la matière.

Cette branche des mathématiques appliquées qui ne suscitait l'intérêt que d'un nombre très restreint de scientifiques dans la première moitié du vingtième siècle, compte actuellement des milliers de chercheurs. De plus, pour mener à bien leurs travaux qui nécessitent un ensemble de connaissances et de compétences, les biomathématiciens ont souvent besoin de la participation de spécialistes de différents domaines. Ainsi, beaucoup de résultats dans ce champ de recherche ont été le fruit de collaborations entre biomathématiciens, mathématiciens, biologistes, physiciens, biophysiciens, biochimistes, ingénieurs biologistes, physiologistes, cancérologues, biologistes moléculaires, généticiens, embryologistes, zoologistes, etc. D'autre part, les disciplines mathématiques trouvant leurs applications dans la biologie et dans les autres sciences du vivant sont de plus en plus nombreuses, on peut citer : les systèmes dynamiques, les équations différentielles et les équations aux dérivées partielles, le calcul scientifique, les probabilités, les statistiques, l'algèbre linéaire, la théorie des graphes, la géométrie algébrique, l'analyse combinatoire, la théorie des groupes, etc.

Par conséquent, et de façon similaire à la physique, les sciences du vivant sont devenues un véritable catalyseur pour le développement des mathématiques. Michael Mackey, ex-président de la société de biologie mathématique, affirme ce fait : "Dans mes 40 années et plus de recherche, je continue à constater que les problèmes en biomathématique nous dévoilent presque toujours des domaines inexplorés et inexploités des mathématiques", et Mackey est allé encore plus loin : "Pour plusieurs années, la physique a été la source principale d'inspiration pour l'innovation en mathématiques appliquées, mais à mon avis, c'est les sciences biologiques, dans leur sens le plus large, qui joueront ce rôle dans ce siècle" [41]. Ainsi, et pour ne citer que quelques exemples, les équations de Hodgkin-Huxley [37] et le papier de Turing sur la morphogénèse [89] ont inspiré beaucoup de travaux concernant les équations de réaction–diffusion, la formation des motifs et les ondes progressives [27, 79]. Les statistiques algébriques, nouveau champ de recherche des mathématiques appliquées, a été fortement stimulé par des problèmes en biologie [68]. Des progrès considérables en géométrie de l'information ont été le résultat de son utilisation pour analyser différent phénomènes biologiques [41]. En paléontologie, la question de savoir comment comparer les dents a conduit à de nouvelles perspectives en géométrie conforme [51]. Pour d'autres exemples de l'influence de la biologie sur le développement des mathématiques, nous référons à [73].

Cependant, malgré cette diversité des disciplines intervenant dans la biologie mathématique, le but ultime de celle-ci reste dans la plupart des cas, la modélisation mathématique et son corollaire, la simulation numérique. L'utilité de la modélisation mathématique dans les sciences du vivant peut se résumer en trois points essentiels :

- Meilleure compréhension des phénomènes biologiques : et cela, par l'étude quantitative des différents paramètres régissant un système biologique, ainsi que les relations existant entre eux. Cette conceptualisation ou théorisation du système vivant, a besoin néanmoins d'une validation expérimentale, et peut être éventuellement révisée ou rejetée. Elle a aussi parfois comme objectif de proposer de nouvelles approches expérimentales pour approfondir les connaissances concernant le phénomène étudié.

- Prédiction et anticipation : L'intérêt des modèles mathématiques apparaît aussi dans leur habileté à prédire les événements, même dans les situations les plus délicates. C'est le cas par exemple de la modélisation de la propagation des maladies infectieuses [32]. Le nombre de cas d'infection, le nombre de malades qui doivent être hospitalisés, le pourcentage de la population qui doit être vaccinée et bien d'autres informations que peuvent procurer ces modèles, permettent d'élaborer des politiques de santé publique efficaces. Il faut toutefois signaler que ces prédictions restent approximatives, et peuvent être totalement trompeuses si les hypothèses adoptées pour la construction du modèle sont incompatibles avec le système étudié.

- Contrôles des phénomènes : les principaux paramètres déterminant l'évolution d'un système biologique étant représentés dans le modèle, différentes stratégies de contrôles peuvent être développer en agissant sur ceux-ci. L'impact de chaque stratégie peut être ensuite évalué par des simulations numériques.

1.1.2 Motilité cellulaire et taxie

La cellule est une structure microscopique complexe, constitutive de tous les êtres vivants, et caractérisée par son pouvoir d'assimilation. Comme tout organisme vivant, le mouvement est une caractéristique essentielle de la cellule. Lors du déplacement de celle-ci, on peut distinguer deux cas différents :

- **Déplacement passif :** dans cette situation, le déplacement de la cellule se fait sans consommation d'énergie, c'est le cas par exemple d'un globule rouge se déplaçant d'une manière spontanée dans un vaisseau sanguin, poussé par la force motrice de la circulation.

- **Déplacement actif :** ce cas se caractérise par une consommation d'énergie lors du déplacement. Le déplacement d'un spermatozoïde dans un milieu liquide est un déplacement actif, puisque sa force motrice est due à l'action de son flagelle.

En biologie, la motilité cellulaire réfère à la capacité d'une cellule à se déplacer spontanément ou sous l'effet d'un stimulus (facteur naturel déclenchant une réaction physiologique).

Toutes les cellules sont en mouvement à au moins une période de leur cycle de vie, même les cellules des tissus de l'organisme connaissent plusieurs migrations en phase embryonnaire.

Quand le déplacement actif d'une cellule est lié à un signal extérieur, on parle de la taxie ou du tactisme. La taxie est un mouvement plus ou moins automatique qui rapproche ou éloigne d'un stimulus (par exemple : s'approcher d'une source de chaleur ou s'en éloigner). Le terme "taxie" (ou "tactisme") est souvent précédé d'un préfixe qui indique la nature du stimulus auquel les organismes d'un système donné répondent. Il existe énormément de types de taxie, dans ce qui suit, nous présentons quelques unes :

- **Aérotaxie :** migration d'un organisme vivant en fonction de la variation de la concentration en oxygène. Certaines bactéries se servent de l'aérotaxie pour atteindre un milieu avec une concentration d'oxygène optimale pour leur croissance.

- **Chimiotaxie :** phénomène qui permet aux organismes vivants de diriger leurs mouvements en réponse à des substances chimiques du monde extérieur.

- **Electrotaxie :** type de taxie référant au mouvement dirigé de cellules (ou d'organismes pluricellulaires) sous l'effet d'un champ électrique.

- **Magnétotaxie :** capacité à se déplacer le long de lignes d'un champ magnétique. Les premières bactéries magnétotactiques ont été découvertes en 1975.

- **Phototaxie :** processus par lequel les organismes vivants dirigent leurs mouvements sous l'influence de la lumière présente dans leur environnement. Lorsqu'il y a attraction vers la lumière, la phototaxie est dite positive. Lorsqu'il y a répulsion, elle est dite négative.

Parmi toutes les formes de taxies qui existent, la chimiotaxie reste le mécanisme le mieux compris, et celui qui suscitent le plus d'intérêt en raison de son rôle clé dans un large éventail de phénomènes biologiques. La chimiotaxie joue un rôle crucial durant le cycle de vie : les spermatozoïdes sont ainsi attirés vers l'ovule par des substances chimiques qu'elle secrète [35] ; dans la phase du développement embryonnaire, elle intervient par exemple dans la migration cellulaire au cours de la gastrulation [21] ou durant la formation de motifs sur la peau de certains animaux [58] ; ce même mécanisme est rencontré dans de nombreux autres phénomènes tels que la cicatrisation des plaies [20], la croissance tumorale [74], l'angiogenèse [82] etc.

La chimiotaxie est aussi liée à un champ de recherche qui suscite l'intérêt d'un très grand nombre de chercheurs : la formation des motifs dans les systèmes biologiques. En effet, les substances chimioattractantes qu'émettent plusieurs types de bactéries leur permettent de s'organiser et de s'agréger formant une multitude de types de structures. C'est le cas des formations hexagonales d'une bactérie de type *Vibrio* présentées dans la Figure 1.1, des anneaux continus générés par la bactérie *Salmonella typhimurium* (Figure 1.2) ou des motifs sous forme de spots remarquablement symétriques (voir Figure 1.3) formés par la bactérie *Escherichia coli.* Le mécanisme de chimiotaxie, supposé se déclencher à un stade du développement embryonnaire, intervient aussi dans l'organisation des cellules mélanocytes. Ces cellules sécrètent un pigment appelé mélanine, responsable de la couleur que prend la peau. Les motifs remarqués sur la peau ou sur le pelage de certains animaux (voir les Figures 1.4 et 1.5) sont le résultat de la présence ou de l'absence des cellules mélanocytes, et par conséquent de la mélanine.

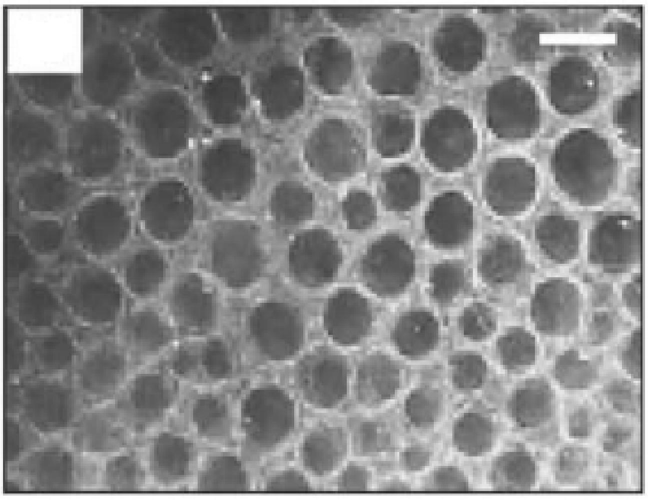

Figure 1.1 Exemple de motifs générés par des bactéries de type *Vibrio* (figure provenant de la référence [88]).

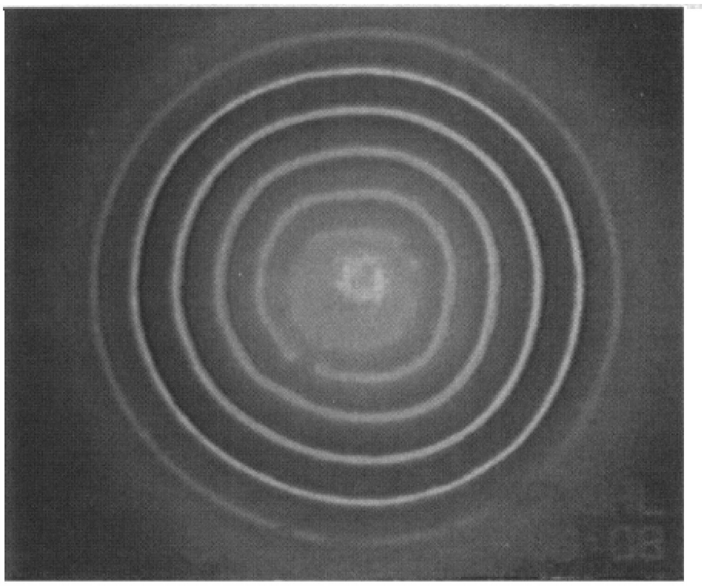

Figure 1.2 Anneaux continus générés par la bactérie *Salmonella typhimurium* (figure provenant de la référence [94]).

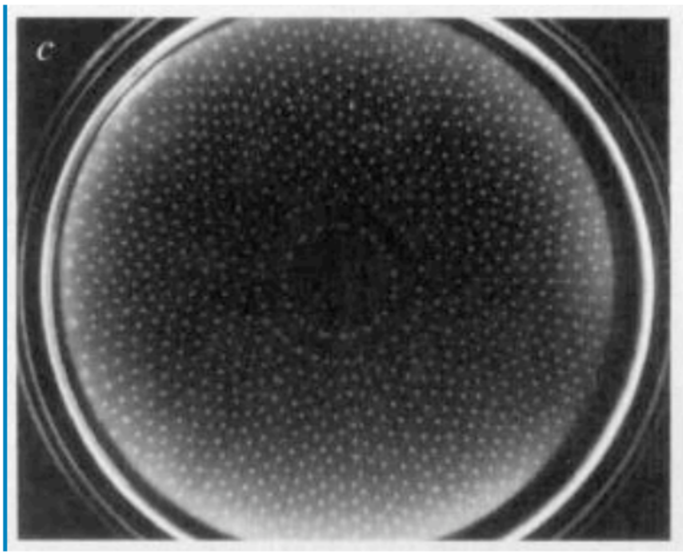

Figure 1.3 Exemple de motifs générés par la bactérie *Escherichia coli* (figure provenant de la référence [12]).

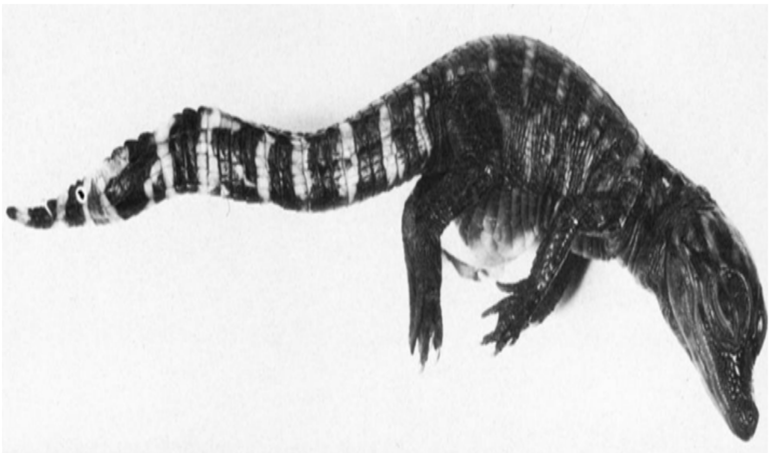

Figure 1.4 Motifs sous forme de bandes sur le tégument d'un nouveau-né alligator (figure provenant de la référence [58]).

Figure 1.5 Exemple de motifs sur la peau d'un serpent.

1.2 Le modèle de Keller-Segel

1.2.1 Présentation du modèle

La modélisation mathématique en chimiotaxie joue un rôle très important dans la compréhension des observations expérimentales. Elle date des travaux pionniers de Patlak en 1953 [72], et de Keller et Segel en 1970 [44, 45]. Le système de (Patlak-) Keller-Segel ainsi que ses nombreuses variantes restent, jusqu'à aujourd'hui, la classe de modèles la plus populaire dans l'étude des mouvements cellulaires dirigés selon des gradients chimiques. Sous sa forme classique, le modèle de Keller-Segel se présente comme suit :

$$\begin{cases} \partial_t u = D_u \, \Delta u - \nabla \cdot (\chi \, u \nabla c) \,, \\ \partial_t c = D_c \, \Delta c + \alpha u - \beta c. \end{cases} \tag{1.1}$$

Dans le système ci-dessus, D_u, D_c, χ, α et β sont des constantes strictement positives. La densité cellulaire ainsi que la concentration de la substance chimioattractante sont représentées respectivement par $u = u(x,t)$ et $c = c(x,t)$, où $x \in \Omega \subset \mathbb{R}^N$ et $t > 0$. Le système est complété par les conditions initiales positives (u_0, c_0), et dans le cas où Ω est un borné de $\mathbb{R}^N$, par des conditions de type Neumann homogène :

$$\nabla u \cdot \nu = \nabla c \cdot \nu = 0.$$

Concernant les paramètres du modèle (1.1), D_u représente le coefficient de diffusion des cellules et D_c le coefficient de diffusion de la substance chimioattractante, χ désigne la sensibilité des cellules au chimioattractant, α et β sont respectivement les taux de production et de dégradation de $c(x,t)$.

Le modèle de Keller-Segel a été initialement proposé pour l'étude du phénomène d'agrégation de certains êtres vivants unicellulaires appelés amibes, et plus précisément des amibes *Dictyostelium discoideum*. Celles-ci, en cas de manque de nutriments, sécrètent une substance chimioattractante qui attirent les autres amibes pour former une entité pluricellulaire. Ce nouveau organisme se développe en une structure pouvant produire des spores qui seront alors disperser dans le but de trouver un environnement plus viable.

Dans le système (1.1), le terme $\nabla \cdot (\chi\, u \nabla c)$ modélise le mécanisme de chimioattraction. Par opposition, si $\chi < 0$, on dit qu'il y a chimiorépulsion. Le modèle de Keller-Segel présenté reflète un nombre d'hypothèses qu'il est important de souligner :

- L'hypothèse de diffusion a été prise en compte aussi bien pour les cellules que pour le chimioattractant. Il est à noter cependant que, dans plusieurs cas, la diffusion de la substance chimioattractante est beaucoup plus rapide que celle des cellules, de ce fait, la deuxième équation du modèle peut être remplacée par l'équation elliptique

$$0 = D_c \Delta c + \alpha u - \beta c. \tag{1.2}$$

- La deuxième équation du modèle modélise le fait que le chimioattractant est produit par les cellules elles-même, et qu'il y a une dégradation de ce produit chimique au cours du temps.

- La masse de la densité cellulaire est conservé au cours du temps dans la première équation du modèle. Cela est compatible avec le fait que l'amibe *Dictyostelium discoideum* arrête sa prolifération lors du stade d'agrégation. Le modèle reste aussi applicable lors de l'étude de l'agrégation des bactéries, puisque la majorité des expériences in vitro ne durent que quelques minutes ou quelques heures, par suite la prolifération et la mort des bactéries restent négligeables. Cependant, certaines expériences durent des jours, et la prise en compte de la croissance bactérienne devient nécessaire [69].

Plusieurs variantes du système (1.1) ont été développées pour modéliser différents phénomènes. Nous référons à [69] pour une discussion sur ces différents modèles. Une forme générale enveloppant la plupart de ces variantes se présente comme suit

$$\begin{cases} \partial_t u = \nabla \cdot (k_1(u,c)\nabla u - k_2(u,c) u \nabla c) + k_3(u,c), \\ \partial_t c = D_c \Delta c + k_4(u,c) - k_5(u,c) c, \end{cases} \tag{1.3}$$

où $k_1(u,c)$ désigne le coefficient de diffusion des cellules, $k_2(u,c)$ représente la sensibilité des cellules à la substance chimioattractante, $k_3(u,c)$ décrit la prolifération et la mort des cellules, enfin, $k_4(u,c)$ et $k_5(u,c)$ sont des fonctions décrivant la production et la dégradation du chimioattractant.

1.2.2 Existence et prévention d'explosion en temps fini

De point de vue mathématique, le modèle de Keller-Segel (1.1) ainsi que sa forme parabolique-elliptique (où la deuxième équation de (1.1) est remplacée par (1.2)) ont été l'objet de plusieurs travaux, spécialement en raison de l'explosion en temps fini de leurs solutions. En effet, sous certaines conditions, la diffusion cellulaire est dominée par le terme de chimiotaxie, ce déséquilibre implique une explosion de la densité des cellules. Dans ce contexte, il convient donc de savoir sous quelles conditions la solution explose-t-elle dans un borné Ω de $\mathbb{R}^N$, ou dans l'espace tout entier de $\mathbb{R}^N$?

Pour le modèle en dimension 1, l'existence de solutions globales en temps et uniformément bornées est établie [66]. La situation est différente en dimension 2, plusieurs résultats ont été prouvés dans ce cas, on se contente de présenter quelques uns :

- Si $\Omega \subset \mathbb{R}^2$ est un borné connexe de frontière C^2, alors la solution du système parabolique-elliptique existe globalement en temps et est uniformément bornée si $\alpha\chi\|u_0\|_{L^1(\Omega)} < 4\pi$. Dans le cas où $\alpha\chi\|u_0\|_{L^1(\Omega)} > 4\pi$, la solution explose si $\displaystyle\int_\Omega u_0(x)|x-q|^2\,\mathrm{d}x$ est assez petit [63], où q est un point appartenant à la frontière de Ω.

- Si $\Omega \subset \mathbb{R}^2$ est une boule et si la condition initiale u_0 présente une symétrie radiale, alors la solution du système parabolique-elliptique existe globalement en temps et est uniformément bornée si la condition $\alpha\chi\|u_0\|_{L^1(\Omega)} < 8\pi$ est vérifiée. Dans le cas où $\alpha\chi\|u_0\|_{L^1(\Omega)} > 8\pi$, la solution explose si $\displaystyle\int_\Omega u_0(x)|x|^2\,\mathrm{d}x$ est assez petit [64, 61].

- Si $\Omega = \mathbb{R}^2$, le seuil critique pour l'explosion est aussi $\alpha\chi\|u_0\|_{L^1(\Omega)} = 8\pi$ [62].

- Si $\Omega \subset \mathbb{R}^2$ est simplement connexe, et pour tout $m > 4\pi$ tel que $m \notin \{4k\pi \setminus k \in \mathbb{N}\}$, il existe une condition initiale (u_0, c_0) avec $\alpha\chi\|u_0\|_{L^1(\Omega)} = m$ tel que la solution du système parabolique-parabolique (1.1) explose en temps fini ou infini [40, 80]. Par contre, si $\alpha\chi\|u_0\|_{L^1(\Omega)} < 4\pi$ la solution existe globalement en temps et est uniformément bornée [64, 29].

En dimension supérieur, si $\Omega \subset \mathbb{R}^N$ $(N \geq 3)$ est une boule, alors pour toute masse $m > 0$ avec $\alpha\chi\|u_0\|_{L^1(\Omega)} = m$, il existe u_0 et v_0 tel que la solution du problème (1.1) explose en temps fini ou infini [92]. Pour plus de résultats concernant l'analyse mathématique du modèle de Keller-Segel, nous référons à titre d'exemple à [38, 39].

L'explosion en temps fini de la densité cellulaire n'étant pas un phénomène biologiquement acceptable, plusieurs approches ont été proposées pour éviter ce comportement pathologique de la solution du système de Keller-Segel. Une approche consiste à considérer une diffusion non-linéaire pour la densité cellulaire en remplaçant la première équation du système par :

$$\partial_t u = \nabla \cdot (D(u)\,\nabla u - \chi\, u \nabla c)\,,$$

où $D(u) \geq \delta u^p$, avec $\delta > 0$. Dans ce cas, si Ω est un ouvert borné de frontière C^2 appartenant à $\mathbb{R}^N$, l'existence d'une solution faible bornée a été établie sous la condition $p > 2 - 4/N$ [46]. L'existence d'une solution bornée est aussi vérifiée pour les modèles avec effet de remplissage de volume (*volume-filling models*). Dans ces modèles, on modifie la sensibilité du chimioattractant par une fonction $\chi(u)$ qui s'annule lorsqu'elle atteint un certain seuil u_m [19]. En particulier, on peut éviter l'explosion en temps fini de la solution en adoptant l'équation suivante

$$\partial_t u = D_u\,\Delta u - \nabla \cdot (\,u(1-u)\nabla c)\,.$$

D'autres approches visant à empêcher l'explosion en temps fini de la solution consistent à ajouter un terme modélisant la croissance cellulaire à la première équation

$$\partial_t u = D_u\,\Delta u - \nabla\cdot(\chi\, u\nabla c) + f(u),$$

où f vérifie certaines propriétés [86, 93]. L'existence d'une solution bornée est aussi vérifiée en dimension 2 dans le cas du modèle de Keller-Segel avec diffusion croisée, c'est à dire en remplaçant la deuxième équation de (1.1) par

$$\partial_t c = D_c\,\Delta c + \delta\Delta u + \alpha u - \beta c,$$

avec $\delta > 0$, ou par la forme elliptique de l'équation ci-dessus [36]. Enfin, nous signalons que si le terme source dans la deuxième équation du système (1.1) est borné, l'explosion en temps fini n'a pas lieu. Un exemple de cette situation est de considérer l'équation suivante [18]

$$\partial_t c = D_c\,\Delta c + \alpha\frac{u}{u+1} - \beta c.$$

1.2.3 Approximation numérique du modèle de Keller-Segel et de ses variantes

Une littérature significative existe en ce qui concerne les méthodes numériques pour la résolution du modèle de Keller-Segel (1.1) et de ses variantes.

Des schémas différences finies ont été introduits dans [77] par Saito et Suzuki, et dans [90] par Tyson et al.. Concernant la méthode des éléments finis, nous mentionnons le schéma conservatif avec correction de flux de Strehl et al. [84, 83]. Une méthode éléments finis mixtes a été aussi proposée par Marrocco [53]. Un schéma volumes finis d'ordre deux a été développé par Chertock et Kurganov [17]. D'autres méthodes ont été introduites pour différentes versions du modèle de Keller-Segel, à savoir la méthode de maillages mobiles adaptatifs développé par Budd et al. [10], la méthode des différences de potentiels de Epshteyn [22] et la méthode de lattice Boltzmann proposée par Yang et al. [95].

D'autres travaux se sont intéressés aussi à l'analyse mathématique des schémas proposés. Des estimations d'erreurs pour un schéma élément finis conservatif ont été obtenues par Saito [75, 76]. Des estimations d'erreurs ont été aussi prouvées pour un schéma de Galerkin discontinu par Epshteyn et Kurganov [23] pour le cas semi-discret et par Epshteyn et Izmirlioglu [24] avec prise en compte de la discrétisation temporelle. Une analyse de convergence a été aussi établie par Blanchet et al. pour un schéma de plus forte pente [9], et par Haškovec et Schmeiser pour une approximation particulaire stochastique [33, 34].

Concernant les schémas volumes finis, une analyse de convergence d'un schéma volumes finis de type upwind pour la version parabolique-elliptique de (1.1) a été réalisée par Filbet [28]. Pour la preuve de convergence aucune hypothèse de régularité sur la solution du système n'a été supposée. Cependant, à cause de l'explosion en temps fini de la solution du modèle, la convergence du schéma n'a été établie que dans le cas où $C_\Omega\chi\|u_0\|_{L^1(\Omega)} < 1$, où C_Ω est une constante liée à l'inégalité de Gagliardo-Nirenberg-Sobolev. Une généralisation de ce schéma pour le modèle de Keller-Segel avec diffusion croisée a été analysée par Bessemoulin-Chatard et Jüngel [7]. Aucune condition sur $\|u_0\|_{L^1(\Omega)}$ n'est obligatoire dans ce cas puisque le terme de diffusion croisée prévient l'explosion de la solution (voir Section 1.2.2). La non-nécessité d'une telle condition apparaît aussi dans le papier d'Andreianov et al. [5], où les auteurs ont prouvé la convergence d'un schéma volumes finis pour un modèle de Keller-Segel avec effet de remplissage de volume et une diffusion non-linéaire. Des schémas volumes-éléments finis ont été aussi analysés pour ce dernier modèle par Chamoun et al. [15], et Ibrahim et Saad [42], mais dans le cas de présence d'un tenseur anisotrope dans les termes de diffusion.

Enfin, nous notons que dans tous ces travaux, différentes approches pour la discrétisation temporelles ont été adoptées. Pour les schémas explicites, on peut consulter par exemple les travaux [17, 22, 24, 34]. Pour les schémas implicites, on peut citer par exemple [53, 28, 7, 9], et pour des discrétisations semi-implicites, on mentionne à titre d'exemple [84, 42, 5, 15].

1.3 Schéma volumes finis pour l'équation de diffusion

Dans cette section, nous donnons une présentation simplifiée du principe de la discrétisation volumes finis. Nous nous restreindrons pour cela à des problèmes de diffusion elliptiques en deux et trois dimensions d'espace, les équations paraboliques de convection-diffusion seront traitées plus en détails dans les chapitres qui suivent, étant donné que les modèles de chimiotaxie se basent sur de telles équations.

On considère le problème suivant :

$$\begin{cases} -\Delta u = f & \text{sur } \Omega, \\ \nabla u \cdot \nu = 0 & \text{sur } \partial\Omega, \end{cases} \tag{1.4}$$

où Ω est un ouvert de $\mathbb{R}^d$, $d = 2$ (resp. $d = 3$), de frontière assez régulière $\partial\Omega$, et ν est la normale à $\partial\Omega$ extérieure à Ω. Comme on peut voir, les conditions aux limites considérées sont de type Neumann homogène. Les modèles de chimiotaxie qui vont être étudiés dans cette thèse adoptent tous le même type de conditions aux limites.

Avant de construire le schéma volumes finis pour le problème (1.4), nous commençons d'abord par donner la définition d'un maillage admissible $\mathcal{T}$ de Ω, au sens de [26, Définition 5.1].

Définition 1.3.1. *Soit Ω un ouvert borné polygonal (resp. polyédrique) de $\mathbb{R}^d$, $d = 2$ (resp. $d = 3$). Un maillage volumes finis admissible $\mathcal{T}$ de Ω, est constitué d'une famille de volumes de contrôle, aussi notée par $\mathcal{T}$, d'une famille d'arêtes $\mathcal{E}$ (resp. de faces), et d'une famille de points distincts $\mathcal{P}$, vérifiant les propriétés suivantes :*

- *La clôture de l'union de tous les volumes de contrôle est $\overline{\Omega}$.*
- *Si K et L sont deux volumes de contrôle voisins, alors $\overline{K} \cap \overline{L}$ est un sommet ou une arête (resp. une face).*
- *Pour tout $K \in \mathcal{T}$, K est un ouvert convexe polygonal (resp. polyédrique) et $\partial K = \overline{K} \setminus K \subset \mathcal{E}$.*
- *La famille de points $\mathcal{P} = (x_K)_{K\in\mathcal{T}}$ telle que $x_K \in \overline{K}$ (pour tout $K \in \mathcal{T}$) vérifie la propriété suivante : si $\sigma = K|L$ est l'arête entre K et L, alors la droite passant par x_K et x_L est orthogonale à σ (voir Figure 1.6). Dans le cas où $\sigma \in \mathcal{E}$, $\sigma \subset \partial\Omega$ et $x_K \notin \sigma$, alors l'existence d'une droite passant par x_K et orthogonale σ est vérifiée.*

Pour tout $K \in \mathcal{T}$, on désigne par $\mathcal{N}(K)$ l'ensemble des volumes de contrôle voisins de K, et par $\mathcal{E}_K$ l'ensemble des arêtes du volume K. On notera par h le diamètre maximal de tous les volumes de contrôle appartenant à $\mathcal{T}$.

On notera par m la mesure de Lebesgue d-dimensionnelle ou $(d-1)$-dimensionnelle.

Pour tout $\sigma \in \mathcal{E}_K$, on définit

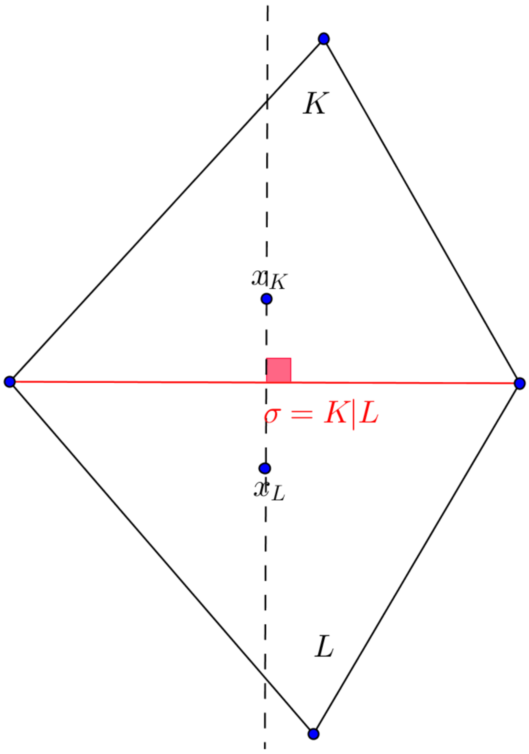

Figure 1.6 Volumes de contrôle triangulaires en deux dimensions.

$$d_\sigma = \begin{cases} \mathrm{d}(x_K, \sigma) & \text{si } \sigma \subset \partial\Omega, \\ \mathrm{d}(x_K, x_L) & \text{sinon }, \ \sigma = K|L, \end{cases}$$

où d est la distance Euclidienne. On notera aussi par τ_σ la transmissibilité à travers l'interface σ :

$$\tau_\sigma = \frac{\mathrm{m}(\sigma)}{d_\sigma}, \quad \sigma \in \mathcal{E}.$$

En intégrant l'équation aux dérivées partielles du problème (1.4) sur $K \in \mathcal{T}$, on obtient

$$\int_K -\Delta u \, dx = \int_K f \, dx.$$

En utilisant la formule de Green-Gauss, on a alors

$$-\int_{\partial K} \nabla u(x) \cdot \nu_K d\gamma(x) = \int_K f(x) \, dx.$$

Dans la formule ci-dessus, ν_K est la normale unitaire à ∂K extérieur à K. En utilisant le fait que K est polygonal (ou polyédrique si $d = 3$), on peut écrire :

$$-\sum_{\sigma \in \mathcal{E}_K} \int_\sigma \nabla u \cdot \nu_{K,\sigma} d\sigma = \int_K f(x) \, dx,$$

où $\nu_{K,\sigma}$ représente la normale unitaire à σ extérieur à K. La dernière étape de la construction du schéma est l'approximation du flux $-\displaystyle\int_\sigma \nabla u \cdot \nu_{K,\sigma} d\sigma$. Dans le cas où $\sigma \subset \partial\Omega$, on peut facilement voir que le flux s'annule en utilisant les conditions aux limites de Neumann homogène.

Dans le cas où $\sigma = K|L \not\subset \partial\Omega$, on adopte l'approximation suivante qui est consistante dans le cas d'un maillage admissible :

$$-\int_\sigma \nabla u \cdot \nu_{K,\sigma} d\sigma \approx -\tau_\sigma \left(u_L - u_K\right),$$

où $(u_K)_{K\in\mathcal{T}}$ sont les inconnues discrètes supposées approximer la valeur moyenne de u sur chaque K. Le schéma final du problème (1.4) se présente comme suit :

$$-\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma Du_{K,\sigma} = \mathrm{m}(K) f_K,$$

avec

$$Du_{K,\sigma} = \begin{cases} 0 & \text{si } \sigma \subset \partial\Omega, \\ u_L - u_K & \text{sinon }, \ \sigma = K|L, \end{cases}$$

et

$$f_K = \frac{1}{\mathrm{m}(K)} \int_K f\, dx.$$

Soit φ une bijection de l'ensemble $\mathcal{T}$ vers $\{1, 2, ..., l\}$, où l est le nombre d'éléments de $\mathcal{T}$ (i.e le nombre de volumes de contrôle). On utilise cette fonction pour numéroter les volumes de contrôle. Similairement au problème unidimensionnel, le schéma présenté peut s'écrire sous la forme matricielle suivante :

$$AU = b,$$

où pour tout $K \in \mathcal{T}$, les vecteurs U et b sont définis par

$$U_{\varphi(K)} = u_K, \quad b_{\varphi(K)} = \mathrm{m}(K) f_K.$$

La matrice du système A se présente comme suit :

$$\begin{aligned} &A_{\varphi(K),\varphi(K)} = \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma\not\subset\partial\Omega}} \tau_\sigma, \\ &A_{\varphi(K),\varphi(L)} = -\tau_\sigma, \quad \text{si } L \in \mathcal{N}(K), \\ &A_{\varphi(K),\varphi(L)} = 0, \quad \text{sinon.} \end{aligned}$$

1.4 Plan de la thèse et présentation des travaux

Nous nous intéressons dans ce manuscrit à l'étude et au développement de schémas volumes finis implicites linéarisés pour des systèmes d'équations aux dérivées partielles non-linéaires basés sur le modèle de Keller-Segel. L'avantage principal de tels schémas réside dans leur faible coût de calcul. L'objectif de cette thèse peut être résumé en deux points :

- L'analyse de convergence de différents schémas volumes finis implicites linéarisés : cet objectif a été réalisé pour les chapitres 2 et 3 où des schémas upwind combinés avec des discrétisations temporelles implicites linéarisés usuelles ont été utilisés, et pour les chapitres 4 et 5 où on a utilisé respectivement des schémas exponentiels et des schémas hybrides combinés avec de nouveaux types de discrétisations temporelles.

- Le développement de nouveaux schémas implicites linéarisés : notre apport en ce qui concerne la conception des schémas apparaît dans les chapitres 4 et 5, où de nouvelles discrétisations temporelles ont été présentées. Ainsi, un schéma semi-exponentiel en temps a été développé dans le chapitre 4 et un nouveau schéma découplé avec un terme de correction a été présenté dans le chapitre 5. Les simulation numériques qui ont été effectuées montrent que ces deux schémas sont plus performants que les schémas linéarisés usuels.

Le travail présenté dans cette thèse a fait l'objet de trois articles [2, 3, 4] publiés dans les revues : **Indagationes Mathematicae**, **Calcolo**, et **Journal of Computational and Applied Mathematics**.

Dans ce qui suit, un résumé des travaux effectués est présenté. Nous adopterons aussi les notations de la Section 1.3. Les détailles de chaque chapitre se trouvent dans la version anglaise.

Chapitre 2 : Schémas volumes finis semi-implicites pour un modèle de chimiotaxie-croissance

Dans ce chapitre nous nous intéressons à l'équation de chimiotaxie-croissance suivante :

$$\begin{cases} \partial_t u = \Delta u - \nabla \cdot (\chi\, u \nabla c) + f(u) \quad \text{dans } \Omega_T, \\ 0 = \Delta c + u - c \quad \text{dans } \Omega_T, \end{cases} \tag{1.5}$$

avec les conditions aux limites et initiales :

$$\nabla u \cdot \nu = \nabla c \cdot \nu = 0 \quad \text{sur } \partial\Omega \times (0,T), \quad u(.,0) = u_0(x) \quad \text{dans } \Omega, \tag{1.6}$$

où $\Omega_T := \Omega \times (0,T)$, Ω est un ouvert borné de $\mathbb{R}^d$, $d = 2$ (resp. $d = 3$), de frontière assez régulière $\partial\Omega$. $T > 0$ représente le temps final, et $f(u) = au(b - u^{\alpha-1})$ avec $a > 0$, $b > 0$ et $\alpha \geq 3$. Le système (1.5) ressemble à la version parabolic-elliptic du modèle classique de Keller-Segel (première équation de (1.1) avec l'équation (1.2)), hormis le faite que la croissance cellulaire est prise en compte cette fois par l'ajout de la source logistique $f(u)$. Ce terme a une importance primordiale dans la modélisation des expériences in vitro de longue durée. En effet, dans le cas par exemple de la mise en culture de la bactérie *Escherichia coli*, on remarque la formation de spots bactériens d'une symétrie remarquable [11, 12] (voir Figure1.3). Le mécanisme par lequel ces motifs sont générés peut être expliqué par trois phénomènes : la chimiotaxie, la diffusion et la croissance.

Nous considérons un maillage admissible au sens de la Définition 1.3.1. Pour des raisons techniques liées à la démonstration de la convergence du schéma qui va être proposé, nous considérons une extension de l'intervalle de temps $(0, T + \epsilon)$ où $\epsilon > 0$. Soit $0 = t_0 < t_1 < ... < t_N = T < t_{N+1} = T + \epsilon$ une partition de $(0, T + \epsilon)$ avec $N \in \mathbb{N}$, le pas de temps est défini par $\Delta t_n = t_{n+1} - t_n$, $n = 0, ..., N$. Nous notons $\Delta t = \max_{0 \leq n \leq N} \Delta t_n$ et $\delta = \max(\Delta t, h)$.

Pour la résolution numérique du système (1.5)–(1.6), nous proposons deux schémas numériques. Le premier schéma est donné par les équations suivantes :

pour tout $K \in \mathcal{T}$ et $n = 0, ..., N$

$$
\begin{aligned}
&\mathrm{m}(K)\frac{u_K^{n+1}-u_K^n}{\Delta t_n} - \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma D u_{K,\sigma}^{n+1} \\
&+ \chi \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \tau_\sigma \Big((Dc_{K,\sigma}^{n+1})^+ u_K^{n+1} - (Dc_{K,\sigma}^{n+1})^- u_L^{n+1} \Big) - \mathrm{m}(K)\, f\left(u_K^n\right) = 0, && (1.7)\\
&- \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\, Dc_{K,\sigma}^{n+1} = \mathrm{m}(K)\, \left(u_K^n - c_K^{n+1}\right), && (1.8)
\end{aligned}
$$

complétées par la condition initiale :

$$
u_K^0 = \frac{1}{\mathrm{m}(K)} \int_K u_0(x)\, dx, \tag{1.9}
$$

avec

$$
z^+ = \max(z,0),\ \ z^- = \max(-z,0)
$$

$$
v_K^n \approx \frac{1}{\mathrm{m}(K)} \int_K v(x,t^n)\, dx.
$$

Le deuxième schéma consiste à remplacer l'équation (1.7) dans le schéma (1.7)–(1.9) par l'équation suivante :

$$
\begin{aligned}
&\mathrm{m}(K)\frac{u_K^{n+1}-u_K^n}{\Delta t_n} - \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma D u_{K,\sigma}^{n+1} \\
&+ \chi \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \tau_\sigma \Big((Dc_{K,\sigma}^{n+1})^+ u_K^{n+1} - (Dc_{K,\sigma}^{n+1})^- u_L^{n+1} \Big) + \mathrm{m}(K)\, a u_K^{n+1} \Big((u_K^n)^{\alpha-1} - b \Big) = 0. && (1.10)
\end{aligned}
$$

On définit $(u_\delta,\ c_\delta)$, l'approximation volumes finis de (u,c) par :

$$
u_\delta(x,t) = u_K^{n+1}, \quad c_\delta(x,t) = c_K^{n+1}, \quad x\in K,\ t\in[t^n,t^{n+1}). \tag{1.11}
$$

Les deux schémas proposés utilisent une approximation upwind pour discrétiser le terme de convection. La discrétisation temporelle utilisée dans (1.8) ainsi que les schémas semi-implicites proposés pour la première équation de (1.5) permettent de construire des schémas découplés et linéaires. A chaque itération, nous commençons donc par résoudre (1.8) pour calculer c_K^{n+1}, puis on résout (1.7) ou (1.10) pour obtenir u_K^{n+1}. Cette stratégie est beaucoup moins coûteuse par rapport à la résolution du schéma non-linéaire couplé. On mentionne dans ce contexte les travaux intéressants [50, 31, 49], où on démontre en utilisant de nouvelles techniques que les schémas éléments finis implicites linéarisés pour les équations de Joule ne nécessitent pas de conditions CFL pour converger. La même remarque reste valable pour les schémas semi-implicites volumes finis appliqués au système de dérive diffusion [13, 14].

Comme on peut voir, la seule différence entre le schéma (1.7) et le schéma (1.10) est la discrétisation du terme de croissance, qui est explicite dans le cas du schéma (1.7), et semi-implicite dans le cas du schéma (1.10).

Définition 1.4.1. *Soit $\psi \in \mathcal{D}(\Omega \times [0,T))$. On appelle solution faible du problème* (1.5)–(1.6) *un couple $(u,c) \in L^2(0,T;H^1(\Omega))^2$ solution de :*

$$\int_0^T \int_\Omega (u\,\partial_t \psi - \nabla u \cdot \nabla \psi + \chi\, u \nabla c \cdot \nabla \psi + f(u)\psi)\ dxdt + \int_\Omega u_0\,\psi(x,0)\,dx = 0,$$
$$\int_0^T \int_\Omega \nabla c \cdot \nabla \psi\ dxdt = \int_0^T \int_\Omega (u-c)\psi\,dxdt.$$

Les deux propositions suivantes donnent des conditions suffisantes pour la stabilité et la positivité des solutions des deux schémas.

Proposition 1.4.1. *Supposons que $u_0 \geq 0$, et que la condition suivante sur le pas de temps est satisfaite*

$$\Delta t < \frac{1}{ab}. \tag{1.12}$$

Alors, il existe une unique solution $\{(u_K^{n+1}, c_K^{n+1}),\ K \in \mathcal{T},\ 0 \leq n \leq N\}$ au schéma (1.10),(1.8)–(1.9)*, telle que $u_K^{n+1} \geq 0$ and $c_K^{n+1} \geq 0$ pour tout $K \in \mathcal{T}$, $0 \leq n \leq N$.*

Proposition 1.4.2. *Supposons que $u_0 \geq 0$, et que pour tout $0 \leq n \leq N$*

$$a\Delta t_n \left(\left(\max_{K\in\mathcal{T}} u_K^n \right)^{\alpha-1} - b \right) \leq 1. \tag{1.13}$$

Alors, le schéma (1.7)–(1.9) *admet une unique solution positive $\{(u_K^{n+1}, c_K^{n+1}),\ K \in \mathcal{T},\ 0 \leq n \leq N\}$.*

On peut remarquer que la condition de stabilité du premier schéma ne dépend que du pas de temps et des paramètres du modèle, tandis que la condition de stabilité du deuxième schéma dépend aussi de la solution du schéma.

Le résultat principal de ce chapitre concernant la convergence du schéma (1.7)–(1.9) se présente comme suit :

Théorème 1.4.1. *Supposons que $u_0 \in L^2(\Omega)$ et que $u_0 \geq 0$. Supposons que la condition* (1.13) *est vérifiée. Alors la solution volumes finis (u_δ, c_δ) du schéma* (1.7)–(1.9) *converge vers (u,c) quand $\delta \to 0$, où (u,c) est une solution faible du système* (1.5)–(1.6) *au sens de la Définition 1.4.1.*

Dans la dernière partie de ce chapitre, nous présentons des simulations numériques afin d'étudier la capacité des deux schémas à capturer les formations bactériennes de motifs. Une comparaison entre les coûts de calcul des deux schémas est aussi présentée.

Chapitre 3 : Un schéma volumes finis pour un modèle de chimiotaxie issu de l'embryologie

L'objectif principal de ce chapitre est l'analyse mathématique d'un schéma volumes finis découplé pour le modèle suivant

$$\begin{cases} \partial_t u = \mu\Delta u - a\nabla\cdot(\,u\nabla c) & \text{dans } \Omega_T, \\ 0 = \Delta c + \dfrac{u}{u+1} - c & \text{dans } \Omega_T, \end{cases} \tag{1.14}$$

muni des conditions aux limites et initiales suivantes :

$$\nabla u \cdot \nu = \nabla c \cdot \nu = 0 \quad \text{sur } \partial\Omega \times (0,T), \quad u(.,0) = u_0 \quad \text{dans } \Omega, \tag{1.15}$$

où $\Omega_T := \Omega \times (0,T)$, Ω est un ouvert borné de $\mathbb{R}^2$, de frontière assez régulière $\partial\Omega$. $T > 0$ représente le temps final, et μ et a sont deux constantes strictement positives.

Dans [58], Murray et al. ont proposé ce modèle pour modéliser la formation des motifs sous forme de bandes qui se développent sur le tégument des alligators (voir Figure 1.4), principalement durant la période embryonnaire. La formation de ces bandes est le résultat de secrétions chimioattractantes dont la concentration est représentée dans le modèle par la variable c, la densité des cellules mélanocytes responsables du changement de la couleur de la peau est représentée par la variable u.

Le maillage considéré dans ce chapitre est un maillage admissible au sens de la Définition 1.3.1. Cependant, nous aurons aussi besoin de supposer que le maillage satisfasse la condition suivante : il existe $\xi > 0$ tel que

$$\mathrm{d}(x_K, \sigma) \geq \xi \, \mathrm{d}(x_K, x_L), \tag{1.16}$$

cette condition sera utilisée lors de la dérivation des estimations a priori nécessaires pour prouver la convergence du schéma. Contrairement au Chapitre 2, nous considérons une partition uniforme de $(0,T)$: $0 = t_0 < t_1 < ... < t_N = T$ où $N \in \mathbb{N}$ et $t_n = n\Delta t$ pour $n = 0, ..., N$. Nous notons aussi $\delta = \max(\Delta t, h)$.

Le schéma volumes finis proposé pour le système (1.14)–(1.15) est :

pour tout $K \in \mathcal{T}$ and $n = 0, ..., N-1$

$$\mathrm{m}(K)\frac{u_K^{n+1} - u_K^n}{\Delta t} - \mu \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma D u_{K,\sigma}^{n+1}$$

$$+ \, a \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \tau_\sigma \left((D c_{K,\sigma}^{n+1})^+ u_K^{n+1} - (D c_{K,\sigma}^{n+1})^- u_L^{n+1} \right) = 0, \tag{1.17}$$

$$- \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma \, D c_{K,\sigma}^{n+1} = \mathrm{m}(K) \, \frac{u_K^n}{u_K^n + 1} - \mathrm{m}(K) \, c_K^{n+1}, \tag{1.18}$$

$$u_K^0 = \frac{1}{\mathrm{m}(K)} \int_K u_0(x) \, dx. \tag{1.19}$$

La solution volumes finis $(u_\delta, \, c_\delta)$ calculée à partir de la solution du schéma (1.17)–(1.19) est définie par :

$$u_\delta(x,t) = u_K^{n+1}, \quad c_\delta(x,t) = c_K^{n+1}, \quad x \in K, \, t \in [t^n, t^{n+1}). \tag{1.20}$$

Nous définissons ci-dessous la solution faible du problème (1.14)–(1.15).

Définition 1.4.2. *Soit $\psi \in \mathcal{D}(\Omega \times [0,T))$. On appelle solution faible du problème* (1.14)–(1.15) *un couple $(u,c) \in L^2(0,T;H^1(\Omega))^2$ solution de :*

$$\int_0^T \int_\Omega (u \, \partial_t \psi - \mu \nabla u \cdot \nabla \psi + a \, u \nabla c \cdot \nabla \psi) \, dx dt + \int_\Omega u_0 \, \psi(x,0) \, dx = 0,$$

$$\int_0^T \int_\Omega \nabla c \cdot \nabla \psi \, dx dt = \int_0^T \int_\Omega \left(\frac{u}{u+1} - c \right) \psi \, dx dt.$$

Comme il a été déjà mentionné dans la Section 1.2.3, la convergence du schéma volumes finis proposé dans [28] pour la version classique du modèle de Keller-Segel est obtenue à condition que la masse cellulaire initiale ne dépasse pas un certain seuil. De plus, la démonstration proposée suppose que la donnée initiale est strictement positive. Dans ce chapitre, nous nous inspirons des techniques de démonstration de [28] et nous démontrons que la bornitude du terme source $\dfrac{u}{u+1}$ dans la deuxième équation du système (1.14) permet d'éviter de telles restrictions.

Théorème 1.4.2. *Supposons que* $u_0 \in L^2(\Omega)$ *et que* $u_0 \geq 0$. *Supposons que la condition sur le maillage spatial* (1.16) *est vérifiée. Alors la solution volumes finis* (u_δ, c_δ) *du schéma* (1.17)–(1.19) *converge vers* (u, c) *quand* $\delta \to 0$, *où* (u, c) *est une solution faible du système* (1.14)–(1.15) *au sens de la Définition 1.4.2.*

Des simulations numériques en deux dimensions d'espace sont proposées en fin de chapitre.

Chapitre 4 : Un schéma semi-exponentiel en temps pour des modèles de chimiotaxie-croissance

Dans ce chapitre, nous nous intéressons à des modèles de chimiotaxie avec un terme de croissance polynomial. Nous nous proposons de développer pour de tels modèles un schéma numérique ayant les propriétés suivantes : le schéma ne requiert que la résolution de systèmes linéaires découplés tout en étant stable sous aucune contraintes liée au pas de temps. La positivité des solutions numériques est aussi obtenue inconditionnellement.

Nous considérons en premier lieu le système :

$$\begin{cases} \partial_t u = \Delta u - \nabla \cdot (\, u \nabla c) + u\,(1-u) \quad \text{dans } \Omega_T, \\ 0 = \Delta c + \dfrac{u}{u+1} - c \quad \text{dans } \Omega_T, \end{cases} \tag{1.21}$$

avec les conditions aux limites et initiales suivantes :

$$\nabla u \cdot \nu = \nabla c \cdot \nu = 0 \quad \text{sur } \partial\Omega \times (0, T_f), \quad u(.,0) = u_0(x) \quad \text{dans } \Omega, \tag{1.22}$$

où $\Omega_T := \Omega \times (0, T_f)$, Ω est un ouvert borné de $\mathbb{R}^2$, de frontière assez régulière $\partial\Omega$. $T_f > 0$ représente le temps final. Le système ci-dessus a été proposé dans [59] pour modéliser la formation de motifs sur la peau des serpents (voir par exemple Figure 1.5). La discrétisation des domaines spatial et temporel est similaire à celle du Chapitre 3, la restriction sur le maillage (1.16) est supposée aussi vérifiée. Un schéma numérique pour le système (1.21)–(1.22) vérifiant les propriétés désirées est le schéma semi-implicite suivant :

pour tout $K \in \mathcal{T}$ and $n = 0, ..., N-1$

$$\mathrm{m}(K)\frac{u_K^{n+1} - u_K^n}{\Delta t} + \sum_{\sigma \in \mathcal{E}_K} F_{K,\sigma}^{n+1} - \mathrm{m}(K)\, u_K^n \left(1 - u_K^{n+1}\right) = 0, \tag{1.23}$$

$$-\sum_{\sigma \in \mathcal{E}_K} \tau_\sigma\, Dc_{K,\sigma}^{n+1} = \mathrm{m}(K)\,\frac{u_K^n}{u_K^n + 1} - \mathrm{m}(K)\, c_K^{n+1}, \tag{1.24}$$

$$u_K^0 = \frac{1}{\mathrm{m}(K)} \int_K u_0(x)\, dx. \tag{1.25}$$

Le flux numérique $F_{K,\sigma}^{n+1}$ choisit dans ce chapitre est défini par

$$F_{K,\sigma}^{n+1} = \begin{cases} 0 & \text{si } \sigma \subset \partial\Omega, \\ \tau_\sigma \left(B\left(-Dc_{K,\sigma}^{n+1}\right) u_K^{n+1} - B\left(Dc_{K,\sigma}^{n+1}\right) u_L^{n+1} \right) & \text{sinon, } \sigma = K|L, \end{cases} \tag{1.26}$$

où B est la fonction de Bernoulli définie par

$$B(0) = 1,\ B(x) = \frac{x}{\exp(x) - 1}\ \forall x \in \mathbb{R} \setminus \{0\}. \tag{1.27}$$

Dans le schéma ci-dessus, le flux $\displaystyle\int_\sigma (\nabla u(x,t) - u(x,t)\nabla c(x,t)) \cdot \nu_{K,\sigma}\, d\sigma$ a été approché par le schéma exponentiel d'Il'in [43], contrairement aux deux chapitres précédents où on a utilisé un schéma upwind. Le schéma semi-exponentiel en temps qu'on propose pour la première équation de (1.21), combine l'idée du schéma d'Il'in initialement proposé pour la discrétisation spatiale, avec l'idée de la discrétisation semi-implicite. Le schéma se présente comme suit :

pour tout $K \in \mathcal{T}$ and $n = 0, ..., N-1$

$$T_K^{n+1} + \sum_{\sigma \in \mathcal{E}_K} F_{K,\sigma}^{n+1} = 0. \tag{1.28}$$

avec

$$T_K^{n+1} = \mathrm{m}(K) \frac{B\left(\Delta t\left(1 - u_K^n\right)\right) u_K^{n+1} - B\left(-\Delta t\left(1 - u_K^n\right)\right) u_K^n}{\Delta t}, \tag{1.29}$$

$F_{K,\sigma}^{n+1}$ et B sont définis respectivement par (1.26) et (1.27). On définit la solution volumes finis par :

$$u_\delta(x,t) = u_K^{n+1}, \quad c_\delta(x,t) = c_K^{n+1}, \quad x \in K,\ t \in [t^n, t^{n+1}). \tag{1.30}$$

Théorème 1.4.3. *Supposons que* $u_0 \in L^2(\Omega)$ *et que* $u_0 \geq 0$. *Supposons que la condition sur le maillage spatial* (1.16) *est vérifiée. Alors la solution volumes finis* (u_δ, c_δ) *du schéma* (1.28),(1.24)–(1.25) *converge vers* (u,c) *quand* $\delta \to 0$*, où* (u,c) *est une solution faible du système* (1.21)–(1.22) *au sens de :*

$$\int_0^T \int_\Omega \left(u \frac{\partial \psi}{\partial t} - \nabla u \cdot \nabla \psi + u \nabla c \cdot \nabla \psi + u\left(1-u\right)\psi \right) dxdt + \int_\Omega u_0\, \psi(x,0)\, dx = 0,$$

$$\int_0^T \int_\Omega \nabla c \cdot \nabla \psi\ dxdt = \int_0^T \int_\Omega \left(\frac{u}{u+1} - c \right) \psi\, dxdt.$$

Une comparaison entre le schéma (1.23)–(1.25) et le schéma semi-exponentiel en temps (1.28),(1.24)–(1.25) est effectuée dans la dernière partie du chapitre. Dans les tests numériques présentés, on démontre qu'on peut atteindre une précision donnée en moins de temps à l'aide du schéma semi-exponentiel. La même comparaison est aussi réalisée dans le cas d'un modèle de chimiotaxie-croissance en bactériologie. L'efficacité du schéma pour la capture des motifs bactériens est aussi étudiée par des exemples numériques.

Chapitre 5 : Un schéma découplé corrigé pour des modèles de chimiotaxie

Le but de ce chapitre est de proposer un nouveau schéma numérique découplé pour des modèles de chimiotaxie. Comme il a été déjà mentionné, les schémas découplés présentent un avantage considérable au niveau du coût de calcul par rapport aux schémas couplés, mais peuvent être parfois très insatisfaisants au niveau de la précision. Notre objectif sera donc d'améliorer la précision des schémas découplés classiques, sans avoir à augmenter leur coût de calcul.

Une attention particulière sera donnée au modèle (1.14)–(1.15) pour qui le schéma proposé sera analysé. La discrétisation des domaines spatial et temporel est similaire à celle du Chapitre 3. Nous commençons par présenter un schéma découplé classique pour (1.14)–(1.15) :

pour tout $K \in \mathcal{T}$ and $n = 0, ..., N-1$

$$\begin{aligned} &\mathrm{m}(K)\frac{u_K^{n+1}-u_K^n}{\Delta t} - \mu \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma D u_{K,\sigma}^{n+1} \\ &+ a \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \tau_\sigma \left(S\left(Dc_{K,\sigma}^{n+1}\right) u_K^{n+1} - S\left(-Dc_{K,\sigma}^{n+1}\right) u_L^{n+1}\right) = 0, \end{aligned} \tag{1.31}$$

$$- \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \, Dc_{K,\sigma}^{n+1} + \mathrm{m}(K)\, c_K^{n+1} = \mathrm{m}(K)\, \frac{u_K^n}{u_K^n+1}, \tag{1.32}$$

$$u_K^0 = \frac{1}{\mathrm{m}(K)} \int_K u_0(x)\, dx, \tag{1.33}$$

avec

$$S(x) = \begin{cases} 0, & \text{if } x < 2\left(-\mu+\varepsilon\right)/a, \\ \dfrac{x}{2}, & \text{if } |x| \le 2\left(\mu-\varepsilon\right)/a, \\ x, & \text{if }, x > 2\left(\mu-\varepsilon\right)/a, \end{cases}$$

où ε est une constante telle que $\varepsilon \ge 0$ et $\varepsilon << \mu$. Dans le cas où $\varepsilon = 0$, on retrouve le schéma hybride de Spalding [81] utilisant le schéma centré dans son domaine de stabilité et le schéma upwind en dehors de ce domaine. Il est clair qu'on peut obtenir une bien meilleure précision en utilisant le schéma couplé non-linéaire, c'est à dire en remplaçant l'équation (1.32) dans le schéma (1.31)–(1.33) par :

$$- \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \, Dc_{K,\sigma}^{n+1} + \mathrm{m}(K)\, c_K^{n+1} = \mathrm{m}(K)\, \frac{u_K^{n+1}}{u_K^{n+1}+1}. \tag{1.34}$$

Cependant, un tel schéma nécessite le résolution d'un large système non-linéaire à chaque itération, ce qui peut s'avérer très coûteux en temps de calcul.

Afin de présenter notre schéma découplé corrigé, nous définissons d'abord pour tout $K \in \mathcal{T}$ et $n = 0, ..., N-1$, la différence T_K^{n+1} par :

$$T_K^{n+1} = \mathrm{m}(K)\left(\frac{u_K^{n+1}}{u_K^{n+1}+1} - \frac{u_K^n}{u_K^n+1} \right), \quad T_K^0 = 0. \tag{1.35}$$

L'idée du schéma vient du faite que l'équation (1.34) peut être écrite sous la forme suivante :

$$-\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \, Dc_{K,\sigma}^{n+1} + \mathrm{m}(K)\, c_K^{n+1} = \mathrm{m}(K)\, \frac{u_K^n}{u_K^n+1} + T_K^{n+1}. \tag{1.36}$$

La seule différence entre l'équation ci-dessus et (1.32) est le terme T_K^{n+1}. Nous supposons donc qu'en ajoutant un terme de correction à l'équation (1.32) approximant T_K^{n+1}, on pourra améliorer la précision du schéma découplé. Notre approche consiste alors à remplacer (1.32) par l'équation suivante :

$$-\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \, Dc_{K,\sigma}^{n+1} + \mathrm{m}(K)\, c_K^{n+1} = \mathrm{m}(K)\, \frac{u_K^n}{u_K^n+1} + \beta_n T_K^{n}, \tag{1.37}$$

la constante β_n est choisi de tel façon que le membre de droite de l'équation reste positive afin de ne pas affecter la positivité de c_K^{n+1}. En pratique, cela est généralement vérifiée pour le choix le plus naturel de β_n i.e. $\beta_n = 1$. La solution volumes finis du schéma (1.31),(1.37),(1.33) est défini par

$$u_\delta(x,t) = u_K^{n+1}, \quad c_\delta(x,t) = c_K^{n+1}, \quad x \in K,\ t \in [t^n, t^{n+1}). \tag{1.38}$$

Théorème 1.4.4. *Supposons que $u_0 \in L^2(\Omega)$ et que $u_0 \geq 0$. Supposons que la condition suivante sur le pas de temps est satisfaite : il existe $\alpha > 0$ tel que $1 - 2a\Delta t \geq \alpha$, ou que la condition sur le maillage spatial* (1.16) *est vérifiée et $\varepsilon > 0$. Alors la solution volumes finis (u_δ, c_δ) du schéma* (1.31),(1.37),(1.33) *converge vers (u,c) quand $\delta \to 0$, où (u,c) est une solution faible du système* (1.14)–(1.15) *au sens de la Définition 1.4.1.*

En fin de ce chapitre, nous comparons notre approche avec les schémas découplés usuels en effectuant des tests numériques sur trois systèmes de chimiotaxie différents. Les résultats numériques obtenus montrent que le schéma découplé corrigé est beaucoup plus précis que les schémas découplés classiques.

Chapter 2

Semi-implicit finite volume schemes for a chemotaxis-growth model

Sommaire

2.1 Introduction

One of the crucial features of living organisms is their ability to move and migrate. Chemotaxis, the phenomenon that enables cells (or organisms) to migrate in response to chemical gradients, is the most encountered mechanism that drives living cell motility. Chemotaxis plays a key-role in aggregation and self-organization of bacteria. It plays also a fundamental role in several physiological and pathological process such as : embryogenesis, wound healing, leukocytes moving in response to bacterial inflammation and tumor growth. In the last decades, many mathematical models have been developed to describe several biological phenomena where chemotaxis is involved. The most known one has been introduced by Patlak [72], and Keller and Segel [44]. The classical form of this model reads :

$$\begin{cases} \partial_t u = \Delta u - \nabla \cdot (\chi\, u \nabla c) \quad \text{in } \Omega_T, \\ \partial_t c = \Delta c - c + u \quad \text{in } \Omega_T, \end{cases} \tag{2.1}$$

with homogeneous Neumann boundary conditions and initial conditions. Herein, $\Omega_T := \Omega \times (0,T)$, Ω is an open bounded domain in $\mathbb{R}^d$ ($d \in \mathbb{N} \setminus \{0\}$) and $T > 0$. In this model, $u(x,t)$ and $c(x,t)$ denote respectively the cell density and the chemoattractant concentration in $(x,t) \in \Omega \times (0,T)$, and χ is a positive constant describing the chemotactic sensitivity. In an other version of the (Patlak-) Keller-Segel model, the second equation in (2.1) has been replaced by the elliptic equation

$$\Delta c - c + u = 0 \quad \text{in } \Omega \times (0,T). \tag{2.2}$$

This substitution can be justified by the assumption that the diffusion of the chemoattractant is much faster than that of the cell .

This model was originally proposed to describe the aggregation of slim molds in response to chemical signal. The Keller-Segel system (2.1) and its parabolic-elliptic form (by considering (2.2)) may blow up in finite time. This blow up occurs for specific initial conditions for u. The occurrence of the blow-up depends also on χ. Several works are devoted to mathematical analysis of these systems (see, e.g., [38, 39]).

In the literature, a number of numerical methods for the Keller-Segel system (2.1) and its variants have been proposed, such as the central-upwind finite volume method [17], finite difference methods [77], finite element methods [84, 83, 53], moving mesh methods [10], lattice Boltzmann method [95], and upwind-difference potentials method [22]. Some other works are devoted to the convergence analysis of numerical schemes for the classical Keller-Segel model and its paraobolic-elliptic version. Saito proved error estimates for a conservative finite element method [75, 76]. In [28], a fully implicit finite volume scheme was proposed by Filbet. After proving the existence and uniqueness of the numerical solution, he shows that the numerical approximation converges to the weak solution of the problem under a smallness condition on the initial cell density for $d = 2$. A generalization of this scheme for a Keller-Segel model with additional cross-diffusion was studied by Bessemoulin-Chatard and Jüngel [7]. In [23], a numerical analysis of a semidiscrete discontinuous Galerkin scheme was performed by Epshteyn and Kurganov. The convergence analysis of the fully discrete scheme was then shown by Epshteyn and Izmirlioglu [24]. A stochastic interacting particle scheme [34, 33] and a steepest descent scheme [9] were studied for a more simplified Keller-Segel model. We note also the works of Andreianov et al. [5], Chamoun et al. [15], and Ibrahim and Saad [42], which are devoted to the numerical approximation of volume-filling chemotaxis problems.

In the model (2.1), the total mass is conserved in time, which means that growth of bacteria (or cell) is ignored. However, growth can have an important effect on the dynamics of bacteria [47]. In [11] and [12], Budrene and Berg found that motile cells of *Escherichia coli* aggregate to form stable patterns of remarkable regularity when grown from a single point on certain substrates. One of the models which describes this process of pattern formation reads

$$\begin{cases} \partial_t u = D_u \, \Delta u - \nabla \cdot (\chi \, u \nabla c) + f(u) & \text{in } \Omega_T, \\ \partial_t c = D_c \, \Delta c - \beta c + \mu u & \text{in } \Omega_T, \end{cases} \tag{2.3}$$

where $f(u)$ is a logistic source representing the growth rate of u, and χ, D_u, D_c, β, μ are positive constants. In this chapter, we are concerned with the parabolic-elliptic version of the above model with the generalized logistic source (see, e.g., [65]), $f(u) = au(b - u^{\alpha-1})$ defined for $u > 0$, with $a > 0$ and $b > 0$. We focus in particular on the case $\alpha \geq 3$, and we may assume without loss of generality that

$$D_u = D_c = \beta = \mu = 1. \tag{2.4}$$

So, the studied model reads :

$$\begin{cases} \partial_t u = \Delta u - \nabla \cdot (\chi\, u \nabla c) + f(u) & \text{in } \Omega_T, \\ 0 = \Delta c - c + u & \text{in } \Omega_T, \end{cases} \tag{2.5}$$

with zero flux boundary conditions and initial conditions

$$\nabla u \cdot \nu = \nabla c \cdot \nu = 0 \quad \text{on } \partial\Omega \times (0,T), \quad u(.,0) = u_0(x) \quad \text{in } \Omega, \tag{2.6}$$

where $\Omega \subset \mathbb{R}^d$, $d = 2$ (resp. $d = 3$) is an open bounded subset with sufficiently regular boundary $\partial\Omega$, and ν denotes the outward unit-normal on the boundary $\partial\Omega$.

The systems (2.3),(2.6) and (2.5)–(2.6) were studied for different quadratic and cubic forms of $f(u)$ (see the references below). These models were theoretically discussed in several papers (see, e.g., [93, 86, 55, 66, 1]), and it is shown that the logistic source may prevent their solutions from blowing up in finite time. It is also shown in [86] that, under some assumptions on the logistic source, there exists a weak solution of (2.5)–(2.6) for $d \geq 1$, without any restriction on the parameters of the system. From the numerical viewpoint, there are few works devoted to describe numerical methods for the system (2.3),(2.6). In [42], Ibrahim and Saad proposed and analyzed an implicit control volume finite element scheme for a volume-filling version of (2.3),(2.6), with nonlinear degenerate diffusion and anisotropic diffusion tensors. Finite element method was developed in [84] for two spatial dimensions and in [83] for three spatial dimensions with simulations, but a convergence analysis was not given. Other papers are concerned with the numerical investigation of the same parabolic-parabolic model (see, e.g., [70, 1]).

In this chapter, we propose two semi-implicit finite volume schemes to solve (2.5)–(2.6). The existence of nonnegative solution of each scheme is proved under time step conditions. However, for one of these schemes, the condition depends on the discrete solution, so we need to adjust the step size throughout the calculations. Such strategy have been used in the framework of finite difference [77] and finite element approximations [76] to the simplified Keller-Segel model. For this scheme, we show that the numerical approximation converges to a weak solution of (2.5)–(2.6), without needing to impose a threshold on $\|u_0\|_{L^1(\Omega)}$ to prove the convergence. We note that (2.4) is supposed only to alleviate the formulae, and does not affect the mathematical analysis of the finite volume scheme.

This chapter is organized as follows. In the next section, we present our finite volume schemes to approximate the solution of (2.5)–(2.6) and state our main result. Existence, uniqueness and nonnegativity of the numerical solutions are shown in Section 2.3. A priori estimates are given in Section 2.4. In Section 2.5, we use these estimates to prove that the approximate solution converges to a weak solution of the studied model. Finally, numerical simulations are performed in Section 2.6.

2.2 Finite volume schemes and main result

2.2.1 Finite volume approximation

Let $\Omega \subset \mathbb{R}^d$, $d = 2$ (resp. $d = 3$) be an open bounded polygonal (resp. polyhedral) subset. We consider an admissible finite volume mesh of Ω, denoted by $\mathcal{T}$, which verifies the properties of Definition 5.1 in [26] (see also Définition 1.3.1). This mesh is given by :

- A family of control volumes which is commonly denoted by the same notation of the mesh $\mathcal{T}$. All control volumes are open and convex polygons (resp. polyhedra).
- A family $\mathcal{E}$ of edges (resp. sides), where the set of edges of any control volume $K \in \mathcal{T}$ is denoted by $\mathcal{E}_K$. We denote also by $\sigma = K|L$ the interface between K and L ($\sigma \in \mathcal{E}_K$ and $\sigma \notin \partial\Omega$)
- A family of points $(x_K)_{K\in\mathcal{T}}$ such that $x_K \in \overline{K}$ (for all $K \in \mathcal{T}$). The straight line going through x_K and x_L must be orthogonal to $\sigma = K|L$.

For all $K \in \mathcal{T}$, we denote by $\mathcal{N}(K)$ the set of control volumes which are neighbors to K, and by m the measure in $\mathbb{R}^d$ or $\mathbb{R}^{d-1}$.

For all $\sigma \in \mathcal{E}_K$, we define

$$d_\sigma = \begin{cases} \mathrm{d}(x_K, \sigma) & \text{if } \sigma \subset \partial\Omega, \\ \mathrm{d}(x_K, x_L) & \text{otherwise , } \sigma = K|L, \end{cases}$$

where d is the Euclidean distance, and we denote by τ_σ the transmissibility coefficient given by :

$$\tau_\sigma = \frac{\mathrm{m}(\sigma)}{d_\sigma}, \quad \sigma \in \mathcal{E}.$$

For technical reasons, we extend the time interval to $(0, T+\epsilon)$ where $\epsilon > 0$. Let $0 = t_0 < t_1 < ... < t_N = T < t_{N+1} = T + \epsilon$ be a partition of $(0, T+\epsilon)$ with $N \in \mathbb{N}$, and let $\Delta t_n = t_{n+1} - t_n$, $n = 0, ..., N$. We define $\Delta t = \max_{\,0\leq n\leq N} \Delta t_n$.

We denote by h the maximal size (diameter) of the control volumes included in $\mathcal{T}$, and we define

$$\delta = \max(\Delta t, h).$$

We present now the finite volume schemes proposed for the problem (2.5)–(2.6). The first scheme reads as follows :

for all $K \in \mathcal{T}$ and $n = 0, ..., N$

$$\begin{aligned} &\mathrm{m}(K)\frac{u_K^{n+1} - u_K^n}{\Delta t_n} - \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma D u_{K,\sigma}^{n+1} \\ &+ \chi \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \tau_\sigma \left((Dc_{K,\sigma}^{n+1})^+ u_K^{n+1} - (Dc_{K,\sigma}^{n+1})^- u_L^{n+1} \right) - \mathrm{m}(K)\, f\,(u_K^n) = 0, \qquad (2.7) \\ &- \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\, Dc_{K,\sigma}^{n+1} = \mathrm{m}(K)\, \left(u_K^n - c_K^{n+1} \right), \qquad (2.8) \end{aligned}$$

with the compatible initial condition

$$u_K^0 = \frac{1}{\mathrm{m}(K)} \int_K u_0(x)\, dx, \tag{2.9}$$

and the following notations

$$z^+ = \max(z, 0),\ \ z^- = \max(-z, 0)$$

$$v_K^n \approx \frac{1}{\mathrm{m}(K)} \int_K v(x, t^n)\, dx,$$

and for all $\sigma \in \mathcal{E}_K$

$$Dv^n_{K,\sigma} = \begin{cases} 0 & \text{for } \sigma \subset \partial\Omega, \\ v^n_L - v^n_K & \text{otherwise }, \ \sigma = K|L. \end{cases}$$

For the second scheme, we replace the equation (2.7) in the scheme (2.7)–(2.9) by the following equation :

$$\begin{aligned} &\mathrm{m}(K)\frac{u_K^{n+1} - u_K^n}{\Delta t_n} - \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma D u^{n+1}_{K,\sigma} \\ &+ \chi \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \tau_\sigma \left((Dc^{n+1}_{K,\sigma})^+ u_K^{n+1} - (Dc^{n+1}_{K,\sigma})^- u_L^{n+1} \right) + \mathrm{m}(K)\, a u_K^{n+1} \left((u_K^n)^{\alpha-1} - b \right) = 0. \end{aligned} \qquad (2.10)$$

The proposed finite volume schemes use a linearized semi-implicit time discretization, which is well known to be more efficient in term of computational cost than a fully implicit one. At each time step, the numerical solution can be obtained by solving only two decoupled linear systems. As in [28], the convective term is discretizated using an upwind technique. The only difference between the two schemes is the discretization of the logistic source, which is explicit for (2.7)–(2.9), and semi-implicit for (2.10),(2.8)–(2.9).

We define u_δ, c_δ, the finite volume approximations of u and c by :

$$u_\delta(x,t) = u_K^{n+1}, \quad c_\delta(x,t) = c_K^{n+1}, \quad x \in K,\, t \in [t^n, t^{n+1}). \qquad (2.11)$$

Let $X(\mathcal{T})$ be the set of functions from Ω to $\mathbb{R}$ which are constant over each control volume of the mesh. For $1 \le p < \infty$, and for $v \in X(\mathcal{T})$, the classical discrete L^p norm reads

$$\| v \|_p = \left(\sum_{K\in\mathcal{T}} \mathrm{m}(K)\, |v_K|^p \right)^{1/p},$$

where $v(x) = v_K$ for all $x \in K$ and for all $K \in \mathcal{T}$.

Following [8], we define also the discrete $W^{1,p}$ seminorm and the discrete $W^{1,p}$ norm :

$$| v |_{1,p,\mathcal{T}} = \left(\sum_{\sigma\in\mathcal{E}} \frac{\mathrm{m}(\sigma)}{d_\sigma^{p-1}} |D_\sigma v|^p \right)^{1/p},$$

$$\| v \|_{1,p,\mathcal{T}} = \| v \|_p + | v |_{1,p,\mathcal{T}}\,,$$

where for all $\sigma \in \mathcal{E}$, $D_\sigma v = 0$ if $\sigma \subset \partial\Omega$ and $D_\sigma v = |v_K - v_L|$ otherwise, with $\sigma = K|L$.

2.2.2 Weak solution and main result

Definition 2.2.1. *A weak solution of the chemotaxis-growth problem* (2.5)–(2.6) *is a pair of nonnegative functions* $(u,c) \in L^2(0,T;H^1(\Omega))^2$ *which verifies the following identities for all test functions* $\psi \in \mathcal{D}(\Omega \times [0,T))$ *:*

$$\int_0^T \int_\Omega \left(u\,\frac{\partial\psi}{\partial t} - \nabla u\cdot\nabla\psi + \chi\, u\nabla c\cdot\nabla\psi + f(u)\psi \right) dxdt + \int_\Omega u_0\, \psi(x,0)\, dx = 0, \qquad (2.12)$$

$$\int_0^T \int_\Omega \nabla c\cdot\nabla\psi \; dxdt = \int_0^T \int_\Omega (u-c)\psi\, dxdt. \qquad (2.13)$$

To prove the convergence of the scheme, the strong compactness of the finite volume solution is necessary, and it will be proved in Section 2.5. The proof requires the definition of an approximate gradient. Therefore, we begin by defining $T_{K,\sigma}$, which is the cell formed from the extremities of σ , x_K and x_L if $\sigma = K|L \not\subset \partial\Omega$, and from the extremities of σ and x_K if $\sigma \subset \partial\Omega$ (see Figure 2.1).

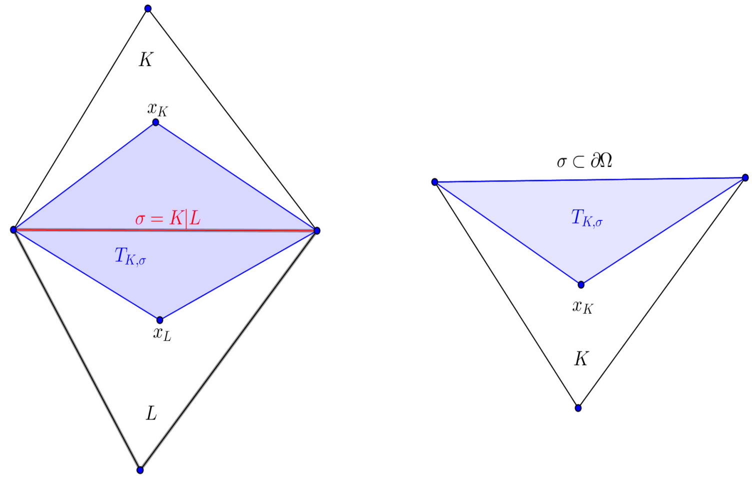

Figure 2.1 Construction of $T_{K,\sigma}$ for two-dimensional triangular mesh

Following [13], we define dv_δ the approximation of ∇v by

$$dv_\delta(x,t) = \frac{\mathrm{m}(\sigma)}{\mathrm{m}(T_{K,\sigma})} Dv^{n+1}_{K,\sigma}\ \nu_{K,\sigma}, \quad x \in T_{K,\sigma},\ t \in [t^n, t^{n+1}), \tag{2.14}$$

for all $K \in \mathcal{T}$ and $n = 0, ..., N-1$, where $\nu_{K,\sigma}$ denotes the unit normal on σ which is outward to K.

We now recall the discrete Gronwall inequality which will be used in the a priori analysis in Section 2.4.

Lemma 2.2.1 (Discrete Gronwall inequality). *Let M, b_n and a_n (for integers $n \geq 0$) be nonnegative numbers such that $a_{n+1} \leq M + \sum_{k=0}^{n} a_k b_k$. Then*

$$a_{n+1} \leq M \exp\left(\sum_{k=0}^{n} b_k\right).$$

Before we state our main result, we note that the nonnegativity of the solution is necessary to prove the convergence of the scheme. Consequently, the time step should satisfy the following condition (see Proposition 2.3.2) : for $0 \leq n \leq N$, we have

$$a\Delta t_n \left(\left(\max_{K\in\mathcal{T}} u^n_K\right)^{\alpha-1} - b\right) \leq 1. \tag{2.15}$$

We present now the main result of this chapter :

Theorem 2.2.1. *Assume that $u_0 \in L^2(\Omega)$ and $u_0 \geq 0$. Assume that the condition* (2.15) *is satisfied. Then the finite volume solution (u_δ, c_δ) to the scheme* (2.7)–(2.9) *converges to (u,c) as $\delta \to 0$, where (u,c) is a weak solution of* (2.5)–(2.6) *in the sense of Definition 2.2.1 .*

2.3 Existence, uniqueness, and nonnegativity of finite volume solutions

In this section, we prove existence, uniqueness and nonnegativity of solutions to the proposed schemes.

Proposition 2.3.1. *Assume that $u_0 \geq 0$, and assume that the following time step condition is fulfilled*

$$\Delta t < \frac{1}{ab}. \tag{2.16}$$

Then there exists a unique solution $\{(u_K^{n+1}, c_K^{n+1}),\ K \in \mathcal{T},\ 0 \leq n \leq N\}$ to the scheme (2.10),(2.8)–(2.9)*, where for all $K \in \mathcal{T}$, $0 \leq n \leq N$, we have $u_K^{n+1} \geq 0$ and $c_K^{n+1} \geq 0$.*

Proof. Let φ be a bijection from the finite set $\mathcal{T}$ to the finite set $\{1, 2, ..., l\}$, where l is the number of elements in $\mathcal{T}$ (i.e the number of control volumes). The aim of this function is to number the control volumes. The decoupled scheme (2.10),(2.8)–(2.9) can be written equivalently as

$$A^n\, U^{n+1} = F^n, \tag{2.17}$$

$$BC^{n+1} = G^n, \tag{2.18}$$

where for all $K \in \mathcal{T}$, the vectors of (2.17)–(2.18) are defined by : $U_{\varphi(K)}^{n+1} = u_K^{n+1}$, $F_{\varphi(K)}^n = \mathrm{m}(K) u_K^n / \Delta t_n$, $C_{\varphi(K)}^{n+1} = c_K^{n+1}$ and $G_{\varphi(K)}^n = \mathrm{m}(K) u_K^n$. The associated system matrices are defined by :

$$\begin{aligned}
A_{\varphi(K),\varphi(K)}^n &= \mathrm{m}(K)/\Delta t_n + \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma \not\subset \partial\Omega}} \tau_\sigma \left(1 + \chi\, (D c_{K,\sigma}^{n+1})^+\right) + \mathrm{m}(K)\, a((u_K^n)^{\alpha-1} - b), \\
A_{\varphi(K),\varphi(L)}^n &= -\tau_\sigma \left(1 + \chi\, (D c_{K,\sigma}^{n+1})^-\right), \quad \text{if } L \in \mathcal{N}(K), \\
A_{\varphi(K),\varphi(L)}^n &= 0, \quad \text{otherwise.}
\end{aligned}$$

and by

$$\begin{aligned}
B_{\varphi(K),\varphi(K)} &= \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma \not\subset \partial\Omega}} \tau_\sigma + \mathrm{m}(K), \\
B_{\varphi(K),\varphi(L)} &= -\tau_\sigma, \quad \text{if } L \in \mathcal{N}(K), \\
B_{\varphi(K),\varphi(L)} &= 0, \quad \text{otherwise.}
\end{aligned}$$

To prove the Proposition 2.3.1, we proceed by induction on n. Taking into account that $u_0 \geq 0$ and $G^0 \geq 0$, the first step of the induction (i.e. $n = 0$) can be shown similarly as the inductive step.

Let us assume now, that U^n and C^n are known, and that $U^n \geq 0$ and $C^n \geq 0$, we have for all $K \in \mathcal{T}$:

$$|B_{\varphi(K),\varphi(K)}| - \sum_{\substack{L \in \mathcal{T} \\ L \neq K}} |B_{\varphi(K),\varphi(L)}| = \mathrm{m}(K) > 0,$$

then the matrix B is strictly diagonally dominant by rows, and since B has nonpositive off-diagonal and positive diagonal entries, B is a nonsingular M-matrix. Finally, by the induction supposition, we have $G^n \geq 0$, which implies that there exists a unique vector C^{n+1} satisfying (2.18), such that $C^{n+1} \geq 0$.

On the other hand, since (2.16) holds and $u_K^n \geq 0$ for all $K \in \mathcal{T}$ (by the induction assumption), we can simply verify that A^n has nonpositive off-diagonal and positive diagonal entries. Furthermore, since for all $\sigma = K|L \in \mathcal{E}_K$, $Dc_{L,\sigma}^{n+1} = -Dc_{K,\sigma}^{n+1}$ i.e. $(Dc_{L,\sigma}^{n+1})^+ = (Dc_{K,\sigma}^{n+1})^-$, and under the condition (2.16), we have for all $K \in \mathcal{T}$

$$|A^n_{\varphi(K),\varphi(K)}| - \sum_{\substack{L\in\mathcal{T}\\ L\neq K}} |A^n_{\varphi(L),\varphi(K)}| = \mathrm{m}(K)/\Delta t_n + \mathrm{m}(K)\, a((u_K^n)^{\alpha-1} - b) > 0.$$

We conclude that A^n is a nonsingular M-matrix, then there exists a unique vector U^{n+1} which satisfies (2.17), moreover, since F^n and $(A^n)^{-1}$ are positive (A^n is an M-matrix), we have $U^{n+1} \geq 0$, which finishes the proof.

Proposition 2.3.2. *Suppose that $u_0 \geq 0$, and assume that* (2.15) *holds. Then there exists a unique nonnegative solution $\{(u_K^{n+1}, c_K^{n+1}),\ K \in \mathcal{T},\ 0 \leq n \leq N\}$ to the scheme* (2.7)–(2.9).

Proof. The proof is similar to that of Proposition 2.3.1. Using the same notations as the previous proof , the equation (2.7) can be written as

$$A^n\, U^{n+1} = F^n, \tag{2.19}$$

where for all $K \in \mathcal{T}$,

$$\begin{aligned}
A^n_{\varphi(K),\varphi(K)} &= \mathrm{m}(K)/\Delta t_n + \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma\not\subset\partial\Omega}} \tau_\sigma \left(1 + \chi\,(Dc_{K,\sigma}^{n+1})^+\right),\\
A^n_{\varphi(K),\varphi(L)} &= -\tau_\sigma \left(1 + \chi\,(Dc_{K,\sigma}^{n+1})^-\right), \quad \text{if}\, L \in \mathcal{N}(K),\\
A^n_{\varphi(K),\varphi(L)} &= 0, \quad \text{otherwise}.\\
F^n_{\varphi(K)} &= \mathrm{m}(K) u_K^n \left(1/\Delta t_n + a\left(b - (u_K^n)^{\alpha-1}\right)\right).
\end{aligned}$$

Following the guidelines of the previous proof, we can simply deduce that A^n is a nonsingular M-matrix unconditionally. And since (2.15) is satisfied and $u_K^n \geq 0$ (by induction supposition), F^n is nonnegative. Equation (2.8) is treated exactly in the same manner as in the proof of Proposition 2.3.1. Then, we obtain the existence of a nonnegative unique solution to the scheme (2.7)–(2.9).

Remark 2.3.1. *In contrast to [28] where a fully implicit scheme is used, the time step restrictions* (2.15) *and* (2.16) *ensuring stability of the schemes do no depend on the spatial mesh size h. In general, schemes with CFL condition become costly if the mesh is very fine.*

2.4 A priori estimates

In this section, we provide some a priori estimates for the approximate solution (u_δ, c_δ) of the scheme (2.7)–(2.9). We recall that Ω is an open bounded polygonal (resp. polyhedral) subset in $\mathbb{R}^d$, $d=2$ (resp. $d=3$), and we assume that $u_0 \in L^2(\Omega)$ with $u_0 \geq 0$. Since the arguments used in this section are based on the nonnegativity of the solution, we suppose also that (2.15) holds. Finally, we denote by C a generic positive constant which is always independent of δ, and which may take different values.

Proposition 2.4.1. *Let $\{(u_K^{n+1}, c_K^{n+1}),\ K \in \mathcal{T},\ 0 \leq n \leq N\}$ be the solution of the scheme (2.7)–(2.9). Then there exists a constant $C>0$, depending on Ω, T, a, b and $\|u_0\|_1$ such that*

$$\sum_{n=0}^{N}\sum_{K\in\mathcal{T}} \Delta t_n \ \mathrm{m}(K)\left|c_K^{n+1}\right|^2 + \sum_{n=0}^{N}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t_n \ \tau_\sigma \left|Dc_{K,\sigma}^{n+1}\right|^2 \leq C, \tag{2.20}$$

$$\sum_{n=0}^{N}\sum_{K\in\mathcal{T}} \Delta t_n \ \mathrm{m}(K)\left|u_K^{n}\right|^\alpha \leq C. \tag{2.21}$$

Proof. We begin by summing (2.7) over $K \in \mathcal{T}$, which gives

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(u_K^{n+1} - u_K^n\right) = ab\sum_{K\in\mathcal{T}} \Delta t_n \ \mathrm{m}(K) u_K^n - a\sum_{K\in\mathcal{T}} \Delta t_n \ \mathrm{m}(K)\left|u_K^n\right|^\alpha .$$

Then, summing over $n = 0, 1, ..., p$, with $p \leq N$, we find that

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K) \ u_K^{p+1} + a\sum_{n=0}^{p}\sum_{K\in\mathcal{T}} \Delta t_n \ \mathrm{m}(K)\left|u_K^n\right|^\alpha = ab\sum_{n=0}^{p}\sum_{K\in\mathcal{T}} \Delta t_n \ \mathrm{m}(K) \ u_K^n + \sum_{K\in\mathcal{T}} \mathrm{m}(K) \ u_K^0.$$

Using Lemma 2.2.1, we have

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K) \ u_K^{p+1} \leq \ \|u_0\|_1 \exp(abT),$$

which, together with the last equality, implies the estimate (2.21).

Now, multiplying the equation (2.8) by $\Delta t_n \ c_K^{n+1}$, summing over $K \in \mathcal{T}$, and using a summation by parts, we have

$$\frac{1}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t_n \ \tau_\sigma \left|Dc_{K,\sigma}^{n+1}\right|^2 + \sum_{K\in\mathcal{T}} \Delta t_n \ \mathrm{m}(K)\left|c_K^{n+1}\right|^2 = \sum_{K\in\mathcal{T}} \Delta t_n \ \mathrm{m}(K) c_K^{n+1} u_K^n.$$

Summing over $n = 0, 1, ..., N$, and using a Young inequality, it follows that

$$\sum_{n=0}^{N}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t_n \ \tau_\sigma \left|Dc_{K,\sigma}^{n+1}\right|^2 + \sum_{n=0}^{N}\sum_{K\in\mathcal{T}} \Delta t_n \ \mathrm{m}(K)\left|c_K^{n+1}\right|^2 \leq \sum_{n=0}\sum_{K\in\mathcal{T}} \Delta t_n \ \mathrm{m}(K)\left|u_K^n\right|^2 .$$

Since $\alpha \geq 3$, and since we have (2.21), we can use the Hölder inequality to control the right-hand side of the latter inequality

$$\sum_{n=0}^{N}\sum_{K\in\mathcal{T}} \Delta t_n \ \mathrm{m}(K)\left|u_K^n\right|^2 \leq |T\mathrm{m}(\Omega)|^{(\alpha-2)/\alpha}\left(\sum_{n=0}^{N}\sum_{K\in\mathcal{T}} \Delta t_n \ \mathrm{m}(K)\left|u_K^n\right|^\alpha\right)^{2/\alpha} .$$

This completes the proof.

Proposition 2.4.2. *Let $\{(u_K^{n+1}, c_K^{n+1}),\ K \in \mathcal{T},\ 0 \le n \le N\}$ be the solution of the scheme (2.7)–(2.9). There exists a constant $C > 0$, depending on Ω, T, a, b, χ and $\|u_0\|_2$ such that, for all $n = 0, 1, ..., p$, with $p \le N-1$:*

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|u_K^{p+1}\right|^2 + \sum_{n=0}^{p} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t_n\ \tau_\sigma \left|Du_{K,\sigma}^{n+1}\right|^2 \le C. \tag{2.22}$$

Proof. We begin by multiplying the equation (2.7) by $\Delta t_n\, u_K^{n+1}$, and summing over $K \in \mathcal{T}$, we find that

$$G_1 + G_2 = G_3 + G_4, \tag{2.23}$$

where

$$\begin{aligned}
G_1 &= \sum_{K\in\mathcal{T}} \mathrm{m}(K) u_K^{n+1} \left(u_K^{n+1} - u_K^n\right),\\
G_2 &= -\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t_n\ \tau_\sigma\ u_K^{n+1} Du_{K,\sigma}^{n+1},\\
G_3 &= -\chi \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \Delta t_n\ \tau_\sigma u_K^{n+1} \left((Dc_{K,\sigma}^{n+1})^+ u_K^{n+1} - (Dc_{K,\sigma}^{n+1})^- u_L^{n+1}\right),\\
G_4 &= ab \sum_{K\in\mathcal{T}} \Delta t_n\ \mathrm{m}(K)\, |u_K^n| \left|u_K^{n+1}\right| - a \sum_{K\in\mathcal{T}} \Delta t_n\ \mathrm{m}(K)\, |u_K^n|^\alpha \left|u_K^{n+1}\right|.
\end{aligned}$$

Using the inequality $a(a-b) \ge \dfrac{1}{2}\left(a^2 - b^2\right)$

$$G_1 \ge \frac{1}{2} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(\left|u_K^{n+1}\right|^2 - |u_K^n|^2\right).$$

Then, we perform a summation by parts on G_2. This yields,

$$G_2 = \frac{1}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t_n\ \tau_\sigma\ \left|Du_{K,\sigma}^{n+1}\right|^2.$$

To deal with the term G_3, we begin by establishing the inequality :

$$F_1 \le \frac{1}{2}\left(F_2 + F_3\right), \tag{2.24}$$

where

$$\begin{aligned}
F_1 &= \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \tau_\sigma \left((Dc_{K,\sigma}^{n+1})^+ u_K^{n+1} - (Dc_{K,\sigma}^{n+1})^- u_L^{n+1}\right)\left(u_L^{n+1} - u_K^{n+1}\right),\\
F_2 &= \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \tau_\sigma u_K^{n+1} (Dc_{K,\sigma}^{n+1}) \left(u_L^{n+1} - u_K^{n+1}\right),\\
F_3 &= \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \tau_\sigma u_L^{n+1} (Dc_{K,\sigma}^{n+1}) \left(u_L^{n+1} - u_K^{n+1}\right).
\end{aligned}$$

Since $Dc_{K,\sigma}^{n+1} = (Dc_{K,\sigma}^{n+1})^+ - (Dc_{K,\sigma}^{n+1})^-$, we have

$$F_1 - F_2 = \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \tau_\sigma (Dc_{K,\sigma}^{n+1})^- \left(u_L^{n+1} - u_K^{n+1}\right)\left(u_K^{n+1} - u_L^{n+1}\right) \leq 0,$$

$$F_1 - F_3 = \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \tau_\sigma (Dc_{K,\sigma}^{n+1})^+ \left(u_K^{n+1} - u_L^{n+1}\right)\left(u_L^{n+1} - u_K^{n+1}\right) \leq 0,$$

which gives (2.24).

Now, multiplying (2.8) by $\left|u_K^{n+1}\right|^2$, and sum by parts, we get

$$\frac{1}{2}\sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \tau_\sigma Dc_{K,\sigma}^{n+1} \left(\left|u_L^{n+1}\right|^2 - \left|u_K^{n+1}\right|^2\right) = \sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^{n+1}\right|^2 \left(u_K^n - c_K^{n+1}\right).$$

which implies that

$$\frac{1}{2}\sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \tau_\sigma \left(u_K^{n+1} + u_L^{n+1}\right) Dc_{K,\sigma}^{n+1} Du_{K,\sigma}^{n+1} \leq \sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^n\right|\left|u_K^{n+1}\right|^2. \tag{2.25}$$

A summation by parts on G_3 gives

$$G_3 = \frac{\chi}{2}\sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t_n\ \tau_\sigma \left((Dc_{K,\sigma}^{n+1})^+ u_K^{n+1} - (Dc_{K,\sigma}^{n+1})^- u_L^{n+1}\right) Du_{K,\sigma}^{n+1}.$$

From the above equality, (2.25) and (2.24), we deduce that

$$\begin{aligned} G_3 &\leq \frac{\chi\Delta t_n}{2}\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^n\right|\left|u_K^{n+1}\right|^2 \\ &\leq \frac{\chi\Delta t_n}{2}\left(\frac{1}{3}\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^n\right|^3 + \frac{2}{3}\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^{n+1}\right|^3\right) \end{aligned}$$

Using a Young inequality for G_4 we obtain

$$\begin{aligned} G_4 \leq \frac{ab}{2}&\left(\sum_{K\in\mathcal{T}} \Delta t_n\ \mathrm{m}(K)\left|u_K^n\right|^2 + \sum_{K\in\mathcal{T}} \Delta t_n\ \mathrm{m}(K)\left|u_K^{n+1}\right|^2\right) \\ &- a\sum_{K\in\mathcal{T}} \Delta t_n\ \mathrm{m}(K)\left|u_K^n\right|^\alpha \left|u_K^{n+1}\right|. \end{aligned}$$

Employing the previous inequalities in (2.23), and summing over $n = 0, 1, ..., p$, with $p \leq N-1$, we obtain

$$\begin{aligned} &\frac{1}{2}\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^{p+1}\right|^2 + \frac{1}{2}\sum_{n=0}^{p}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t_n\ \tau_\sigma \left|Du_{K,\sigma}^{n+1}\right|^2 \\ &+ a\sum_{n=0}^{p}\sum_{K\in\mathcal{T}} \Delta t_n\ \mathrm{m}(K)\left|u_K^n\right|^\alpha\left|u_K^{n+1}\right| \leq \frac{1}{2}\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^0\right|^2 \\ &+ \frac{\chi}{2}\sum_{n=0}^{p+1}\sum_{K\in\mathcal{T}} \Delta t_n \mathrm{m}(K)\left|u_K^n\right|^3 + ab\sum_{n=0}^{p+1}\sum_{K\in\mathcal{T}} \Delta t_n\ \mathrm{m}(K)\left|u_K^n\right|^2 \end{aligned}$$

From the fact that $u_0 \in L^2(\Omega)$, and since we have from (2.21) that u_δ is bounded in $L^\alpha(\Omega_T)$ ($\alpha \geq 3$), the right-hand side of the latter inequality is bounded, which finishes the proof.

2.5 Convergence of the scheme

This section is devoted to prove that the approximate solution (u_δ, c_δ) given by the scheme (2.7)–(2.9), converges to a weak solution of (2.5)–(2.6). We suppose naturally that the assumptions of Section 2.4 hold. The key of the proof is the discrete Aubin-Simon lemma [30, Theorem 3.4]. The proof is based also on Lemma 4.4 in [13], and uses some techniques from the proof of Proposition 4.1 in [28] and Section 5 in [7].

Proposition 2.5.1. *There exists a subsequence of $(u_\delta, c_\delta)_\delta$, not relabeled, and a function $(u,c) \in L^2(0,T;H^1(\Omega))^2$ such that as δ goes to zero we have*

$$\begin{aligned} u_\delta \to u \quad &\textit{strongly in } L^2(\Omega_T), \\ du_\delta \rightharpoonup \nabla u \quad &\textit{weakly in } L^2(\Omega_T), \\ c_\delta \rightharpoonup c, \quad dc_\delta \rightharpoonup \nabla c \quad &\textit{weakly in } L^2(\Omega_T). \end{aligned}$$

Proof. We define $\phi \in H^q(\Omega)$ with $q > 2$ and we denote by ϕ_K the average of ϕ in the control volume K. Multiplying equation (2.7) by ϕ_K, we have :

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(u_K^{n+1} - u_K^n\right)\phi_K = I_1 + I_2 + I_3,$$

where

$$\begin{aligned} I_1 &= \frac{\Delta t_n}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma D u_{K,\sigma}^{n+1} D\phi_{K,\sigma} \\ &\leq \frac{\Delta t_n}{2} \|\phi_\delta\|_{1,2,\mathcal{T}} \|u_\delta\|_{1,2,\mathcal{T}}. \end{aligned}$$

$$\begin{aligned} I_2 &= -\frac{\chi}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t_n\ \tau_\sigma \left((Dc_{K,\sigma}^{n+1})^+ u_K^{n+1} - (Dc_{K,\sigma}^{n+1})^- u_L^{n+1} \right) D\phi_{K,\sigma} \\ &\leq \frac{\chi}{2}\Delta t_n \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t_n\ \tau_\sigma u_K^{n+1} \left| Dc_{K,\sigma}^{n+1} \right| |D\phi_{K,\sigma}| \\ &\leq \frac{\chi}{2}\Delta t_n \|c_\delta\|_{1,2,\mathcal{T}} \|u_\delta\|_2 \|\phi_\delta\|_{1,\infty,\mathcal{T}}. \end{aligned}$$

$$\begin{aligned} I_3 &= \sum_{K\in\mathcal{T}} \Delta t_n\ \mathrm{m}(K)\, f\left(u_K^n\right)\phi_K \\ &\leq \Delta t_n \left(ab\left\|u_\delta^n\right\|_1 + a\left\|u_\delta^n\right\|_\alpha^\alpha\right)\|\phi_\delta\|_\infty . \end{aligned}$$

Now, using the Holder inequality, the estimates obtained in Proposition 2.4.1 and Proposition 2.4.2, and the appropriate Sobolev embedding, we obtain the following uniform estimate for the time translation of u_δ in $L^2\left(0,T;(H^q(\Omega))'\right)$

$$\int_0^{T-\tau} \int_\Omega \left(u_\delta(x,t+\tau) - u_\delta(x,t)\right)\phi(x,t)\,dx\,dt \leq C\,\tau \|\phi\|_{L^2(0,T;H^q(\Omega))},$$

where $\tau \in (0, \Delta t)$ and C is independent of δ. Since $H^1(\Omega) \hookrightarrow\hookrightarrow L^2(\Omega) \hookrightarrow (H^q(\Omega))'$), we conclude from the discrete Aubin-Simon lemma [30, Theorem 3.4] (see also Remark 4 in [30]), that there exists a subsequence of $(u_\delta)_\delta$, not relabeled, such that

$$u_\delta \to u \quad \text{strongly in } L^2(\Omega_T),$$

and since du_δ is bounded in $L^2(\Omega_T)$ from (2.22), there exists a subsequence of $(du_\delta)_\delta$ not relabeled, such that

$$du_\delta \rightharpoonup k \quad \text{weakly in } L^2(\Omega_T),$$

where from [13, Lemma 4.4] we have $k = \nabla u$.

Finally, from the boundedness of c_δ in $L^2(0,T;H^1(\Omega))$ and using [13, Lemma 4.4], there exists a subsequence, which is not relabeled, such that

$$c_\delta \rightharpoonup c, \quad dc_\delta \rightharpoonup \nabla c \quad \text{weakly in } L^2(\Omega_T).$$

Proposition 2.5.2. *The function* (u,c) *defined in Proposition 2.5.1 is a weak solution of* (2.5)–(2.6) *in the sense of Definition 2.2.1 .*

Proof. The proof is similar to that of Theorem 5.1 and Theorem 5.2 in [13], or that of Proposition 4.2 and Proposition 4.3 in [28]. We recall it here since it is important to understand why we consider that the straight line going through x_K and x_L must be orthogonal to $\sigma = K|L$.

Let $\psi \in \mathcal{D}(\Omega \times [0,T))$, we set

$$\begin{aligned} J_{10}(\delta) &= -\int_0^T \int_\Omega dc_\delta \cdot \nabla\psi \; dx\,dt, \\ J_{20}(\delta) &= \int_0^T \int_\Omega \left(\frac{u_\delta}{u_\delta + 1} - c_\delta\right) \psi \; dx\,dt, \end{aligned}$$

with $\lambda_1(\delta) = J_{10}(\delta) + J_{20}(\delta)$.

From the weak convergence of $(dc_\delta)_\delta$ to ∇c, the weak convergence of $(c_\delta)_\delta$ to c and the strong convergence of $(u_\delta)_\delta$ to u in $L^2(\Omega_T)$ with the continuity of the function $x \mapsto \frac{x}{x+1}$ if $x \geq 0$, we have when $\delta \to 0$

$$\lambda_1(\delta) \to -\int_0^T \int_\Omega \nabla c \cdot \nabla\psi \; dxdt + \int_0^T \int_\Omega \left(\frac{u}{u+1} - c\right) \psi \, dxdt.$$

Let $\psi_K^n = \psi(t^n, x_K)$ for all $K \in \mathcal{T}$ and $n = 0, 1, ..., N$. By multiplying the equation (2.8) by $\Delta t_n \, \psi_K^n$ and summing for K and n, we get

$$J_1(\delta) + J_2(\delta) = 0,$$

where

$$\begin{aligned} J_1(\delta) &= \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t_n \; \tau_\sigma \; Dc_{K,\sigma}^{n+1} \psi_K^n, \\ J_2(\delta) &= \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \Delta t_n \; \left(\frac{u_K^n}{u_K^n + 1} - \mathrm{m}(K) c_K^{n+1}\right) \psi_K^n. \end{aligned}$$

Now, we need to prove that $J_j(\delta) - J_{j0}(\delta) \to 0$ as $\delta \to 0$ for $j = 1, 2$.

For $j = 2$, we have

$$\begin{aligned} J_2(\delta) - J_{20}(\delta) = & \sum_{n=0}^{N-1} \sum_{K \in \mathcal{T}} \int_{t_n}^{t_{n+1}} \int_K \frac{u_K^{n+1}}{u_K^{n+1}+1} \left(\psi_K^{n+1} - \psi(t,x)\right) dxdt \\ & - \sum_{n=0}^{N-1} \sum_{K \in \mathcal{T}} \int_{t_n}^{t_{n+1}} \int_K c_K^{n+1} \left(\psi_K^n - \psi(t,x)\right) dxdt \\ & + \sum_{K \in \mathcal{T}} \Delta t_n \, \mathrm{m}(K) \frac{u_K^0}{u_K^0+1} \psi_K^0 - \int_{t_{N-1}}^{T} \int_\Omega \frac{u_\delta}{u_\delta+1} \psi \, dx \, dt. \end{aligned}$$

Since c_δ is uniformly bounded in $L^2(\Omega_T)$ and since ψ is smooth, we have

$$|J_2(\delta) - J_{20}(\delta)| \le \|\psi\|_{C^1(\Omega_T)} \left(\|c_\delta\|_1 + T\mathrm{m}(\Omega)\right) \delta + C\Delta t \to 0, \quad \text{as } \delta \to 0,$$

where $C > 0$ is a constant.

Using a summation by parts, we have

$$J_1(\delta) = -\frac{\Delta t_n}{2} \sum_{n=0}^{N-1} \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma D c_{K,\sigma}^{n+1} D\psi_{K,\sigma}^n.$$

From the definition of J_{10}, we can write

$$J_{10}(\delta) = -\frac{1}{2} \sum_{n=0}^{N-1} \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \int_{t_n}^{t_{n+1}} \int_{T_{K,\sigma}} dc_\delta \cdot \nabla\psi \, dxdt.$$

The last two equalities imply that

$$\begin{aligned} J_1(\delta) - J_{10}(\delta) = & -\frac{1}{2} \sum_{n=0}^{N-1} \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \mathrm{m}(\sigma) \, D c_{K,\sigma}^{n+1} \times \\ & \int_{t_n}^{t_{n+1}} \left(\frac{D\psi_{K,\sigma}^n}{\mathrm{d}(x_K, x_L)} - \frac{1}{\mathrm{m}(T_{K,\sigma})} \int_{T_{K,\sigma}} \nabla\psi \cdot \nu_{K,\sigma} \, dx \right) dt. \end{aligned}$$

Now, using the fact that the straight line going through x_K and x_L is orthogonal to $\sigma = K|L$, we have $x_K - x_L = \mathrm{d}(x_K, x_L) \, \nu_{K,\sigma}$, which implies that

$$\begin{aligned} \frac{D\psi_{K,\sigma}^n}{\mathrm{d}(x_K, x_L)} & = \nabla\psi(t^n, x_L) \, \nu_{K,\sigma} + \mathcal{O}(h) \\ & = \nabla\psi(t, x) \, \nu_{K,\sigma} + \mathcal{O}(\delta), \quad (t, x) \in [t^n, t^{n+1}) \times T_{K,\sigma}. \end{aligned}$$

It follows then that there exists a constant $C > 0$ such that

$$\left| \int_{t_n}^{t_{n+1}} \left(\frac{D\psi_{K,\sigma}^n}{\mathrm{d}(x_K, x_L)} - \frac{1}{\mathrm{m}(T_{K,\sigma})} \int_{T_{K,\sigma}} \nabla\psi \cdot \nu_{K,\sigma} \, dx \right) dt \right| \le C\Delta t \delta.$$

Since the mesh $\mathcal{T}$ is regular, there exists a constant $C > 0$, only depending on the geometry of $\mathcal{T}$, such that for $\sigma = K|L$

$$\mathrm{d}\left(x_K, x_L\right) \mathrm{m}(\sigma) \leq C \mathrm{m}\left(T_{K,\sigma}\right).$$

which implies that

$$\begin{aligned}
\mathrm{m}(\sigma)\left|c_L^{n+1} - c_K^{n+1}\right| &= \sqrt{\tau_\sigma}\left|c_L^{n+1} - c_K^{n+1}\right| \sqrt{\mathrm{d}\left(x_K, x_L\right) \mathrm{m}(\sigma)} \\
&\leq \sqrt{\tau_\sigma}\left|c_L^{n+1} - c_K^{n+1}\right| \sqrt{C \mathrm{m}\left(T_{K,\sigma}\right)}
\end{aligned}$$

Finally, using the Cauchy-Schwarz inequality with the estimate (2.20), we can conclude that

$$|J_1(\delta) - J_{10}(\delta)| \leq C\delta \to 0, \quad \text{as } \delta \to 0,$$

where C is a positive constant.

We define now

$$\begin{aligned}
E_{10}(\delta) &= -\int_0^T \int_\Omega u_\delta \frac{\partial \psi}{\partial t} dx\, dt - \int_\Omega u_\delta(x,0)\, \psi(x,0)\, dx \\
E_{20}(\delta) &= \int_0^T \int_\Omega du_\delta \cdot \nabla \psi \; dx\, dt \\
E_{30}(\delta) &= -\chi \int_0^T \int_\Omega u_\delta \; dc_\delta \cdot \nabla \psi \; dx\, dt \\
E_{40}(\delta) &= -\int_0^T \int_\Omega f(u_\delta) \psi \; dx\, dt
\end{aligned}$$

and $\lambda_2(\delta) = -\left(E_{10}(\delta) + E_{20}(\delta) + E_{30}(\delta) + E_{40}(\delta)\right)$.

Multiplying the scheme (2.7) by $\Delta t_n \psi_K^n$ and summing for K and n, we have

$$E_1(\delta) + E_2(\delta) + E_3(\delta) + E_4(\delta) = 0$$

where

$$\begin{aligned}
E_1(\delta) &= \sum_{n=0}^{N-1} \sum_{K \in \mathcal{T}} \mathrm{m}(K)\left(u_K^{n+1} - u_K^n\right) \psi_K^n, \\
E_2(\delta) &= -\sum_{n=0}^{N-1} \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \Delta t_n \; \tau_\sigma \; D u_{K,\sigma}^{n+1} \psi_K^n, \\
E_3(\delta) &= \chi \sum_{n=0}^{N-1} \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \Delta t_n \; \tau_\sigma \left((D c_{K,\sigma}^n)^+ u_K^{n+1} - (D c_{K,\sigma}^n)^- u_L^{n+1}\right) \psi_K^n, \\
E_4(\delta) &= -\sum_{n=0}^{N-1} \sum_{K \in \mathcal{T}} \Delta t_n \; \mathrm{m}(K) f\left(u_K^n\right) \psi_K^n.
\end{aligned}$$

From the weak convergence of $(dc_\delta)_\delta$ to ∇c, $(du_\delta)_\delta$ to ∇u, and the strong convergence of $(u_\delta)_\delta$ to u in $L^2\left(\Omega_T\right))$, with the continuity of the function f, we have when $\delta \to 0$

$$\lambda_2(\delta) \to \int_0^T \int_\Omega \left(u \frac{\partial \psi}{\partial t} - \nabla u \cdot \nabla \psi + \chi\, u \nabla c \cdot \nabla \psi + f(u) \psi\right) dx dt + \int_\Omega u_0\, \psi(x,0)\, dx.$$

As before, we need to prove that $E_j(\delta) - E_{j0}(\delta) \to 0$ as $\delta \to 0$ for $j = 1, 2, 3, 4$. The proof of $E_2(\delta) - E_{20}(\delta) \to 0$ as $\delta \to 0$ is exactly similar to the proof of $J_1(\delta) - J_{10}(\delta) \to 0$ as $\delta \to 0$. $E_4(\delta) - E_{40}(\delta) \to 0$ is obvious from the continuity of the function f, the strong convergence of the sequence $(u_\delta)_\delta$, and the estimate (2.21).

For $j = 1$, we have

$$\begin{aligned}
E_1(\delta) &= \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \mathrm{m}(K) u_K^{n+1} \left(\psi_K^n - \psi_K^{n+1}\right) - \sum_{K\in\mathcal{T}} \mathrm{m}(K) u_K^0 \psi_K^0 \\
&= \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \int_{t_n}^{t_{n+1}} \int_K u_K^{n+1} \frac{\partial \psi}{\partial t}(t, x_K) dx dt - \sum_{K\in\mathcal{T}} \int_K u_K^0 \psi(0, x_K) dx, \\
E_{10}(\delta) &= \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \int_{t_n}^{t_{n+1}} \int_K u_K^{n+1} \frac{\partial \psi}{\partial t} dx dt - \sum_{K\in\mathcal{T}} \int_K u_K^0 \psi(0, x) dx.
\end{aligned}$$

Then, from the regularity of ψ, we obtain

$$|E_1(\delta) - E_{10}(\delta)| \le \left((T+1)\, \mathrm{m}(\Omega) \|u_\delta\|_{L^2(\Omega_T)} \|\psi\|_{C^2(\Omega_T)}\right) h \to 0, \quad \text{as } \delta \to 0.$$

Now, using the identity (for $\sigma = K|L$) :

$$\left(Dc_{K,\sigma}^{n+1}\right)^+ u_K^{n+1} - \left(Dc_{K,\sigma}^{n+1}\right)^- u_L^{n+1} = -\frac{1}{2}\left|Dc_{K,\sigma}^{n+1}\right| Du_{K,\sigma}^{n+1} + \frac{1}{2}\left(u_K^{n+1} + u_L^{n+1}\right) Dc_{K,\sigma}^{n+1}.$$

and using a summation by parts, we can write $E_3(\delta) = E_{31}(\delta) + E_{32}(\delta)$, with

$$\begin{aligned}
E_{31}(\delta) &= \frac{\chi}{4} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t_n\, \tau_\sigma \left|Dc_{K,\sigma}^{n+1}\right| Du_{K,\sigma}^{n+1} D\psi_{K,\sigma}^n, \\
E_{32}(\delta) &= -\frac{\chi}{2} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t_n\, \tau_\sigma u_K^{n+1} Dc_{K,\sigma}^{n+1} D\psi_{K,\sigma}^n.
\end{aligned}$$

For $E_{30}(\delta)$, it is easy to see that

$$\begin{aligned}
E_{30}(\delta) &= -\frac{1}{2} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \int_{t_n}^{t_{n+1}} \int_{T_{K,\sigma}} u_\delta\, dc_\delta \cdot \nabla\psi\, dx dt \\
&= E_{310}(\delta) + E_{320}(\delta).
\end{aligned}$$

with

$$\begin{aligned}
E_{310}(\delta) &= -\frac{1}{2} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \int_{t_n}^{t_{n+1}} \int_{S_{L,\sigma}} Du_{K,\sigma}^{n+1}\, dc_\delta \cdot \nabla\psi\, dx dt \\
E_{320}(\delta) &= -\frac{1}{2} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \int_{t_n}^{t_{n+1}} \int_{T_{K,\sigma}} u_K^{n+1}\, dc_\delta \cdot \nabla\psi\, dx dt.
\end{aligned}$$

where $S_{L,\sigma} = L \cap T_{K,\sigma}$. Since we have

$$E_{32}(\delta) - E_{320}(\delta) = -\frac{\chi}{2} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \mathrm{m}\,(\sigma)\, u_K^{n+1} Dc_{K,\sigma}^{n+1} \times$$
$$\int_{t_n}^{t_{n+1}} \left(\frac{D\psi_{K,\sigma}^n}{\mathrm{d}\,(x_K, x_L)} - \frac{1}{\mathrm{m}\,(T_{K,\sigma})} \int_{T_{K,\sigma}} \nabla\psi \cdot \nu_{K,\sigma}\, dx \right) dt,$$

the proof of $E_{32}(\delta) - E_{320}(\delta) \to 0$ as $\delta \to 0$ is similar to that of $J_2(\delta) - J_{20}(\delta) \to 0$ as $\delta \to 0$. Using the estimates (2.20) and (2.22), we get

$$|E_{31}(\delta)| \leq \frac{\chi\, h}{4} \|\psi\|_{C^1(\Omega_T)} \|c_\delta\|_{1,2,\mathcal{T}} \|u_\delta\|_{1,2,\mathcal{T}} \to 0, \quad \text{as } \delta \to 0.$$

Finally, using the definition (2.14) and since $S_{L,\sigma} \subset T_{K,\sigma}$, it follows from the Cauchy-Schwarz inequality.

$$\begin{aligned} |E_{310}(\delta)| &\leq \frac{\chi}{2} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t_n \mathrm{d}\,(x_K, x_L)\, \tau_\sigma \left| Du_{K,\sigma}^{n+1} \right| \left| Dc_{K,\sigma}^{n+1} \right| \|\psi\|_{C^1(\Omega_T)} \\ &\leq \chi\, h\, \|\psi\|_{C^1(\Omega_T)} \|c_\delta\|_{1,2,\mathcal{T}} \|u_\delta\|_{1,2,\mathcal{T}} \to 0, \quad \text{as } \delta \to 0. \end{aligned}$$

This ends the proof of Theorem 2.2.1.

2.6 Numerical results

In this section, we show some numerical results for the chemotaxis-growth model (2.5)–(2.6) in two space dimensions, using the finite volume schemes presented in this chapter. To compute the discrete solution, we need only to solve two linear systems at each time step. For example, in the case of (2.7)–(2.8), and at each time step, we begin by solving (2.8) to compute c_K^{n+1} and then, we compute u_K^{n+1} from (2.7). The same iterative procedure is used for the scheme (2.10),(2.8).

The schemes used in this section include the parameters omitted by the assumption (2.4). In all numerical experiments, unless otherwise stated, we assume that $\alpha = 3$, $a = 1$ and $b = 1$. We adopt also the following parameter values taken from [1] : $D_u = 0.0625$, $D_v = 1$, $\mu = 1$, $\beta = 32$. The solution is very sensitive to the choice of parameters. It is also sensitive to the spatial domain Ω and to the initial conditions (see [70]). We set $\Omega = (-4, 4)^2$, and we use a uniform mesh grid 100×100.

For the initial cell density, we consider that an inoculum of cells is added to the center of the domain at $t = 0$, so we write

$$u_0(x) = \begin{cases} \text{random}[0, 1] & \text{if } \|x\|_2 < 0.5, \\ 0 & \text{otherwise .} \end{cases}$$

Numerical simulations using an other pattern formation model with such initial datum can be found in [90].

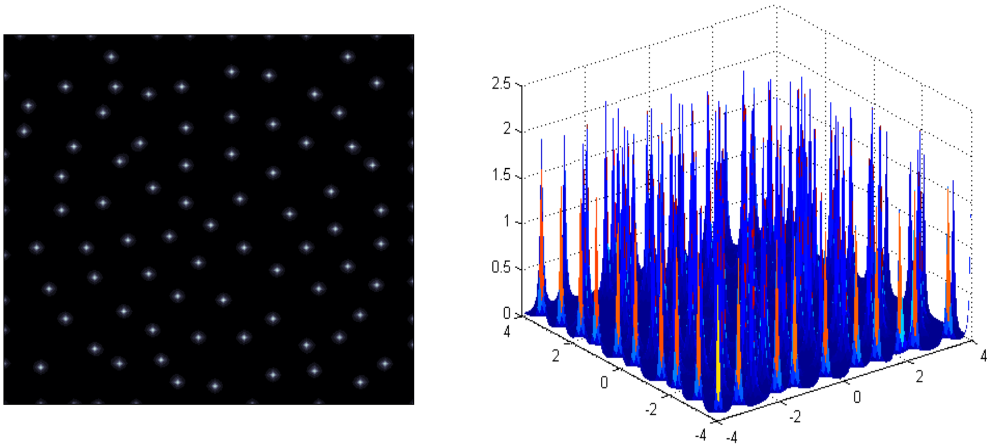

Figure 2.2 Solution (u) at t=100 using the scheme (2.7)–(2.8), with the step size control (2.26).

A uniform time step $\Delta t = 0.1$ is a reasonable step size for such problem. It is clear that for this value, the condition (2.16) holds true, but in the other side, we have verified throughout the calculations that the time step condition (2.15) may be violated for certain values of χ. So, to ensure the nonnegativity of the solution and the convergence of the scheme (2.7)–(2.8), we use the following step size control :

$$\Delta t_n = \begin{cases} \Delta t & \text{if } \lambda^2 - 1 \leq 0, \\ \min\left(\Delta t, \ \dfrac{1}{a\left(\lambda^2 - b\right)}\right) & \text{otherwise,} \end{cases} \tag{2.26}$$

with $\lambda = \max\limits_{K \in \mathcal{T}} u_K^n$.

In Figure 2.2 and Figure 2.3, the cell density is plotted at $t = 100$ for a chemotactic sensitivity $\chi = 60$, using the schemes (2.7)–(2.8) and (2.10),(2.8) respectively. We see that both numerical solutions have no negative values, moreover, spot pattern formations are shown in whole domain.

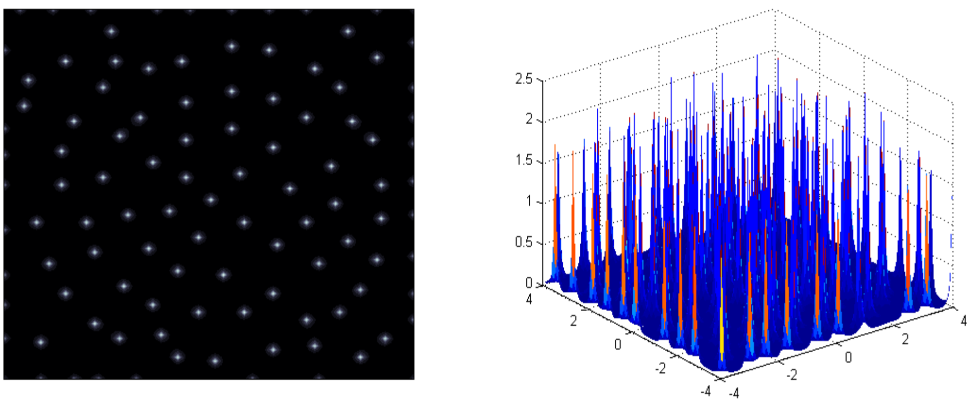

Figure 2.3 Solution (u) at t=100 using the scheme (2.10),(2.8), with a uniform time step $\Delta t = 0.1$.

We recall that in the case of the classical Patlak-keller-Segel (i.e. $a = 0$) and, for such

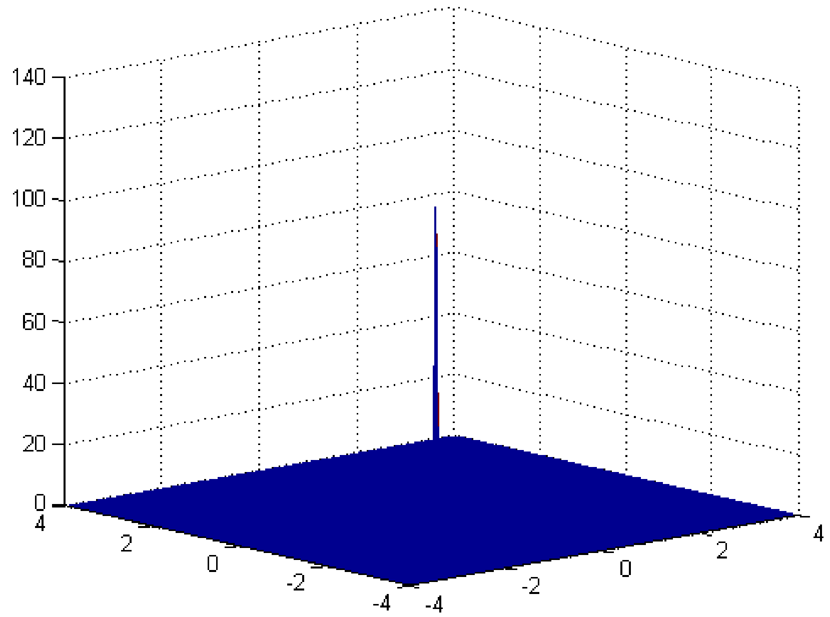

Figure 2.4 Solution (u) when $a = 0$ at $t = 100$, using a finite volume scheme with a uniform time step $\Delta t = 0.1$.

initial conditions, only one peak is observed (see Figure 2.4), which means that all cells migrate towards one point.

In Figure 2.5, we examine the influence of the parameter χ on the solution using the scheme (2.7)–(2.8) with the step size control (2.26). For sufficiently small values of χ, the numerical solution converges to the homogeneous steady state $\overline{u} = 1$. This fact is illustrated in Figure 2.5 (left top) in the case $\chi = 4$. When $\chi = 8.8$, we observe that hexagonal pattern (also called honeycomb pattern) appears. This pattern is similar to that of *Vibrio* bacteria (see Figure 1.1). Then, the solution changes its structure to chaotic spots pattern with the increasing of χ. Finally, we note that the symmetry of these spots increases for high values of χ (see Figure 2.5 (right bottom)).

The results obtained by the scheme (2.10),(2.8) for the same values of the chemotactic sensitivity χ are presented in Figure 2.6. As we can see, the numerical solution exhibits the same patterns shown in Figure 2.5.

χ		8.8	30	60	100
Scheme (2.7)–(2.8)	Number of time steps	1000	1171	1058	1047
	CPU time (s)	504.38	610.69	551.53	536.65
Scheme (2.10),(2.8)	Number of time steps	1000	1000	1000	1000
	CPU time (s)	523.95	523.99	520.78	530.58

Table 2.1 Number of time steps and computational time needed to reach the final time $T = 100$, for the two schemes (2.7)–(2.8) and (2.10),(2.8), in the case when $\Delta t = 0.1$.

In Table 2.1 and Table 2.2, we depict the number of time steps and CPU time needed to reach the final time $T = 100$, for different values of χ. As in the previous numerical examples,

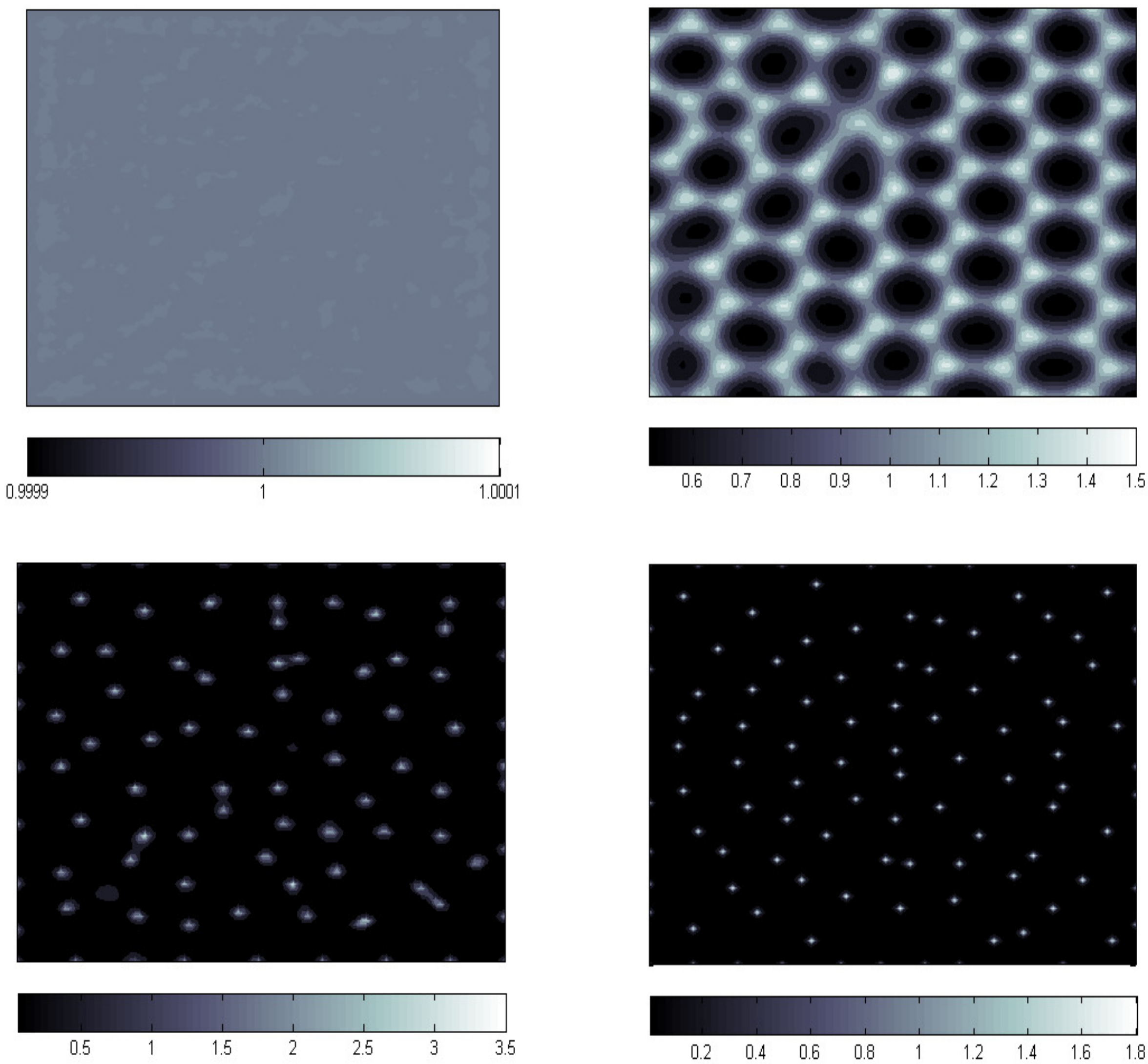

Figure 2.5 Cell density by the scheme (2.7)–(2.8) at $t = 100$, for $\chi = 4$ (left top), $\chi = 8.8$ (right top), $\chi = 30$ (left bottom), $\chi = 100$ (right bottom).

the results of Table 2.1 are computed for $\Delta t = 0.1$, and by using the step size control (2.26) to ensure the stability of the scheme (2.7)–(2.8). For Table 2.2, the time step is chosen to be $\Delta t = 0.5$.

We observe from Table 2.1 that for small and high values of χ ($\chi = 8.8$, 60, 100), there is not a significant difference between our two schemes. The case when $\chi = 30$ is slightly different, since the condition (2.15) is somehow restrictive and, consequently, the scheme (2.10),(2.8) is a little faster than (2.7)–(2.8). However, for the large time step size $\Delta t = 0.5$, Table 2.2 shows that the scheme (2.7)–(2.8) becomes much more costly than (2.10),(2.8), the condition (2.15) becoming too restrictive. We note that if the step size control (2.26) is not used to adjust the time step, negative values appear at certain times and the solver breaks down very early.

Finally, in Figure 2.7, we show temporal evolution of the solution for $\chi = 60$. As we can see, the cells propagate throughout the whole domain in a wave-like manner. These numerical results seem in good agreement with the experimental data reported in [12] (see Figure 1.3).

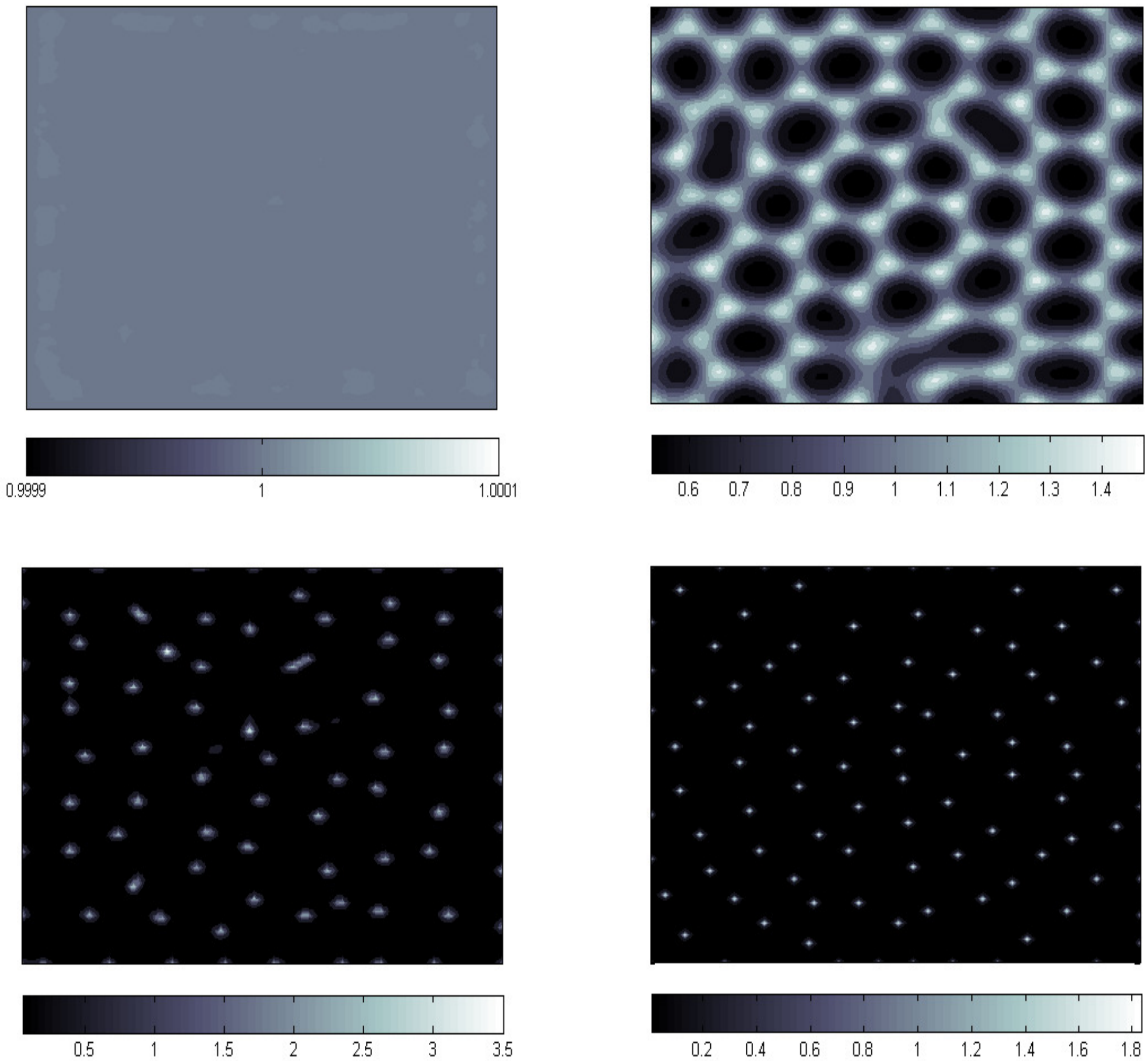

Figure 2.6 Cell density by the scheme(2.10),(2.8) at $t = 100$, for $\chi = 4$ (left top), $\chi = 8.8$ (right top), $\chi = 30$ (left bottom), $\chi = 100$ (right bottom).

χ		8.8	30	60	100
Scheme (2.7)–(2.8)	Number of time steps	203	1070	592	435
	CPU time (s)	109.79	560.45	303.20	218.86
Scheme (2.10),(2.8)	Number of time steps	200	200	200	200
	CPU time (s)	106.57	108.97	105.94	108.75

Table 2.2 Number of time steps and computational time needed to reach the final time $T = 100$, for the two schemes (2.7)–(2.8) and (2.10),(2.8), in the case when $\Delta t = 0.5$.

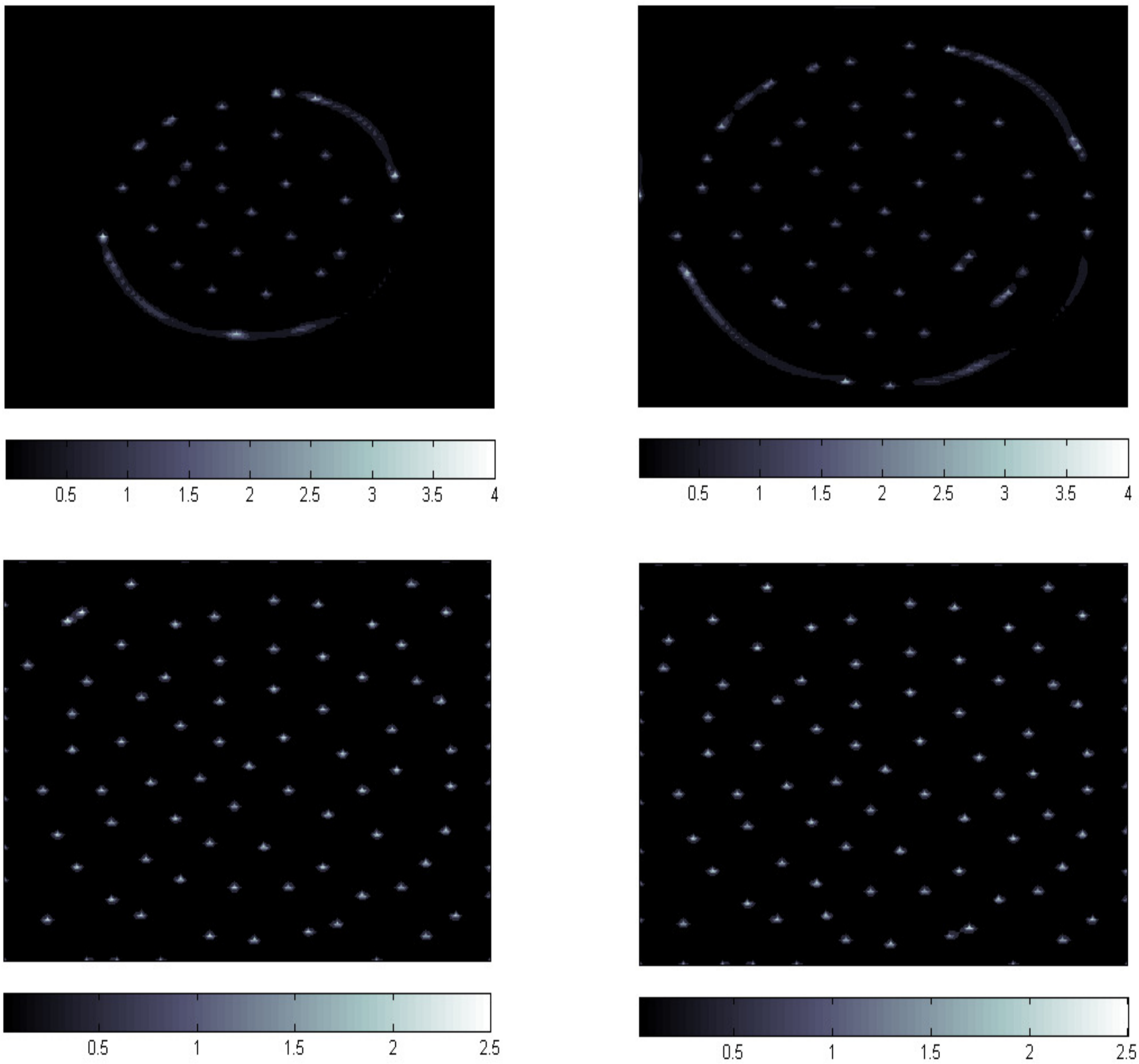

Figure 2.7 Cell density for $\chi = 60$ by the scheme (2.7)–(2.8) at $t = 10$ (left top), $t = 12$ (right top), $t = 20$ (left bottom), $t = 60$ (right bottom).

Chapter 3

A finite volume scheme for a chemotaxis model arising in embryology

Sommaire

3.1 Introduction

The mechanism which generates integumental patterns in animals has sparked the interest of many scientists. The first approach proposed to produce such biological patterns is the reaction diffusion theory, due to Turing in 1952 [89]. In embryology, several applications of reaction diffusion systems in pattern formation problems can be found in literature (see, e.g., [57, 56, 54]).

In [67], Oster and Murray discussed a cell-chemotaxis model involving motile cells that respond to a chemoattractant secreted by the cells themselves. In its dimensionless form, the model reads

$$\begin{cases} \partial_t u = \mu\Delta u - a\nabla\cdot(\,u\nabla c), \\ \partial_t c = \Delta c + \dfrac{u}{u+1} - c\,, \end{cases} \tag{3.1}$$

where μ and a are positive constants, u is the cell density and c is the concentration of chemoattractant.

The above system is based on the Keller-Segel model (2.1), which is the most popular model for chemotaxis. The migration of cells is assumed to be governed by Fickian diffusion and chemotaxis, and the mass of cells is conserved. The chemoattractant is assumed also to diffuse, but it increases with cell density in Michaelis-Menten way and undergoes decay through simple degradation.

In [58], Murray *et al.* suggest that the presented cell-chemotaxis model is an appropriate mechanism for the formation of stripe patterns on the dorsal integument of embryonic and hatchling alligators (*Alligator mississippiensis*) (see Figure 1.4). These skin pigment patterns is associated with the density of melanocyte cells : melanocytes are abundant in the regions where the black stripes appear, and are insufficient in the regions of the white stripes. The formation of these regions is a result of a chemoattractant secretion which starts at the anterior end and progresses towards the tail. The system (3.1) is known to produce propagating pattern of standings peaks and troughs in cell density in the case of one-dimensional space. This patterning process was numerically and analytically investigated by Myerscough and Murray in [60].

In this chapter, we are concerned with the following parabolic-elliptic system

$$\begin{cases} \partial_t u = \mu\Delta u - a\nabla\cdot(\,u\nabla c) & \text{in } \Omega_T, \\ 0 = \Delta c + \dfrac{u}{u+1} - c & \text{in } \Omega_T, \end{cases} \tag{3.2}$$

with the homogeneous Neumann boundary and initial conditions

$$\nabla u\cdot\nu = \nabla c\cdot\nu = 0 \quad \text{on } \partial\Omega\times(0,T), \quad u(.,0) = u_0 \quad \text{in } \Omega, \tag{3.3}$$

where $\Omega_T := \Omega\times(0,T)$, $\Omega\subset\mathbb{R}^2$ is an open bounded subset with sufficiently regular boundary $\partial\Omega$, $T>0$ is a fixed time and ν denotes the outward unit-normal on the boundary $\partial\Omega$. As we can see, the system (3.2) is a simpler version of the original model (3.1) in which the second equation of the system is elliptic, using the reasonable assumption that the diffusion of the chemoattractant is much faster than that of the cell. Such a model was theoretically discussed in [18], where the existence and uniqueness of global solution are established.

As mentioned in Chapter 2, Filbet established the convergence of a fully implicit finite volume scheme for the parabolic-elliptic version of the Keller-Segel model (2.1), but only for a sufficiently small and positive initial cell density. This result is in good agreement with the behavior of the solution in the continuous case, when the solution of the problem may blow-up in finite time, and only an $L^1(\Omega)$-boundedness of u is known to exist [61](unless one imposes a threshold on $\|u_0\|_{L^1(\Omega)}$). So, the discrete a priori estimates needed to prove the convergence of the finite volume solution cannot be established for all values of the initial cell density. However, it is proved in [7], that by adding an additional cross-diffusion term to the chemoattractant equation (which prevents the blow-up [36]), we do not need this restriction to obtain the convergence of the scheme. The same remark is valid for the volume-filling Keller-Segel model [5, 15, 42].

This chapter is devoted to the numerical analysis of a semi-implicit finite volume scheme for the problem (3.2)–(3.3). We show that if we modify the simplified Keller-Segel model such that the cells produce the chemoattractant in Michaelis-Menten way (which is the case of (3.2)), we do not need to impose a smallness condition to the initial datum to establish the convergence of the finite volume scheme. Then, the convergence of the approximate solution can be obtained for all nonnegative initial cell density $u_0 \in L^2(\Omega)$, without any time step restriction.

This chapter is organized as follows. In the next section, we present our finite volume scheme to approximate the solution of (3.2)–(3.3) and state our main result. In Section 3.3, we prove the existence and uniqueness of the solution of the proposed scheme. Positivity preservation and mass conservation are also shown in this section. A priori estimates are given in Section 3.4. In Section 3.5, we use these estimates to prove that the approximate solution converges to a weak solution of the studied model. Finally, numerical simulations are performed in Section 3.6.

3.2 Finite volume scheme and main result

3.2.1 Finite volume approximation

Let $\Omega \subset \mathbb{R}^2$ be an open bounded polygonal subset. We consider in this chapter an admissible finite volume mesh of Ω, denoted by $\mathcal{T}$, which verifies the properties of Definition 5.1 in [26] (see also Définition 1.3.1). We will also use the same notations related to the spatial discretization and discrete norms as in Section 2.2.1.

Let $0 = t_0 < t_1 < ... < t_N = T$ be a uniform partition of $(0, T)$ with $N \in \mathbb{N}$ and $t_n = n\Delta t$ for $n = 0, ..., N$.

We denote by h the maximal size (diameter) of the control volumes included in $\mathcal{T}$, and we define

$$\delta = \max(\Delta t, h).$$

We assume that the mesh satisfies this additional constraint : there exists $\xi > 0$ such that

$$\mathrm{d}(x_K, \sigma) \geq \xi\, \mathrm{d}(x_K, x_L), \tag{3.4}$$

which is specially needed to apply the discrete Gagliardo-Nirenberg-Sobolev inequality (see Lemma 3.2.1).

The proposed finite volume scheme for the problem (3.2)–(3.3) is given by the following discrete equations :

for all $K \in \mathcal{T}$ and $n = 0, ..., N$

$$\begin{aligned} &\mathrm{m}(K)\frac{u_K^{n+1} - u_K^n}{\Delta t} - \mu \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma D u_{K,\sigma}^{n+1} \\ &+ a \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \tau_\sigma \left((Dc_{K,\sigma}^{n+1})^+ u_K^{n+1} - (Dc_{K,\sigma}^{n+1})^- u_L^{n+1} \right) = 0, \end{aligned} \tag{3.5}$$

$$- \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma\, Dc_{K,\sigma}^{n+1} = \mathrm{m}(K)\, \frac{u_K^n}{u_K^n + 1} - \mathrm{m}(K)\, c_K^{n+1}, \tag{3.6}$$

with the compatible initial condition

$$u_K^0 = \frac{1}{\mathrm{m}(K)} \int_K u_0(x)\,dx, \tag{3.7}$$

and the following notations

$$z^+ = \max(z,0),\ z^- = \max(-z,0),$$

and for all $\sigma \in \mathcal{E}_K$

$$Dv_{K,\sigma}^n = \begin{cases} 0 & \text{for } \sigma \subset \partial\Omega, \\ v_L^n - v_K^n & \text{otherwise },\ \sigma = K|L. \end{cases}$$

The terms u_K^n and c_K^n denote respectively the approximations of $\frac{1}{\mathrm{m}(K)} \int_K u(x,t^n)\,dx$ and $\frac{1}{\mathrm{m}(K)} \int_K c(x,t^n)\,dx$. We use a linearized implicit discretization in time which leads to a decoupled linear system i.e. : at each time step, we need only to solve two decoupled linear systems.

We define u_δ, c_δ, the finite volume approximations of u and c by :

$$u_\delta(x,t) = u_K^{n+1}, \quad c_\delta(x,t) = c_K^{n+1}, \quad x \in K,\ t \in [t^n, t^{n+1}). \tag{3.8}$$

3.2.2 Definitions and main result

We define a weak solution of the system (3.2) with boundary and initial conditions (3.3) as follows :

Definition 3.2.1. *A weak solution of the initial-boundary value problem* (3.2)–(3.3) *is a pair of nonnegative functions* $(u,c) \in L^2(0,T;H^1(\Omega))^2$ *which verify the following identities for all test functions* $\psi \in \mathcal{D}(\Omega \times [0,T))$ *:*

$$\int_0^T \int_\Omega (u\,\partial_t\psi - \mu\nabla u\cdot\nabla\psi + a\,u\nabla c\cdot\nabla\psi)\ dxdt + \int_\Omega u_0\,\psi(x,0)\,dx = 0, \tag{3.9}$$

$$\int_0^T \int_\Omega \nabla c\cdot\nabla\psi\ dxdt = \int_0^T \int_\Omega \left(\frac{u}{u+1} - c\right)\psi\,dxdt. \tag{3.10}$$

We now recall the discrete Gagliardo-Nirenberg-Sobolev inequality (see [8]), which will be useful to establish a priori estimates in Section 3.4.

Lemma 3.2.1. *Let Ω be an open bounded polyhedral domain of $\mathbb{R}^d$, $d \geq 2$. Let $\mathcal{T}$ a mesh satisfying* (3.4) *and $v \in X(\mathcal{T})$.*

- *If $1 \leq p < d$, let $1 \leq s \leq m \leq p^* = pd/(d-p)$,*
- *If $p \geq d$, let $1 \leq s \leq m < +\infty$.*

Then there exists a constant $C > 0$ only depending on p, s, m, d and Ω such that

$$\|\, v \,\|_m \leq \frac{C}{\xi^{(p-1)\theta/p}} \,\|\, v \,\|_{1,p,\mathcal{T}}^{\theta} \,\|\, v \,\|_s^{1-\theta},\ \forall v \in X(\mathcal{T}),$$

where θ is defined by

$$\theta = \frac{1/s - 1/m}{1/s + 1/d - 1/p}$$

Now, we define approximations of the gradients of u and c. To this end, we begin by defining $T_{K,\sigma}$, which is the cell formed from the vertices of σ, x_K and x_L if $\sigma = K|L \not\subset \partial\Omega$, and from the vertices of σ and x_K if $\sigma \subset \partial\Omega$.

Following [13], we define the discrete gradient dv_δ which is the approximations of ∇v by

$$dv_\delta(x,t) = \frac{\mathrm{m}(\sigma)}{\mathrm{m}(T_{K,\sigma})}\, Dv_{K,\sigma}^{n+1}\ \nu_{K,\sigma}, \quad x \in T_{K,\sigma},\ t \in [t^n, t^{n+1}),$$

for all $K \in \mathcal{T}$ and $n = 0, ..., N-1$, where $\nu_{K,\sigma}$ denotes the unit normal on σ which is outward to K.

We present now our main result :

Theorem 3.2.1. *Let $\mathcal{T}$ be an admissible discretization of Ω. Assume that* (3.4) *is fulfilled and that $u_0 \in L^2(\Omega)$ with $u_0 \geq 0$. Then the finite volume approximation (u_δ, c_δ) given by the scheme* (3.5)–(3.7) *converges as $\delta \to 0$ to (u, c) which is a weak solution of* (3.2)–(3.3) *in the sense of Definition 3.2.1 .*

3.3 Existence and uniqueness of a discrete solution

In this section, we prove existence and uniqueness of the solution of the proposed scheme. We show also that the scheme is mass conserving and positivity preserving.

Proposition 3.3.1. *Assume that $u_0 \geq 0$. Then, the scheme* (3.5)–(3.7) *admits a unique solution $\{(u_K^{n+1}, c_K^{n+1}),\ K \in \mathcal{T},\ n = 0, ..., N-1\}$, which satisfies the following properties*

$$u_K^{n+1} \geq 0 \quad \textit{and} \quad c_K^{n+1} \geq 0 \quad \textit{for all } K \in \mathcal{T},\ n = 0, ..., N-1, \tag{3.11}$$

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\, u_K^n = \sum_{K\in\mathcal{T}} \mathrm{m}(K)\, u_K^0 = \|u_0\|_{L^1(\Omega)}, \quad \textit{for all } n = 0, ..., N. \tag{3.12}$$

Proof. Let φ be a bijection from the finite set $\mathcal{T}$ to the finite set $\{1, 2, ..., l\}$, where l is the number of elements in $\mathcal{T}$ (i.e the number of control volumes). The aim of this function is to number the control volumes.

Now, for all $n = 0, ..., N-1$, we define the following l-dimensional vectors :

$$\begin{aligned} U_{\varphi(K)}^{n+1} &= u_K^{n+1}, \quad F_{\varphi(K)}^n = \mathrm{m}(K) u_K^n/\Delta t, \\ C_{\varphi(K)}^{n+1} &= c_K^{n+1}, \quad G_{\varphi(K)}^n = \mathrm{m}(K)\frac{u_K^n}{u_K^n+1}. \end{aligned}$$

We also define for $n = 0, ..., N-1$, the matrix $A^n = \left(A_{i,j}^n\right)_{l\times l}$ by

$$\begin{aligned} A_{\varphi(K),\varphi(K)}^n &= \mathrm{m}(K)/\Delta t + \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma\not\subset\partial\Omega}} \tau_\sigma \left(\mu + a\,(Dc_{K,\sigma}^{n+1})^+\right), \\ A_{\varphi(K),\varphi(L)}^n &= -\tau_\sigma \left(\mu + a\,(Dc_{K,\sigma}^{n+1})^-\right), \quad \text{if } L \in \mathcal{N}(K),\ \sigma = K|L, \\ A_{\varphi(K),\varphi(L)}^n &= 0, \quad \text{otherwise.} \end{aligned}$$

And we define the matrix $B = (B_{i,j})_{l\times l}$ by

$$\begin{aligned} B_{\varphi(K),\varphi(K)} &= \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma\not\subset\partial\Omega}} \tau_\sigma + \mathrm{m}(K),\\ B_{\varphi(K),\varphi(L)} &= -\tau_\sigma, \quad \text{if } L\in\mathcal{N}(K),\ \sigma = K|L,\\ B_{\varphi(K),\varphi(L)} &= 0, \quad \text{otherwise.} \end{aligned}$$

The decoupled scheme (3.5)–(3.7) can be then written equivalently as

$$A^n\,U^{n+1} = F^n, \tag{3.13}$$

$$BC^{n+1} = G^n. \tag{3.14}$$

To prove the existence of a discrete solution satisfying (3.11) , we proceed by induction on n. For $n = 0$, $u_K^0 \geq 0$ is given by the initial condition (3.7) for each $K\in\mathcal{T}$. Then, the existence and nonnegativity of u_K^1 and c_K^1 for all $K\in\mathcal{T}$ can be shown similarly as the inductive step.

Let us assume now, that U^n and C^n are known, and that $U^n\geq 0$ and $C^n\geq 0$. We have for all $K\in\mathcal{T}$:

$$|B_{\varphi(K),\varphi(K)}| - \sum_{\substack{L\in\mathcal{T}\\ L\neq K}} |B_{\varphi(K),\varphi(L)}| = \mathrm{m}(K) > 0,$$

then the matrix B is strictly diagonally dominant by rows, and since B has nonpositive off-diagonal and positive diagonal entries, B is a nonsingular M-matrix. Finally, by the induction supposition, we have $G^n\geq 0$, which implies that there exists a unique vector C^{n+1} satisfying (3.14), such that $C^{n+1}\geq 0$.

On the other hand, since for all $\sigma = K|L\in\mathcal{E}_K$, $Dc_{L,\sigma}^{n+1} = -Dc_{K,\sigma}^{n+1}$ i.e. $(Dc_{L,\sigma}^{n+1})^+ = (Dc_{K,\sigma}^{n+1})^-$, we have for all $K\in\mathcal{T}$

$$|A^n_{\varphi(K),\varphi(K)}| - \sum_{\substack{L\in\mathcal{T}\\ L\neq K}} |A^n_{\varphi(L),\varphi(K)}| = \mathrm{m}(K)/\Delta t > 0.$$

Then A^n is strictly diagonal dominant by columns, moreover since A^n has nonpositive off-diagonal and positive diagonal entries, we conclude that A^n is a nonsingular M-matrix, then there exists a unique vector U^{n+1} which satisfies (3.13). Furthermore F^n and $(A^n)^{-1}$ are positive (A^n is an M-matrix), which implies that $U^{n+1}\geq 0$. This establishes (3.11).

Now, summing (3.5) over $K\in\mathcal{T}$, we have :

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(u_K^{n+1} - u_K^n\right) = 0.$$

Then, summing over $n = 0, 1, ..., p$, with $p\leq N-1$ we get (3.12) :

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\,u_K^{p+1} = \sum_{K\in\mathcal{T}} \mathrm{m}(K)\,u_K^0 = \|u_0\|_{L^1(\Omega)}, \quad \text{for all } p = 0, ..., N-1\,,$$

which ends the proof.

3.4 A priori estimates

In this section, we assume that $u_0 \in L^2(\Omega)$ with $u_0 \geq 0$. We denote by C a generic positive constant which is always independent of δ, and which may take different values.

Proposition 3.4.1. *Let* $\{(u_K^{n+1}, c_K^{n+1}),\ K \in \mathcal{T},\ n = 0, ..., N-1\}$ *be the solution of the scheme* (3.5)–(3.7)*. Then, we have for all* $n = 0, ..., N-1$, $K \in \mathcal{T}$

$$c_K^{n+1} \leq 1, \tag{3.15}$$

$$\sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma \left| Dc_{K,\sigma}^{n+1} \right|^2 \leq \mathrm{m}\,(\Omega)\,. \tag{3.16}$$

Proof. Now, for all $n = 0, ..., N-1$, we fix K such that $c_K^{n+1} = \max\{c_L^{n+1}\}_{L \in \mathcal{T}}$. Multiplying the equation (3.6) by $\left(c_K^{n+1} - 1\right)^+$, we get

$$-\sum_{\sigma \in \mathcal{E}_K} \tau_\sigma Dc_{K,\sigma}^{n+1} \left(c_K^{n+1} - 1\right)^+ \leq \mathrm{m}(K) \left(1 - c_K^{n+1}\right) \left(c_K^{n+1} - 1\right)^+.$$

The right-hand side of the above inequality is clearly nonpositive. Moreover, according to the choice of K, we have $Dc_{K,\sigma}^{n+1} \leq 0$ for all $\sigma \in \mathcal{E}_K$, which implies that

$$-\sum_{\sigma \in \mathcal{E}_K} \tau_\sigma Dc_{K,\sigma}^{n+1} \left(c_K^{n+1} - 1\right)^+ \geq 0.$$

Thus, we can deduce that $\left(c_K^{n+1} - 1\right) \leq 0$, which implies the estimate (3.15).

The estimate (3.16), is obtained by multiplying the equation (3.6) by c_K^{n+1}, summing over $K \in \mathcal{T}$, using a summation by parts, and applying the Young inequality on the right hand side :

$$\frac{1}{2} \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma \left| Dc_{K,\sigma}^{n+1} \right|^2 + \frac{1}{2} \sum_{K \in \mathcal{T}} \mathrm{m}(K) \left| c_K^{n+1} \right|^2 \leq \frac{1}{2} \mathrm{m}(\Omega).$$

This concludes the proof.

Lemma 3.4.1. *Let* $\{(u_K^{n+1}, c_K^{n+1}),\ K \in \mathcal{T},\ n = 0, ..., N-1\}$ *be the solution of the scheme* (3.5)–(3.7)*. Assume that* (3.4) *is fulfilled. Then, there exists a constant* $C > 0$ *only depending on* Ω, T,u_0, a, μ *and* ξ *such that*

$$\sum_{n=0}^{N} \sum_{K \in \mathcal{T}} \Delta t\ \mathrm{m}(K) \left| u_K^n \right|^2 \leq C. \tag{3.17}$$

Proof. Performing the following change of variable :

$$\tilde{u}_K^n = u_K^n + 1 \quad \text{for all } K \in \mathcal{T},\ n = 0, ..., N, \tag{3.18}$$

the scheme (3.5)–(3.7) becomes

$$
\begin{aligned}
&\mathrm{m}(K)\frac{\tilde{u}_K^{n+1}-\tilde{u}_K^{n}}{\Delta t} - \mu\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma D\tilde{u}_{K,\sigma}^{n+1}\\
&+a\sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}}\tau_\sigma\left((Dc_{K,\sigma}^{n+1})^{+}\tilde{u}_K^{n+1}-(Dc_{K,\sigma}^{n+1})^{-}\tilde{u}_L^{n+1}\right)-a\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma Dc_{K,\sigma}^{n+1}=0, \qquad (3.19)\\
&-\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma\, Dc_{K,\sigma}^{n+1}=\mathrm{m}(K)\,\frac{\tilde{u}_K^{n}-1}{\tilde{u}_K^{n}}-\mathrm{m}(K)\,c_K^{n+1}, \qquad (3.20)\\
&\tilde{u}_K^{0}=\frac{1}{\mathrm{m}(K)}\int_K u_0(x)+1\,dx, \qquad (3.21)
\end{aligned}
$$

and from (3.12), it is clear that

$$
\sum_{K\in\mathcal{T}}\mathrm{m}(K)\,\tilde{u}_K^{n}=\sum_{K\in\mathcal{T}}\mathrm{m}(K)\,\tilde{u}_K^{0}=\|u_0+1\|_{L^1(\Omega)},\quad \text{for all } n=0,...,N. \qquad (3.22)
$$

From (3.20) we can easily see that

$$
-\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma\, Dc_{K,\sigma}^{n+1}\geq -\mathrm{m}(K)\,c_K^{n+1}.
$$

Using the above inequality in (3.19), we obtain

$$
\begin{aligned}
&\mathrm{m}(K)\frac{\tilde{u}_K^{n+1}-\tilde{u}_K^{n}}{\Delta t} - \mu\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma D\tilde{u}_{K,\sigma}^{n+1}\\
&\leq -a\sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}}\tau_\sigma\left((Dc_{K,\sigma}^{n+1})^{+}\tilde{u}_K^{n+1}-(Dc_{K,\sigma}^{n+1})^{-}\tilde{u}_L^{n+1}\right)+a\,\mathrm{m}(K)\,c_K^{n+1}.
\end{aligned}
$$

Let us now multiply the above inequality by $\Delta t\,\log\left(\tilde{u}_K^{n+1}\right)$, and summing over $K\in\mathcal{T}$, we find that

$$
I_1+I_2\leq I_3+I_4, \qquad (3.23)
$$

where

$$
\begin{aligned}
I_1&=\sum_{K\in\mathcal{T}}\mathrm{m}(K)\left(\tilde{u}_K^{n+1}-\tilde{u}_K^{n}\right)\log\left(\tilde{u}_K^{n+1}\right),\\
I_2&=-\mu\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K}\Delta t\,\tau_\sigma D\tilde{u}_{K,\sigma}^{n+1}\log\left(\tilde{u}_K^{n+1}\right),\\
I_3&=-a\sum_{K\in\mathcal{T}}\sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}}\Delta t\,\tau_\sigma\left((Dc_{K,\sigma}^{n+1})^{+}\tilde{u}_K^{n+1}-(Dc_{K,\sigma}^{n+1})^{-}\tilde{u}_L^{n+1}\right)\log\left(\tilde{u}_K^{n+1}\right),\\
I_4&=a\sum_{K\in\mathcal{T}}\Delta t\,\mathrm{m}(K)\,c_K^{n+1}\log\left(\tilde{u}_K^{n+1}\right).
\end{aligned}
$$

From the expression of I_1, we can see that

$$
\begin{aligned}
I_1=&\sum_{K\in\mathcal{T}}\mathrm{m}(K)\left(\tilde{u}_K^{n+1}\log\left(\tilde{u}_K^{n+1}\right)-\tilde{u}_K^{n}\log\left(\tilde{u}_K^{n}\right)\right)\\
&-\sum_{K\in\mathcal{T}}\mathrm{m}(K)\tilde{u}_K^{n}\left(\log\left(\tilde{u}_K^{n+1}\right)-\log\left(\tilde{u}_K^{n}\right)\right),
\end{aligned}
$$

and since we have from a Taylor expansion of log

$$\tilde{u}_K^n \left(\log\left(\tilde{u}_K^{n+1}\right) - \log\left(\tilde{u}_K^n\right)\right) \leq \tilde{u}_K^{n+1} - \tilde{u}_K^n,$$

we deduce using the mass conservation property (3.22) that

$$I_1 \geq \sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(\tilde{u}_K^{n+1}\log\left(\tilde{u}_K^{n+1}\right) - \tilde{u}_K^n \log\left(\tilde{u}_K^n\right)\right).$$

By an integration by parts on I_2, we get

$$I_2 = \frac{\mu}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma D\tilde{u}_{K,\sigma}^{n+1}\left(\log\left(\tilde{u}_L^{n+1}\right) - \log\left(\tilde{u}_K^{n+1}\right)\right).$$

Then, by a Taylor expansion of log we get

$$I_2 = \frac{\mu}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma \left|\frac{D\tilde{u}_{K,\sigma}^{n+1}}{\sqrt{\theta_\sigma^{n+1}}}\right|^2,$$

where $\theta_\sigma^{n+1} = t_\sigma \tilde{u}_K^{n+1} + (1-t_\sigma)\,\tilde{u}_L^{n+1}$ with $t_\sigma \in (0,1)$. Hence, using the fact that

$$\frac{\left|D\tilde{u}_{K,\sigma}^{n+1}\right|}{\sqrt{\theta_\sigma^{n+1}}} = \frac{\sqrt{\tilde{u}_K^{n+1}} + \sqrt{\tilde{u}_L^{n+1}}}{\sqrt{\theta_\sigma^{n+1}}}\left|D\left(\sqrt{\tilde{u}^{n+1}}\right)_{K,\sigma}\right| \geq \left|D\left(\sqrt{\tilde{u}^{n+1}}\right)_{K,\sigma}\right|,$$

we infer that

$$I_2 \geq \frac{\mu}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma \left|D\left(\sqrt{\tilde{u}^{n+1}}\right)_{K,\sigma}\right|^2.$$

Now, performing the same computations as in [28, p. 468], we get

$$I_3 \leq \frac{a}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma \theta_\sigma^{n+1} Dc_{K,\sigma}^{n+1}\left(\log\left(\tilde{u}_L^{n+1}\right) - \log\left(\tilde{u}_K^{n+1}\right)\right),$$

which leads using the Taylor expansion of log to

$$I_3 \leq \frac{a}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma Dc_{K,\sigma}^{n+1} D\tilde{u}_{K,\sigma}^{n+1}.$$

Multiplying (3.20) by $\Delta t\, \tilde{u}_{K,\sigma}^{n+1}$ and using an integration by parts, we can easily see that

$$\frac{1}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma Dc_{K,\sigma}^{n+1} D\tilde{u}_{K,\sigma}^{n+1} \leq \sum_{K\in\mathcal{T}} \Delta t\ \mathrm{m}(K)\tilde{u}_K^{n+1}.$$

From the two last inequalities, and due to the mass conservation (3.22), we deduce that

$$I_3 \leq a\,\Delta t \|u_0 + 1\|_{L^1(\Omega)}.$$

Next, using the $L^\infty(\Omega)$ bound (3.20), it is clear that

$$I_4 \leq a\,\Delta t \|u_0 + 1\|_{L^1(\Omega)}.$$

Collecting the estimates obtained for I_1, I_2, I_3 and I_4, and summing over $n = 0, ..., N-1$, we obtain

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\tilde{u}_K^N \log\left(\tilde{u}_K^N\right) + \frac{\mu}{2}\sum_{n=0}^{N-1}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\,\tau_\sigma \left|D\left(\sqrt{\tilde{u}^{n+1}}\right)_{K,\sigma}\right|^2$$
$$\leq 2a\,T\|u_0+1\|_{L^1(\Omega)} + \sum_{K\in\mathcal{T}} \mathrm{m}(K)\tilde{u}_K^0 \log\left(\tilde{u}_K^0\right).$$

Therefore, we have

$$\sum_{n=0}^{N-1}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\,\tau_\sigma \left|D\left(\sqrt{\tilde{u}^{n+1}}\right)_{K,\sigma}\right|^2 \leq C, \tag{3.24}$$

where $C > 0$ is a constant depending on Ω, T,u_0, a and μ .

Now, we denote by $\tilde{u}_\delta^{n+1} \in X(\mathcal{T})$ the function defined by : $\tilde{u}_\delta^{n+1}(x) = \tilde{u}_K^{n+1}$ for all $x \in K$. and $n = 0, ..., N-1$,

From Lemma 3.2.1, we have

$$\left\|\sqrt{\tilde{u}_\delta^{n+1}}\right\|_4 \leq \frac{C}{\xi^{1/4}} \left\|\sqrt{\tilde{u}_\delta^{n+1}}\right\|_{1,2,\mathcal{T}}^{1/2} \left\|\sqrt{\tilde{u}_\delta^{n+1}}\right\|_2^{1/2},$$

which implies that

$$\|\tilde{u}_\delta^{n+1}\|_2^2 \leq \frac{C}{\xi} \left\|\sqrt{\tilde{u}_\delta^{n+1}}\right\|_{1,2,\mathcal{T}}^2 \|\tilde{u}_\delta^{n+1}\|_1,$$

where $C > 0$ depends on Ω.

Finally, gathering the above inequality with (3.22), (3.24) and (3.18), we deduce that u_δ is bounded in $L^2(\Omega_T)$. This completes the proof of the lemma.

Proposition 3.4.2. *Let* $\{(u_K^{n+1}, c_K^{n+1}),\ K \in \mathcal{T},\ n = 0, ..., N-1\}$ *be the solution of the scheme* (3.5)–(3.7). *Then, there exists a constant* $C > 0$ *depending on* Ω, T,u_0, a, μ *and* ξ *such that, for all* $n = 0, 1, ..., p$, *with* $p \leq N-1$ *:*

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^{p+1}\right|^2 + \sum_{n=0}^{p}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\ \tau_\sigma \left|Du_{K,\sigma}^{n+1}\right|^2 \leq C. \tag{3.25}$$

Proof.

Multiplying the equation (3.5) by $\Delta t\, u_K^{n+1}$, and summing over $K \in \mathcal{T}$, we get

$$F_1 + F_2 = F_3, \tag{3.26}$$

where

$$F_1 = \sum_{K\in\mathcal{T}} \mathrm{m}(K)u_K^{n+1}\left(u_K^{n+1} - u_K^n\right),$$
$$F_2 = -\mu \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\ \tau_\sigma\ u_K^{n+1} Du_{K,\sigma}^{n+1},$$
$$F_3 = -a \sum_{K\in\mathcal{T}}\sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \Delta t\ \tau_\sigma u_K^{n+1}\left((Dc_{K,\sigma}^{n+1})^+ u_K^{n+1} - (Dc_{K,\sigma}^{n+1})^- u_L^{n+1}\right),$$

Using the inequality $a(a-b) \geq \frac{1}{2}\left(a^2-b^2\right)$

$$F_1 \geq \frac{1}{2} \sum_{K \in \mathcal{T}} \mathrm{m}(K) \left(\left|u_K^{n+1}\right|^2 - \left|u_K^n\right|^2 \right).$$

Next, we perform a summation by parts on F_2. This yields,

$$F_2 = \frac{\mu}{2} \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \Delta t \ \tau_\sigma \ \left| Du_{K,\sigma}^{n+1} \right|^2.$$

To deal with the term F_3, we begin by establishing the inequality :

$$E_1 \leq \frac{1}{2}\left(E_2 + E_3\right), \tag{3.27}$$

where

$$E_1 = \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \tau_\sigma \left((Dc_{K,\sigma}^{n+1})^+ u_K^{n+1} - (Dc_{K,\sigma}^{n+1})^- u_L^{n+1} \right) \left(u_L^{n+1} - u_K^{n+1} \right),$$

$$E_2 = \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \tau_\sigma u_K^{n+1} (Dc_{K,\sigma}^{n+1}) \left(u_L^{n+1} - u_K^{n+1} \right),$$

$$E_3 = \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \tau_\sigma u_L^{n+1} (Dc_{K,\sigma}^{n+1}) \left(u_L^{n+1} - u_K^{n+1} \right).$$

Since $Dc_{K,\sigma}^{n+1} = (Dc_{K,\sigma}^{n+1})^+ - (Dc_{K,\sigma}^{n+1})^-$, we have

$$E_1 - E_2 = \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \tau_\sigma (Dc_{K,\sigma}^{n+1})^- \left(u_L^{n+1} - u_K^{n+1} \right) \left(u_K^{n+1} - u_L^{n+1} \right) \leq 0,$$

$$E_1 - E_3 = \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \tau_\sigma (Dc_{K,\sigma}^{n+1})^+ \left(u_K^{n+1} - u_L^{n+1} \right) \left(u_L^{n+1} - u_K^{n+1} \right) \leq 0,$$

which gives (3.27).

Now, multiplying (3.6) by $\left|u_K^{n+1}\right|^2$, and sum by parts, we obtain

$$\frac{1}{2} \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \tau_\sigma Dc_{K,\sigma}^n \left(\left|u_L^{n+1}\right|^2 - \left|u_K^{n+1}\right|^2 \right) = \sum_{K \in \mathcal{T}} \mathrm{m}(K) \left|u_K^{n+1}\right|^2 \left(\frac{u_K^n}{u_K^n + 1} - c_K^n \right).$$

which leads to

$$\frac{1}{2} \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \tau_\sigma \left(u_K^{n+1} + u_L^{n+1} \right) Dc_{K,\sigma}^{n+1} Du_{K,\sigma}^{n+1} \leq \sum_{K \in \mathcal{T}} \mathrm{m}(K) \left|u_K^{n+1}\right|^2. \tag{3.28}$$

A summation by parts on F_3 gives

$$F_3 = \frac{a}{2} \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \Delta t \ \tau_\sigma \left((Dc_{K,\sigma}^{n+1})^+ u_K^{n+1} - (Dc_{K,\sigma}^{n+1})^- u_L^{n+1} \right) Du_{K,\sigma}^{n+1}.$$

From the above equality, (3.27) and (3.28), we deduce that

$$F_3 \leq \frac{a\,\Delta t}{2} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|u_K^{n+1}\right|^2 .$$

Employing the previous estimates for F_1, F_2 and F_3 in (3.26), and summing over $n = 0, 1, ..., p$, with $p \leq N-1$, we obtain

$$\begin{aligned} &\frac{1}{2}\sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|u_K^{p+1}\right|^2 + \frac{\mu}{2}\sum_{n=0}^{p}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\ \tau_\sigma \left|Du_{K,\sigma}^{n+1}\right|^2 \\ &\leq \frac{1}{2}\sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|u_K^0\right|^2 + \frac{a}{2}\sum_{n=0}^{p}\sum_{K\in\mathcal{T}} \Delta t\,\mathrm{m}(K) \left|u_K^{n+1}\right|^2 . \end{aligned} \tag{3.29}$$

It follows from Lemma 3.4.1 that the right hand side of the above inequality is bounded, which gives the desired result.

3.5 Convergence of the finite volume scheme

Proposition 3.5.1. *There exists a subsequence of $(u_\delta, c_\delta)_\delta$, not relabeled, and a function $(u, c) \in L^2(0, T; H^1(\Omega))^2$ such that as δ goes to zero we have*

$$\begin{aligned} u_\delta &\to u \quad \textit{strongly in } L^2(\Omega_T), \\ du_\delta &\rightharpoonup \nabla u \quad \textit{weakly in } L^2(\Omega_T), \\ c_\delta \rightharpoonup c, \quad dc_\delta &\rightharpoonup \nabla c \quad \textit{weakly in } L^2(\Omega_T). \end{aligned}$$

Proof. The proof will be omitted since it is similar to that of Proposition 2.5.1.

Proposition 3.5.2. *The function (u, c) constructed in Proposition 3.5.1 is a weak solution of* (3.2)–(3.3) *in the sense of Definition 3.2.1.*

Proof. The proof is similar to that of Proposition 2.5.2. The treatment of the term $\dfrac{u_K^n}{u_K^n + 1}$ is straightforward, since it is bounded for all $K \in \mathcal{T}$ and $n = 0, ..., N-1$.

3.6 Numerical results

We now show some numerical results in two dimensions with the proposed finite volume scheme. This section gives some insights on the effect of the initial condition on the behavior of the cell density given by the the system (3.2)–(3.3). In all numerical tests, we consider the problem in a rectangular domain $\Omega = (-9/2, 9/2) \times (-75, 75)$, discretized via a uniform mesh of 33750 control volumes. The time step is $\Delta t = 1$, and we take the following data used in [58, 60] : $\mu = 0.25$, $a = 2$.

Test 1. In this test, we consider the following initial condition :

$$u_0(x) = \begin{cases} 1 - \epsilon(x) & \text{if } x \in (-9/2, 9/2) \times (-75, -73), \\ 1 & \text{otherwise, with } x \in \Omega, \end{cases} \tag{3.30}$$

where $\epsilon(x)$ is a positive perturbation such that $\epsilon(x) < 1$.

The above initial condition is in concordance with the case of stripe patterning on the dorsal of the embryonic alligator. As mentioned in Section 3 in [58], the default colour of the embryo is the dark pigmented form, and the the white pigmentation appears first in the head region. The white stripes (see Figure 1.4) are the result of an insufficient density of Melanocyte cells u. Note that the long rectangular domain reflects the geometry of the alligator dorsal.

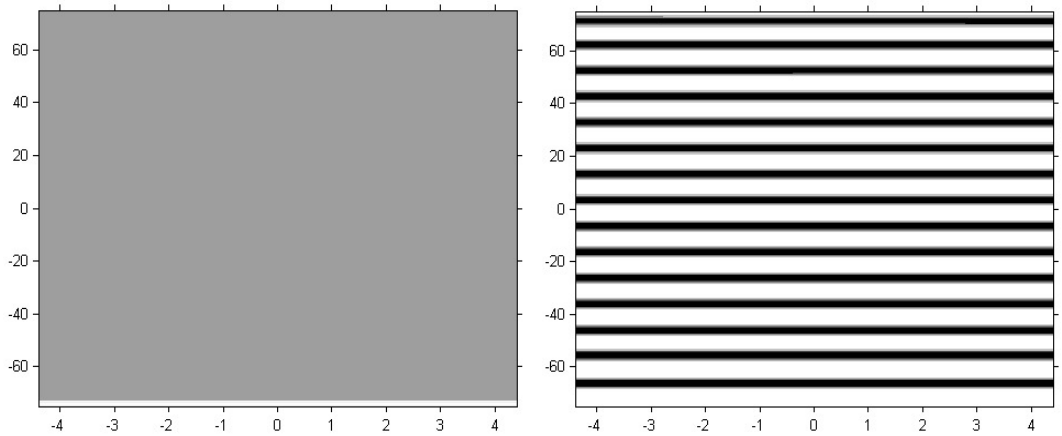

Figure 3.1 Initial condition of the cell density u_0 (3.30) (left), and the solution (u) at time $t = 800$ (right).

In Figure 3.1 (left), we show the initial condition (3.30). The strip patterning predicted by the model at $t = 800$ can be observed in Figure 3.1(left). In both figures, we supposed that in the region when $u = 2$, there is a sufficient Melanocytes to have a black colour. The 3D view of this patterning and its contour along the line $L = \{0\} \times (-75, 75)$ are shown in Figure 3.2.

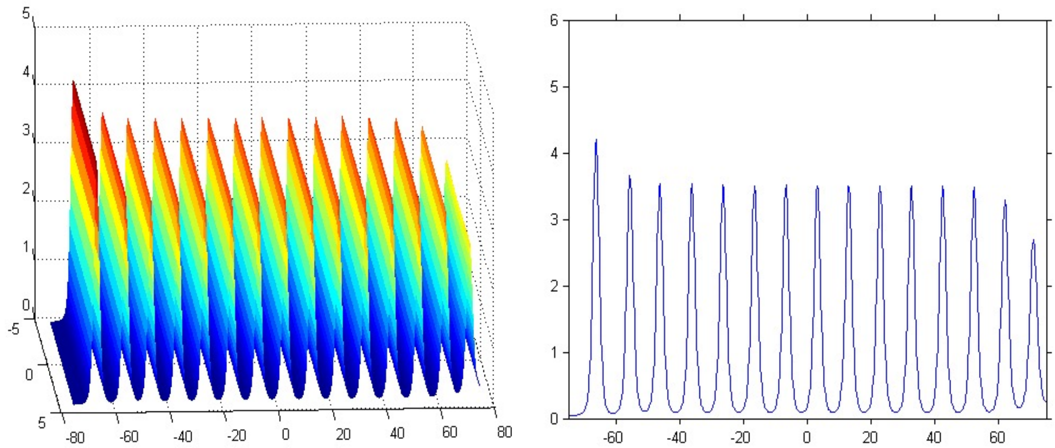

Figure 3.2 Three dimensional plot of the solution (u) from Figure 3.1 (right), and its contour along the line L (left).

In Figure 3.3, we show the evolution of the cell density by plotting the contour of u along the line L at different times. We can observe that the wave patterning moves to the right which reflects the fact that the pigment starts in the anterior part of the embryo and proceeds down the body in time.

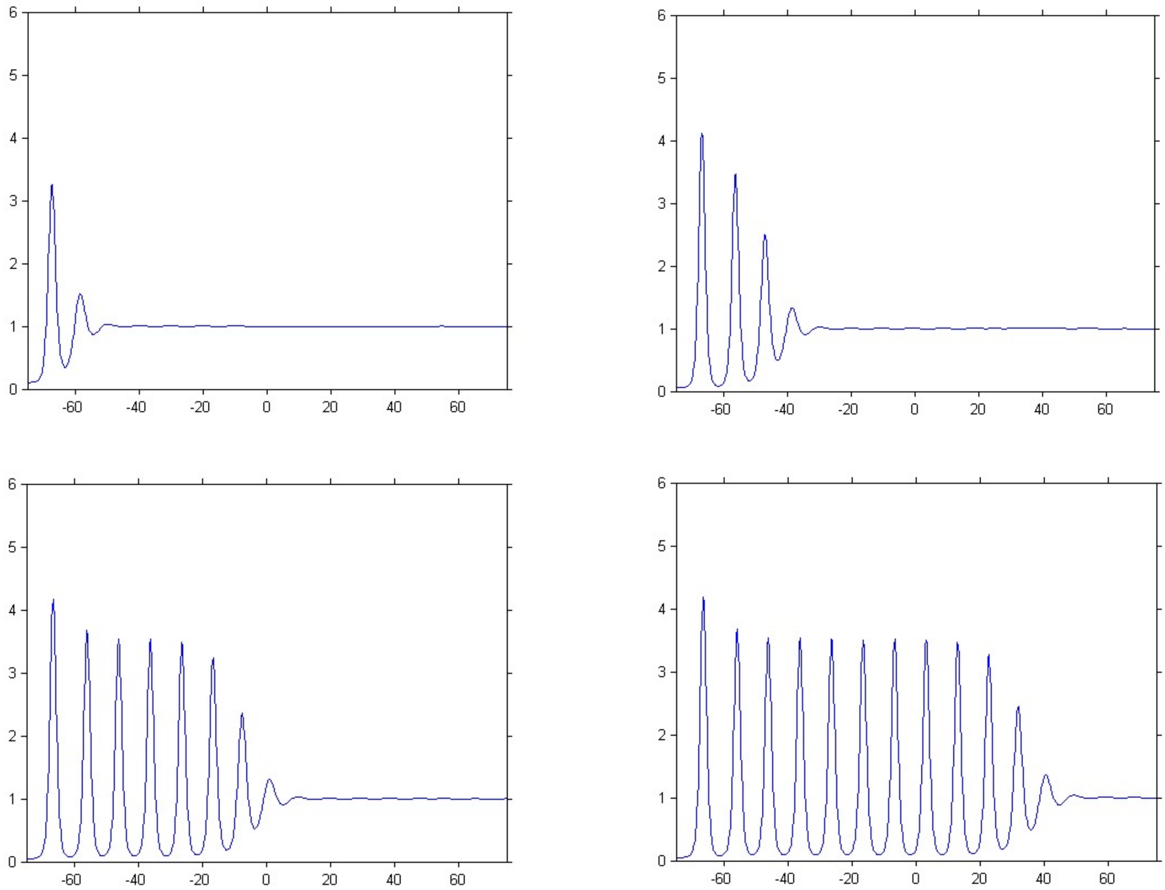

Figure 3.3 Contours along the line L of the cell density (u) computed from the initial condition (3.30) at $t = 100$ (left top), $t = 200$ (right top), $t = 400$ (left bottom), $t = 600$ (right bottom).

Test 2. Now, a positive perturbation is made in cell density as follow :

$$u_0(x) = \begin{cases} 1 + \epsilon(x) & \text{if } x \in (-9/2, 9/2) \times (-5/2, 5/2), \\ 1 & \text{otherwise, with } x \in \Omega. \end{cases} \tag{3.31}$$

The above initial condition and the computed solution at $t = 400$ are shown in Figure 3.4. We can see the ability of the model to generate stripe patterning for such initial condition. Contours along the line L of u at $t = 200$ and $t = 400$ are plotted in Figure 3.5 (left) and Figure 3.6 (left). In this case, we observe that the patterning begins at the middle of L and progresses to the boundaries symmetrically.

Test 3. In this test, the initial condition is similar to the previous test, except that the perturbation in initial cell density is done on a thinner strip :

$$u_0(x) = \begin{cases} 1 + \epsilon(x) & \text{if } x \in (-9/2, 9/2) \times (-1/2, 1/2), \\ 1 & \text{otherwise, with } x \in \Omega. \end{cases} \tag{3.32}$$

Figure 3.5 (right) and Figure 3.6 (right) show that the propagation pattern is similar to the previous test, but with a difference in peaks.

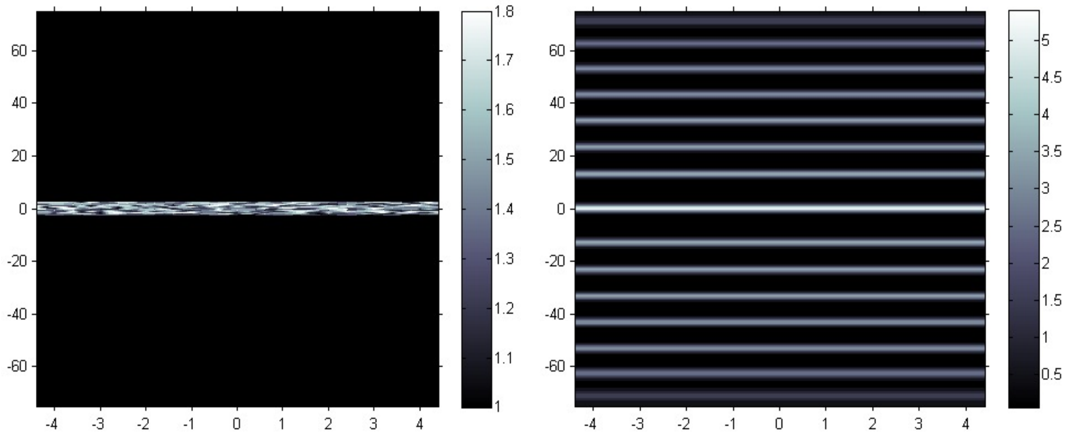

Figure 3.4 Initial condition of the cell density u_0 (3.31) (left), and the solution (u) at time $t = 400$ (right).

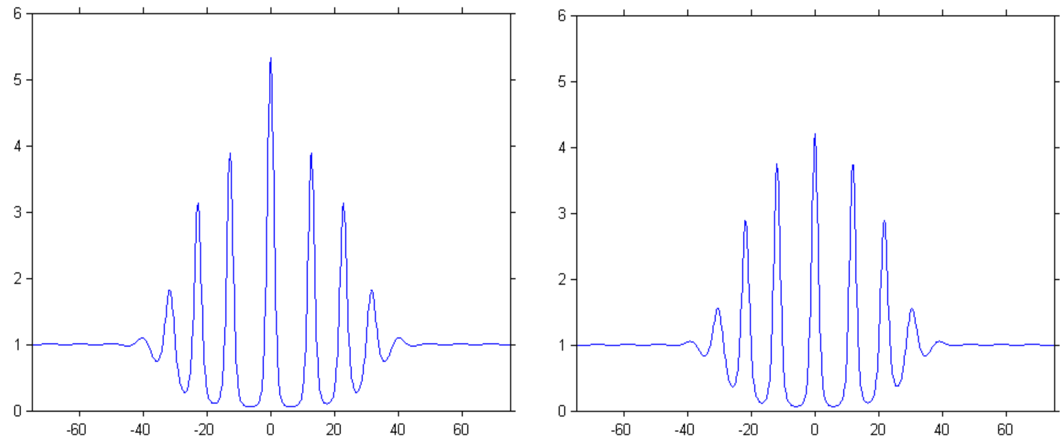

Figure 3.5 Contours along the line L of the cell density (u) at $t = 200$, computed from the initial condition (3.31) (left), and from the initial condition (3.32) (right).

Test 4. Finally, we consider a zero initial condition with a perturbation in a small strip :

$$u_0(x) = \begin{cases} \epsilon(x) & \text{if } x \in (-9/2, 9/2) \times (-1/2, 1/2), \\ 0 & \text{otherwise, with } x \in \Omega. \end{cases} \tag{3.33}$$

The computed cell density at times $t = 10$ and $t = 1000$ are plotted in Figure 3.5. As we can see, the model does not generate any spatial patterns in this case.

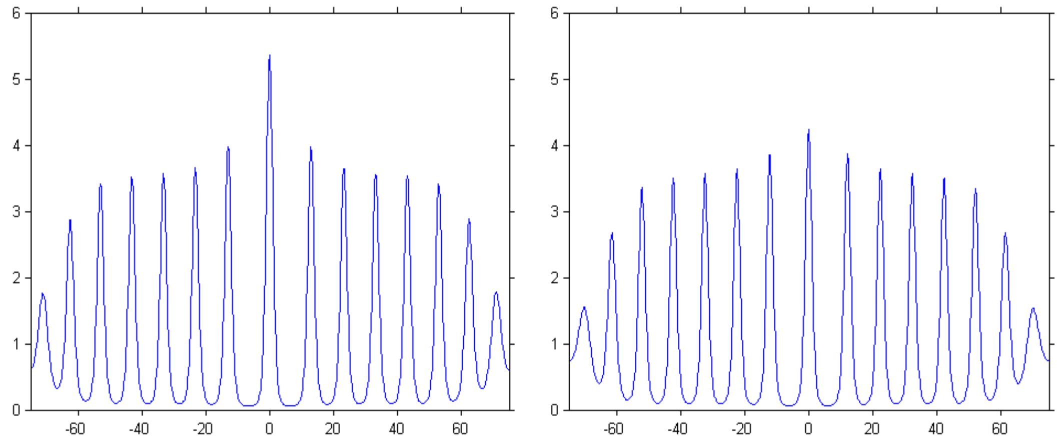

Figure 3.6 Contours along the line L of the cell density (u) at $t = 400$, computed from the initial condition (3.31) (left), and from the initial condition (3.32) (right).

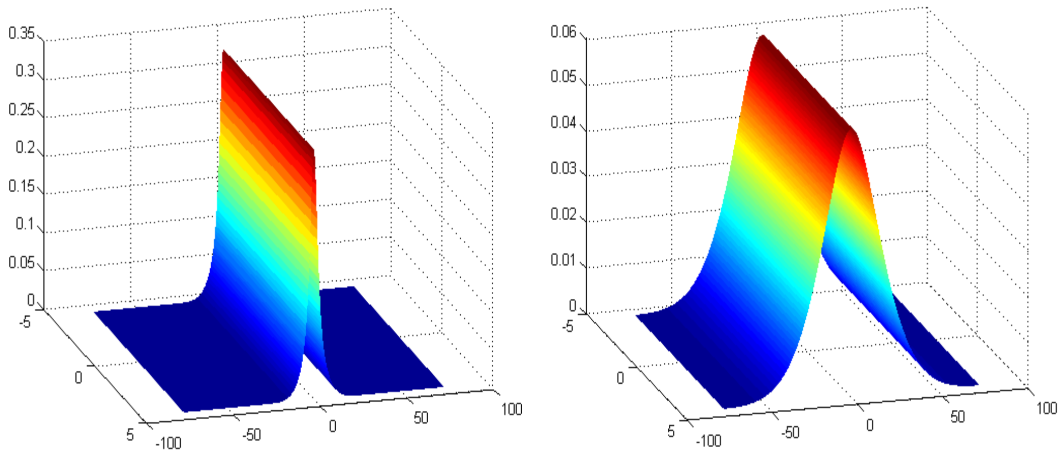

Figure 3.7 Solution (u) computed from the initial condition (3.33) at $t = 10$ (left), and at $t = 1000$ (right).

Chapter 4

A time semi-exponentially fitted scheme for chemotaxis-growth models

Sommaire

4.1 Introduction

In this chapter, we are interested in the following general form of the Keller-Segel model

$$\begin{cases} \partial_t u = \nabla \cdot (D_u \nabla u - u\,\chi(c)\,\nabla c) + f(u), \\ \partial_t c = D_c \Delta c + g(u,c), \end{cases} \tag{4.1}$$

where $u(x,t)$ denotes the cell density at spatial position x and time t, $c(x,t)$ is the concentration of the chemoattractant substance, D_u is the diffusion coefficient of the cells, $\chi(c)$ is the chemotactic sensitivity, $f(u)$ is a function describing cell proliferation and reduction, D_c is the chemical diffusion coefficient while $g(u,c)$ is a kinetic function that describes production and degradation of the chemical signal.

Several chemotaxis models are derived from the above generic system. The classical Keller-Segel model (2.1) can be obtained from (4.1) by taking :

$$\chi(c) = \chi, \; f(u) = 0, \; g(u,c) = u - c,$$

with χ is a positive constant. Many other parabolic systems based on (4.1) are used to model complicated phenomena by incorporating new equations, variables, and terms. In this work, we focus on such models, and in particular on the chemotaxis models with a polynomial cellular growth term $f(u)$ (the growth term may depend on other variables). Such function is widely used in modeling of cell proliferation and reduction in chemotaxis systems (see, e.g., [16, 69, 57, 85]) and other population dynamics models.

In Chapter 2, we give an overview of existing papers focused on numerical treatment of chemotaxis models. In these works, different approaches for time discretization have been adopted. Explicit schemes were proposed by many authors [17, 22, 24, 34]. The main drawback of these schemes is the CFL condition needed to ensure stability, which becomes too restrictive when a fine mesh is used. To avoid such constraint, implicit schemes were developed in several works such as [53, 28, 7, 9]. However, such time discretization requires the numerical resolution of a large nonlinear algebra system, which is in general too expensive. The third approach which can be used is the linearized semi-implicit discretization, used in chapters 2 and 3. In terms of computational cost, this discretization is more advantageous compared with the fully implicit one, since we only need to solve decoupled linear systems. This strategy was used in [77, 75] for the classical Keller-Segel model. In [84, 83], a semi-implicit finite element scheme was derived for two and three dimensional chemotaxis-growth system. This scheme is unconditionally stable and positivity-preserving if the growth term in the weak form of the system is positive [84, Theorem 3.2]. Unfortunately, this last condition is not satisfied in general.

In this chapter, a finite volume scheme combined with a semi-exponentially fitted time discretization is presented. The scheme is applied to two chemotaxis-growth models. However, it may be easily adapted to a class of chemotaxis models with polynomial growth based on (4.1). The method uses Il'in discretization for the approximation of the convection-diffusion fluxes. The time discretization, which is the main goal of this chapter, is developed in a such way that the numerical scheme verifies the following properties : first, it may only require to solve decoupled linear systems, which excludes the fully implicit schemes. Second, the scheme should be stable without any time-step restriction. Finally, it should be unconditionally positivity preserving. As mentioned previously, the explicit and semi-implicit methods that have been proposed do not fulfill the last two properties.

In the framework of embryonic pattern formation, Oster and Murray proposed in [67] a cell-chemotaxis model based on (4.1). In its dimensionless form, the model reads [59, 52]

$$\begin{cases} \partial_t u = \mu \Delta u - \nabla \cdot (\chi\, u \nabla c) + sru\,(1-u)\,, \\ \partial_t c = \Delta c + s\left(\dfrac{u}{u+1} - c\right), \end{cases} \tag{4.2}$$

where μ, χ, s and r are positive constants. The model is completed with the zero-flux boundary conditions.

The above model can generate different forms of spatial patterns [52]. It describes also the mechanism which generates the condensations of cells leading to embryonic pattern formations. In [59], several numerical simulations were presented to demonstrate the ability of (4.2), in its steady state form, to reproduce the patterns found on the skin of snakes (see for example Figure 1.5).

In this chapter, a particular attention will be given to the following model

$$\begin{cases} \partial_t u = \Delta u - \nabla \cdot (\, u \nabla c) + u\,(1-u) \quad \text{in } \Omega_T, \\ 0 = \Delta c + \dfrac{u}{u+1} - c \quad \text{in } \Omega_T, \end{cases} \tag{4.3}$$

with the homogeneous Neumann boundary and initial conditions

$$\nabla u \cdot \nu = \nabla c \cdot \nu = 0 \quad \text{on } \partial\Omega \times (0, T_f), \quad u(.,0) = u_0(x) \quad \text{in } \Omega. \tag{4.4}$$

Herein, $\Omega_T := \Omega \times (0, T_f)$, $\Omega \subset \mathbb{R}^2$ is an open bounded subset with sufficiently regular boundary $\partial\Omega$, $T_f > 0$ is a fixed time and ν denotes the outward unit-normal on the boundary $\partial\Omega$.

The above model was theoretically discussed in [18], where the existence and uniqueness of global solution are established for suitable initial data. As we can see, the system (4.3) is a modified version of the original model (4.2) in which the second equation of the system is elliptic, using the reasonable assumption that the diffusion of the chemoattractant is much faster than that of the cell. Without loss of generality, we have also assumed that

$$\mu = \chi = s = r = 1. \tag{4.5}$$

Contrary to the classical Keller-Segel model, the solutions of the system (4.3) do not blow-up in finite time, since the source term in the elliptic equation is bounded. It also should be noted that the reactive term $u\,(1-u)$ prevents the blow-up of the solutions, even if the second equation of the model is similar to that of the classical Keller-Segel model [86, 93].

The chapter is organized as follows. In section 4.2, we derive our semi-exponentially fitting in time and finite volume in space scheme for the system (4.3)–(4.4). Some semi-implicit Euler schemes are also presented. The existence and uniqueness of a discrete solution for the presented schemes are shown is Section 4.3. The advantage of our time discretization over linearized backward Euler schemes will be underlined in this section. A priori estimates on the discrete solution of the new scheme are given in Section 4.4. In Section 4.5, we use these estimates to prove that the approximate solution converges to a weak solution of the studied model. In Section 4.6, we present some numerical experiments that show the efficiency of the scheme. The scheme is then applied to another chemotaxis-growth model to demonstrate its ability to capture bacterial pattern formations.

4.2 Presentation of the numerical schemes

Let $\Omega \subset \mathbb{R}^2$ be an open bounded polygonal subset. We consider in this chapter an admissible finite volume mesh of Ω, denoted by $\mathcal{T}$, which verifies the properties of Definition 5.1 in [26] (see also Définition 1.3.1). We will also use the same notations related to the spatial discretization and discrete norms as in Section 2.2.1.

The mesh is assumed to satisfy this additional constraint : there exists $\xi > 0$ such that

$$\mathrm{d}(x_K, \sigma) \geq \xi\, \mathrm{d}(x_K, x_L). \tag{4.6}$$

Let $0 = t_0 < t_1 < ... < t_N = T_f$ be a uniform partition of $(0, T_f)$ with $N \in \mathbb{N}$ and $t_n = n\Delta t$ for $n = 0, ..., N$.

We denote by h the maximal size (diameter) of the control volumes included in $\mathcal{T}$, and we define

$$\delta = \max(\Delta t, h).$$

4.2.1 Semi-implicit finite volume schemes

We present some numerical schemes for the cell density equation, in which we apply a finite volume approximation in space, and semi-implicit Euler discretizations in time.

Let $x \in K$ and $t \in [t_n, t_{n+1})$. We denote by u_K^n an approximation of $\frac{1}{\mathrm{m}(K)} \int_K u(x,t^n)\,dx$, and similarly for c_K^n. We begin by integrating the first equation of the system over the control volume K, then by applying the Green-Gauss formula, we get for all $K \in \mathcal{T}$

$$\begin{aligned} &\int_K \partial_t u(x,t)\,dx - \sum_{\sigma\in\mathcal{E}_K} \int_\sigma (\nabla u(x,t) - u(x,t)\nabla c(x,t)) \cdot \nu_{K,\sigma}\,d\sigma \\ &\quad - \int_K u(x,t)\,(1-u(x,t))\,dx = 0, \end{aligned} \tag{4.7}$$

where $\nu_{K,\sigma}$ denotes the unit normal on σ which is outward to K. Now, by approximating the average of the product by the product of the averages in the reaction term, and by using a backward Euler discretization for the time derivative and the convection-diffusion flux, with a semi-implicit discretization for the last term, we obtain the following scheme :
for all $K \in \mathcal{T}$ and $n = 0, ..., N-1$

$$\mathrm{m}(K)\frac{u_K^{n+1} - u_K^n}{\Delta t} + \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^{n+1} - \mathrm{m}(K)\, u_K^n \left(1 - u_K^{n+1}\right) = 0, \tag{4.8}$$

where $F_{K,\sigma}^{n+1}$ is an approximation of the convection-diffusion flux at $t = t^{n+1}$. In the equation (4.8), a semi-implicit discretization is used to linearize the reaction term. It should be noted, however, that in order to achieve this purpose, two other schemes can be proposed

$$\mathrm{m}(K)\frac{u_K^{n+1} - u_K^n}{\Delta t} + \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^{n+1} - \mathrm{m}(K)\, u_K^n \left(1 - u_K^{n}\right) = 0, \tag{4.9}$$

$$\mathrm{m}(K)\frac{u_K^{n+1} - u_K^n}{\Delta t} + \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^{n+1} - \mathrm{m}(K)\, u_K^{n+1} \left(1 - u_K^{n}\right) = 0. \tag{4.10}$$

In this chapter, we focus on the scheme (4.8), since it is unconditionally stable, as it is proved in Section 4.3.

For the numerical flux $F_{K,\sigma}^{n+1}$, we will consider the following exponential fitting approximation

$$F_{K,\sigma}^{n+1} = \begin{cases} 0 & \text{for } \sigma \subset \partial\Omega, \\ \tau_\sigma \left(B\left(-Dc_{K,\sigma}^{n+1}\right) u_K^{n+1} - B\left(Dc_{K,\sigma}^{n+1}\right) u_L^{n+1} \right) & \text{otherwise }, \ \sigma = K|L, \end{cases} \tag{4.11}$$

where B is the Bernoulli function defined by

$$B(0) = 1, \; B(x) = \frac{x}{\exp(x) - 1} \, \forall x \in \mathbb{R} \setminus \{0\}, \tag{4.12}$$

and where for all $\sigma \in \mathcal{E}_K$, $Dc_{K,\sigma}^{n+1}$ is defined by

$$Dc_{K,\sigma}^{n+1} = \begin{cases} 0 & \text{for } \sigma \subset \partial\Omega, \\ c_L^{n+1} - c_K^{n+1} & \text{otherwise} \;, \; \sigma = K|L. \end{cases}$$

Using the definition (4.11), the same approximation can also be expressed by

$$F_{K,\sigma}^{n+1} = \tau_\sigma \left(Dc_{K,\sigma}^{n+1} \frac{u_K^{n+1} + u_L^{n+1}}{2} - \mathfrak{L}\left(\frac{Dc_{K,\sigma}^{n+1}}{2} \right) Du_{K,\sigma}^{n+1} \right), \tag{4.13}$$

where $Du_{K,\sigma}^{n+1}$ is defined similarly to $Dc_{K,\sigma}^{n+1}$, and $\mathfrak{L}$ is the 1-Lipschitz continuous function :

$$\mathfrak{L}(0) = 1, \; \mathfrak{L}(x) = x \coth(x) \, \forall x \in \mathbb{R} \setminus \{0\}, \tag{4.14}$$

a plot of this function is presented in Figure 4.1.

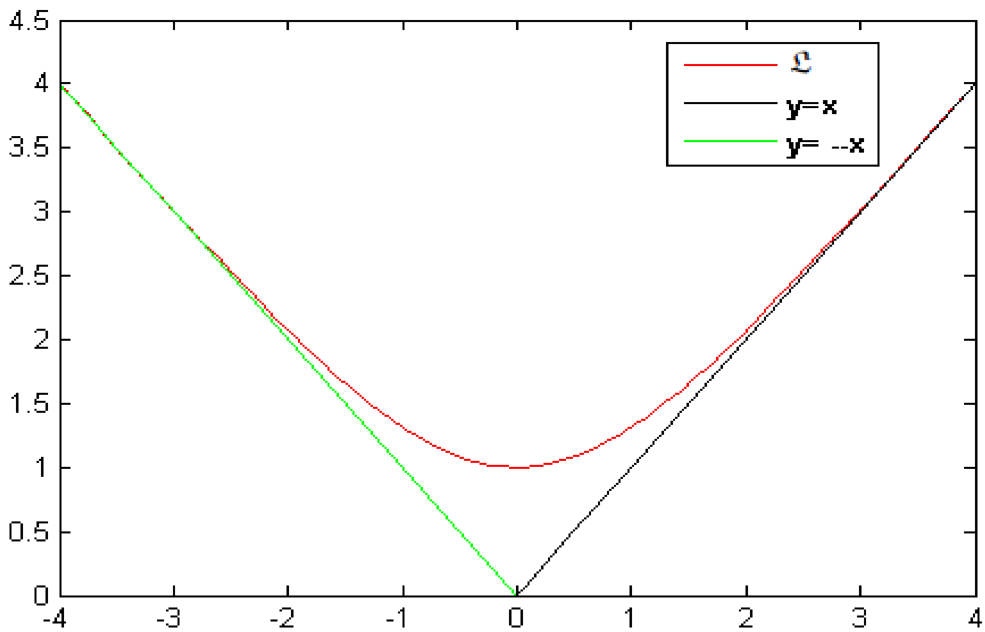

Figure 4.1 Plot of the function $\mathfrak{L}$.

The numerical scheme (4.11), which has been developed to approximate the convection-diffusion flux, is called Il'in scheme [43] (or Scharfetter-Gummel scheme [78]). It is derived from an appropriate one dimensional local Dirichlet problem. The Il'in scheme is suitable to solve convective dominated problems and it is second order accurate [48]. The spatial finite volume approximation (4.11) has been generalized to the case of nonlinear diffusion [6], and when both convection and diffusion terms are nonlinear [25].

The finite volume scheme for the second equation of the system (4.3) will be presented in the next section.

4.2.2 Time semi-exponentially fitted scheme for the chemotaxis problem

The time discretization, which will be introduced in this section, is inspired by the Il'in scheme and the semi-implicit scheme (4.10). We focus first on the volume integrated time derivative and reaction terms obtained from (4.7)

$$\int_K \partial_t u(x,t)\,dx - \int_K u(x,t)\,(1-u(x,t))\;dx.$$

Let $t \in [t_n, t_{n+1})$. We approximate the above nonlinear term by

$$\mathrm{m}(K)\left(u_K' - u_K\,(1-u_K^n)\right) \quad \text{where } u_K(t) = \frac{1}{\mathrm{m}(K)}\int_K u(x,t)\,dx.$$

We recall that u_K^n is an approximation of the mean value of $u(x,t^n)$ on K. The last expression is then approximated by a constant value T_K^{n+1}, using the solution of the following local boundary value problem :

$$\begin{cases} (u_K' - u_K\,(1-u_K^n))' = 0 \quad \text{on } (0,\Delta t), \\ u_K(0) = u_K^n, \quad u_K(\Delta t) = u_K^{n+1}. \end{cases}$$

Hence, the numerical approximation T_K^{n+1} may be defined by the following expression

$$T_K^{n+1} = \mathrm{m}(K)\frac{B\left(\Delta t\,(1-u_K^n)\right)u_K^{n+1} - B\left(-\Delta t\,(1-u_K^n)\right)u_K^n}{\Delta t}, \tag{4.15}$$

which, by using the definition of the Bernoulli function (4.11), can also be written as

$$T_K^{n+1} = \frac{\mathrm{m}(K)}{\Delta t}\left(-\Delta t\,(1-u_K^n)\,\frac{u_K^n + u_K^{n+1}}{2} + \mathfrak{L}\left(\frac{\Delta t\,(1-u_K^n)}{2}\right)\left(u_K^{n+1} - u_K^n\right)\right). \tag{4.16}$$

The convection-diffusion flux is discretized implicitly similarly to the schemes presented in Section 4.2.1. The final time semi-exponentially fitted scheme reads :
for all $K \in \mathcal{T}$ and $n = 0, ..., N-1$

$$T_K^{n+1} + \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^{n+1} = 0. \tag{4.17}$$

For the concentration equation, we adopt the following classical finite volume scheme

$$-\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\, Dc_{K,\sigma}^{n+1} + \mathrm{m}(K)\,c_K^{n+1} = \mathrm{m}(K)\,\frac{u_K^n}{u_K^n+1}. \tag{4.18}$$

This allows us to avoid coupling and nonlinearity of the numerical discrete system (4.17)–(4.18). We note that all schemes defined in this section and in Section 4.2.1 should be completed by the following compatible initial data :

$$u_K^0 = \frac{1}{\mathrm{m}(K)}\int_K u_0(x)\,dx \quad \text{for all } K \in \mathcal{T}. \tag{4.19}$$

Finally, the piecewise constant finite volume approximations of u denoted by u_δ and its discrete gradient du_δ (see [13]) are defined by :

$$\begin{aligned}
u_\delta(x,t) &= u_K^{n+1}, \quad c_\delta(x,t) = c_K^{n+1}, \quad x \in K,\, t \in [t^n, t^{n+1}). \\
du_\delta(x,t) &= \frac{\mathrm{m}(\sigma)}{\mathrm{m}(M_{K,\sigma})}\, Du_{K,\sigma}^{n+1}\, \nu_{K,\sigma}, \quad x \in M_{K,\sigma},\ t \in [t^n, t^{n+1}),
\end{aligned}$$

for all $K \in \mathcal{T}$ and $n = 0, ..., N-1$, where $M_{K,\sigma}$ is the cell formed from the vertices of σ, x_K and x_L if $\sigma = K|L \not\subset \partial\Omega$, and from the vertices of σ and x_K if $\sigma \subset \partial\Omega$. The same notations hold for the variable c.

4.3 Existence, uniqueness, and nonnegativity of finite volume solutions

Proposition 4.3.1. *Assume that* $u_0 \geq 0$. *Then, the schemes* (4.17)–(4.19) *and* (4.8),(4.18)–(4.19) *admit a unique solution* $\{(u_K^{n+1}, c_K^{n+1}),\, K \in \mathcal{T},\, n = 0, ..., N-1\}$, *where for all* $K \in \mathcal{T}$, $0 \leq n \leq N-1$, *we have* $u_K^{n+1} \geq 0$ *and* $c_K^{n+1} \geq 0$.

Proof. At each time-step, the decoupled scheme (4.17)–(4.18) can be written equivalently as

$$A^n\, U^{n+1} = S^n, \tag{4.20}$$

$$JC^{n+1} = G^n, \tag{4.21}$$

where for all $K \in \mathcal{T}$, the vectors of (4.20)–(4.21) are defined by : $U_K^{n+1} = u_K^{n+1}$, $S_K^n = \mathrm{m}(K) B\left(-\Delta t\,(1 - u_K^n)\right) u_K^n / \Delta t$, $C_K^{n+1} = c_K^{n+1}$ and $G_K^n = \mathrm{m}(K) \frac{u_K^n}{u_K^n + 1}$. The associated system matrices are defined by :

$$\begin{aligned}
A_{K,K}^n &= \mathrm{m}(K) B\left(\Delta t\,(1 - u_K^n)\right)/\Delta t + \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma \not\subset \partial\Omega}} \tau_\sigma B\left(-Dc_{K,\sigma}^{n+1}\right), \\
A_{K,L}^n &= -\tau_\sigma B\left(Dc_{K,\sigma}^{n+1}\right), \quad \text{if } L \in \mathcal{N}(K) \text{ with } \sigma = K|L, \\
A_{K,L}^n &= 0, \quad \text{otherwise,}
\end{aligned}$$

and by

$$\begin{aligned}
J_{K,K} &= \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma \not\subset \partial\Omega}} \tau_\sigma + \mathrm{m}(K), \\
J_{K,L} &= -\tau_\sigma, \quad \text{if } L \in \mathcal{N}(K) \text{ with } \sigma = K|L, \\
J_{K,L} &= 0, \quad \text{otherwise.}
\end{aligned}$$

To prove the uniqueness and nonnegativity of the solution of the scheme (4.17)–(4.18), we proceed by induction on n. Since we have from (4.19) that $u_K^0 \geq 0$ for each $K \in \mathcal{T}$, the existence and nonnegativity of u_K^1 and c_K^1 for all $K \in \mathcal{T}$ can be shown similarly as the inductive step.

Let us assume now that U^n and C^n are known, and that $U^n \geq 0$ and $C^n \geq 0$. Then, we have for all $K \in \mathcal{T}$:

$$|J_{K,K}| - \sum_{\substack{L\in\mathcal{T}\\ L\neq K}} |J_{K,L}| = \mathrm{m}(K) > 0,$$

which implies that the matrix J is strictly diagonally dominant by rows, and since J has nonpositive off-diagonal and positive diagonal entries, J is a nonsingular M-matrix. It follows that J^{-1} is positive. Consequently, since $G^n \geq 0$ by the induction supposition, there exists a unique nonnegative vector C^{n+1} satisfying (4.21).

On the other hand, since for all $\sigma = K|L \in \mathcal{E}_K$, $Dc^{n+1}_{L,\sigma} = -Dc^{n+1}_{K,\sigma}$, we get for all $K \in \mathcal{T}$

$$|A^n_{K,K}| - \sum_{\substack{L\in\mathcal{T}\\ L\neq K}} |A^n_{L,K}| = \mathrm{m}(K)B\left(\Delta t\left(1-u^n_K\right)\right)/\Delta t > 0.$$

Thus, since the function B (defined in (4.12)) is positive, A^n is strictly diagonal dominant by columns. Furthermore, the matrix A^n has nonpositive off-diagonal and positive diagonal entries. Therefore, A^n is a nonsingular M-matrix. Hence, since S^n and $(A^n)^{-1}$ are positive (A^n is an M-matrix), there exists a unique nonnegative vector U^{n+1} which satisfies (4.20).

Finally, the uniqueness and nonnegativity of the solution of the scheme (4.8),(4.18)–(4.19) can be established in the same way, using the fact that the scheme (4.8) can be written as :

$$A^n\, U^{n+1} = S^n,$$

with $S^n_K = \mathrm{m}(K)u^n_K\left(1/\Delta t + 1\right)$ and

$$\begin{aligned}
A^n_{K,K} &= \mathrm{m}(K)\left(1/\Delta t + u^n_K\right) + \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma\not\subset\partial\Omega}} \tau_\sigma B\left(-Dc^{n+1}_{K,\sigma}\right),\\
A^n_{K,L} &= -\tau_\sigma B\left(Dc^{n+1}_{K,\sigma}\right), \quad \text{if } L \in \mathcal{N}(K) \text{ with } \sigma = K|L,\\
A^n_{K,L} &= 0, \quad \text{otherwise.}
\end{aligned}$$

□

Remark 4.3.1. *As mentioned in Section 4.2.1, the linear schemes* (4.9),(4.18)–(4.19) *and* (4.10),(4.18)–(4.19) *require time-step restrictions to obtain stability and nonnegativity of the discrete solutions. Arguing similarly to the proof of Proposition 4.3.1, we can show that the time-step condition in the case of the scheme based on* (4.9) *depends on the solution, while the time-step condition in the case of the scheme based on* (4.10) *depends only on the parameters of the model.*

4.4 A priori estimates

In this section, we will derive the a priori estimates on the nonnegative approximate solution (u_δ, c_δ), which are needed to prove the convergence of the scheme (4.17)–(4.19). To this end, we assume that $u_0 \in L^2(\Omega)$ with $u_0 \geq 0$. In this and following sections, we denote by C a generic positive constant which is always independent of δ, and which may take different values.

Proposition 4.4.1. *Let* $\{(u_K^{n+1}, c_K^{n+1})$ *,* $K \in \mathcal{T}$*,* $n = 0, ..., N-1\}$ *be the solution of the scheme* (4.17)–(4.19)*. Then there exists a constant* $C > 0$*, only depending on* Ω*,* T_f *and* u_0*, such that for all* $n = 0, ..., N-1$*, and* $K \in \mathcal{T}$

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\, u_K^{n+1} \le C, \tag{4.22}$$

$$\sum_{n=0}^{N-1}\sum_{K\in\mathcal{T}} \mathrm{m}(K) B\left(-\Delta t\left(1-u_K^n\right)\right)\left(u_K^{n+1}-u_K^n\right) \le C, \tag{4.23}$$

$$c_K^{n+1} \le 1 \ \textit{and} \ \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \left|Dc_{K,\sigma}^{n+1}\right|^2 \le \mathrm{m}\left(\Omega\right). \tag{4.24}$$

Proof. Since the Bernoulli function B is decreasing, using (4.11) and (4.15), and summing (4.17) over $K \in \mathcal{T}$, we get $n = 0, ..., N-1$

$$B\left(\Delta t\right)\sum_{K\in\mathcal{T}} \mathrm{m}(K) u_K^{n+1} - B\left(-\Delta t\right)\sum_{K\in\mathcal{T}} \mathrm{m}(K) u_K^n \le 0.$$

Now, multiplying the above equation by $1/B\left(\Delta t\right)$, and using the fact that $B(-x)/B(x) = \exp(x)$, we can easily see by induction that

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K) u_K^{n+1} \le \exp(T_f)\sum_{K\in\mathcal{T}} \mathrm{m}(K) u_K^0,$$

which gives the estimate (4.22).

Summing (4.17) over $K \in \mathcal{T}$, it is easy to see that

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K) u_K^{n+1}\left(B\left(\Delta t\left(1-u_K^n\right)\right) - B\left(-\Delta t\left(1-u_K^n\right)\right)\right) + \sum_{K\in\mathcal{T}} \mathrm{m}(K) B\left(-\Delta t\left(1-u_K^n\right)\right)\left(u_K^{n+1}-u_K^n\right) = 0.$$

Next, summing over $n = 0, ..., N-1$, and since B verifies

$$B(x) - B(-x) = -x, \tag{4.25}$$

we get

$$\sum_{n=0}^{N-1}\Delta t\sum_{K\in\mathcal{T}} \mathrm{m}(K) u_K^{n+1} u_K^n + \sum_{n=0}^{N-1}\sum_{K\in\mathcal{T}} \mathrm{m}(K) B\left(-\Delta t\left(1-u_K^n\right)\right)\left(u_K^{n+1}-u_K^n\right) = \sum_{n=0}^{N-1}\Delta t\sum_{K\in\mathcal{T}} \mathrm{m}(K) u_K^{n+1}.$$

Together with (4.22), we deduce (4.23).

The proof of the estimate (4.24) is similar to that of Proposition 3.4.1.

The second estimates of (4.24), is obtained by multiplying the equation (4.18) by c_K^{n+1}, summing over $K \in \mathcal{T}$, using a summation by parts, and applying the Young inequality on the right hand side :

$$\frac{1}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma\left|Dc_{K,\sigma}^{n+1}\right|^2+\frac{1}{2}\sum_{K\in\mathcal{T}}\mathrm{m}(K)\left|c_K^{n+1}\right|^2\leq\frac{1}{2}\mathrm{m}(\Omega).$$

This concludes the proof.

□

Lemma 4.4.1. *Assume that* (4.6) *is fulfilled. Let* $\{(u_K^{n+1},c_K^{n+1}),\ K\in\mathcal{T},\ n=0,...,N-1\}$ *be the solution of the scheme* (4.17)–(4.19)*, and let* a *be a constant defined by* $0<a<B(1)$*. Then, there exists a constant* $C>0$ *only depending on* Ω*,* T_f*,* u_0*,* a *and* ξ *such that*

$$\sum_{n=0}^{N}\Delta t\sum_{K\in\mathcal{T}}\mathrm{m}(K)\left|u_K^n\right|^2\leq C. \tag{4.26}$$

Proof. For all $K\in\mathcal{T}$, $n=0,...,N$, we set

$$\overline{u}_K^n=u_K^n+1, \tag{4.27}$$

and for all $x\in\mathbb{R}$, we define

$$\overline{B}(x)=B(x)-a,\quad\text{where } 0<a<B(1). \tag{4.28}$$

By multiplying the equation (4.17) by $\Delta t\log\left(\overline{u}_K^{n+1}\right)$, summing over $K\in\mathcal{T}$, and performing the changes of variable (4.28) (only for the expression of $F_{K,\sigma}^{n+1}$) and (4.27), we get

$$I_1^{n+1}+I_2^{n+1}=I_3^{n+1}+I_4^{n+1},$$

where

$$\begin{aligned}
I_1^{n+1}&=\sum_{K\in\mathcal{T}}\mathrm{m}(K)\left(B\left(\Delta t\left(2-\overline{u}_K^n\right)\right)\overline{u}_K^{n+1}-B\left(-\Delta t\left(2-\overline{u}_K^n\right)\right)\overline{u}_K^n\right)\log\left(\overline{u}_K^{n+1}\right),\\
I_2^{n+1}&=-a\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K}\Delta t\,\tau_\sigma D\overline{u}_{K,\sigma}^{n+1}\log\left(\overline{u}_K^{n+1}\right),\\
I_3^{n+1}&=-\sum_{K\in\mathcal{T}}\sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}}\Delta t\,\tau_\sigma\left(\overline{B}\left(-Dc_{K,\sigma}^{n+1}\right)\overline{u}_K^{n+1}-\overline{B}\left(Dc_{K,\sigma}^{n+1}\right)\overline{u}_L^{n+1}\right)\log\left(\overline{u}_K^{n+1}\right),\\
I_4^{n+1}&=\sum_{K\in\mathcal{T}}\Delta t\,\mathrm{m}(K)\left(\overline{u}_K^n+\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma Dc_{K,\sigma}^{n+1}-2\right)\log\left(\overline{u}_K^{n+1}\right),
\end{aligned}$$

where we have used the definitions (4.11) and (4.15), and the identity (4.25).

For the term I_1^{n+1}, using the fact that B is a decreasing function, we have

$$\begin{aligned}
I_1^{n+1}\geq&\sum_{K\in\mathcal{T}}\mathrm{m}(K)\left(B\left(2\Delta t\right)\overline{u}_K^{n+1}\log\left(\overline{u}_K^{n+1}\right)-B\left(-2\Delta t\right)\overline{u}_K^n\log\left(\overline{u}_K^n\right)\right)\\
&-\sum_{K\in\mathcal{T}}\mathrm{m}(K)B\left(-\Delta t\left(2-\overline{u}_K^n\right)\right)\overline{u}_K^n\left(\log\left(\overline{u}_K^{n+1}\right)-\log\left(\overline{u}_K^n\right)\right).
\end{aligned}$$

Then, since we have from a Taylor expansion of log

$$\overline{u}_K^n \left(\log\left(\overline{u}_K^{n+1}\right) - \log\left(\overline{u}_K^n\right)\right) \leq \overline{u}_K^{n+1} - \overline{u}_K^n, \tag{4.29}$$

we deduce that

$$\begin{aligned} I_1^{n+1} \geq & \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(B\left(2\Delta t\right) \overline{u}_K^{n+1} \log\left(\overline{u}_K^{n+1}\right) - B\left(-2\Delta t\right) \overline{u}_K^n \log\left(\overline{u}_K^n\right)\right) \\ & - \sum_{K\in\mathcal{T}} \mathrm{m}(K) B\left(-\Delta t\left(2 - \overline{u}_K^n\right)\right)\left(\overline{u}_K^{n+1} - \overline{u}_K^n\right). \end{aligned}$$

By a summation by parts on I_2^{n+1}, we get

$$I_2^{n+1} = \frac{a}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t\, \tau_\sigma D\overline{u}_{K,\sigma}^{n+1} \left(\log\left(\overline{u}_L^{n+1}\right) - \log\left(\overline{u}_K^{n+1}\right)\right).$$

Next, using the Taylor expansion

$$\log\left(\overline{u}_K^{n+1}\right) = \log\left(\overline{u}_L^{n+1}\right) + \frac{\overline{u}_K^{n+1} - \overline{u}_L^{n+1}}{\theta_\sigma^{n+1}}, \tag{4.30}$$

where $\theta_\sigma^{n+1} = t_\sigma \overline{u}_K^{n+1} + (1-t_\sigma)\,\overline{u}_L^{n+1}$ with $t_\sigma \in (0,1)$, we can write

$$I_2^{n+1} = \frac{a}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma \left| \frac{D\overline{u}_{K,\sigma}^{n+1}}{\sqrt{\theta_\sigma^{n+1}}} \right|^2.$$

Now, since

$$\frac{\left|D\overline{u}_{K,\sigma}^{n+1}\right|}{\sqrt{\theta_\sigma^{n+1}}} = \frac{\sqrt{\overline{u}_K^{n+1}} + \sqrt{\overline{u}_L^{n+1}}}{\sqrt{\theta_\sigma^{n+1}}} \left| D\left(\sqrt{\overline{u}^{n+1}}\right)_{K,\sigma} \right| \geq \left| D\left(\sqrt{\overline{u}^{n+1}}\right)_{K,\sigma} \right|,$$

it follows that

$$I_2^{n+1} \geq \frac{a}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma \left| D\left(\sqrt{\overline{u}^{n+1}}\right)_{K,\sigma} \right|^2.$$

Using a summation by parts, we obtain

$$\begin{aligned} I_3^{n+1} = \frac{1}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t\, \tau_\sigma & \left(\overline{B}\left(-Dc_{K,\sigma}^{n+1}\right) \overline{u}_K^{n+1} - \overline{B}\left(Dc_{K,\sigma}^{n+1}\right) \overline{u}_L^{n+1}\right) \\ & \times \left(\log\left(\overline{u}_L^{n+1}\right) - \log\left(\overline{u}_K^{n+1}\right)\right). \end{aligned}$$

Let us now define $\tilde{I}_3^{n+1}$

$$\tilde{I}_3^{n+1} = \frac{1}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t\, \tau_\sigma \theta_\sigma^{n+1} Dc_{K,\sigma}^{n+1} \left(\log\left(\overline{u}_L^{n+1}\right) - \log\left(\overline{u}_K^{n+1}\right)\right).$$

Then, using (4.25) and the last expression of I_3^{n+1}, it yields

$$\begin{aligned}
I_3^{n+1} - \tilde{I}_3^{n+1} =& \frac{1}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t\, \tau_\sigma \overline{B}\left(-Dc_{K,\sigma}^{n+1}\right)\left(\overline{u}_K^{n+1} - \theta_\sigma^{n+1}\right) \\
& \times \left(\log\left(\overline{u}_L^{n+1}\right) - \log\left(\overline{u}_K^{n+1}\right)\right) + \frac{1}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t\, \tau_\sigma \overline{B}\left(Dc_{K,\sigma}^{n+1}\right) \\
& \times \left(\theta_\sigma^{n+1} - \overline{u}_L^{n+1}\right)\left(\log\left(\overline{u}_L^{n+1}\right) - \log\left(\overline{u}_K^{n+1}\right)\right) \\
=& \frac{1-t_\sigma}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t\, \tau_\sigma \overline{B}\left(-Dc_{K,\sigma}^{n+1}\right)\left(\overline{u}_K^{n+1} - \overline{u}_L^{n+1}\right) \\
& \times \left(\log\left(\overline{u}_L^{n+1}\right) - \log\left(\overline{u}_K^{n+1}\right)\right) + \frac{t_\sigma}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t\, \tau_\sigma \overline{B}\left(Dc_{K,\sigma}^{n+1}\right) \\
& \times \left(\overline{u}_K^{n+1} - \overline{u}_L^{n+1}\right)\left(\log\left(\overline{u}_L^{n+1}\right) - \log\left(\overline{u}_K^{n+1}\right)\right).
\end{aligned}$$

From the nonnegativity of c_K^{n+1} for all $K \in \mathcal{T}$ and (4.24), it is easy to see that $B\left(\pm\, Dc_{K,\sigma}^{n+1}\right) \geq B(1)$ (B is decreasing), which implies that $\overline{B}\left(\pm\, Dc_{K,\sigma}^{n+1}\right) > 0$. Thus, from the above equality, we obtain $I_3^{n+1} \leq \tilde{I}_3^{n+1}$, which leads using (4.30) to

$$I_3^{n+1} \leq \frac{1}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma Dc_{K,\sigma}^{n+1} D\overline{u}_{K,\sigma}^{n+1}.$$

Therefore, by multiplying (4.18) by $\overline{u}_K^{n+1}$ and summing by parts we can easily deduce that

$$I_3^{n+1} \leq \sum_{K\in\mathcal{T}} \Delta t\, \mathrm{m}(K)\, \overline{u}_K^{n+1}.$$

Finally, using (4.29), and since we have from (4.18) and (4.24) that $\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma Dc_{K,\sigma}^{n+1} \leq \mathrm{m}(K)$, we deduce that

$$I_4^{n+1} \leq \Delta t \left(\sum_{K\in\mathcal{T}} \mathrm{m}(K) \overline{u}_K^n \log\left(\overline{u}_K^n\right) + \sum_{K\in\mathcal{T}} \mathrm{m}(K) \overline{u}_K^{n+1} + \sum_{K\in\mathcal{T}} \mathrm{m}(K) \log\left(\overline{u}_K^{n+1}\right) \right).$$

Collecting the previous inequalities, summing over $n = 0, ..., p$ with $p \leq N-1$, and using (4.25), we obtain

$$\begin{aligned}
& B\left(2\,T_f\right) \sum_{K\in\mathcal{T}} \mathrm{m}(K) \overline{u}_K^{p+1} \log\left(\overline{u}_K^{p+1}\right) + \frac{a}{2} \sum_{n=0}^{p} \Delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \left| D\left(\sqrt{\overline{u}^{n+1}}\right)_{K,\sigma} \right|^2 \\
& \leq 3 \sum_{n=1}^{p} \Delta t \sum_{K\in\mathcal{T}} \mathrm{m}(K) \overline{u}_K^n \log\left(\overline{u}_K^n\right) + 2 \sum_{n=0}^{p} \Delta t \sum_{K\in\mathcal{T}} \mathrm{m}(K) \overline{u}_K^{n+1} \\
& \quad + \sum_{n=0}^{p} \Delta t \sum_{K\in\mathcal{T}} \mathrm{m}(K) \log\left(\overline{u}_K^{n+1}\right) + \left(B\left(-2\,T_f\right) + 1\right) \sum_{K\in\mathcal{T}} \mathrm{m}(K) \overline{u}_K^0 \log\left(\overline{u}_K^0\right) \\
& \quad + \sum_{n=0}^{p} \sum_{K\in\mathcal{T}} \mathrm{m}(K) B\left(-\Delta t\left(2 - \overline{u}_K^n\right)\right)\left(\overline{u}_K^{n+1} - \overline{u}_K^n\right).
\end{aligned}$$

From (4.22), (4.23) and since $u_0 \in L^2(\Omega)$, we can deduce by applying Lemma 2.2.1 that : there exists a positive constant C only depending on Ω, T_f and u_0, such that

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\overline{u}_K^{p+1} \log\left(\overline{u}_K^{p+1}\right) \leq C,$$

which implies the existence of a positive constant C only depending on Ω, T_f, u_0 and a, such that

$$\sum_{n=0}^{p} \Delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \left| D\left(\sqrt{\overline{u}^{n+1}}\right)_{K,\sigma}\right|^2 \leq C. \tag{4.31}$$

We denote by $\overline{u}_\delta^{n+1} \in X(\mathcal{T})$ the function defined by : $\overline{u}_\delta^{n+1}(x) = \overline{u}_K^{n+1}$ for all $x \in K$ and $n = 0, ..., N-1$. An application of Lemma 3.2.1 gives

$$\left\| \sqrt{\overline{u}_\delta^{n+1}} \right\|_4 \leq \frac{C}{\xi^{1/4}} \left\| \sqrt{\overline{u}_\delta^{n+1}} \right\|_{1,2,\mathcal{T}}^{1/2} \left\| \sqrt{\overline{u}_\delta^{n+1}} \right\|_2^{1/2}.$$

So, it follows that

$$\| \overline{u}_\delta^{n+1} \|_2^2 \leq \frac{|C|^4}{\xi} \left\| \sqrt{\overline{u}_\delta^{n+1}} \right\|_{1,2,\mathcal{T}}^2 \| \overline{u}_\delta^{n+1} \|_1,$$

where $C > 0$ is a constant depending only on Ω.

Therefore, from the above inequality, (4.31) and (4.22), the statement of the lemma follows.

□

Proposition 4.4.2. *There exists a constant $C > 0$ only depending on Ω, T_f, u_0, a (see (4.28)) and ξ (see (4.6)) such that, for all $n = 0, 1, ..., p$, with $p \leq N-1$:*

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K) \left| u_K^{p+1} \right|^2 + \sum_{n=0}^{p} \Delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \left| Du_{K,\sigma}^{n+1} \right|^2 \leq C, \tag{4.32}$$

$$\sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(u_K^{n+1} - u_K^n\right)^2 \leq C. \tag{4.33}$$

Proof. Multiplying the equation (4.17) by $\Delta t\, u_K^{n+1}$, using (4.13) and (4.15), and summing over $K \in \mathcal{T}$, we get

$$F_1^{n+1} + F_2^{n+1} = F_3^{n+1}, \tag{4.34}$$

where

$$F_1^{n+1} = \sum_{K\in\mathcal{T}} \mathrm{m}(K) \Bigg(B\left(\Delta t\left(1 - u_K^n\right)\right) u_K^{n+1} - B\left(-\Delta t\left(1 - u_K^n\right)\right) u_K^n \Bigg) u_K^{n+1},$$

$$F_2^{n+1} = -\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma \mathfrak{L}\left(\frac{Dc_{K,\sigma}^{n+1}}{2}\right) u_K^{n+1} Du_{K,\sigma}^{n+1},$$

$$F_3^{n+1} = -\frac{1}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t\, \tau_\sigma Dc_{K,\sigma}^{n+1} \left(u_K^{n+1} + u_L^{n+1}\right) u_K^{n+1}.$$

Using Young inequality, we have

$$\begin{aligned} F_1^{n+1} \geq & \frac{1}{2} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|u_K^{n+1}\right|^2 \Bigg(B\left(\Delta t\left(1-u_K^n\right)\right) - B\left(-\Delta t\left(1-u_K^n\right)\right) \Bigg) \\ & + \frac{1}{2} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(B\left(\Delta t\left(1-u_K^n\right)\right) \left|u_K^{n+1}\right|^2 - B\left(-\Delta t\left(1-u_K^n\right)\right) \left|u_K^n\right|^2 \right). \end{aligned}$$

Using the identity (4.25) on the first term on the right-hand side, and since B is decreasing, we get

$$\begin{aligned} F_1^{n+1} \geq & \frac{1}{2} \sum_{K\in\mathcal{T}} \Delta t\, \mathrm{m}(K) \left|u_K^{n+1}\right|^2 \left(u_K^n - 1\right) \\ & + \frac{1}{2} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(B\left(\Delta t\right) \left|u_K^{n+1}\right|^2 - B\left(-\Delta t\right) \left|u_K^n\right|^2 \right). \end{aligned} \tag{4.35}$$

Using a summation by parts, and since $\mathfrak{L}(x) \geq 1$ ($x\coth(x) > 1$ for all $x \in \mathbb{R}\setminus\{0\}$, and $\mathfrak{L}(0) = 1$), we obtain

$$F_2^{n+1} \geq \frac{1}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma \left| D u_{K,\sigma}^{n+1} \right|^2. \tag{4.36}$$

By using a summation by parts, we have

$$F_3^{n+1} = \frac{1}{4} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t\, \tau_\sigma D c_{K,\sigma}^{n+1} \left(\left|u_L^{n+1}\right|^2 - \left|u_K^{n+1}\right|^2 \right).$$

Now, multiplying the equation (4.18) by $\Delta t \left|u_K^{n+1}\right|^2$ and summing by parts, we deduce that

$$F_3^{n+1} \leq \frac{1}{2} \sum_{K\in\mathcal{T}} \Delta t\, \mathrm{m}(K) \left|u_K^{n+1}\right|^2. \tag{4.37}$$

Gathering (4.34), (4.35), (4.36) and (4.37), and summing over $n = 0, ..., p$, with $p \leq N-1$, we obtain

$$\begin{aligned} & \frac{B(T_f)}{2} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|u_K^{p+1}\right|^2 + \frac{1}{2} \sum_{n=0}^{p} \Delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \left| D u_{K,\sigma}^{n+1} \right|^2 \\ & + \frac{1}{2} \sum_{n=0}^{p} \Delta t \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|u_K^{n+1}\right|^2 u_K^n \\ & \leq \frac{1}{2} \sum_{n=0}^{p} \Delta t \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|u_K^{n+1}\right|^2 + \frac{1}{2} \sum_{n=1}^{p} \Delta t \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|u_K^n\right|^2 \\ & + \frac{B(-T_f)}{2} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|u_K^0\right|^2. \end{aligned}$$

In view of the estimate (4.26), (4.32) follows from the above inequality.

Now, we set

$$\tilde{B}(x) = B(x) - 1. \tag{4.38}$$

Using the above definition, we have

$$F_1^{n+1} = J_1^{n+1} + J_2^{n+1}, \tag{4.39}$$

where

$$J_1^{n+1} = \sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(\left|u_K^{n+1}\right|^2 - u_K^{n+1}u_K^n\right)$$
$$J_2^{n+1} = \sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(\tilde{B}\left(\Delta t\left(1-u_K^n\right)\right)u_K^{n+1} - \tilde{B}\left(-\Delta t\left(1-u_K^n\right)\right)u_K^n\right)u_K^{n+1}.$$

Since $\tilde{B}$ is decreasing, we can easily see that $\tilde{B}(\Delta t) < 0$ and $\tilde{B}(-\Delta t) > 0$ for $\Delta t > 0$, so using the Young inequality and (4.25), we get

$$\begin{aligned} J_2^{n+1} &\geq \sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(\tilde{B}\left(\Delta t\right)u_K^{n+1} - \tilde{B}\left(-\Delta t\right)u_K^n\right)u_K^{n+1} \\ &\geq -\frac{1}{2}\sum_{K\in\mathcal{T}} \Delta t\, \mathrm{m}(K)\left|u_K^{n+1}\right|^2 \\ &\quad + \frac{1}{2}\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(\tilde{B}\left(\Delta t\right)\left|u_K^{n+1}\right|^2 - \tilde{B}\left(-\Delta t\right)\left|u_K^n\right|^2\right). \end{aligned}$$

From the above inequality, the definition of J_1^{n+1}, (4.39), (4.34), and summing over $n = 0, ..., N-1$, we can easily see that

$$\begin{aligned} &\frac{1}{2}\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^N\right|^2 + \frac{1}{2}\sum_{n=0}^{N-1}\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(u_K^{n+1} - u_K^n\right)^2 \\ &\quad + \frac{1}{2}\sum_{n=0}^{N-1}\Delta t \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\left|Du_{K,\sigma}^{n+1}\right|^2 \\ &\leq \frac{1}{2}\sum_{n=0}^{N-1}\Delta t\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^{n+1}\right|^2 - \frac{\tilde{B}(T_f)}{2}\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^N\right|^2 \\ &\quad + \frac{1}{2}\sum_{n=1}^{N-1}\Delta t\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^n\right|^2 + \frac{\tilde{B}(-T_f)+1}{2}\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^0\right|^2. \end{aligned}$$

Hence, (4.33) follows from the estimate (4.32). This concludes the proof of Proposition 4.4.2.

□

4.5 Convergence of the scheme

This section is devoted to prove the convergence of the finite volume solution (u_δ, c_δ) of the scheme (4.17)–(4.19), to a weak solution of (4.3)–(4.4). The a priori estimates established in Section 4.4 are crucial to prove the convergence of the scheme, so, we suppose naturally that the assumptions of Section 4.4 hold.

Proposition 4.5.1. *There exists a subsequence of $(u_\delta, c_\delta)_\delta$, not relabeled, and a function $(u, c) \in L^2(0,T;H^1(\Omega))^2$ such that as δ goes to zero we have*

$$\begin{gathered} u_\delta \to u \quad \textit{strongly in } L^2(\Omega_T), \\ du_\delta \rightharpoonup \nabla u \quad \textit{weakly in } L^2(\Omega_T), \\ c_\delta \rightharpoonup c, \quad dc_\delta \rightharpoonup \nabla c \quad \textit{weakly in } L^2(\Omega_T). \end{gathered}$$

Proof. The key of the proof is the discrete Aubin-Simon lemma [30, Theorem 3.4]. The proof is based also on Lemma 4.4 in [13], and uses some techniques from the proof of Proposition 4.1 in [28].

Let $\phi \in H^q(\Omega)$ with $q > 1$ and let ϕ_K be the average of ϕ in the control volume K. Multiplying the equation (4.17) by ϕ_K, using (4.13) and (4.16), and summing over $K \in \mathcal{T}$, we obtain

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(u_K^{n+1} - u_K^n\right) \phi_K = E_1 + E_2 + E_3 + E_4,$$

where

$$\begin{aligned} E_1 &= \frac{\Delta t}{2} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(1 - u_K^n\right) \left(u_K^n + u_K^{n+1}\right) \phi_K, \\ E_2 &= -\sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(\mathfrak{L}\left(\frac{\Delta t\,(1-u_K^n)}{2}\right) - 1\right) \left(u_K^{n+1} - u_K^n\right) \phi_K, \\ E_3 &= -\frac{\Delta t}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \tau_\sigma D c_{K,\sigma}^{n+1} \left(u_K^{n+1} + u_L^{n+1}\right) \phi_K, \\ E_4 &= \Delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \mathfrak{L}\left(\frac{D c_{K,\sigma}^{n+1}}{2}\right) D u_{K,\sigma}^{n+1} \phi_K. \end{aligned}$$

Now, using the Cauchy-Schwarz inequality, and since u_δ is bounded in $L^\infty(0,T;L^2(\Omega))$, we can easily see that

$$|E_1| \le \frac{\Delta t}{2} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|1 - u_K^n\right| \left|u_K^n + u_K^{n+1}\right| |\phi_K| \le C\Delta t \|\phi\|_\infty.$$

Similarly, and since $\mathfrak{L}$ (defined in (4.14)) is a 1-Lipschitz continuous function, we have

$$|E_2| \le \frac{\Delta t}{2} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|1 - u_K^n\right| \left|u_K^{n+1} - u_K^n\right| |\phi_K| \le C\Delta t \|\phi\|_\infty.$$

For the term E_3, we can be write

$$
\begin{aligned}
|E_3| &= \left| -\frac{\Delta t}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma Dc^{n+1}_{K,\sigma} Du^{n+1}_{K,\sigma} \phi_K - \Delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma Dc^{n+1}_{K,\sigma} u^{n+1}_K \phi_K \right| \\
&\le \Delta t \|\phi\|_\infty \left(\frac{1}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \left| Dc^{n+1}_{K,\sigma} \right| \left| Du^{n+1}_{K,\sigma} \right| + \sum_{K\in\mathcal{T}} \left| \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma Dc^{n+1}_{K,\sigma} \right| u^{n+1}_K \right).
\end{aligned}
$$

Using the Cauchy-Schwarz inequality, (4.24) and (4.18), it is easy to see that

$$
|E_3| \le C \Delta t \|\phi\|_\infty \left(\|u_\delta\|_{1,2,\mathcal{T}} + 1 \right).
$$

Using a summation by parts, the Cauchy-Schwarz inequality, the estimate (4.32) and since we have from (4.24) that $\left| Dc^{n+1}_{K,\sigma} \right| \le 1$, we can write

$$
\begin{aligned}
|E_4| &\le \frac{\Delta t}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \left| Du^{n+1}_{K,\sigma} \right| |D\phi_{K,\sigma}| \\
&\quad + \frac{\Delta t}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \left| \mathfrak{L}\left(\frac{Dc^{n+1}_{K,\sigma}}{2} \right) - 1 \right| \left| Du^{n+1}_{K,\sigma} \right| |D\phi_{K,\sigma}| \\
&\le C \Delta t \|u_\delta\|_{1,2,\mathcal{T}} \|\phi_\delta\|_{1,2,\mathcal{T}}.
\end{aligned}
$$

Now, similarly to the proof of Proposition 4.1 in [28], and using the estimates obtained for E_j, with $j = 1,2,3,4$, we obtain the following uniform estimate of the time translation of u_δ in $L^2(0,T;(H^q(\Omega))')$:

$$
\int_0^{T-\tau} \int_\Omega \left(u_\delta(x,t+\tau) - u_\delta(x,t) \right) \phi(x,t)\, dx\, dt \le C\, \tau \|\phi\|_{L^2(0,T;H^q(\Omega))},
$$

where $\tau \in (0,\Delta t)$. Since $H^1(\Omega) \hookrightarrow\hookrightarrow L^2(\Omega) \hookrightarrow (H^q(\Omega))'$), we conclude from the discrete Aubin-Simon lemma [30, Theorem 3.4] (see also Remark 4 in [30]), that there exists a subsequence of $(u_\delta)_\delta$, not relabeled, such that

$$
u_\delta \to u \quad \text{strongly in } L^2(\Omega_T).
$$

Finally, using the estimates (4.32) and (4.24), and following the proof of Lemma 4.4 in [13], we can prove the existence of a subsequence still labeled $(u_\delta, c_\delta)_\delta$ such that

$$
\begin{aligned}
du_\delta &\rightharpoonup \nabla u \quad \text{weakly in } L^2(\Omega_T), \\
c_\delta \rightharpoonup c, &\quad dc_\delta \rightharpoonup \nabla c \quad \text{weakly in } L^2(\Omega_T).
\end{aligned}
$$

This ends the proof.

□

Theorem 4.5.1. *The function* (u,c) *defined in Proposition 4.5.1 is a weak solution of the system* (4.3)–(4.4) *in the sense of :*

$$
\begin{aligned}
\int_0^T \int_\Omega \left(u \frac{\partial \psi}{\partial t} - \nabla u \cdot \nabla \psi + u \nabla c \cdot \nabla \psi + u(1-u)\psi \right) dxdt & \qquad (4.40) \\
+ \int_\Omega u_0\, \psi(x,0)\, dx = 0, &
\end{aligned}
$$

$$
\int_0^T \int_\Omega \nabla c \cdot \nabla \psi \; dxdt = \int_0^T \int_\Omega \left(\frac{u}{u+1} - c \right) \psi\, dxdt, \qquad (4.41)
$$

for all test functions $\psi \in \mathcal{D}(\Omega \times [0,T))$.

Proof. Let $\psi \in \mathcal{D}(\Omega \times [0,T))$, and let

$$\begin{aligned}
G_{10}(\delta) &= -\int_0^T \int_\Omega u_\delta \frac{\partial \psi}{\partial t} dx\, dt - \int_\Omega u_\delta(x,0)\, \psi(x,0)\, dx,\\
G_{20}(\delta) &= \int_0^T \int_\Omega du_\delta \cdot \nabla\psi \; dx\, dt,\\
G_{30}(\delta) &= -\int_0^T \int_\Omega u_\delta \; dc_\delta \cdot \nabla\psi \; dx\, dt,\\
G_{40}(\delta) &= -\int_0^T \int_\Omega u_\delta\, (1-u_\delta)\, \psi \; dx\, dt,
\end{aligned}$$

and $\varepsilon(\delta) = -\left(G_{10}(\delta) + G_{20}(\delta) + G_{30}(\delta) + G_{40}(\delta)\right)$.

Let $\psi_K^n = \psi\left(t^n, x_K\right)$ for all $K \in \mathcal{T}$ and $n = 0,1,...,N$. Multiplying the scheme (4.17) by $\Delta t \psi_K^n$, using (4.13) and (4.16), and summing for K and n, we get

$$G_1(\delta) + G_2(\delta) + G_3(\delta) + G_4(\delta) = 0,$$

where

$$\begin{aligned}
G_1(\delta) &= \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(\mathfrak{L}\left(\frac{\Delta t\,(1-u_K^n)}{2}\right) \left(u_K^{n+1} - u_K^n\right)\right) \psi_K^n,\\
G_2(\delta) &= -\frac{1}{2} \sum_{n=0}^{N-1} \Delta t \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left((1-u_K^n)\left(u_K^n + u_K^{n+1}\right)\right) \psi_K^n,\\
G_3(\delta) &= -\sum_{n=0}^{N-1} \Delta t \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \left(\mathfrak{L}\left(\frac{Dc_{K,\sigma}^{n+1}}{2}\right) Du_{K,\sigma}^{n+1}\right) \psi_K^n,\\
G_4(\delta) &= \frac{1}{2} \sum_{n=0}^{N-1} \Delta t \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \tau_\sigma Dc_{K,\sigma}^{n+1} \left(u_K^{n+1} + u_L^{n+1}\right) \psi_K^n.
\end{aligned}$$

From the weak convergence of $(dc_\delta)_\delta$ to ∇c, $(du_\delta)_\delta$ to ∇u, and the strong convergence of $(u_\delta)_\delta$ to u in $L^2(\Omega_T)$, with the continuity of the function $x \mapsto x\,(1-x)$, we have when $\delta \to 0$

$$\varepsilon(\delta) \to \int_0^T \int_\Omega \left(u \frac{\partial \psi}{\partial t} - \nabla u \cdot \nabla \psi + u \nabla c \cdot \nabla\psi + u\,(1-u)\,\psi \right) dxdt + \int_\Omega u_0\, \psi(x,0)\, dx.$$

To prove the convergence of the scheme (4.17), we need to prove that $G_j(\delta) - G_{j0}(\delta) \to 0$ as $\delta \to 0$ for $j = 1,2,3,4$.

We begin by rewriting $G_1(\delta)$ as : $G_1(\delta) = G_{11}(\delta) + G_{12}(\delta)$, with

$$G_{11}(\delta) = \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(u_K^{n+1} - u_K^n\right) \psi_K^n,$$

$$G_{12}(\delta) = \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(\left(\mathfrak{L}\left(\frac{\Delta t\,(1-u_K^n)}{2} \right) - 1 \right) \left(u_K^{n+1} - u_K^n\right) \right) \psi_K^n.$$

The proof of $G_{11}(\delta) - G_{10}(\delta) \to 0$ as $\delta \to 0$ is discussed in the proof of the Proposition 2.5.2. Now, to prove that $G_1(\delta) - G_{10}(\delta) \to 0$ as $\delta \to 0$, it suffices to prove that $G_{12}(\delta) \to 0$ as $\delta \to 0$. Thus, since $\mathfrak{L}$ is a 1-Lipschitz continuous function, and using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
|G_{12}(\delta)| &\le \frac{1}{2} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \mathrm{m}(K)\,\Delta t\ |(1-u_K^n)| \left| \left(u_K^{n+1} - u_K^n\right) \right| |\psi_K^n| \\
&\le \|\psi\|_\infty \frac{\sqrt{\Delta t}}{2} \left(\sum_{n=0}^{N-1} \Delta t \sum_{K\in\mathcal{T}} \mathrm{m}(K)\,|1-u_K^n|^2 \right)^{1/2} \\
&\quad\times \left(\sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(u_K^{n+1} - u_K^n\right)^2 \right)^{1/2}.
\end{aligned}$$

Using the $L^2(\Omega_T)$ bound of u_δ and the estimate (4.33), we conclude that $G_{12}(\delta) \to 0$ as $\delta \to 0$.

For G_2, we have $G_2(\delta) = G_{21}(\delta) + G_{22}(\delta)$, with

$$G_{21}(\delta) = -\sum_{n=0}^{N-1} \Delta t \sum_{K\in\mathcal{T}} \mathrm{m}(K)\, u_K^{n+1} \left(1 - u_K^{n+1}\right) \psi_K^n,$$

$$\begin{aligned}
G_{22}(\delta) = &-\Delta t \sum_{K\in\mathcal{T}} \mathrm{m}(K)\, u_K^0 \left(1-u_K^0\right) \psi_K^n + \Delta t \sum_{K\in\mathcal{T}} \mathrm{m}(K)\, u_K^N \left(1-u_K^N\right) \psi_K^n \\
&-\frac{\Delta t}{2} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \mathrm{m}(K)\ (1-u_K^n) \left(u_K^{n+1} - u_K^n\right) \psi_K^n.
\end{aligned}$$

Now, since we have

$$G_{21}(\delta) - G_{20}(\delta) = \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \int_{t_n}^{t_{n+1}} \int_K u_K^{n+1} \left(1-u_K^{n+1}\right) \left(\psi_K^n - \psi\,(t,x)\right) dxdt,$$

we obtain, using the $L^2(\Omega_T)$ bound of u_δ and the regularity of ψ

$$|G_{21}(\delta) - G_{20}(\delta)| \le C\delta \|\psi\|_{C^1(\Omega_T)} \to 0, \quad \text{as } \delta \to 0.$$

Next, using the $L^\infty(0,T;L^2(\Omega))$ bound of u_δ and arguing similarly to the estimate of $|G_{12}(\delta)|$, we can easily see that

$$G_{22}(\delta) \le C\,\|\psi\|_\infty \left(\Delta t + \sqrt{\Delta t}\right) \to 0, \quad \text{as } \delta \to 0.$$

Hence, we prove that $G_2(\delta) - G_{20}(\delta) \to 0$ as $\delta \to 0$.

For the proof of $G_j(\delta) - G_{j0}(\delta) \to 0$ as $\delta \to 0$ when $j = 3, 4$, we refer to [6, Theorem 1]. The limit $\delta \to 0$ in the equation (4.18) is performed as in the proof of Proposition 2.5.2 . The treatment of the term $\dfrac{u_K^n}{u_K^n + 1}$ is straightforward, since it is bounded for all $K \in \mathcal{T}$ and $n = 0, ..., N-1$.

This ends the proof of the theorem.

□

4.6 Numerical experiments

4.6.1 Chemotaxis-growth model for embryonic pattern formation

At first, we compare the numerical results obtained from the two unconditionally stable and positivity preserving schemes : (4.8),(4.18)–(4.19) and (4.17)–(4.19). The numerical solutions are obtained by solving two linear systems at each time-step. For example, in the case of (4.17)–(4.19), and at each time-step, we begin by solving (4.18) to compute c_K^{n+1} and then, we compute u_K^{n+1} from (4.17). The same procedure is used for the semi-implicit Euler scheme (4.8),(4.18)–(4.19).

We consider that the two studied schemes include the parameters omitted by the assumption (4.5). Thus, the expression of the Il'in flux becomes :

$$F_{K,\sigma}^{n+1} = \tau_\sigma \mu \left(B\left(\frac{-\chi\, Dc_{K,\sigma}^{n+1}}{\mu} \right) u_K^{n+1} - B\left(\frac{\chi\, Dc_{K,\sigma}^{n+1}}{\mu} \right) u_L^{n+1} \right).$$

In all numerical experiments presented in this subsection, we consider a rectangular computational domain $\Omega = (0, 4) \times (0, 10)$, discretized via a uniform mesh of 25000 control volumes, and we take $T_f = 10$. We restrict us to the following parameter values which are picked in [59] : $\mu = 0.25$, $\chi = 13.4$, $s = 1$ and $r = 1.52$. The steady-state form of the system (4.2) generates stripe patterns when these values are adopted.

We consider also the following initial condition :

$$u_0(x) = \begin{cases} 1 + \epsilon(x) & \text{if } x \in (0,4) \times (4.8, 5.2), \\ 1 & \text{otherwise, with } x \in \Omega, \end{cases} \tag{4.42}$$

where $\epsilon(x)$ is a random perturbation which is constant on each cell of the mesh. In our test, it is equal on each control volume to the average of ten uniformly distributed random values in $[0, 1]$.

All computations of this section are performed in MATLAB 7.8.0. To reduce the cost of computations, the identity (4.25) is taken into account in the implementation of the proposed scheme. We note also that since the mesh considered is very fine, the value of $Dc_{K,\sigma}^{n+1}$ can be very close to 0. This may cause a severe instability if the term $B\left(\dfrac{\chi\, Dc_{K,\sigma}^{n+1}}{\mu} \right)$ is computed in a naive way. Indeed, the quantity $\exp(x) - 1$ can be 0 for a small $|x|$. This problem can be resolved by using the MATLAB function *expm1(x)*, which computes $\exp(x) - 1$ accurately when x is close to 0.

It is well known that, compared with the classical upwind discretization, a greater accuracy can be obtained by using the Il'in scheme, albeit at the expense of increasing complexity. For the time discretization, and contrary the two linearized implicit schemes (4.9),(4.18)–(4.19) and (4.10),(4.18)–(4.19), the proposed semi-exponentially fitted scheme has the advantage to ensure the stability and the positivity of the numerical solution unconditionally (see Section 4.3). However, these advantages are also provided by using the linearized backward Euler scheme (4.8),(4.18)–(4.19), which is less expensive in terms of computational cost. Indeed, it is easy to see that an exponentially fitted discretization is more time consuming than an Euler one because of the involved exponentials. Therefore, an efficiency comparison must be performed to assess the interest of the proposed scheme.

Now, to compare the accuracy of the two schemes, and since the exact solution of the chemotaxis system (4.3)–(4.4) is unavailable, we compute the reference solution via the analysed scheme (4.17)–(4.19) on a very fine time-stepping $\Delta t = 10^{-4}$ up to T_f. The relative L^2 errors are then calculated for the two schemes by choosing two sequences of time-stepping : $\Delta t = 10^{-j}$ with $j = 0, 1, 2, 3$ and $\Delta t = 5.10^{-j}$ with $j = 1, 2, 3$.

Δt	L^2-error exponential	Rate	L^2-error Euler	Rate
1	1.320×10^{-1}	—	1.372×10^{-1}	—
5.10^{-1}	1.029×10^{-1}	0.359	1.225×10^{-1}	0.164
10^{-1}	3.619×10^{-2}	0.649	6.196×10^{-2}	0.423
5.10^{-2}	1.985×10^{-2}	0.866	3.779×10^{-2}	0.713
10^{-2}	4.262×10^{-3}	0.956	9.089×10^{-3}	0.885
5.10^{-3}	2.130×10^{-3}	1.00	4.642×10^{-3}	0.969
10^{-3}	3.940×10^{-4}	1.05	9.137×10^{-4}	1.01

Table 4.1 Relative L^2-errors and convergence rates obtained for (u) by using the time semi-exponentially fitted scheme (4.17)–(4.19) and the linearized backward Euler scheme (4.8),(4.18)–(4.19).

We observe from Table 4.1 (see also Figure 4.2) that the semi-exponentially fitted scheme is more accurate than the linearized backward Euler scheme (4.8),(4.18)–(4.19). The presented results show also that, for both schemes, the approximations are already acceptable even for coarse time-steps. The two schemes are first order convergent for $\Delta t \to 0$.

In the tests performed in Table 4.1, the Euler scheme requires between 91.18 and 95.12% of the time needed by our scheme to compute the approximate solution. The question then arises is : to what extent does the accuracy of our time semi-exponential discretization justifies its complexity ? The results reported in Table 4.2 respond clearly to this question, and demonstrate the efficiency of such discretization. The presented results show that for each numerical solution computed via the linearized backward Euler scheme using Δt (see Table 4.1), we can find a suitable time-step Δt^*, for which our scheme is more accurate and needs less computational time. The ratio $\Gamma_{\Delta t, \Delta t^*}$ used in Table 4.2 is defined by :

$$\Gamma_{\Delta t^*, \Delta t} = \frac{\text{CPU}^*}{\text{CPU}} \times 100,$$

where CPU^* is the computational time for the scheme (4.17)–(4.19) using Δt^*, and CPU is the computational time for the scheme (4.8),(4.18)–(4.19) using Δt.

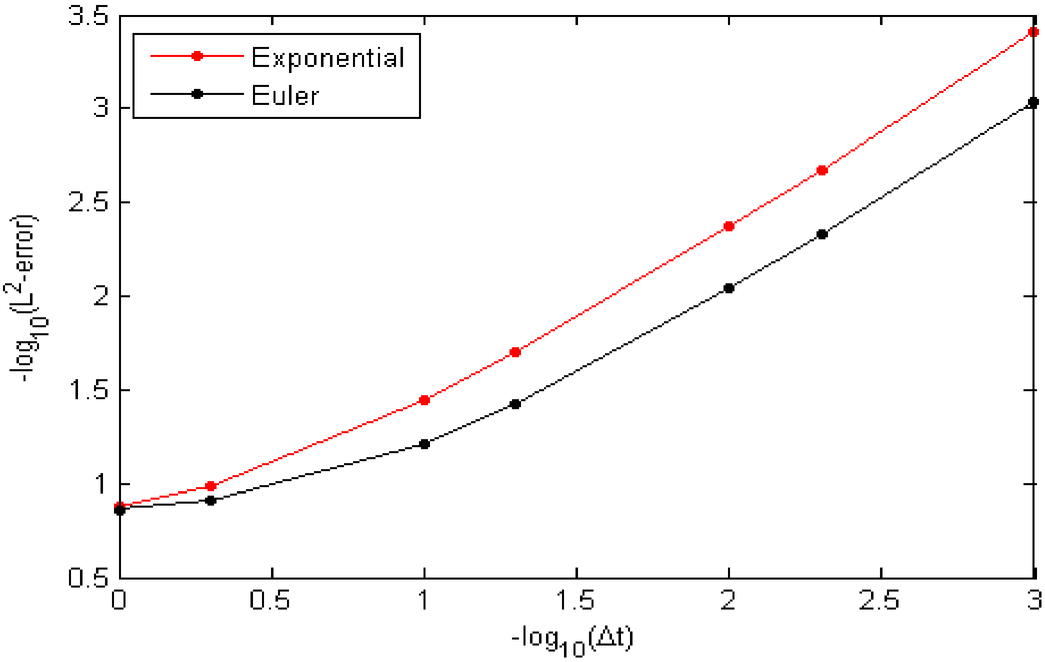

Figure 4.2 Relative L^2-errors and convergence rates obtained for (u) by using the time semi-exponentially fitted scheme (4.17)–(4.19) and the linearized backward Euler scheme (4.8),(4.18)–(4.19).

Δt^*	L^2-error	CPU* (s)	Δt	CPU (s)	$\Gamma_{\Delta t^*,\Delta t}$
1.111	1.357×10^{-1}	11.9	1	12.1	98.35%
6.25×10^{-1}	1.129×10^{-1}	17.5	5.10^{-1}	19.5	89.74%
2.10^{-1}	6.112×10^{-2}	46.5	10^{-1}	83.6	55.62%
10^{-1}	3.619×10^{-2}	86.0	5.10^{-2}	154.1	55.84%
2.10^{-2}	8.399×10^{-3}	499.8	10^{-2}	939.5	53.20%
10^{-2}	4.262×10^{-3}	997.4	5.10^{-3}	1876.1	53.16%
2.10^{-3}	8.304×10^{-4}	5041.1	10^{-3}	8960.8	56.25%

Table 4.2 Relative L^2-errors obtained for (u) by using the time semi-exponentially fitted scheme (4.17)–(4.19) for Δt^*. The computational time and the obtained results for the ratio $\Gamma_{\Delta t^*,\Delta t}$ are also presented for different values of Δt^* and Δt.

The strip patterns generated by the model (4.3)–(4.4) can be observed in Figure 4.3. The cell density u is computed at the final time $T_f = 10$, using the two compared schemes, with $\Delta t = 10^{-3}$. In Figure 4.4, three-dimensional plots illustrating the time evolution of the reference solution of u are presented. As we can see, the obtained numerical solution does not exhibit oscillatory behavior and has no negative values.

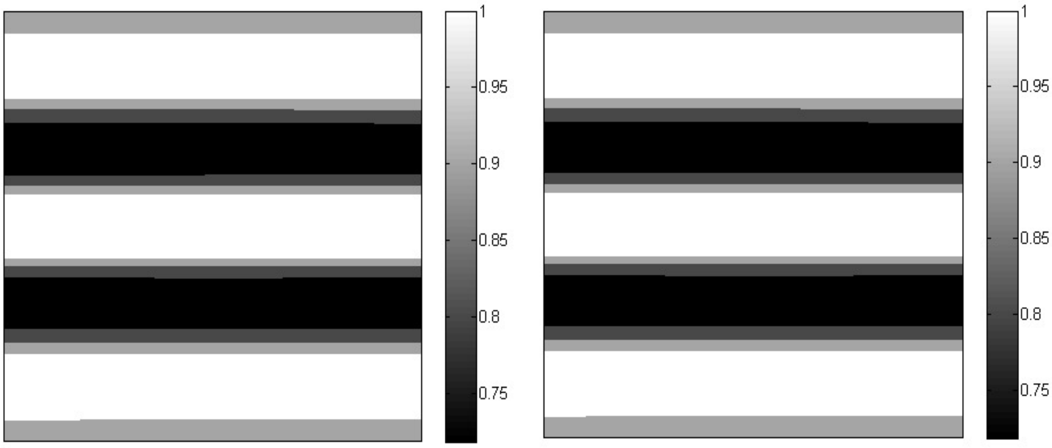

Figure 4.3 Cell density (u) at final time T_f computed via the scheme (4.17)–(4.19) (left) and the scheme (4.8),(4.18)–(4.19) (right) with $\Delta t = 10^{-3}$. White denotes high cell density ($u \geq 1$).

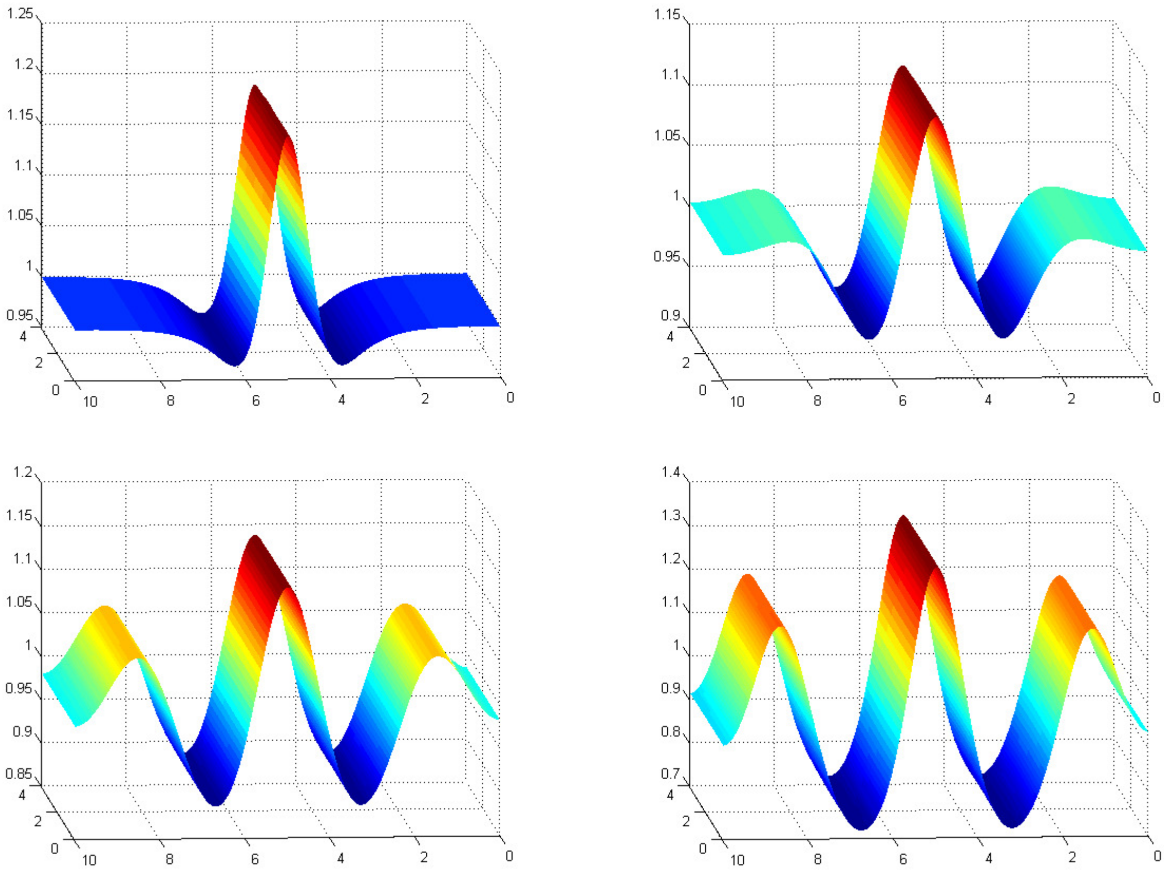

Figure 4.4 Reference solution of (u) at $t = 0.5$ (left top), $t = 2$ (right top), $t = 5$ (left bottom), T_f (right bottom).

4.6.2 Chemotaxis-growth model for bacterial pattern formation

In this subsection, we are concerned with the following chemotaxis system :

$$\begin{cases} \partial_t u = \mu\Delta u - \chi\nabla\cdot(\,u\nabla c) + f(u) \quad \text{in } \Omega\times(0,T_f), \\ \partial_t c = \Delta c - sc + u \quad \text{in } \Omega\times(0,T_f), \end{cases} \tag{4.43}$$

with the homogeneous Neumann boundary conditions. The term $f(u)$ is a logistic source representing the growth rate of u, and μ, χ and s are positive constants. The above model has the ability to describe the process of pattern formation performed by bacteria [1]. In all numerical tests, the computational domain is the square $\Omega = (-8,8)^2$, and is discretized with a uniform mesh grid 100×100. For all x in Ω, we choose the following initial conditions :

$$u(x,0) = \begin{cases} 1+\epsilon(x) & \text{if } \|x\|_2 < 0.7, \\ 1 & \text{otherwise,} \end{cases} \tag{4.44}$$

and $c(x,0) = 1/32$, $\epsilon(x)$ is defined as in (4.42).

Test 1. For this test, we take the following data : $\mu = 0.0625$, $\chi = 6$, $s = 16$, $f(u) = 2u(1-u)$ and $T_f = 30$. As before, we compare the numerical results obtained using the time semi-exponentially fitted scheme and the linearized backward Euler scheme. For the first equation of (4.43), the two schemes are similar to (4.17) and (4.8). The second equation of the model is approximated using an implicit Euler scheme with an explicit discretization of the cell density reaction term. The reference solution is computed via the time semi-exponentially fitted scheme with $\Delta t = 10^{-4}$.

The comparison between the two schemes is performed similarly as in the previous test and leads to the same conclusions. The results presented in Table 4.3 (see also Figure 4.5) show that the time semi-exponentially fitted scheme is more accurate than the linearized backward Euler scheme. Moreover, it appears from Table 4.4 that we need less computational time to reach a better accuracy using the exponentially fitted scheme. The definitions of CPU, CPU* and $\Gamma_{\Delta t^*,\Delta t}$ are similar to that of the previous subsection.

Δt	L^2-error exponential	Rate	L^2-error Euler	Rate
1	2.324×10^{-2}	—	2.489×10^{-2}	—
5.10^{-1}	1.720×10^{-2}	0.434	2.194×10^{-2}	0.182
10^{-1}	5.953×10^{-3}	0.659	1.096×10^{-2}	0.431
5.10^{-2}	3.257×10^{-3}	0.870	6.662×10^{-3}	0.718
10^{-2}	6.978×10^{-4}	0.957	1.594×10^{-3}	0.889
5.10^{-3}	3.488×10^{-4}	1.00	8.137×10^{-4}	0.970
10^{-3}	6.456×10^{-5}	1.05	1.604×10^{-4}	1.01

Table 4.3 Relative L^2-errors and convergence rates obtained for the solution (u) of (4.43) (*Test 1*) by using the time semi-exponentially fitted scheme and the linearized backward Euler scheme.

In Figure 4.6 , we plot the numerical cell density u of the model (4.43) computed using both schemes with $\Delta t = 10^{-3}$. As we can see, the solution forms a periodic arrays of continuous rings. Such formations are similar to the patterns formed by *Salmonella typhimurium* (see Figure 1.2).

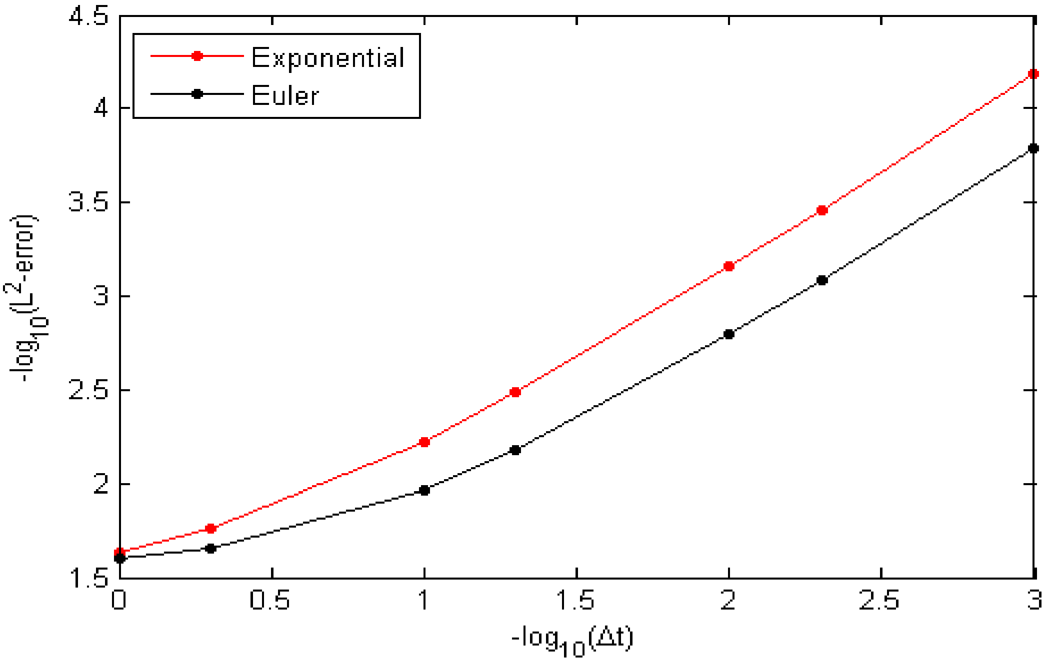

Figure 4.5 Relative L^2-errors and convergence rates obtained for the solution (u) of (4.43) (*Test 1*) by using the time semi-exponentially fitted scheme and the linearized backward Euler scheme.

Δt^*	L^2-error	CPU* (s)	Δt	CPU (s)	$\Gamma_{\Delta t^*,\Delta t}$
7.5×10^{-1}	2.041×10^{-2}	13.1	5.10^{-1}	17.2	76.16%
2.10^{-1}	1.010×10^{-2}	43.4	10^{-1}	71.2	60.95%
10^{-1}	5.953×10^{-3}	83.4	5.10^{-2}	139.5	59.78%
2.10^{-2}	1.376×10^{-3}	403.6	10^{-2}	679.5	59.4%
10^{-2}	6.978×10^{-4}	820.5	5.10^{-3}	1402.9	58.49%
2.10^{-3}	1.360×10^{-4}	3987.3	10^{-3}	7061.1	56.47%

Table 4.4 Relative L^2-errors obtained for the solution (u) of (4.43) (*Test 1*) by using the time semi-exponentially fitted scheme for Δt^*. The computational time and the obtained results for the ratio $\Gamma_{\Delta t^*,\Delta t}$ are also presented for different values of Δt^* and Δt.

Test 2. In some situations, capturing solutions of the model (4.43) may be very challenging due to their spiky behavior. To demonstrate the ability of our numerical approach for capturing such solutions, we consider in this test the following data picked in [1] : $\mu = 0.0625$, $s = 32$ and $f(u) = u^2\,(1-u)$. For the final time, we take $T_f = 100$.

The time semi-exponentially fitted scheme with the Il'in discretization of the convection-diffusion fluxes are then used for the approximation of the solutions of the system (4.43). Since the logistic source has now a cubic form, we adopt the following approximation :

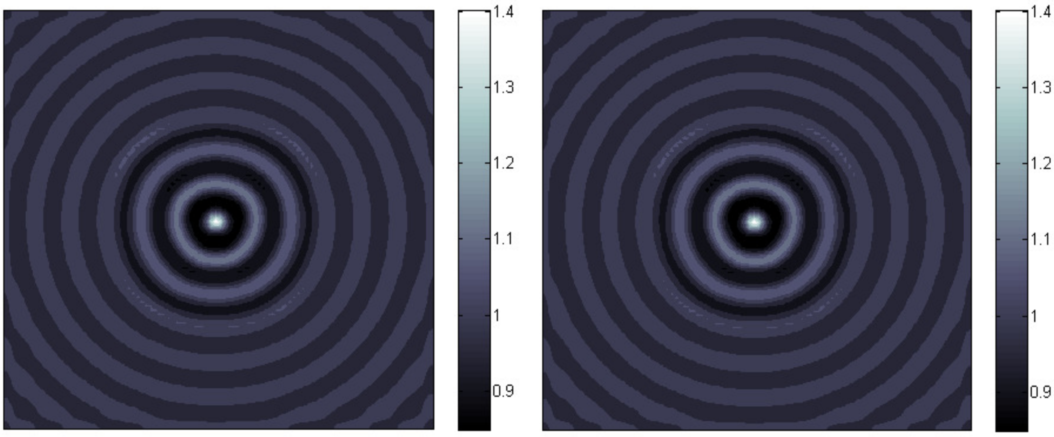

Figure 4.6 Solution (u) of (4.43) (*Test 1*) at final time T_f computed via the time semi-exponentially fitted scheme (left) and the linearized backward Euler scheme (right) with $\Delta t = 10^{-3}$.

$$\int_K \partial_t u(x,t)\,dx - \int_K u(x,t)^2\,(1-u(x,t))\;dx \approx \mathrm{m}(K)\frac{B\left(\Delta t\, u_K^n\,(1-u_K^n)\right)u_K^{n+1} - B\left(-\Delta t\, u_K^n\,(1-u_K^n)\right)u_K^n}{\Delta t},$$

for all $t \in [t_n, t_{n+1})$ and $x \in \Omega$.

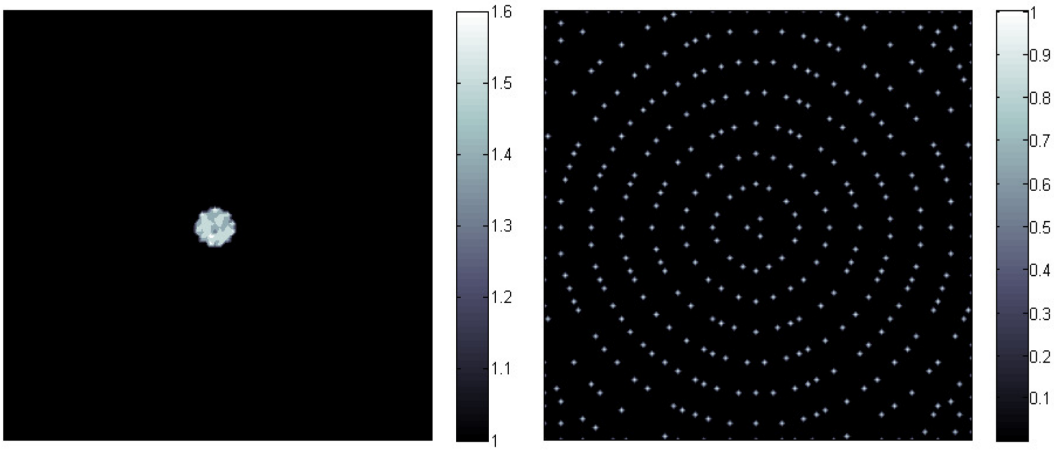

Figure 4.7 Initial condition (4.44) (left), and solution (u) of (4.43) (*Test 2*) by the time semi-exponentially fitted scheme at the final time T_f (right), for $\chi = 60$.

In all numerical tests, we take for the step size $\Delta t = 10^{-1}$. The initial cell density is shown in Figure 4.7 (left). The solution at the final time T_f is plotted in Figure 4.7 (right) in the case of $\chi = 60$. As one can see, spot patterns of remarkable regularity are shown in whole domain. The three-dimensional plots of the solution at $t = 1$ and at the final time are presented in Figure 4.8. The expected spiky behavior of the solution is obtained, and no negative numerical values have been developed.

In Figure 4.9, we examine the influence of the parameter χ on the solution. When $\chi = 8.8$, a

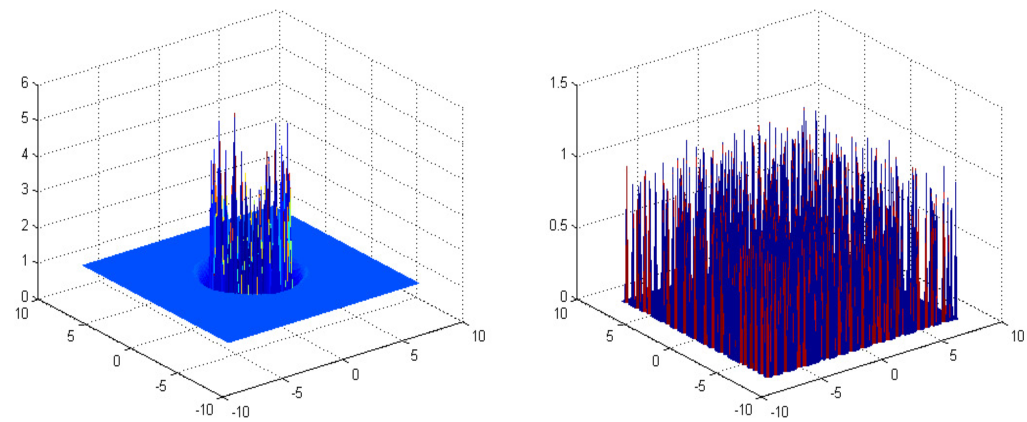

Figure 4.8 Three dimensional plots of the cell density (u) of (4.43) (*Test 2*) by the time semi-exponentially fitted scheme at $t = 1$ (left) and at the final time T_f (right), for $\chi = 60$.

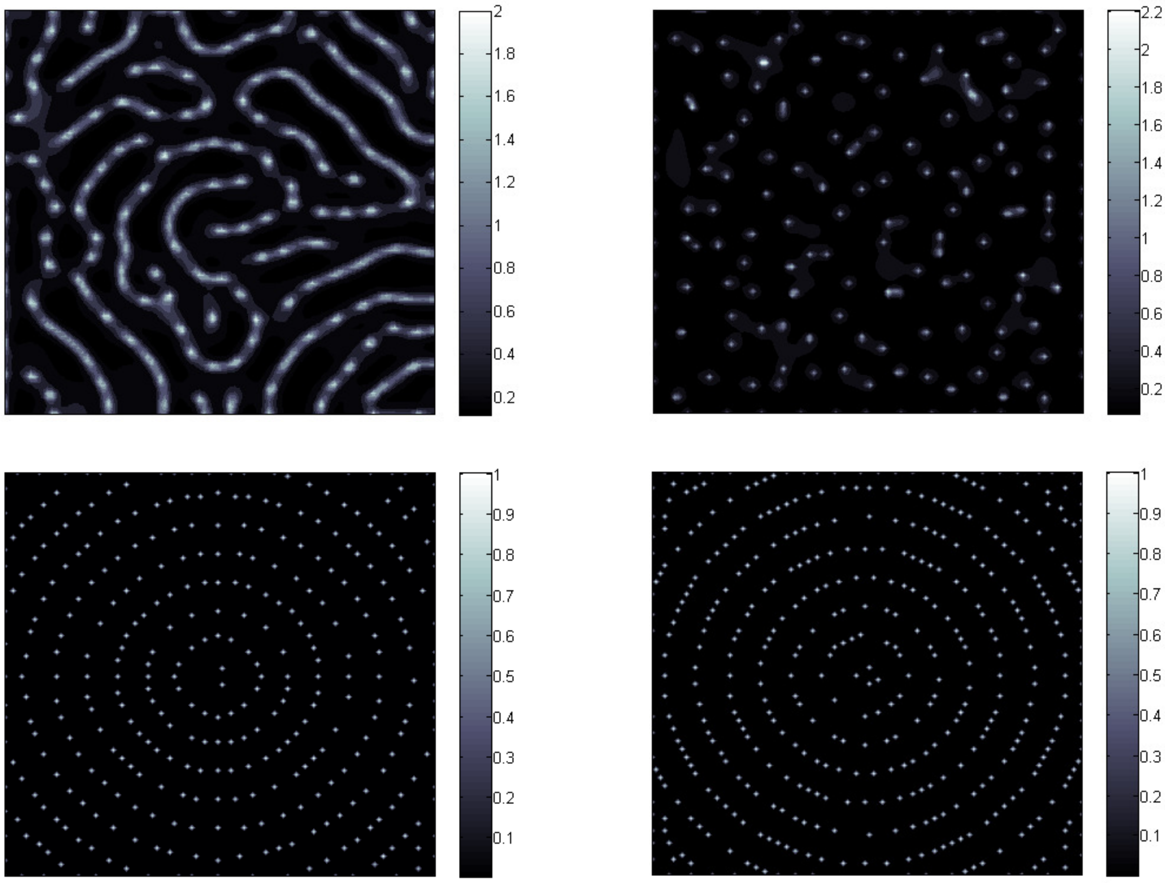

Figure 4.9 Cell density (u) of (4.43) (*Test 2*) by the time semi-exponentially fitted scheme at final time T_f, for $\chi = 8.8$ (left top), $\chi = 15$ (right top), $\chi = 50$ (left bottom), $\chi = 70$ (right bottom).

moving perforated labyrinthine pattern is observed. Then, the solution changes its structure to chaotic spots pattern for $\chi = 15$. Finally, Figure 4.9 (left bottom) and Figure 4.9 (right bottom) show that the symmetry of these spots increases for high values of χ. These symmetric spots seem in good agreement with the *Escherichia coli* patterns reported in [12] (see Figure 1.3).

Chapter 5

A corrected decoupled scheme for chemotaxis models

Sommaire

5.1 Introduction

In this chapter, we develop a decoupled finite volume scheme which can be applied to a class of chemotaxis models. For the convection-diffusion term, the approximation used is quite similar to the hybrid scheme of Spalding [81]. Concerning the time discretization, which is the main aim of this chapter, it is developed such that the scheme only requires to solve decoupled systems, which excludes fully implicit discretizations. We require also that the scheme converges without needing to fulfill any CFL condition, which is not the case of the fully explicit schemes. In the literature, a number of decoupled methods for the Keller-Segel model and its variants have been proposed (see, e.g., [76, 75, 84, 5, 42, 15]). In all these works, the time discretization is based on the classical backward Euler scheme with an explicit approximation of some terms to avoid coupling of the system. However, it is well known that the main drawback of this strategy is its lack of accuracy. A more efficient approach will be presented in this chapter.

This chapter provides also a convergence analysis of the proposed scheme applied to the chemotaxis system studied in Chapter 3 :

$$\begin{cases} \partial_t u = \mu\Delta u - a\nabla\cdot(\,u\nabla c) & \text{in } \Omega_T, \\ 0 = \Delta c + \dfrac{u}{u+1} - c & \text{in } \Omega_T, \end{cases} \tag{5.1}$$

with the homogeneous Neumann boundary and initial conditions

$$\nabla u\cdot\nu = \nabla c\cdot\nu = 0 \quad \text{on } \partial\Omega\times(0,T_f), \quad u(.,0) = u_0 \quad \text{in } \Omega, \tag{5.2}$$

where μ and a are positive constants, $\Omega_T := \Omega\times(0,T_f)$, $\Omega\subset\mathbb{R}^2$ is an open bounded subset with sufficiently regular boundary $\partial\Omega$, $T_f>0$ is a fixed time and ν denotes the outward unit-normal on the boundary $\partial\Omega$. It is proved that the convergence of the approximate solution can be obtained for any nonnegative initial cell density $u_0\in L^2(\Omega)$.

The outline of this chapter is as follows. In the next section, we present our corrected decoupled finite volume scheme to approximate the solution of (5.1)–(5.2). In section 5.3, we prove the existence and uniqueness of the solution of the proposed scheme. Positivity preservation and mass conservation are also shown in this section. A priori estimates are given in Section 5.4. In Section 5.5, we use these estimates to prove that the approximate solution converges to a weak solution of the studied model. Finally, we present some numerical tests in Section 5.6 and compare the accuracy of our approach with that of more usual decoupled schemes.

5.2 Presentation of the numerical scheme

We assume that $\Omega\subset\mathbb{R}^2$ is an open bounded polygonal subset. We consider in this chapter an admissible finite volume mesh of Ω, denoted by $\mathcal{T}$, which verifies the properties of Definition 5.1 in [26] (see also Définition 1.3.1). We will also use the same notations related to the spatial discretization and discrete norms as in Section 2.2.1.

The time discretization of $(0,T_f)$ is given by a uniform partition : $0=t_0<t_1<...<t_N=T_f$ with $N\in\mathbb{N}$ and $t_n=n\Delta t$ for $n=0,...,N$.

We denote by h the maximal size (diameter) of the control volumes included in $\mathcal{T}$, and we define

$$\delta = \max(\Delta t, h).$$

In Section 5.4, the following time-step condition will be used : there exists $\alpha>0$ such that

$$1-2a\Delta t \geq \alpha. \tag{5.3}$$

We will also need this additional constraint on the mesh : there exists $\xi>0$ such that

$$\mathrm{d}(x_K,\sigma) \geq \xi\,\mathrm{d}(x_K,x_L), \tag{5.4}$$

which is specially needed to apply the discrete Gagliardo-Nirenberg-Sobolev inequality (see Lemma 3.2.1).

We define a weak solution of the system (5.1) with boundary and initial conditions (5.2) as follows :

Definition 5.2.1. *A weak solution of the initial-boundary value problem* (5.1)–(5.2) *is a pair of nonnegative functions* $(u,c) \in L^2(0,T;H^1(\Omega))^2$ *which verify the following identities for all test functions* $\psi \in \mathcal{D}(\Omega \times [0,T_f))$ *:*

$$\int_0^T \int_\Omega (u\,\partial_t\psi - \mu\nabla u\cdot\nabla\psi + a\,u\nabla c\cdot\nabla\psi)\,dxdt + \int_\Omega u_0\,\psi(x,0)\,dx = 0, \tag{5.5}$$

$$\int_0^T \int_\Omega \nabla c\cdot\nabla\psi\ dxdt = \int_0^T \int_\Omega \left(\frac{u}{u+1} - c\right)\psi\,dxdt. \tag{5.6}$$

Now, we recall the classical decoupled finite volume scheme studied in Chapter 3 for the problem (5.1)–(5.2) :

for all $K \in \mathcal{T}$ and $n = 0, ..., N-1$

$$\mathrm{m}(K)\frac{u_K^{n+1} - u_K^n}{\Delta t} - \mu \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma D u_{K,\sigma}^{n+1}$$
$$+ a \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \tau_\sigma \left(S\left(Dc_{K,\sigma}^{n+1}\right)u_K^{n+1} - S\left(-Dc_{K,\sigma}^{n+1}\right)u_L^{n+1}\right) = 0, \tag{5.7}$$

$$-\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\, Dc_{K,\sigma}^{n+1} + \mathrm{m}(K)\,c_K^{n+1} = \mathrm{m}(K)\,\frac{u_K^n}{u_K^n+1}, \tag{5.8}$$

with the compatible initial condition

$$u_K^0 = \frac{1}{\mathrm{m}(K)}\int_K u_0(x)\,dx, \tag{5.9}$$

and where for all $\sigma \in \mathcal{E}_K$

$$Dv_{K,\sigma}^n = \begin{cases} 0 & \text{for } \sigma \subset \partial\Omega, \\ v_L^n - v_K^n & \text{otherwise },\ \sigma = K|L. \end{cases}$$

In the above scheme, the function S is defined by

$$S(x) = \begin{cases} 0, & \text{if } x < 2\,(-\mu+\varepsilon)\,/a, \\ \dfrac{x}{2}, & \text{if } |x| \le 2\,(\mu-\varepsilon)\,/a, \\ x, & \text{if },\, x > 2\,(\mu-\varepsilon)\,/a, \end{cases} \tag{5.10}$$

where ε is a small constant such that $\varepsilon \ge 0$ and $\varepsilon << \mu$.

The terms u_K^n and c_K^n denote respectively the approximations of $\frac{1}{\mathrm{m}(K)}\int_K u(x,t^n)\,dx$ and $\frac{1}{\mathrm{m}(K)}\int_K c(x,t^n)\,dx$. The discretization used for $\nabla\cdot(u\nabla c)$ is equivalent to the second order central difference scheme when $\left|Dc_{K,\sigma}^{n+1}\right| \le 2\,(\mu-\varepsilon)\,/a$ and to the first order upwind scheme when $Dc_{K,\sigma}^{n+1} < 2\,(-\mu+\varepsilon)\,/a$ or $Dc_{K,\sigma}^{n+1} > 2\,(\mu-\varepsilon)\,/a$. When $\varepsilon = 0$, the scheme is identical to that of Spalding [81] (see also [71]).

It is clear that we can obtain a best accuracy if we replace (5.8) by the equation

$$-\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\, Dc_{K,\sigma}^{n+1} + \mathrm{m}(K)\,c_K^{n+1} = \mathrm{m}(K)\,\frac{u_K^{n+1}}{u_K^{n+1}+1}, \tag{5.11}$$

however, it will be expensive in term of computational cost to find the solution of the scheme (5.7),(5.11) since we have to solve a large nonlinear system at each time step.

Now, for all $K \in \mathcal{T}$ and $n = 0, ..., N-1$, we define

$$T_K^{n+1} = \mathrm{m}(K) \left(\frac{u_K^{n+1}}{u_K^{n+1}+1} - \frac{u_K^n}{u_K^n+1} \right), \quad T_K^0 = 0. \tag{5.12}$$

The equation (5.11) can then be written as

$$- \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma \, Dc_{K,\sigma}^{n+1} + \mathrm{m}(K)\, c_K^{n+1} = \mathrm{m}(K)\, \frac{u_K^n}{u_K^n+1} + T_K^{n+1}. \tag{5.13}$$

As we can see, the only difference between (5.13) and (5.8) is the term T_K^{n+1}, so we conjecture that we can improve the accuracy of the decoupled scheme (5.7)–(5.8) if we add to the right hand side of (5.8) a correction term which approximates T_K^{n+1}. Hence, we propose to replace (5.8) with the following scheme :

$$- \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma \, Dc_{K,\sigma}^{n+1} + \mathrm{m}(K)\, c_K^{n+1} = \mathrm{m}(K)\, \frac{u_K^n}{u_K^n+1} + \beta_n T_K^n, \tag{5.14}$$

where $\beta_n > 0$. The purpose of β_n is to ensure the nonnegativity of the right-hand side of (5.14), to do not affect the nonnegativity of c_K^{n+1}. Then, by supposing that $u_K^n \geq 0$ (this will be proved in Section 5.3), we define β_n for $n = 0, ..., N-1$ as follows :

$$\beta_n = \begin{cases} 1, & \text{if } \mathcal{T}_n^* = \emptyset, \\ \min_{K \in \mathcal{T}_n^*} \dfrac{\dfrac{u_K^n}{u_K^n+1}}{\dfrac{u_K^{n-1}}{u_K^{n-1}+1} - \dfrac{u_K^n}{u_K^n+1}}, & \text{otherwise ,} \end{cases} \tag{5.15}$$

where $\mathcal{T}_n^* = \left\{ K \in \mathcal{T} \;\middle|\; 2\dfrac{u_K^n}{u_K^n+1} - \dfrac{u_K^{n-1}}{u_K^{n-1}+1} < 0 \right\}$ for $n = 0, ..., N-1$ and $\mathcal{T}_0^* = \emptyset$. We can easily verify that $0 < \beta_n \leq 1$. We mention however that, in practice, we will take $\beta_n = 1$, which seems the most natural choice. Indeed, several numerical tests are performed and, as expected, the right-hand side of (5.14) is always positive for this value unless the time-step size is extremely large.

We define u_δ, c_δ, the finite volume approximations of u and c by :

$$u_\delta(x,t) = u_K^{n+1}, \quad c_\delta(x,t) = c_K^{n+1}, \quad x \in K,\, t \in [t^n, t^{n+1}). \tag{5.16}$$

We define also approximations of the gradients of u and c. To this end, we begin by defining $M_{K,\sigma}$, which is the cell formed from the vertices of σ, x_K and x_L if $\sigma = K|L \not\subset \partial\Omega$, and from the vertices of σ and x_K if $\sigma \subset \partial\Omega$.

Following [13], we define the discrete gradient dv_δ which is the approximation of ∇v by

$$dv_\delta(x,t) = \frac{\mathrm{m}(\sigma)}{\mathrm{m}(M_{K,\sigma})}\, Dv_{K,\sigma}^{n+1}\, \nu_{K,\sigma}, \quad x \in M_{K,\sigma},\; t \in (t^n, t^{n+1}),$$

for all $K \in \mathcal{T}$ and $n = 0, ..., N-1$, where $\nu_{K,\sigma}$ denotes the unit normal on σ which is outward to K.

5.3 Existence and uniqueness of a discrete solution

In this section, we prove existence and uniqueness of the solution of the proposed scheme. We show also that the scheme is mass conserving and positivity preserving.

Proposition 5.3.1. *Assume that $u_0 \geq 0$. Then there exists a unique solution $\{(u_K^{n+1}, c_K^{n+1}),\, K \in \mathcal{T},\, n = 0, ..., N-1\}$ to the scheme (5.7),(5.14) which satisfies the following properties :*

$$u_K^{n+1} \geq 0 \quad \textit{and} \quad c_K^{n+1} \geq 0 \quad \textit{for all } K \in \mathcal{T},\, n = 0, ..., N-1, \tag{5.17}$$

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\, u_K^{n+1} = \sum_{K\in\mathcal{T}} \mathrm{m}(K)\, u_K^0 = \|u_0\|_{L^1(\Omega)}, \quad \textit{for all } n = 0, ..., N-1. \tag{5.18}$$

Proof. For all $n = 0, ..., N-1$, we define the vectors $U^{n+1} = \left(U_K^{n+1}\right)_{K\in\mathcal{T}}$, $F^n = (F_K^n)_{K\in\mathcal{T}}$, $C^{n+1} = \left(C_K^{n+1}\right)_{K\in\mathcal{T}}$ and $G^n = (G_K^n)_{K\in\mathcal{T}}$, for which :

$$\begin{aligned}
U_K^{n+1} &= u_K^{n+1}, \quad F_K^n = \mathrm{m}(K) u_K^n / \Delta t, \\
C_K^{n+1} &= c_K^{n+1}, \quad G_K^n = \mathrm{m}(K) \frac{u_K^n}{u_K^n + 1} + \beta_n T_K^n.
\end{aligned}$$

We also define the matrix $A^n = \left(A_{K,K}^n\right)_{K\in\mathcal{T}}$ and $B = (B_{K,K})_{K\in\mathcal{T}}$ by

$$\begin{aligned}
A_{K,K}^n &= \mathrm{m}(K)/\Delta t + \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma\not\subset\partial\Omega}} \tau_\sigma \left(\mu + a\, S\left(D c_{K,\sigma}^{n+1}\right)\right), \\
A_{K,L}^n &= -\tau_\sigma \left(\mu + a\, S\left(-D c_{K,\sigma}^{n+1}\right)\right), \quad \text{if } L \in \mathcal{N}(K) \text{ with } \sigma = K|L, \\
A_{K,L}^n &= 0, \quad \text{otherwise,}
\end{aligned}$$

and

$$\begin{aligned}
B_{K,K} &= \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma\not\subset\partial\Omega}} \tau_\sigma + \mathrm{m}(K), \\
B_{K,L} &= -\tau_\sigma, \quad \text{if } L \in \mathcal{N}(K) \text{ with } \sigma = K|L, \\
B_{K,L} &= 0, \quad \text{otherwise.}
\end{aligned}$$

The decoupled scheme (5.7),(5.14) can be then written equivalently as

$$A^n\, U^{n+1} = F^n, \tag{5.19}$$

$$B C^{n+1} = G^n. \tag{5.20}$$

To prove the desired results, we proceed by induction on n. The proof in the case of $n = 0$ is similar to that of the inductive step. We argue now that the vectors U^n and C^n are defined and nonnegative. We have

$$|B_{K,K}| - \sum_{\substack{L\in\mathcal{T}\\ L\neq K}} |B_{K,L}| = \mathrm{m}(K) > 0,$$

which means the matrix B is strictly diagonally dominant by rows. Moreover, since the diagonal elements of B are positive and its off-diagonal entries are nonpositive, we can conclude that B is a nonsingular M-matrix, which implies the unique solvability of (5.20) with $B^{-1} > 0$. Therefore, in view of the nonnegativity of G^n (by the definition of β_n (5.15)), we deduce that $C^{n+1} \geq 0$. Since for all $\sigma = K|L \in \mathcal{E}_K$, $Dc^{n+1}_{L,\sigma} = -Dc^{n+1}_{K,\sigma}$, we have for all $K \in \mathcal{T}$

$$|A^n_{K,K}| - \sum_{\substack{L\in\mathcal{T}\\ L\neq K}} |A^n_{L,K}| = \mathrm{m}(K)/\Delta t > 0.$$

Then A^n is strictly diagonal dominant by columns. Now, since $\mu + a\,S\,(x) \geq 0$ for all $x \in \mathbb{R}$ (by the definition (5.10)), the matrix A^n has nonpositive off-diagonal and positive diagonal entries, which implies that A^n is a nonsingular M-matrix. Consequently, the existence and uniqueness of U^{n+1} is proved, and since $F^n \geq 0$, it is clear that $U^{n+1} \geq 0$.

Now, summing (5.7) over $K \in \mathcal{T}$, we have :

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(u^{n+1}_K - u^n_K\right) = 0.$$

Then, summing over $n = 0, 1, ..., p$, with $p \leq N-1$ we get (5.18) :

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\, u^{p+1}_K = \sum_{K\in\mathcal{T}} \mathrm{m}(K)\, u^0_K = \|u_0\|_{L^1(\Omega)}, \quad \text{for all } p = 0, ..., N-1\,,$$

which ends the proof.

5.4 A priori estimates

In all this section, we assume that $u_0 \in L^2(\Omega)$ and $u_0 \geq 0$. We assume also that $\{\left(u^{n+1}_K, c^{n+1}_K\right),\ K \in \mathcal{T},\ n = 0, ..., N-1\}$ is the solution of the scheme (5.7),(5.14).

Proposition 5.4.1. *For all $n = 0, ..., N-1$ and for all $K \in \mathcal{T}$, we have*

$$c^{n+1}_K \leq 2, \tag{5.21}$$

$$\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma \left|Dc^{n+1}_{K,\sigma}\right|^2 \leq C. \tag{5.22}$$

where C is a positive constant which only depends on Ω.

Proof.

Let K be the control volume which verifies $c^{n+1}_K = \max\{c^{n+1}_L\}_{L\in\mathcal{T}}$. Multiplying the equation (5.14) by $\left(c^{n+1}_K - 2\right)^+$, and using the fact that $0 < \beta_n \leq 1$, we get

$$-\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma Dc^{n+1}_{K,\sigma} \left(c^{n+1}_K - 2\right)^+ \leq \mathrm{m}(K)\left(2 - c^{n+1}_K\right)\left(c^{n+1}_K - 2\right)^+,$$

it follows that

$$-\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma Dc_{K,\sigma}^{n+1}\left(c_K^{n+1}-2\right)^+ \leq 0.$$

In view of the choice of K, $Dc_{K,\sigma}^{n+1}\leq 0$ for all $\sigma\in\mathcal{E}_K$, which leads to

$$-\sum_{\sigma\in\mathcal{E}_K} \tau_\sigma Dc_{K,\sigma}^{n+1}\left(c_K^{n+1}-2\right)^+ \geq 0.$$

Hence, we deduce that $\left(c_K^{n+1}-2\right)\leq 0$. This establish (5.21).

Now, multiplying the equation (5.14) by c_K^{n+1}, summing over $K\in\mathcal{T}$, using a summation by parts and (5.21), we get

$$\frac{1}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma\left|Dc_{K,\sigma}^{n+1}\right|^2+\sum_{K\in\mathcal{T}}\mathrm{m}(K)\left|c_K^{n+1}\right|^2\leq 4\mathrm{m}(\Omega),$$

which gives (5.22).

Lemma 5.4.1. *Assume that* (5.4) *is fulfilled and that* $\varepsilon>0$ *(see* (5.10)*). Then, there exists a constant* $C>0$ *only depending on* Ω*,* T_f*,*u_0*,* a*,* ε *and* ξ *such that*

$$\sum_{n=0}^{N}\sum_{K\in\mathcal{T}}\Delta t\ \mathrm{m}(K)\left|u_K^n\right|^2\leq C. \tag{5.23}$$

Proof. Performing the following changes of variable :

$$\tilde{u}_K^n=u_K^n+1 \quad \text{for all } K\in\mathcal{T},\ n=0,...,N, \tag{5.24}$$

and

$$\tilde{S}(x)=S(x)+\left(\mu-\varepsilon\right)/a, \tag{5.25}$$

and using the identity

$$S(x)-S(-x)=x, \tag{5.26}$$

the scheme (5.7) becomes

$$\begin{aligned}
&\mathrm{m}(K)\frac{\tilde{u}_K^{n+1}-\tilde{u}_K^n}{\Delta t}-\varepsilon\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma D\tilde{u}_{K,\sigma}^{n+1}\\
&+a\sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}}\tau_\sigma\left(\tilde{S}\left(Dc_{K,\sigma}^{n+1}\right)\tilde{u}_K^{n+1}-\tilde{S}\left(-Dc_{K,\sigma}^{n+1}\right)\tilde{u}_L^{n+1}\right)-a\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma Dc_{K,\sigma}^n=0.
\end{aligned} \tag{5.27}$$

From (5.14), using (5.21) and since $0<\beta_n\leq 1$, we can easily see that

$$\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma\, Dc_{K,\sigma}^n\leq 3\mathrm{m}(K).$$

Using the above inequality in (5.27), we obtain

$$\begin{aligned}
&\mathrm{m}(K)\frac{\tilde{u}_K^{n+1}-\tilde{u}_K^n}{\Delta t}-\varepsilon\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma D\tilde{u}_{K,\sigma}^{n+1}\\
&\leq -a\sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}}\tau_\sigma\left(\tilde{S}\left(Dc_{K,\sigma}^{n+1}\right)\tilde{u}_K^{n+1}-\tilde{S}\left(-Dc_{K,\sigma}^{n+1}\right)\tilde{u}_L^{n+1}\right)+3a\,\mathrm{m}(K)\,.
\end{aligned}$$

Let us now multiply the above inequality by $\Delta t\, \log\left(\tilde{u}_K^{n+1}\right)$, and summing over $K \in \mathcal{T}$, we find that

$$E_1 + E_2 \leq E_3 + E_4, \tag{5.28}$$

where

$$\begin{aligned}
E_1 &= \sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(\tilde{u}_K^{n+1} - \tilde{u}_K^n\right)\log\left(\tilde{u}_K^{n+1}\right),\\
E_2 &= -\varepsilon \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma D\tilde{u}_{K,\sigma}^{n+1}\log\left(\tilde{u}_K^{n+1}\right),\\
E_3 &= -a \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \Delta t\, \tau_\sigma \left(\tilde{S}\left(Dc_{K,\sigma}^{n+1}\right)\tilde{u}_K^{n+1} - \tilde{S}\left(-Dc_{K,\sigma}^{n+1}\right)\tilde{u}_L^{n+1}\right)\log\left(\tilde{u}_K^{n+1}\right),\\
E_4 &= 3a\sum_{K\in\mathcal{T}} \Delta t\, \mathrm{m}(K)\log\left(\tilde{u}_K^{n+1}\right).
\end{aligned}$$

From the expression of E_1, we can see that

$$\begin{aligned}
E_1 = &\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(\tilde{u}_K^{n+1}\log\left(\tilde{u}_K^{n+1}\right) - \tilde{u}_K^n\log\left(\tilde{u}_K^n\right)\right)\\
&- \sum_{K\in\mathcal{T}} \mathrm{m}(K)\tilde{u}_K^n\left(\log\left(\tilde{u}_K^{n+1}\right) - \log\left(\tilde{u}_K^n\right)\right),
\end{aligned}$$

and since we have from a Taylor expansion of log

$$\tilde{u}_K^n\left(\log\left(\tilde{u}_K^{n+1}\right) - \log\left(\tilde{u}_K^n\right)\right) \leq \tilde{u}_K^{n+1} - \tilde{u}_K^n,$$

we deduce using the mass conservation property (5.18) that

$$E_1 \geq \sum_{K\in\mathcal{T}} \mathrm{m}(K)\left(\tilde{u}_K^{n+1}\log\left(\tilde{u}_K^{n+1}\right) - \tilde{u}_K^n\log\left(\tilde{u}_K^n\right)\right).$$

By an integration by parts on E_2, we get

$$E_2 = \frac{\varepsilon}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma D\tilde{u}_{K,\sigma}^{n+1}\left(\log\left(\tilde{u}_L^{n+1}\right) - \log\left(\tilde{u}_K^{n+1}\right)\right).$$

Then, by a Taylor expansion of log we get

$$E_2 = \frac{\varepsilon}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma \left|\frac{D\tilde{u}_{K,\sigma}^{n+1}}{\sqrt{\theta_\sigma^{n+1}}}\right|^2,$$

where $\theta_\sigma^{n+1} = t_\sigma \tilde{u}_K^{n+1} + (1-t_\sigma)\,\tilde{u}_L^{n+1}$ with $t_\sigma \in (0,1)$. Hence, using the fact that

$$\frac{\left|D\tilde{u}_{K,\sigma}^{n+1}\right|}{\sqrt{\theta_\sigma^{n+1}}} = \frac{\sqrt{\tilde{u}_K^{n+1}} + \sqrt{\tilde{u}_L^{n+1}}}{\sqrt{\theta_\sigma^{n+1}}}\left|D\left(\sqrt{\tilde{u}^{n+1}}\right)_{K,\sigma}\right| \geq \left|D\left(\sqrt{\tilde{u}^{n+1}}\right)_{K,\sigma}\right|,$$

we infer that

$$E_2 \geq \frac{\varepsilon}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma \left|D\left(\sqrt{\tilde{u}^{n+1}}\right)_{K,\sigma}\right|^2.$$

Using a summation by parts, we obtain

$$E_3^{n+1} = \frac{a}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \Delta t\, \tau_\sigma \left(\tilde{S}\left(Dc_{K,\sigma}^{n+1}\right) \tilde{u}_K^{n+1} - \tilde{S}\left(-Dc_{K,\sigma}^{n+1}\right) \tilde{u}_L^{n+1} \right)$$
$$\times \left(\log\left(\tilde{u}_L^{n+1}\right) - \log\left(\tilde{u}_K^{n+1}\right)\right).$$

Now, we define $\tilde{E}_3^{n+1}$

$$\tilde{E}_3^{n+1} = \frac{a}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \Delta t\, \tau_\sigma \theta_\sigma^{n+1} Dc_{K,\sigma}^{n+1} \left(\log\left(\tilde{u}_L^{n+1}\right) - \log\left(\tilde{u}_K^{n+1}\right)\right).$$

Next, using the identity (5.26) and the last expression of E_3^{n+1}, we can write

$$\begin{aligned}
E_3^{n+1} - \tilde{E}_3^{n+1} =& \frac{a}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \Delta t\, \tau_\sigma \tilde{S}\left(Dc_{K,\sigma}^{n+1}\right) \left(\tilde{u}_K^{n+1} - \theta_\sigma^{n+1}\right)\\
&\times \left(\log\left(\tilde{u}_L^{n+1}\right) - \log\left(\tilde{u}_K^{n+1}\right)\right) + \frac{a}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \Delta t\, \tau_\sigma \tilde{S}\left(-Dc_{K,\sigma}^{n+1}\right)\\
&\times \left(\theta_\sigma^{n+1} - \tilde{u}_L^{n+1}\right) \left(\log\left(\tilde{u}_L^{n+1}\right) - \log\left(\tilde{u}_K^{n+1}\right)\right)\\
=& \frac{a\left(1-t_\sigma\right)}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \Delta t\, \tau_\sigma \tilde{S}\left(Dc_{K,\sigma}^{n+1}\right) \left(\tilde{u}_K^{n+1} - \tilde{u}_L^{n+1}\right)\\
&\times \left(\log\left(\tilde{u}_L^{n+1}\right) - \log\left(\tilde{u}_K^{n+1}\right)\right) + \frac{a\, t_\sigma}{2} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \Delta t\, \tau_\sigma \tilde{S}\left(-Dc_{K,\sigma}^{n+1}\right)\\
&\times \left(\tilde{u}_K^{n+1} - \tilde{u}_L^{n+1}\right) \left(\log\left(\tilde{u}_L^{n+1}\right) - \log\left(\tilde{u}_K^{n+1}\right)\right).
\end{aligned}$$

Since we have from (5.10) that $S(x) \geq (-\mu+\varepsilon)/a$ $\forall x\in\mathbb{R}$, it follows that $\tilde{S}(x) \geq 0$. Consequently, the above equality implies that $E_3^{n+1} \leq \tilde{E}_3^{n+1}$, which yields by a Taylor expansion of log to

$$E_3^{n+1} \leq \frac{a}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma Dc_{K,\sigma}^{n+1} D\tilde{u}_{K,\sigma}^{n+1}.$$

Multiplying (5.14) by $\tilde{u}_K^{n+1}$ and summing by parts we can easily see that

$$E_3^{n+1} \leq 2a\Delta t \left\|u_0+1\right\|_{L^1(\Omega)}.$$

Finally for E_4, we have

$$E_4 \leq 3a\, \Delta t \|u_0+1\|_{L^1(\Omega)}.$$

Collecting the estimates obtained for E_1, E_2, E_3 and E_4, and summing over $n = 0, ..., N-1$, we obtain

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K) \tilde{u}_K^N \log\left(\tilde{u}_K^N\right) + \frac{\varepsilon}{2} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma \left| D\left(\sqrt{\tilde{u}^{n+1}}\right)_{K,\sigma} \right|^2$$
$$\leq 5a\, T \|u_0+1\|_{L^1(\Omega)} + \sum_{K\in\mathcal{T}} \mathrm{m}(K) \tilde{u}_K^0 \log\left(\tilde{u}_K^0\right).$$

Therefore, we have

$$\sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t\, \tau_\sigma \left| D\left(\sqrt{\tilde{u}^{n+1}}\right)_{K,\sigma} \right|^2 \leq C, \tag{5.29}$$

where $C > 0$ is a constant depending on Ω, T_f,u_0, a and ε .

Now, we denote by $\tilde{u}_\delta^{n+1} \in X(\mathcal{T})$ the function defined by : $\tilde{u}_\delta^{n+1}(x) = \tilde{u}_K^{n+1}$ for all $x \in K$ and $n = 0, ..., N-1$,

From Lemma 3.2.1, we have

$$\left\| \sqrt{\tilde{u}_\delta^{n+1}} \right\|_4 \leq \frac{C}{\xi^{1/4}} \left\| \sqrt{\tilde{u}_\delta^{n+1}} \right\|_{1,2,\mathcal{T}}^{1/2} \left\| \sqrt{\tilde{u}_\delta^{n+1}} \right\|_2^{1/2},$$

which implies that

$$\| \tilde{u}_\delta^{n+1} \|_2^2 \leq \frac{C}{\xi} \left\| \sqrt{\tilde{u}_\delta^{n+1}} \right\|_{1,2,\mathcal{T}}^2 \| \tilde{u}_\delta^{n+1} \|_1,$$

where $C > 0$ depends on Ω.

Finally, gathering the above inequality with (5.29) and (5.18), we deduce that u_δ is bounded in $L^2(\Omega_T)$. This completes the proof of the lemma.

Proposition 5.4.2. *Assume that* (5.4) *is fulfilled and that* $\varepsilon > 0$. *Then, there exists a constant* $C > 0$ *depending on* Ω, T_f,u_0, a, μ *and* ξ *such that, for all* $n = 0, 1, ..., p$, *with* $p \leq N-1$ *:*

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K) \left| u_K^{p+1} \right|^2 + \sum_{n=0}^{p} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t\ \tau_\sigma \left| Du_{K,\sigma}^{n+1} \right|^2 \leq C, \tag{5.30}$$

$$\sum_{n=0}^{p} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(u_K^{n+1} - u_K^n \right)^2 \leq C. \tag{5.31}$$

Proof.

Multiplying the equation (5.7) by $\Delta t\, u_K^{n+1}$, and summing over $K \in \mathcal{T}$, we have

$$I_1 + I_2 = I_3, \tag{5.32}$$

where

$$I_1 = \sum_{K\in\mathcal{T}} \mathrm{m}(K) u_K^{n+1} \left(u_K^{n+1} - u_K^n \right),$$

$$I_2 = -\mu \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t\ \tau_\sigma\ u_K^{n+1} Du_{K,\sigma}^{n+1},$$

$$I_3 = -a \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t\ \tau_\sigma u_K^{n+1} \left(S\left(Dc_{K,\sigma}^{n+1}\right) u_K^{n+1} - S\left(-Dc_{K,\sigma}^{n+1}\right) u_L^{n+1} \right).$$

Performing a summation by parts on I_2, yields,

$$I_2 = \frac{\mu}{2} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t\ \tau_\sigma\ \left| Du_{K,\sigma}^{n+1} \right|^2.$$

A summation by parts on I_3 gives

$$I_3 = \frac{a}{2} \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \Delta t \; \tau_\sigma \left(S\left(Dc_{K,\sigma}^{n+1}\right) u_K^{n+1} - S\left(-Dc_{K,\sigma}^{n+1}\right) u_L^{n+1} \right) Du_{K,\sigma}^{n+1}.$$

Now, using (5.26), we can easily verify that

$$\begin{aligned}
&2I_3 - \frac{a}{2} \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \Delta t \, \tau_\sigma Dc_{K,\sigma}^{n} \left(u_L^{n+1} + u_K^{n+1}\right) Du_{K,\sigma}^{n+1} \\
&\leq \frac{a}{2} \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \tau_\sigma \Delta t \left(S\left(Dc_{K,\sigma}^{n+1}\right) + S\left(-Dc_{K,\sigma}^{n+1}\right) \right) \left(u_K^{n+1} - u_L^{n+1}\right) Du_{K,\sigma}^{n+1} \\
&\leq 0,
\end{aligned}$$

and since by multiplying (5.14) by $\left|u_K^{n+1}\right|^2$, and sum by parts, we have

$$\frac{1}{2} \sum_{K \in \mathcal{T}} \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \tau_\sigma Dc_{K,\sigma}^{n+1} \left(\left|u_L^{n+1}\right|^2 - \left|u_K^{n+1}\right|^2 \right) \leq 2 \sum_{K \in \mathcal{T}} \mathrm{m}(K) \left|u_K^{n+1}\right|^2,$$

we deduce that

$$I_3 \leq a \Delta t \sum_{K \in \mathcal{T}} \mathrm{m}(K) \left|u_K^{n+1}\right|^2.$$

Employing the previous estimates for I_2 and I_3 in (5.32), and summing over $n = 0, 1, ..., p$, with $p \leq N - 1$, we obtain

$$\begin{aligned}
&\frac{1}{2} \sum_{K \in \mathcal{T}} \mathrm{m}(K) \left|u_K^{p+1}\right|^2 + \frac{1}{2} \sum_{n=0}^{p} \sum_{K \in \mathcal{T}} \mathrm{m}(K) \left(u_K^{n+1} - u_K^{n}\right)^2 \\
&+ \frac{\mu}{2} \sum_{n=0}^{p} \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \Delta t \; \tau_\sigma \left|Du_{K,\sigma}^{n+1}\right|^2 \\
&\leq \frac{1}{2} \sum_{K \in \mathcal{T}} \mathrm{m}(K) \left|u_K^{0}\right|^2 + a \sum_{n=1}^{p+1} \sum_{K \in \mathcal{T}} \Delta t \, \mathrm{m}(K) \left|u_K^{n}\right|^2. \qquad (5.33)
\end{aligned}$$

It follows from Lemma 5.4.1 that the right hand side of the above inequality is bounded, which gives the desired results.

Proposition 5.4.3. *Assume that the time step condition* (5.3) *is fulfilled. Then, the estimates* (5.30) *and* (5.31) *hold for a constant* $C > 0$ *depending on* Ω, T_f, u_0, a , μ *and* α.

Proof. From (5.33), and since (5.3) is verified, we have

$$\sum_{K \in \mathcal{T}} \mathrm{m}(K) \left|u_K^{p+1}\right|^2 \leq \sum_{n=1}^{p} \frac{2a\Delta t}{\alpha} \sum_{K \in \mathcal{T}} \mathrm{m}(K) \left|u_K^{n}\right|^2 + \frac{\|u_0\|_2}{\alpha}.$$

Using Lemma 2.2.1, we get

$$\sum_{K\in\mathcal{T}} \mathrm{m}(K)\left|u_K^{p+1}\right|^2 \le \frac{\|u_0\|_2}{\alpha}\exp\left(\frac{2aT_f}{\alpha}\right),$$

which with (5.33) completes the proof.

5.5 Convergence of the finite volume scheme

Proposition 5.5.1. *Assume that $u_0 \in L^2(\Omega)$ and $u_0 \ge 0$. Assume that the time step condition (5.3) is fulfilled or that (5.4) is fulfilled and $\varepsilon > 0$ (see (5.10)). Then, there exists a subsequence of $(u_\delta, c_\delta)_\delta$, not relabeled, and a function $(u, c) \in L^2(0, T; H^1(\Omega))^2$ such that as δ goes to zero we have*

$$\begin{aligned} u_\delta \to u \quad &\textit{strongly in } L^2(\Omega_T),\\ du_\delta \rightharpoonup \nabla u \quad &\textit{weakly in } L^2(\Omega_T),\\ c_\delta \rightharpoonup c, \quad dc_\delta \rightharpoonup \nabla c \quad &\textit{weakly in } L^2(\Omega_T). \end{aligned}$$

Proof. The proof will be omitted since it is exactly similar to that of Proposition 4.1 in [28]. We need only to note that $S(x) \le |x|$ for all $x \in \mathbb{R}$.

Theorem 5.5.1. *Assume that $u_0 \in L^2(\Omega)$ and $u_0 \ge 0$. Assume that the time step condition (5.3) is fulfilled or that (5.4) is fulfilled and $\varepsilon > 0$ (see (5.10)). The function (u, c) constructed in Proposition 5.5.1 is a weak solution of (5.1)–(5.2) in the sense of Definition 5.2.1 .*

Proof. Let $\psi \in \mathcal{D}(\Omega \times [0, T_f))$, and let

$$\begin{aligned} J_{10}(\delta) &= -\int_0^T\int_\Omega u_\delta \frac{\partial\psi}{\partial t} dx\,dt - \int_\Omega u_\delta(x,0)\,\psi(x,0)\,dx,\\ J_{20}(\delta) &= \mu\int_0^T\int_\Omega du_\delta\cdot\nabla\psi\;dx\,dt,\\ J_{30}(\delta) &= -a\int_0^T\int_\Omega u_\delta\;dc_\delta\cdot\nabla\psi\;dx\,dt, \end{aligned}$$

with $\lambda(\delta) = -\left(J_{10}(\delta) + J_{20}(\delta) + J_{30}(\delta)\right)$.

Let $\psi_K^n = \psi\left(x_K, t^n\right)$ for all $K \in \mathcal{T}$ and $n = 0, 1, ..., N$. Multiplying the scheme (5.7) by $\Delta t\psi_K^n$ and summing for K and n, we have

$$J_1(\delta) + J_2(\delta) + J_3(\delta) = 0$$

where

$$J_1(\delta) = \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left(u_K^{n+1} - u_K^n\right) \psi_K^n,$$

$$J_2(\delta) = -\mu \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t \ \tau_\sigma \ Du_{K,\sigma}^{n+1} \psi_K^n,$$

$$J_3(\delta) = a \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\substack{\sigma\in\mathcal{E}_K \\ \sigma=K|L}} \Delta t \ \tau_\sigma \left(S\left(Dc_{K,\sigma}^{n+1}\right) u_K^{n+1} - S\left(-Dc_{K,\sigma}^{n+1}\right) u_L^{n+1}\right) \psi_K^n.$$

From the weak convergence of $(dc_\delta)_\delta$ to ∇c, $(du_\delta)_\delta$ to ∇u, and the strong convergence of $(u_\delta)_\delta$ to u in $L^2(\Omega_T)$, we have when $\delta \to 0$

$$\lambda(\delta) \to \int_0^T \int_\Omega \left(u \frac{\partial \psi}{\partial t} - \mu \nabla u \cdot \nabla \psi + a\, u \nabla c \cdot \nabla \psi\right) dxdt + \int_\Omega u_0\, \psi(x,0)\, dx.$$

To prove that (u,c) defined in Proposition 5.5.1 verifies (5.5), we need to prove that $J_i(\delta) - J_{i0}(\delta) \to 0$ as $\delta \to 0$ for $i = 1,2,3$. The proof is exactly similar to that of Proposition 2.5.2, we simply have to note that, using the identity (for $\sigma = K|L$) :

$$S\left(Dc_{K,\sigma}^{n+1}\right) u_K^{n+1} - S\left(-Dc_{K,\sigma}^{n+1}\right) u_L^{n+1} = -S\left(-Dc_{K,\sigma}^{n+1}\right) Du_{K,\sigma}^{n+1} + u_K^{n+1} Dc_{K,\sigma}^{n+1}.$$

we can write $J_3(\delta) = J_{31}(\delta) + J_{32}(\delta)$, with

$$J_{31}(\delta) = \frac{a}{2} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t \ \tau_\sigma S\left(-Dc_{K,\sigma}^{n+1}\right) Du_{K,\sigma}^{n+1} D\psi_{K,\sigma}^n,$$

$$J_{32}(\delta) = -\frac{a}{2} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \Delta t \ \tau_\sigma u_K^{n+1} Dc_{K,\sigma}^{n+1} D\psi_{K,\sigma}^n.$$

The limit $\delta \to 0$ in the equation (5.14) is performed as in the proof of Proposition 2.5.2. The treatment of the term $\dfrac{u_K^n}{u_K^n+1}$ is obvious, since it is bounded for all $K \in \mathcal{T}$ and $n = 0, ..., N-1$. For the correction term, using the Cauchy-Schwarz inequality and the fact that $0 < \beta_n \leq 1$, we obtain

$$\begin{aligned}\sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \Delta t \beta_n T_K^n \psi_K^n \leq & \sqrt{\Delta t} \left(\sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \Delta t\, \mathrm{m}(K) \left|\psi_K^n\right|^2\right)^{1/2} \\ & \times \left(\sum_{n=0}^{N-1} \sum_{K\in\mathcal{T}} \mathrm{m}(K) \left|\frac{u_K^n}{u_K^n+1} - \frac{u_K^{n-1}}{u_K^{n-1}+1}\right|^2\right)^{1/2}.\end{aligned}$$

Now, since the function $x \longrightarrow \dfrac{x}{x+1}$ is smooth for all $x \geq 0$, there exists a constant C such that

$$\left|\frac{u_K^n}{u_K^n+1} - \frac{u_K^{n-1}}{u_K^{n-1}+1}\right| \leq C \left|u_K^n - u_K^{n-1}\right|.$$

From the two last inequalities, and from the regularity of ψ and the estimate (5.31), it is easy to see that

$$\sum_{n=0}^{N-1} \sum_{K \in \mathcal{T}} \Delta t \beta_n T_K^n \psi_K^n \to 0 \quad \text{as } \delta \to 0.$$

This completes the proof of Theorem 5.5.1.

5.6 Numerical experiments

In this section, we use the proposed corrected decoupled scheme to solve a number of two-dimensional chemotaxis systems. The obtained numerical results are compared with those of more usual decoupled methods. We mention that the computational time of all compared schemes is almost the same, so the comparison is only performed with respect to accuracy. For the value of ε, reference solutions are computed with $\varepsilon = 0$, and we take $\varepsilon = 10^{-6}$ for the other computed solutions. Finally, as mentioned in Section 5.2, $\beta_n = 1$ in all numerical tests.

Test 1. This test deals with the comparison between the corrected decoupled scheme (5.7),(5.14) and the decoupled scheme (5.7)–(5.8). The spatial domain is $\Omega = (-7/2, 7/2) \times (-35, 35)$ and is discretized via a uniform mesh of 12250 control volumes, whereas the final time is $T_f = 150$. We adopt the following parameters used in [60] : $\mu = 0.25$ and $a = 2$, and we consider the following initial condition :

$$u_0(x) = \begin{cases} 1 + \epsilon(x) & \text{if } x \in (-7/2, 7/2) \times (-1, 1), \\ 1 & \text{otherwise, with } x \in \Omega, \end{cases} \tag{5.34}$$

where $\epsilon(x)$ is a positive perturbation which is constant on each control volume. It is equal on each cell of the mesh to the average of ten uniformly distributed random values in $[0, 1]$.

Since the exact cell density solution u of the studied system (5.1)–(5.2) is unavailable, we compute a reference solution by the proposed corrected decoupled scheme on a very fine time-stepping $\Delta t = 10^{-3}$, so the number of time-steps needed to reach the final time T_f is $150,000$. We then use the obtained reference solution to compute the relative L^2-errors for the two schemes (5.7),(5.14) and (5.7)–(5.8). These errors are presented in Table 5.1 (see also Figure 5.1). From this table, we can see that both schemes are first-order accurate and stable for any time step Δt used in this test. However, the corrected decoupled scheme is about three to five times more accurate than the decoupled scheme (5.7)–(5.8).

The initial condition (5.34) and the reference solution at final time T_f are plotted in Figure 5.2. The presented figure demonstrates the ability of the model (5.1)–(5.2) to generate stripe patterns. Three-dimensional plots of the cell density using both studied schemes at final time with $\Delta t = 10^{-2}$ are presented in Figure 5.3. Finally, in Figure 5.4, we plot the contour along the line $L = \{0\} \times (-35, 35)$ of the reference solution and of the solutions computed using both schemes with $\Delta t = 1$. We can see from this figure that even for a large time-step size, the corrected decoupled scheme is in close agreement with the reference solution, which is not the case of the scheme (5.7)–(5.8).

Δt	L^2-error corrected decoupled	Rate	L^2-error decoupled	Rate
5	1.320×10^{-1}	—	4.042×10^{-1}	—
1	2.923×10^{-2}	0.937	1.435×10^{-1}	0.643
5.10^{-1}	1.703×10^{-2}	0.779	7.767×10^{-2}	0.886
10^{-1}	3.817×10^{-3}	0.929	1.630×10^{-2}	0.970
5.10^{-2}	1.918×10^{-3}	0.993	8.205×10^{-3}	0.990
10^{-2}	3.566×10^{-4}	1.05	1.672×10^{-3}	0.988

Table 5.1 Relative L^2-errors and time convergence orders obtained for (u) using the corrected decoupled scheme (5.7),(5.14) and the decoupled scheme (5.7)–(5.8).

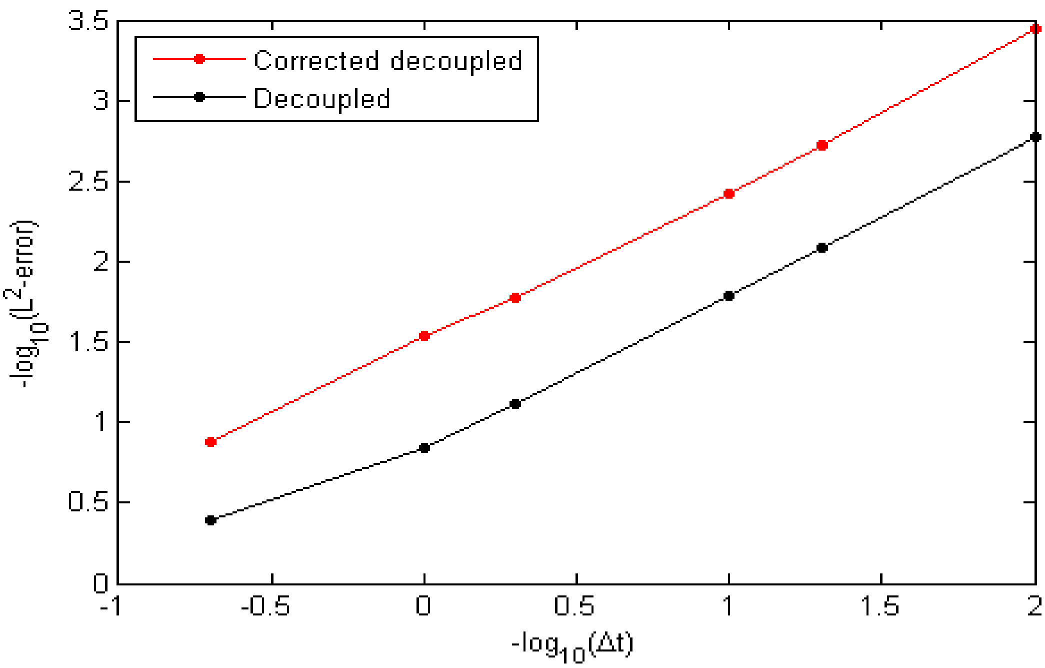

Figure 5.1 Relative L^2-errors and time convergence orders obtained for (u) using the corrected decoupled scheme (5.7),(5.14) and the decoupled scheme (5.7)–(5.8).

Test 2. The purpose of this test is to investigate the efficiency of the corrected decoupled scheme in the case of the parabolic-parabolic version of the model (5.1)–(5.2).

We consider the same data used in the previous test. Nevertheless, we need to define an initial condition for the concentration of chemoattractant, we take then $c(x, 0) = 1/32$. The schemes adopted are similar to those of the previous test, we only need to replace the equation (5.8) in the scheme (5.7)–(5.8) by

$$\mathrm{m}(K)\frac{c_K^{n+1} - c_K^n}{\Delta t} - \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma \, Dc_{K,\sigma}^{n+1} + \mathrm{m}(K)\, c_K^{n+1} = \mathrm{m}(K)\, \frac{u_K^n}{u_K^n + 1}, \tag{5.35}$$

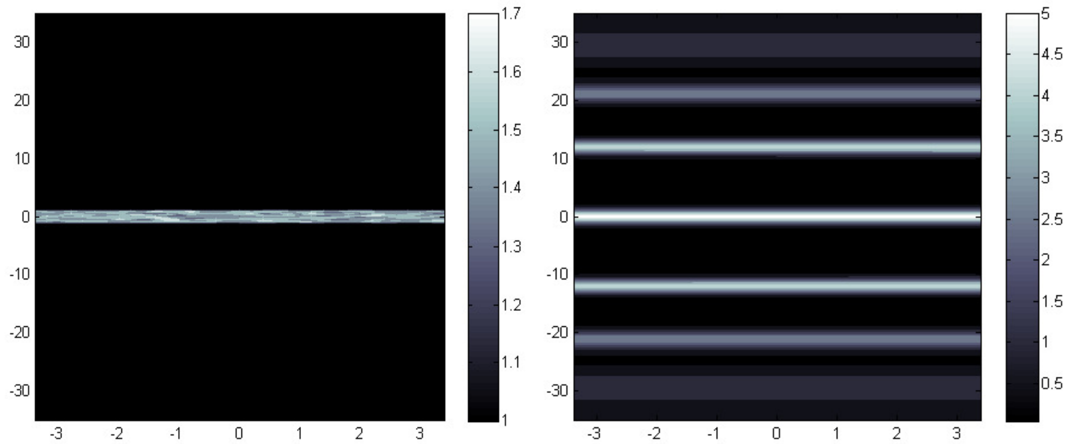

Figure 5.2 Initial cell density (5.34) (left) and cell density (u) at final time T_f computed via the corrected decoupled scheme (5.7),(5.14) with $\Delta t = 10^{-2}$ (right).

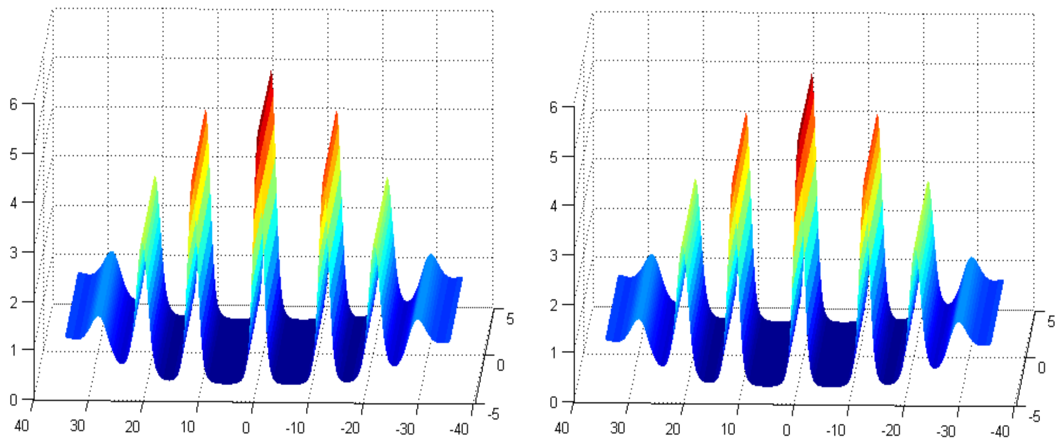

Figure 5.3 Three-dimensional plot of the cell density (u) at final time T_f computed via the corrected decoupled scheme (5.7),(5.14) (left) and the decoupled scheme (5.7)–(5.8) (right) with $\Delta t = 10^{-2}$.

and we substitute the equation (5.14) in the scheme (5.7),(5.14) by (we recall that $\beta_n = 1$)

$$\mathrm{m}(K)\frac{c_K^{n+1} - c_K^n}{\Delta t} - \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\, Dc_{K,\sigma}^{n+1} + \mathrm{m}(K)\, c_K^{n+1} = \mathrm{m}(K)\,\frac{u_K^n}{u_K^n + 1} + T_K^n. \tag{5.36}$$

The following decoupled scheme is also investigated in this test :

$$\mathrm{m}(K)\frac{u_K^{n+1} - u_K^n}{\Delta t} - \mu \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma D u_{K,\sigma}^{n+1}$$

$$+ a \sum_{\substack{\sigma\in\mathcal{E}_K\\ \sigma=K|L}} \tau_\sigma \left(S\left(Dc_{K,\sigma}^n\right) u_K^{n+1} - S\left(-Dc_{K,\sigma}^n\right) u_L^{n+1}\right) = 0, \tag{5.37}$$

$$\mathrm{m}(K)\frac{c_K^{n+1} - c_K^n}{\Delta t} - \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\, Dc_{K,\sigma}^{n+1} + \mathrm{m}(K)\, c_K^{n+1} = \mathrm{m}(K)\,\frac{u_K^{n+1}}{u_K^{n+1} + 1}, \tag{5.38}$$

such scheme is used for example in [84].

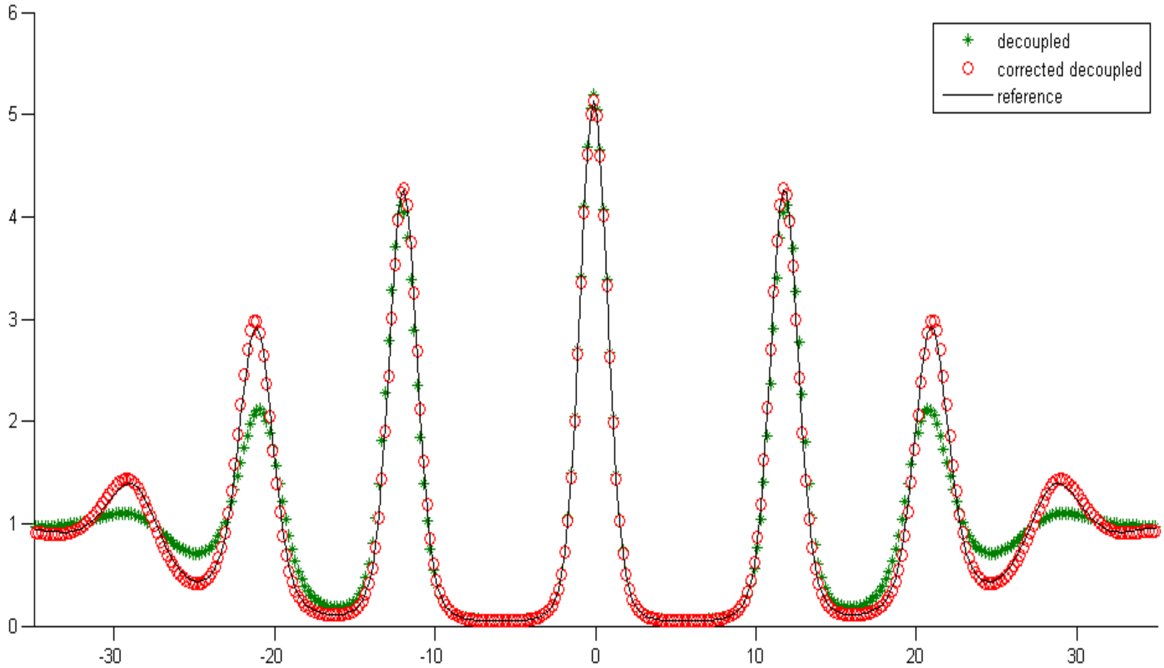

Figure 5.4 Contours along the line L at final time T_f of the reference solution, the solution computed via the corrected decoupled scheme (5.7),(5.14) with $\Delta t = 1$, and the solution computed via the decoupled scheme (5.7)–(5.8) with $\Delta t = 1$.

The reference solution is computed similarly to the previous test. We observe from Table 5.2 (see also Figure 5.5) that the scheme (5.37)–(5.38) is the less accurate one. We can see also that the corrected decoupled scheme (5.7),(5.36) is about four to five times more accurate than the scheme (5.7),(5.35).

Δt	L^2-error corrected decoupled	Rate	L^2-error (5.7),(5.35)	Rate	L^2-error (5.37)–(5.38)	Rate
5	8.450×10^{-2}	—	3.775×10^{-1}	—	4.231×10^{-1}	—
1	2.323×10^{-2}	0.802	1.117×10^{-1}	0.756	1.243×10^{-1}	0.761
5.10^{-1}	1.344×10^{-2}	0.790	6.109×10^{-2}	0.871	6.846×10^{-2}	0.861
10^{-1}	2.971×10^{-3}	0.938	1.314×10^{-2}	0.955	1.477×10^{-2}	0.953
5.10^{-2}	1.490×10^{-3}	0.996	6.634×10^{-3}	0.986	7.456×10^{-3}	0.986
10^{-2}	2.765×10^{-4}	1.05	1.354×10^{-3}	0.987	1.519×10^{-3}	0.988

Table 5.2 Relative L^2-errors and time convergence orders obtained for (u) using the corrected decoupled scheme (5.7),(5.36), the decoupled scheme (5.7),(5.35), and the decoupled scheme (5.37)–(5.38).

In Figure 5.6, we show the evolution of the cell density by plotting the contour of the reference solution along the line $L = \{0\} \times (-35, 35)$ at different times.

Test 3. The purpose of this test is to study the accuracy of our corrected decoupled scheme when the source term in the equation for concentration c is linear. To this end, we consider the

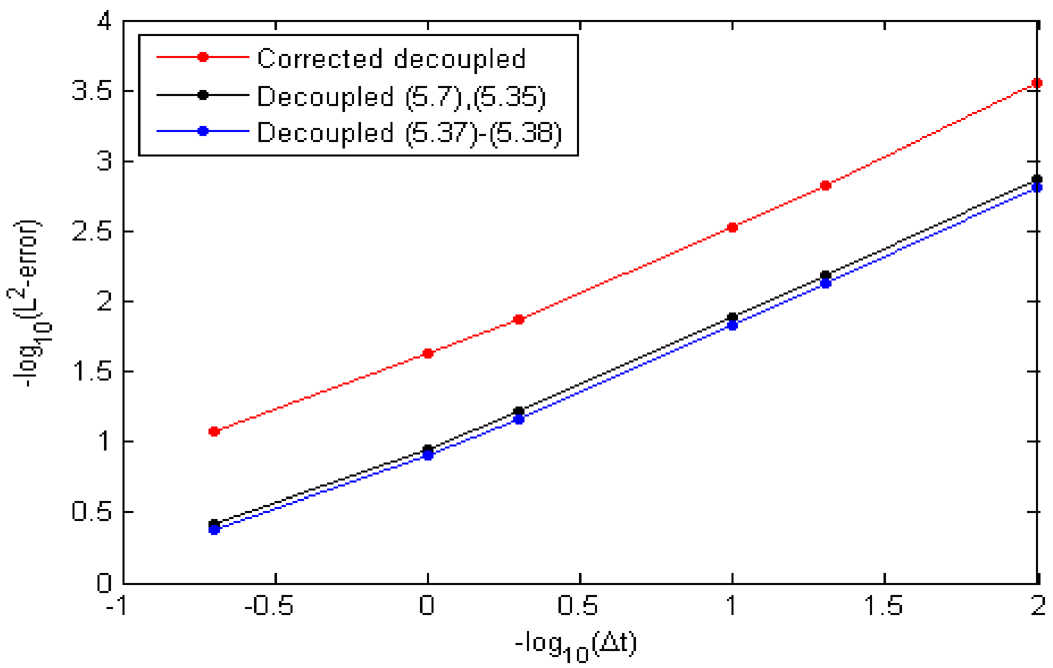

Figure 5.5 Relative L^2-errors and time convergence orders obtained for (u) using the corrected decoupled scheme (5.7),(5.36), the decoupled scheme (5.7),(5.35), and the decoupled scheme (5.37)–(5.38).

following chemotaxis-growth model for bacterial pattern formation studied in Section 4.6.2 :

$$\begin{cases} \partial_t u = \mu \Delta u - \chi \nabla \cdot (u \nabla c) + f(u) & \text{in } \Omega \times (0, T_f), \\ \partial_t c = \Delta c - \gamma c + u & \text{in } \Omega \times (0, T_f), \end{cases} \tag{5.39}$$

endowed with the homogeneous Neumann boundary conditions. We consider the following data : $\mu = 0.0625$, $\chi = 6$, $\gamma = 16$, $f(u) = 2u(1-u)$ and $T_f = 30$. The computational domain is the square $\Omega = (-8, 8)^2$ discretized with a uniform mesh grid 100×100. For all $x \in \Omega$, we choose the following initial conditions :

$$u(x, 0) = \begin{cases} 1 + \epsilon(x) & \text{if } \|x\|_2 < 0.7, \\ 1 & \text{otherwise,} \end{cases}$$

and $c(x, 0) = 1/32$. The random perturbation $\epsilon(x)$ is defined as in (5.34).

In this test, we compare the numerical results obtained from the decoupled scheme :

$$\begin{aligned} &\mathrm{m}(K) \frac{u_K^{n+1} - u_K^n}{\Delta t} - \mu \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma D u_{K,\sigma}^{n+1} \\ &+ \chi \sum_{\substack{\sigma \in \mathcal{E}_K \\ \sigma = K|L}} \tau_\sigma \left(S\left(D c_{K,\sigma}^{n+1}\right) u_K^{n+1} - S\left(-D c_{K,\sigma}^{n+1}\right) u_L^{n+1} \right) - 2\mathrm{m}(K)\, u_K^n \left(1 - u_K^{n+1}\right) = 0, \end{aligned} \tag{5.40}$$

$$\mathrm{m}(K) \frac{c_K^{n+1} - c_K^n}{\Delta t} - \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma\, D c_{K,\sigma}^{n+1} + \gamma \mathrm{m}(K)\, c_K^{n+1} = \mathrm{m}(K)\, u_K^n, \tag{5.41}$$

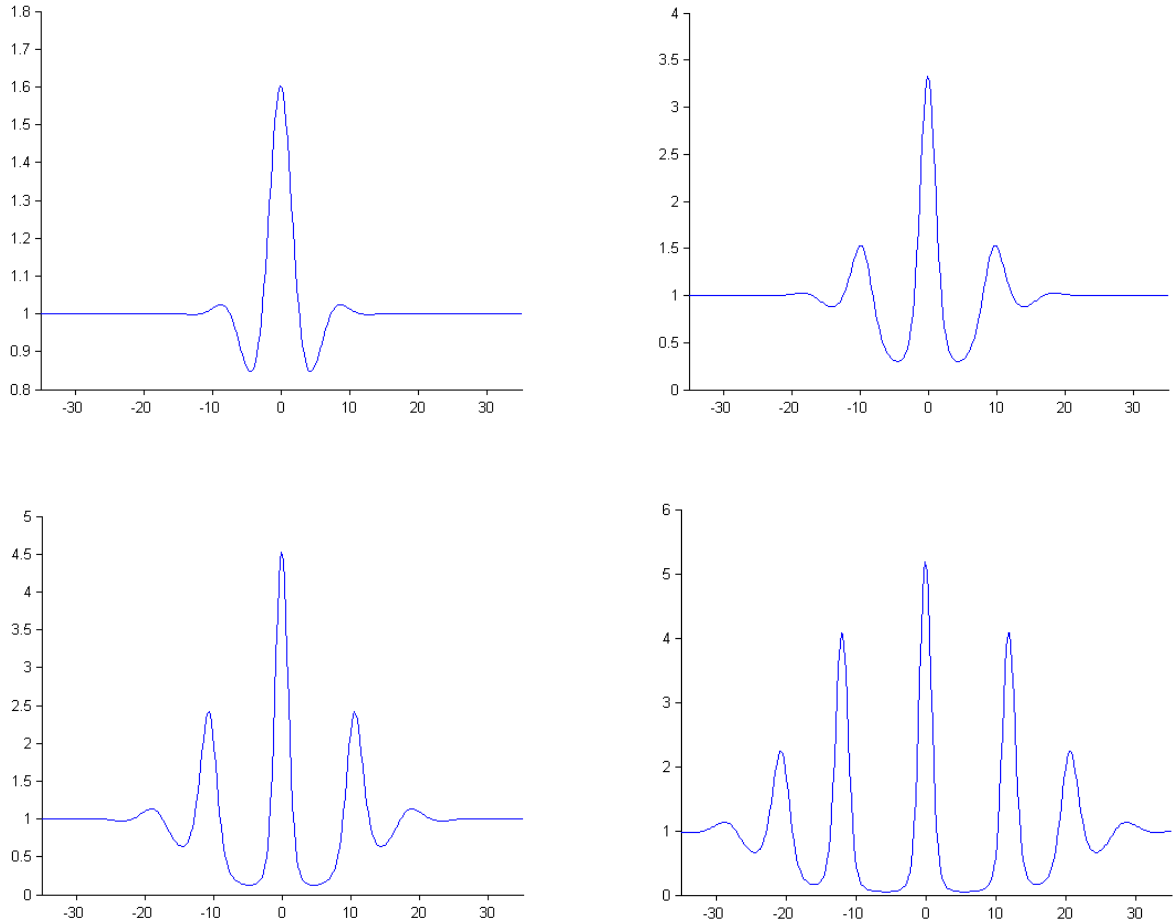

Figure 5.6 Contours along the line L of the reference solution (u) at $t = 25$ (left top), $t = 75$ (right top), $t = 100$ (left bottom), T_f (right bottom).

with those of the corrected decoupled scheme, consisting of (5.40) and the following equation

$$\mathrm{m}(K)\frac{c_K^{n+1} - c_K^n}{\Delta t} - \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma \, Dc_{K,\sigma}^{n+1} + \gamma \mathrm{m}(K)\, c_K^{n+1} = \mathrm{m}(K)\, u_K^n + T_K^n. \tag{5.42}$$

where for all $K \in \mathcal{T}$ and $n = 1, ..., N-1$

$$T_K^n = \mathrm{m}(K)\left(u_K^n - u_K^{n-1}\right), \quad T_K^0 = 0.$$

The reference solution is computed by the corrected decoupled scheme using a very fine time-step size $\Delta t = 10^{-4}$. The results presented in Table 5.3 (see also Figure 5.7) show that the corrected decoupled scheme is highly accurate compared to the scheme (5.40)–(5.41). In the case when $\Delta t = 10^{-3}$, the corrected decoupled scheme is about 32 times more accurate than the scheme (5.40)–(5.41).

The numerical cell density u of the model computed using both schemes with $\Delta t = 10^{-3}$ is shown in Figure 5.8. As we can see, the solution forms periodic arrays of continuous rings which match well with the patterns formed by *Salmonella typhimurium* [94] (see Figure 1.2).

Δt	L^2-error corrected decoupled	Rate	L^2-error decoupled	Rate
5.10^{-1}	4.216×10^{-3}	—	2.234×10^{-2}	—
10^{-3}	6.022×10^{-4}	1.21	1.119×10^{-2}	0.430
5.10^{-2}	2.947×10^{-4}	1.03	6.806×10^{-3}	0.717
10^{-2}	5.863×10^{-5}	1.00	1.636×10^{-3}	0.886
5.10^{-3}	2.907×10^{-5}	1.01	8.385×10^{-4}	0.964
10^{-3}	5.347×10^{-6}	1.05	1.707×10^{-4}	0.989

Table 5.3 Relative L^2-errors and time convergence orders obtained for (u) using the corrected decoupled scheme (5.40),(5.42) and the decoupled scheme (5.40)–(5.41).

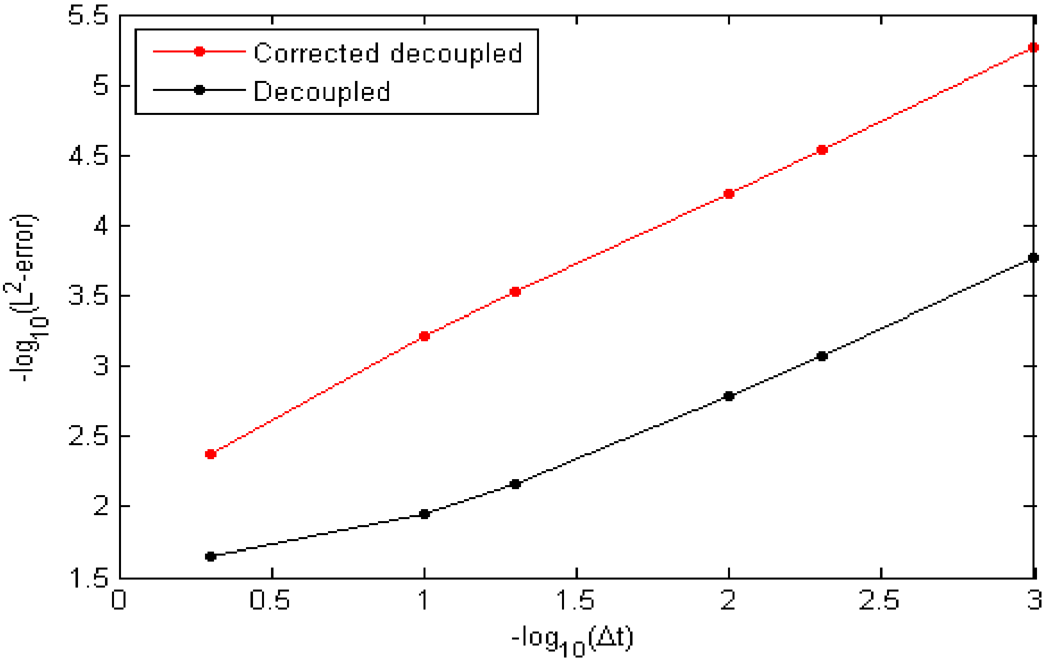

Figure 5.7 Relative L^2-errors and time convergence orders obtained for (u) using the corrected decoupled scheme (5.40),(5.42) and the decoupled scheme (5.40)–(5.41).

Test 4. In this test, we present some numerical simulations which illustrate the ability of the presented corrected decoupled finite volume scheme to capture different forms of bacterial spatial patterns. For this purpose, we consider the chemotaxis model (5.39) with the following data used in [1] : $\mu = 0.0625$, $\gamma = 32$ and $f(u) = u^2(1-u)$. The domain is the square $\Omega = (-10, 10)^2$, which is discretized via a uniform mesh grid 150×150, and the time-step used is $\Delta t = 10^{-1}$. For the final time, we take $T_f = 150$ and we consider the following initial conditions

$$u(x,0) = \begin{cases} 1+\epsilon(x) & \text{if } \|x\|_2 < 1, \\ 1 & \text{otherwise,} \end{cases}$$

and $c(x,0) = 1/32$.

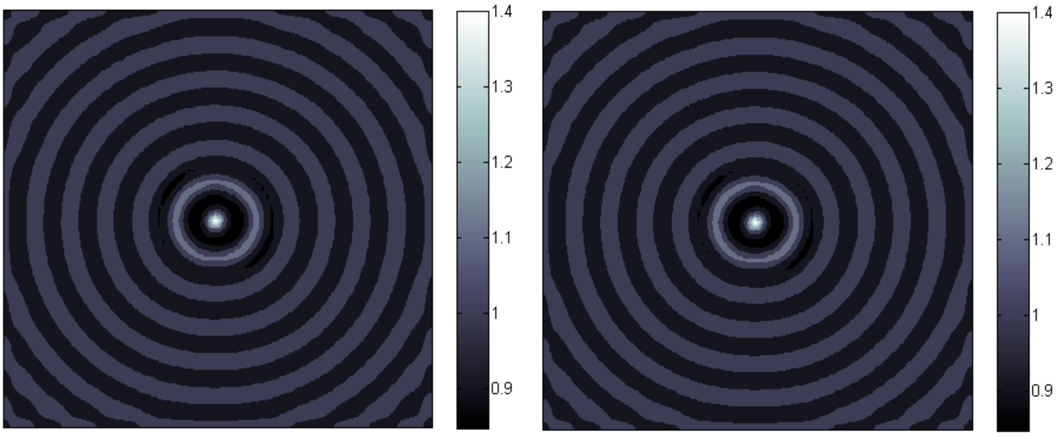

Figure 5.8 Cell density (u) at final time T_f computed via the corrected decoupled scheme (5.40),(5.42) (left) and the decoupled scheme (5.40)–(5.41) (right) with $\Delta t = 10^{-3}$.

The corrected decoupled scheme used is similar to that of the previous section. However, since the logistic source $f(u)$ has now a cubic form, we use the following linearized finite volume discretization :

$$\int_K u(x,t)^2 \,(1 - u(x,t))\; dx \approx \mathrm{m}(K)\, u_K^{n+1} u_K^n \,(1 - u_K^n)\,.$$

For $\chi = 80$, the computed solution at $t = 30$ is shown in Figure 5.9. We observe from Figure 5.9 (left) the formation of symmetrical spots in whole domain. The 3D view of this patterning is presented in Figure 5.9 (right).

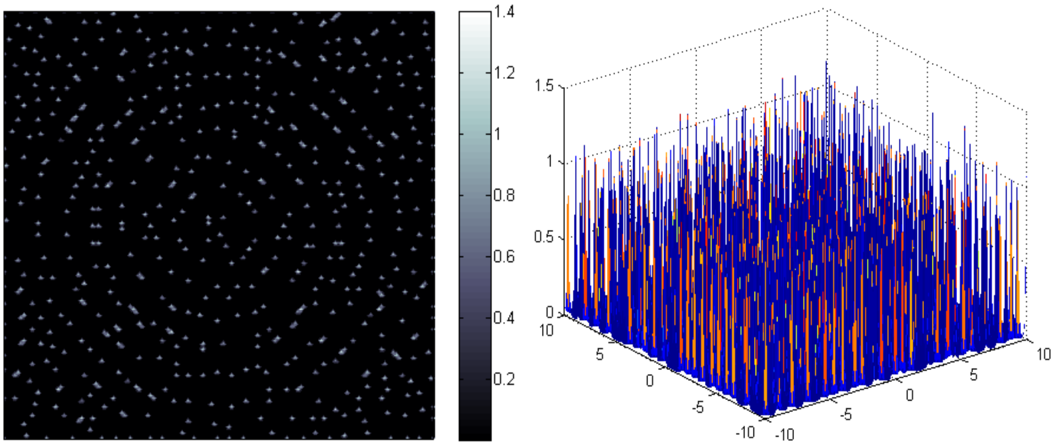

Figure 5.9 Cell density (u) for $\chi = 80$ at $t = 30$ computed via the corrected decoupled.

In Figure 5.10, we examine the effect of the parameter χ on the numerical solution. When $\chi = 6$, a honeycomb pattern is observed, which is similar to that of *Vibrio* bacteria (see Figure 1.1). Then, the solution changes its structure to continuous rings for $\chi = 7.4$. When $\chi = 20$, chaotic spots appear. The symmetry of these spots increases for high values of χ (see Figure 5.10 (right bottom)). These symmetric spots seem in good agreement with the *Escherichia coli* patterns reported in [12] (see Figure 1.3).

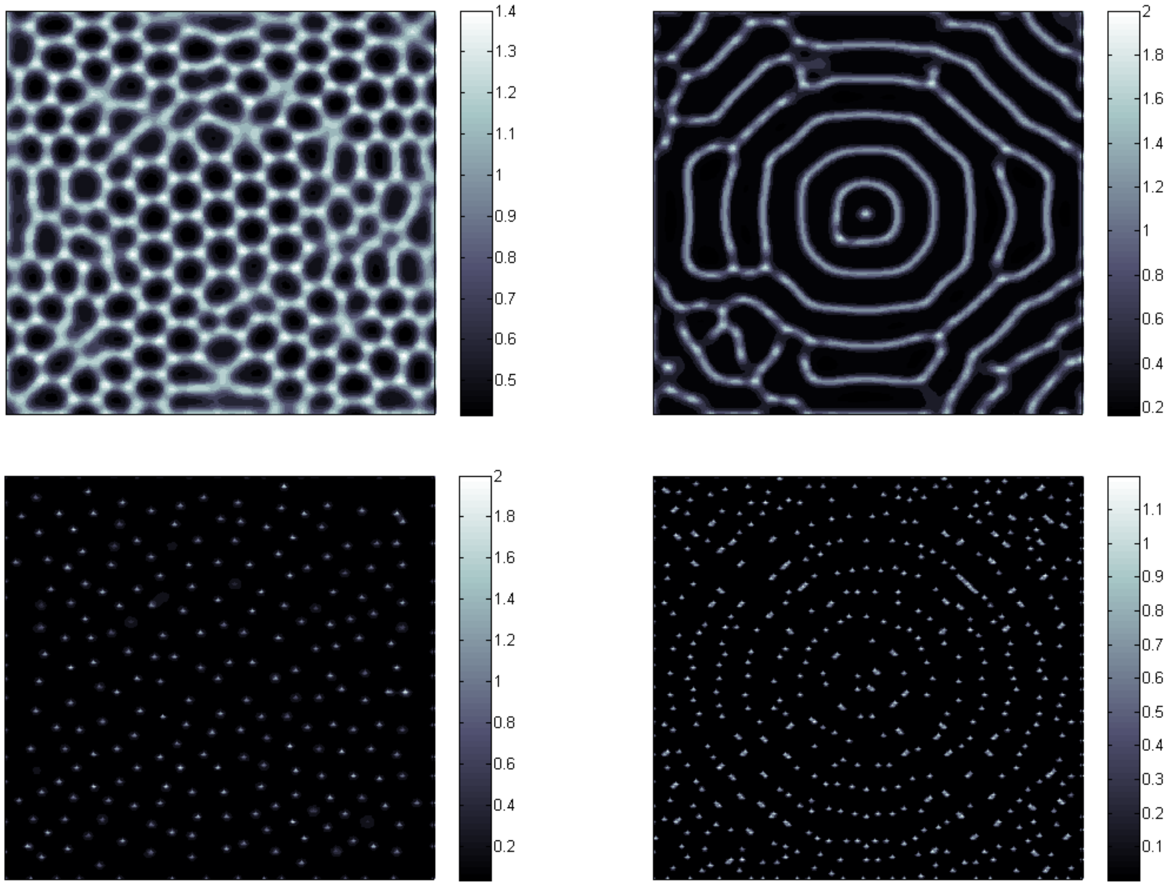

Figure 5.10 Cell density (u) computed via the corrected decoupled scheme at final time T_f, for $\chi = 6$ (left top), $\chi = 7.4$ (right top), $\chi = 20$ (left bottom), $\chi = 70$ (right bottom).

Conclusion et perspectives

Dans cette thèse, nous avons proposé et analysé différents schémas numériques pour des modèles de chimiotaxie. Les schémas qui ont été présentés sont de type volumes finis en espace utilisant différentes approximations : upwind, exponentielle et hybride (upwind/centré), afin de maintenir la positivité de la solution numérique et d'assurer sa stabilité. En ce qui concerne la discrétisation temporelle, nous nous sommes intéressés à des approximations implicites linéarisés qui permettent d'obtenir des schémas découplés. Ce type de schémas est sans doute moins précis par rapport aux schémas couplés, mais reste très avantageux au niveau du coût de calcul. La facilité d'implémentation de tels schémas est aussi un atout qui explique leur large utilisation.

Dans les chapitres 2 et 3, nous nous sommes penchés sur l'analyse de convergence de schémas implicites linéarisés usuels, utilisant une discrétisation volumes finis en espace. Dans les deux derniers chapitres, nous avons développé deux nouveaux schémas : un schéma découplé semi-exponentiel en temps dans le chapitre 4 et un schéma découplé corrigé dans le chapitre 5. Les résultats numériques obtenus à l'aide de ces deux schémas montrent qu'ils sont plus précis et plus performants que les schémas implicites linéarisés classiques. Enfin, en utilisant les schémas proposés dans cette thèse, nous avons étudié l'habileté d'un nombre de modèles de chimiotaxie issus de l'embryologie et de la bactériologie à générer différentes formes de motifs biologiques.

Nous concluons cette thèse en présentant plusieurs points qui peuvent être encore abordés ainsi que les pistes de recherche qui se dégagent des travaux effectués. Nous commençons par signaler que l'analyse de convergence des schémas volumes finis pour les modèles de chimiotaxie étudiés dans [28] et [7], ainsi que pour ceux présentés dans cette thèse, n'a été effectuée que pour les systèmes paraboliques-elliptiques. Dans la littérature, les schémas volumes finis pour les systèmes paraboliques-paraboliques n'ont été analysés que dans le cas des modèles avec effet de remplissage de volume (voir par exemple [5]). Il serait aussi intéressant de généraliser le résultat de convergence obtenu pour le modèle de chimiotaxie-croissance du chapitre 2 au cas d'une source logistique quadratique. Pour se faire, il faudrait essayer d'obtenir l'analogue discret des résultats qui ont été prouvés pour le cas continu dans [86]. D'autre part, nous pourrions étendre les résultats obtenus dans les chapitres 3, 4 et 5 pour des conditions aux limites plus générales, comme celles considérées dans [18]. En ce qui concerne le schéma semi-exponentiel en temps proposé dans le chapitre 4, nous pouvons envisager de développer un schéma complètement exponentiel en temps. On pourrait pour cela s'inspirer de [87] où un schéma complètement exponentiel en espace a été développé en incorporant le terme de dérivée temporelle dans la discrétisation exponentielle spatiale. Enfin, nous pensons qu'il serait très intéressant de généraliser l'idée du schéma découplé corrigé proposé dans le chapitre 5 à d'autres systèmes d'équations aux dérivées partielles tels que les équations de Navier-Stokes, les modèles d'écoulement en milieu poreux, le système de dérive-diffusion pour les semi-conducteurs, etc...

Bibliographie

[1] M. Aida, T. Tsujikawa, M. Efendiev, A. Yagi, and M. Mimura. Lower estimate of the attractor dimension for a chemotaxis growth system. *Journal of the London Mathematical Society*, 74(2) :453–474, 2006.

[2] M. Akhmouch and M. Benzakour Amine. Semi-implicit finite volume schemes for a chemotaxis-growth model. *Indagationes Mathematicae*, 27(3) :702–720, 2016.

[3] M. Akhmouch and M. Benzakour Amine. A time semi-exponentially fitted scheme for chemotaxis-growth models. *Calcolo*, doi :10.1007/s10092-016-0201-4, 2016.

[4] M. Akhmouch and M. Benzakour Amine. A corrected decoupled scheme for chemotaxis models. *Journal of Computational and Applied Mathemetics*, http ://dx.doi.org/10.1016/j.cam.2017.04.001, 2017.

[5] B. Andreianov, M. Bendahmane, and M. Saad. Finite volume methods for degenerate chemotaxis model. *Journal of Computational and Applied Mathematics*, 235 (14) :4015–4031, 2011.

[6] M. Bessemoulin-Chatard. A finite volume scheme for convection–diffusion equations with nonlinear diffusion derived from the scharfetter–gummel scheme. *Numerische Mathematik*, 121(4) :637–670, 2012.

[7] M. Bessemoulin-Chatard and A. Jüngel. A finite volume scheme for a Keller-Segel model with additional cross-diffusion. *IMA journal of numerical analysis*, 34 (1) :96–122, 2014.

[8] M. Bessemoulin-Chatard, C. Chainais-Hillairet, and F. Filbet. On discrete functional inequalities for some finite volume schemes. *IMA journal of numerical analysis*, 35 (3) :doi : 10.1093/imanum/dru032, 2014.

[9] A. Blanchet, V. Calvez, and J. A. Carrillo. Convergence of the mass-transport steepest descent scheme for the subcritical patlak-keller-segel model. *SIAM Journal on Numerical Analysis*, 46(2) :691–721, 2008.

[10] C. J. Budd, R. Carretero-González, and R. D. Russell. Precise computations of chemotactic collapse using moving mesh methods. *Journal of Computational Physics*, 202(2) :463–487, 2005.

[11] E. O. Budrene and H. C. Berg. Complex patterns formed by motile cells of escherichia coli. *Nature*, 349 :630–633, 1991.

[12] E. O. Budrene and H. C. Berg. Dynamics of formation of symmetrical patterns of chemotactic bacteria. *Nature*, 376 :49–53, 1995.

[13] C. Chainais-Hillairet, J.-G. Liu, and Y.-J. Peng. Finite volume scheme for multidimensional drift-diffusion equations and convergence analysis. *Mathematical Modelling and Numerical Analysis*, 37 :319–338, 2003.

[14] C. Chainais-Hillairet and Y.-J. Peng. Finite volume approximation for degenerate drift-diffusion system in several space dimensions. *Mathematical Models and Methods in Applied Sciences*, 14(03) :461–481, 2004.

[15] G. Chamoun, M. Saad, and R. Talhouk . Monotone combined edge finite volume-finite element scheme for anisotropic keller-segel model. *Numererical Methods For Partial Differential Equation*, 30(3) :1030–1065, 2014.

[16] M. A. J. Chaplain and G. Lolas. Mathematical modelling of cancer invasion of tissue : dynamic heterogeneity. *Networks and Heterogeneous Media*, 1(3) :399–439, 2006.

[17] A. Chertock and A. Kurganov. A second-order positivity preserving central-upwind scheme for chemotaxis and haptotaxis models. *Numerische Mathematik*, 111(2) :169–205, 2008.

[18] M. Delgado, C. Morales-Rodrigo, A. Suárez, and J. I. Tello. On a parabolic–elliptic chemotactic model with coupled boundary conditions. *Nonlinear Analysis : Real World Applications*, 11(5) :3884–3902, 2010.

[19] M. Di Francesco and J. Rosado. Fully parabolic keller–segel model for chemotaxis with prevention of overcrowding. *Nonlinearity*, 21(11) :2715, 2008.

[20] R. F. Diegelmann and M. C. Evans. Wound healing : an overview of acute, fibrotic and delayed healing. *Frontiers in Bioscience*, 9(1) :283–289, 2004.

[21] D. Dormann and C. J. Weijer. Chemotactic cell movement during dictyostelium development and gastrulation. *Current opinion in genetics & development*, 16(4) :367–373, 2006.

[22] Y. Epshteyn. Upwind-Difference Potentials Method for Patlak-Keller-Segel Chemotaxis Model. *Journal of Scientific Computing*, 53 :689–713, 2012.

[23] Y. Epshteyn and A. Kurganov. New Interior Penalty Discontinuous Galerkin Methods for the Keller-Segel Chemotaxis Model. *SIAM Journal on Numerical Analysis*, 47(1) :386–408, 2008.

[24] Y. Epshteyn and A. Izmirlioglu. Fully discrete analysis of a discontinuous finite element method for the keller-segel chemotaxis model. *Journal of Scientific Computing*, 40(1-3) :211–256, 2009.

[25] R. Eymard, J. Fuhrmann, and K. Gärtner. A finite volume scheme for nonlinear parabolic equations derived from one-dimensional local dirichlet problems. *Numerische Mathematik*, 102(3) :463–495, 2006.

[26] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. *In Handbook of numerical analysis*, volume VII :713–1020, 2000.

[27] P. C. Fife. *Mathematical aspects of reacting and diffusing systems*, volume 28. Springer Science & Business Media, 2013.

[28] F. Filbet. A finite volume scheme for the Patlak-Keller-Segel chemotaxis model. *Numerische Mathematik*, 104(4) :457–488, 2006.

[29] H. Gajewski, K. Zacharias, and K. Gröger. Global behaviour of a reaction-diffusion system modelling chemotaxis. *Mathematische Nachrichten*, 195(1) :77–114, 1998.

[30] T. Gallouët and J.-C. Latché. Compactness of discrete approximate solutions to parabolic pdes-application to a turbulence model. *Communications on Pure and Applied Analysis*, 11(6) :21, 2012.

[31] H. Gao. Optimal error analysis of galerkin fems for nonlinear joule heating equations. *Journal of Scientific Computing*, 58(3) :627–647, 2014.

[32] N. C. Grassly and C. Fraser. Mathematical models of infectious disease transmission. *Nature Reviews Microbiology*, 6(6) :477–487, 2008.

[33] J. Haškovec and C. Schmeiser. Stochastic particle approximation for measure valued solutions of the 2d keller-segel system. *Journal of Statistical Physics*, 135(1) :133–151, 2009.

[34] J. Haškovec and C. Schmeiser. Convergence of a stochastic particle approximation for measure solutions of the 2d keller-segel system. *Communications in Partial Differential Equations*, 36(6) :940–960, 2011.

[35] E. Hildebrand and U. Kaupp. Sperm chemotaxis : a primer. *Annals of the New York Academy of Sciences*, 1061(1) :221–225, 2005.

[36] S. Hittmeir and A. Jüngel. Cross diffusion preventing blow-up in the two-dimensional keller-segel model. *SIAM Journal on Mathematical Analysis*, 43(2) :997–1022, 2011.

[37] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4) :500–544, 1952.

[38] D. Horstmann. From 1970 until now : The Keller-Segel model in chemotaxis and its consequences I. *Jahresberichte der DMV*, 105 :103–165, 2003.

[39] D. Horstmann. From 1970 until now : The Keller-Segel model in chemotaxis and its consequences II. *Jahresberichte der DMV*, 106 :51–69, 2004.

[40] D. Horstmann and G. Wang. Blow-up in a chemotaxis model without symmetry assumptions. *European Journal of Applied Mathematics*, 12(02) :159–177, 2001.

[41] P. Hunter. Biology is the new physics. *EMBO reports*, 11(5) :350–352, 2010.

[42] M. Ibrahim and M. Saad. On the efficacy of a control volume finite element method for the capture of patterns for a volume-filling chemotaxis model. *Computers & Mathematics with Applications*, 68 :1032—-1051, 2014.

[43] A. Il'in. Differencing scheme for a differential equation with a small parameter affecting the highest derivative. *Matematicheskie Zametki*, 6 :237–248, 1969.

[44] E. Keller and L.A. Segel. Initiation of slime mold aggregation viewed as an instability. *Journal of Theoretical Biology*, 26 :399–415, 1970.

[45] E. Keller and L.A. Segel. Traveling bands of chemotactic bacteria : a theoretical analysis. *Journal of Theoretical Biology*, 30(2) :235–248, 1971.

[46] R. Kowalczyk and Z. Szymańska. On the global existence of solutions to an aggregation model. *Journal of Mathematical Analysis and Applications*, 343(1) :379–398, 2008.

[47] D. Lauffenburger, R. Aris, and K. Keller. Effects of random motility on growth of bacterial populations. *Microbial Ecology*, 7(3) :207–227, 1981.

[48] R. D. Lazarov, I. D. Mishev, and P. S. Vassilevski. Finite volume methods for convection-diffusion problems. *SIAM Journal on Numerical Analysis*, 33(1) :31–55, 1996.

[49] B. Li, H. Gao, and W. Sun. Unconditionally optimal error estimates of a crank–nicolson galerkin method for the nonlinear thermistor equations. *SIAM Journal on Numerical Analysis*, 52(2) :933–954, 2014.

[50] B. Li and W. Sun. Error analysis of linearized semi-implicit galerkin finite element methods for nonlinear parabolic equations. *International Journal of Numerical Analysis and Modeling*, 10(3) :622–633, 2013.

[51] Y. Lipman and I. Daubechies. Conformal wasserstein distances : Comparing surfaces in polynomial time. *Advances in Mathematics*, 227(3) :1047–1077, 2011.

[52] P. K. Maini, M. R. Myerscough, K. H. Winters, and J. D. Murray. Bifurcating spatially heterogeneous solutions in a chemotaxis model for biological pattern formation. *Bulletin of mathematical biology*, 53 :701–719, 1991.

[53] A. Marrocco. Numerical simulation of chemotactic bacteria aggregation via mixed finite elements. *ESAIM : Mathematical Modelling and Numerical Analysis*, 37(4) :617–630, 2003.

[54] H. Meinhardt. *Models of biological pattern formation*, volume 6. Academic Press London, 1982.

[55] M. Mimura and T. Tsujikawa. Aggregation pattern dynamics in a chemotaxis model including growth. *Physica A*, 230 :499–543, 1996.

[56] J. D. Murray. On pattern formation mechanisms for lepidopteran wing patterns and mammalian coat markings. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 295(1078) :473–496, 1981.

[57] J. D. Murray. *Mathematical Biology, third edition*. Springer, New York, 2002.

[58] J. D. Murray, D. C. Deeming, and M. W. J. Ferguson. Size-dependent pigmentation-pattern formation in embryos of alligator mississippiensis : time of initiation of pattern generation mechanism. *Proceedings of the Royal Society of London B : Biological Sciences*, 239(1296) :279–293, 1990.

[59] J. D. Murray and M. R. Myerscough. Pigmentation pattern formation on snakes. *Journal of theoretical biology*, 149 :339–360, 1991.

[60] M. R. Myerscough and J. D. Murray. Analysis of propagating pattern in a chemotaxis system. *Bulletin of mathematical biology*, 54(1) :77–94, 1992.

[61] T. Nagai. Blow-up of radially symmetric solutions to a chemotaxis system. *Advances in Mathematical Sciences and Applications*, 5 :581–601, 1995.

[62] T. Nagai. Behavior of solutions to a parabolic-elliptic system modelling chemotaxis. *Journal of the Korean Mathematical Society*, 37(5) :721–733, 2000.

[63] T. Nagai. Blowup of nonradial solutions to parabolic-elliptic systems modeling chemotaxis in two-timensional tomains. *Journal of inequalities and applications*, 6 :37–55, 2001.

[64] T. Nagai, T. Senba, and K. Yoshida. Application of the trudinger-moser inequality to a parabolic system of chemotaxis. *Funkcialaj Ekvacioj*, 40 :411–433, 1997.

[65] J. Nelder. The fitting of a generalization of the logistic curve. *Biometrics*, 17 :89–110, 1961.

[66] K. Osaki and A. Yagi. Finite dimensional attractor for one-dimensional keller-segel equations. *Funkcialaj Ekvacioj*, 44(3) :441–470, 2001.

[67] G. F. Oster and J. D. Murray. Pattern formation models and developmental constraints. *Journal of Experimental Zoology*, 251(2) :186–202, 1989.

[68] L. Pachter and B. Sturmfels. *Algebraic statistics for computational biology*, volume 13. Cambridge university press, 2005.

[69] K. J. Painter and T. Hillen. A user's guide to pde models for chemotaxis. *Journal of mathematical biology*, 58(1-2) :183–217, 2009.

[70] K. J. Painter and T. Hillen. Spatio-temporal chaos in a chemotaxis model. *Physica D*, 240 :363–375, 2011.

[71] S. V. Patankar. *Numerical heat transfer and fluid flow.* Hemisphere Publishing Corporation, Taylor and Francis Group, New York, 1990.

[72] C. Patlak. Random walk with persistence and external bias. *The bulletin of mathematical biophysics*, 15 :311–338, 1953.

[73] M. C. Reed. Mathematical biology is good for mathematics. *Notices of the AMS*, 62(10), 2015.

[74] E. T. Roussos, J. S. Condeelis, and A. Patsialou. Chemotaxis in cancer. *Nature Reviews Cancer*, 11(8) :573–587, 2011.

[75] N. Saito. Conservative upwind finite-element method for a simplified Keller-Segel system modelling chemotaxis. *IMA journal of numerical analysis*, 27 :332–365, 2007.

[76] N. Saito. Error analysis of a conservative finite-element approximation for the Keller-Segel system of chemotaxis. *Communications on Pure and Applied Analysis*, 11(1) :339–364, 2012.

[77] N. Saito and T. Suzuki. Notes on finite difference schemes to a parabolic-elliptic system modelling chemotaxis. *Applied mathematics and computation*, 171 :72–90, 2005.

[78] D. L. Scharfetter and H. K. Gummel. Large-signal analysis of a silicon read diode oscillator. *IEEE Transactions on electron devices*, 16(1) :64–77, 1969.

[79] H. P. Schwan and R. FitzHugh. *Biological engineering*, volume 9. McGraw-Hill Companies, 1969.

[80] T. Senba and T. Suzuki. Parabolic system of chemotaxis : blowup in a finite and the infinite time. *Methods and applications of analysis*, 8(2) :349–367, 2001.

[81] D. B. Spalding. A novel finite difference formulation for differential expressions involving both first and second derivatives. *International Journal for Numerical Methods in Engineering*, 4 :551–559, 1972.

[82] C. L. Stokes and D. A. Lauffenburger. Analysis of the roles of microvessel endothelial cell random motility and chemotaxis in angiogenesis. *Journal of theoretical biology*, 152(3) :377–403, 1991.

[83] R. Strehl, A. Sokolov, D. Kuzmin, D. Horstmann, and S. Turek. A positivity-preserving finite element method for chemotaxis problems in 3D. *Journal of Computational and Applied Mathematics*, 239 :290–303, 2013.

[84] R. Strehl, A. Sokolov, D. Kuzmin, and S. Turek. A flux-corrected finite element method for chemotaxis problems. *Computational Methods in Applied Mathematics*, 10(2) :118–131, 2010.

[85] J. Tello and M. Winkler. Stabilization in a two-species chemotaxis system with a logistic source. *Nonlinearity*, 25(5) :1413, 2012.

[86] J. I. Tello and M. Winkler. A chemotaxis system with logistic source. *Communications in Partial Differential Equations*, 32(6) :849–877, 2007.

[87] J. H. M. ten Thije Boonkkamp and M. J. H. Anthonissen. The finite volume-complete flux scheme for advection-diffusion-reaction equations. *Journal of Scientific Computing*, 46(1) :47–70, 2011.

[88] R. Thar and M. Kühl. Complex pattern formation of marine gradient bacteria explained by a simple computer model. *FEMS microbiology letters*, 246(1) :75–79, 2005.

[89] A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London B : Biological Sciences*, 237(641) :37–72, 1952.

[90] R. Tyson, L. G. Stern, and R. J. LeVeque. Fractional step methods applied to a chemotaxis model. *Journal of mathematical biology*, 41(5) :455–475, 2000.

[91] E. P. Wigner. The unreasonable effectiveness of mathematics in the natural sciences. *Communications on pure and applied mathematics*, 13(1) :1–14, 1960.

[92] M. Winkler. Aggregation vs. global diffusive behavior in the higher-dimensional keller–segel model. *Journal of Differential Equations*, 248(12) :2889–2905, 2010.

[93] M. Winkler. Boundedness in the higher-dimensional parabolic-parabolic chemotaxis system with logistic source. *Communications in Partial Differential Equations*, 35(8) :1516–1537, 2010.

[94] D. Woodward, R. Tyson, M. Myerscough, J. Murray, E. Budrene, and H. Berg. Spatio-temporal patterns generated by salmonella typhimurium. *Biophysical journal*, 68(5) :2181–2189, 1995.

[95] X. Yang, B. Shi, and Z. Chai. Coupled lattice boltzmann method for generalized keller-segel chemotaxis model. *Computers & Mathematics with Applications*, 68(12) :1653–1670, 2014.