Schémas d'ordre élevé pour des simulations réalistes en électrophysiologie cardiaque

Charlie Douanla Lontsi

université de BORDEAUX

THÈSE

PRÉSENTÉE À

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET D'INFORMATIQUE

par **Douanla Lontsi Charlie**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES ET CALCUL SCIENTIFIQUE

Schémas d'ordre élevé pour des simulations réalistes en électrophysiologie cardiaque

---

Rapporteurs : Stéphanie Salmon (Pr, Université de Reims Champagne-Ardenne),
Omar Lakkis (Pr, Free University of Bozen/Bolzano).

Date de soutenance : 15 Novembre 2017

Devant la commission d'examen composée de :

Florence HUBERT .	Pr, Institut de Mathématiques de Marseille ..	président du jury
Muriel BOULAKIA .	MC HDR, Université Pierre et Marie Curie ..	examinateur
Stéphanie SALMON	Pr, Université de Reims Champagne-Ardenne	rapporteur
Omar LAKKIS	Pr, Free University of Bozen/Bolzano	rapporteur
Yves COUDIÈRE ...	Pr, Université de Bordeaux	directeur
Charles PIERRE ...	IR, Université de Pau et des pays de l'Adour .	co-encadrant

– 2017 –

THESIS

PRESENTED TO

THE UNIVERSITY OF BORDEAUX

DOCTORAL SCHOOL OF MATHEMATICS AND COMPUTER SCIENCE

by **Douanla Lontsi Charlie**

TO OBTAIN THE DEGREE OF

DOCTOR OF PHILOSOPHY

SPECIALISATION : APPLIED MATHEMATICS AND SCIENTIFIC COMPUTING

High order schemes for realistic simulations in cardiac electrophysiology

---

Reviewers : Stéphanie Salmon (Pr, Université de Reims Champagne-Ardenne),
Omar Lakkis (Pr, Free University of Bozen/Bolzano).

Defence date : 15th November 2017
In front of the jury members :

Florence HUBERT .	Pr, Institut de Mathématiques de Marseille ..	president of jury
Muriel BOULAKIA .	MC HDR, Université Pierre et Marie Curie ..	jury member
Stéphanie SALMON	Pr, Université de Reims Champagne-Ardenne	reviewer
Omar LAKKIS	Pr, Free University of Bozen/Bolzano	reviewer
Yves COUDIÈRE ...	Pr, Université de Bordeaux	adviser
Charles PIERRE ...	IR, Université de Pau et des pays de l'Adour .	co-adviser

– 2017 –

Remerciements

Je remercie mes directeurs de thèse Yves Coudière et Charles Pierre pour le soutien inestimable qu'ils m'ont apporté durant cette thèse. Je les remercie en particulier de m'avoir initier à la recherche et de m'avoir donné tous les conseils dont j'avais besoin.

Je remercie tous les membres de l'équipe Carmen pour la convivialité qu'ils ont fait régner dans le bureau commun et dans la vie de l'équipe en générale. Je remercie particulièrement Andjela, Mehdi, Marc, Antoine, Pauline pour leur soutien et les différents échanges très utiles que nous avons eu durant le déroulement de cette thèse. Je remercie tout ceux qui ont participé de près ou de loin au bon déroulement de ce travail.

Je remercie mon épouse Lontsi Darelle qui a été toujours là quand j'ai eu besoin d'elle. Je remercie mon père Douanla Thomas et ma mère Nobou Jeannette pour les conseils qu'ils n'ont jamais cessés de me donner depuis que je suis tout petit.

Funding

This thesis was entirely founded by the ANR-HR-CEM project (ANR-13-MONU-0004) : 36 months of phd salary, 2 weeks of study visit in the University of Ottawa (Canada), 5 days in Tunis (Tunisia) for CARI (Colloque Africain sur la Recherche en Informatique et Mathématiques Appliquées), 5 days in Voss (Norway) for ENUMATH (European Conference of Numerical Mathematics and Advanced Applications).
The thesis was performed in Carmen team at Inria Bordeaux Sud Ouest Research Center and IHU Liryc.

Résumé

Les simulations numériques réalistes en électrophysiologie cardiaque ont un coût de calcul extrêmement élevé. Ce coût s'explique en grande partie par la raideur, à la fois en temps et en espace, d'une onde de « potentiel d'action » (PA). Par ailleurs, les phénomènes observés sont très instationnaires et s'étudient en temps long. Une description précise de la dynamique des PA est cruciale pour construire des modèles numériques pertinents d'un point de vue médical ou clinique. Cet aspect fondamental ne peut être contourné dans les études numériques réalistes.

La raideur de l'onde de PA ne peut être captée numériquement qu'en ayant recours à des maillages très fins. Ces maillages très fins induisent un coût de calcul très important, et introduisent aussi des erreurs supplémentaires : les systèmes linéaires à résoudre deviennent très mal conditionnés. Au final, les erreurs numériques peuvent être particulièrement grandes dans les simulations alors que leur contrôle est évidemment essentiel pour assurer la fiabilité des résultats. Jusqu'à présent, très peu de résultats sont disponibles pour assurer cette fiabilité. Dans les faits, les erreurs sont la plupart du temps contrôlées par des procédés empiriques. Il existe quelques résultats théoriques étudiant la convergence et la stabilité des schémas numériques associés. En pratique, en plus d'avoir un contrôle de l'erreur sur le potentiel, il est aussi nécessaire d'avoir un contrôle de l'erreur sur des quantités macroscopiques décrivant la dynamique de l'onde de PA : temps d'activation, durée du PA, propriétés de restitution... Ces quantités ont en effet une interprétation physiologique qui permet de caractériser le caractère arythmogène des tissus.

Les modèles sont des systèmes d'EDP de réaction-diffusion couplés avec des systèmes d'équations différentielles pouvant être très raides, les modèles ioniques. Ils sont actuellement discrétisés par éléments finis conforme (Lagrange) et par des schémas en temps d'ordre un ou deux. Dans ce travail, nous concevons et évaluons l'intérêt d'utiliser des méthodes d'ordre supérieure pour ces systèmes. Parallèlement nous introduisons d'une part une nouvelle classe de schémas appelé schémas exponentiel Adams Bashforth intégral (IEAB), et d'autre part des schémas Rush Larsen (RL) d'ordre élevé. Ces nouveaux schémas sont des schémas multipas de type exponentiels. Nous montrons qu'ils possèdent des bonnes propriétés de stabilité et permettent de faire face efficacement à la raideur des modèles ioniques. Les schémas que nous proposons sont comparés numériquement (en terme de précision, coût en temps de calcul et stabilité) à plusieurs schémas classiques, ainsi qu'aux schémas exponentiels (RL1, RL2) communément utilisés pour des simulations

en électrophysiologie cardiaque. Nous proposons des techniques permettant de calculer avec précision les quantités d'intérêts cliniques (temps d'activation, de récupération, durée du potentiel d'action). Des résultats théoriques de convergence en temps et de convergence globale (espace et temps) sont énoncés et prouvés. Ces résultats sont ensuite illustrés numériquement à travers le modèle monodomaine et les modèles ioniques de Beeler Reuter, de Ten Tusscher et al. L'intérêt d'utiliser des schémas d'ordre élevés est aussi évalué sur des ondes spirales en 2D et 3D.

Abstract

Realistic numerical simulations in cardiac electrophysiology have a computational cost of extremely high. This cost is largely explained by the stiffness both in time and space, of the action potential (AP) wave. Moreover, the observed phenomena are very unsteady and are studied in long time. A precise description of the dynamic of AP is crucial for constructing relevant numerical models, from a medical or clinical perspective. This fundamental aspect can not be circumvented in realistic numerical studies.
The stiffness of AP wave can only be captured numerically, by using very fine meshes. In addition to the high computational cost, these very fine meshes also introduce additional errors : the linear systems to solve become very badly conditioned. In the end, the numerical errors can be particularly large whereas their control is obviously essential to ensure the reliability of the results. So far very few results are available to ensure this reliability. In practice, the errors are mostly controlled by empirical processes. In practice, in addition of having a control of the error on the potential, it is also necessary to have an error control on macroscopic quantities describing the dynamics of the AP wave : activation time, AP duration, properties of restitution ... These quantities have indeed a physiological interpretation which allows to characterize the arrhythmogenic character of the tissues.
The models are systems of reaction diffusion PDE coupled with systems of differential equations that can be very stiffs (ionic models). They are currently discretized by conforming finite elements (Lagrange finite elements methods) and by schemes in time of order one or two. In this work, we design and evaluate the interest of using higher order methods for these systems. At the same time, we introduce on the one hand, a new class of schemes called Integral Exponential Adams Bashforth (IEAB) schemes and, on the other hand, high order Rush Larsen (RL) schemes. These new schemes are exponential time-stepping schemes. We show that they have good stability properties and can efficiently cope with the stiffness of ionic models. The schemes we propose are numerically compared (in terms of accuracy, CPU time and stability) with several classical schemes, as well as with the exponential schemes (RL1, RL2), commonly used for cardiac electrophysiology simulations. We propose good techniques for accurately calculating quantities of clinical interest (activation time, recovery time, duration of action potential). Theoretical results of convergence in time and global convergence (in space and time) are stated and proved. These results are then illustrated numerically through the monodomain model and the ionic models

of Beeler Reuter, Ten Tusscher et al. The advantage of using high order schemes is also evaluated on spiral waves in 2D and 3D.

Table des matières

Table des figures

Liste des tableaux

1 Introduction générale et motivation de la thèse

Depuis les années 350 A.J.C, le cœur est classé parmi les organes les plus importants du corps humain. Durant ces dernières décennies, la recherche sur le cœur a été motivée d'une part par le désir de découvrir son secret en temps qu'organe vital, d'autre part par son importance en médecine et en santé des populations. En effet une partie non négligeable de la population mondiale est touchée par la mort due à un dysfonctionnement cardiaque. D'après l'organisation mondiale de la santé [1], 13% (chiffres en 2012) des décès dans le monde sont causés par un dysfonctionnement cardiaque. La compréhension du mode de fonctionnement du cœur pourrait amener à découvrir des nouvelles techniques permettant de diagnostiquer et de régler certains de ces problèmes.

La fonction principale du cœur est de pomper le sang dans l'organisme. Cette fonction est assurée par un système électrique très complexe qui malgré les avancées sur sa compréhension, n'est pas encore totalement compris. De nombreux progrès sont encore nécessaires pour améliorer des dépistages et pour comprendre le système électrique du cœur. Cette mission rassemble des chercheurs de plusieurs domaines de la science. Notamment les médecins, les ingénieurs, les mathématiciens, les biologistes et beaucoup d'autres. Chacun selon son domaine essaye d'apporter son expertise afin de faire avancer la recherche sur le sujet. Ceci se fait en général à travers des expériences cliniques ou bien la construction des modèles mathématiques. À cause des contraintes financières induites par des expériences cliniques, les chercheurs sont de plus en plus tournés vers la construction de modèles mathématiques. Ceci leur permet de faire moins d'expériences cliniques et par conséquent de faire des économies financières.

La construction et l'exploitation des modèles décrivant l'activité électrique du cœur sont très complexes et se heurtent à de nombreuses difficultés. Une de ces difficultés est la résolution numérique efficace des équations différentielles obtenues dans les modèles. Cette difficulté est dû à la raideur des équations, car celles-ci décrivent des phénomènes rapides, instables non-linéaires et multi-échelles. Leur résolution numérique nécessite donc des schémas robustes pouvant faire face efficacement à la raideur et à la non-linéarité, tout en restant précis. Ceci n'est pas toujours évident dans la mesure où les schémas explicites classiques sont limités par leurs propriétés de stabilité et n'offrent donc pas cette robustesse face à la raideur. Par ailleurs, à cause de la non-linéarité des équations, les

schémas implicites classiques coûtent très cher en temps de calcul. La technique utilisée en cardiologie pour faire face à cette difficulté est l'utilisation des schémas Implicites-explicites ainsi qu'exponentiels.
Jusqu'ici, ces techniques sont limitées à l'ordre 1 et 2 de précision. Ceci contraint à l'utilisation des discrétisations très fines lorsque l'on veut faire des simulations réalistes. La discrétisation fine n'est pas sans conséquence, elle est très coûteuse en temps de calcul et est à l'origine des erreurs additionnelles d'arrondi et de phénomènes d'artefacts. L'objectif de cette thèse est donc d'apporter une amélioration aux techniques de résolution numériques déjà présentes en construisant des schémas d'intégration en temps précis, stables, moins chers en temps de calcul et simple à mettre en œuvre.

Le manuscrit est constitué de 8 chapitres, le 1^{er} étant l'introduction et le $8^{ème}$ la conclusion.
Dans le Chapitre 2, nous introduisons quelques notions générales sur les cellules cardiaques. Nous décrivons les deux modèles les plus populaires décrivant la propagation électrique du potentiel dans le tissus cardiaque. Il s'agit du modèle bidomaine et monodomaine. Nous y décrivons aussi les modèles ioniques de Beeler-Reuter [2], Ten Tusscher et al [5]. Ces modèles font partie des modèles évolués décrivant l'activité chimio-électrique des cellules cardiaques. Leur description sera assez brève, vu que l'objectif de cette thèse n'est pas centrée sur le développement des modèles mathématiques en eux-même, mais sur des modèles numériques permettant de les résoudre. Les méthodes numériques que nous développerons sont principalement liée à l'évolution des modèles ioniques qui n'ont cessé de s'améliorer, en augmentant en parallèle les difficultés de résolution numérique. Nous allons faire un état des lieux sur les modèles ioniques, les méthodes de résolution numérique des équations aux dérivées partielles. Nous mettrons un accent particulier sur les modèles mathématiques en cardiologie.

Dans le Chapitre 3, nous proposerons un nouveau schéma d'intégration en temps des EDO. Ce schéma sera nommé "Integral Exponential Adams Bashforth " (IEAB), dont le schéma Exponentiel Adams Bashforth (EAB) en est un cas particulier. Les propriétés numériques de ce schéma seront étudiées. Par la consistance et la stabilité sur perturbation, nous allons énoncer et montrer en particulier, des résultats de convergence de ces schémas. Ces résultats seront illustrés à travers le modèle ionique de Beeler-Reuter [2]. Nous ferons aussi une étude de la stabilité au sens de Dahlquist. Cette étude nous permettra de construire les domaines de stabilité des schémas IEAB et EAB. Nous ferons le calcul des pas de temps critiques tout en les comparant à ceux de certains schémas classiques. Ceci sera fait sur les modèles ioniques de Beeler-Reuter, de Ten Tusscher et al. . Le calcul des pas de temps critiques nous permettra d'évaluer la robustesse des schémas IEAB et EAB par rapport à la raideur des problèmes étudiés.

Le chapitre 4 sera destiné à la généralisation des schémas Rush-Larsen (RL) [4, 3], qui offrent une simplicité en terme d'implémentation et une amélioration en terme de coût par rapport aux IEAB et EAB. Dans ce chapitre, nous allons proposer une méthode qui permet de construire des schémas de type Rush-Larsen à l'ordre arbitraire. Cette méthode nous permettra d'introduire de nouveaux schémas notamment les schémas $RL3$ et $RL4$. Comme pour les schémas IEAB et EAB, les propriétés numériques de ces schémas seront étudiées. Nous énoncerons et montrerons en particulier, des résultats de convergence suite à l'étude de la consistance et la stabilité sous perturbation. Les résultats de convergences seront vérifiés numériquement à travers le modèle ionique de Beeler-Reuter [2]. Nous ferons aussi une étude de la stabilité au sens de Dahlquist qui nous permettra de représenter les domaines de stabilités des schémas RL. Nous ferons le calcul des pas de temps critiques tout en les comparant à ceux des schémas EAB. Ceci sera fait sur les modèles ioniques de Beeler-Reuter [2] et celui de ten Tusscher et al. [5] . Ce calcul nous permettra d'évaluer la robustesse des schémas RL par rapport à la raideur des problèmes étudiés.

Dans le Chapitre 5, nous faisons une évaluation de l'efficacité des schémas EAB, RL et d'autre schémas classiques utilisés en cardiologie, notamment, les schémas Runge Kutta (RK), Backward differentiation (BDF), Crank Nicolson (CN) et Adams Bashforth. Nous décrivons pour cela des outils d'analyse numérique et des critères de comparaisons qui prennent en compte à la fois le coût d'un schéma et sa précision. Nous montrons d'ailleurs par ce critère que le schéma $RL3$ offre une alternative très intéressante pour résoudre les équations des modèles de la membrane cellulaire. Nous proposons aussi dans ce chapitre une méthode permettant de calculer sans dégrader la précision certaines quantités à intérêt particulier en cardiologie. Il s'agit en occurrence du temps d'activation (t_a), du temps de récupération (t_r) et de la durée du potentiel d'action (APD). Nous montrons que sur ces quantités physiologiques, il est possible d'observer l'ordre du schéma numérique utilisé dans la résolution du modèle.

Dans le chapitre 6, nous montrons comment il faut utiliser les schémas EAB et RL pour le modèle monodomaine. Nous faisons des combinaisons des méthodes d'éléments finis de type Lagrange ($\mathbb{P}_1$ et $\mathbb{P}_2$) avec les schémas RL et $S-BDF$ (Semi-Backward Differentiation). Nous énonçons et prouvons théoriquement sur certaines hypothèses de régularité, des théorèmes de stabilité, de convergence pour les schémas d'intégration en espace et en temps $\mathbb{P}_1 + RL1 + FBE$ et $\mathbb{P}_2 + RL2 + SBDF2$. Dans ces théorèmes, les estimations des erreurs sont faites en norme L^2.

Le chapitre 7 est consacré aux résultats numériques sur le modèle monodomaine. Dans ce chapitre, nous montrons numériquement que les solution calculées par les schémas numériques que nous avons obtenus en faisant des combinaisons des schémas $\mathbb{P}_r$, RL et $SBDF$ convergent vers la même solution. Une étude numérique quantitative en 2D et qualitative en 3D des précisions sur des ondes spirales est aussi présentées dans ce chapitre, pour évaluer la précision de ces schémas pour des simulations des phénomènes

répétitifs comme plusieurs battements de cœur.

Les chapitres 3 et 4 ont fait l'objet de deux communications scientifiques et sont d'ailleurs présentés sous forme d'articles. Il s'agit des articles référencés par,

> COUDIÈRE, YVES, CHARLIE DOUANLA LONTSI, AND CHARLES PIERRE. "Exponential Adams Bashforth ODE solver for stiff problems." (2016).
>
> COUDIÈRE, YVES, CHARLIE DOUANLA LONTSI, AND CHARLES PIERRE. "Rush Larsen time stepping methods of high order for stiff problems in cardiac electrophysiology." (2017).

Le chapitre 5 a fait l'objet d'un proceeding accepté et publié dans la conférence africaine pour la recherche en informatique et mathématiques appliquées (CARI).

> CHARLIE DOUANLA LONTSI, COUDIÈRE, YVES, AND CHARLES PIERRE. Efficient high order schemes for stiff ODEs in cardiac electrophysiology. *Colloque africain pour la recherche en informatique et mathématiques appliquées*, Hammamet, Tunis (October 2016), P. 312-319.

Cette thèse, a été aussi l'objet de plusieurs activités scientifiques qui n'apparaissent pas directement dans le manuscrit. Il s'agit notamment des abstracts présentés dans plusieurs conférences scientifiques, des séminaires dans des laboratoires scientifiques internationaux et des contributions à du développement de logiciel de calculs hautes performances. Comme abstract présenté à divers journées scientifiques, conférences et séminaire, nous avons par exemple :

> CHARLIE DOUANLA LONTSI, COUDIÈRE, YVES, AND CHARLES PIERRE. High-order ODE solver for cardiac electrophysiology, IHU Liryc workshop on cardiologie, Bordeaux(France), Sep 2016.
>
> CHARLIE DOUANLA LONTSI, COUDIÈRE, YVES, AND CHARLES PIERRE. High-order Rush-Larsen time-stepping methods for cardiac electrophysiology, European Numerical Mathematics and Advanced Application, Voss, Norway, Sep 2017.
>
> CHARLIE DOUANLA LONTSI. High order time-stepping methods for stiff ODEs in cardiology, Seminar in the departement of mathematics and statistic, University of Ottawa (Canada), April 2017.
>
> CHARLIE DOUANLA LONTSI. Schémas d'ordre élevé pour des EDPs raides en Cardiologie, Séminaire au département de mathématiques, Université de Nantes (France), Mai 2017.

Pour le travail de développement durant cette thèse, les schémas proposés dans nos travaux ont été implémentés dans le logiciel de calcul CEPS (Cardiac Electrophysiology Simulator) qui est un code de calcul développé en C++ dans l'équipe Carmen permettant de faire des calculs sur des modèles de cardiologie.

Références

[1] http://www.who.int/mediacentre/factsheets/fs317/fr/ (cf. p. 1).

[2] G.W. BEELER et H. REUTER. „Reconstruction of the Action Potential of Ventricular Myocardial Fibres“. English. In : *J. Physiol.* 268 (1977), p. 177–210 (cf. p. 2, 3).

[3] L. GERARDO-GIORDA, M. PEREGO et A. VENEZIANI. „Optimized Schwarz coupling of bidomain and monodomain models in electrocardiology“. In : *M2AN* (2010) (cf. p. 3).

[4] S RUSH et H LARSEN. „A practical algorithm for solving dynamic membrane equations.“ In : *IEEE Trans Biomed Eng* 25.4 (juil. 1978), p. 389–92 (cf. p. 3).

[5] KHWJ TEN TUSSCHER, D NOBLE, PJ NOBLE et Alexander V PANFILOV. „A model for human ventricular tissue“. In : *American Journal of Physiology-Heart and Circulatory Physiology* 286.4 (2004), H1573–H1589 (cf. p. 2, 3).

2 Électrophysiologie cellulaire, propagation et schémas numériques

Contents

2.1 Électrophysiologie cellulaire

2.1.1 La membrane cellulaire.

La cellule est entourée par la membrane cellulaire dont l'épaisseur est d'environ 7.5 à 10 nm ([31], chap. 2) . La membrane cellulaire sépare le milieu intérieur de la cellule (le cytoplasme) du milieu extérieur. Elle contrôle l'ensemble des échanges entre le cytoplasme et le milieu extérieur. Du point de vue structurel, la membrane cellulaire est formée de deux couches de molécules appelées phospholipides. L'une est intérieure et l'autre est extérieure. Ces deux couches sont systématiquement opposées. Un phospholipide est constitué d'une tête hydrophile (qui est attirée par l'eau) et d'une queue hydrophobe (qui est repoussée par l'eau). La composition aqueuse du cytoplasme et du milieu extérieur de la cellule force les queues des phospholipides constituant chaque couche à pointer vers l'intérieur de la membrane cellulaire. Ceci permet à la membrane de la cellule d'assurer son

imperméabilité et sa cohésion mécanique. Le flux des substances à travers la membrane de la cellule est assuré par des protéines qui transpercent la membrane cellulaire. Le transport d'une substance par ces protéines est le résultat d'un processus chimique qui lui permet d'absorber la substance d'une de ses extrémité et de la libérer par l'autre extrémité. Chacune de ces protéines est en charge du transport dans un seul sens (de l'intérieur vers l'extérieur de la cellule ou dans le sens contraire) d'une substance précise. Ce transport peut avoir un comportement actif ou passif. Dans le cas du comportement actif, l'activité des protéines est alimentée énergiquement par le métabolisme cellulaire de sorte que le transport des substances puisse se faire à contre courant des gradients de concentration. La composition chimique du cytoplasme est différente de celle du milieu extérieur. Par rapport au milieu extérieur, le cytoplasme est fortement concentré en ion potassium K^+ et faiblement concentré en ion sodium Na^+ et calcium Ca^{2+}. Ces différences de concentrations induisent une différence de potentiel (le potentiel transmembranaire) entre le cytoplasme et le milieu extérieur, et l'on dit que la cellule est polarisée.

La différence de potentiel électrique entre les deux faces de la membrane cellulaire est

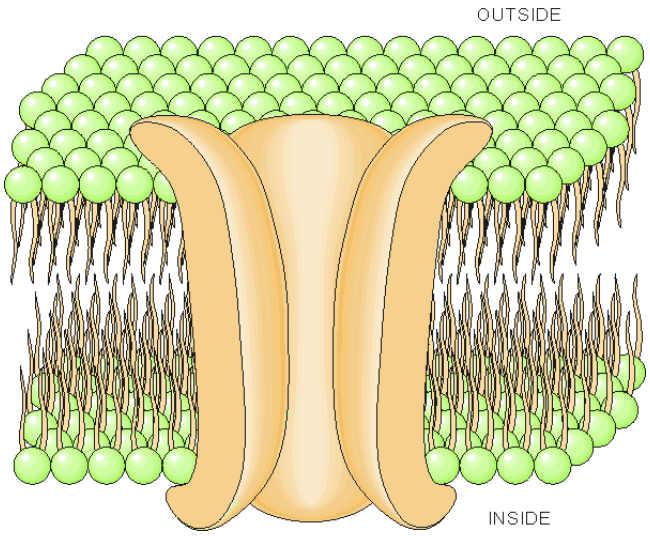


FIGURE 2.1: Construction d'une membrane. Deux couches de phospholipides, avec des queues hydrophobes pointant à l'intérieur de la membrane. Pores macromoléculaires (jonctions) dans la membrane de la cellule provenant des canaux ioniques qui rendent possible les phénomène bioélectriques. ([31], chap. 2)

d'autant variable que les différences des concentrations en ions sodium, potassium et calcium entre les milieux intra et extracellulaire. Ces différences de concentrations varient en fonction des modifications des conductances membranaires à ces ions. La membrane constitue une fine couche isolante séparant les milieux intra et extracellulaire, qui sont des conducteurs. Elle peut dès lors être considérée comme un condensateur de capacité C_m . En effet, la charge totale transférée q permettant d'obtenir une différence de potentiel u est telle que,

$$q = C_m u \tag{2.1}$$

Par dérivation de (2.1), on a

$$\frac{dq}{dt} = C_m \frac{du}{dt} = -I_{ion}, \tag{2.2}$$

où I_{ion} est la somme de tous les courants circulant dans la membrane cellulaire. Cette équation constitue la base des modèles décrivant l'activité électrique des cellules cardiaques.

2.1.2 Excitabilité des cellules cardiaques

Les cellules cardiaques et les cellules nerveuses font partie des cellules dont le potentiel membranaire peut varier au cours du temps. Ces cellules sont capables de répondre activement à un stimulus extérieur par une modification de leur potentiel membranaire. On dit alors qu'elles sont excitables. Cette propriété d'excitabilité est régie par des propriétés dynamiques de la membrane. Le stimulus permettant d'exciter des cellules musculaires peut être une modification du potentiel extérieur ou l'application d'un courant électrique. Au repos, le potentiel membranaire des cellules est négatif. Sous l'effet d'un stimulus, ce potentiel croit vers des valeurs positives, on parle de la dépolarisation de la cellule. Le changement du potentiel transmembranaire sous l'effet d'un stimulus dépend du degré de stimulation. Jusqu'à un certain degré, la dépolarisation est proportionnée à la stimulation. Mais si la stimulation amène la cellule à se dépolariser au delà d'une certaine valeur appelée potentiel seuil, cette dépolarisation n'est plus de nature proportionnée. On assiste alors au potentiel d'action cellulaire (voir figure 2.2). Dans ce cas, la valeur du potentiel membranaire croît d'abord rapidement et brusquement, ensuite se stabilise autour d'une valeur dite valeur de plateau où elle varie très faiblement et lentement, avant de revenir à son potentiel de repos. On dit que la cellule se repolarise. Plus spécifiquement, le potentiel d'action d'une cellule cardiaque suit une dynamique qui peut se découper en quatre phases numérotées de 0 à 4 (voir figure 2.2).

— **Phase 0 :** Encore appelée montée du potentiel d'action, elle correspond au moment où la cellule se dépolarise rapidement (durée de l'ordre de la milliseconde).

— **Phase 1 :** Encore appelée pointe ou spike, elle correspond à la phase de repolarisation initiale rapide plus ou moins marquée selon le tissu cardiaque considéré.

— **Phase 2 :** Elle est aussi appelée phase de repolarisation lente et est caractérisée par un "plateau" plus ou moins rectangulaire selon le tissus cardiaque et l'espèce animale.

— **Phase 3 :** elle correspond à la phase de repolarisation finale.

Le courant rentrant d'ions sodium Na^+ et le courant sortant d'ions potassium K^+ sont les principaux courants responsables du potentiel d'action cellulaire. Le courant d'ions sodium Na^+ est le principal acteur de la repolarisation. Lorsque le stimulus force la cellule à se dépolariser au delà du potentiel seuil, la membrane devient perméable à l'ion sodium Na^+. On assiste à un flux rentrant d'ions Na^+. Ce flux est responsable de la croissance rapide et brusque du potentiel transmembranaire à la dépolarisation. Après la dépolarisation, la

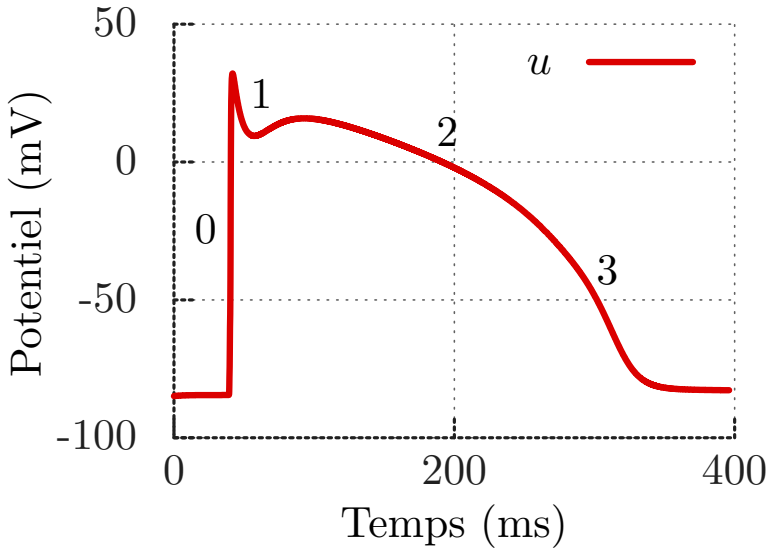


FIGURE 2.2: Potentiel d'action transmembranaire avec ses quatre phases

membrane de la cellule devient perméable à l'ion K^+. Cette perméabilité initie un flux d'ions potassium qui permet de faire décroître le potentiel membranaire vers sa valeur de repos. Même si les ions sodium Na^+ et potassium K^+ jouent un rôle très important dans le début et la fin du potentiel d'action, ce ne sont pas les seuls ions intervenant dans la biochimie électrique qui a lieu dans la membrane cellulaire lors du potentiel d'action. L'ion calcium joue aussi un rôle très important dans ce processus, notamment dans la phase plateau. Nous apporterons beaucoup plus de détails sur ce rôle dans la section 2.1.4.
Les échanges ioniques entre le milieu intra et extracellulaire se font par des transports ioniques. Ces transports sont variés et se font par des voies différentes et spécifiques.

2.1.3 Transports ioniques

Les protéines de la membrane chargées du transport ionique peuvent être classées en deux catégories. Les canaux ioniques et les transporteurs. La dynamique des différents flux ioniques intervenant dans le transport ionique membranaire est un phénomène complexe qui n'est pas encore bien compris. Il n'est donc pas envisageable de le décrire entièrement. On se limite à la description du transport ionique de trois principaux ions : le sodium Na^+, le potassium K^+ et le calcium Ca^{2+}.

Canaux ioniques

Un canal ionique a pour rôle de laisser passer dans un sens une espèce ionique conformément à son gradient électrochimique. Son comportement est modélisé par une résistance. Ce transport est passif du fait qu'il ne requiert aucun apport en énergie. Cependant, la conductivité d'un canal ionique est variable selon les conditions extérieures. Ce

comportement fait donc de lui un conducteur actif. Plus précisément, le courant I_X d'un ion X à travers le canal ionique associé est fonction de son gradient électrochimique et défini par la loi d'Ohm suivante,

$$I_X = g_X \Psi(u - u_X), \tag{2.3}$$

où g_X est la conductance, u le potentiel transmembranaire, u_X le potentiel électrochimique de l'ion X et Ψ une fonction pouvant être linéaire ou non suivant le type d'ion. Le potentiel u_X est donné par la loi de Nersnt en fonction des concentrations intra et extracellulaire $[X_i]$ et $[X_e]$ respectivement en ion X.

$$u_X = \frac{RT}{z_X F} ln \frac{[X_i]}{[X_e]}, \tag{2.4}$$

où R, T, F et z_X sont respectivement la constante des gaz parfaits, la température, la constante de Faraday et la valence de X. L'état courant d'un canal ionique revêt une certaine marge d'incertitude. Les protéines étant réparties de façon dense sur la membrane, le comportement d'un type d'ion est influencé par celui des autres. Les comportements de deux canaux de même type peuvent être différents. Ceci induit un aspect probabiliste sur le comportement global d'une cellule. On introduit donc la variable $0 \leq \omega \leq 1$ appelée variable de porte qui représente la probabilité d'ouverture du canal d'un type d'ion X et sa conductivité maximale $\overline{g}_X$. La conductance g_X est définie par,

$$g_X = \overline{g}_X \omega. \tag{2.5}$$

Dans la majorité des modèles ioniques construits pour les cellules cardiaques, en particulier les plus récents [40, 37, 8], la variable ω est décrite par l'équation différentielle ordinaire,

$$\frac{d\omega}{dt} = \frac{\omega_\infty(u) - \omega(u)}{\tau_\omega(u)}, \tag{2.6}$$

où ω_∞ et τ_ω sont respectivement la probabilité d'ouverture du canal à l'état stationnaire et la constante de temps caractéristique.

Les transporteurs

Les transporteurs ont pour rôle principal de ramener, entre deux potentiels membranaires successifs, les concentrations des espèces ionique à leurs valeurs normales. Il en existe deux types : les pompes et les échangeurs.

— **Les pompes :** Ce sont des protéines membranaires ayant la capacité de faire entrer et sortir des espèces ioniques à contre-courant de leur gradients électrochimiques. L'énergie nécessaire pour alimenter leur transport est directement fournie par le métabolisme cellulaire. En effet, cette énergie est fournie par les molécules d'ATP (Adénosine Tri-Phosphate) dont l'hydrolyse constitue la source énergétique de la

cellule. Comme exemple de pompe, nous avons la pompe Na^+/K^+-ATPase. Pour une molécule d'ATP hydrolysée, cette pompe fait simultanément rentrer deux ions potassium K^+ et sortir trois ions sodium Na^+. Lorsque la cellule est au repos, la concentration en ions potassium est forte tandis que celle en ion sodium est faible. À la fin du potentiel d'action la situation est inversée. La cellule est riche en ions potassium et pauvre en ions sodium. La restitution des concentrations initiales de la cellule en ion potassium et sodium est rendu possible par l'activation de la pompe Na^+/K^+. L'activité de cette pompe est une fonction des concentrations $[K]_{i,e}$ et $[Na]_{i,e}$, à l'ATP et au potentiel transmembranaire. Le courant traversant cette pompe est noté $I_{Na/K}$. On a donc,

$$I_{Na/K} = f(u, [Na]_{i,e}, [K]_{i,e})$$

Les courants en ion sodium $I_{Na,Na/K}$, et en ion potassium $I_{K,Na/K}$ qui traversent cette pompe sont calculées à partir du courant $I_{Na/K}$ par les relations,

$$I_{Na,Na/K} = 3I_{Na/K} \quad , \quad I_{K,Na/K} = -2I_{Na/K}.$$

Ceci permet de réactualiser les concentrations intra et extracellulaires en sodium et potassium.

— **Les échangeurs :** Les transporteurs n'utilisent pas directement l'énergie provenant du métabolisme cellulaire dans leur tache de transport ionique. Ils réalisent donc un transport ionique actif à travers la membrane de la cellule. L'énergie qui alimente leur fonction est fournie par un ion qui suit son gradient électrochimique. Ce phénomène couple un canal ionique à un transporteur membranaire et utilise l'énergie de l'un pour activer l'autre. Ce phénomène est appelé transport couplé. Un exemple d'échangeur est l'échangeur Na^+/Ca^{2+}. Il permet d'échanger trois ions sodium Na^+ contre un ion calcium Ca^{2+}. Son activation permet à la cellule de retrouver ses concentrations initiales en sodium et calcium. Le flux ionique à travers cet échangeur est calculé par une loi explicite prenant en compte les concentrations intra et extracellulaire des ions qui le traversent ainsi que le potentiel membranaire.

$$I_{NaCa} = \phi(u, [Ca]_{i,e}, [Na]_{i,e})$$

Les courants en ions sodium $I_{Na,NaCa}$ et en ions calcium $I_{Ca,NaCa}$ à travers ce transporteur sont calculés à partir de ce courant.

Le courant qui circule dans la membrane cellulaire et noté I_{ion} et est la somme des courants ioniques I_X circulant dans les canaux, pompes ou transporteurs X considérés.

$$I_{ion} = \sum_X I_X.$$

2.1.4 Les stocks de calcium

La dynamique du calcium joue un rôle important dans l'initiation de la contraction du muscle cardiaque. En effet, dans le milieu intracellulaire de la cellule, le calcium est toxique et sa concentration doit rester faible (environ $10^{-7}M$, $M = mol/L$). Une augmentation de $10^{-6}M$ initie une contraction. Bien que l'entrée du calcium dans la cellule soit à l'origine de cette augmentation, la majeure partie provient du stock de calcium dans le milieu intracellulaire. On distingue plusieurs mécanismes de régulation de cette concentration.

— **Les tampons :** Les tampons sont des protéines ayant la capacité de stocker le calcium libre dans le milieu intracellulaire. Comme exemple de tampon, on a la troponine et la calmoduline. Ces deux tampons jouent le rôle de capteur d'ions calcium lorsque la concentration du milieu intracellulaire en calcium est au delà d'un certain niveau. Dans le cas échéant, leur action initie une contraction. Cette dynamique est traduite par l'équation

$$\frac{dCa_{Tn}}{dt} = k_{on}^{Tn} Ca_i (B_{max}^{Tn} - Ca_{Tn}) - k_{off}^{Tn} Ca_{Tn}, \tag{2.7}$$

où k_{on}^{Tn} et k_{off}^{Tn} sont les taux de réactions directe et inverse, et B_{max}^{Tn} est la capacité maximale de stockage de calcium sur la troponine. D'autres tampons comme la myosine qui a une grande affinité avec le calcium et le magnésium, peuvent être en compétition. Leurs dynamiques sont décrites par les équations suivantes.

$$\frac{dCa_{M-Ca}}{dt} = k_{on}^{M} Ca_i (B_{max}^{M} - Ca_{M-Ca} - Ca_{M-Mg}) - k_{off}^{M} Ca_{M-Ca}, \tag{2.8}$$

$$\frac{dCa_{M-Mg}}{dt} = k_{on}^{M} Mg_i (B_{max}^{M} - Ca_{M-Ca} - Ca_{M-Mg}) - k_{off}^{M} Ca_{M-Mg}. \tag{2.9}$$

— **Le réticulum sarcoplasmique (RS) :** C'est un organite qui représente un grand réservoir de calcium et contient une molécule tampon (la calséquestrine). Pendant la dépolarisation, un flux entrant de Ca^{2+} pénètre la cellule. En même temps, le RS libère du calcium dans le milieu intracellulaire par l'usage de deux canaux ioniques. Ces canaux sont des canaux de largage et de fuite de calcium. Les mouvements cordonnés d'entrées et de sorties simultanées des ions calcium Ca^{2+} dans la cellule entraînent ensuite une contraction. Ceci est dû au raccourcissement des sarcomères provoqué par des ponts actine-myosine formé par le calcium à l'origine. Lors de la repolarisation, le RS recupère le calcium par la pompe Ca^{2+}. À l'intérieur du RS, le calcium est stocké par la calséquestrine. Le contenu du RS en calcium est quantifiable et sa dynamique est décrite par l'équation,

$$\frac{dCa_{SRT}}{dt} = J_{up} - J_{rel} - J_{leak}, \tag{2.10}$$

2.1.5 Exemples de modèles de la membrane cellulaire

On trouve dans la littérature plusieurs types de modèles de la membrane cellulaire. Comme souligné dans [17], ces modèles peuvent être classifiés en deux grands groupes : Les modèles physiologiques et les modèles réduits de la membrane.

Modèles physiologiques

Ce sont des modèles qui sont construits de façon à intégrer au maximum des informations qui expliquent les phénomènes électrophysiologiques les plus importants. Ce type de modèles a pour but d'être utilisé par exemple dans des études pharmacologiques ou comme outils prédictifs pour des nouvelles pathologies. Comme exemples de modèles physiologiques de la membrane, nous avons le modèle de Beeler et Reuter [8], le modèle de Ten Tusscher et al. [40].

— **Le modèle de Beeler et Reuter [8] :** C'est un modèle ionique physiologique qui décrit le potentiel d'action ventriculaire d'un mammifère. Ce modèle prend en compte quatre types de courants comme on peut observer sur la figure 2.3. Une de ses propriétés importante est la description de la dynamique du calcium intracellulaire. Ce modèle se traduit par le système différentiel à 8 équations suivantes,

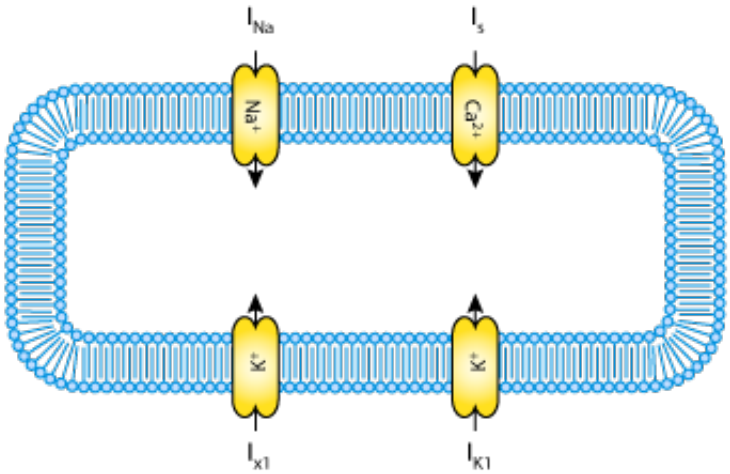


FIGURE 2.3: Différents courants ioniques pris en compte par le modèle ionique de Beeler et Reuter.

$$\begin{cases} \dfrac{d\omega}{dt} = \dfrac{\omega_\infty - \omega}{\tau_\omega}, & \omega \in \{x_1, m, h, j, d, t\}, \\ \dfrac{dCa_i}{dt} = -10^{-7} i_s + 0.07(10^{-7} - Ca), \\ \dfrac{du}{dt} = -\dfrac{I_{ion} + I_{stim}}{C_m}. \end{cases} \tag{2.11}$$

Avec $I_{ion} = I_{K1} + I_{x1} + I_{Na} + I_{Ca}$ et I_{stim} un courant extérieur qui sert à stimuler la cellule. Notons qu' à cause de la complexité de ses équations et pour des raisons de lisibilité, le système ci-dessus nous donne juste la forme contracté du modèle. Pour plus de détails, on peut consulter [8].

— **Le modèle de Ten Tusscher et al. [40] :** C'est un modèle physiologique de cellule cardiaque ventriculaire humaine. Il est construit à base des données expérimentales humaines et animales. La plupart des courants ioniques est établie sur des mesures expérimentales sur les tissus humains et inclut une dynamique de base pour le calcium. Le courant de calcium de type L (celui des canaux calciques voltage-dépendants) ainsi que les équations décrivant la dynamique du calcium intracellulaire sont obtenus sur la base des mesures expérimentales sur des animaux. Il contient douze courants ioniques provenant des canaux ioniques et des transporteurs. Les ions intervenant dans ce modèles peuvent être observés sur la figure 2.4 Le modèle mathématique

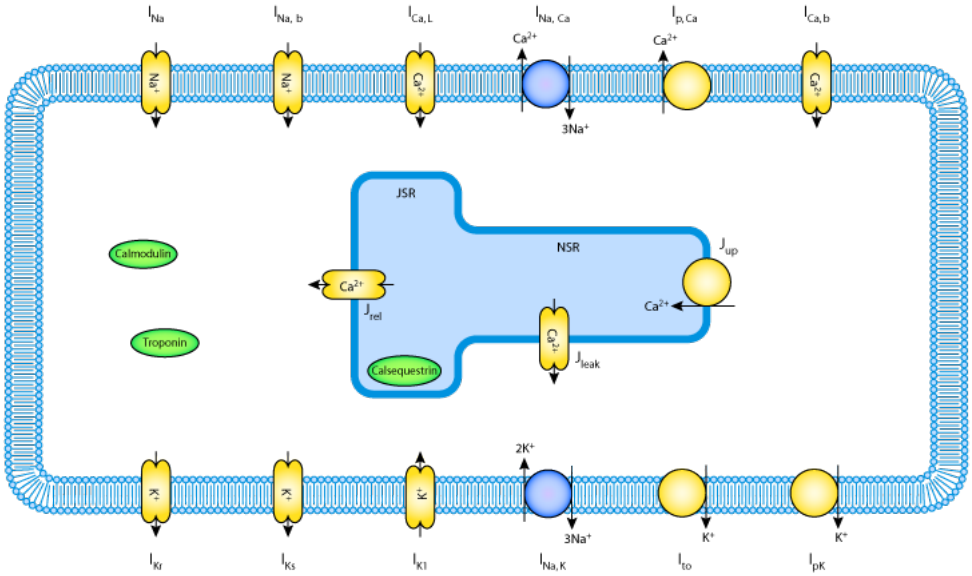


FIGURE 2.4: Modèle schématique du modèle de Ten Tusscher et al (2004). Modèle mathématique du myocyte ventriculaire humain

simplifié dont tous les détails sont dans [40] se traduit par le système différentiel à 17 équations,

$$\begin{cases} \dfrac{d\omega}{dt} = \dfrac{\omega_\infty - \omega}{\tau_\omega}, \qquad \omega \in \{d, f, f_{Ca}, g, h, m, j, r, s, x_s, x_{r1}, x_{r2}\} \\ \dfrac{dK_i}{dt} = -\dfrac{I_{K1} + I_{to} + I_{Kr} + I_{Ks} - 2I_{NaK} + I_{pK} + I_{stim} - I_{ax}}{FV_C}, \\ \dfrac{dNa_i}{dt} = -\dfrac{I_{Na} + I_{bNa} + 3I_{NaK} + 3I_{NaCa}}{FV_C}, \\ \dfrac{dCa_{itotal}}{dt} = -\dfrac{I_{CaL} + I_{pCa} + I_{bCa} - 2I_{NaCa}}{2FV_C} + J_{leak} - J_{up} + J_{rel}, \\ \dfrac{Ca_{srtotal}}{dt} = \dfrac{V_C}{V_{SR}}(-J_{leak} + J_{up} - J_{rel}), \\ \dfrac{du}{dt} = -\dfrac{I_{ion} + I_{stim}}{C_m}. \end{cases} \tag{2.12}$$

Avec $I_{ion} = I_{Na}+I_{K1}+I_{to}+I_{Kr}+I_{Ks}+I_{CaL}+I_{NaCa}+I_{NaK}+I_{pCa}+I_{pK}+I_{bCa}+I_{bNa}$ et I_{stim} un courant extérieur qui sert à stimuler la cellule.

Du fait de l'évolution des technologies et des méthodes expérimentales, la récolte des données expérimentales devient de plus en plus efficace avec une richesse très croissante d'informations. Ceci accroît considérablement la complexité des modèles physiologiques, entraînant ainsi la motivation de la création des modèles plus simples.

2.1.6 Modèles simplifiés

Les modèles simplifiés sont des modèles issus des modèles phénoménologiques. Ils sont basés sur la reproduction du caractère excitable ou inexcitable de la cellule. Ils sont plus simples que les modèles physiologiques. Par une approche géométrique, ils reproduisent le comportement des modèles physiologiques. Par rapport aux modèles physiologiques, ces modèles ont l'avantage d'intégrer un petit nombre de paramètres. Comme exemples de ce type de modèle, nous avons le modèle de Fitz Hugh et Nagumo [19] qui est une simplification du modèle de Hodgkin et Huxley [25], le modèle d'Aliev et Panfilov [5].

— *Modèle de Fitz Hugh et Nagumo [19]* : C'est un modèle constitué de deux inconnues, le potentiel transmembranaire u et une variable de recouvrement w. Ce modèle est décrit par le système d'équation,

$$\begin{aligned} \frac{du}{dt} &= c_1 u(u-a)(1-u) - c_2 w + I_{stim}, \\ \frac{dw}{dt} &= b(u - c_3 w). \end{aligned}$$

Avec a, b, c_1, c_2, c_3 des paramètres données par le modèle, pouvant être ajustés pour simuler différents types de cellules. C'est un modèle qui reproduit les caractéristiques les plus importantes du potentiel d'action.

— **Modèle d'Aliev et Panfilov [5] :** C'est une version modifiée du modèle de Fitzhugh Nagumo proposé en 1996 par Aliev et Panfilov. Cette version permet de décrire

adéquatement la dynamique du potentiel d'action cardiaque. Elle est assez simple pour être utilisée dans des études numériques. Ce modèle est constitué de deux paramètres dont le potentiel transmembranaire et une variable de recouvrement w. Ce modèle est décrit par le système d'équation suivant.

$$\begin{aligned}\frac{du}{dt} &= -k(u-a)(u-1) - uw + I_{stim},\\ \frac{dw}{dt} &= \varepsilon(u,w)(-w - ku(u-a-1)).\end{aligned}$$

Avec $\varepsilon(u,w) = \varepsilon_0 + \mu_1 w/(u+\mu_2)$ et I_{stim} un courant de stimulation. $k, a et \varepsilon_0$ sont des constantes positives données par le modèle. μ_1 et μ_2 sont des paramètres pouvant être modifiés en fonction de la cellule.

2.1.7 Formulation abstraite et difficultés numériques de résolution

Comme on peut l'observer sur les différents modèles de membranes précédemment décrits, les modèles ioniques sont du point de vue mathématique des systèmes d'équations différentielles ordinaires du premier ordre. On peut les écrire sous la forme,

$$\begin{cases}\frac{dw_i}{dt} = \frac{w_{\infty,i}(u)-w_i}{\tau_i(u)},\\ \frac{dc}{dt} = g(w,c,u),\\ \frac{du}{dt} = -I_{\text{ion}}(w,c,u) + I_{stim}(t),\end{cases} \tag{2.13}$$

où $w = (w_1,\dots,w_p) \in \mathbb{R}^p$ est le vecteur des variables de portes, $c \in \mathbb{R}^q$ est le vecteur contenant les concentrations ioniques ou autres variables d'états, $u \in \mathbb{R}$ est le potentiel membranaire. Les quatre fonctions, $w_{\infty,i}(u)$, $\tau_i(u)$, $g(w,c,u)$ et $I_{\text{ion}}(w,c,u)$ sont des termes de réactions donnés par le modèle. La fonction $I_{stim}(t)$ est un terme source.
On peut écrire (2.13) sous des formes plus contractées. Soit $y = (w,c,u) \in \mathbb{R}^N$ (avec $N = p+q+1$), le vecteur contenant toutes les variables décrites par un modèle de membrane donné. On peut écrire (2.13) sous la forme,

$$\frac{dy}{dt} = f(t,y), \tag{2.14}$$

avec,

$$f(t,y) = \begin{pmatrix}(w_\infty(u)-w)/\tau(u)\\ g(y)\\ -I_{\text{ion}}(y) + I_{stim}(t)\end{pmatrix}, \quad (w_\infty(u)-w)/\tau(u) = \left(\frac{w_{\infty,i}(u)-w_i}{\tau_i(u)}\right)_{i=1\dots p}.$$

On peut aussi exploiter la forme quasi-linéaire des équations sur les variables de portes et écrire (2.13) sous la forme synthétique,

$$\frac{dy}{dt} = a(y)y + b(t,y). \tag{2.15}$$

avec,

$$a(y) = \begin{pmatrix} -1/\tau(u) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad b(t,y) = \begin{pmatrix} w_\infty(u)/\tau(u) \\ g(y) \\ -I_{\text{ion}}(y) + I_{stim}(t) \end{pmatrix},$$

et $-1/\tau(u) = \text{diag}\left(-1/\tau_i(u)\right)_{i=1\ldots p}$.

Les deux formulations (2.14) et (2.15) ont pour intérêt de faciliter l'étude et la résolution numérique des équations provenant des modèles ioniques. La formulation (2.14) est adaptée à l'implémentation des schémas numériques linéaires. Elle permet de faire le calcul des valeurs propres de la Jacobienne du système. Ces valeurs propres permettent en effet d'étudier la raideur du système.

La formulation (2.15) permet d'écrire le système comme somme d'une partie "linéaire" ($a(y)y$) et d'une partie non-linéaire ($b(t,y)$). La jacobienne de $a(y)y$ contient la plus grande raideur du système. Le terme $b(t,y)$ n'est pas très raide mais reste difficile à intégrer à cause de la non-linéarité. La formulation (2.15) est pratique pour les mise en œuvre des schémas nécessitant une linéarisation. Notamment les schémas de type exponentiels. En effet, le terme $a(y)$ est diagonal et contient une partie de la Jacobienne suffisante pour représenter la raideur du système. Ce terme ($a(y)$) peut donc représenter de façon efficace la Jacobienne du système.

Les difficultés rencontrées lors de la résolution numérique des équations qui décrivent les modèles ioniques sont principalement liés à la raideur et la non-linéarité des équations. Si les modèles simplifiés comme le modèle de Fitz Hugh et Nagumo [19] et celui d'Aliev et Panfilov [5] précédemment décrits, sont moins contraignants en terme de raideur, les modèles physiologiques comme celui de Beeler et Reuter [8] ou de Ten Tusscher et al. [40] présentent une raideur beaucoup plus sévère. Les modèles simplifiés peuvent être résolus efficacement par des schémas numériques explicites classiques (linéaires). La forme (2.14) est alors bien adaptée pour la résolution numérique des schémas simplifiés. Par contre, dans le cas des modèles physiologiques, les schémas explicites classiques sont contraints à l'utilisation des petits pas de temps pour assurer leur stabilité. L'utilisation des petits pas de temps induit un coût en temps de calcul très élevé. Les fonctions qui décrivent les modèles ioniques physiologiques sont constituées d'une bonne quantité d'exponentielles. Le nombre d'évaluation de ces fonctions doit donc rester le plus faible que possible. L'usage des schémas implicites doit être évité. Les schémas explicites doivent être privilégiés. Des schémas numériques permettant de faire faces à la raideur des modèles ioniques physiologiques ont déjà été proposé par plusieurs chercheurs dans le domaine de la cardiologie. Notamment Rush et Larsen [38], Perego et Veneziani, [36], Sundness [35]. Ces schémas sont basés sur la formulation (2.15) et sont principalement des schémas de type exponentiels d'ordre 1 ou 2. Nous apporterons plus de détails sur leur description dans la section 2.3.

2.2 Propagation électrique cardiaque

Les cellules du muscle cardiaque sont connectées entre elles de telle sorte qu'une cellule stimulée peut passer le signal électrique aux cellules voisines. Le transfert du signal d'une cellule aux cellules voisines se fait à travers une initialisation du potentiel d'action des cellules voisines. Cette capacité de communication permet à une stimulation électrique appliquée à une partie du cœur, de se propager à travers tout le muscle cardiaque et d'activer tout le cœur. Le phénomène de propagation électrique dans le cœur est un phénomène difficile à modéliser même dans le cas où on considère un petit nombre de cellules. Étant conscient du nombre trop grand de cellules dans le cœur, on est contraint pour des raisons mathématiques et numériques d'approximer le tissus cardiaque par des modèles continus. En plus des raisons liées au nombre trop grand des cellules, s'ajoute le fait de la non accessibilité de la géométrie précise du réseau cellulaire microscopique.

2.2.1 Modèle bidomaine

Le modèle bidomaine est basé sur une approche de type milieu continu. Le modèle utilise l'idée de la séparation du tissu cardiaque en deux domaines tous continus : le milieu intracellulaire et le milieu extracellulaire. Le domaine intracellulaire est supposé connexe à cause des jonctions lacunaires. Les jonctions lacunaires sont des petits canaux (voir 2.1) incorporés dans la membrane de la cellule, permettant le contact direct entre son milieu interne et celui des cellules voisines. Grâce à ces jonctions lacunaires, des ions ou des petites molécules peuvent passer directement d'une cellule à l'autre sans entrer dans l'espace entre les deux cellules (le milieu extracellulaire). Pour un point considéré dans chaque domaine, on y définit le potentiel comme étant une quantité moyennée sur un petit volume l'entourant. Chaque point du domaine est à la fois dans les deux domaines, soit l'intracellulaire et l'extracellulaire. On y associe alors les potentiels des deux milieux : le potentiel intracellulaire u^i et le potentiel extracellulaire u^e. Puisque les deux domaines sont connexes, la membrane (la réunion de toutes les membranes des cellules) qui les sépare doit être aussi supposée connexe. Cette membrane agit comme un isolant entre les deux milieux ; ce qui induit une différence de potentiel $u = u^e - u^i$ (le potentiel dont il a été question dans la section précédente). Le modèle bidomaine complet est un système

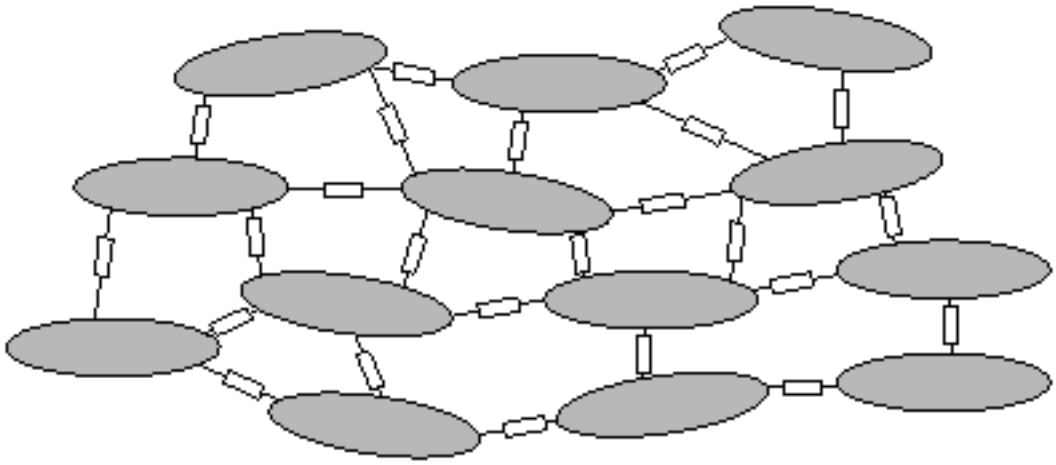


FIGURE 2.5: Vue schématique du milieu intra et extracellulaire ([31], chap. 9)

d'EDP parabolique dont la forme générale peut être écrite comme suit,

$$\frac{\partial w}{\partial t} = g(w,u) \quad \text{dans} \quad]0,T[\times\Omega, \tag{2.16}$$

$$C_m\frac{\partial u}{\partial t} - \frac{1}{\chi}\nabla\cdot(\sigma_i\nabla u_i) - I_{ion}(w,u) = -I_{stim,i} \quad \text{dans} \quad]0,T[\times\Omega, \tag{2.17}$$

$$C_m\frac{\partial u}{\partial t} - \frac{1}{\chi}\nabla\cdot(\sigma_e\nabla u_e) - I_{ion}(w,u) = I_{stim,e} \quad \text{dans} \quad]0,T[\times\Omega, \tag{2.18}$$

$$n\cdot(\sigma_i\nabla(u+u_e)) = 0 \quad \text{sur} \quad]0,T[\times\partial\Omega, \tag{2.19}$$

$$n\cdot(\sigma_e\nabla u_e) = 0 \quad \text{sur} \quad]0,T[\times\partial\Omega, \tag{2.20}$$

$$w(.,0) = w^0 \quad \text{dans} \quad \Omega, \tag{2.21}$$

$$u_i(.,0) = u_i^0 \quad \text{dans} \quad \Omega, \tag{2.22}$$

$$u_e(.,0) = u_e^0 \quad \text{dans} \quad \Omega, \tag{2.23}$$

où σ_i , σ_e sont respectivement les tenseurs de conductivités intra et extracellulaire. C_m est la conductivité de la membrane par unité d'aire, χ est l'aire de la membrane par unité de volume. Les fonctions g, I_{ion} sont données par le modèle ionique et I_{stim_i} (rep : I_{stim_e}) est un courant de stimulation extérieur. Les propriétés de conductivité du tissu cardiaque sont fortement anisotropes. En effet le muscle cardiaque est constitué des fibres et la conductivité est plus élevée dans la direction des fibres que dans la direction transversale des fibres. Ceci explique en pourquoi les quantités σ_i , σ_e sont des tenseurs. En un point donné x, on peut définir un repère orthonormé (a_l, a_t, a_n), où a_l est la direction le long des fibres. Dans ce repère les tenseurs de conductivités sont des matrices diagonales.

$$\sigma_{i,e}(x) = Diag(\sigma_{i,e}^l, \sigma_{i,e}^t, \sigma_{i,e}^n).$$

2.2.2 Modèle monodomaine

En posant certaines conditions sur les tenseurs de conductivités des milieux intra et extra cellulaire σ_i et σ_e respectivement, il est possible de simplifier le modèle bidomaine à une équation scalaire décrivant seulement la dynamique du potentiel transmembranaire u. En effet, si on suppose les ratio d'anisotropies égaux, c'est-à-dire

$$\sigma_e = \lambda \sigma_i, \tag{2.24}$$

avec λ un scalaire constant, alors σ_e peut être éliminé dans (2.17)-(2.20). On obtient alors le modèle monodomaine,

$$\frac{\partial w}{\partial t} = g(w,u) \quad \text{dans} \quad]0,T[\times\Omega, \tag{2.25}$$

$$C_m \frac{\partial u}{\partial t} - \frac{1}{\chi}\nabla\cdot(\sigma\nabla u) - I_{ion}(w,u) = I_{stim} \quad \text{dans} \quad]0,T[\times\Omega, \tag{2.26}$$

$$n\cdot(\sigma\nabla u) = 0 \quad \text{sur} \quad]0,T[\times\partial\Omega, \tag{2.27}$$

$$w(.,0) = w^0 \quad \text{dans} \quad \Omega, \tag{2.28}$$

$$u(.,0) = u^0 \quad \text{dans} \quad \Omega, \tag{2.29}$$

où u est le potentiel transmambranaire, I_{stim} un courant externe permettant de stimuler les cellules, $\sigma = \frac{\lambda}{1+\lambda}\sigma_i$. Le tenseur σ_i, les constantes C_m et χ sont les mêmes que dans la section précédente. Les fonctions g et I_{ion} sont décrites par le modèle ionique. En particulier pour le modèle ionique Beeler Reuter [8], w sera défini par un 7-uplet constitué des variables de portes et de la concentration en ion calcium du milieu intracellulaire. La fonction g sera donc donnée par les deux premières équations du système (2.11). I_{ion} sera donné par la dernière équation du même système.

2.2.3 Formulation abstraite et difficultés numérique de résolution

Tout au long de notre travail, nos études seront basées sur le modèle monodomaine. Nous nous intéresserons donc particulièrement à ce modèle dans cette partie.
Une fois le modèle ionique donné, l'équation du monodomaine couplée au modèle ionique peut être formulée de façon générale par,

$$\begin{cases} \frac{\partial w}{\partial t} = \frac{w_\infty(u)-w}{\tau(u)}, \\ \frac{\partial c}{\partial t} = g(w,c,u), \\ \frac{\partial u}{\partial t} = \mathrm{div}(G(x)\nabla u) + f_1(w,c,u,t,x), \end{cases} \tag{2.30}$$

avec w, c, u des fonctions dépendant des variables en espace et en temps $x \in \Omega \subset \mathbb{R}^d$, $t \in \mathbb{R}$. $w(0,.) = w^0$, $c(0,.) = c^0$, $u(0,.) = u^0$ et la condition au bord (2.27) est maintenue. La fonction $u : \Omega \times \mathbb{R} \to \mathbb{R}$ désigne le potentiel membranaire, les fonctions $w : \Omega \times \mathbb{R} \to \mathbb{R}^p$ et $c : \Omega \times \mathbb{R} \to \mathbb{R}^{N-p-1}$ désignent respectivement les variables de

portes et les concentrations provenant du modèle ionique. La fonction f_1 est définie par $f_1(w,c,u,t,x) = \frac{1}{C_m}(I_{\text{ion}}(w,c,u) + I_{stim}(t,x))$ et le tenseur $G(x) = \frac{1}{C_m \chi}\sigma(x)$.
Sous une forme plus synthétique, (2.30) peut s'écrire,

$$\frac{\partial y}{\partial t} = a(y)y + b(t,x,y) \tag{2.31}$$

avec, $y = (w,c,u)$, $-1/\tau(u) = \text{diag}\,(-1/\tau_i(u))_{i=1\ldots p}$,

$$a(y) = \begin{pmatrix} -1/\tau(u) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & A \end{pmatrix}, \quad b(t,y) = \begin{pmatrix} w_\infty(u)/\tau(u) \\ g(y) \\ f_1(y,t,x) \end{pmatrix}.$$

A désigne l'opérateur linéaire spatial $div(G(x)\nabla .)$.
On peut aussi écrire (2.30) sous la forme plus générale,

$$\frac{\partial z}{\partial t} = f_2(z,u), \tag{2.32}$$

$$\frac{\partial u}{\partial t} = Au + f_1(z,u,x,t), \tag{2.33}$$

avec $z = (w,c)$ et $f_2(z,u) = \left(\frac{w_\infty(u)-w}{\tau(u)}, g(z,u)\right)$. Le terme Au est un terme de diffusion, les fonctions f_1 et f_2 sont des termes de réactions particulièrement non-linéaires. On obtient donc un problème de type réaction diffusion.
La résolution numérique du problème (2.32) doit faire face à plusieurs difficultés. Ces difficultés proviennent de la raideur du système. Cette raideur provient d'une part, de la discrétisation en espace de l'opérateur A et d'autre part de la raideur de la fonction f_2 et de la non linéarité des fonctions f_1 et f_2. En effet, après la discrétisation en espace du système (2.32)-(2.33), la matrice approximant l'opérateur A est une matrice possédant des valeurs propres à parties réelles négativement grandes. La plus grande partie réelle négativement de ces valeurs propres tend vers l'infini au fur et à mesure que la discrétisation en espace devient fine. La raideur de l'équation (2.33) provient donc de la discrétisation de Au et dépend du maillage. Elle repose sur la condition CFL venant de f_1.

2.3 Schémas Numériques en temps pour les EDO et les EDP en Cardiologie

La résolution numérique des équations mathématiques provenant des modèles est incontournable pour une exploitation effective. Quel que soit le domaine scientifique, on a besoin des résultats numériques pour valider un modèle. Pour un type de problème donné, il existe une large variété de méthodes de résolution numérique. Résoudre numériquement une équation différentielle consiste à utiliser un ou plusieurs schémas d'intégration numérique pour approcher sa solution exacte. Chaque schéma numérique est utilisé selon son efficacité à produire le résultat souhaité. Dans le domaine de la cardiologie, la

diversité des schémas numériques d'intégration utilisés pour des simulations est assez vaste et dédiée principalement à la résolution des équations des modèles monodomaine et bidomaine [42, 27]. Comme nous avons décrit en section 2.2.3, ces modèles sont des équations aux dérivées partielles paraboliques et font partie de la classe des équations de réaction-diffusion. Ils sont formés d'une partie EDO provenant du modèle ionique et une partie EDP provenant de la propagation du potentiel membranaire dans le tissu cardiaque. Nous avons vu en particulier (voir section 2.2.3) que le monodomaine pouvait être formulé sous la forme,

$$\frac{\partial z}{\partial t} = f_2(z, u), \tag{2.34}$$

$$\frac{\partial u}{\partial t} = Au + f_1(z, u, x, t), \tag{2.35}$$

où Au est le terme de diffusion et f_1, f_2 sont des termes de réaction. La relation (2.34) représente la partie EDO et la relation (2.35) représente la partie EDP. La variable z est une variable vectorielle qui permet de coupler l'EDP à l'EDO. Les simulations numériques par ces modèles sont à objectifs variés et peuvent être à plusieurs usages. Certaines simulations comme dans [9, 32, 11, 12, 28] ont pour but de produire des outils permettant de donner des réponses à des questions de recherches scientifiques, notamment médicales. D'autres simulations comme cela sera le cas dans nos travaux ont pour but d'investiguer des résultats d'analyse numérique ou d'améliorer des techniques de simulation [28, 7, 10, 15]. L'intégration numérique en temps de ces équations nécessite au préalable une intégration en espace. Dans notre travail, nous utiliserons pour la discrétisation en espace les méthodes d'éléments finis Lagrange $\mathbb{P}_1$ et $\mathbb{P}_2$ [chap. 6, sec. 6.3].
Une fois une EDP discrétisée en espace, on obtient une équation différentielle ordinaire que l'on peut résoudre par des méthodes numériques d'intégration en temps. Il existe aussi une grande variété de schémas d'intégration en temps. Ceux qui sont utilisés en cardiologie sont explicites, implicites, semi-implicites ou exponentiels. Nous allons décrire et commenter ces schémas à pas de temps fixe Δt, et pour une suite d'instants $(t_n = t_0 + n\Delta t)_{n\in\mathbb{N}}$ donné.

2.3.1 Schémas explicites et implicites

Il existe en général trois types de schémas explicites : Les schémas à 1 pas, les schémas multipas et les schémas multiétages. Soit $(v^n)_{n\in 0,\dots,N}$ un $N+1$-uplet représentant une solution numérique définie par un schéma et d un entier donné.

— **Schéma explicite à un pas.** Un schéma explicite à un pas est une relation de récurrence de la forme

$$v^{n+1} = v^n + \Delta t\phi(t^n, v^n, \Delta t), \tag{2.36}$$

où ϕ est une fonction de $\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^+$ dans $\mathbb{R}^d$ et v^0 est donné. L'exemple basique d'un tel schéma est le schéma d'Euler explicite. Appliqué au problème (2.34), le schéma d'Euler explicite se formule par,

$$z^{n+1} = z^n + \Delta t f_2(z^n, u^n).$$

— **Schéma explicite à k pas.** Un schéma explicite à k pas est une relation de récurrence de la forme

$$\sum_{j=0}^{k} \alpha_j v^{n+j} = \Delta t \phi(t^n, v^n, \ldots, v^{n-k+1}, \Delta t), \tag{2.37}$$

où ϕ est une fonction de $\mathbb{R}^+ \times (\mathbb{R}^d)^k \times \mathbb{R}^+$ dans $\mathbb{R}^d$, α_j et v^j, $j = 0, \ldots, k-1$ sont donnés. Un exemple de ce type de schéma est le schéma d'Adams-Bashforth (voir [20], Chap III). Appliqué au problème (2.34) le schéma d'Adams-Bashforth à deux pas se formule par,

$$z^{n+1} - z^n = \Delta t \left(\frac{3}{2} f_2(z^n, u^n) - \frac{1}{2} f_2(z^{n-1}, u^{n-1}) \right).$$

— **Schéma explicite à p étages.** un schéma explicite à p étages est une relation de récurrence de la forme (2.36) dont la fonction ϕ nécessite p autres relations de la même forme, avant que le calcul final de v^{n+1} soit possible. Un exemple bien connu de ce type de schéma est le schéma de Runge-Kutta d'ordre 2. C'est une schéma à deux étages. Appliqué au problème (2.34) où on omet la variable u pour simplifier, il peut se formuler comme suit,

$$\begin{aligned} z_{n1} &= z^n, \\ z_{n2} &= z^n + \Delta t f_2(z_{n1}), \\ z^{n+1} &= z^n + \frac{\Delta t}{2}(f_2(z_{n1}) + f_2(z_{n2})). \end{aligned}$$

Une façon très commode de représentation de ces schémas se fait par le tableau de Butcher (voir [21], chap 4).

Les schémas explicites à un pas sont les schémas les plus simples que l'on peut trouver lors de la résolution numérique d'une EDO. C'est en général par des schémas classiques à un pas comme le schéma d'Euler explicite que l'on fait ses tous premiers pas dans l'intégration numérique des EDO. Ce schéma est utilisé en Cardiologie pour faire des tests sur des modèles simples. Ses limites en terme de précision, ne favorisent pas toujours son utilisation. Les schémas explicites multipas et multiétages sont une solution pour surmonter le problème de précision dont souffre le schéma d'Euler explicite. Cependant, à cause du caractère raide des problèmes de cardiologie, l'utilisation directe (sans modification) des schémas explicites classiques est parfois impossible en pratique. En effet, la raideur des équations différentielles en cardiologie provient à la fois de la discrétisation de la diffusion dans le modèle physiologique et des modèles ioniques. Dans ce contexte, les

schémas explicites sont d'une part soumis à la condition CFL comme dans la résolution numérique des EDP classiques, et d'autre part à la raideur du système d'EDO qui décrit les modèles ioniques. Ces contraintes se traduisent en pratique par l'obligation de choisir un pas de temps Δt et un diamètre de maillage h suivant une certaine condition. Cette condition dépend en général du schéma utilisé. Par exemple pour le schéma Euler explicite associé à un schéma d'intégration en espace, la contrainte est $\Delta t \leq Ch^2$ [18] (C donné par la condition CFL imposé par l'EDP) et $\Delta t \leq 1/|\lambda|$ (λ valeur propre représentant la raideur du modèle ionique). C n'est en général pas très grand. Par ailleurs $|\lambda|$ peut être très grand comme dans le cas des modèles ioniques évolués en cardiologie. La conséquence en pratique est l'utilisation de Δt très petit, induisant donc des coût en temps de calcul très élevés.
Une alternative pour utiliser Δt grand est l'utilisation des schémas implicites. Comme pour le schéma explicite, il en existe en général trois types : Les schémas à 1 pas, les schémas multipas et les schémas multiétages.

— **Schéma implicite à un pas.** un schéma implicite à un pas est une relation de récurrence de la forme

$$v^{n+1} = v^n + \Delta t \phi(t^{n+1}, v^{n+1}, \Delta t), \tag{2.38}$$

où ϕ est une fonction de $\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^+$ dans $\mathbb{R}^d$ et v^0 est donné. Un exemple basique de ce type de schéma est le schéma d'Euler implicite. Appliqué au problème (2.34), il s'écrit comme suit,

$$z^{n+1} = z^n + \Delta t f_2(z^{n+1}, u^{n+1}).$$

— **Schéma implicite à k pas.** un schéma implicite à k pas est une relation de récurrence de la forme

$$\sum_{j=0}^{k} \alpha_j v^{n+j} = \Delta t \phi(t^{n+1}, v^{n+1}, \ldots, v^{n-k+1}, \Delta t), \tag{2.39}$$

où ϕ est une fonction de $\mathbb{R}^+ \times (\mathbb{R}^d)^k \times \mathbb{R}^+$ dans $\mathbb{R}^d$, α_j et v^j, $j = 0, \ldots, k-1$ sont donnés. La famille des schémas d'Adams-Moulton est un exemple de ce type de schémas. Le schéma d'Adams-Moulton d'ordre 2 est un schéma à deux pas. Appliqué au problème (2.34), il s'écrit,

$$z^{n+1} - z^n = \Delta t \left(\frac{3}{2} f_2(z^{n+1}, u^{n+1}) - \frac{1}{2} f_2(z^n, u^n \right).$$

— **Schéma implicite à p étages.** un schéma implicite à p étages est une relation de récurrence de la forme (2.38) dont la fonction ϕ nécessite p autres relations de la même forme, avant que le calcul final de v^{n+1} soit possible. Un exemple de ce type de schéma est la famille des schéma DIRK (Diagonal implicit Runge Kutta). L'ordre 2 de

cette famille de schémas appliqué à (2.34), où on omet la variable u pour simplifier est donné par,

$$
\begin{aligned}
z_{n1} &= z^n + \frac{\Delta t}{3} f_2(z_{n1}), \\
z_{n2} &= z^n + \Delta t \left(\frac{3}{4} f_2(z_{n1}) + \frac{1}{4} f_2(z_{n2}) \right), \\
z^{n+1} &= z^n + \Delta t \left(\frac{3}{4} f_2(z_{n1}) + \frac{1}{4} f_2(z_{n2}) \right).
\end{aligned}
$$

Les schémas implicites à un pas sont les schémas les plus simples que l'on peut trouver dans la famille des schémas implicites. Un exemple est le schéma de Euler implicite. C'est en général le schéma par lequel on fait ses premiers pas en implicite. Il est équivalent au schémas Euler explicite par rapport à la précision. Les schémas multipas et multiétages implicites offrent une meilleure précision par rapport aux schémas à un pas. Par ailleurs, ils ne souffrent pas tous des conditions de stabilité [18] précédemment décrites sur les schémas explicites. En pratique, l'utilisation des schémas implicites nécessite l'usage d'une méthode itérative pour le solveur implicite. Cet usage doit faire face à la non linéarité des modèles ioniques qui sont pour la plupart des fonctions exponentielles et rendent donc très coûteuses les simulations à grand pas de temps. En face de toutes ces difficultés liées aux schémas explicites et implicites, en exploitant l'efficacité des schémas implicites à résoudre les problèmes linéaires, la communauté du numérique en cardiologie s'est tourné vers les schémas implicites-explicites pour pouvoir améliorer ses simulations.

2.3.2 Schémas implicites-explicites (IMEX)

Lors de la résolution numérique d'un problème pouvant se mettre sous la forme (2.35), la fonction f_1 peut être non linéaire et avoir certaines propriétés ne favorisant pas son intégration par un schéma implicite. Dans les cas qui vont nous intéresser, la fonction f_1 provient des modèles ioniques et est constituée d'un grand nombre de fonctions exponentielles. Pour éviter de faire des calculs très coûteux, nous allons utiliser des schémas permettant d'évaluer le moins que possible la fonction f_1. Pour cela, nous utiliserons des schémas explicites pour l'intégration de f_1.
Le terme Au représente une raideur très forte. Cette raideur est dominante par rapport à celle de f_1. Nous allons intégrer le terme Au implicitement pour éviter d'être soumis à l'utilisation d'une condition CFL en h^2. Le terme linéaire Au est un terme de diffusion caractérisé par une matrice symétrique, creuse et définie positive. Les systèmes linéaires définies par des telles matrices peuvent être résolus efficacement par des méthodes itératives [43], notamment la méthode du gradient conjugué [26]. De plus, la matrice A étant constante, peut être pré-conditionnée efficacement. Ceci donne donc une raison tout à fait logique de l'usage des schémas explicites pour le terme f_1 et implicites pour le terme Au. Les schémas basés sur ce procédé sont appelés schémas implicites-explicites et IMEX en abrégé.
Nous nous référons à [6] pour dire que jusqu'à 1995, le schéma IMEX le plus populaire était

celui obtenu par la combinaison du schémas Adams Bashforth 2 pour la partie explicite (en f_1) et le schéma Crank Nicolson pour la partie implicite (en Au) [1, 16]. Plus précisément, il s'agit du schémas CNAB2 qui appliqué à (2.35) donne,

$$\frac{u^{n+1}-u^n}{\Delta t} = \frac{1}{2}A(u^{n+1}+u^n) + \frac{3}{2}f_1^n + \frac{1}{2}f_1^{n-1}. \tag{2.40}$$

À l'ordre 1 et 2, on énumère plusieurs combinaisons aboutissant à des schémas IMEX. Notamment le schéma Euler implicite explicite (FBE) ordre 1, le schéma Adams-Bashforth Crank-Nicolson modifié (MCNAB2 "*modified Crank-Nicolson*") ordre 2, le schéma Crank-Nicolson saute mouton (CNLF "Crank Nicolson Leap Frog") ordre 2. Tous ces schémas peuvent être trouvés dans les travaux de Ascher et Al [6]. Une autre combinaison très populaire est celle qui conduit au schéma différence finie semi rétrograde SBDF. Il s'obtient en intégrant la partie explicite par une formule d'extrapolation explicite et la partie implicite par la formule de différence finie rétrograde. Le schéma SBDF d'ordre 1 est le schéma FBE. Un avantage des schémas SBDF par rapport aux autres précédemment cités est qu'il a des versions allant jusqu'à l'ordre 5. Les schémas IMEX d'ordre 1 ou 2 peuvent aussi être obtenus par des méthodes de splitting ou à pas fractionnaires. Nous ne développerons pas ce type de schémas dans ce travail, mais pour une description détaillée, on peut voir [30, 29].

Plusieurs auteurs notamment Ascher, Crouzeix, Akrivis [16, 4, 2, 3, 6] ont travaillé sur les propriétés numériques des schémas IMEX. De tous ces travaux, il ressort que les schémas SBDF sont parmi les meilleurs en terme de stabilité. Dans les solveurs les plus utilisés en cardiologie notamment CARP [44], Chaste [33], les schémas IMEX sont utilisés à l'ordre 1 et 2. La raideur des modèles ioniques ne permet pas d'utiliser efficacement les schémas SBDF à l'ordre 3 ou plus. En effet, à l'ordre 1 et 2, on a des schémas qui permettent de faire face à la raideur des variables de portes que contiennent les équations de (2.34). Ces schémas appliqués à (2.34) et combinés aux schémas IMEX d'ordre 1 et 2 (pour l'équation (2.35)) permettent de résoudre le système (2.34)-(2.35) tout en restant robuste à la raideur du système. Cette robustesse permet d'utiliser des grands pas de temps et de faire des calculs à des petits coûts en temps de calcul. Cependant, à l'ordre 3 ou plus, on n'a pas des schémas permettant de faire face à la raideur des variables de portes, tout en gardant des faibles coûts en temps de calculs. Un des objectifs de ce travail sera de fabriquer des schémas d'ordre 3 ou 4 capables de faire face à la raideur des variables de portes, et à des faibles coûts. Ainsi on pourra combiner ces schémas aux schémas SBDF pour résoudre efficacement le système (2.34)-(2.35) par des schémas d'ordre 3 ou 4. Nous allons à cet effet nous intéresser aux schémas exponentiels. Mais avant de parler des schémas exponentiels, il est important de faire quelques commentaires sur l'initialisation des schémas multipas.

2.3.3 Initialisation des schémas multipas

Un des problèmes que rencontrent les schémas multipas réside au niveau de l'initialisation. En pratique, on a toujours accès à une valeur initiale. Pourtant les schémas multipas ont besoin des solutions à plusieurs instants selon le schéma, pour que leur mise en œuvre effective soit possible. Une technique assez commode est d'utiliser un schéma du même ordre mais à un-pas, pour calculer les solutions dont on a besoin pour l'initialisation. Cette pratique n'est pas toujours évidente et doit faire face à plusieurs difficultés. En effet, le schéma utilisé pour l'initialisation doit avoir le même ordre de précision que celui que l'on veut initialiser. De plus, il doit être de préférence plus stable pour ne pas détruire les propriétés de stabilité de l'autre schéma. Mais si les deux schémas ont les mêmes propriétés de stabilité, cela peut suffire pour l'initialisation. Pour assurer la stabilité, on utilise en général des schémas implicites stables à un-pas pour l'initialisation. Notamment les schémas Runge Kutta implicites. La conséquence de ceci est l'ajout d'un coût supplémentaire en temps de calcul à celui du schéma lui même. Il existe d'autre méthodes notamment l'utilisation des schémas exponentiels à un-pas mais multiétages (voir [24, 13]). En cardiologie, il est pratique de contourner le recours à l'usage d'autres schémas pour l'initialisation. En effet, les cellules cardiaques restent au repos pendant plusieurs millisecondes. Un état d'équilibre pendant lequel les différentes variables intervenant dans l'activité électrique restent constantes. Ceci permet donc d'utiliser les valeurs correspondantes à cet état pour l'initialisation des schémas multipas.

2.3.4 Schémas exponentiels

Les schémas exponentiels comme décrits dans [22] sont des intégrateurs numériques qui font intervenir des fonctions exponentiels dans leurs définitions. Ils permettent particulièrement d'intégrer numériquement des problèmes raides de la forme ,

$$\begin{cases} \dfrac{dy}{dt} = f(t,y), \quad t \in]0,T], \\ y(0) = y^0. \end{cases} \tag{2.41}$$

L'idée consiste à trouver sur chaque intervalle de discrétisation $[t_n, t_{n+1}]$, $t_n = n\Delta t$, une équation différentielle prototype de (2.41), qui possède des propriétés de raideur similaires, et qui peut être intégrée de façon exacte. Pour obtenir l'équation différentielle prototype, on écrit d'abord (2.41) sur chaque intervalle $[t_n, t_{n+1}]$, comme somme d'une partie linéaire et d'une partie non linéaire.

$$\begin{cases} \dfrac{dy}{dt} = J_n y + g_n(t,y), \quad t \in]t_n, t_{n+1}], \\ y(t_n) = y^n. \end{cases} \tag{2.42}$$

$g_n(t,y) = f(t,y) - J_n y$, J_n est une matrice dont les valeurs propres sont proches de celles de la jacobienne $\partial_y f(t_n, y(t_n))$ de f en $(t_n, y(t_n))$. Il est commode de prendre directement

$J_n = \partial_y f(t_n, y(t_n))$.
Le prototype est ensuite obtenu en approximant la fonction $g_n(t, y)$ dans (2.42) par une fonction $\tilde{g}_n(t)$. Ce prototype s'écrit donc,

$$\begin{cases} \dfrac{dz}{dt} = J_n z + \tilde{g}_n(t), \quad t \in]t_n, t_{n+1}], \\ z(t_n) = y^n. \end{cases} \tag{2.43}$$

Les schémas exponentiels sont alors obtenus en intégrant (2.43) par la formule de variation de la constante. L'approximation de y^{n+1} de $y(t_{n+1})$ par un schéma exponentiel est donné par $z(t_{n+1})$. Plus précisément,

$$y^{n+1} = z(t_{n+1}) = e^{J_n \Delta t} y^n + \Delta t \int_0^1 e^{(1-\tau) J_n \Delta t} \tilde{g}_n(t_n + \tau \Delta t) d\tau. \tag{2.44}$$

Le schéma exponentiel le plus simple est le schéma exponentiel Euler, il est obtenu en prenant $\tilde{g}_n(t) = f_n$. Il est donc défini par,

$$y^{n+1} = e^{J_n \Delta t} y^n + \Delta t \varphi_1(J_n \Delta t) f_n, \tag{2.45}$$

avec $\varphi_1(J_n \Delta t) = \int_0^1 e^{(1-\tau) J_n \Delta t} d\tau$, $f_n = f(t_n, y^n)$.
Le schéma exponentiel Euler est le premier schéma exponentiel introduit dans le domaine de la cardiologie. Ce schéma a été introduit dans le domaine en 1978 par Rush et Larsen [38] avec une mise en œuvre adaptée aux modèles rencontrés en cardiologie.
Rush et Larsen ont exploité le fait que les équations sur les variables de portes des modèles ioniques se présentent sous forme d'une partie "linéaire" $a(y)y$ et d'une partie non linéaire $b(t, y)$ (voir (2.14) de la section 2.1.7). Leur idée a été de prendre dans (2.44), $J_n = a(y^n) = a^n$ et $\tilde{g}_n(t) = b(t_n, y^n) = b^n$. Ils ont obtenu le schéma,

$$y^{n+1} = y^n + \Delta t \varphi_1(a^n \Delta t)(a^n y^n + b^n). \tag{2.46}$$

Le choix $J_n = a^n$ est adapté aux modèles ioniques et présente beaucoup d'avantages. Puisque a^n est une matrice diagonale, le calcul de $e^{a^n \Delta t}$ pour chaque intervalle de discrétisation $[t_n, t_{n+1}]$ se ramène au calcul de l'exponentiel de plusieurs scalaires. Le calcul de y^{n+1} par le schéma (2.46) est donc explicite. Par ailleurs, la matrice a^n est constituée de la partie de la jacobienne de $a(y)y + b(t, y)$ (de l'équation (2.14) des modèles ioniques) la plus raide sur $[t_n, t_{n+1}]$. a^n représente donc efficacement la raideur du système, dans le sens où ses valeurs propres sont proches de celles de $a(y)y + b(t, y)$.
Le schéma de Rush et Larsen (2.46) a eu beaucoup de succès dans le domaine de la cardiologie grâce à sa simplicité et sa capacité à faire face à la raideur des modèles ioniques. Le problème de cette méthode est qu'elle se limite à l'ordre 1 et nécessite donc l'utilisation de pas de temps très petits pour améliorer la précision. C'est donc un schéma qui coûte cher pour des simulations réalistes. Pour surmonter ce problème Perego et Veneziani ont proposé en 2009 [36] l'ordre 2 d'un schéma multipas exponentiel suivant la même idée que Rush et Larsen. Sundness et al.[35] ont aussi proposé la même année un schéma

d'ordre 2 nommé GRL. Le schéma GRL est plus stable que le schéma RL mais entre dans la classe des schémas à deux étages et est plus complexe que celui proposé par Peregot et Veneziani.

Les schémas de Perego et Veneziani [36] ainsi que le schéma de Sundness et al.[35] sont limités à l'ordre 2 de précision. Un des objectifs de ce travail sera de proposer des alternatives pour avoir des schémas exponentiels simples et d'ordre élevé du même type que celui de Rush et Larsen. Nous allons aussi nous intéresser aux schémas exponentiels Adams-Bashforth (EAB). Le schéma EAB a été introduit pour la première fois en 1960 par Certaine [14] et sous une forme plus systématique par Norset en 1969 [39]. La motivation de la construction de ce schéma était de donner une version du schéma classique d'Adams-Bashforth qui soit A-stable. À cette époque, les ordinateurs n'étaient pas assez performants pour calculer de façon efficace les exponentiels de matrices. Puisque la méthode nécessitait de tels calculs, la communauté du numérique a jugé ce schéma inutilisable en pratique. De nos jours, les technologies des ordinateurs et du calcul scientifique ont évolué considérablement par rapport à cette époque. Les ordinateurs offrent des mémoires capables de sauvegarder de très grands volumes de données. De plus les techniques du calcul parallèle avec des supercalculateurs croissent et prouvent leur efficacité. Pour ces raisons plusieurs auteurs s'intéressent à nouveau aux schémas exponentiels ces dernières années [24, 22, 45, 23]. L'utilisation des schémas exponentiels pour la résolution des EDP est fortement liée au calcul d'exponentiels de matrices. Il existe plusieurs méthodes d'approximation d'exponentiel de matrices. On a par exemple (Voir [34]) : la méthode des séries, l'approximation de Padé, "scaling and squaring", l'approximation rationnelle de Chebyshev, la méthode par les équations aux dérivées ordinaires, la méthode des polynômes, la méthode de décomposition des matrices et bien d'autres. Comme démontré dans [34], ces méthodes sont toutes insatisfaisantes pour des problèmes à grandes échelles car leur coût en temps de calcul est très élevé. Une méthode plus récente et qui est prometteuse est l'utilisation des espaces de Krylov pour approximer le produit exponentiel de matrice et vecteur au lieu de calculer la matrice [41, 46]. On s'arrange pour ne pas avoir ce problème en proposant des schémas dont l'exponentiel est facile à calculer.

Bibliographie

[1] L ABIA et JM SANZ-SERNA. „The spectral accuracy of a fully-discrete scheme for a nonlinear third order equation“. In : *Computing* 44.3 (1990), p. 187–196 (cf. p. 27).

[2] Georgios AKRIVIS et Michel CROUZEIX. „Linearly implicit methods for nonlinear parabolic equations“. In : *Mathematics of computation* 73.246 (2004), p. 613–635 (cf. p. 27).

[3] Georgios AKRIVIS, Michel CROUZEIX et Charalambos MAKRIDAKIS. „Implicit-explicit multistep finite element methods for nonlinear parabolic problems“. In : *Mathematics of Computation of the American Mathematical Society* 67.222 (1998), p. 457–477 (cf. p. 27).

[4] Georgios AKRIVIS et Yiorgos-Sokratis SMYRLIS. „Implicit–explicit BDF methods for the Kuramoto–Sivashinsky equation“. In : *Applied Numerical Mathematics* 51.2-3 (2004), p. 151–169 (cf. p. 27).

[5] Rubin R. ALIEV et Alexander V. PANFILOV. „A simple two-variable model of cardiac excitation“. In : *Chaos, Solitons & Fractals* 7.3 (1996), p. 293–301 (cf. p. 16, 18).

[6] Uri M ASCHER, Steven J RUUTH et Brian TR WETTON. „Implicit-explicit methods for time-dependent partial differential equations“. In : *SIAM Journal on Numerical Analysis* 32.3 (1995), p. 797–823 (cf. p. 26, 27).

[7] Ezio BARTOCCI, Elizabeth M CHERRY, James GLIMM et al. „Toward real-time simulation of cardiac dynamics“. In : *Proceedings of the 9th International Conference on Computational Methods in Systems Biology*. ACM. 2011, p. 103–112 (cf. p. 23).

[8] G.W. BEELER et H. REUTER. „Reconstruction of the Action Potential of Ventricular Myocardial Fibres“. English. In : *J. Physiol.* 268 (1977), p. 177–210 (cf. p. 11, 14, 15, 18, 21).

[9] Omer BERENFELD et José JALIFE. „Purkinje-muscle reentry as a mechanism of polymorphic ventricular arrhythmias in a 3-dimensional model of the ventricles“. In : *Circulation Research* 82.10 (1998), p. 1063–1077 (cf. p. 23).

[10] Miguel O BERNABEU, Pras PATHMANATHAN, Joe PITT-FRANCIS et David KAY. „Stimulus protocol determines the most computationally efficient preconditioner for the bidomain equations“. In : *IEEE Transactions on Biomedical Engineering* 57.12 (2010), p. 2806–2815 (cf. p. 23).

[11] Martin J BISHOP, Patrick M BOYLE, Gernot PLANK, Donald G WELSH et Edward J VIGMOND. „Modeling the role of the coronary vasculature during external field stimulation“. In : *IEEE Transactions on Biomedical Engineering* 57.10 (2010), p. 2335–2345 (cf. p. 23).

[12] Martin J BISHOP, David J GAVAGHAN, Natalia A TRAYANOVA et Blanca RODRIGUEZ. „Photon scattering effects in optical mapping of propagation and arrhythmogenesis in the heart“. In : *Journal of electrocardiology* 40.6 (2007), S75–S80 (cf. p. 23).

[13] M. P. CALVO et C. PALENCIA. „A class of explicit multistep exponential integrators for semilinear problems“. In : *Numer. Math.* 102.3 (2006), p. 367–381 (cf. p. 28).

[14] J. CERTAINE. „The solution of ordinary differential equations with large time constants“. In : *Mathematical methods for digital computers*. Wiley, New York, 1960, p. 128–132 (cf. p. 30).

[15] Elizabeth M CHERRY, Henry S GREENSIDE et Craig S HENRIQUEZ. „Efficient simulation of three-dimensional anisotropic cardiac tissue using an adaptive mesh refinement method“. In : *Chaos : An Interdisciplinary Journal of Nonlinear Science* 13.3 (2003), p. 853–865 (cf. p. 23).

[16] Michel CROUZEIX. „Une méthode multipas implicite-explicite pour l'approximation des équations d'évolution paraboliques“. In : *Numerische Mathematik* 35.3 (1980), p. 257–276 (cf. p. 27).

[17] Karima DJABELLA. „Modélisation de l'activité électrique du coeur et de sa régulation par le système nerveux autonome“. Thèse de doct. Université Paris Sud-Paris XI, 2008 (cf. p. 14).

[18] Marc ETHIER et Yves BOURGAULT. „Semi-implicit time-discretization schemes for the bidomain model“. In : *SIAM Journal on Numerical Analysis* 46.5 (2008), p. 2443–2468 (cf. p. 25, 26).

[19] Richard FITZHUGH. „Impulses and physiological states in theoretical models of nerve membrane“. In : *Biophysical journal* 1.6 (1961), p. 445–466 (cf. p. 16, 18).

[20] E. HAIRER, S. P. NORSETT et G. WANNER. *Solving ordinary differential equations. I*. Second. T. 8. Springer Series in Computational Mathematics. Nonstiff problems. Springer-Verlag, Berlin, 1993, p. xvi+528 (cf. p. 24).

[21] E. HAIRER et G. WANNER. *Solving ordinary differential equations. II*. T. 14. Springer Series in Computational Mathematics. Stiff and differential-algebraic problems, Second revised edition, paperback. Springer-Verlag, Berlin, 2010, p. xvi+614 (cf. p. 24).

[22] M. HOCHBRUCK et A. OSTERMANN. „Exponential integrators“. In : *Acta Numer.* 19 (2010), p. 209–286 (cf. p. 28, 30).

[23] Marlis HOCHBRUCK. „A short course on exponential integrators“. In : *Matrix functions and matrix equations*. T. 19. Ser. Contemp. Appl. Math. CAM. Higher Ed. Press, Beijing, 2015, p. 28–49 (cf. p. 30).

[24] Marlis HOCHBRUCK et Alexander OSTERMANN. „Exponential multistep methods of Adams-type“. In : *BIT* 51.4 (2011), p. 889–908 (cf. p. 28, 30).

[25] Alan L HODGKIN et Andrew F HUXLEY. „A quantitative description of membrane current and its application to conduction and excitation in nerve“. In : *The Journal of physiology* 117.4 (1952), p. 500 (cf. p. 16).

[26] Franck JEDRZEJEWSKI. *Introduction aux méthodes numériques*. Springer Science & Business Media, 2006 (cf. p. 26).

[27] James KEENER et James SNEYD. *Mathematical Physiology, Interdisciplinary Applied Mathematics 8*. 1998 (cf. p. 23).

[28] Rikkert H KELDERMANN, Martyn P NASH, Hanneke GELDERBLOM, Vicky Y WANG et Alexander V PANFILOV. „Electro-Mechanical wavebreak in a model of the human left ventricle“. In : *American Journal of Physiology-Heart and Circulatory Physiology* (2010) (cf. p. 23).

[29] John KIM et Parviz MOIN. „Application of a fractional-step method to incompressible Navier-Stokes equations“. In : *Journal of computational physics* 59.2 (1985), p. 308–323 (cf. p. 27).

[30] Pierre-Louis LIONS et Bertrand MERCIER. „Splitting algorithms for the sum of two nonlinear operators“. In : *SIAM Journal on Numerical Analysis* 16.6 (1979), p. 964–979 (cf. p. 27).

[31] Jaakko MALMIVUO et Robert PLONSEY. *Bioelectromagnetism*. `http://www.bem.fi/book/` (cf. p. 7, 8, 20).

[32] Walter T MILLER et David B GESELOWITZ. „Simulation studies of the electrocardiogram. I. The normal heart.“ In : *Circulation Research* 43.2 (1978), p. 301–315 (cf. p. 23).

[33] Gary R MIRAMS, Christopher J ARTHURS, Miguel O BERNABEU et al. „Chaste : an open source C++ library for computational physiology and biology“. In : *PLoS computational biology* 9.3 (2013), e1002970 (cf. p. 27).

[34] Cleve MOLER et Charles VAN LOAN. „Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later“. In : *SIAM review* 45.1 (2003), p. 3–49 (cf. p. 30).

[35] Sundnes J. Artebrant R. Skavhaug O et Tveito A. „A second-order Algorithm for solving Dynamic cell Membrane Equations“. In : *IEEE Transactions On Biomedical Engineering* 26 (oct. 2009), p. 2546–2548 (cf. p. 18, 29, 30).

[36] M. PEREGO et A. VENEZIANI. „An efficient generalization of the Rush-Larsen method for solving electro-physiology membrane equations“. In : *ETNA* 35 (2009), p. 234–256 (cf. p. 18, 29, 30).

[37] Bernardo RUDY et Linda E IVERSON. *Ion channels*. T. 207. Gulf Professional Publishing, 1997 (cf. p. 11).

[38] S RUSH et H LARSEN. „A practical algorithm for solving dynamic membrane equations.“ In : *IEEE Trans Biomed Eng* 25.4 (juil. 1978), p. 389–92 (cf. p. 18, 29).

[39] P SYVERT et NØRSETT. „An A-stable modification of the Adams-Bashforth methods“. In : *Conference on the Numerical Solution of Differential Equations*. T. 109. Lecture Notes in Mathematics. Springer, Berlin, 1969, p. 214–219 (cf. p. 30).

[40] KHWJ TEN TUSSCHER, D NOBLE, PJ NOBLE et Alexander V PANFILOV. „A model for human ventricular tissue“. In : *American Journal of Physiology-Heart and Circulatory Physiology* 286.4 (2004), H1573–H1589 (cf. p. 11, 14–16, 18).

[41] Mayya TOKMAN, John LOFFELD et Paul TRANQUILLI. „New Adaptive Exponential Propagation Iterative Methods of Runge–Kutta Type“. In : *SIAM Journal on Scientific Computing* 34.5 (2012), A2650–A2669 (cf. p. 30).

[42] Leslie TUNG. „A bi-domain model for describing ischemic myocardial dc potentials.“ Thèse de doct. Massachusetts Institute of Technology, 1978 (cf. p. 23).

[43] Richard S VARGA. *Matrix iterative analysis*. T. 27. Springer Science & Business Media, 2009 (cf. p. 26).

[44] Edward J VIGMOND, Matt HUGHES, G PLANK et L Joshua LEON. „Computational tools for modeling electrical activity in cardiac tissue“. In : *Journal of electrocardiology* 36 (2003), p. 69–74 (cf. p. 27).

[45] L. VU THAI et A. OSTERMANN. „Explicit exponential Runge-Kutta methods of high order for parabolic problems“. In : *J. Comput. Appl. Math.* 256 (2014), p. 168–179 (cf. p. 30).

[46] Hao ZHUANG. „Exponential Time Integration for Transient Analysis of Large-Scale Circuits“. In : (2016) (cf. p. 30).

3

Schémas Exponentiel Adams Bashforth intégral et Exponentiel Adams Bashforth.

Dans ce chapitre, nous analyserons le recours aux schémas exponentiels de types Adams Bashforth pour des simulations en électrophysiologie. Cette analyse sera détaillée dans l'article titré *Exponential Adams Bashforth integrators for stiff ODEs, application to cardiac electrophysiology*. Mais avant nous ferons un bref résumé soulignant les idées et résultats importants ayant fait l'objet de cet article.

Les modèles qui décrivent le potentiel de la membrane dû aux différents échanges chimiques entre le milieu intra et extracellulaire sont des EDO très raides et non linéaires appelé modèles ioniques. À cause de cette non-linéarité constituée essentiellement des fonctions exponentielles, l'évaluation à chaque itération des fonctions constituant ces EDO coûtent très cher en temps de calcul. L'objectif est de développer des schémas d'ordre élevés explicites et stables, capables de faire face à la raideur et à la non-linéarité des modèles ioniques. Notre attention s'est portée à cet effet aux schémas exponentiels de type Adams Bashforth. Cette attention a été motivée d'une part par le caractère multi-pas de tels schémas qui nécessitent une seule évaluation des fonctions du modèle à chaque pas de temps. D'autre part, par leur idée de conception qui consiste à utiliser la formule de la variation de la constante après avoir divisé le problème en une partie linéaire et une partie non linéaire. Dans l'usage des schémas exponentiels classiques, la partie linéaire est représentée par la jacobienne de la fonction décrivant le modèle. Ceci induit en général des coûts très élevés lorsqu'on a à faire à des applications complexes, en l'occurrence les modèles ioniques physiologiques. Ce coût supplémentaire peut être réduit significativement en remplaçant la jacobienne de la fonction du modèle par une approximation ou une partie de celle-ci. Les modèles ioniques ont une forme particulière qui permettent d'identifier facilement des termes qui peuvent représenter efficacement les parties linéaires et non-linéaires, sans toutefois être obligé de calculer la jacobienne de la fonction décrivant le modèle.

Nous avons proposé une nouvelle famille de schémas appelé *Integral Exponential Adams Bashforth* (IEAB), dont le schéma *Exponential Adams Bashforth* (EAB) en est un cas particulier. Cette nouvelle famille de schéma est basée sur une idée où la jacobienne de la fonction du modèle est remplacée par une quantité plus générale appelée stabilisateur. Le concept du stabilisateur offre donc une grande flexibilité dans le choix de la partie linéaire du problème. Nous avons dans ce contexte énoncé et prouvé des résultats de convergence pour les schémas IEAB et EAB. Ces résultats ont été confirmés numériquement sur un cas

test à travers le modèle ionique de Beeler et Reuter (BR). Par ce même cas test, nous avons montré que le schéma IEAB est légèrement plus précis que le schéma EAB pendant que le schéma IEAB est plus cher en terme de coût par rapport au schéma EAB. Nous avons fait une étude de la stabilité au sens de Dahlquist pour le schéma EAB. Il en est ressorti que ce schéma est A(0)-stable sur certaines conditions sur le stabilisateur. Par ailleurs, les domaines de stabilités de ce schéma ont été produits et sont assez grands sur certaines condition sur le stabilisateur. Ces domaines donnent une idée très pratique sur l'ordre de grandeur du pas de temps qu'il faut utiliser pour mener à bien une simulation tout en maintenant la stabilité. En effet, si on connaît la valeur propre de la jacobienne de la fonction du modèle la plus grande négativement, on est peut utiliser le domaine de stabilité pour calculer le pas de temps qui permet de maintenir la stabilité du schéma.
La robustesse des schémas IEAB et EAB par rapport à la raideur a été évaluée sur les modèles ioniques BR et TNNP, sachant que le modèle ionique TNNP est environ 15 fois plus raide que le modèle ionique BR, bien que le modèle BR est lui même assez raide. Ceci s'est fait par le calcul des pas de temps critiques. Il a été aussi question de comparer ces pas de temps critiques à ceux de certains schémas classiques (AB, RK, BDF). Il en est ressortit que les schémas IEAB et EAB présentaient à peu près les même pas critiques que les schémas implicites classiques.
Après l'étude des schémas IEAB et EAB nous pouvons conclure que ces schémas sont stables, précis et offrent une alternative pour des simulations réalistes en cardiologie. Cependant, il est à signaler qu'une des limites de ces schémas reste la multitude des étapes et le calcul des fonctions exponentielles nécessaires à leur mise en œuvre. Cet aspect sera largement amélioré dans le chapitre suivant où nous proposerons une méthode d'ordre élevé plus simple à mettre en œuvre.

Exponential Adams Bashforth integrators for stiff ODEs, application to cardiac electrophysiology

Yves Coudière[a,b,c], Charlie Douanla Lontsi[a,c,b], Charles Pierre[d]

[a] *IHU Liryc, Electrophysiology and Heart Modeling, 33600 Pessac, France*
[b] *IMB, UMR 5251 CNRS, Université de Bordeaux, Bordeaux-INP, 33400 Talence, France*
[c] *Carmen team, Inria Bordeaux Sud-Ouest, 33400 Talence, France*
[d] *LMAP, UMR 5142 CNRS, Université de Pau, 64000 Pau, France*

Abstract

Models in cardiac electrophysiology are coupled systems of reaction diffusion PDE and of ODE. The ODE system displays a very stiff behavior. It is non linear and its upgrade at each time step is a preponderant load in the computational cost. The issue is to develop high order explicit and stable methods to cope with this situation.

In this article, is is analyzed the resort to exponential Adams Bashforth (EAB) integrators in cardiac electrophysiology. The method is presented in the framework of a general and varying stabilizer, that is well suited in this context. Stability under perturbation (or 0-stability) is proven. It provides a new approach for the convergence analysis of the method. The Dahlquist stability properties of the method is performed. It is presented in a new framework that incorporates the discrepancy between the stabilizer and the system Jacobian matrix. Provided this discrepancy is small enough, the method is shown to be A(alpha)-stable. This result is interesting for an explicit time-stepping method. Numerical experiments are presented for two classes of stiff models in cardiac electrophysiology. They include performances comparisons with several classical methods. The EAB method is observed to be as stable as implicit solvers and cheaper at equal level of accuracy.

Keywords: stiff equations, explicit high-order multistep methods, exponential integrators of Adams type, stability and convergence, Dahlquist stability
2000 MSC: 65L04, 65L99, 65L06, 65L20

1. Introduction

Computations in cardiac electrophysiology have to face two constraints. Firstly the stiffness due to heterogeneous time and space scales. This is usually dealt with by considering very fine grids. This strategy is associated with large computational costs, still challenging in dimension three. Secondly, the resolution of the reaction terms from the ionic models has an important cost. This resolution occur at each grid node. The total amount of evaluation of the reaction terms has to be maintained as low as possible. For this reason, implicit solvers are usually avoided.

*Inria, 200 Avenue Vieille Tour, 33400 Talence, France
Email address: yves.coudiere@u-bordeaux.fr (Yves Coudière)

Exponential integrators are well adapted to cope with these two constraints. Actually they allow an explicit resolution of the reaction term, and display strong stability properties. In this article, we study and analyze exponential time-stepping methods dedicated to the resolution of reaction equations.

Models for the propagation of the cardiac action potential are evolution reaction diffusion equations coupled with ODE systems. The widely used monodomain model [1, 2, 3] formulates as $\frac{\partial v}{\partial t} = Av + f_1(v, w, x, t)$ and $\frac{\partial w}{\partial t} = f_2(v, w, x, t)$, with space and time variables $x \in \Omega \subset \mathbb{R}^d$ and $t \in \mathbb{R}$. The unknowns are the functions $v(t, x) \in \mathbb{R}$ (the transmembrane voltage) and $w(t, x) \in \mathbb{R}^N$ (a vector that gathers variables describing pointwise the electrophysiological state of the heart cells). In the monodomain model, the diffusion operator is $A(:= \mathrm{div}(g(x)\nabla\cdot))$, and the reaction terms are the nonlinear functions f_1, f_2. These functions model the cellular electrophysiology. They are called ionic models. Ionic models are of complex nature, see e.g. [4, 5, 6, 7]. A special attention has to be paid to the number of evaluations of the functions f_1 and f_2, and implicit solvers are usually avoided. Though we ultimately use an implicit/explicit method to solve the PDE, we need an efficient, fast and robust method to integrate the reaction terms. Therefore, this article focuses on the time integration of the stiff ODE system

$$\frac{dy}{dt} = f(t, y), \quad y(0) = y_0, \tag{1}$$

in the special cases where $f(t, y)$ is an ionic model from cellular electrophysiology. In this case, stiffness is due to the co-existence of fast and slow variables. Fast variables are given in (1) by equations of the form,

$$\frac{dy_i}{dt} = f_i(t, y) = a_i(t, y)y_i + b_i(t, y). \tag{2}$$

Here $a_i(t, y) \in \mathbb{R}$ is provided by the model. This scalar rate of variation will be inserted in the numerical method to stabilize its resolution.

Exponential integrators are a class of explicit methods meanwhile exhibiting strong stability properties. They have motivated many studies along the past 15 years, among which we quote e.g. [8, 9, 10, 11, 12, 13] and refer to [14, 15, 16] for general reviews. They have already been used in cardiac electrophysiology, as e.g. in [17, 18]. Exponential integrators are based on a reformulation of (1) as,

$$\frac{dy}{dt} = a(t, y)y + b(t, y), \quad y(0) = y_0, \tag{3}$$

(with $f = ay + b$) where the linear part $a(t, y)$ is used to stabilize the resolution. Basically $a(t, y)$ is assumed to capture the stiffest modes of the Jacobian matrix of system (1). Stabilization is brought by performing an exact integration of these modes. This exact integration involves the computation of the exponential $\exp(a(t_n, y_n)h)$ at the considered point. This computation is the supplementary cost for exponential integrators as compared to other time stepping methods.

Exponential integrators of Adams type are explicit multistep exponential integrators. They were first introduced by Certaine [19] in 1960 and Nørsett [20] in 1969 for a constant linear part $A = a(t, y)$ in (3). The schemes are derived using a polynomial interpolation of the non linear term $b(t, y)$. It recently received an increasing interest [21, 22, 23] and various convergence analysis have been completed in this particular case [24, 25, 26]. Non constant linear parts have been less studied. Lee and Preiser [27] in 1978 and by Chu [28] in 1983 first suggested to rewrite the equation (1) at each time instant t_n as, rewritten as,

$$\frac{dy}{dt} = a_n y + g_n(t, y), \quad y(t_n) = y_n, \tag{4}$$

with $a_n = a(t_n, y_n)$ and $g_n(t, y) = b(t, y) + (a(t, y) - a_n)y$. In the sequel, a_n is referred to as the *stabilizer*. It is updated after each time step. Recently, Ostermann *et al* [24, 26] analyzed the linearized exponential Adams method, where the stabilizer a_n is set to the Jacobian matrix of $f(t, y)$ in (1). This choice requires the computation of a matrix exponential at every time step. Anyway, when the fast variables of the system are known, stabilization can be performed only on these variables. Considering the full Jacobian as the stabilizer implies unnecessary computational efforts. To avoid these problems, an alternative is to set the stabilizer as a part or as an approximation of the Jacobian. This has been analyzed in [29] and in [30] for exponential Rosenbrock and exponential Runge Kutta type methods respectively. This strategy is well adapted to cardiac electrophysiology, where a diagonal stabilizer associated with the fast variables is directly provided by the model with equation (2). The present contribution is to analyze general varying $a(t, y)$ in (3) for exponential integrators of Adams type, referred to as *exponential Adams Bashforth*, and shortly denoted EAB. Together with the EAB scheme, we introduce a new variant, that we called *integral exponential Adams Bashforth*, denoted I-EAB.

The convergence analysis held in [24] extends to the case of general varying stabilizers. However there is a lack of results concerning the stability in this case: for instance, consider the simpler exponential Euler method, defined by $y_{n+1} = s(t_n, y_n, h) := \mathrm{e}^{a_n h} y_n + h\varphi_1(a_n h)b_n$ with $\varphi_1(z) = (\mathrm{e}^z - 1)/z$. Stability under perturbation (also called 0-stability) can be easily proven provided that the scheme generator $s(t, y, h)$ is globally Lipschitz in y with a constant bounded by $1 + Ch$. Therefore stability under perturbation is classically studied by analyzing the partial derivative $\partial_y s$. This can be done in the case where $a(t, y)$ is either a constant operator or a diagonal varying matrix. In the general case however things turn out to be more complicated. Indeed the general expansion $e^{M+\varepsilon N} \neq e^M + \varepsilon e^M N + O(\varepsilon^2)$ does not hold, unless the two matrices M and N are commuting. *As a consequence differentiating* $\mathrm{e}^{a(t,y)h}$ *in y cannot be done without very restrictive assumptions on* $a(t, y)$. We present here a stability analysis for general varying stabilizers. This will be done by introducing relaxed stability conditions on the scheme generator $s(t, y, h)$. Together with a consistency analysis, it provides a new proof for the convergence of the EAB schemes, in the spirit of [24].

Stability under perturbation provides results of qualitative nature. In addition, the Dahlquist stability analysis strengthens these results. It is a practical tool that allows to dimension the time step h with respect to the variations of $f(t, y)$ in equation (1). The analysis is made by setting $f(t, y) = \lambda y$ in (1). For exponential integrators with general varying stabilizer, the analysis must incorporate the decomposition of $f(t, y) = \lambda y$ used in (3). The stability domain of the considered method will depend on the relationship between λ and $a(t, y)$, following a concept first introduced in [17]. We numerically establish that EAB methods are $A(\alpha)$ stable provided that the stabilizer is sufficiently close to the system Jacobian matrix (precise definitions are in section 5). Moreover the angle α approaches $\pi/2$ when the stabilizer goes to the system Jacobian matrix. In contrast, there exists no $A(0)$ stable explicit linear multistep method (see [31, chapter V.2]. This property is remarkable for explicit methods.

Numerical experiments for the EAB and I-EAB scheme are provided in section 6, in the context of cardiac electrophysiology. Robustness to stiffness is studied with this choice. It is numerically shown to be comparable to implicit methods both in terms of accuracy and of stability condition on the time step. We conclude that EAB methods are well suited for solving stiff differential problems. In particular they allow computations at large time step with good accuracy properties and cheap cost.

The article is organized as follows. The EAB and I-EAB methods are introduced in section 2. The general stability and convergence results are stated and proved in section 3. The EAB and

I-EAB stability under perturbation and convergence are proved in section 4. The Dahlquist stability is investigated in section 5, and the numerical experiments end the article, in section 6.

In all this paper, $h > 0$ is a constant time-step and $t_n = nh$ are the time instants associated with the numerical approximate y_n of the solution of the ODE (1).

2. Scheme definitions

2.1. The EAB$_k$ method

The exact solution at time t_{n+1} to the equation (4) (with $a_n = a(t_n, y_n)$) is given by the variation of the constants formula

$$y(t_{n+1}) = e^{a_n h}\left(y(t_n) + \int_0^h e^{-a_n \tau} g_n\left(t_n + \tau, y(t_n + \tau)\right) d\tau\right). \tag{5}$$

Using the k approximations $y_{n-j} \simeq y(t_{n-j})$ for $j = 0 \dots k-1$, we build the Lagrange polynomial $\tilde{g}_n$ of degree at most $k-1$ that satisfies,

$$\tilde{g}_n(t_{n-j}) = g_{nj} := g(t_{n-j}, y_{n-j}), \quad 0 \le j \le k-1. \tag{6}$$

It provides the numerical approximation $y_{n+1} \simeq y(t_{n+1})$ as

$$y_{n+1} = \mathrm{e}^{a_n h}\left(y_n + \int_0^h \mathrm{e}^{-a_n \tau}\, \tilde{g}_n(t_n + \tau) d\tau\right). \tag{7}$$

The Taylor expansion of the polynomial $\tilde{g}_n$ is $\tilde{g}_n(t_n + \tau) = \sum_{j=1}^{k} \frac{\gamma_{nj}}{(j-1)!} (\tau/h)^{j-1}$, where the coefficients γ_{nj} are uniquely determined by (6), and actually given in table 1 for $k = 1, 2, 3, 4$. An exact integration of the integral in equation (7) may be performed:

$$y_{n+1} = \mathrm{e}^{a_n h}\, y_n + h \sum_{j=1}^{k} \varphi_j(a_n h) \gamma_{nj}, \tag{8}$$

where the functions φ_j, originally introduced in [20], are recursively defined (for $j \ge 0$) by,

$$\varphi_0(z) = \mathrm{e}^z, \quad \varphi_{j+1}(z) = \frac{\varphi_j(z) - \varphi_j(0)}{z} \quad \text{and} \quad \varphi_j(0) = \frac{1}{j!}. \tag{9}$$

The equation (8) defines the Exponential Adams Bashforth method of order k, denoted by EAB$_k$.

Table 1: Coefficients γ_{nj} for the EAB$_k$ schemes

k	1	2	3	4
γ_{n1}	g_n	g_n	g_n	g_n
γ_{n2}		$g_n - g_{n-1}$	$\frac{3}{2}g_n - 2g_{n-1} + \frac{1}{2}g_{n-2}$	$\frac{11}{6}g_n - 3g_{n-1} + \frac{3}{2}g_{n-2} - \frac{1}{3}g_{n-3}$
γ_{n3}			$g_n - 2g_{n-1} + g_{n-2}$	$2g_n - 5g_{n-1} + 4g_{n-2} - g_{n-3}$
γ_{n4}				$g_n - 3g_{n-1} + 3g_{n-2} - g_{n-3}$

Remark 1. When $a(t, y) = \mathrm{diag}(d_i)$ is a diagonal matrix, $\varphi_k(a_n h) = \mathrm{diag}(\varphi_k(d_i))$ can be computed component-wise. Its computation is straightforward.

Remark 2. With the definition (9), the functions φ_k are analytic on the whole complex plane. Therefore the EAB_k scheme definition (8) makes sense for a matrix term $a(t, y)$ in equation (3) without particular assumption.

Remark 3. The computation of y_{n+1} in the formula (8) requires the computation of $\varphi_j(a_n h)$ for $j = 0, \ldots, k$. This computational effort can be reduced with the recursive definition (9). In practice only $\varphi_0(a_n h)$ needs to be computed. This is detailed in section 6.1.

2.2. A variant: the I-EAB$_k$ method

If the matrix $a(t, y)$ is diagonal, we can take advantage of the following version for the variation of the constants formula

$$y(t_{n+1}) = \mathrm{e}^{A_n(h)} \left(y(t_n) + \int_0^h \mathrm{e}^{-A_n(\tau)}\, b(y(t_n + \tau), t_n + \tau) d\tau \right),$$

where $A_n(\tau) = \int_0^\tau a(t_n+\sigma, y(t_n+\sigma))d\sigma$. An attempt to improve the EAB_k formula (8) is to replace $a(t, y)$ and $b(t, y)$ in the integral above by their Lagrange interpolation polynomials. At time t_n, we define the two polynomials $\tilde{a}_n$ and $\tilde{b}_n$ of degree at most $k - 1$ so that $\tilde{a}_n(t_{n-j}) = a(t_{n-j}, y_{n-j})$, and $\tilde{b}_n(t_{n-j}) = b(t_{n-j}, y_{n-j})$, for $j = 0 \ldots k-1$, and the primitive $\tilde{A}_n(\tau) = \int_0^\tau \tilde{a}(t_n + \sigma, y(t_n + \sigma))d\sigma$. The resulting approximate solution at time t_{n+1} is finally given by the formula

$$y_{n+1} = \mathrm{e}^{\tilde{A}_n(h)} \left(y_n + \int_0^h \mathrm{e}^{-\tilde{A}_n(\tau)}\, \tilde{b}_n(t_n + \tau) d\tau \right). \tag{10}$$

The method is denoted I-EAB_k, for integral EAB_k. Dislike for the formula (7), no exact integration formula is available, because of the term $\mathrm{e}^{-\tilde{A}_n(\tau)}$. A quadrature rule is required for the actual numerical computation of the integral in formula (10). Implementation details are given in section 6.1.

3. Stability conditions and convergence

The equation (1) is considered on a finite dimensional vector space E with norm $|\cdot|_E$. We fix a final time $T > 0$ and assume that equation (1) has a solution y on $[0, T]$. We adopt the general settings for the analysis of k-multistep methods following [32]. The space E^k is equipped with the maximum norm $|Y|_\infty = \max_{1 \le i \le k} |y_i|_E$ with $Y = (y_1, \ldots, y_k) \in E^k$. A k-multistep scheme is defined by a mapping $s : (t, Y, h) \in \mathbb{R} \times E^k \times \mathbb{R}^+ \mapsto s(t, Y, h) \in E$. For instance, the EAB_k scheme rewrites with $Y = (y_{n-k+1}, \ldots, y_n)$ in the formula (8), and $s(t_n, Y, h) = y_{n+1}$. The scheme generator is the mapping S given by $S : (t, Y, h) \in \mathbb{R} \times E^k \times \mathbb{R}^+ \mapsto (y_2, \ldots y_k s(t, Y, h)) \in E^k$. A numerical solution is a sequence (Y_n) in E^k for $n \ge k - 1$ so that

$$Y_{n+1} = S(t_n, Y_n, h) \quad \text{for} \quad n \ge k - 1, \tag{11}$$

and $Y_{k-1} = (y_0, \ldots, y_{k-1})$ is a given initial data. A perturbed numerical solution is a sequence (Z_n) in E^k for $n \ge k - 1$ such that,

$$Z_{k-1} = Y_{k-1} + \xi_{k-1}, \quad Z_{n+1} = S(t_n, Z_n, h) + \xi_{n+1} \quad \text{for} \quad n \ge k - 1, \tag{12}$$

with $(\xi_n) \in E^k$ for $n \geq k-1$. The scheme is said to be stable under perturbation (or 0-stable) if, for any numerical solution (Y_n) as in (11), there exists a (stability) constant $L_s > 0$ such that, for any perturbation (Z_n) as defined in (12), we have,

$$\max_{k-1 \leq n \leq T/h} |Y_n - Z_n|_\infty \leq L_s \sum_{k-1 \leq n \leq T/h} |\xi_n|_\infty. \tag{13}$$

Proposition 1. *Assume that there exists constants $C_1 > 0$ and $C_2 > 0$ such that,*

$$1 + |S(t, Y, h)|_\infty \leq (1 + |Y|_\infty)(1 + C_1 h), \tag{14}$$

$$|S(t, Y, h) - S(t, Z, h)|_\infty \leq |Y - Z|_\infty (1 + C_2 h(1 + |Y|_\infty)), \tag{15}$$

for $0 \leq t \leq T$, and for $Y, Z \in E^k$. Then, the numerical scheme is stable under perturbation with the constant L_s in (13) *given by,*

$$L_s = \mathrm{e}^{C^\star T}, \quad C^\star := C_2 \, \mathrm{e}^{C_1 T} (1 + |Y_{k-1}|_\infty). \tag{16}$$

Proof. Consider a numerical solution (Y_n) in (11). A recursion on condition (14) gives,

$$1 + |Y_n|_\infty \leq (1 + |Y_{k-1}|_\infty)(1 + C_1 h)^{n-k+1} \leq \mathrm{e}^{C_1 T} (1 + |Y_{k-1}|_\infty),$$

since $(1 + x)^p \leq \mathrm{e}^{px}$ (for $x \geq 0$), and $(n - k + 1)h \leq nh \leq T$. Now, consider a perturbation (Z_n) of (Y_n) given by (12). Using the condition (15) together with the previous inequality,

$$\begin{aligned} |Y_{n+1} - Z_{n+1}|_\infty &\leq |S(t_n, Y_n, h) - S(t_n, Z_n, h)|_\infty + |\xi_{n+1}|_\infty \leq |Y_n - Z_n|_\infty (1 + C_2 h(1 + |Y_n|_\infty)) + |\xi_{n+1}|_\infty \\ &\leq |Y_n - Z_n|_\infty \left(1 + C_2 \, \mathrm{e}^{C_1 T} (1 + |Y_{k-1}|_\infty) h\right) + |\xi_{n+1}|_\infty \leq |Y_n - Z_n|_\infty \left(1 + C^\star h\right) + |\xi_{n+1}|_\infty, \end{aligned}$$

where $C^\star := C_2 \, \mathrm{e}^{C_1 T} (1 + |Y_{k-1}|_\infty)$. By recursion we get,

$$\begin{aligned} |Y_n - Z_n|_\infty &\leq \left(1 + C^\star h\right)^{n-k+1} |Y_{k-1} - Z_{k-1}|_\infty + \sum_{i=0}^{n-k} \left(1 + C^\star h\right)^i |\xi_{n-i}|_\infty \\ &\leq \left(1 + C^\star h\right)^n \sum_{i=k-1}^{n} |\xi_i|_\infty \leq \mathrm{e}^{c^\star T} \sum_{i=k-1}^{n} |\xi_i|_\infty, \end{aligned}$$

which ends the proof. □

Like in the classical cases, stability under perturbation together with consistency ensures convergence. Let us specify this point. For the considered solution $y(t)$ of problem (1) on $[0, T]$, we define,

$$Y(t) = (y(t - (k-1)h), \ldots y(t)) \in E^k \quad \text{for } 0 \leq (k-1)h \leq t \leq T. \tag{17}$$

The local error at time t_n is,

$$\varepsilon(t_n, h) = Y(t_{n+1}) - S(t_n, Y(t_n), h). \tag{18}$$

The scheme is said to be consistent of order p if there exists a (consistency) constant $L_c > 0$ only depending on $y(t)$ such that, $\max_{k-1 \leq n \leq T/h} |\varepsilon(t_n, h)|_\infty \leq L_c h^{p+1}$.

Corollary 1. *If the scheme satisfies the stability conditions* (14) *and* (15)*, and is consistent of order p, then a numerical solution* (Y_n) *given by* (11) *satisfies,*

$$\max_{k-1\le n\le T/h} |Y(t_n) - Y_n|_\infty \le L_s L_c T h^p + L_s |\xi_0|_\infty, \tag{19}$$

where $\xi_0 = Y(t_{k-1}) - Y_{k-1}$ *denotes the error on the initial data, and the constant* L_s *is as in equation* (16)*.*

Remark 4. Note that the stability constant L_s in (16) depends on $|Y_{k-1}|_\infty$, and then on h. This is not a problem since L_s can be bounded uniformly as $h \to 0$ for Y_{k-1} in a neighborhood of y_0.

Proof. We have $Y(t_{k-1}) = Y_{k-1} + \xi_0$ and $Y(t_{n+1}) = S(t_n, Y(t_n), h) + \varepsilon(t_n, h)$. Therefore the sequence $(Y(t_n))$ is a perturbation of the numerical solution (Y_n) in the sense of (12). As a consequence, proposition 1 shows that

$$\max_{k-1\le n\le T/h} |Y_n - Y(t_n)|_E \le L_s \left(|\xi_0| + \sum_{k\le n\le T/h} |\varepsilon(t_n, h)| \right) \le L_s |\xi_0| + L_s L_c \left(\sum_{k\le n\le T/h} h \right) h^p,$$

and the convergence result follows. □

4. EAB$_k$ and I-EAB$_k$ scheme analysis

The space E is assumed to be $E = \mathbb{R}^N$ with its canonical basis and with $|\cdot|_E$ the maximum norm. The space of operators on E is equipped with the associated operator norm, and associated to $N\times N$ matrices. Thus $a(t, y)$ is a $N\times N$ matrix and its norm $|a(t, y)|$ is the matrix norm associated to the maximum norm on $\mathbb{R}^N$.

It is commonly assumed for the numerical analysis of ODE solvers that f in the equation (1) is uniformly Lipschitz in its second component y. With the formulation (3), the following assumptions will be needed: on $\mathbb{R} \times E$,

$$|a(t, y)| \le M_a, \quad a(t, y),\ b(t, y) \text{ and } f(t, y) \text{ uniformly Lipschitz in } y. \tag{20}$$

We denote by K_f, K_a and K_b the Lipschitz constant for f, a and b respectively.

Theorem 1. *With the assumptions* (20)*, the EAB$_k$ and I-EAB$_k$ schemes are stable under perturbations. Moreover, if a and b are C^k regular on $\mathbb{R} \times E$, then the EAB$_k$ and I-EAB$_k$ schemes are consistent of order k. Therefore they converge with order k in the sense of inequality* (19)*, by applying corollary 1.*

The stability and consistency are proved in sections 4.3 and 4.4, respectively. Preliminary tools and definitions are provided in the sections 4.1 and 4.2.

4.1. Interpolation results

Consider a function $x : \mathbb{R} \times E \to \mathbb{R}$ and a triplet $(t, Y, h) \in \mathbb{R} \times E^k \times \mathbb{R}^+$ with $Y = (y_1, \dots, y_k)$. We set to $\tilde{x}_{[t,Y,h]}$ the polynomial with degree at most $k - 1$ so that

$$\tilde{x}_{[t,Y,h]}(t - ih) = x(t - ih, y_{k-i}), \quad 0 \le i \le k - 1.$$

We then extend component-wise this definition to vector valued or matrix valued functions x (e.g. the functions a or b).

Lemma 1. *There exists an (interpolation) constant $L_i > 0$ such that, for any function $x : \mathbb{R}\times E \mapsto \mathbb{R}$, and for any vectors $Y, Z \in E^k$,*

$$\sup_{t\le\tau\le t+h} \left|\tilde{x}_{[t,Y,h]}(\tau)\right| \le L_i \max_{0\le i\le k-1} |x(t-ih, y_{k-i})|\,, \tag{21}$$

$$\sup_{t\le\tau\le t+h} \left|\tilde{x}_{[t,Y,h]}(\tau) - \tilde{x}_{[t,Z,h]}(\tau)\right| \le L_i \max_{0\le i\le k-1} |x(t-ih, y_{k-i}) - x(t-ih, z_{k-i})|\,. \tag{22}$$

Consider a function $y : [0,T] \to E$ and assume that x and y have a C^k regularity. Then, when $[t-(k-1)h, t+h] \subset [0,T]$,

$$\sup_{t\le\tau\le t+h} \left|x(\tau, y(\tau)) - \tilde{x}_{[t,Y(t),h]}(\tau)\right|_E \le \sup_{[0,T]} \left|\frac{d^k}{dt^k}\left(f(t, y(t))\right)\right| h^k, \tag{23}$$

with $Y(t)$ defined in (17)*.*

For a vector valued function in $\mathbb{R}^d$ the previous inequalities hold when considering the max norm on $\mathbb{R}^d$. For a matrix valued function in $\mathbb{R}^d \times \mathbb{R}^d$ this is also true for the operator norm on $\mathbb{R}^d \times \mathbb{R}^d$ when multiplying the constants in the inequalities (21), (22) and (23) by d.

Proof. The space $\mathbb{P}_{k-1}$ of the polynomials p with degree at most $k-1$ is equipped with the norm $\sup_{[0,1]} |p(\tau)|$. We associate to the $R = (r_1, \dots, r_k) \in \mathbb{R}^k$ its Lagrange interpolation polynomial $\mathcal{L}R \in \mathbb{P}_{k-1}$, uniquely determined by $\mathcal{L}R(-i) = r_{k-1}$ for $i = 0 \dots k-1$. The mapping $\mathcal{L}$ is linear. Let $C_{\mathcal{L}}$ be its continuity constant (it only depends on k).

We fix the function $x : \mathbb{R}\times E \to \mathbb{R}$ and $(t,h) \in \mathbb{R}\times\mathbb{R}^+$. Consider the vector $Y = (y_1, \dots, y_k) \in E^k$ and define the vector $R = (x(t-(k-1)h, y_1), \dots, x(t, y_k)) \in \mathbb{R}^k$. We have $\tilde{x}_{[t,Y,h]}(t+\tau) = \mathcal{L}R(\tau/h)$. The relation (21) is exactly the continuity of $\mathcal{L}$ and $L_i = C_{\mathcal{L}}$.

Consider $Y_1, Y_2 \in E^k$ and the associated vectors R_1, R_2 as above. We have $(x_{[t,Y_1,h]} - x_{[t,Y_2,h]})(t+\tau) = \mathcal{L}(R_1 - R_2)(\tau/h)$. Again, relation (22) is derived from the continuity of $\mathcal{L}$.

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a C^k function, its interpolation polynomial $\tilde{\varphi}$ at the points $t-(k-1)h, \dots, t$ is considered. A classical result on Lagrange interpolation applied to φ states that, for all $\tau \in (t, t+h)$, there exists $\xi \in (t-(k-1)h, t+h)$, such that $(\varphi - \tilde{\varphi})(\tau) = \frac{1}{k!}\varphi^{(k)}(\xi)\pi(\tau)$, where $\pi(\tau) = \prod_{i=1}^k (\tau - t_i)$. For $\tau \in (t, t+h)$, we have $|\pi(\tau)| \le k!\, h^k$. This proves (23) by setting $\varphi(t) = x(t, y(t))$.

For a vector valued function $x : \mathbb{R}\times E \to \mathbb{R}^d$, these three inequalities holds by processing component-wise and when considering the max norm on $\mathbb{R}^d$.

For a matrix valued function $x : \mathbb{R}\times E \to \mathbb{R}^d \times \mathbb{R}^d$, the extension is direct when considering the max norm $|\cdot|_\infty$ on $\mathbb{R}^d \times \mathbb{R}^d$ (*i.e.* the max norm on the matrix entries). The operator norm $|\cdot|$ is retrieved with the inequality $|\cdot|_\infty \le d|\cdot|$ □

4.2. Scheme generators

Let us consider $(t, Y, h) \in \mathbb{R} \times E^k \times \mathbb{R}^+$ with $Y = (y_1, \dots, y_k)$. With the notations used in the previous subsection, we introduce the interpolations $\tilde{a}_{[t,Y,h]}$ and $\tilde{b}_{[t,Y,h]}$ for the functions a and b (in (3)). Thanks to its definition (10), the I-EAB$_k$ scheme generator is defined by,

$$s(t, Y, h) = z(t+h) \quad \text{with } \frac{dz}{d\tau} = \tilde{a}_{[t,Y,h]}(\tau)z(\tau) + \tilde{b}_{[t,Y,h]}(\tau), \quad z(t) = y_k. \tag{24}$$

We introduce the polynomial $\bar{g}_{[t,Y,h]}$ with degree at most $k-1$ that satisfies,

$$\bar{g}_{[t,Y,h]}(t-ih) = f(t-ih, y_{k-i}) - a(t, y_k)y_{k-i}, \quad i = 0 \dots k-1.$$

The function $\tilde{g}_n$ in (6) is given by $\tilde{g}_n = \bar{g}_{[t_n,Y_n,h]}$ with $Y_n = (y_{n-k+1} \dots, y_n)$. With the definition (7), the EAB_k scheme generator is defined by,

$$s(t, Y, h) = z(t + h) \quad \text{with } \frac{dz}{d\tau} = a(t, y_k)z(\tau) + \bar{g}_{[t,Y,h]}(\tau), \quad z(t) = y_k. \tag{25}$$

We will use the fact that $\bar{g}_{[t,Y,h]}$ is the Lagrange interpolation polynomial of the function $g_{t,y_k} : (\tau, \xi) \mapsto f(\tau, \xi) - a(t, y_k)\xi$.

These scheme generator definitions will allow us to use the following Gronwall's inequality (see [33, Lemma 196, p.150]).

Lemma 2. *Suppose that $z(t)$ is a C^1 function on E. If there exist $\alpha > 0$ and $\beta > 0$ such that $|z'(t)|_E \leq \alpha t + \beta$ for all $t \in [t_0, t_0 + h]$, then:*

$$|z(t)|_E \leq |z(t_0)|_E \, \mathrm{e}^{\alpha h} + \beta h \, \mathrm{e}^{\alpha h} \quad \text{for } t \in [t_0, t_0 + h]. \tag{26}$$

4.3. Stability

According to proposition 1, we have to prove the stability conditions (14) and (15). It is sufficient to prove these relations for $h \leq h_0$ for some constant $h_0 > 0$ since the limit $h \to 0$ is of interest here.

4.3.1. Case of the I-EAB$_k$ scheme

Consider $(t, h) \in \mathbb{R} \times \mathbb{R}^+$ and $Y = (y_1, \dots, y_k) \in E^k$. We simply denote $\tilde{a} = \tilde{a}_{[t,Y,h]}$ and $\tilde{b} = \tilde{b}_{[t,Y,h]}$. The scheme generator is given by (24). We first have to bound $z(t + h)$ where z is given by, $z' = \tilde{a}z + \tilde{b}$, and $z(t) = y_k$. Firstly, with the interpolation bound (21), $\sup_{t\leq\tau\leq t+h} |\tilde{a}(\tau)| \leq L_i \max_{0\leq i\leq k-1} |a(t - ih, y_{k-1})| \leq L_i M_a := \alpha$. Secondly, the function $b(t, y)$ is globally Lipschitz in y and thus can be bounded by $|b(t, y)|_E \leq |b(t, 0)|_E + K_b|y|_E \leq R_b(|y|_E + 1)$, for $0 \leq t \leq T$ and for some constant R_b only depending on K_b and on T. Then with the bound (21), $\sup_{t\leq\tau\leq t+h} |\tilde{b}(\tau)|_E \leq L_i \max_{0\leq i\leq k-1} R_b \, (|y_{k-i}|_E + 1) \leq L_i R_b(|Y|_\infty + 1) := \beta$. By applying the Gronwall inequality (26) with these α and β, for $0 \leq \tau \leq h$, $|z(t + \tau)|_E \leq \mathrm{e}^{L_i M_a h} \, (|y_k|_E + hL_i R_b(|Y|_\infty + 1))$. Thus, there exists a constant C_1 only depending on L_i, M_a, R_b and h_0 such that, for $0 \leq \tau \leq h$ and $0 \leq h \leq h_0$,

$$|z(t + \tau)|_E \leq C_1 h + |Y|_\infty(1 + C_1 h). \tag{27}$$

This gives the condition (14), by taking $\tau = h$.

For j=1, 2 We consider $Y_j = (y_{j,1}, \dots y_{j,k}) \in E^k$ and denote $\tilde{a}_j = \tilde{a}_{[t,Y_j,h]}$ and $\tilde{b}_j = \tilde{b}_{[t,Y_j,h]}$ the interpolations of the functions a and b. With the definition (24) of the I-EAB$_k$scheme, we have $|s(t, Y_1, h) - s(t, Y_2, h)|_E = |\delta(t + h)|$ with $\delta = z_1 - z_2$ and with z_j given by $z_j' = \tilde{a}_j z_j + \tilde{b}_j$, and $z_j(t) = y_{j,k}$. We have then $\delta' = \tilde{a}_1\delta + r$, and $r := (\tilde{a}_1 - \tilde{a}_2)z_2 + (\tilde{b}_1 - \tilde{b}_2)$. Using that a and b are Lipschitz in y and with the interpolation bound (22), $\sup_{t\leq\tau\leq t+h} |\tilde{b}_1(\tau) - \tilde{b}_2(\tau)|_E \leq L_i K_b|Y_1 - Y_2|_\infty$, and $\sup_{t\leq\tau\leq t+h} |\tilde{a}_1(\tau) - \tilde{a}_2(\tau)| \leq L_i K_a|Y_1 - Y_2|_\infty$. With the upper bound (27), for $t \leq \tau \leq t + h \leq T$ and for $h \leq h_0$,

$$|r(\tau)|_E \leq L_i|Y_1 - Y_2|_\infty \, (K_b + K_a \, (C_1 h + |Y_2|_\infty(1 + C_1 h))) \leq C|Y_1 - Y_2|_\infty \, (1 + |Y_2|_\infty) \, . \tag{28}$$

For a constant C only depending on h_0, K_a, K_b, L_i and C_1. We finally apply the Gronwall inequality (26). It yields $|\delta(t + h)| \leq \mathrm{e}^{L_i M_a h} \, (|y_{1,k} - y_{2,k}|_E + Ch|Y_1 - Y_2|_\infty \, (1 + |Y_2|_\infty)) \leq |Y_1 - Y_2|_\infty \, \mathrm{e}^{L_i M_a h} \, (1 + Ch \, (1 + |Y_2|_\infty))$, This implies the second stability condition (15) for $h \leq h_0$.

4.3.2. Case of the EAB$_k$ scheme

Consider $(t,h) \in \mathbb{R}\times\mathbb{R}^+$ and a vector $Y = (y_1,\dots,y_k) \in E^k$. Following the definition of the EAB$_k$ scheme given in section 2.1, we denote $\bar{a} = a(t,y_k)$, g the function $g(\tau,\xi) = b(\tau,\xi) + (a(\tau,\xi)-\bar{a})\,\xi = f(\tau,\xi) - \bar{a}\xi$ and $\bar{g} = \bar{g}_{[t,Y,h]}$. We have that $\bar{g}$ is the Lagrange interpolation polynomial of g, specifically $\bar{g} = \tilde{g}_{[t,Y,h]}$. The scheme generator is then given by the equation (25): $s(t,Y,h) = z(t+h)$ with $z' = \bar{a}z+\bar{g}$, and $z(t) = y_k$. We first have the bound $|\bar{a}| \le M_a$. As in the previous subsection, f being globally Lipschitz in y, one can find a constant R_f so that for $0 \le t \le T$, $|f(t,y)|_E \le R_f(1+|y|_E)$. It follows that $|g(\tau,\xi)|_E \le R_f(|y|_E+1) + M_a|y|_E \le C_0/L_i(|y|_E+1)$, with $C_0/L_i = R_f + M_a$. Therefore, with the interpolation bound (21), $\sup_{t\le\tau\le t+h}|\bar{g}(\tau)|_E \le L_i \max_{0\le i\le k-1}|g(t-ih,y_{k-i})|_E \le C_0(|y|_E+1)$. By applying the Gronwall inequality (26), for $0\le\tau\le h$, $|z(t+\tau)|_E \le \mathrm{e}^{M_a h}(|y_k|_E + hC(|Y|_\infty+1))$. Thus, there exists a constant C_1 only depending on M_a and C_0 such that, for $0\le\tau\le h$ and $0\le h\le h_0$, the bound (27) holds. This gives the condition (14).

We now consider Y_1, $Y_2 \in E^k$ for $j = 1,2$, and denote as previously, $\bar{a}_j = a(t,y_{j,k})$, g_j the function $g_j(\tau,\xi) = f(\tau,\xi) - \bar{a}_j\xi$ and $\bar{g}_j = \bar{g}_{[t,Y_j,h]}$. With (25), $|s(t,Y_1,h) - s(t,Y_2,h)|_E = |\delta(t+h)|$ with $\delta = z_1 - z_2$ and with z_j given by, $z_j' = \bar{a}_j z_j + \bar{g}_j$, and $z_j(t) = y_{j,k}$. The function δ satisfies the ODE, $\delta' = \bar{a}_1\delta + r(t)$, with $r(t) := (\bar{a}_1-\bar{a}_2)z_2 + (\bar{g}_1-\bar{g}_2)$.

Now, we have, $|g_1(\tau,y_{1,i}) - g_2(\tau,y_{2,i})|_E \le |f(\tau,y_{1,i}) - f(\tau,y_{2,i})|_E + |\bar{a}_1||y_{1,i}-y_{2,i}|_E + |\bar{a}_1-\bar{a}_2||y_{2,i}|_E \le |Y_1-Y_2|_\infty\left(K_f + M_a + K_a|Y_2|_\infty\right)$. Thus, with the bound (22), for some $C>0$, $\sup_{t\le\tau\le t+h}|\bar{g}_1(\tau) - \bar{g}_2(\tau)|_E \le L_i\max_{0\le i\le k-1}|g_1(t-ih,y_{1,k-i}) - g_2(t-ih,y_{2,k-i})|_E \le C|Y_1-Y_2|_\infty(1+|Y_2|_\infty)$.

Meanwhile we have the upper bound (27) that gives, for $t\le\tau\le t+h\le T$ and $h\le h_0$, $|(\bar{a}_1-\bar{a}_2)z_2|_E\, M_a\,|Y_1-Y_2|_\infty\,|z_2(\tau)|_E \le M_a\,|Y_1-Y_2|_\infty\,(C_1h + |Y_2|_\infty(1+C_1h))$.

Altogether, we retrieve the upper bound (28) on $r(t)$. We can end the proof as for the I-EAB$_k$ case and conclude that the stability condition (15) holds for the EAB$_k$ scheme.

4.4. Consistency

Consider a solution $y \in C^1([0,T])$ to the problem (1). The functions a and b in (3) are assumed to be C^k regular so that y is C^{k+1} regular.

4.4.1. Case of the EAB$_k$ scheme

The local error (18) for the EAB$_k$ scheme has been analyzed in [24]. The analysis remains valid for the case presented here and we only briefly recall it. The local error is obtained by subtracting (7) to (5).

$$|\varepsilon(t_n,h)|_E \le \int_0^h \mathrm{e}^{M_a(h-\tau)}\,|g_n(t+\tau,y(t+\tau)) - \tilde{g}_n(t+\tau)|_E\,d\tau \le h\varphi_1(M_ah)h^k \sup_{[0,T]}\left|\frac{d^k}{dt^k}\left(g_n(t,y(t))\right)\right|,$$

thanks to the interpolation error estimate (23). Finally, with the upper bound M_a on a_n, the last term can be bounded independently of n, for $h\le h_0$.

4.4.2. Case of the I-EAB$_k$ scheme

We denote $\tilde{a} = \tilde{a}_{[t_n,Y(t_n),h]}$ and $\tilde{b} = \tilde{b}_{[t_n,Y(t_n),h]}$. The local error (18) for the I-EAB$_k$ scheme satisfies $\varepsilon(t_n,h) = |\delta(t_{n+1})|_E$ with $\delta = y-z$ and where z is defined by, $z' = \tilde{a}z+\tilde{b}$, and $z(t_n) = y(t_n)$, so that with (24) we have $s(t_n,Y(t_n),h) = z(t_{n+1})$. The function δ is defined with $\delta(t_n) = 0$ and $\delta' = \tilde{a}\delta + r$, with $r(\tau) := (a(\tau,y(\tau)) - \tilde{a}(\tau))y(\tau) + (b(\tau,y(\tau)) - \tilde{b}(\tau))$. The following constants only depend on the considered exact solution y, on the functions a and b in problem (3) and on T, $C_y = \sup_{[0,T]}|y|_E$, $C_{a,y} = \sup_{[0,T]}\left|\frac{d^k}{dt^k}a(t,y(t))\right|$,and $C_{b,y} = \sup_{[0,T]}\left|\frac{d^k}{dt^k}b(t,y(t))\right|$.

With the interpolation bound (23), $|r(\tau)|_E \leq Ch^k$ on $[t_n, t_{n+1}]$ with $C = C_{a,y}C_y + C_{b,y}$. It has already been showed in section 4.3 that $\sup_{[t_n,t_{n+1}]} |\tilde{a}(\tau)| \leq L_i M_a$. Therefore, with the Gronwall inequality (26), $\varepsilon(t_n, h) = |\delta(t_{n+1})|_E \leq \mathrm{e}^{L_i M_a h}\, hCh^K$. Thus the EAB_k scheme is consistent of order k.

5. Dahlquist stability

5.1. Background

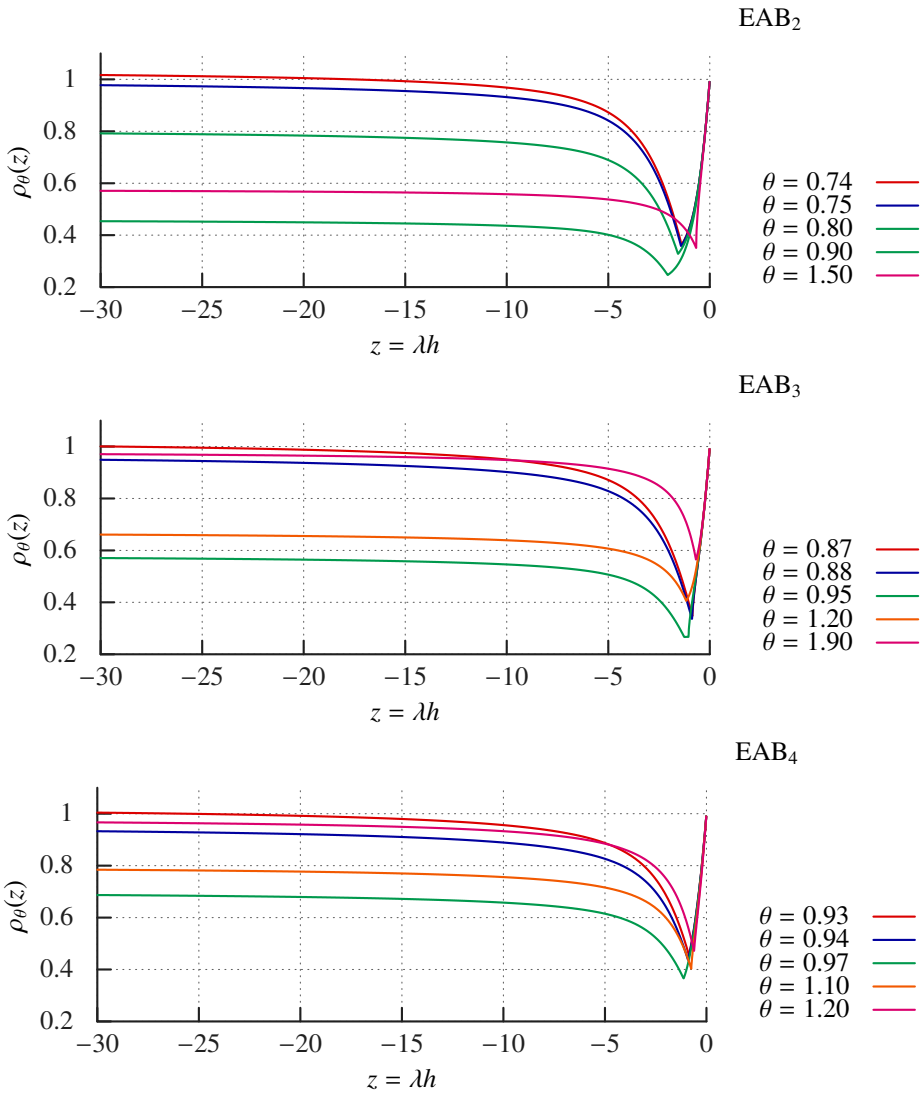


Figure 1: Stability function $\rho_\theta(z)$ for $z \in \mathbb{R}^-$, for various values of θ and for the three schemes EAB_2, EAB_3 and EAB_4.

The classical framework for the Dahlquist stability analysis is to set $f(t, y)$ in problem (1) to $f = \lambda y$. For linear multistep methods, see e.g. [31], the numerical solutions satisfy $|y_{n+1}/y_n| \leq \rho(\lambda h)$, where $\rho : \mathbb{C} \to \mathbb{R}^+$. The function ρ is the stability function. It is defined point wise by the maximum root modulus of a family of polynomial depending on $z = \lambda h$. The stability domain is defined by $D = \{z \in \mathbb{C}, \quad \rho(z) < 1\}$. The scheme is said to be:

- A stable if $\mathbb{C}^- \subset D$,
- $A(\alpha)$ stable if D contains the cone with axis $\mathbb{R}^-$ and with half angle α,
- $A(0)$ stable if $\mathbb{R}^- \subset D$,
- stiff stable if D contains a half plane $\mathrm{Re}\, z < x \in \mathbb{R}^-$.

For exponential integrators, when setting $a(t, y) = \lambda$ in the reformulation (3) of problem (1), the scheme is exact, and therefore also A stable. Such an equality does not hold in general. Then for exponential integrators the Dahlquist stability analysis has to incorporate the relationship between the stabilization term $a(t, y)$ in (3) and the test function $f = \lambda y$. This is done here by considering the splitting,

$$f = \lambda y = ay + b, \quad a = \theta\lambda \quad \text{and} \quad b = \lambda(1 - \theta)y,$$

The parameter $\theta > 0$ controls with what accuracy the exact linear part of f in equation 1 is captured by a in equation 3. In practice $\theta \neq 1$, though we may hope that $\theta - 1$ is small. In this framework, the stability function and the stability domain depend on θ, following the idea of Perego and Venezziani in [17]. For a fixed θ, the stability function is ρ_θ so that

$$\left|\frac{y_{n+1}}{y_n}\right| \leq \rho_\theta(\lambda h),$$

and the stability domain is $D_\theta = \{z \in \mathbb{C}, \quad \rho_\theta(z) < 1\}$.

5.2. A(0) stability

The stability functions $\rho_\theta(z)$ are numerically studied for $z \in \mathbb{R}^-$. These functions have been plotted for different values of the parameter θ. The results are depicted on figure 1. A limit $\lim_{-\infty} |\rho_\theta|$ is always observed. The scheme is $A(0)$ stable when this limit is lower than 1. From Figure 1,

- EAB_2 scheme is $A(0)$ stable if $\theta \geq 0.75$,
- EAB_3 scheme is $A(0)$ stable if $0.88 \leq \theta \leq 1.9$,
- EAB_4 scheme is $A(0)$ stable if $0.94 \leq \theta \leq 1.2$.

Roughly speaking, $A(0)$ stability holds for the EAB_k scheme if the exact linear part of $f(t, y)$ in problem (1) is approximated with an accuracy of 75 %, 85 % or 95% for $k = 2$, 3 or 4 respectively.

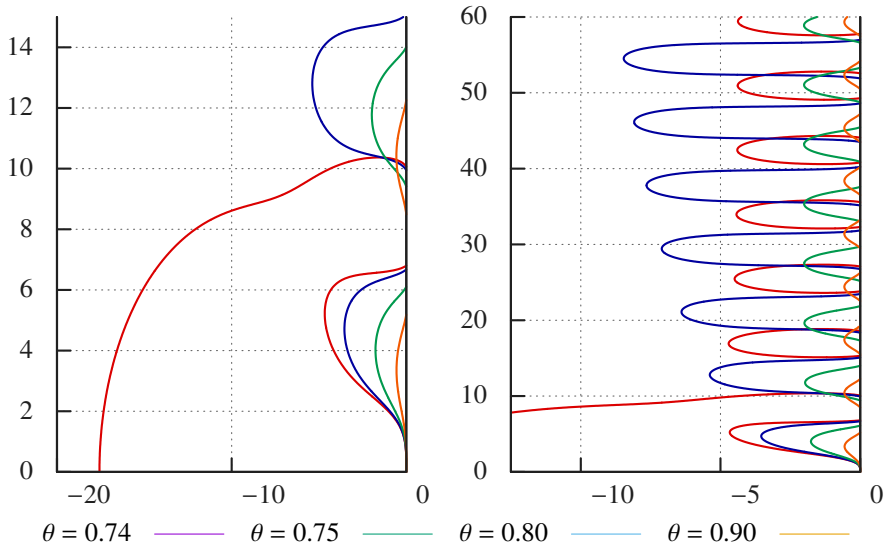


Figure 2: EAB_2: isolines $\rho_\theta(z) = 1$ for two different ranges. The stability domain D_θ is on the left of the isoline.

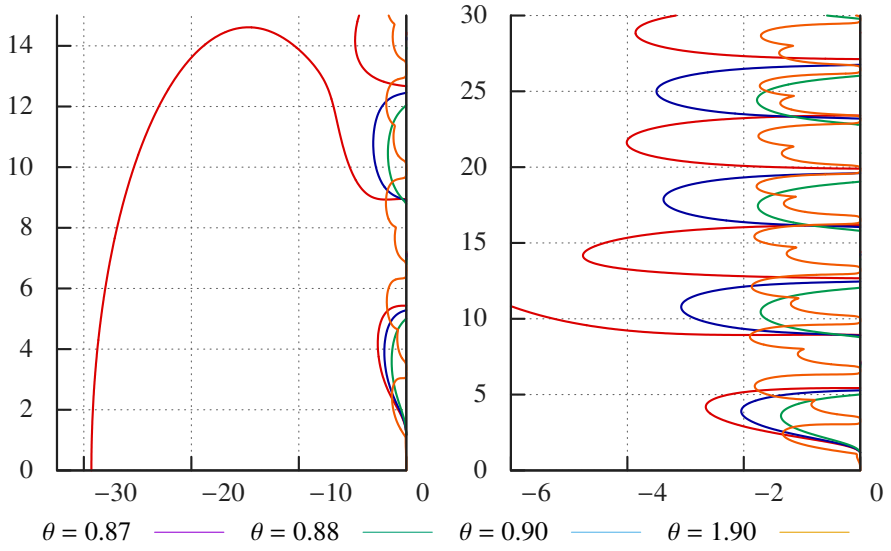


Figure 3: Same thing as figure 2 for the EAB_3 scheme.

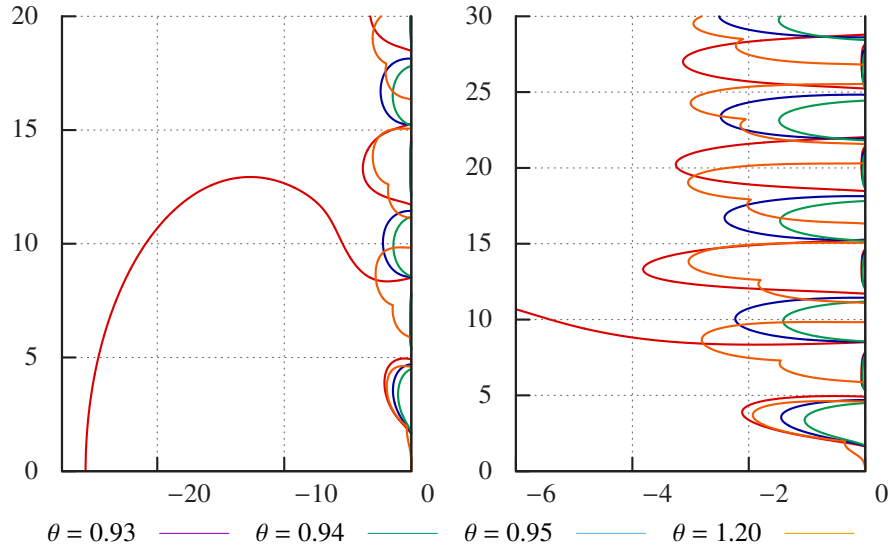


Figure 4: Same thing as figure 2 for the EAB_4 scheme.

5.3. $A(\alpha)$ stability

The stability domains D_θ have been plotted for various values of θ taken from figure 1. The results are depicted on the figures 2 to 4 for $k = 2$ to 4 respectively. Each figure shows the isolines $\rho_\theta(z) = 1$. The stability domain D_θ is on the left of these curves.

- Figure 2 shows that the EAB_2 scheme is $A(\alpha)$ stable when $\theta = 0.75$, 0.8 and 0.9 with $\alpha \simeq 50$, 60 and 80 angle degrees respectively.
- Figure 3 displays $A(\alpha)$ stability with $\alpha \simeq 60$, 70 and 60 angle degrees for $\theta = 0.88$, 0.9 and 1.9 respectively for the EAB_3 scheme.
- For the EAB_4 scheme eventually, $A(\alpha)$ stability holds with an angle α approximately of 65, 70 and 60 degrees for $\theta = 0.94$, 0.95 and 1.2 respectively, as shown on figure 4.

In all cases, when $A(\alpha)$ stability is observed, the unstable region inside $\mathbb{C}^-$ is made of a discrete collection of uniformly bounded sets located along the imaginary axes. Hence, the stability domain D_θ also contains half planes of the form $\mathrm{Re}(z) \leq a < 0$. We conjecture that, when θ is so that the EAB_k scheme is $A(\alpha)$-stable, then it is also stiff stable.

5.4. Conclusion

For explicit linear multistep methods, $A(0)$ stability cannot occur, see [31, chapter V.2]. In contrast, EAB_k and I-EAB_k methods exhibit much better stability properties. When θ is close enough to 1, they are $A(\alpha)$ stable and stiff stable. Such stability properties are comparable with those of implicit linear multistep methods. In practice, these properties will hold if the stabilization term $a(t, y)$ in (3) approximates the Jacobian of $f(t, y)$ in (1) with an absolute discrepancy lower than 25 %, 10 % and 5 % for k =2, 3 and 4 respectively.

6. Numerical results

We present in this section numerical experiments that investigate the convergence, accuracy and stability properties of the I-EAB$_k$ and EAB$_k$ schemes. The membrane equation in cardiac electrophysiology is considered for two ionic models, the Beeler-Reuter (BR) and to the Ten-Tusscher *et al.* (TNNP) models. We refer to [4] and [6] for the definition of the models. The stiffness of these two models is due to the presence of different time scales ranging from 1 ms to 1 s, as depicted on figure 5. The stabilizer a_n always is a diagonal matrix in this section.

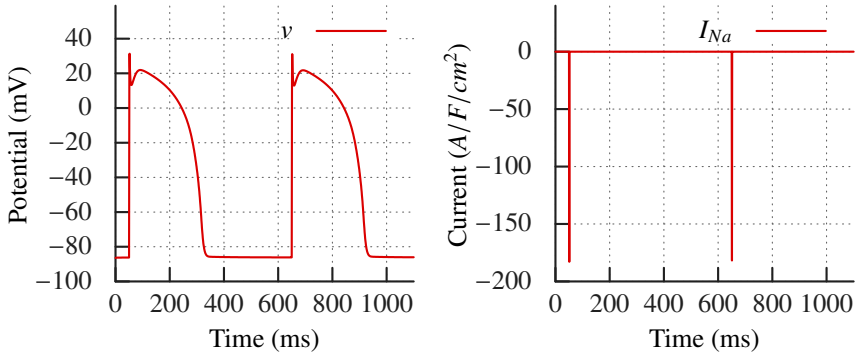


Figure 5: Two consecutive action potentials for the TNNP model: transmembrane potential v (left) and the fast sodium current I_{Na} (right), that is the main component of I_{ion} during the fast upstroke of the action potential.

The membrane equation has the general form, see [34, 4, 5, 6]:

$$\frac{dw_i}{dt} = \frac{w_{\infty,i}(v) - w_i}{\tau_i(v)}, \quad \frac{dc}{dt} = g(w, c, v), \quad \frac{dv}{dt} = -I_{\text{ion}}(w, c, v) + I_{\text{st}}(t), \tag{29}$$

where $w = (w_1, \ldots, w_p) \in \mathbb{R}^p$ is the vector of the gating variables, $c \in \mathbb{R}^q$ is a vector of ionic concentrations or other state variables, and $v \in \mathbb{R}$ is the cell membrane potential. These equations model the evolution of the transmembrane potential of a single cardiac cell. The four functions $w_{\infty,i}(v)$, $\tau_i(v)$, $g(w, c, v)$ and $I_{\text{ion}}(w, c, v)$ are given reaction terms. They characterize the cell model. The function $I_{\text{st}}(t)$ is a source term. It represents a stimulation current.

The formulation (3) is recovered with,

$$a(t, y) = \begin{pmatrix} -1/\tau(v) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad b(t, y) = \begin{pmatrix} w_\infty(v)/\tau(v) \\ g(y) \\ -I_{\text{ion}}(y) + I_{\text{st}}(t) \end{pmatrix},$$

for $y = (w, c, v) \in \mathbb{R}^N$ (with $N = p + q + 1$) and where $-1/\tau(v) = \text{diag}\,(-1/\tau_i(v))_{i=1\ldots p}$.

6.1. Implementation and computational cost

The computation of y_{n+1} with the I-EAB$_k$ and EAB$_k$ schemes requires the data y_{n-i}, $a_{n-i} := a(t_{n-i}, y_{n-i})$, and $b_{n-i} := b(t_{n-i}, y_{n-i})$ for $i = 0 \ldots k$.

EAB$_k$ practical implementation

Firstly, the $g_{n-i} = b_{n-i} + (a_{n-i} - a_n)y_{n-i}$ are updated at each time step. Then the coefficients γ_{nj} in table 1 are computed. Secondly, the computation of y_{n+1} by formula (8) also requires the computation of the $\varphi_j(a_n h)\gamma_{nj}$. This is a matrix-vector product in general.

In the present case of a diagonal stabilizer, it becomes a scalar-scalar product per row. The $\varphi_j(a_n h)$ are computed on all diagonal entries of $a_n h$. This computation simply necessitates to compute $\varphi_0(a_n h) = \mathrm{e}^{a_n h}$ (one exponential per non zero diagonal entry) thanks to the recursion rule (9).

In general, the relation (9) can be used to replace the computation of the $\varphi_j(a_n h)\gamma_{nj}$ for $j = 0 \dots k$ by the computation of a single product $\varphi_k(a_n h)w_k$. Denoting by $w_1 = a_n y_n + b_n$ and $w_j = \gamma_{nj} + a_n h w_{j-1}$:

$$\begin{aligned}
\mathrm{EAB}_2 &: y_{n+1} = y_n + h\left(w_1 + \varphi_2(a_n h)w_2\right), \\
\mathrm{EAB}_3 &: y_{n+1} = y_n + h\left(w_1 + w_2/2 + \varphi_3(a_n h)w_3\right), \\
\mathrm{EAB}_4 &: y_{n+1} = y_n + h\left(w_1 + w_2/2 + w_3/6 + \varphi_4(a_n h)w_4\right).
\end{aligned}$$

I-EAB$_k$ practical implementation

In addition, the I-EAB$_k$ method (10) requires a quadrature rule of sufficient order to preserve the scheme accuracy and convergence order. We used the Simpson quadrature rule for the cases $k = 2$, 3 and the three point Gaussian quadrature rule for k=4. We point out that a_n is assumed diagonal here so that the matrix exponentials below actually are scalar exponential.

The I-EAB$_k$ method with Simpson quadrature rule reads,

$$y_{n+1} = \mathrm{e}^{\tilde{g}_1}\left(y_n + b_n h/6\right) + \left(\tilde{b}_1 + 4\,\mathrm{e}^{\delta}\,\tilde{b}_{1/2}\right)h/6,$$

where (with the notations of section 2.2) $\tilde{g}_1 = \tilde{g}_n(t_{n+1})$, $\delta = \tilde{g}1 - \tilde{g}_n(t_n + h/2)$, $\tilde{b}_1 = \tilde{b}_n(t_{n+1})$ and $\tilde{b}_{1/2} = \tilde{b}_n(t_n + h/2)$. These coefficients are given for $k = 2$ by,

$$\tilde{g}_1 = (3a_n - a_{n-1})\,h/2, \quad \delta = (7a_n - 3a_{n-1})\,h/8, \quad \tilde{b}_1 = 2b_n - b_{n-1}, \quad \tilde{b}_{1/2} = (3b_n - b_{n-1})/2,$$

and for $k = 3$ by,

$$\begin{aligned}
\tilde{g}_1 = (23a_n - 16a_{n-1} + 5a_{n-2})\,h/12, \quad \delta = (29a_n - 25a_{n-1} + 8a_{n-2})\,h/24, \\
\tilde{b}_1 = 3b_n - 3b_{n-1} + b_{n-2}, \quad \tilde{b}_{1/2} = (15b_n - 10b_{n-1} + 3b_{n-2})\,/8.
\end{aligned}$$

The I-EAB$_k$ method with the three point Gaussian quadrature rule reads,

$$y_{n+1} = \mathrm{e}^{\tilde{g}_1}\left(y_n + \frac{h}{18}\left(5\tilde{b}_l\,\mathrm{e}^{-\tilde{g}_l} + 8\tilde{b}_0\,\mathrm{e}^{-\tilde{g}_0} + 5\tilde{b}_r\,\mathrm{e}^{-\tilde{g}_r}\right)\right),$$

with $\tilde{b}_s = \tilde{b}_n(t_s)$, $\tilde{g}_s = \tilde{g}_n(t_s)$ for $s \in \{l, 0, r\}$ where $t_l = t_n + (1 - \sqrt{3/5}))h/2$, $t_0 = t_n + h/2$, $t_r = t_n + (1 + \sqrt{3/5}))h/2$ and with $\tilde{g}_1 = \tilde{g}_n(t_{n+1})$. These parameters are linear combination of the data a_{n-i}, b_{n-i} for $i = 0 \dots k-1$ with fixed coefficients. Formula for $k = 4$ follow. The parameters $\tilde{b}_s$ are given by

$$16\tilde{b}_0 = 35b_n - 35b_{n-1} + 21b_{n-2} - 5b_{n-3},$$

and

$$40\tilde{b}_r = (95 + 179\sqrt{15}/15)b_n - (107 + 119\sqrt{15}/5)b_{n-1} + (69 + 79\sqrt{15}/5)b_{n-2} - (17 + 59\sqrt{15}/15)b_{n-3},$$

and $\tilde{b}_l$ is the radical conjugate of $\tilde{b}_r$ (the radical conjugate of $x + \sqrt{y}$ is $x - \sqrt{y}$). Finally, the parameters $\tilde{g}_s$ definition is

$$24/h\tilde{g}_1 = 55a_n - 59a_{n-1} + 37a_{n-2} - 9a_{n-3},$$
$$384/h\tilde{g}_0 = 297a_n - 187a_{n-1} + 107a_{n-2} - 25a_{n-3},$$

and

$$200/h\tilde{g}_r = \left(797/4 + 45\sqrt{15}\right)a_n - \left(2233/12 + 47\sqrt{15}\right)a_{n-1} + \left(1373/12 + 29\sqrt{15}\right)a_{n-2} - \left(331/12 + 7\sqrt{15}\right)a_{n-3},$$

and $\tilde{g}_l$ is the radical conjugate of $\tilde{g}_r$.

Computational cost

Consider an ODE system (1) whose numerical resolution cost is dominated by the computation of $(t, y) \mapsto f(t, y)$. This might be the case in general for *"large and complex models"*. For such problems explicit multistep methods are relevant since they will require one such operation per time step. In contrast, implicit methods, associated to a non linear solver, may necessitate a lot of these operations, especially for large time steps when convergence is harder to reach.

In addition, the I-EAB$_k$ and EAB$_k$ schemes need several specific operations. In the case of a diagonal function $a(t, y)$ they have been previously described: the EAB$_k$ require one scalar exponential computation per non zero row of $a(t, y)$, the I-EAB$_k$ with Simpson rule needs twice more and the I-EAB$_3$ with 3 point Gaussian quadrature rule four times more. Such a cost is not negligible, but is at worst of same order than computing $(t, y) \mapsto f(t, y)$ for complex models. For the TNNP model considered here, computing $(t, y) \mapsto f(t, y)$ costs 50 scalar exponentials whereas the EAB$_k$ implementation adds 7 supplementary scalar exponentials per time step. In terms of cost per time step, the EAB$_k$ method is rather optimal. The relationship between accuracy and cost of the EAB$_k$ method has been investigated in [35]: more details are available in section 6.4.

6.2. Convergence

For the chosen application, no theoretical solution is available. Convergence properties are studied by computing a reference solution y_{ref} for a reference time step h_{ref} with the Runge Kutta 4 scheme. Numerical solutions y are computed to y_{ref} for coarsest time steps $h = 2^p h_{ref}$ for increasing p. Any numerical solution y consists in successive values y_n at the time instants $t_n = nh$. On every interval (t_{3n}, t_{3n+3}) the polynomial $\overline{y}$ of degree at most 3 so that $\overline{y}(t_{3n+i}) = y_{3n+i}$, $i = 0 \dots 4$ is constructed. On $(0, T)$, $\overline{y}$ is a piecewise continuous polynomial of degree 3. Its values at the reference time instants nh_{ref} are computed. This provides a projection $P(y)$ of the numerical solution y on the reference grid. Then $P(y)$ can be compared with the reference solution y_{ref}. The numerical error is defined by,

$$e(h) = \frac{\max |v_{\text{ref}} - P(v)|}{\max |v_{\text{ref}}|}, \tag{30}$$

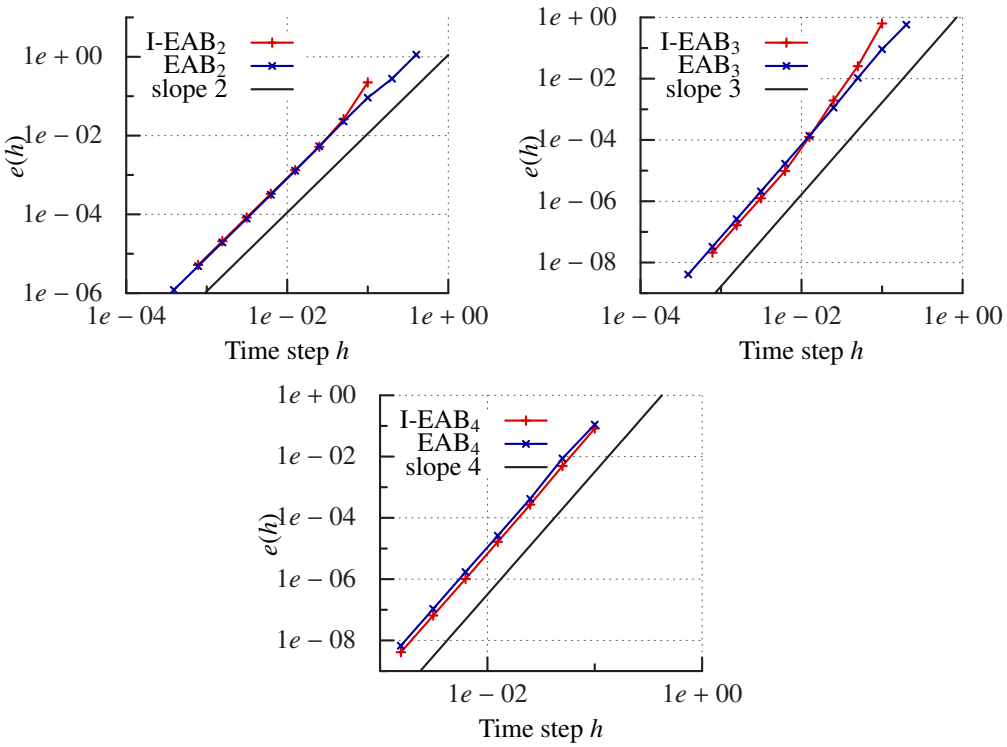


Figure 6: Relative L^∞ error $e(h)$ for the I-EAB$_k$ and EAB$_k$ schemes, $k = 2$, 3 and 4, and for the BR model.

where the potential v is the last and stiffest component of y in equation 29.

The convergence graphs for the BR model are plotted on figure 6. All the schemes display the expected asymptotic behavior $e(h) = O(h^k)$ as $h \to 0$, as proved in theorem 1.

6.3. Stability

The stiffness of the BR and TNNP models along one cellular electrical cycle (as depicted on figure 5) has been evaluated in [36]. The largest negative real part of the eigenvalues of the Jacobian matrix during this cycle is of -1170 and -82 for the TNNP and BR models respectively. This means that the TNNP model is 15 times stiffer than the BR model ($15 \simeq 1170/82$).

We want to evaluate the impact of this increase of stiffness in terms of stability for the EAB_k and I-EAB_k schemes and to provide a comparison with some other classical time stepping methods. To this aim we consider the *critical time step* Δt_0. It is defined as the largest time step such that the numerical simulation runs without overflow nor non linear solver failure for $h < \Delta t_0$. The numerical evaluation of Δt_0 is easy for explicit methods. For implicit methods, the choice of the non linear solver certainly impacts Δt_0. Without considering more deeply this problem, we just carefully set up the non linear solver, so as to provide the largest Δt_0. In practice, we have been using a Jacobian free Krylov Newton method.

Table 2: Critical time step Δt_0

(a) Classical methods

	BR	TNNP
AB_2	0.124×10^{-1}	0.850×10^{-3}
BDF_2	0.306	0.158
AB_3	0.679×10^{-2}	0.464×10^{-3}
BDF_3	0.362	0.181
AB_4	0.372×10^{-2}	0.255×10^{-3}
RK_4	0.338×10^{-1}	0.255×10^{-2}
BDF_4	0.423	0.201

(b) I-EAB_k and EAB_k exponential methods

	BR	TNNP
I-EAB_2	0.121	0.103
EAB_2	0.424	0.233
I-EAB_3	0.103	0.123
EAB_3	0.203	0.108
I-EAB_4	0.133	0.106
EAB_4	0.122	0.756×10^{-1}

Results are on table 2. The Adams Bashforth (AB_k) and the backward differentiation (BDF_k) methods of order k have been considered, together with the RK_4 scheme.

The AB_k and the RK_4 schemes have bounded stability domain (see [31, p. 243]). Then it is expected for the critical time step to be divided by a factor close to 15 between the BR and TNNP models. Results presented in table 2 show this behavior.

The BDF_2 scheme is A-stable whereas the BDF_3 and BDF_4 are $A(\alpha)$-stable with large angle α (see [31, p. 246]). Hence the critical time step is expected to remain unchanged between the two models. Table 2 shows that the Δt_0 actually are divided by approximately 2.

The critical time steps for the I-EAB_k and EAB_k models are presented in table 2. The critical time steps for the I-EAB_k schemes remain almost unchanged from the BR to the TNNP model. For the EAB_k, they are divided by approximately 2, which behavior is similar as for the BDF_k method.

As a conclusion, for the present application, the EAB_k and I-EAB_k methods are as robust to stiffness than the implicit BDF_k schemes, though being explicit. As a matter of fact, section 5 shows that the stability domains for the I-EAB_k and EAB_k schemes depend on the discrepancy

between the complete Jacobian matrix and $a(t, y)$. In the present case, $a(t, y)$ only contains a part of the Jacobian matrix diagonal. It is very interesting to notice that robustness to stiffness is actually achieved with this choice. It is finally also interesting to see that the critical time steps of implicit and exponential methods are of the same order.

6.4. Accuracy

In terms of accuracy, the schemes can be compared using the relative error $e(h)$ in equation (30). The EAB_k and I-EAB_k schemes can be compared with the AB_k methods only at very small

Table 3: Accuracy $e(h)$ for the AB_k,I-EAB_k and EAB_k schemes: using the BR model and fixed time step $h = 10^{-3}$

	$k = 2$	$k = 3$	$k = 4$
AB_k	5.32×10^{-6}	4.33×10^{-8}	8.69×10^{-10}
I-EAB_k	8.55×10^{-6}	4.44×10^{-8}	7.30×10^{-10}
EAB_k	7.90×10^{-6}	7.00×10^{-8}	1.16×10^{-9}

time steps, because of the lack of stability of AB_k schemes (see table 2). In table 6.4 are given the accuracies of these methods for a given time step $h = 10^{-3}$ and for the BR model. Ii is observed that the same level of accuracy is obtained with AB_k and EAB_k at fixed k. These figures illustrate that inside the asymptotic convergence region, EAB_k, I-EAB_k and AB_k schemes are equivalent in terms of accuracy.

Table 4: Accuracy for the TNNP model

(a) EAB_k

h	$k = 2$	$k = 3$	$k = 4$
0.1	0.351	0.530	
0.05	9.01×10^{-2}	5.59×10^{-2}	8.93×10^{-2}
0.025	2.14×10^{-2}	7.34×10^{-3}	8.34×10^{-3}

(b) BDF_k

h	$k = 2$	$k = 3$	$k = 4$
0.1			0.129
0.05	3.57×10^{-2}	1.15×10^{-2}	1.44×10^{-2}
0.025	1.10×10^{-2}	2.58×10^{-3}	2.38×10^{-3}

Comparison at large time steps between the EAB_k and BDF_k for the TNNP model is shown in table 4. These figures show that for large time steps BDF_k is more accurate than EAB_k. A gain in accuracy of factor 2.5, 5 and 6 is observed for $h = 0.05$ and for k=2, 3 and 4 respectively. However, compare row 3 for EAB_k ($h = 0.025$) with row 2 for BDF_k ($h = 0.05$). It shows that the numerical solutions with an accuracy close to 0.01 are obtained when dividing the time step by (at most) 2 between BDF_k and EAB_k. Meanwhile, EAB_k with $h = 0.025$ costs less than BDF_k with $h = 0.05$, as developed in section 6.1.

We conclude that EAB_k schemes provide a cheaper way to compute numerical solutions at large time step for a given accuracy. The same conclusion also holds for the BR model, see table

5. A deeper analysis of the relationship between accuracy and computational cost for the EAB_k scheme as compared to other methods is available in [35] with the same conclusion.

Table 5: Accuracy for the BR model

(a) EAB_k

h	$k=2$	$k=3$	$k=4$
0.2	0.284	0.516	
0.1	9.26×10^{-2}	9.17×10^{-2}	0.119
0.05	8.20×10^{-2}	1.09×10^{-2}	8.96×10^{-3}

(b) $BDFk_k$

h	$k=2$	$k=3$	$k=4$
0.2	9.74×10^{-2}	4.09×10^{-2}	4.98×10^{-2}
0.1	3.44×10^{-2}	1.04×10^{-2}	1.27×10^{-2}
0.05	9.74×10^{-3}	2.29×10^{-3}	2.02×10^{-3}

In table 5 are given the accuracies at large time step now considering the BR model. Comparison with table 4 shows that accuracy is preserved by dividing h by 2 when switching from the BR to the TNNP model. As already said, the TNNP model is 15 times stiffer than the BR model.

We conclude that the EAB_k schemes also exhibit a large robustness to stiffness in terms of accuracy. This robustness is equivalent as for the implicit BDF_k schemes. This is remarkable for an explicit scheme, as for the robustness to stiffness in terms of critical time step discussed in the previous subsection.

7. Acknowledgments

This study received financial support from the French Government as part of the "Investissement d'avenir" program managed by the National Research Agency (ANR), Grant reference ANR-10-IAHU-04. It also received fundings of the project ANR-13-MONU-0004-04.

[1] J. Clements, J. Nenonen, P. Li, B. Horacek, Activation dynamics in anisotropic cardiac tissue via decoupling, Ann. Biomed. Eng. 32 (7) (2004) 984–990.

[2] P. Colli-Franzone, L. Pavarino, B. Taccardi, Monodomain simulations of excitation and recovery in cardiac blocks with intramural heterogeneity, in: A. F. Frangi, P. I. Radeva, A. Santos, M. Hernandez (Eds.), Functional Imaging and Modeling of the Heart, Vol. 3504 of Theoretical Computer Science and General Issues, Springer-Verlag Berlin Heidelberg, 2005, pp. 267–277.

[3] P. C. Franzone, L. Pavarino, B. Taccardi, Simulating patterns of excitation, repolarization and action potential duration with cardiac bidomain and monodomain models, Mathematical Biosciences 197 (1) (2005) 35 – 66.

[4] G. W. Beeler, H. Reuter, Reconstruction of the action potential of ventricular myocardial fibres, J. Physiol. 268 (1) (1977) 177–210.

[5] C. H. Luo, Y. Rudy, A dynamic model of the cardiac ventricular action potential. I. Simulations of ionic currents and concentration changes., Circ. Res. 74 (6) (1994) 1071–1096.

[6] K. H. W. J. ten Tusscher, D. Noble, P. J. Noble, A. V. Panfilov, A model for human ventricular tissue, Am. J. Physiol. Heart Circ. Physiol. 286 (4) (2004) H1573–H1589.

[7] V. Iyer, R. Mazhari, R. L. Winslow, A computational model of the human left-ventricular epicardial myocyte, Biophys. J. 87 (3) (2004) 1507–1525.

[8] M. Hochbruck, C. Lubich, H. Selhofer, Exponential integrators for large systems of differential equations, SIAM J. Sci. Comput. 19 (5) (1998) 1552–1574.

[9] S. M. Cox, P. C. Matthews, Exponential time differencing for stiff systems, J. Comput. Phys. 176 (2) (2002) 430–455.
[10] M. Hochbruck, A. Ostermann, Explicit exponential Runge-Kutta methods for semilinear parabolic problems, SIAM J. Numer. Anal. 43 (3) (2005) 1069–1090.
[11] M. Hochbruck, A. Ostermann, J. Schweitzer, Exponential Rosenbrock-type methods, SIAM J. Numer. Anal. 47 (1) (2009) 786–803.
[12] M. Tokman, J. Loffeld, P. Tranquilli, New adaptive exponential propagation iterative methods of Runge-Kutta type, SIAM J. Sci. Comput. 34 (5) (2012) A2650–A2669.
[13] V. T. Luan, A. Ostermann, Explicit exponential Runge-Kutta methods of high order for parabolic problems, J. Comput. Appl. Math. 256 (2014) 168–179.
[14] B. Minchev, W. M. Wright, A review of exponential integrators for first order semi-linear problems, Preprint Numerics 2/2005, Norges Teknisk-Naturvitenskapelige Universitet (2005).
[15] M. Hochbruck, A. Ostermann, Exponential integrators, Acta Numer. 19 (2010) 209–286.
[16] M. Hochbruck, A short course on exponential integrators, in: Z. Bai, W. Gao, Y. Su (Eds.), Matrix Functions and Matrix Equations, Vol. 19 of Contemp. Appl. Math., Higher Ed. Press, Beijing, 2015, pp. 28–49.
[17] M. Perego, A. Veneziani, An efficent generalization of the Rush-Larsen method for solving electro-physiology membrane equations, ETNA 35 (2009) 234–256.
[18] C. Börgers, A. R. Nectow, Exponential time differencing for Hodgkin-Huxley-like ODEs, SIAM J. Sci. Comput. 35 (3) (2013) B623–B643.
[19] J. Certaine, The solution of ordinary differential equations with large time constants, in: A. Ralston, H. S. Wilf (Eds.), Mathematical Methods for Digital Computers, John Wiley & Sons, 1960, pp. 128–132.
[20] S. P. Norsett, An A-stable modification of the Adams-Bashforth methods, in: J. L. Morris (Ed.), Conference on the Numerical Solution of Differential Equations: Held in Dundee/Scotland, June 23–27, 1969, Springer, Berlin, Heidelberg, 1969, pp. 214–219.
[21] M. Tokman, Efficient integration of large stiff systems of ODEs with exponential propagation iterative (EPI) methods, J. Comput. Phys. 213 (2) (2006) 748–776.
[22] M. P. Calvo, C. Palencia, A class of explicit multistep exponential integrators for semilinear problems, Numer. Math. 102 (3) (2006) 367–381.
[23] A. Ostermann, M. Thalhammer, W. M. Wright, A class of explicit exponential general linear methods, BIT 46 (2) (2006) 409–431.
[24] M. Hochbruck, A. Ostermann, Exponential multistep methods of Adams-type, BIT 51 (4) (2011) 889–908.
[25] W. Auzinger, M. Lapińska, Convergence of rational multistep methods of Adams-Padé type, Tech. rep., ASC Report, Vienna University of Technology (2011).
[26] A. Koskela, A. Ostermann, Exponential Taylor methods: analysis and implementation, Comput. & Math. with Appl. 65 (3) (2013) 487–499.
[27] D. Lee, S. Preiser, A class of non linear multistep A-stable numerical methods for solving stiff differential equations, Comput. & Math with Appl. 4 (1978) 43–51.
[28] M. T. Chu, An automatic multistep method for solving stiff initial value problems, J. Comput. Appl. Math. 9 (3) (1983) 229–238.
[29] P. Tranquilli, A. Sandu, Rosenbrock-Krylov methods for large systems of differential equations, SIAM J. Sci. Comput. 36 (3) (2014) A1313–A1338.
[30] G. Rainwater, M. Tokman, A new class of split exponential propagation iterative methods of Runge-Kutta type (sEPIRK) for semilinear systems of ODEs, J. Comput. Phys. 269 (2014) 40–60.
[31] E. Hairer, G. Wanner, Solving Ordinary Differential Equations II, Vol. 14 of Springer Series in Computational Mathematics, Springer, Berlin, Heidelberg, 1996.
[32] E. Hairer, S. P. Nørsett, G. Wanner, Solving Ordinary Differential Equations I, Vol. 8 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1993.
[33] S. S. Dragomir, Some Gronwall type inequalities and applications, Nova Science Publishers, Inc., Hauppauge, New York, 2003.
[34] A. Hodgkin, A. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, J. Physiol. 117 (1952) 500–544.
[35] C. Douanla Lontsi, Y. Coudière, C. Pierre, Efficient high order schemes for stiff ODEs in cardiac electrophysiology, in: M. Lo, E. Badouel, N. Gmati (Eds.), Proceedings of CARI 2016, Hammamet, Tunisia, 2016, pp. 312–319.
[36] R. J. Spiteri, C. D. Ryan, Stiffness analysis of cardiac electrophysiological models, Ann. Biomed. Eng. 38 (2010) 3592–3604.

4 Généralisation des schémas Rush Larsen

Ce chapitre est consacré à la généralisation des méthodes de type Rush-Larsen. Très utilisé en électrophysiologie pour la résolution des équations différentielles sur les variables de portes des modèles ioniques, les schémas Rush-Larsen (RL) doivent leur popularité à leur facilité de mise en œuvre et leur efficacité à faire face à la raideur des équations constituant cette partie du modèle ionique. Ces méthodes étaient jusqu'ici limitées à l'ordre 1 et 2 de précision. Mais nous introduirons par l'article qui va suivre une méthode qui permet de construire l'ordre supérieur de cette méthode. Cette méthode sera utilisée en particulier pour construire les schémas Rush-Larsen d'ordre 3 et 4 que nous noterons simplement RL_k k étant l'ordre du schéma. Les schémas RL s'inscrivent dans la classe des schémas exponentiels multi-pas. Comme dans le cas des schémas IEAB et EAB, les schémas RL sont construits à base de la formule de variation de la constante. Sa particularité par rapport aux schémas IEAB et EAB est au niveau du choix du stabilisateur et la partie non linéaire. En effet, contrairement à ces deux familles de schémas, le stabilisateur et la partie linéaire sont des constantes convenablement choisies de sorte que l'ordre de précision voulu soit atteint. Toujours dans le contexte où on peut prendre autre chose que la jacobienne de la fonction de l'équation comme stabilisateurs, nous avons énoncé et prouvé des résultats de convergence pour les schémas RL. Ces résultats ont été confirmés numériquement en appliquant cette famille de schémas au modèle ionique de Beeler et Reuter. La robustesse de ce schéma par rapport à la raideur a été évaluée par le calcul des pas de temps critiques. Ces pas de temps critiques ont été mis en comparaison avec ceux des schémas EAB. Il en est ressorti que ces méthodes permettent d'utiliser des grands pas de temps tout en produisant des bonnes précisions. De plus, elles résistent à la raideur aussi bien que les schémas EAB.
Nous avons aussi étudié numériquement les propriété de stabilité au sens de Dahlquist. Cette étude nous a permis de produire des domaines de stabilités. Ces domaines permettent de faire le choix du pas de temps à utiliser par rapport à la raideur du problème que l'on veut résoudre. Ces domaines de stabilités sont assez grands et confirment ainsi l'ordre de grandeur des pas de temps critiques qui ont été évalués. Nous avons montré qu'en terme de précision, le schéma EAB est plus précis que le schéma RL à l'ordre 2. À l'ordre 3, le schéma RL est meilleur pour des grands pas de temps. Mais pour des pas de temps petits (< 0.025) le schéma EAB est légèrement plus précis. À l'ordre 4, le schéma RL est légèrement plus précis que le schéma EAB. Au final, ces deux schémas présentent asymptotiquement presque la même précision. Cependant la facilité de mise en œuvre du schéma RL fait de lui le meilleur.

Rush Larsen time stepping methods of high order for stiff problems in cardiac electrophysiology

Y. COUDIÈRE, C. DOUANLA LONTSI, AND C. PIERRE

Abstract. The development of efficient solvers in cardiac electrophysiology requires high order (semi) explicit and stable time stepping methods. In this paper are introduced two new exponential integrators of orders 3 and 4. They generalize the order 2 Rush Larsen scheme derived by Perego and Veneziani [24] in 2009. They have been named Rush Larsen of order k, shortly RL_k. The RL_k schemes are explicit exponential multistep integrators. They display a simple general formulation and an easy implementation.
The RL_k schemes are shown to be stable under perturbation (or 0-stable) and convergent of order k. Their Dahlquist stability analysis is performed. They have a very large stability domain provided that the stabilizer associated with the method captures well enough the stiff modes of the problem. The RL_k method is numerically studied as applied to the membrane equation in cardiac electrophysiology.

1. Introduction

The monodomain model in cardiac electrophysiology [3, 4, 5], formulates as a coupling between an evolution reaction diffusion equation and an ODE system. On the heart domain Ω and on the time interval $[0,T]$, it reads:

$$\frac{\partial v}{\partial t} = Av + f_1(v,w) + s(x,t), \quad \frac{\partial w}{\partial t} = f_2(v,w), \tag{1}$$

where A is a diffusion operator. The first unknown $v : \Omega \times [0,T] \to \mathbb{R}$ is the cellular transmembrane potential. The second unknown $w : \Omega \times [0,T] \to \mathbb{R}^N$ gathers variables describing the cellular membrane state: it incorporates ionic concentrations and gating variables. The source term $s(x,t)$ allows to apply stimulation currents to the system. The reaction terms f_1 and f_2 are cell membrane models for ionic currents and voltage, that are named ionic models. Ionic models originally have been developed by Hodgkin and Huxley [17] in 1952 for the squid axon. Several

Date: July the 10^{th} 2017.

2010 *Mathematics Subject Classification.* 65L04, 65L06, 65L20, 65L99.

Key words and phrases. stiff equations, explicit high-order multistep methods, exponential integrators of Adams type, stability and convergence, Dahlquist stability .

This work was supported by the ANR, INRIA and the CNRS.

ionic models have been especially designed for cardiac cells, such as the Beeler Reuter model [1], the Luo and Rudy models [21, 22] or the TNNP model [27].
Numerical simulations in cardiac electrophysiology face two difficulties. The first one is the stiffness displayed by the solutions of (1). This is commonly coped with by resorting to very fine space and time grids, associated with high computational costs. Stiffness is due to the coexistence of fast and slow variables. Fast variables in ionic models are described by equations in the ODE system in (1) of the form,

$$\frac{\partial w_i}{\partial t} = f_{2,i}(v,w) = a_i(v)w_i + b_i(v). \tag{2}$$

This feature of ionic models will be exploited here. The rate of variations $a_i(v)$ will be inserted in the numerical method in order to allow stable computations at large time step.
The second difficulty is the nature of the reaction terms f_1 and f_2 in (1). It is non linear and the operation $(v,w) \to f_i(v,w)$ has a significant cost. For example, this operation for the TNNP model [27] involves the computation of 50 exponentials. These operations need to be performed at every node of the grid. They represent a large computational load. Their total amount needs to be maintained as low as possible. Fully implicit time stepping methods (that require a non linear solver) therefore are avoided.

Our objective for the numerical resolution of (1) is to go towards high order methods, in order to reduce the grid size. A high order time stepping method is required that fulfills two conditions. It must have strong stability properties. It has to be explicit for the reaction terms. To this aim, we will focus in this paper on the time integration of stiff ODE systems,

$$\frac{dy}{dt} = f(t,y), \quad y(0) = y_0, \tag{3}$$

for which a reformulation of the following kind is available,

$$\frac{dy}{dt} = a(t,y)y + b(t,y), \quad y(0) = y_0. \tag{4}$$

The linear part $a(t,y)$ will be referred to as the stabilizer. Exponential integrators fulfill these two conditions. We refer to [23, 14, 12] for general reviews. They have been widely studied for the quasilinear equations, $\partial_t y = Ay + b(t,y)$, see *e.g.* [13, 7, 11, 16, 28, 20]. The basic idea is to use the exact solution of the linearized equation in order to stabilize the numerical scheme. In general this implies to compute a matrix exponential. This is the supplementary cost associated to exponential integrators.
The targeted problem (4) displays a non constant linear part $a(t,y)$. Exponential integrators have been less studied in that case. Exponential integrators of Adams type for a non constant linear part have been first considered by Lee and Preiser

[19] in 1978 and by Chu [2] in 1983. Recently, Ostermann *et al* [15, 18] developed and analyzed the linearized exponential Adams method. The original problem (3) is reformulated after each time step as,

$$\frac{dy}{dt} = J_n y + c_n(t,y), \quad J_n = \partial_y f(t_n, y_n), \quad c_n(t,y) = f(t,y) - J_n y.$$

The Jacobian matrix J_n is used as a stabilizer. This requires the computation of matrix exponentials. Moreover, when the fast variables of the system are known, stabilization can be performed only on these variables. Considering the full Jacobian as the stabilizer implies unnecessary computational efforts. To avoid these problems, an alternative is to set the stabilizer as a part or as an approximation of the Jacobian. This has been analyzed in [29], [25] and [6] for exponential Rosenbrock, exponential Runge Kutta and exponential Adams type methods respectively. For exponential Adams type methods, equation (4) is reformulated after each time step as,

$$\frac{dy}{dt} = a_n y + c_n(t,y), \quad a_n = a(t_n, y_n), \quad c_n(t,y) = f(t,y) - a_n y.$$

The resulting scheme is (see details in [15, 6]),

$$y_{n+1} = y_n + h\left[\varphi_1(a_n h)\left(a_n y_n + \gamma_1\right) + \varphi_2(a_n h)\gamma_2 + \ldots + \varphi_k(a_n h)\gamma_k\right], \tag{5}$$

where γ_i are the coefficients of the Lagrange interpolation polynomial of $c_n(t,y)$ (in a classical k-step setting) and where the functions φ_j are given by,

$$\varphi_0(z) = e^z, \quad \varphi_{j+1}(z) = \frac{\varphi_j(z) - 1/j!}{z}. \tag{6}$$

Independently, Perego and Veneziani [24] presented in 2009 a new exponential integrator of order 2, of a different nature:

$$y_{n+1} = y_n + h\varphi_1(\alpha_n h)\left(\alpha_n y_n + \beta_n\right). \tag{7}$$

The two constants α_n and β_n are updated after each time step. They are defined with $\alpha_n = 3/2a_n - 1/2a_{n-1}$ and $\beta_n = 3/2b_n - 1/2b_{n-1}$ with $a_j = a(t_j, y_j)$ and $b_j = b(t_j, y_j)$. The numerical solution y_{n+1} in (7) satisfies,

$$y_{n+1} = z(t_{n+1}) \quad \text{with} \quad z' = \alpha_n z + \beta_n, \quad z(t_n) = y_n. \tag{8}$$

The ODE (8) involves two constant terms only. It is interesting to obtain the order 2 when approximating the original ODE (4) on $[t_n, t_{n+1}]$ with that simple ODE (8).

In this paper we will study the general formulation (7). It results in schemes with a very simple definition. It is in particular simpler than the exponential Adams integrators (5. We will show that such schemes also exist at the orders 3 and 4, for an explicit definition of the two constants α_n and β_n. These schemes will be referred to as Rush Larsen schemes of order k, shortly denoted RL_k. They will be shown to be stable under perturbation (or 0-stable) and convergent of order k. The

Dahlquist stability analysis for the RL_k schemes is also performed. It is a practical tool that allows to dimension the time step h with respect to the variations of $f(t,y)$ in problem (3), see *e.g.* [10]. When considering varying stabilizers, the stability domain depends on how $f(t,y)y$ is decomposed in equation (4), following [24]. The stability domains are numerically computed and shown to be much larger than in the absence of stabilization (*i.e.* when $a(t,y)=0$) provided that $a(t,y)$ captures well enough the variations of $f(t,y)$. The performances of the RL_k method are evaluated for the membrane equation in cardiac electrophysiology. They are compared with the exponential Adams integrators (5). The two methods have a very similar robustness to stiffness. They both allow stable computations on coarse time grids. At lareg time step, the RL_3 and RL_4 schemes are slightly more accurate, meanwhile with a simpler implementation.

The paper is organized as follows. The RL_k schemes are derived in section 2 and their numerical analysis is made in sections 2 and 3. The Dahlquist stability analysis is in section 4. The numerical results are presented in section 5. The paper ends with the conclusion section 6.

In the sequel h denotes the time step and $t_n = nh$ the associated time instants.

2. RL_k scheme definition and consistency

Definition 1. The RL_k scheme is an explicit k-step method. It is defined with the formulation (7) for the following setting of α_n and of β_n:

$$
\begin{aligned}
RL_2: \qquad & \alpha_n = \frac{3}{2}a_n - \frac{1}{2}a_{n-1}, \quad \beta_n = \frac{3}{2}b_n - \frac{1}{2}b_{n-1},\\
RL_3: \qquad & \alpha_n = \frac{1}{12}(23a_n - 16a_{n-1} + 5a_{n-2}),\\
& \beta_n = \frac{1}{12}(23b_n - 16b_{n-1} + 5b_{n-2}) + \frac{h}{12}(a_nb_{n-1} - a_{n-1}b_n).\\
RL_4: \qquad & \alpha_n = \frac{1}{24}(55a_n - 59a_{n-1} + 37a_{n-2} - 9a_{n-3}),\\
& \beta_n = \frac{1}{24}(55b_n - 59b_{n-1} + 37b_{n-2} - 9b_{n-3})\\
& \qquad + \frac{h}{12}(a_n(3b_{n-1} - b_{n-2}) - (3a_{n-1} - a_{n-2})b_n),
\end{aligned}
$$

where $a_j = a(t_j, y_j)$ and $b_j = b(t_j, y_j)$.

A solution $y(t)$ of equation (4) on a time interval $[0,T]$ is fixed. It is recalled that the scheme (7) is consistent of order k if:

- being given a time step h and a time instant $kh \le t_n \le T - h$,

- being given the numerical approximation y_{n+1} in (7) computed with $y_{n-j} = y(t_{n-j})$ for $j = 0 \dots k-1$,

we have $|y_{n+1} - y(t_n + h)| \leq Ch^{k+1}$, for a constant C only depending on the problem (4) data a, b, y_0 and on T.

Proposition 1. *Assume that the functions $a(t, y)$ and $b(t, y)$ in problem (4) are C^k regular. Moreover assume that $a(t, y)$ either is a diagonal matrix or a constant linear operator.*
Then the RL_k scheme is consistent of order k.

Remark 1. In the case of a constant linear part $a(t, y) = A$, we always have $\alpha_n = A$. The definition of β_n also simplifies at the order 3 and 4,

$$RL_3: \qquad \beta_n = \frac{1}{12}(23b_n - 16b_{n-1} + 5b_{n-2}) - \frac{h}{12}A(b_n - b_{n-1}).$$
$$RL_4: \qquad \beta_n = \frac{1}{24}(55b_n - 59b_{n-1} + 37b_{n-2} - 9b_{n-3}) - \frac{h}{12}A(2b_n - 3b_{n-1} + b_{n-2}).$$

Remark 2. The assumption *"$a(t, y)$ either is a diagonal matrix or a constant linear operator"* in proposition 1 has the following origin. To analyze the scheme consistency we will derive a Taylor expansion in h of (7). That series is computed with the help of Taylor expansions in h for α_n and β_n.
Assume the simple form $\alpha_n = \alpha_0 + h\alpha_1$. We need to expand $\varphi_1(\alpha_n h)$ as a series in h. The function φ_1 is analytic on $\mathbb{C}$. However in the matrix case, the equality, $\varphi_1(M + N) = \varphi_1(M) + \varphi_1'(M)N + \ldots + \varphi_1^{(i)}(M)N^i/i! + \ldots$ holds if M and N are commutative matrices. Therefore one cannot expand $\varphi_1(\alpha_n h)$ without the assumptions that α_0 and α_1 are commutative.
That difficulty vanishes if $a(t, y)$ is constant or scalar or, equivalently, a diagonal matrix.

The proof of proposition 1 is based on the following result.

Lemma 2. *With the same assumption as in proposition 1, the scheme (7) is consistent of order k if:*

$$
\begin{aligned}
k=2: \quad & \alpha_n = a_n + \frac{1}{2}a'_n h + O(h^2), \quad \beta_n = b_n + \frac{1}{2}b'_n h + O(h^2).\\
k=3: \quad & \alpha_n = a_n + \frac{1}{2}a'_n h + \frac{1}{6}a''_n h^2 + O(h^3),\\
& \beta_n = b_n + \frac{1}{2}b'_n h + \frac{1}{12}(a'_n b_n - a_n b'_n)h^2 + O(h^3).\\
k=4: \quad & \alpha = a_n + \frac{1}{2}a'_n h + \frac{1}{6}a''_n h^2 + \frac{1}{24}a'''_n h^3 + O(h^4),\\
& \beta = b_n + \frac{1}{2}b'_n h + \frac{1}{12}(a'_n b_n - a_n b'_n)h^2 + \left(\frac{1}{24}b'''_n + \frac{1}{24}(a''_n b_n - a_n b''_n)\right) h^3 + O(h^4).
\end{aligned}
$$

Where a'_n, a''_n, a'''_n and b'_n, b''_n, b'''_n denote the successive derivatives at time t_n of the functions $t \mapsto a(t, y(t))$ and $t \mapsto b(t, y(t))$.

Proof of lemma 2. By assumption the functions a and b in problem (4) are C^k regular. Therefore a solution y of problem (4) on a closed time interval $[0, T]$ is C^{k+1} regular. Its derivatives up to order $k+1$ can be bounded by constants only depending on the problem (4) data and on T. The Taylor expansion of y at time instant t_n is,

$$y(t_n + h) = y(t_n) + \sum_{j=1}^{k} \frac{s_j}{j!} h^j + O(h^{k+1}),$$

with $s_j = y^{(i)}(t_n)$. Using that $y' = ay + b$ we get,

$$
\begin{aligned}
s_1 =& a_n y_n + b_n,\\
s_2 =& (a'_n + a_n^2) y_n + a_n b_n + b'_n,\\
s_3 =& (a''_n + 3a_n a'_n + a_n^3) y_n + b''_n + a_n b'_n + 2a'_n b_n + a_n^2 b_n,\\
s_4 =& (a'''_n + 4a''_n a_n + 3a_n^{'2} + 6a'_n a_n^2 + a_n^4) y_n\\
& + b'''_n + b''_n a_n + 3a''_n b_n + 5a'_n a_n b_n + 3a'_n b'_n + a_n^3 b_n + a_n^2 b'_n.
\end{aligned}
$$

A series expansion in h for α_n and for β_n is formally introduced,

$$
\begin{aligned}
\alpha_n &= \alpha_{n,0} + \alpha_{n,1} h + \cdots + \alpha_{n,k-1} h^{k-1} + O(h^k),\\
\beta_n &= \beta_{n,0} + \beta_{n,1} h + \cdots + \beta_{n,k-1} h^{k-1} + O(h^k).
\end{aligned}
$$

With the assumption that $a(t, y)$ is either constant or a diagonal matrix (see remark 2), a Taylor expansion of the numerical solution y_{n+1} in (7) can be performed,

$$y_{n+1} = y(t_n) + \sum_{j=1}^{k} \frac{r_j}{j!} h^j + O(h^{k+1}).$$

A direct computation of the r_j gives,

$$\begin{aligned}
r_1 =& \alpha_{n,0} y_n + \beta_{n,0},\\
r_2 =& (2\alpha_{n,1} + \alpha_{n,0}^2) y_n + 2\beta_{n,1} + \alpha_{n,0}\beta_{n,0},\\
r_3 =& (6\alpha_{n,2} + \alpha_{n,0}^3 + 6\alpha_{n,0}\alpha_{n,1}) y_n + 3\alpha_{n,1}\beta_{n,0} + 6\beta_{n,2} + \alpha_{n,0}^2\beta_{n,0} + 3\alpha_{n,0}\beta_{n,1},\\
r_4 =& (24\alpha_{n,0}\alpha_{n,2} + 24\alpha_{n,3} + 12\alpha_{n,1}\alpha_{n,0}^2 + 12\alpha_{n,1}^2 + \alpha_{n,0}^4) y_n\\
&+ 12\alpha_{n,2}\beta_{n,0} + 24\beta_{n,3} + 12\alpha_{n,0}\beta_{n,2} + 12\alpha_{n,1}\beta_{n,1} + 4\alpha_{n,0}^2\beta_{n,1} + 8\alpha_{n,0}\alpha_{n,1}\beta_{n,0} + \alpha_{n,0}^3\beta_{n,0}.
\end{aligned}$$

The condition to be consistent of order k is: $r_i = s_i$ for $1 \leq i \leq k$. Lemma 2 consistency conditions are obtained by solving recursively these relations. □

Proof of proposition 1. It is a direct and simple consequence of the backwards differentiation formula, that we first recall. Consider a real function f, its derivatives can be approximated as follows (with obvious notations). For the first derivative,

$$\begin{aligned}
f'_n &= \frac{f_n - f_{n-1}}{h} + O(h),\\
&= \frac{1}{2h}(3f_n - 4f_{n-1} + f_{n-2}) + O(h^2),\\
&= \frac{1}{6h}(11f_n - 18f_{n-1} + 9f_{n-2} - 2f_{n-3}) + O(h^3).
\end{aligned}$$

For the second derivative,

$$\begin{aligned}
f''_n &= \frac{1}{h^2}(f_n - 2f_{n-1} + f_{n-2}) + O(h),\\
&= \frac{1}{h^2}(2f_n - 5f_{n-1} + 4f_{n-2} - f_{n-3}) + O(h^2).
\end{aligned}$$

For the third derivative,

$$f'''_n = \frac{1}{h^3}(f_n - 3f_{n-1} + 3f_{n-2} - f_{n-3}) + O(h),$$

With these formula, the consistency condition at order 3 on α_n in lemma 2 becomes,

$$\begin{aligned}\alpha_n &= a_n + \frac{1}{2}a'_n h + \frac{1}{6}a''_n h^2 + O(h^3),\\ &= a_n + \frac{1}{4}\left(3a_n - 4a_{n-1} + a_{n-2}\right) + \frac{1}{6}\left(a_n - 2a_{n-1} + a_{n-2}\right) + O(h^3)\\ &= \frac{1}{12}(23a_n - 16a_{n-1} + 5a_{n-2}) + O(h^3).\end{aligned}$$

We retrieve the definition of α_n for the RL_3 scheme. The same holds for β_n and the RL_3 scheme then is consistent order 3.
The proof is the same at the other oreders. □

3. Stability under perturbation and convergence

We refer to [9, Ch. III-8] for the definitions of convergence and of stability under perturbation (or 0-stability). For the analysis of ODE numerical integrators, it is commonly assumed that f in (3) is uniformly Lipschitz in its second variable y. That hypothesis will be replaced by assimptions based on the formulation (4). Precisely it will be assumed that,

$$a(t,y) \text{ is bounded}, \quad a(t,y),\ b(t,y) \text{ are uniformly Lipschitz in } y. \tag{9}$$

The Lipschitz constants for a and b are denoted L_a and L_b respectively. The upper bound on $|a(t,y)|$ is denoted M_a.

Proposition 3. *With the assumptions (9) the RL_k scheme is stable under perturbation.*

Corollary 4. *Assume that $a(t,y)$ and $b(t,y)$ are C^k regular and that $a(t,y)$ either is a diagonal or a constant matrix. In addition assume (9). Then the RL_k scheme is convergent of order k.*

Stability under perturbation together with consistency implies convergence, see *e.g.* [9] or [6] where the current setting has been detailed. Therefore corollary 4 is an immediate consequence of the propositions 1 and 3. Before to prove the proposition 3 definitions are needed.
Equation (3) is considered on $E = \mathbb{R}^N$ with the max norm $|\cdot|$. A final time $T > 0$ is considered. The space of $N \times N$ matrices is equipped with the operator norm $\|\cdot\|$ associated to $|\cdot|$. The space E^k is equipped with the max norm $|Y|_\infty = \max_{1\le i\le k} |y_i|$ with $Y = (y_1, \dots, y_k)$.
The RL_k scheme is defined with the mapping,

$$s_{t,h} :\ Y = (y_1, \dots, y_k) \in E^k \longrightarrow s_{t,h}(Y) \in E,$$

with,

$$s_{t,h} = y_k + h\varphi_1(\alpha_{t,h}(Y)h)\left(\alpha_{t,h}(Y)y_k + \beta_{t,h}(Y)\right),$$

in such a way that $y_{n+1} = s_{t_n,h}(y_{n-k+1}, \ldots, y_n)$ in (7). The functions $\alpha_{t,h}$ and $\beta_{t,h}$ are given in definition 1. For instance, $\alpha_{t,h}(Y)$ for the RL_3 scheme reads,

$$\alpha_{t,h}(Y) = \frac{1}{12}(23a(t, y_3) - 16a(t - h, y_2) + 5a(t - 2h, y_1)), \quad Y = (y_1, y_2, y_3).$$

A first way to prove the stability under perturbation is to show that $s_{t,h}$ is globally Lipschitz in Y. For this the derivative $\partial_Y s_{t,h}$ has to be analyzed. As developed in the remark 2, this will imply restrictions on $a(t, y)$: either diagonal or constant.
A second way is to prove the two following stability conditions,

$$|s_{t,h}(Y) - s_{t,h}(Z)| \leq |Y - Z|_\infty \left(1 + Ch(|Y|_\infty + 1)\right), \tag{10}$$

$$|s_{t,h}(Y)| \leq |Y|_\infty(1 + Ch) + Ch \tag{11}$$

where C is a constant only depending on the data a, b, y_0 in equation (4) and on the final time T. These are sufficient conditions for the stability under perturbation, as proved in [6].
That second way will be used here, for it is more general and giving rise to less computations. The core of the proof is the following property of the RL_k scheme. For $Y = (y_1, \ldots, y_k) \in E^k$:

$$s_{t,h}(Y) = z(t + h) \quad \text{for} \quad z' = \alpha_{t,h}(Y)z + \beta_{t,h}(Y), \quad z(t) = y_k. \tag{12}$$

It will be used together with the following Gronwall inequality (see [8, Lemma 196, p.150]). Suppose that $z(t)$ is a C^1 function and that there exist $M_1,\ M_2 > 0$ such that $|z'(t)| \leq M_1(t - t_0) + M_2$ for all $t \in [t_0, t_0 + h]$. Then,

$$\forall\ t \in [t_0, t_0 + h], \quad |z(t)| \leq \mathrm{e}^{M_1(t-t_0)}\left(|z(t_0)| + M_2(t - t_0)\right). \tag{13}$$

Proof of proposition 3. In this proof it is always assumed that $0 \leq h, t \leq T$. We will denote by C_i a constant only depending on the problem (4) data a and b and on T.

With the assumptions (9) and the definition 1, the function $\alpha_{t,h}$ is uniformly Lipschitz with a Lipschitz constant L_α. Moreover we have a uniform bound $\|\alpha_{t,h}\| \leq M_\alpha$. Since $b(t, y)$ is uniformly Lipschitz in y and since $0 \leq t \leq T$, there exists a constant K_b so that,

$$|b(t, y)| \leq K_b(1 + |y|). \tag{14}$$

Consider the RL_3 scheme,

$$\begin{aligned} |\beta_{t,h}(Y)|_\infty &\leq \frac{11}{3}K_b(1 + |Y|_\infty) + \frac{h}{12}M_a 2K_b(1 + |Y|_\infty) \\ &\leq C_1(1 + |Y|_\infty). \end{aligned}$$

The same inequality holds for the RL_2 and RL_4 schemes. With (12) we have $s_{t,h}(Y) = z(t+h)$ and,

$$|z'| = |\alpha_{t,h}(Y)z + \beta_{t,h}(Y)| \leq M_\alpha |z| + C_1(1 + |Y|_\infty).$$

The initial state is $|z(t)| = |y_k| \leq |Y|_\infty$. With the Gronwall inequality (13) we obtain for $t \leq \tau \leq t+h$,

$$\begin{aligned} |z(\tau)| &\leq \mathrm{e}^{M_\alpha h}\left(|Y|_\infty + hC_1(1+|Y|_\infty)\right) \\ &\leq \mathrm{e}^{M_\alpha h}\left(|Y|_\infty(1+C_1h) + C_1h\right) \\ &\leq |Y|_\infty(1+C_2h) + C_2h, \end{aligned} \tag{15}$$

by bounding the exponential with an affine function for $0 \leq h \leq T$. This gives the stability condition (11) for $\tau = t+h$.

For the RL_2 scheme $\beta_{t,h}$ is uniformly Lipschitz.
For the RL_3 scheme, consider $Y = (y_1, y_2, y_3)$ and $Z = (z_1, z_2, z_3)$ in E^3. We have,

$$\begin{aligned} |\beta_{t,h}(Y) - \beta_{t,h}(Z)|_\infty \leq \frac{11}{3}L_b|Y-Z|_\infty + \frac{h}{12}\big(&|a(t,y_3)b(t-h,y_2) - a(t,z_3)b(t-h,z_2)| \\ &+ |a(t-h,y_2)b(t,y_3) - a(t-h,z_2)b(t,z_3)|\big) \end{aligned}$$

Let us bound the Lipschitz constant for a function of the type $F(Y) = a(\xi, y_2)b(\tau, y_3)$ for $0 \leq \tau, \xi \leq T$.

$$\begin{aligned} |F(Y) - F(Z)| &= |a(\xi,y_3)\left(b(\tau,y_2) - b(\tau,z_2)\right) + \left(a(\xi,y_3) - a(\xi,z_3)\right)b(\tau,z_2)| \\ &\leq M_aL_b|Y-Z|_\infty + L_a|Y-Z|_\infty|b(\tau,z_2)|. \end{aligned}$$

With (14), this yields for $0 \leq \tau, \xi \leq T$ and for $Y,\ Z \in E^k$,

$$|F(Y) - F(Z)| \leq C_3|Y-Z|_\infty(1+|Z|_\infty),$$

As a result,

$$|\beta_{t,h}(Y) - \beta_{t,h}(Z)|_\infty \leq C_4|Y-Z|_\infty\left(1+|Z|_\infty\right)$$

The same inequality holds for the RL_4 scheme.

Finally consider $Y_1,\ Y_2 \in E^k$ and denote $\alpha_i = \alpha_{t,h}(Y_i)$, $\beta_i = \beta_{t,h}(Y_i)$. With the property (12), $s_{t,h}(Y_1) - s_{t,h}(Y_2) = (z_1 - z_2)(t+h)$ where z_i is the solution to,

$$z_i' = \alpha_i z_i + \beta_i, \quad z_i(t) = Y_{i,k}.$$

On the first hand, with the inequality (15), we have $|z_2(\tau)| \leq C_5(1+|Y_2|_\infty)$ for $t \leq \tau \leq t+h$.

On the second hand, on $[t, t+h]$,

$$\begin{aligned}|(z_1 - z_2)'| &\leq |\alpha_1||z_1 - z_2| + |\alpha_1 - \alpha_2||z_2| + |\beta_1 - \beta_2| \\ &\leq M_\alpha |z_1 - z_2| + L_\alpha |Y_1 - Y_2|_\infty C_5(1 + |Y_2|_\infty) + C_4|Y_1 - Y_2|_\infty(1 + |Y_2|_\infty) \\ &\leq M_\alpha |z_1 - z_2| + C_6|Y_1 - Y_2|_\infty(1 + |Y_2|_\infty)\end{aligned}$$

The initial condition is $|(z_1 - z_2)(t)| = |Y_{1,k} - Y_{2,k}| \leq |Y_1 - Y_2|_\infty$. Then the Gronwall inequality (13) yields,

$$\begin{aligned}|(z_1 - z_2)(t+h)| &\leq \mathrm{e}^{M_\alpha h}\left(|Y_1 - Y_2|_\infty + hC_6|Y_1 - Y_2|_\infty(1 + |Y_2|_\infty)\right) \\ &\leq \mathrm{e}^{M_\alpha h}\,|Y_1 - Y_2|_\infty\left(1 + C_6 h(1 + |Y_2|_\infty)\right).\end{aligned}$$

This last inequality implies the stability condition (10), again by bounding the exponential with an affine function for $0 \leq h \leq T$. □

4. Dahlquist stability

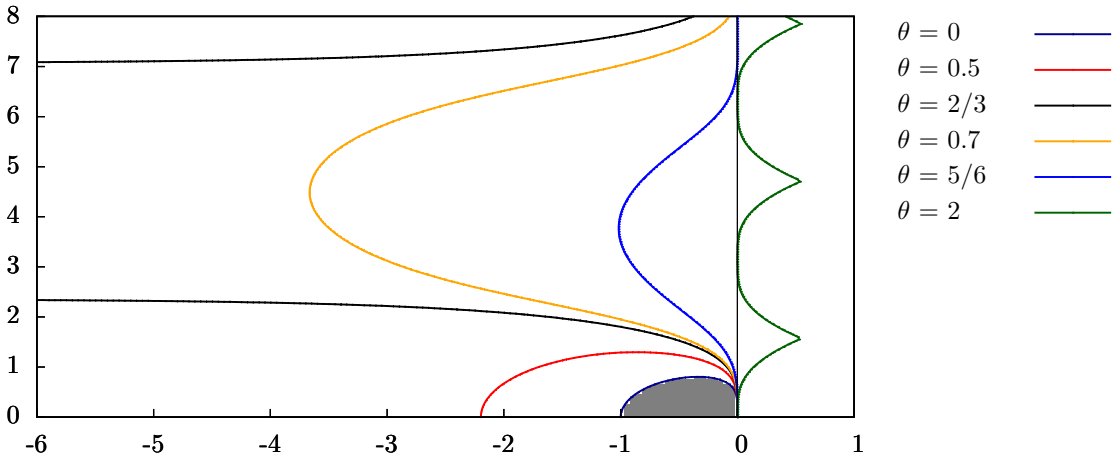


Figure 1. Stability domain D_θ for the RL_2 scheme for various values of θ. The stability domain for the particular case $\theta = 0$ (no stabilization) is in grey, corresponding to the Adams Bashforth scheme of order 2.

For the general definition concerning the Dahlquist stability we refer to [10]. The background for the Dahlquist stability of exponential integrators with a general varying stabilizer $a(t, y)$ has been developed in [6], following the ideas of Perego

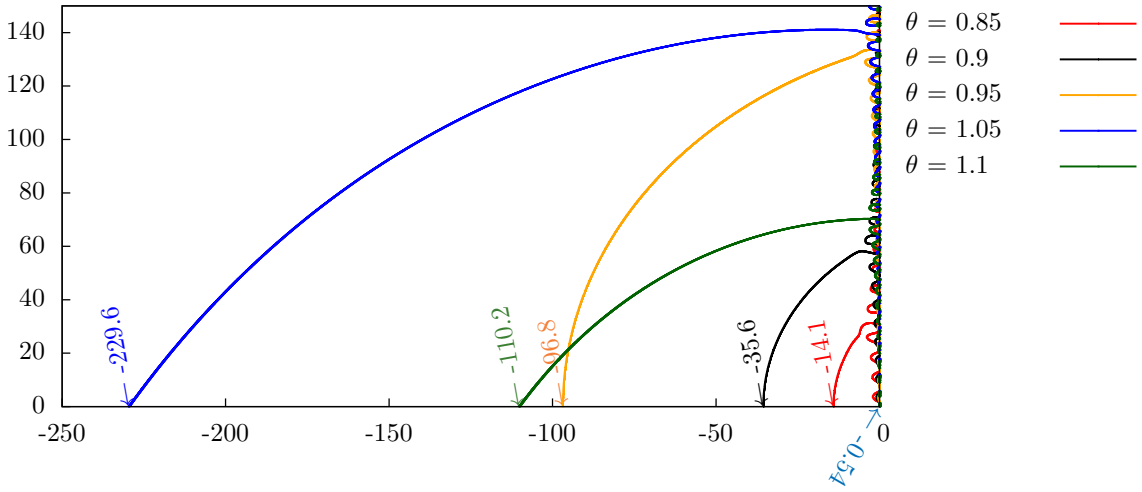


FIGURE 2. Stability domain D_θ for the RL_3 scheme. In the particular case $\theta = 0$ (no stabilization, corresponding to the Adams Bashforth scheme of order 3), the stability domain crosses the x-axis at $x \simeq -0.54$ (dark blue arrow).

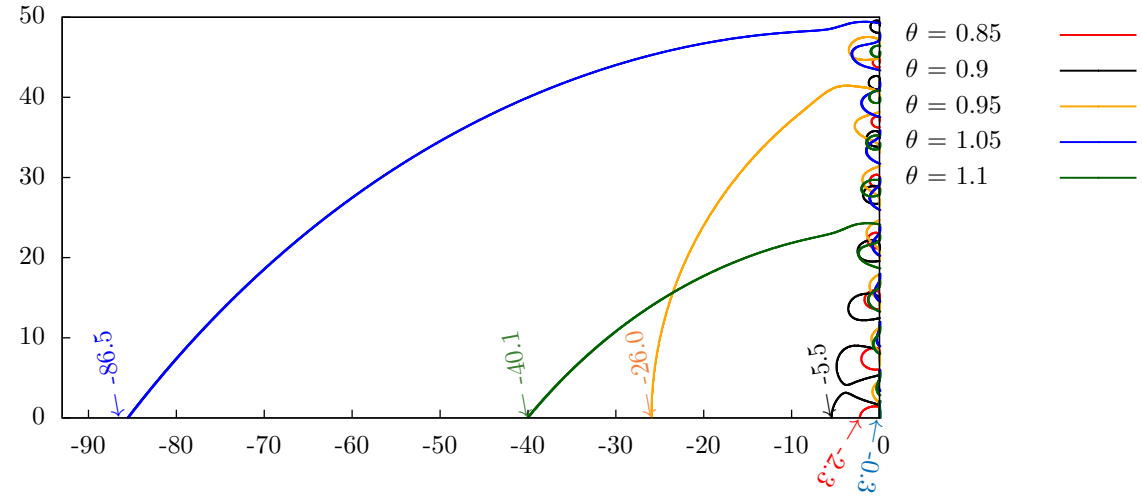


FIGURE 3. Stability domain D_θ for the RL_4 scheme. In the particular case $\theta = 0$ (no stabilization, corresponding to the Adams Bashforth scheme of order 3), the stability domain crosses the x-axis at $x \simeq -0.3$ (dark blue arrow).

and Veneziani [24]. Problem (3) is considered with the Dahlquist test function $f(t, y) = \lambda y$ that is decomposed in (4) as $f(t, y) = a(t, y)y + b(t, y)$ with,

$$a(t, y) = \theta\lambda, \quad b(t, y) = \lambda(1 - \theta)y.$$

When $\theta \simeq 1$, the exact linear part of $f(t, y)$ in equation (3) is well approximated by $a(t, y)$. The stability domain depends on θ. It is denoted D_θ. At a fixed value of θ, it is given by the modulus of a stability function, with the same definition as for multistep methods, see *e.g.* [10]. That stability function has been numerically computed pointwise on a grid inside the complex plane $\mathbb{C}$.

Order 2 Rush Larsen. The stability domain for the RL_2 scheme has been analyzed in [24]. The situation for this scheme is interesting and we reproduced the results on figure 1.

- If $0 \leq \theta < 2/3$ the stability domain D_θ is bounded. Its size increases with θ, starting from the Adams Bashforth scheme of order 2 stability domain when no stabilization occurs ($\theta = 0$).
- If $\theta = 2/3$, D_θ contains the negative real axis: the method is $A(0)$ stable. The domain boundary is asymptotically parallel to the real axis so that the method is not $A(\alpha)$ stable.
- If $\theta > 2/3$, the stability domain is located around the y-axis: the method is $A(\alpha)$ stable. The angle α increases with θ, it goes to $\pi/2$ as $\theta \to 1^-$.

Rush Larsen of order 3 and 4. The situation is different for the Rush Larsen methods of order 3 and 4. The stability domains D_θ for various values of θ have been numerically computed and depicted on figures 2 and 3.
Excepted for the case $\theta = 1$, the stability domains are always bounded: the schemes are not A-stable. However, this stability domain, for values of $\theta \simeq 1$ are much larger than the $D_{\theta|\theta=0}$ stability domain when no stabilization occurs (corresponding to the Adams Bashforth scheme of order 3 or 4). For the order 3 case, the stability domain for $\theta = 0.85$ is 25 times wider on the left than $D_{\theta|\theta=0}$, and for $\theta = 1.05$ it is 400 times wider. For the order 4 case, $D_{\theta|\theta=1.05}$ is almost 300 times wider on the left than $D_{\theta|\theta=0}$.

5. Numerical results

In this section are presented numerical experiments in order to investigate the performances of the RL_k method. It will be compared to the exponential integrator of Adams type of order $k5$), shortly denoted EAB_k here. The EAB_k schemes have been numerically studied in [6] as compared to several classical methods. It had been shown to be a good candidate for the resolution of the stiff membrane equation in cardiac electrophysiology.

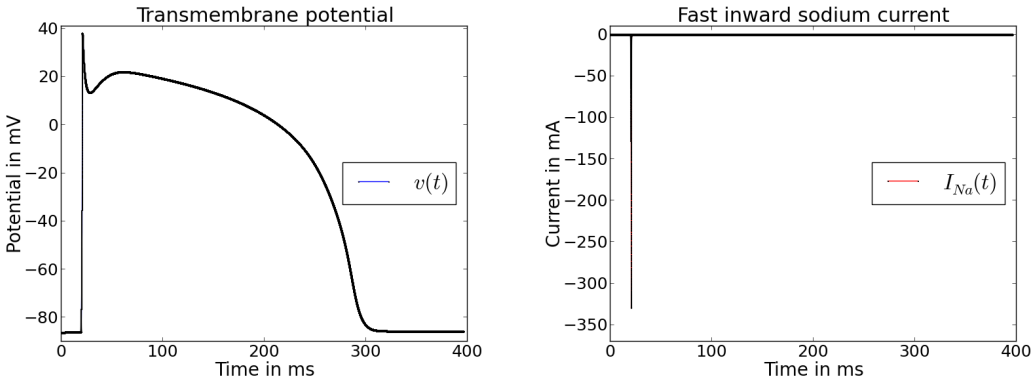


FIGURE 4. $TNNP$ model illustration. Left, cellular action potential: starting at a (negative) rest value, the membrane potential $v(t)$ has a stiff depolarization followed by a plateau and a repolarization to the rest value. Right, depolarization is induced by an ionic sodium current I_{Na}, with visible stiffness.

5.1. **The membrane equation.** We consider a class of models in cardiac electrophysiology. As illustrated on figure 4, these models display a stiff behaviour characterized by the presence of heterogeneous time scales. The models used to simulate the electrical activity of cardiac cells are ODE systems of the form, see [17, 1, 22, 27]:

$$\begin{aligned} \frac{dw_i}{dt} &= \frac{w_{\infty,i}(v) - w_i}{\tau_i(v)}, \qquad \frac{dc}{dt} = g(w,c,v), \\ \frac{dv}{dt} &= -I_{ion}(w,c,v) + I_{st}(t), \end{aligned} \tag{16}$$

where $w = (w_1, \ldots, w_p) \in \mathbb{R}^p$ is the vector of the gating variables, $c \in \mathbb{R}^q$ is a vector of ionic concentrations or other state variables, and $v \in \mathbb{R}$ is the cell membrane potential. These equations model the evolution of the transmembrane potential of a single cardiac cell. The four functions $w_{\infty,i}(v)$, $\tau_i(v)$, $g(w,c,v)$ and $I_{ion}(w,c,v)$ are given reaction terms. They characterize the cell model. The function $I_{st}(t)$ is a source term. It represents a stimulation current. Problem (16) reformulates into problem (4) form with:

$$a(t,y) = \begin{pmatrix} -1/\tau(v) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad b(t,y) = \begin{pmatrix} w_\infty(v)/\tau(v) \\ g(y) \\ -I_{ion}(y) + I_{st}(t) \end{pmatrix}, \tag{17}$$

for $y = (w,c,v) \in \mathbb{R}^N$ ($N = p+q+1$) and where $-1/\tau(v)$ the $p \times p$ diagonal matrix with diagonal entries $(-1/\tau_i(v))_{i=1\ldots p}$. The resulting matrix $a(t,y)$ is diagonal.

We will consider two such models: the *BeelerReuter* model (BR) [1] or the TNNP model [27] for human cardiac cells.

5.2. **Convergence.** No theoretical solution are available for the chosen application. A reference solution y_{ref} for a reference time step h_{ref} is computed with the Runge Kutta 4 scheme to analyze the convergence properties of the RL_k scheme. Numerical solutions y are computed to y_{ref} for coarsest time steps $h = 2^p h_{ref}$ for increasing p. Any numerical solution y consists in successive values y_n at the time instants $t_n = nh$. On every interval (t_{3n}, t_{3n+3}) the polynomial $\overline{y}$ of degree at most 3 so that $\overline{y}(t_{3n+i}) = y_{3n+i}$, $i = 0 \dots 4$ is constructed. On $(0, T)$, $\overline{y}$ is a piecewise continuous polynomial of degree 3. Its values at the reference time instants nh_{ref} are computed. This provides a projection $P(y)$ of the numerical solution y on the reference grid. Then $P(y)$ can be compared with the reference solution y_{ref}. The numerical error is defined by,

$$e(h) = \frac{\max |v_{ref} - P(v)|}{\max |v_{ref}|}, \tag{18}$$

where the potential v is the last and stiffest component of y in equation (16). The

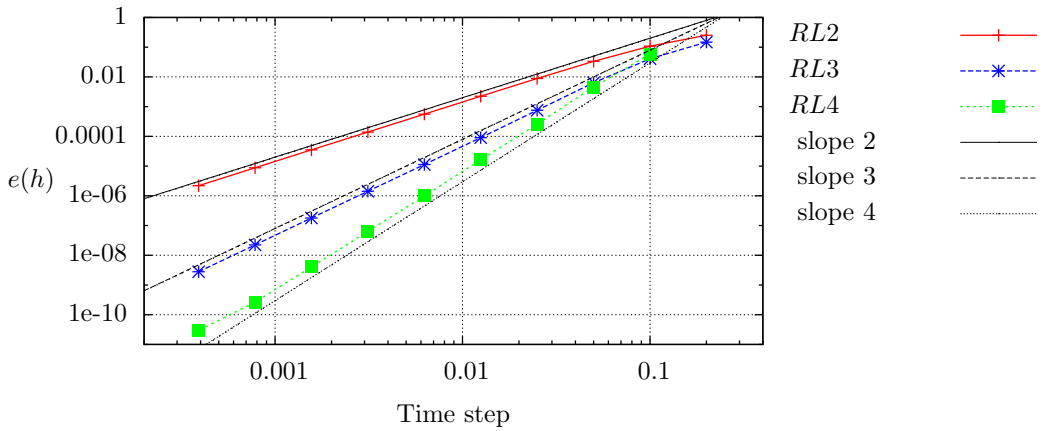


FIGURE 5. Relative error $e(h)$ (definition (18)) as a function of the time step h for the RL_k schemes, for $k = 2$ to 4 and in Log/Log scale.

numerical convergence graphs for the BR model are plotted on figure 5. All the schemes display the expected asymptotic behaviour $e(h) = O(h^k)$ as $h \to 0$ in corollary 4.

5.3. **Stability robustness to stiffness.** In [26] has been evaluated the stiffness of the BR and TNNP models along one cellular electrical cycle (as depicted on figure 4). The largest negative real part of the eigenvalues of the Jacobian matrix during

this cycle is of -1170 and -82 for the TNNP and BR models respectively. The TNNP model thus is 15 times stiffer than the BR model ($15 \simeq 1170/82$).
Robustness to stiffness for the RL_k scheme is evaluated by comparing the critical time step for these two models.The critical time step Δt_0 is defined as the largest time step such that the numerical simulation runs without overflow for $h < \Delta t_0$. The results are presented in table 1.

method	RL_2	RL_3	RL_4	EAB_2	EAB_3	EAB_4
BR	0.323	0.200	0.149	0.424	0.203	0.123
TNNP	0.120	0.148	0.111	0.233	0.108	$7.56\ 10^{-2}$

TABLE 1. Critical time step Δt_0 for the RL_k and EAB_k schemes

An excellent robustness to stiffness can be observed. The critical time step is divided by 2.7, 2.0 and 1.3 for $k = 2$, 3 and 4 respectively. A comparison with the EAB_k schemes shows that the two scheme display a robustness to stiffness of same order. For a method that is not $A(\alpha)$ stable, it is expected for the critical time step to be divided by 15 in case of an increase of stiffness of magnitude 15. This is not observed here, though the RL_k scheme is not $A(\alpha)$ stable. The reason for this is that the ODE system (16) is only partially stabilized by (17). Loss of stability is induced by the non-stabilized part, whose eigenvalues are less modified between the BR and the $TNNP$ models.

5.4. **Accuracy.** The RL_k scheme is here compared to the EAB_k scheme in terms of accuracy. This comparison is done by computing the relative error $e(h)$ in equation (18). The two BR and $TNNP$ models are considered. We recall than the $TNNP$ model is stiffer by a factor 15. The results are collected in the tables 2 and 3.

h	RL_2	RL_3	RL_4	EAB_2	EAB_3	EAB_4
0.2	0.251	0.147	-	0.284	0.516	-
0.1	0.107	$4.07\ 10^{-2}$	$5.86\ 10^{-2}$	$9.26\ 10^{-2}$	$9.17\ 10^{-2}$	0.119
0.05	$3.35\ 10^{-2}$	$6.34\ 10^{-3}$	$4.58\ 10^{-3}$	$2.31\ 10^{-2}$	$1.09\ 10^{-2}$	$8.96\ 10^{-3}$
0.025	$8.88\ 10^{-3}$	$7.57\ 10^{-4}$	$2.61\ 10^{-4}$	$5.39\ 10^{-3}$	$1.17\ 10^{-3}$	$4.33\ 10^{-4}$

TABLE 2. Relative error $e(h)$ (eq. (18)) for the BR model.

For the RL_2 and the EAB_2 schemes, the accuracies are very close, the EAB_2 scheme being slightly more accurate for the BR model. For the orders 3 and 4, a non negligible difference is observed between the RL and EAB schemes. The RL scheme is more accurate at large time steps. For smaller time steps, accuracies are almost

h	RL_2	RL_3	RL_4	EAB_2	EAB_3	EAB_4
0.1	0.177	0.305	0.421	0.351	0.530	-
0.05	$7.39\ 10^{-2}$	$4.54\ 10^{-2}$	$4.61\ 10^{-2}$	$9.01\ 10^{-2}$	$5.59\ 10^{-2}$	$8.93\ 10^{-2}$
0.025	$2.21\ 10^{-2}$	$6.53\ 10^{-3}$	$5.96\ 10^{-3}$	$2.14\ 10^{-2}$	$7.34\ 10^{-3}$	$8.34\ 10^{-3}$
0.0125	$5.75\ 10^{-3}$	$8.05\ 10^{-4}$	$3.21\ 10^{-4}$	$5.11\ 10^{-3}$	$7.62\ 10^{-4}$	$3.70\ 10^{-4}$

TABLE 3. Relative error $e(h)$ (eq. (18)) for the $TNNP$ model.

the same. This means that RL and EAB schemes are equivalent in terms of accuracy considering the asymptotic convergence region, but outside this region, RL scheme is more precise.

6. Conclusion

We introduced in this paper two new ODE solvers that we named Rush Larsen of order 3 and 4. They are explicit multistep exponential integrators. Their general definition (7) is very simple inducing an easy implementation. We provided a convergence and stability under perturbation analysis of these two schemes. We also performed their Dahlquist stability analysis: they are not $A(0)$ stable but display a very large stability domain for sufficiently precise stabilization. The numerical properties of the schemes are analyzed for a complex and realistic stiff application. The RL_k schemes are as stable as exponential integrators of Adams type, allowing simulations at large time step. The RL_k schemes moreover are more accurate for $k = 3$ and 4 when considering large time steps. The are shown to be robust to stiffness both in terms of stability and of accuracy.

References

[1] G.W. Beeler and H Reuter. Reconstruction of the Action Potential of Ventricular Myocardial Fibres. *J. Physiol.*, 268:177–210, 1977.

[2] M. T. Chu. An automatic multistep method for solving stiff initial value problems. *J. Comput. Appl. Math.*, 9(3):229–238, 1983.

[3] J.C. Clements, J. Nenonen, P.K. Li, and B.M.. Horacek. Activation dynamics in anisotropic cardiac tissue via decoupling. *Ann. Biomed. Eng.*, 32(7):984–990, 2004.

[4] P. Colli-Franzone, L.F. Pavarino, and B. Taccardi. Monodomain simulations of excitation and recovery in cardiac blocks with intramural heterogeneity. *in Functional Imaging and Modeling of the Heart (FIMH05), Lect. Notes Comput. Sci.*, 3504:267–277, 2005.

[5] P. Colli-Franzone, L.F. Pavarino, and B. Taccardi. Simulating patterns of excitation, repolarization and action potential duration with cardiac Bidomain and Monodomain models. *To appear in Math. Biosci.*, 2005.

[6] Y. Coudiére, C. Douanla Lontsi, and C. Pierre. Exponential Adams Bashforth integrators for stiff ODEs, application to cardiac electrophysiology. *HAL Preprint*, 2016.

[7] S. M. Cox and P. C. Matthews. Exponential time differencing for stiff systems. *J. Comput. Phys.*, 176(2):430–455, 2002.
[8] Sever Silvestru Dragomir. *Some Gronwall type inequalities and applications*. Nova Science Publishers, Inc., Hauppauge, NY, 2003.
[9] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving ordinary differential equations. I*, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1993.
[10] E. Hairer and G. Wanner. *Solving ordinary differential equations. II*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2010.
[11] M. Hochbruck and A. Ostermann. Explicit Exponential Runge-Kutta Methods for Semilinear Parabolic Problems. *SIAM J. Numerical Analysis*, 43(3):1069–1090, 2005.
[12] Marlis Hochbruck. A short course on exponential integrators. In *Matrix functions and matrix equations*, volume 19 of *Ser. Contemp. Appl. Math. CAM*, pages 28–49. Higher Ed. Press, Beijing, 2015.
[13] Marlis Hochbruck, Christian Lubich, and Hubert Selhofer. Exponential integrators for large systems of differential equations. *SIAM J. Sci. Comput.*, 19(5):1552–1574 (electronic), 1998.
[14] Marlis Hochbruck and Alexander Ostermann. Exponential integrators. *Acta Numer.*, 19:209–286, 2010.
[15] Marlis Hochbruck and Alexander Ostermann. Exponential multistep methods of Adams-type. *BIT*, 51(4):889–908, 2011.
[16] Marlis Hochbruck, Alexander Ostermann, and Julia Schweitzer. Exponential Rosenbrock-type methods. *SIAM J. Numer. Anal.*, 47(1):786–803, 2008/09.
[17] A.L. Hodgkin and A.F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117:500–544, 1952.
[18] Antti Koskela and Alexander Ostermann. Exponential Taylor methods: analysis and implementation. *Comput. Math. Appl.*, 65(3):487–499, 2013.
[19] D. Lee and S. Preiser. A class of non linear multistep A-stable numerical methods for solving stiff differential equations. *Comp. & maths with appls.*, 4:43–51, 1978.
[20] V. T. Luan and A. Ostermann. Explicit exponential runge kutta methods of high order for parabolic problems. *J. Comput. Appl. Math.*, 256:168–179, 2014.
[21] C.H. Luo and Y. Rudy. A model of the Ventricular Cardiac Action Potential. *Circ. Res.*, 68:1501–1526, 1991.
[22] C.H. Luo and Y. Rudy. A Dynamic Model of the Cardiac Ventricular Action Potential I. Simulations of Ionic Currents and Concentration Changes. *Circ. Res.*, 74:1071–1096, 1994.
[23] B.V. Minchev and W. M. Wright. A review of exponential integrators for first order semi-linear problems. Technical report, Norwegian universtity of science and technology trondhem, 2005.
[24] M. Perego and A. Veneziani. An efficent generalization of the Rush-Larsen method for solving electro-physiology membrane equations. *ETNA*, 35:234–256, 2009.
[25] G. Rainwater and M. Tokman. A new class of split exponential propagation iterative methods of Runge-Kutta type (sEPIRK) for semilinear systems of ODEs. *J. Comput. Phys.*, 269:40–60, 2014.
[26] R. J. Spiteri and C. D. Ryan. Stiffness Analysis of Cardiac Electrophysiological Models. *Annals of Biomedical Engineering*, 38:3592–3604, Dec 2010.
[27] K.H. Ten Tusscher, D. Noble, P.J. Noble, and A.V. Panfilov. A Model for Human Ventricular Tissue. *Am J Physiol Heart Circ Physiol*, 286, 2004.
[28] M. Tokman, J. Loffeld, and P. Tranquilli. New adaptive exponential propagation iterative methods of Runge-Kutta type. *SIAM J. Sci. Comput.*, 34(5):A2650–A2669, 2012.

[29] Paul Tranquilli and Adrian Sandu. Rosenbrock-Krylov methods for large systems of differential equations. *SIAM J. Sci. Comput.*, 36(3):A1313–A1338, 2014.

Yves Coudière, INRIA Bordeaux Sud Ouest, Université de Bordeaux
E-mail address: yves.coudiere@inria.fr

Charlie Douanla Lontsi, INRIA Bordeaux Sud Ouest, Université de Bordeaux
E-mail address: charlie.douanla-lontsi@inria.fr

Charles Pierre, CNRS, Université de Pau, LMAP
E-mail address: charles.pierre@univ-pau.fr

5 Étude numérique des propriétés des schémas classiques et exponentiels sur l'équation de la membrane

Contents

Nous allons dans cette partie faire une étude en 0D de l'efficacité des schémas EAB, RL tout en les comparant d'une part entre eux, et d'autre part en les comparant à certains schémas classiques. Pour cela nous auront besoin pour des raisons de lisibilité de rappeler des formulations qui ont déjà été données dans les chapitres précédents. Il s'agit de la formulation générale des ODE et en particulier des modèles ioniques que nous allons utiliser tout au long de ce chapitre. On considère donc le problème de Cauchy,

$$\frac{dy}{dt} = f(t,y), \quad t \in]0,T], \quad y(0) = y_0 \in \mathbb{R}^N. \tag{5.1}$$

On a vu dans les deux chapitres précédents que les EDO provenant des modèles ioniques qui nous intéressent pouvaient être écrites sous la forme,

$$\frac{dy}{dt} = a(y)y + b(t,y), \quad t \in]0,T], \quad y(0) = y_0 \in \mathbb{R}^N \tag{5.2}$$

avec,

$$a(t,y) = \begin{pmatrix} -1/\tau(v) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad b(t,y) = \begin{pmatrix} w_\infty(v)/\tau(v) \\ g(y) \\ -I_{ion}(y) + I_{st}(t) \end{pmatrix}, \tag{5.3}$$

pour $y = (w, c, v) \in \mathbb{R}^N$ ($N = p + q + 1$) et où $-1/\tau(v)$ est la matrice diagonale $p \times p$ de diagonale $(-1/\tau_i(v))_{i=1\ldots p}$.

5.1 Temps d'activation, de récupération et APD

Soit u le potentiel transmembranaire et u_r le potentiel de repos précédemment décrits au chapitre 2, particulièrement dans les sections 2.1.1 et 2.1.2. On associe au potentiel de repos u_r un potentiel de pic u_p correspondant au maximum du potentiel u en fin d'excitation et un potentiel de seuil u_{th} tel que $u_r < u_{th} < u_p$ (voir figure 5.1). Nous adoptons les définitions suivantes :

— Potentiel seuil u_{th} : Potentiel correspondant à 80% de l'amplitude $u_p - u_r$ du potentiel transmembranaire.

— Temps d'activation t_a : Instant auquel le potentiel trans-membranaire u atteint la valeur u_{th} pour la première fois.

— Temps de récupération t_r : Instant auquel le potentiel trans-membranaire u atteint la valeur u_{th} pour la seconde fois.

— Durée du potentiel d'action APD : Temps moyen pendant lequel le voltage reste au dessus de la valeur u_{th}.

Ces définitions se traduisent par les égalités suivantes,

$$u_{th} = 0.8u_r + 0.2u_p, \quad u(t_a) = u_{th} = u(t_r),\ t_a < t_r \quad \text{et} \quad \text{APD} = t_r - t_a. \tag{5.4}$$

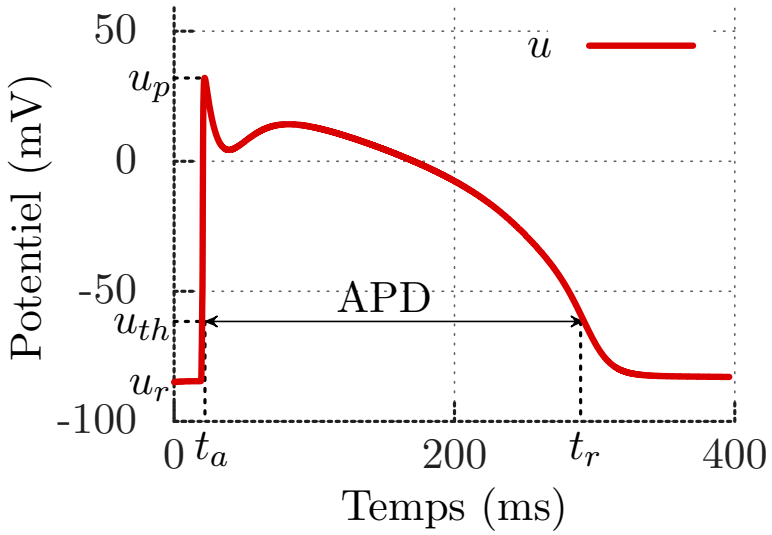


FIGURE 5.1: Temps d'activation t_a, de repolarisation t_r et durée du potentiel d'action APD

5.2 Description des outils d'analyse et méthodologie

Nous présentons dans cette section les outils qui nous permettrons de comparer les différentes méthodes de résolution d'ODE. Nous commençons par lister et commenter brièvement ces concepts et apportons plus de détails par la suite.

1. **Cas test.** La comparaison des solutions numériques se fera avec la solution $y(t)$ de (5.1). Elle sera notée ainsi dans toute la suite pour chaque cas test considéré. Le cas test n'est relatif qu'au choix du modèle. Les comparatifs des schémas se feront dans le cadre restrictif d'un type d'application (l'électrophysiologie cellulaire) et d'un phénomène associé (le potentiel d'action).

2. **Précision.** La notion de précision sera centrale ici et il convient d'en dégager plusieurs aspects. La définition précise de l'erreur sera donnée en section 5.2.5.
   Premièrement la précision est définie comme une erreur entre la solution numérique et la solution exacte $t \to y(t)$ du problème (5.1). Cependant la solution exacte y n'est pas connue dans notre cas. L'erreur sera donc toujours calculée relativement à une *solution de référence* définie en section 5.2.2. Cette solution de référence y_{ref} sera assimilée à la solution exacte y quoiqu'elle soit elle même une solution numérique. On précise que pour des raisons pratiques ce ne sera pas la même solution de référence qui sera utilisée pour tous les schémas, mais ce sera toujours la même méthode qui servira à calculer les solutions de référence.
   La précision est étudiée d'abord pour valider le comportement asymptotique lorsque le pas de temps $\Delta t \longrightarrow 0$. C'est à dire que l'on vérifie pour une erreur donnée e que $e(\Delta t) = O(\Delta t^k)$ où k est l'ordre mathématiquement attendu pour la méthode considérée.
   Il est enfin fondamental d'évaluer la précision des méthodes par rapport à des critères physiologiques. Nous considérerons pour cela les erreurs commises sur les temps

d'activation, de récupération et sur l'APD tels que définis en 5.1.
Les erreurs physiologiques fournissent des critères pertinents en pratique pour comparer les résultats des différentes méthodes.

3. **Coût.** La précision ne prend tout son sens que quand on lui associe un coût, ici il s'agit du coût en temps de calcul. Ce coût est relatif à la machine utilisée, à l'implémentation et à son optimisation.
   La définition du coût en terme de temps de calcul n'est pas la seule alternative pour évaluer le coût. Une autre définition du coût basée sur le nombre d'évaluation des fonctions qui décrivent les modèles sera donnée plus bas.

4. **Souplesse et polyvalence.** La facilité d'utilisation d'une méthode est un autre critère important. Dans l'idéal on recherche des méthodes avec peu de paramètres et surtout robustes par rapport au réglage de ces paramètres.
   Par ailleurs il ne faut pas perdre de vue qu'au delà du problème 0D considéré on vise des extensions pour des systèmes d'EDP paraboliques. La facilité de l'extension des méthodes considérées à ces systèmes d'EDP est fondamentale.
   Avoir un grand pas de temps critique est intéressant à double titre. D'abord on diminue le risque d'échec ou d'artefact numérique sur une plus grande gamme de pas de temps. Mais surtout, dans l'optique du couplage avec un système d'EDP parabolique, travailler à grand pas de temps implique une diminution du coût associé à la résolution des EDP.
   Grand et *petit* pas de temps sont des notions relatives. Pour l'application considérée on donnera comme ordre de grandeur que $\Delta t < 0.01ms$ est un petit pas de temps et que $\Delta t \geq 0.01ms$ est un pas de temps moyen à grand.

5.2.1 Cas test

Dans toute la suite on note $y(t)$ la solution de (5.1) pour le modèle de TNNP [5] ou pour le modèle de BR [1] sur l'intervalle $(0, T)$ avec $T = 396$ ms. Le choix d'une telle valeur pour T permet d'observer un potentiel d'action. Il permet d'avoir tout le temps des maillages de $(0, T)$ imbriqués (voir figure 5.3) les uns dans les autres. Travailler sur des maillages imbriqués nous permet de travailler sur un maillage de référence qui contient tous les points des maillages grossiers que nous utilisons dans nos études. La solution y est uniquement définie une fois fixés la condition initiale y^0 et le courant de stimulation I_{st} dans (5.3) (le terme source) :

— y^0 est l'état de repos qui doit vérifier $f(t, y^0) = 0$. C'est un état d'équilibre pendant lequel toutes les variables du modèle restent à une valeur fixe pendant quelques millisecondes. Cet état d'équilibre est exploité dans notre travail pour initialiser les schémas multipas que nous étudions. La précision sur la valeur de y^0 est donc très importante pour l'ordre de convergence de ces schémas (multipas). Si la valeur de y^0 est trop approximative, les erreurs commises lors de l'initialisation du schéma numérique se répandent dans tous les calculs et ne permettent pas d'observer l'ordre

de convergence. Cependant, les valeurs initiales y^0 données dans les articles ne sont pas toujours assez précises. Pour chaque modèle utilisé, nous avons calculé la valeur initiale y^0 par un algorithme de Newton pour pouvoir observer les ordres de convergence des schémas d'ordre élevé.

— La fonction $I_{st}(t)$ est positive, nulle en dehors de l'intervalle $(t_s - 1, t_s + 1)$, t_s=20 ms, et d'intégrale $\int_0^T I_{st}(t)dt = I_{stim}$, un courant total typique de stimulation fixé par les modèles, de l'ordre de 50 mA.
On impose par ailleurs une régularité C^4 à I_{st} de manière à pouvoir observer une convergence des schémas jusqu'à l'ordre 4.

$$I_{st}(t) = I_{stim}\Psi(t - t_s), \tag{5.5}$$

avec,

$$\Psi(x) = \begin{cases} 1 - 630\left(\frac{1}{9}|x|^9 - \frac{1}{2}|x|^8 + \frac{6}{7}|x|^7 - \frac{2}{3}|x|^6 + \frac{1}{5}|x|^5\right), \text{ si } |x| \leq 1, \\ 0, \qquad \text{si non.} \end{cases}$$

La fonction Ψ est représentée sur la figure 5.2. Elle vérifie en particulier les relations,

$$\begin{aligned} &\Psi(1) = \Psi'(1) = \Psi''(1) = \Psi'''(1) = \Psi^4(1) = 0, \\ &\Psi'(0) = \Psi''(0) = \Psi'''(0) = \Psi^4(0) = 0, \\ &\Psi(0) = 1 \quad et \quad \int_{-1}^{1} \Psi(t)dt = 1. \end{aligned}$$

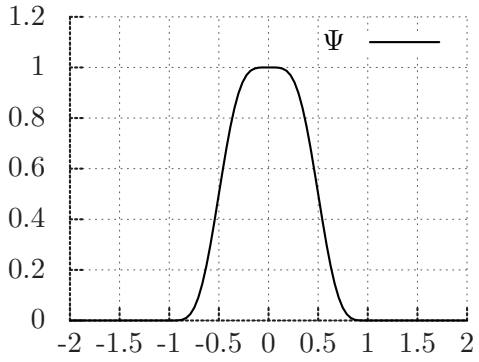


FIGURE 5.2: Fonction C^4 régularisant la stimulation.

5.2.2 Solutions numériques et solution de référence

Soit $m \geq 1$ un entier auquel on associe le pas de temps $\Delta t = T/m$ et le maillage régulier $\mathcal{T}_m = \{t_n = j\Delta t,\ j = 0 \dots m\}$ de l'intervalle $(0, T)$. On appellera solution numérique, que l'on notera (y^n) un élément de l'espace E_m.

$$E_m = \{(y^n)_{0 \leq n \leq m}, \quad y^n \in \mathbb{R}^N\}$$

L'espace E_m des solutions numériques est simplement $(\mathbb{R}^N)^{m+1}$ mais à $(y^n) \in E_m$ est implicitement associé un pas de temps Δt et un maillage $\mathcal{T}_m$, de sorte que chaque valeur y^n, $0 \leq n \leq m$ de $(y^n) \in E_m$ est censée être une approximation de $y(t_n)$.
Pour un cas test donné, puisque nous n'avons pas accès à la solution exacte $y(t)$ associée à l'EDO, nous utilisons une approximation de $y(t)$. Soit m' un entier divisible par 2^r, avec r suffisamment grand. La solution de référence $y_{ref} \in E_{m'}$, est une solution numérique calculée par le schéma de Runge Kutta 4 (5.19) et un pas de temps Δt_{ref}, pour le problème (5.2) associé au cas test considéré. Soit $m = 2^{-l}m'$, $0 \leq l \leq r$ et $(y^n) \in E_m$, une solution numérique calculée par un schéma donné et un pas de temps Δt. On a,

$$\Delta t_{ref} = T/m' = \Delta t/2^l. \tag{5.6}$$

En pratique r est choisi *suffisamment grand*, de telle sorte que l'erreur entre la solution exacte y et y_{ref} soit négligeable devant l'erreur entre la solution numérique (y^n) et y_{ref}.

5.2.3 Interpolation de la solution numérique

La solution de référence est calculée sur un maillage très fin (maillage de référence) par rapport à celui (maillage grossier) sur lequel est calculée une solution numérique quelconque. Le maillage de référence possède donc plus de points que le maillage grossier. Pour rendre possible la comparaison entre la solution de référence et une solution numérique, on doit interpoler la solution numérique. En considérant des maillages imbriqués, le maillage de référence $\mathcal{T}_{m'}$ contient tous les points du maillage grossier $\mathcal{T}_m$ (voir la figure 5.3).
En travaillant sur des maillages imbriqués, on peut se passer de l'interpolation et faire la comparaison uniquement sur les points du maillage grossier $\mathcal{T}_m$. En faisant la comparaison ainsi, une seule erreur est prise en compte : l'erreur sur les points du maillage grossier (erreur de discrétisation). La solution exacte étant définie sur un domaine continu, il est nécessaire d'interpoler la solution numérique l'approximant pour obtenir une fonction définie sur un domaine continue. Cette interpolation induit aussi une erreur dans l'approximation qui doit être prise en compte dans l'évaluation de l'erreur d'approximation. Pour tenir compte de cette erreur d'interpolation, on fait la comparaison sur les points du maillage de référence et non sur les points du maillage grossier, bien que la comparaison uniquement sur les points du maillage grossier soit plus facile. Les valeurs de la solution

numérique aux points du maillage de référence n'appartenant pas à l'ensemble des points du maillage grossier sont calculées en utilisant une interpolation que nous définissons dans la suite. On définit l'interpolateur,

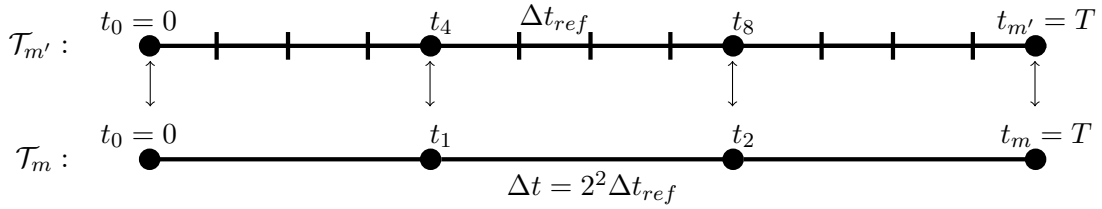


FIGURE 5.3: Deux maillages imbriqués : maillage de référence $\mathcal{T}_{m'}$ au dessus et maillage grossier $\mathcal{T}_m$ en bas. Le maillage $\mathcal{T}_{m'}$ est 4 fois plus fin que $\mathcal{T}_m$ et contient tous les points du maillage de grossier

$$\pi_{m,i} : E_m \longrightarrow C^0(0,T),$$

qui transforme la composante i de la solution numérique $(y^n) \in E_m$ en une fonction continue sur $[0,T]$. On demande par ailleurs à l'interpolant $\pi_{m,i}y^n$ d'être polynomiale de degré 3 par morceaux. Cette contrainte est nécessaire pour observer des convergences jusqu'à l'ordre 4.
Soit $(y^n) \in E_m$, l'interpolant $f = \pi_{m,i}y^n$ est construit de la manière suivante.

1- On décompose l'intervalle $[0,T]$ en une suite de paquets de 3 intervalles (figure 5.4) $P_s = [t_{3s}, t_{3s+1}] \cup [t_{3s+1}, t_{3s+2}] \cup [t_{3s+2}, t_{3(s+1)}]$, pour $s = 0, \ldots, m/3$.

2- On construit l'unique fonction f continue sur $[0,T]$, polynomiale de degré 3 sur chaque P_s et telle que $f(t_n) = y_i^n$ pour tout $n = 0 \ldots m$.

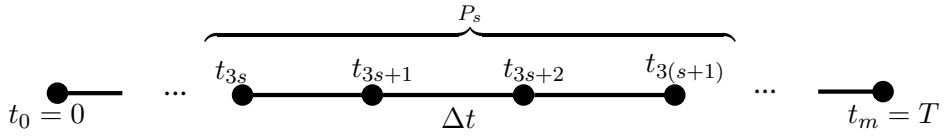


FIGURE 5.4: Bloc P_s de trois intervalles $[t_{3s}, t_{3s+1}], [t_{3s+1}, t_{3s+2}], [t_{3s+2}, t_{3(s+1)}]$ extrait d'un maillage de $[0,T]$ construit avec un pas de temps Δt

Pour des raisons pratiques, l'entier m sera toujours un multiple de 3. Un tel choix facilite le découpage de l'intervalle $[0,T]$ en paquets P_s précédemment décrits.
L'emphase sera ici mise sur le potentiel de membrane $u(t) = y_N(t)$ ($N^{\text{ième}}$ composante de y) et pour plus de simplicité on notera $\pi_m = \pi_{m,N}$ et $\pi = \pi_{m,N}$ en l'absence de confusion.

5.2.4 Calcul des temps d'activation, de repolarisation et de l'APD

On suppose que la solution est un potentiel d'action. On se donne une solution numérique (y^n). Soit $u^n = y_N^n$, $0 \leq n \leq m$, le potentiel de membrane à chaque instant t_n extrait de (y^n). Il existe alors deux indices uniques $n_a < n_r$ tels que,

$$u_{n_a+1} \leq u_{th} < u_{n_a+2} \quad , \qquad u_{n_r+1} \leq u_{th} < u_{n_r+2}.$$

u_{th} étant le potentiel seuil défini en section 5.1.
Sur les intervalles (t_{n_i}, t_{n_i+3}) , $i \in \{a, r\}$, on calcule le polynôme d'interpolation de Lagrange de degré 3 $p_i(t)$ pour les valeurs u^j associées aux instants t_j, $j = n_i, \ldots, n_i + 3$. On calcule alors les temps d'activation t_a et de récupération t_r comme les solutions de,

$$p_a(t_a) = u_{th}, \quad p_r(t_r) = u_{th}.$$

De nouveau le recours à une interpolation à l'ordre 3 est nécessaire pour pouvoir observer des ordres de convergence jusqu'à 4. Dans ce qui précède on suppose que tout est bien défini ce qui est le cas si la solution numérique (y^n) est physiologiquement pertinente (ie : si elle est un potentiel d'action unique.)

5.2.5 Calcul d'erreurs

Soit (y^n) une solution numérique et u_{ref} la solution de référence. On note respectivement par $\overline{u}_{ref} = \pi y_{ref}^n$ et $\overline{u} = \pi y^n$ le potentiel trans-membranaire donné par la composante N de la solution de référence y_{ref}, et l'interpolant du potentiel trans-membranaire donné par la composante N de la solution numérique (y^n). Soit Δt le pas de temps utilisé pour calculer la solution numérique (y^n). On définit l'erreur relative en norme L^∞ :

$$e_\infty(\Delta t) = \frac{\|\overline{u} - \overline{u}_{ref}\|_{L^\infty}}{\|\overline{u}_{ref}\|_{L^\infty}}. \tag{5.7}$$

Il est à noter que le choix du potentiel de membrane V (composante N de la solution numérique (y^n)) est arbitraire et que toute autre composante de (y^n) auraient pu être considérée.
Les erreurs relatives entre les temps d'activation, de repolarisation et l'APD prédites par une solution numérique (y^n) et une solution de référence y_{ref} seront calculées par ,

$$e_{t_a}(\Delta t) = \frac{|t_a - t_{a,ref}|}{|t_{a,ref}|}, \quad e_{t_r}(\Delta t) = \frac{|t_r - t_{r,ref}|}{|t_{r,ref}|}, \quad e_{\text{APD}}(\Delta t) = \frac{|\text{APD} - \text{APD}_{ref}|}{|, \text{APD}_{ref}|}, \tag{5.8}$$

5.2.6 Évaluation du coût

Le coût sera le temps de calcul et évalué par le temps CPU obtenu après une simulation. Il sera particulièrement évalué par notre code de calcul écrit en fortran 90 pour chaque simulation. Comme nous l'avons évoqué au début de cette partie, un autre critère d'évaluation conduisant à des mêmes résultats peut être le nombre d'évaluation de la fonction f définissant le modèle en (5.1),

$$\text{cout} := \text{nombre d'appels de la fonction } (t,y) \longrightarrow f(t,y) \text{ tout au long du calcul.} \quad (5.9)$$

L'examen des schémas numériques en section 5.3 montre en effet que cette opération $(t,y) \longrightarrow f(t,y)$ est centrale. C'est particulièrement évident pour les schémas explicites. Pour les schémas implicites, le recours à un algorithme de résolution de type Newton-Krylov avec calcul approché de la jacobienne par différences finies (voir section 5.3.3), fait que c'est également le cas. D'autres types d'opérations sont pratiquées à chaque pas de temps. En particulier pour les schémas implicites un solveur de Krylov (GMRes) est utilisé à chaque itération de l'algorithme de Newton. Ce critère de coût tendra donc à privilégier les schémas pratiquant beaucoup d'autres opérations que l'évaluation $(t,y) \longrightarrow f(t,y)$ à chaque pas de temps, ce sera en particulier le cas de tous les schémas implicites. Ce critère de coût sera d'autant plus valide que l'opération $(t,y) \longrightarrow f(t,y)$ est lourde. C'est donc dépendant du modèle. En général les modèles ioniques *évolués* font appel à de nombreuses évaluations de fonctions transcendantales :

— 31 évaluations de exp ou Log pour Beeler-Reuter,

— 55 évaluations de exp ou Log pour TNNP,

L'évaluation de telles fonctions est environ 10 fois plus coûteuse qu'une multiplication ou qu'une addition de scalaires.

5.3 Schémas en temps

Soit $(y^n) \in E_m$ une solution numérique, on note :

$$f^n = f(t_n, y^n), \quad a^n = a(t_n, y^n), \quad b^n = b(t_n, y^n), \quad n = 0 \dots m,$$

où la fonction F est celle définie dans (5.1) et les fonctions a, b sont celles définies dans (5.2).
Nous énonçons dans cette section les différents schémas en temps qui feront l'intérêt de notre étude. Nous distinguerons deux classes de schémas :

— les schémas *"classiques"*, nous renvoyons par exemple à [2, 3] pour la bibliographie, qui s'appuient tous sur la formulation (5.1) du problème. Il s'agit des schémas Adams Bashforth (AB), "Backward Differentiation Formula" (BDF) et Runge Kutta (RK) et Crank Nicolson (CN).

— les schémas *"stabilisés* prenant en compte la partie linéaire connue du modèle et s'appuyant donc sur la formulation (5.2). Ces schémas incluent en particulier les schémas implicites-explicites, les schémas Rush-Larsen (RL) et exponentiel Adams-Bashforth (EAB) énoncés respectivement dans les chapitres 3 et 4.

5.3.1 Schémas *"classiques"*

Ordre 1

— Euler explicite (FE)

$$\frac{y^{n+1}-y^n}{\Delta t} = f^n. \tag{5.10}$$

— Euler implicite (BE)

$$\frac{y^{n+1}-y^n}{\Delta t} = f^{n+1}. \tag{5.11}$$

Ordre 2

— Crank Nikolson (CN)

$$\frac{y^{n+1}-y^n}{\Delta t} = \frac{1}{2}f^{n+1} + \frac{1}{2}f^n. \tag{5.12}$$

— $BDF2$ (Gear implicite)

$$\frac{\frac{3}{2}y^{n+1} - 2y^n + \frac{1}{2}y^{n-1}}{\Delta t} = f^{n+1}. \tag{5.13}$$

— $SBDF2$

$$\frac{\frac{3}{2}y^{n+1} - 2y^n + \frac{1}{2}y^{n-1}}{\Delta t} = 2f^n - f^{n-1}. \tag{5.14}$$

— $AB2$

$$\frac{y^{n+1}-y^n}{\Delta t} = \frac{3}{2}f^n - \frac{1}{2}f^{n-1}. \tag{5.15}$$

Ordre 3

— $BDF3$

$$\frac{\frac{11}{6}y^{n+1} - 3y^n + \frac{3}{2}y^{n-1} - \frac{1}{3}y^{n-2}}{\Delta t} = f^{n+1}. \tag{5.16}$$

— $SBDF3$

$$\frac{\frac{11}{6}y^{n+1} - 3y^n + \frac{3}{2}y^{n-1} - \frac{1}{3}y^{n-2}}{\Delta t} = 3f^n - 3f^{n-1} + f^{n-2}. \tag{5.17}$$

— $AB3$

$$\frac{y^{n+1} - y^n}{\Delta t} = \frac{23}{12}f^n - \frac{4}{3}f^{n-1} + \frac{5}{12}f^{n-2}. \tag{5.18}$$

Ordre 4

— $RK4$

$$\begin{cases} k_1 & = f^n, \\ k_2 & = f(t_n + \frac{\Delta t}{2}, y^n + \frac{\Delta t}{2}k_1), \\ k_3 & = f(t_n + \frac{\Delta t}{2}, y^n + \frac{\Delta t}{2}k_2), \\ k_4 & = f(t_n + \Delta t, y^n + \Delta t k_3), \\ y^{n+1} & = y^n + \frac{\Delta t}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right). \end{cases} \tag{5.19}$$

— $BDF4$

$$\frac{\frac{25}{12}y^{n+1} - 4y^n + 3y^{n-1} - \frac{4}{3}y^{n-2} + \frac{1}{4}y^{n-3}}{\Delta t} = f^{n+1} \tag{5.20}$$

— $SBDF4$

$$\frac{\frac{25}{12}y^{n+1} - 4y^n + 3y^{n-1} - \frac{4}{3}y^{n-2} + \frac{1}{4}y^{n-3}}{\Delta t} = 4f^n - 6f^{n-1} + 4f^{n-2} - f^{n-3} \tag{5.21}$$

— $AB4$

$$\frac{y^{n+1} - y^n}{\Delta t} = \frac{55}{24}f^n - \frac{59}{24}f^{n-1} + \frac{37}{24}f^{n-2} - \frac{9}{24}f^{n-3}. \tag{5.22}$$

5.3.2 Schémas stabilisés

Ordre 1

— Euler semi-implicite,

$$\frac{y^{n+1}-y^n}{\Delta t} = a^n y^{n+1} + b^n. \tag{5.23}$$

— Exponentiel Euler.

$$y^{n+1} = e^{a^n \Delta t} y^n + \Delta t \varphi_1(a^n \Delta t) b^n. \tag{5.24}$$

Ordre 2, 3 et 4

Les schémas stabilisés que nous allons utiliser ici seront les schémas RL et EAB développés dans les chapitres 3 et 4. Nous allons donc rappeler uniquement leur formulation générale sachant que les détails pour les coefficients sont donnés dans les chapitre 3 et 4.

— RL

$$y^{n+1} = y^n + \Delta t \varphi_1(\alpha_n \Delta t)(\alpha_n y^n + \beta_n) \tag{5.25}$$

— EAB

$$y^{n+1} = \mathrm{e}^{a^n \Delta t} y^n + \Delta t \sum_{j=0}^{k-1} \gamma_{nj} \varphi_{j+1}(a^n \Delta t). \tag{5.26}$$

5.3.3 Mise en œuvre

Pour tous les schémas, la condition initiale sera y^0 dans le problème (5.1) telle que fixée par la définition du cas test en section 5.2.1. Si les schémas explicites et classiques sont faciles à mettre en œuvre, les schémas implicites et exponentiels nécessitent d'autres outils pour leur mise en œuvre effective. Il s'agit de l'évaluation des fonctions φ_j pour les schémas exponentiels et des solveurs implicites pour les schémas implicites. Il est donc question ici de faire des commentaires par rapport à ces aspects incontournables pour ces schémas.

Schémas exponentiels

Pour les schémas exponentiels on doit calculer des termes de la forme,

$$\exp(\alpha_n \Delta t) X, \quad \varphi_k(\alpha_n \Delta t) X, \tag{5.27}$$

pour les fonctions φ_k définies précédemment dans les sections 5.24 et 5.26, pour $\alpha_n \in \mathbb{R}^{N\times N}$ une matrice et pour $X \in \mathbb{R}^N$. La difficulté est alors de calculer le produit matrice vecteur entre une matrice définie par une exponentielle (ou une fonction rationnelle d'exponentielles) de matrice. Cette difficulté sera toujours évitée ici et on supposera toujours que la matrice α_n est diagonale, auquel cas seulement des exponentiels (ou fonctions rationnelles d'exponentielles) de scalaires sont à calculer. Cette hypothèse est forte : elle est naturelle dans notre contexte applicatif grâce à la forme spécifique décrite dans (5.3), elle limite notre étude au cas des ODEs.

Schémas implicites

Pour tous les schémas implicites un solveur non linéaire est nécessaire : ce sera l'algorithme de Newton. A chaque itération de cet algorithme pour la résolution de $G(X) = 0$, la résolution d'un système linéaire du type $DG(A).y = Z$ (d'inconnue y et de matrice $DG(A)$ la jacobienne de G au point A) est nécessaire. Pour éviter l'assemblage de la matrice $DG(A)$ ce système linéaire est résolu par une méthode de Krylov (ici un GMRes, voir par exemple [4]). L'algorithme de Krylov nécessite de réaliser des produits matrices vecteurs $y \to DG(A).y$ qui ici sont des dérivées directionnelles et qui sont calculées de manière approchées par,

$$DG(A).y \simeq \frac{G(A+\varepsilon y) - G(A)}{\varepsilon},$$

où ε est un petit paramètre. Le produit matrice vecteur est donc en pratique réalisé via par une évaluation de la fonction G.

Initialisation : la convergence de l'algorithme de Newton repose sur une initialisation adroite. L'initialisation par la valeur au pas de temps précédent donne de mauvais résultats (nombre d'itérations élevé voire non convergence). On initialise l'algorithme de Newton en utilisant un prédicteur : ici en utilisant un schéma explicite stable et léger (Euler semi implicite (5.23)).

5.4 Comparaison sur des cas tests

Dans cette section, nous ferons une comparaison entre les schémas sur les cas tests précédemment présentés à la section 5.2.1. Cette comparaison se fera principalement sur la base des notions précédemment décrites. Nous considérerons des schémas explicites, implicites et exponentiels d'ordre 1, 2, 3 et 4. Les schémas explicites et implicites seront des schémas classiques bien connus que nous avons définis à la section 5.3.1. Les schémas exponentiels seront les schémas stabilisés RL et EAB.

5.4.1 Comparaison des schémas en terme de précision

Les erreur relatives $e_\infty(\Delta t)$ (5.7) pour les modèles BR et $TNNP$ sont calculées et représentées respectivement dans les tables 5.1 et 5.2. On peut observer dans la table 5.1 que les schémas explicites pour le modèle BR souffrent d'un problème de stabilité. Cette instabilité est traduite par la contrainte d'utilisation des pas de temps assez petits.

(a) AB_2, RL_2, EAB_2 and CN

	Explicite	Exponentiel		Implicite
Δt	AB_2	RL_2	EAB_2	CN
0.2	–	0.251	0.284	4.11×10^{-2}
0.1	–	0.107	9.26×10^{-2}	1.13×10^{-2}
0.05	–	3.35×10^{-2}	2.31×10^{-2}	2.65×10^{-3}
0.025	–	8.88×10^{-3}	5.39×10^{-3}	6.66×10^{-3}
0.0125	–	2.23×10^{-3}	1.29×10^{-3}	1.68×10^{-4}
6.25×10^{-3}	2.07×10^{-4}	5.6×10^{-4}	3.17×10^{-4}	4.25×10^{-5}

(b) AB_3, RL_3, EAB_3 and BDF_3

	Explicite	Exponentiel		Implicite
Δt	AB_3	RL_3	EAB_3	BDF_3
0.2	–	0.148	0.516	4.09×10^{-2}
0.1	–	4.07×10^{-2}	9.17×10^{-2}	1.04×10^{-2}
0.05	–	6.34×10^{-3}	1.09×10^{-2}	2.29×10^{-3}
0.025	–	7.57×10^{-4}	1.17×10^{-3}	3.84×10^{-4}
0.0125	–	9.07×10^{-5}	1.4×10^{-4}	5.25×10^{-5}
6.25×10^{-3}	1.13×10^{-5}	8.23×10^{-6}	1.72×10^{-5}	2.01×10^{-5}

(c) RK_4, RL_4, EAB_4 and BDF_4

	Explicite	Exponentiel		Implicite
Δt	RK_4	RL_4	EAB_4	BDF_4
0.2	–	–	–	4.98×10^{-2}
0.1	–	5.86×10^{-2}	0.119	1.27×10^{-2}
0.05	–	4.58×10^{-3}	8.96×10^{-3}	2.02×10^{-3}
0.025	4.65×10^{-5}	2.61×10^{-4}	4.33×10^{-4}	1.93×10^{-4}
0.0125	2.67×10^{-6}	1.62×10^{-5}	2.67×10^{-5}	3.52×10^{-5}
6.25×10^{-3}	1.65×10^{-7}	9.94×10^{-7}	1.73×10^{-6}	2.01×10^{-5}

TABLE 5.1: Erreurs $e_\infty(\Delta t)$ (5.7) des schémas classiques et stabilisés pour divers pas de temps Δt. Le modèle ionique utilisé est le modèle de Beeler et Reuter [1] (1977).

Le modèle $TNNP$ étant plus raide (à peu près 15 fois) que le modèle BR, la table 5.2 nous montre une contrainte de pas de temps plus sévère dans le cas du modèle $TNNP$. Cette contrainte rend donc les schémas explicites inutilisables pour des modèles raides comme le $TNNP$. Dans les même tables, on peut observer que les schémas exponentiels (RL, EAB) et implicites (BDF, CN) ne souffrent pas de ce problème d'instabilité. Ces schémas résistent donc efficacement à la raideur des modèles BR et $TNNP$. L'utilisation des pas de temps très petits conduit souvent à des précisions dont on n'a pas forcement besoin. Faire des simulations dans cette situation est synonyme de surcoût en temps de

calcul. La capacité de faire des calculs à des pas de temps assez grands donne la possibilité de faire des calculs à des coût raisonnables et à des précisions convenables. Les tables 5.1a et 5.2a montrent que le schéma CN est à peu près d'un facteur de 10 le plus précis parmi les schémas d'ordre 2 considérés. Le schéma $RL2$ est moins précis que le schéma $EAB2$ il est d'ailleurs le moins précis parmi les schémas d'ordre 2. Le schéma $BDF3$ comme on peut le voir dans les tables 5.1b et 5.2b est environ 10 fois plus précis que les schémas stabilisés ($RL3$, $EAB3$) pour $\Delta t \geq 0.0125$. Par ailleurs le schéma $RL3$ est plus précis que les schémas $EAB3$ et $AB3$ pour le modèle BR (table 5.1b). Dans le cas

(a) AB_2, RL_2, EAB_2 and CN

	Explicite	Exponentiel		Implicite
Δt	AB_2	RL_2	EAB_2	CN
0.1	–	0.190	0.339	3.19×10^{-2}
0.05	–	7.45×10^{-2}	9.10×10^{-2}	1.07×10^{-2}
0.025	–	2.20×10^{-2}	2.14×10^{-2}	3.11×10^{-3}
0.0125	–	5.75×10^{-3}	5.12×10^{-3}	7.83×10^{-4}
6.25×10^{-3}	–	1.45×10^{-3}	1.26×10^{-3}	1.86×10^{-4}

(b) AB_3, RL_3, EAB_3 and BDF_3

	Explicite	Exponentiel		Implicite
Δt	AB_3	RL_3	EAB_3	BDF_3
0.1	–	0.307	0.542	3.77×10^{-2}
0.05	–	4.47×10^{-2}	5.95×10^{-2}	1.13×10^{-2}
0.025	–	6.53×10^{-3}	7.06×10^{-3}	2.59×10^{-3}
0.0125	–	8.03×10^{-4}	7.59×10^{-4}	3.93×10^{-4}
6.25×10^{-3}	–	9.87×10^{-5}	8.23×10^{-5}	6.48×10^{-5}

(c) RK_4, RL_4, EAB_4 and BDF_4

	Explicite	Exponentiel		Implicite
Δt	RK_4	RL_4	EAB_4	BDF_4
0.1	–	0.421	–	5.80×10^{-2}
0.05	–	4.96×10^{-2}	7.96×10^{-2}	1.40×10^{-2}
0.025	–	5.85×10^{-3}	8.45×10^{-3}	2.32×10^{-3}
0.0125	–	3.22×10^{-4}	3.68×10^{-4}	1.95×10^{-4}
6.25×10^{-3}	–	2.37×10^{-5}	2.84×10^{-5}	3.39×10^{-5}

TABLE 5.2: Erreurs $e_\infty(\Delta t)$ (5.7) des schémas classiques et stabilisés pour divers pas de temps Δt. Le modèle ionique utilisé est le modèle de Ten Tusscher et al. [5] (2004).

du modèle $TNNP$ (table 5.1b), les deux schémas ont à peu près la même précision. Le schéma $RK4$ est le plus précis des schémas d'ordre 4 considérés pour le modèle BR (table 5.1c). Pour le modèle $TNNP$ (table 5.2c), les schémas d'ordre 4 ont presque la même précision. En définitif, si on s'intéresse uniquement à la précision, les schémas CN, $BDF3$ et $RK4$ respectivement sont les meilleurs parmi les schémas d'ordre 2, 3 et 4 appliqué au modèle BR. La remarque est la même pour le modèle $TNNP$ mais il faut aller à des pas de temps très petits pour pouvoir observer la précision du schéma $RK4$. À pas de temps grands ($\Delta t \geq 0.1$), les schémas d'ordre 2 sont plus précis que les schémas d'ordre élevés.

En passant de l'ordre 2 à l'ordre 3 on gagne plus de précision à partir d'un certain pas de temps ($\Delta t \leq 0.05$). En particulier, le passage du $RL2$ au $RL3$ fait gagner 10 fois plus de précision. Le passage en général de l'ordre 3 à l'ordre 4 donne un gain mais pas assez considérable. Mais si on veut des erreurs très petites ($\leq 10^{-6}$) ça vaut le coup d'utiliser des schémas d'ordre 4 à la place des schémas d'ordre 3.
Maintenant que nous avons comparé les schémas en terme de précision, il sera intéressant de voir si on a les mêmes conclusions si on associe à chaque précision un coût. La section qui va suivre sera par conséquent dédiée à la comparaison des schémas que nous venons de comparer, mais cette fois avec un critère beaucoup plus fort (précision + coût).

5.4.2 Comparaison des schémas en terme de coût

Le coût sur l'axe des ordonnées et l'erreur $e_\infty(\Delta t)$ défini en (5.7) sur l'axe des abscisses, les résultats pour chaque schéma sont représentés dans les figures 5.5 et 5.6 pour les modèles BR et $TNNP$ respectivement. Puisque nous prenons en compte le coût et la précision, nous pouvons quand c'est possible comparer tous les schémas entre eux sans distinction de type (explicites, implicites, stabilisés) de schémas et d'ordre de convergence. La logique d'interprétation de ces figures est de regarder pour une précision donnée, quel schémas coûte plus chers en temps de calcul. Ainsi, plus la courbe représentant un schéma est en dessous des autres, meilleur est ce schéma. Une vue générale sur les figures 5.5 et 5.6 (au dessus) permet de voir en ce qui concerne les schémas RL et EAB que pour des erreurs inférieures 10%, le gain en temps CPU est très grand (avec un facteur de plus de 10) lorsque l'on utilise des schémas 2 à la place des schémas d'ordre 1. Pour des erreurs entre 0.1% et 10%, ce gain n'est pas très significatif quand on bouge de l'ordre 2 à l'ordre 3. Cependant, pour des erreurs inférieures à 0.1%, ce gain commence à être important (avec un facteur de 5 à peu près). En bougeant de l'ordre 3 à l'ordre 4, on ne gagne presque rien sauf si on va chercher des erreurs de l'ordre de 0.0001%. Maintenant que nous avons regardé l'intérêt de monter en ordre pour les schémas RL et EAB, allons à présent à la comparaison entre tous les types de schémas que nous avons considérés à la section précédente. Dans les figures 5.5 et 5.6 (en bas et à gauche), on observe que les schémas $RL3$ et $EAB3$ sont meilleurs que les schémas implicites ($CN, BDF3$ et $BDF4$). On peut d'ailleurs voir que les schémas implicites aussi bien que les schémas $RL3$ et $EAB3$ sont presque équivalents. Cependant, vu que les schémas RL sont faciles à mettre en œuvre par rapport aux schémas EAB, on peut dire que le schéma $RL3$ est meilleur que le schéma $EAB3$. Dans les figures 5.5 et 5.6 (en bas et à droite), on constate que lorsque la comparaison est possible, le schéma $RL4$ est meilleur par rapport aux schémas explicites ($AB2, AB3$ et $RK4$) et $EAB4$. Puisque nous avons vu qu'il n'y avait pas une grande différence entre les schémas stabilisés (RL et EAB) d'ordre 3 et d'ordre 4, on peut conclure que le meilleur parmi tous ces schémas est le schéma $RL3$ car plus facile à mettre en œuvre que le schéma $RL4$. Ceci vient donc changer la conclusion que l'on a eu à la section précédente où à l'ordre 2, 3 et 4 respectivement, le meilleur était le CN, $BDF3$ et

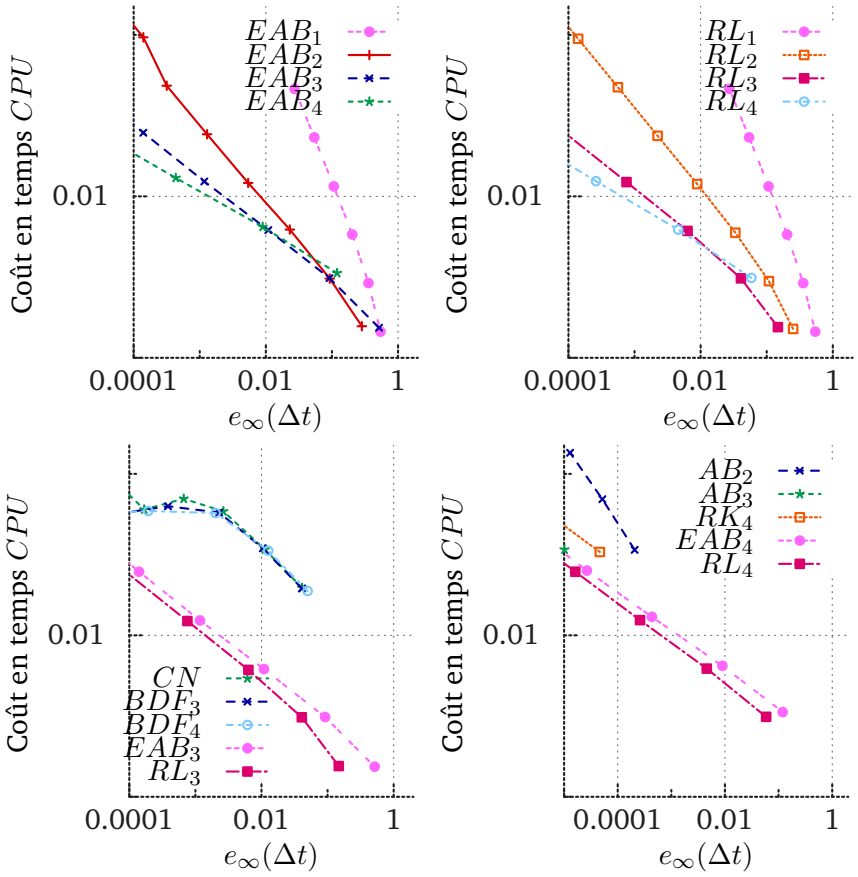


FIGURE 5.5: Temps CPU vs erreurs $e_\infty(\Delta t)$ (5.7) pour divers schémas. Le modèle ionique utilisé est le modèle de Beeler et Reuter [1] (1977). Les courbes sont tracées en échelle Log/Log.

$RK4$. Le schéma $RL3$ est donc une bonne alternative pour faire des simulations réalistes (erreur $\leq 0.01\%$) et à coût raisonnable.

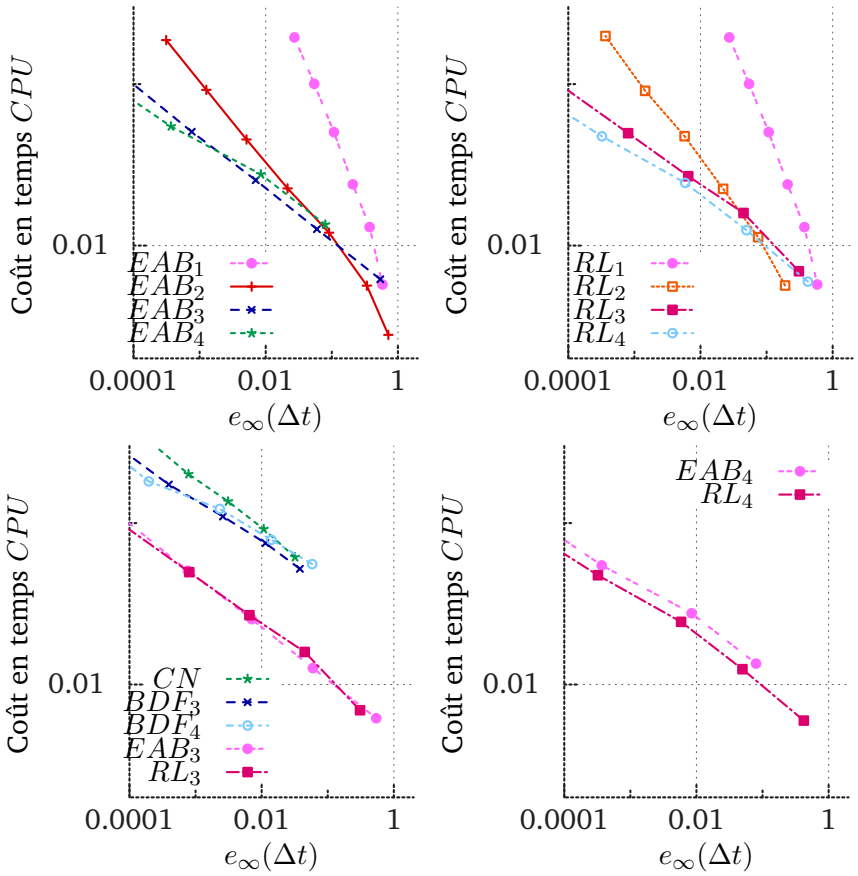


FIGURE 5.6: Temps CPU vs erreurs $e_\infty(\Delta t)$ (5.7) pour divers schémas. Le modèle ionique utilisé est le modèle de Ten Tusscher et al. [5] (2004). Les courbes sont tracées en échelle Log/Log.

5.4.3 Convergence et précision sur les temps d'activation, de repolarisation et de l' APD

Nous investiguons dans cette section la précision sur le calcul de t_a, t_r et APD. Nous avons vu dans la section précédente que les schémas explicites étaient instables et induisaient un coût en temps de calcul très élevé à cause de l'utilisation des pas de temps très petits. Nous considérerons donc pour cette étude les schémas implicites et les schémas stabilisés. Pour un schéma numérique donné, nous utiliserons la méthode de calcul qui a été décrite à la section 5.2.4 et les erreurs seront calculées par les formules données par (5.8). Nous allons nous intéresser particulièrement au cas test du modèle ionique BR. Les figures 5.7-5.9 représentent les courbes en échelle logarithmique des erreurs sur t_a, t_r et APD par rapport au pas de temps Δt. On observe sur ces figures que les erreurs sur t_a, t_r et APD sont à des ordres de grandeurs inférieurs à ceux des erreurs sur le potentiel précédemment étudié. Par ailleurs, on peut voir sur ces figures que les temps d'activation,

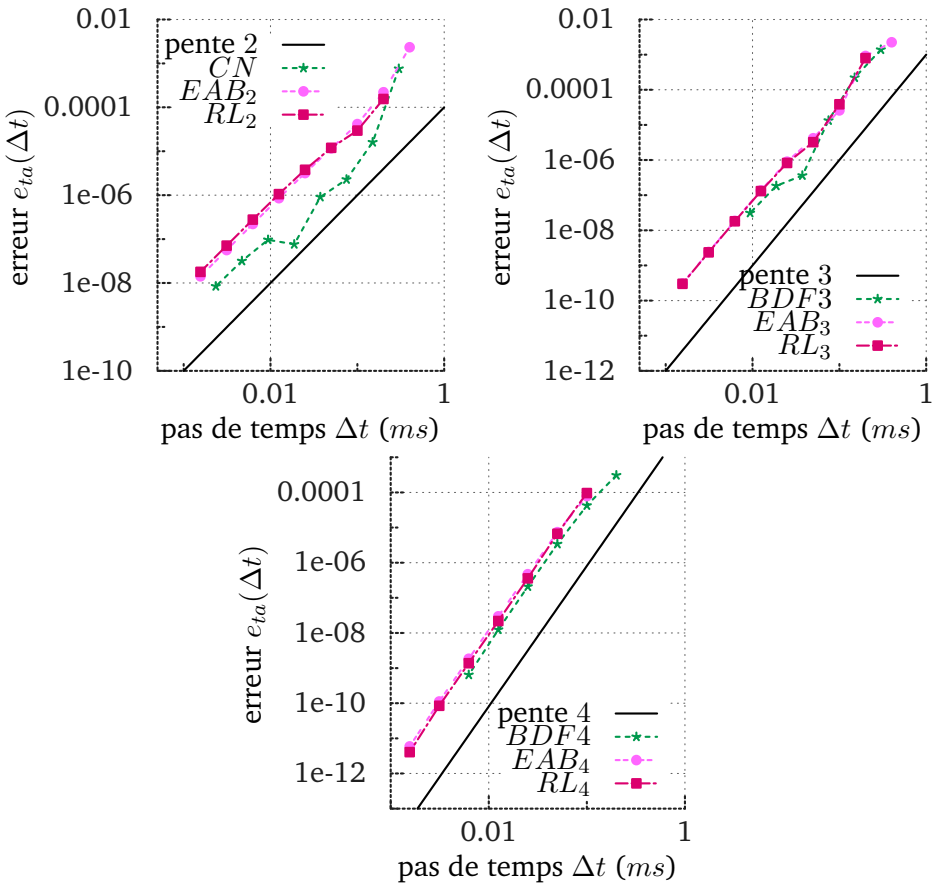


FIGURE 5.7: Erreurs (5.8) $e_{ta}(\Delta t)$ pour les schémas CN, $BDF3$, $BDF4$, EAB_k, RL_k, $k = 2, 3, 4$. Le modèle ionique utilisé est le modèle de Beeler et Reuter [1] (1977).

de récupération et la durée du potentiel d'action prédits par les solutions numériques convergent aux ordres des schémas numériques utilisés respectivement vers les temps d'activation, de récupération et la durée du potentiel d'action prédit par la solution de référence. En pratique, les calculs sur l'étude numérique des ordres de convergence sont coûteux en temps de calculs, en mémoire et assez complexes. Ces coûts et complexités sont un peu allégés sur l'étude de la convergence sur t_a, t_r et APD. On peut donc se servir de cette étude pour avoir une idée sur l'ordre du schéma numérique que l'on utilise.

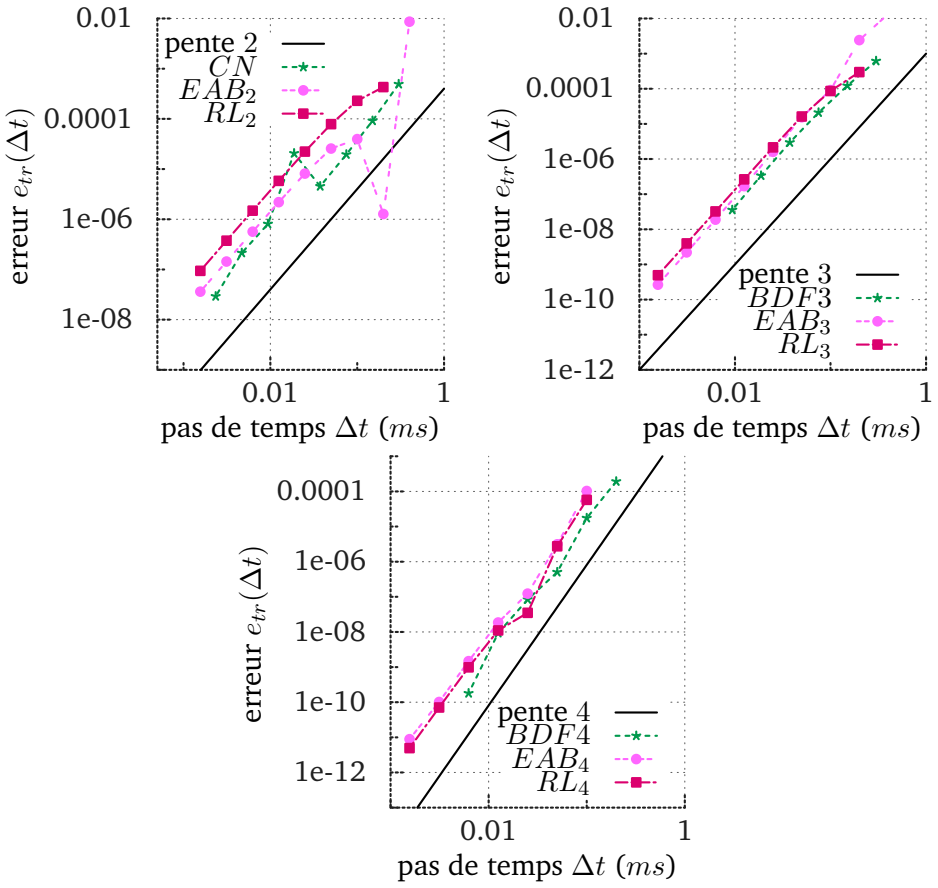


FIGURE 5.8: Erreurs $e_{tr}(\Delta t)$ (5.8) pour les schémas CN, $BDF3$, $BDF4$, EAB_k, RL_k, $k = 2, 3, 4$. Le modèle ionique utilisé est le modèle de Beeler et Reuter [1] (1977).

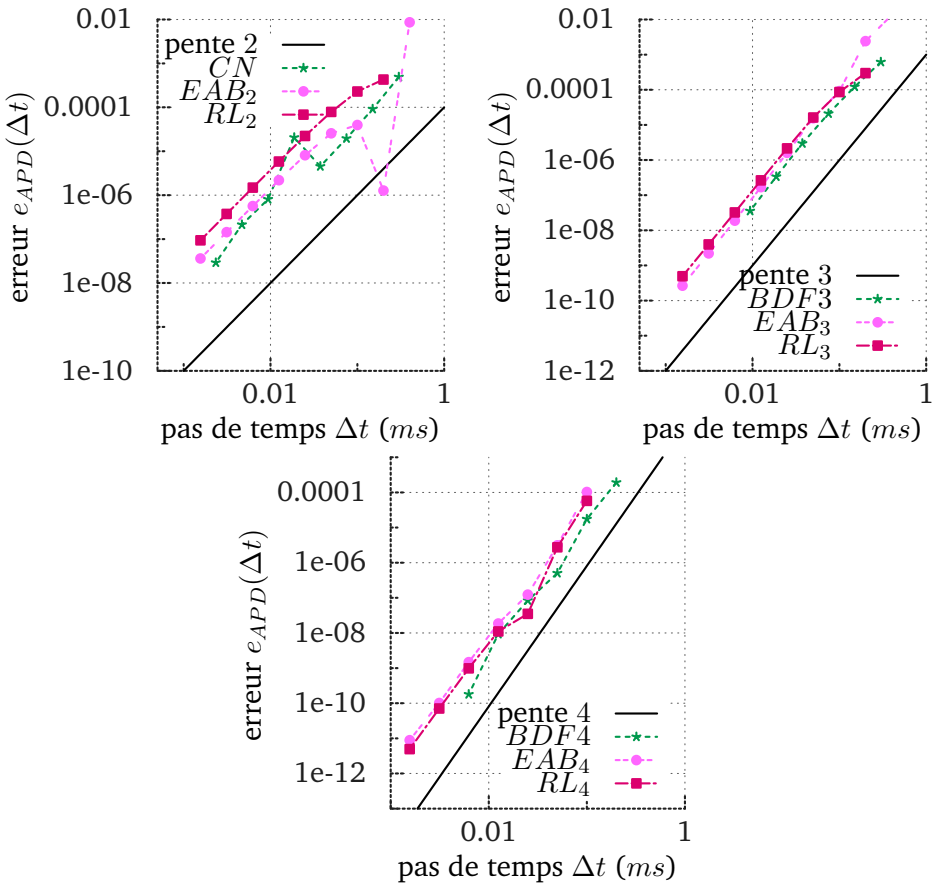


FIGURE 5.9: Erreurs $e_{APD}(\Delta t)$ (5.8) pour les schémas CN, $BDF3$, $BDF4$, EAB_k, RL_k, $k = 2, 3, 4$. Le modèle ionique utilisé est le modèle de Beeler et Reuter [1] (1977).

Bibliographie

[1] G.W. BEELER et H. REUTER. „Reconstruction of the Action Potential of Ventricular Myocardial Fibres“. English. In : *J. Physiol.* 268 (1977), p. 177–210 (cf. p. 82, 92, 95, 97–99).

[2] E. HAIRER, S. P. NORSETT et G. WANNER. *Solving ordinary differential equations. I*. Second. T. 8. Springer Series in Computational Mathematics. Nonstiff problems. Springer-Verlag, Berlin, 1993, p. xvi+528 (cf. p. 87).

[3] E. HAIRER et G. WANNER. *Solving ordinary differential equations. II*. T. 14. Springer Series in Computational Mathematics. Stiff and differential-algebraic problems, Second revised edition, paperback. Springer-Verlag, Berlin, 2010, p. xvi+614 (cf. p. 87).

[4] Y. SAAD. *Iterative methods for sparse linear systems*. Second. Philadelphia, PA : Society for Industrial et Applied Mathematics, 2003, p. xviii+528 (cf. p. 91).

[5] KHWJ TEN TUSSCHER, D NOBLE, PJ NOBLE et Alexander V PANFILOV. „A model for human ventricular tissue“. In : *American Journal of Physiology-Heart and Circulatory Physiology* 286.4 (2004), H1573–H1589 (cf. p. 82, 93, 96).

6 Discrétisation spatiale et temporelle pour le modèle monodomaine

Contents

6.1 Espaces fonctionnels

Soit d un entier non nul ($d = 1, 2, 3$), $T > 0$ et Ω un ouvert borné régulier de $\mathbb{R}^d$ muni de la mesure de Lebesgue. Pour un espace de Banach $(E, \|\cdot\|_E)$ de fonctions mesurables définies dans Ω, on note E' son dual topologique. Soit $v : [0, T] \to E$ une fonction mesurable. On définit la norme

$$\|v\|_{L^2(0,T;E)} = \left(\int_0^T \|v\|_E^2 \, dx \right)^{\frac{1}{2}}, \tag{6.1}$$

et on considère l'espace $L^2(0, T; E) = \{v : [0, T] \to E : \|v\|_{L^2(0,T;E)} < \infty\}$ qui muni de la norme (6.1) est un espace de Banach.
Soit $L^2(\Omega)$ l'espace des fonctions mesurables de carrés sommables dans Ω, muni du produit scalaire $(u, v) = \int_\Omega uvdx$. On note $\|\cdot\|$ la norme dans $L^2(\Omega)$ telle que pour toutes fonctions à valeurs réelles v,

$$\|v\| = \left(\int_\Omega v^2 dx \right)^{\frac{1}{2}}.$$

Soit $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, d un entier positif. On désigne par $D^\alpha = (\partial/\partial x_d)^{\alpha_d} \cdots (\partial/\partial x_1)^{\alpha_d}$, $\alpha = (\alpha_1, \ldots, \alpha_d) \in N^d$ une dérivée arbitraire par rapport à x et d'ordre $|\alpha| = \sum_{j=1}^d \alpha_j$.
On désigne par $H^r(\Omega)$ l'espace de Sobolev des fonctions de $L^2(\Omega)$ admettant des dérivées

d'ordre r dans $L^2(\Omega)$. On munit $H^r(\Omega)$ du produit scalaire $(u,v)_r = \int_\Omega \sum_{|\alpha|\leq r} D^\alpha u D^\alpha v dx$. On note $\|\cdot\|_r$ la norme dans l'espace $H^r(\Omega)$ telle que pour toutes fonctions à valeurs réelles v,

$$\|v\|_r = \left(\int_\Omega \sum_{|\alpha|\leq r} \|D^\alpha v\|^2 dx \right)^{\frac{1}{2}}.$$

On pose $V = H^1(\Omega)$, $H = L^2(\Omega)$ et on note V', H' leurs espaces duaux respectifs. Il existe des injections denses et continues tels que, $V \subset H \simeq H' \subset V'$ (voir [6] Chapitre 5). Ainsi, le crochet de dualité $(,)_{V',V}$ coïncide avec le produit scalaire $(,)$ dans V. On définit l'espace $\mathcal{W}(V,V')$ par,

$$\mathcal{W}(V,V') = \{\omega : [0,T] \to V ; \omega \in L^2(0,T;V), \partial_t \omega \in L^2(0,T;V')\} \tag{6.2}$$

On peut montrer (voir [9]) qu'il existe une injection continue de $\mathcal{W}(V,V')$ dans $C([0,T];H)$.

6.2 Rappel sur le modèle monodomaine

Nous rappelons le modèle monodomaine décrit dans la section 2.2.2. Le modèle ionique utilisé est de type physiologique comme les modèles TNNP et BR décrits en 2.1.5. Une formulation générale des modèles physiologiques, en particulier des modèles TNNP et BR a été donnée dans les chapitres 3 et 4. Le modèle monodomaine est formé d'une EDP parabolique couplée à un système d'équations différentielles ordinaires. Ce système couplé peut se formuler comme suit,

$$\begin{aligned}
\partial_t \begin{pmatrix} w \\ c \end{pmatrix} &= \begin{pmatrix} -1/\tau(u) & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} w \\ c \end{pmatrix} + \begin{pmatrix} w_\infty(u)/\tau(v) \\ g(w,c,u) \end{pmatrix} \quad dans\ [0,T]\times\Omega, \\
\partial_t u &= \operatorname{div}(\sigma \nabla u) + \frac{1}{C_m}(-I_{ion}(w,c,u) + I_{st}) \qquad dans\ [0,T]\times\Omega, \\
n\cdot(\sigma\nabla u) &= 0 \qquad sur\ [0,T]\times\partial\Omega, \\
y(0) &= y_0 \qquad sur\ \Omega,
\end{aligned}$$

où $y = (w,c,u)$, $y(0) = (w_0,c_0,u_0)$ est la condition initiale donnée par le modèle et n est la normale unitaire sortante à $\partial\Omega$. Ce système écrit sous une forme générale donne,

$$\partial_t v = a(u)v + b(v,u) \qquad \text{dans } [0,T]\times\Omega, \tag{6.3}$$

$$\partial_t u = \operatorname{div}(\sigma\nabla u) + f(t,v,u) \qquad \text{dans } [0,T]\times\Omega, \tag{6.4}$$

$$n\cdot(\sigma\nabla u) = 0 \qquad \text{sur } [0,T]\times\partial\Omega, \tag{6.5}$$

$$v(0,x) = v_0(x) \qquad \text{dans } \Omega, \tag{6.6}$$

$$u(0,x) = u_0(x) \qquad \text{dans } \Omega. \tag{6.7}$$

où $v = (w, c)$, $a(u) = diag(-1/\tau(u), 0)$, $f(t, w, u) = \frac{1}{C_m}(-I_{ion}(w, c, u) + I_{stim})$ et $b(v, u) = (w_\infty(u)/\tau(v), g(w, c, u))$. Le symbole σ désigne le tenseur de diffusivité positif qui dépend de l'espace et pas du temps.
La matrice $a(u)$ étant diagonale et ne dépendant que de la fonction u (pas de w ni de c), il n'est alors pas restrictif de traiter le système (6.3) composante par composante. On va donc pour nos études, prendre (6.3) dans le cas où le vecteur v est de dimension 1.
Dans ce cas, le problème faible associé est de trouver $v \in L^2(0, T, L^2(\Omega)) \cap \mathcal{W}(V, V')$ et $u \in L^2(0, T, H^1(\Omega)) \cap \mathcal{W}(V, V')$ telle que ,

$$(\partial_t v, \psi) = (a(u)v + b(v, u), \psi) \qquad \forall \psi \in L^2(\Omega), \tag{6.8}$$

$$(\partial_t u, \varphi) + (\sigma \nabla u, \nabla \varphi) = (f(t, v, u), \varphi) \qquad \forall \varphi \in H^1(\Omega). \tag{6.9}$$

Avec v_0 et u_0 donnés dans $L^2(\Omega)$ et $H^1(\Omega)$ respectivement. La discrétisation du problème (6.8) -(6.9) se fait en deux étapes à savoir la discrétisation en espace et ensuite la discrétisation en temps, qui conduit à la discrétisation totale du problème.

6.3 Discrétisation du problème

Pour être capable de mailler exactement l'ouvert Ω, on va supposer qu'il est polyédrique, c'est à dire que $\overline{\Omega}$ est une réunion finie de polyèdres de $\mathbb{R}^d$, $d = 1, 2, 3$. On notera $\mathbb{P}_r$ l'ensemble des polynômes de degré inférieur ou égal à r par rapport à $x = (x_1, ..., x_d) \in \mathbb{R}^d$.

$$\mathbb{P}_r = \left\{ \sum_{0 \le i_1 + ... + i_d \le r} a_{i_1 ... i_n} x_1^{i_1} ... x_{i_d}^{i_d} \ : \ a_{i_1 ... i_d} \in \mathbb{R}, (i_1, ..., i_d) \in \mathbb{N}^d \right\}. \tag{6.10}$$

Avant de parler de la discrétisation en elle même, nous avons besoin de définir certains outils qui sont incontournables pour la réalisation effective d'une discrétisation en espace. Plus de détails sur ces définitions peuvent être trouvées dans [3]

6.3.1 Préliminaires

Définition 1 (Simplexes)**.** Soient $(n \ge 2)$ et $(p_i)_{1 \le i \le n+1}$ une famille de $n + 1$ points de $\mathbb{R}^d$ telle que la matrice suivante :

$$A = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n+1} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n1} & \cdots & p_{nn+1} \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

soit inversible et où $(p_{ji})_{1\leq j\leq n}$ sont les composantes du vecteur p_i dans $\mathbb{R}^n$. On appelle n-simplexe K de sommets $(p_i)_{1\leq i\leq n+1}$ l'enveloppe convexe des points p_i. On montre que

$$K = \left\{ x \in \mathbb{R}^n : \ x = \sum_{j=1}^{n+1} \lambda_j(x) p_j, \ 0 \leq \lambda_j(x) \leq 1, \ \sum_{j=1}^{n+1} \lambda_j = 1 \right\}.$$

Les scalaires $(\lambda_j(x))_{1\leq j\leq n+1}$ sont appelés coordonnées barycentriques du point x.

Définition 2 (Maillage)**.** On suppose que Ω est connexe. Un maillage ou triangulation de $\overline{\Omega}$ est une famille finie $\mathcal{T}_h = \{K_i\}_{1\leq i\leq m}$ de n-simplexes telle que :

1. $\overline{\Omega} = \bigcup_{K\in\mathcal{T}_h} K$. Pour tout $K \in \mathcal{T}_h$, K est un polyèdre d'intérieur non vide.
2. $\mathring{K}_i \cap \mathring{K}_j = \emptyset, \ pour \ i \neq j.$
3. $K_i \cap K_j$ est soit une face commune, soit une arête ou un sommet commun à $K_i \ et \ K_j$.

Pour tout $K \in \mathcal{T}_h$, on pose h_K =diam$(K) := \sup_{x,y\in K} \mathcal{N}(x-y)$, avec $\mathcal{N}$ la norme euclidienne dans $\mathbb{R}^d$ et $h = \max_{K\in\mathcal{T}_h} h_K$; h est appelé taille du maillage.

Définition 3 (treillis d'ordre r)**.** Soit K un n-simplexe. On appelle treillis d'ordre r l'ensemble (fini)

$$\Sigma_r = \left\{ x \in K \text{ tel que } \lambda_j(x) \in \left\{ 0, \frac{1}{r}, \ldots, \frac{r-1}{r} \right\} \text{ pour } 1 \leq j \leq N \right\}.$$

Les sommets de Σ_r sont notés $(\sigma_j)_{1\leq j\leq n_r}$.
Pour $r = 1$ ces sommets sont exactement les sommets de K et pour $r = 2$ ce sont les sommets et les points milieux des arrêtes reliant deux sommets de K.

Remarque 1. Soit K un n-simplexe. Tout polynôme de $\mathbb{P}_r$ est déterminé de manière unique par ses valeurs aux sommets σ_j de Σ_r. L'ensemble Σ_r est donc unisolvant pour $\mathbb{P}_r$ (voir [3], Lemme 6.3.3 page 189). La conséquence de cette unisolvance est l'existence d'une base $(\psi_j)_{1\leq j\leq n_r}$ de $\mathbb{P}_r$ telle que,

$$\psi_j(\sigma_i) = \delta_{ij}, \qquad 1 \leq i, j \leq n_r$$

Définition 4 (Elément fini Lagrange)**.** Soit $\mathcal{T}_h$ une triangulation de Ω formé de n-simplexes. L'espace des éléments finis Lagrange d'ordre r ou élément fini de type $\mathbb{P}_r$ associé à ce maillage, est le sous-espace vectoriel :

$$V_h = \left\{ v \in \mathcal{C}(\overline{\Omega}) \ : \ v_{|K_i} \in \mathbb{P}_r, \ \forall K \in \mathcal{T}_h \right\}.$$

Les élément de V_h sont des fonctions polynomiales par morceaux. D'autre part V_h est un sous espace de $H^1(\Omega)$ de dimension finie.

Définition 5 (Noeuds et degrés de liberté)**.** Soit $\mathcal{T}_h$ un maillage de Ω formé de n-simplexes. Soit V_h l'espace d'élément fini Lagrange d'ordre r. On appelle nœuds des degrés de liberté

l'ensemble des points (distincts) $(\widehat{p}_i)_{1\leq i\leq n_{dl}}$ des treillis d'ordre r de chacun des n-simplexes $K_i \in \mathcal{T}_h$. On appelle degré de liberté d'une fonction $v \in V_h$ l'ensemble des valeurs v aux nœuds $(\widehat{p}_i)_{1\leq i\leq n_{dl}}$.

Remarque 2. Par la propriété d'unisolvance de Σ_r à $\mathbb{P}_r$ soulignée dans la remarque 1, il existe une base $(\phi_j)_{1\leq j\leq n_{dl}}$ de V_h telle que,

$$\phi_j(\widehat{p}_i) = \delta_{ij}, \qquad 1 \leq i, j \leq n_{dl}.$$

Cette base est appelée base fondamentale de V_h.

6.3.2 Discrétisation spatiale

Soit $\mathcal{T}_h$ une triangulation de Ω telle que définie dans 2, soit n_{dl} le nombre de degrés de libertés et $(\widehat{p}_j)_{1\leq j\leq n_{dl}}$ l'ensemble des nœuds des degrés de liberté. Soit V_h l'espace d'éléments fini Lagrange d'ordre r, $(\phi_i)_{1\leq i\leq n_{dl}}$ sa base fondamentale tel que défini dans la remarque 2. On reformule le problème (6.8)-(6.9) en remplaçant les espaces H et V par l'espace vectoriel de dimension finie V_h de telle sorte que le problème revient à trouver v_h, u_h dans $C^1(]0,T[;V_h)$ tels que,

$$(\partial_t v_h, \psi_h) = (a(u_h)v_h + b(v_h,u_h), \psi_h) \qquad \forall \psi_h \in V_h \tag{6.11}$$

$$(\partial_t u_h, \varphi_h) + (\sigma\nabla u_h, \nabla\varphi_h) = (f(t,v_h,u_h), \varphi_h) \qquad \forall \varphi_h \in V_h, \tag{6.12}$$

Le problème obtenu étant défini sur l'espace de dimension finie V_h, on peut l'écrire dans la base fondamentale de V_h. Cela peut se faire simplement en introduisant dans (6.11)-(6.12) l'écriture de v_h et u_h dans la base $(\phi_i)_{1\leq i\leq n_{dl}}$. Plus précisément, pour

$$v_h(t) = \sum_{i=1}^{n_{dl}} v_i(t)\phi_i \quad \text{et} \quad u_h(t) = \sum_{i=1}^{n_{dl}} u_i(t)\phi_i, \tag{6.13}$$

avec $v_i(t)$ et $u_i(t)$, des réels à déterminer. Les fonctions a, b et f sont en général non linéaires. Par conséquent, $a(u_h)v_h$, $b(v_h,u_h)$ et $f(t,v_h,u_h)$ peuvent ne pas être dans l'espace V_h.
Dans le cas où elles sont toutes dans V_h, le problème (6.11)-(6.12) peut se formuler facilement en utilisant la matrice de masse $M = (\phi_i,\phi_j)_{1\leq i,j\leq n_{dl}}$ et la matrice de raideur $S = (\sigma\nabla\phi_i, \nabla\phi_j)_{1\leq i,j\leq n_{dl}}$. On a en effet,

$$a(u_h)v_h + b(v_h,u_h) = \sum_{i=1}^{n_{dl}} \left(a(u_i)v_i + b(v_i,u_i)\right)\phi_i, \tag{6.14}$$

$$f(t,v_h,u_h) = \sum_{i=1}^{n_{dl}} f(t,v_i,u_i)\phi_i. \tag{6.15}$$

En utilisant (6.13) et (6.14)-(6.15), le problème (6.11)-(6.12) revient à trouver les réels $u_i(t)$ et $v_i(t)$ tel que pour $1 \leq j \leq n_{dl}$,

$$\sum_{i=1}^{n_{dl}}(\phi_i,\phi_j)\partial_t v_i(t) = \sum_{i=1}^{n_{dl}}(\phi_i,\phi_j)\left(a(u_i(t)) + b(v_i(t),u_i(t))\right), \tag{6.16}$$

$$\sum_{i=1}^{n_{dl}}(\phi_i,\phi_j)\partial_t u_i(t) + \sum_{i=1}^{n_{dl}}(\sigma\nabla\phi_i,\nabla\phi_j)u_i(t) = \sum_{i=1}^{n_{dl}}(\phi_i,\phi_j)f(t,v_i(t),u_i(t)). \tag{6.17}$$

En fonction de la matrice de masse M et de la matrice de raideur S, (6.16)-(6.17) donne,

$$M\partial_t V = M\left(a(U)V + b(V,U)\right), \tag{6.18}$$

$$M\partial_t U + SU = Mf(t,V,U), \tag{6.19}$$

où $V=(v_i)$, $U=(u_i)$, $a(U)=(a(u_i))$, $b(V,U)=(b(v_i,u_i))$ et $f(t,V,U)=(f(t,v_i,u_i))$, pour $1 \leq i \leq n_{dl}$.
Dans le cas où les fonctions $a(u_h)v_h,\ b(v_h,u_h)$ et $f(t,v_h,u_h)$ ne sont pas dans V_h, le calcul des degrés de liberté $v_i(t)$, $u_i(t)$ est plus compliqué. En effet, en utilisant (6.13) dans (6.11)-(6.12), le problème qu'on obtient revient à trouver les réels, $v_i(t)$, $u_i(t)$ tels que pour $1 \leq j \leq n_{dl}$,

$$\sum_{i=1}^{n_{dl}}(\phi_i,\phi_j)\partial_t v_i(t) = \left(a\left(\sum_{i=1}^{n_{dl}}u_i(t)\phi_i\right)\sum_{i=1}^{n_{dl}}v_i(t)\phi_i + b\left(\sum_{i=1}^{n_{dl}}v_i(t)\phi_i,\sum_{i=1}^{n_{dl}}u_i(t)\phi_i\right),\phi_j\right), \tag{6.20}$$

$$\sum_{i=1}^{n_{dl}}(\phi_i,\phi_j)\partial_t u_i(t) + \sum_{i=1}^{n_{dl}}(\sigma\nabla\phi_i,\nabla\phi_j)u_i(t) = \left(f\left(t,\sum_{i=1}^{n_{dl}}v_i(t)\phi_i,\sum_{i=1}^{n_{dl}}u_i(t)\phi_i\right),\phi_j\right). \tag{6.21}$$

Le problème (6.20)-(6.21) est d'un point de vue numérique moins pratique à résoudre que le problème (6.18)-(6.19).
Le cas où les fonctions $a(u_h)v_h,\ b(v_h,u_h)$ et $f(t,v_h,u_h)$ ne sont pas dans V_h est celui que nous rencontrons dans la résolution du modèle monodomaine considéré pour notre étude. Pour obtenir un problème plus simple à résoudre numériquement, nous utilisons une fonction d'interpolation pour approximer les fonctions $a(u_h)v_h,\ b(v_h,u_h)$ et $f(t,v_h,u_h)$.
Soit $I_h : C(\Omega) \to V_h$ l'interpolation qui transforme une fonction $g \in C(\Omega)$ en une fonction $I_h g = g_h$ dans V_h par,

$$g_h = \sum_{i=1}^{n_{dl}} g_i\phi_i = \sum_{i=1}^{n_{dl}} g(\widehat{p}_i)\phi_i. \tag{6.22}$$

L'interpolateur I_h est classique et bien documenté dans la littérature pour l'approximation des fonctions continues (voir [7, 10, 5]).
Les interpolations par I_h de $a(u_h)v_h + b(v_h, u_h)$ et $f(t, v_h, u_h)$ donnent respectivement,

$$I_h(a(u_h)v_h + b(v_h, u_h))(t) = \sum_{i=1}^{n_{dl}} (a_i v_i + b_i)(t)\phi_i, \tag{6.23}$$

$$I_h f(t, v_h, u_h) = f_h(t) = \sum_{i=1}^{n_{dl}} f_i(t)\phi_i, \tag{6.24}$$

avec $(a_i v_i + b_i)(t) = a(u_i(t))v_i(t) + b(v_i(t), u_i(t))$ et $f_i(t) = f(t, v_i(t), u_i(t))$. En remplaçant dans (6.11)-(6.12) les fonctions $a(u_h)v_h + b(v_h, u_h)$ et $f(t, v_h, u_h)$ par leurs fonctions interpolées données respectivement par (6.23) et (6.24), on obtient encore le problème (6.18)-(6.19). Cependant, on n'a plus exactement le problème de départ (6.11)-(6.12) mais une approximation. La matrice de masse M étant inversible, l'équation 6.18 peut s'écrire point par points (points des degrés de libertés). Le système 6.18-6.19 devient alors,

$$\partial_t V = a(U)V + b(V, U) \tag{6.25}$$

$$M\partial_t U + SU = Mf(t, V, U). \tag{6.26}$$

Sous une forme plus détaillée on aura donc de façon équivalente,

$$\partial_t v_i = a_i v_i + b_i, \qquad i = 1, \ldots, n_{dl} \tag{6.27}$$

$$M\partial_t U + SU = Mf(t, V, U), \tag{6.28}$$

avec $a_i = a(u_i), b_i = b(v_i, u_i)$. Le système (6.27)-(6.28) ou (6.25)-(6.26) correspond à notre problème semi-discret et est celui que nous utilisons en pratique.
Pour compléter la discrétisation et afin d'avoir une discrétisation complète, il reste à discrétiser en temps. Celle ci se fait à travers des schémas numériques d'intégration en temps.

6.3.3 Discrétisation en temps

Le problème semi-discret (6.27)-(6.28) est constitué de deux systèmes d'équations de types différents. Contrairement à (6.27), le système (6.28) contient une matrice de raideur provenant de la discrétisation du terme de diffusion $-\mathrm{div}(\sigma\nabla u)$. Les fonctions a et b provenant du type de modèles ioniques que nous aurons à étudier sont non linéaires et sont constituées majoritairement de fonctions exponentielles. Pour des raisons de coût en temps de calcul, on n'aimerait pas les intégrer implicitement. On va donc les intégrer explicitement par des schémas RL proposés dans le chapitre 4. Par ailleurs, la matrice S possède des valeurs propres à partie réelle tendant vers $-\infty$ au fur et à mesure que le maillage devient de plus en plus fin. Pour cette raison, on aimerait intégrer la partie linéaire SU implicitement et traiter $f(t, V, U)$ explicitement pour les mêmes raisons que a

et b. Ceci nous amène donc à traiter la deuxième équation par des schémas de type IMEX en l'occurrence, nous avons choisis les schémas $SBDF$.

Soit donc $0 = t_0, \ldots, t_N = T$ une discrétisation de l'intervalle $[0, T]$ tel que $t_n = n\Delta t$, $n = 0, \ldots, N$. On désigne par V^n et U^n les approximations respectives de V et U à l'instant t_n. On pose $A^n = a(U^n)$, $B^n = b(V^n, U^n)$ et $F^n = f(t_n, V^n, U^n)$. Nous proposons donc la combinaison suivante pour l'intégration en temps du problème (6.25)-(6.26).

$$\begin{aligned} &\text{Sur (6.25) :} && V^{n+1} = V^n + \Delta t \varphi_1(\alpha_n \Delta t)\left(\alpha_n V^n + \beta_n\right), \\ &\text{Sur (6.26) :} && \text{Une methode IMEX convenable d'ordre } k \text{ sur } U, \end{aligned}$$

où α_n et β_n sont calculés comme dans le chapitre 4 par la méthode de Rush Larsen à l'ordre k. Plus précisément, on a les schémas suivant,

— Éléments finis $\mathbb{P}_r$, Rush Larsen ordre 1 et Euler implicite-explicite ($\mathbb{P}_r + RL_1 + FBE$).

$$\begin{aligned} V^{n+1} &= V^n + \Delta t \varphi_1(A^n \Delta t)\left(A^n V^n + B^n\right), \\ M\frac{U^{n+1} - U^n}{\Delta t} &= -SU^{n+1} + MF^n. \end{aligned}$$

— Éléments finis $\mathbb{P}_r$, Rush Larsen ordre 2 et différences finies semi-rétrogrades ordre 2 ($\mathbb{P}_r + RL_2 + SBDF_2$).

$$\begin{aligned} V^{n+1} &= V^n + \Delta t \varphi_1(\alpha_n \Delta t)\left(\alpha_n V^n + \beta_n\right), \\ M\frac{\frac{3}{2}U^{n+1} - 2U^n + \frac{1}{2}U^{n-1}}{\Delta t} &= -SU^{n+1} + M(2F^n - F^{n-1}). \end{aligned}$$

$\alpha_n = \frac{3}{2}A^n - \frac{1}{2}A^{n-1}$ et $\beta_n = \frac{3}{2}B^n - \frac{1}{2}B^{n-1}$

— Éléments finis $\mathbb{P}_r$, Rush Larsen ordre 3 et différences finies semi-rétrogrades ordre 3 ($\mathbb{P}_r + RL_3 + SBDF_3$).

$$\begin{aligned} V^{n+1} &= V^n + \Delta t \varphi_1(\alpha_n \Delta t)\left(\alpha_n V^n + \beta_n\right), \\ M\frac{\frac{11}{6}U^{n+1} - 3U^n + \frac{3}{2}U^{n-1} - \frac{1}{3}U^{n-2}}{\Delta t} &= -SU^{n+1} + M(3F^n - 3F^{n-1} + F^{n-2}). \end{aligned}$$

$\alpha_n = \frac{1}{12}(23A^n - 16A^{n-1} + 5A^{n-2})$,
$\beta_n = \frac{1}{12}(23B^n - 16B^{n-1} + 5B^{n-2}) + \frac{\Delta t}{12}(A^n B^{n-1} - A^{n-1}B^n)$.

— Éléments finis $\mathbb{P}_r$, Rush Larsen ordre 4 et différences finies semi-rétrogrades ordre 4 ($\mathbb{P}_r + RL_4 + SBDF_4$).

$$\begin{aligned} V^{n+1} &= V^n + \Delta t \varphi_1(\alpha_n \Delta t)\left(\alpha_n V^n + \beta_n\right), \\ M\frac{\frac{25}{12}U^{n+1} - 4U^n + 3U^{n-1} - \frac{4}{3}U^{n-2} + \frac{1}{4}U^{n-3}}{\Delta t} &= -SU^{n+1} \\ &\quad + M(4F^n - 6F^{n-1} + 4F^{n-2} - F^{n-3}). \end{aligned}$$

$$\alpha_n = \tfrac{1}{24}(55A^n - 59A^{n-1} + 37A^{n-2} - 9A^{n-3}),$$
$$\beta_n = \tfrac{1}{24}(55B^n - 59B^{n-1} + 37B^{n-2} - 9B^{n-3}) + \frac{\Delta t}{12}(A^n(3B^{n-1} - B^{n-2}) - (3A^{n-1} - A^{n-2})B^n).$$

On peux aussi remplacer le schéma RL par le schéma EAB défini dans la section 5.26. Dans ce cas on obtient donc en général le schéma,

(6.25) : $$V^{n+1} = \mathrm{e}^{A^n h}V^n + \Delta t \sum_{j=0}^{k-1} \gamma_{nj}\varphi_{j+1}(A^n\Delta t),$$

(6.26) : Une methode IMEX convenable d'ordre k sur U.

Nous préférons utiliser les schémas RL aux schémas EAB parce que comparés aux schémas EAB, les schémas RL sont plus faciles à mettre en œuvre.

6.4 Résultats principaux

Le problème (6.9) est une EDP parabolique quasi-linéaire. Dans les travaux de Akrivis et Al. [2, 1] l'étude de la convergence globale (temps et espace) pour les EDP paraboliques quasi-linéaires est abordée dans le cas où un schéma $SBDF$ est utilisé pour la discrétisation en temps et un schéma éléments finis est utilisé pour la discrétisation en espace. Dans cette étude, sont montrés des résultats de convergence globale sous certaines conditions dites *faibles* sur le maillage de diamètre h et le pas temps Δt utilisé pour la discrétisation (ie : sous les conditions Δt et $h^{2r}\Delta t$ assez petit ; r ordre de la méthode de discrétisation en espace).
Dans notre cas, nous avons en plus d'une EDP parabolique semi-linéaire, un système d'EDO qui couple l'EDP et que nous intégrons avec un schéma de type exponentiel (Schéma Rush Larsen (RL)). Il est donc question d'étudier la convergence en espace puis la convergence globale (temps et espace) dans le cas où pour l'intégration en temps, les schémas $SBDF$ et RL sont combinés : le schéma $SBDF$ pour l'EDP et le schéma RL pour l'EDO, sachant que l'EDP et l'EDO sont couplées. À notre connaissance, une telle étude n'est pas présente dans la littérature.
Nous allons donc dans cette partie **montrer des résultats de convergence nouveaux** pour le problème (6.8)-(6.9) dans le cas où l'opérateur spatial est discrétisé par une méthode d'éléments finis Lagrange, l'EDP parabolique semi-linéaire (6.9) est intégrée en temps par le schéma $SBDF$ et (6.8) est intégrée par le schéma RL. On va étudier en particulier, la convergence du problème semi-discret obtenu après la discrétisation en espace de (6.8)-(6.9), et la convergence globale des schémas $\mathbb{P}_r + RL1 + FBE$ et $\mathbb{P}_r + RL2 + SBDF2$ pour le problème (6.8)-(6.9). Pour cela, on aura besoin de faire des hypothèses sur les fonctions a, b, f définissant le problème (6.8)-(6.9). Commençons par donner quelques définitions qui nous seront utiles.

Définition 6 (Projection elliptique sur V_h)**.** On appelle projection elliptique sur V_h, l'application $\mathcal{R}_h : H^1(\Omega) \to V_h$ telle que pour $w \in H^1(\Omega)$,

$$(\sigma \nabla \mathcal{R}_h w, \nabla \chi) = (\sigma \nabla w, \nabla \chi) \quad \forall \chi \in V_h \tag{6.29}$$

Définition 7 (Projection L^2 sur V_h)**.** On appelle projection L^2 sur V_h, l'application $P_h : L^2(\Omega) \to V_h$ telle que pour $w \in L^2(\Omega)$,

$$(P_h w, \chi) = (w, \chi) \quad \forall \chi \in V_h \tag{6.30}$$

On fait les hypothèses suivantes sur $\mathcal{R}_h$, P_h, les fonctions a, b, f et la solution (v, u) du problème (6.8)-(6.9) :

Hypothèse 1. On suppose que la solution (v, u) du problème (6.8)-(6.9) existe, est unique et est assez régulière (au moins $C^3([0,T], H^r(\Omega))$). On suppose aussi que v et u sont dans $L^\infty(0,T,L^\infty(\Omega))$ ainsi que leurs dérivées.

Hypothèse 2. On suppose que l'espace V_h est tel que,

$$\|w - \mathcal{R}_h w\| \leq C h^r \|w\|_r \quad \forall w \in H^r(\Omega), \tag{6.31}$$

$$\|w - P_h w\| \leq C h^r \|w\|_r \quad \forall w \in H^r(\Omega). \tag{6.32}$$

Hypothèse 3. On suppose que les fonctions $a : \mathbb{R} \to \mathbb{R}$, $b : \mathbb{R}^2 \to \mathbb{R}$ et $f : \mathbb{R}^3 \to \mathbb{R}$ sont des fonctions assez régulières (au moins C^3).

La question sur l'existence, l'unicité et la régularité de la solution (u, v) du problème (6.8)-(6.9) n'a pas été abordée dans ce travail. Nous avons exprimé cela par l'hypothèse 1. Cette hypothèse nous permet de nous assurer que la solution de notre problème ainsi que ses dérivées restent bornées presque partout sur $[0,T] \times \Omega$. L'hypothèse 2 est un résultat classique de la méthode des élément finis (voir [7, 5, 10]) que nous avons admis pour pouvoir faire nos preuves. Elle exprime l'ordre de convergence de la méthode de discrétisation en espace. L'hypothèse 3 permet de nous assurer que les fonctions composées $a(u)$, $b(v,u)$ et $f(.,v,u)$ ainsi que leurs dérivées restent dans $L^\infty(0,T,L^\infty(\Omega))$ pour toutes fonctions $v, u \in L^\infty(0,T,L^\infty(\Omega))$. Ces hypothèses vont nous permettre de montrer les théorèmes de convergence. Elles vont aussi nous permettre de définir certains outils qui seront utiles dans les preuves.

Définition 8 (Boule autour de (v,u))**.** Sous l'hypothèse 1, on définit autour des fonctions v et u (solutions de (6.8)-(6.9)) les boules de rayon $R > 0$, $\mathcal{B}_v(R)$ et $\mathcal{B}_u(R)$ telles que,

$$\mathcal{B}_v(R) = \{w \in L^\infty(0,T,L^\infty(\Omega));\ \|v - w\|_{L^\infty(0,T,L^\infty(\Omega))} < R\} \tag{6.33}$$

$$\mathcal{B}_u(R) = \{w \in L^\infty(0,T,L^\infty(\Omega) \cap H^1(\Omega));\ \|u - w\|_{L^\infty(0,T,L^\infty(\Omega))} < R\} \tag{6.34}$$

Proposition 1. On suppose que l'hypothèse 3 est vérifiée. Soient $t \in [0,T]$, $(x_1,y_1),(x_2,y_2) \in \mathcal{B}_v(R) \times \mathcal{B}_u(R)$. Alors des constantes L_a, L_b, L_f et L_{aR} strictement positifs et ne dépendant que de u, v et R tels que,

$$\|a(x_1(t)) - a(x_2(t))\| \leq L_a \|x_1(t) - x_2(t)\|, \tag{6.35}$$

$$\|b(x_1(t),y_1(t)) - b(x_2(t),y_2(t)\| \leq L_b \left(\|x_1(t) - x_2(t)\| + \|y_1(t) - y_2(t)\|\right), \tag{6.36}$$

$$\|f(t,x_1(t),y_1(t)) - f(t,x_2(t),y_2(t))\| \leq L_f \left(\|x_1(t) - x_2(t)\| + \|y_1(t) - y_2(t)\|\right), \tag{6.37}$$

$$\|a(y_1(t))x_1(t) - a(y_2(t))x_1(t)\| \leq L_{aR} \left(\|x_1(t) - x_2(t)\| + \|y_1(t) - y_2(t)\|\right) \tag{6.38}$$

Preuve. Soit $t_n \in [0,T]$ et $x_i, y_i \in \mathcal{B}_v(R)$ (resp : $\mathcal{B}_u(R)$), $i = 1,2$. En posant

$$R_{vu} = \|v\|_{L^\infty(0,T,L^\infty(\Omega))} + \|u\|_{L^\infty(0,T,L^\infty(\Omega))} + R, \tag{6.39}$$

on a,

$$|x_i(t,x)| \leq R_{vu} \text{ pour presque tout } t \in [0,T] \text{ et } x \in \Omega \tag{6.40}$$

$$|y_i(t,x)| \leq R_{vu} \text{ pour presque tout } t \in [0,T] \text{ et } x \in \Omega \tag{6.41}$$

Les fonction a, b et f étant localement Lipschitz par l'hypothèse 3, elles le sont en particulier dans $[-R_{vu}, R_{vu}]$ pour a, $[-R_{vu}, R_{vu}]^2$ pour b et $[0,T] \times [-R_{vu}, R_{vu}]^2$ pour f. Ainsi, il existe $La, L_b, L_f > 0$ tels que pour presque tout $t \in [0,T]$ et $x \in \Omega$,

$$|a(x_1(t,x)) - a(x_2(t,x))| \leq L_a |x_1(t,x) - x_2(t,x)|, \tag{6.42}$$

$$|b(x_1(t,x),x_2(t,x)) - a(x_2(t,x),y_2(t,x))| \leq L_b \left(|x_1(t,x) - x_2(t,x)|^2 + |y_1(t,x) - y_2(t,x)|^2\right)^{\frac{1}{2}},$$

$$|f(t,x_1(t,x),x_2(t,x)) - f(t,x_2(t,x),y_2(t,x)) \leq L_f \left(|x_1(t,x) - x_2(t,x)|^2 + |y_1(t,x) - y_2(t,x)|^2\right)^{\frac{1}{2}}.$$

Par une élévation au carré et une intégration dans Ω des relations (6.42), on obtient les inégalités (6.35)-(6.37).
Il reste donc à montrer (6.38).

$$\begin{aligned}
&\|a(y_1(t))x_1(t) - a(y_2(t))x_2(t)\|^2 = \int_\Omega |a(y_1(t,,x))x_1(t,x) - a(y_2(t,x))x_2(t,x)|^2 dx \\
&\leq \int_\Omega \left(|a(y_1(t,x))||x_1(t,x) - x_2(t,x)| + |x_2(t,x)||a(y_1(t,x)) - a(y_2(t,x))|\right)^2 dx \\
&\leq L \int_\Omega \left(|x_1(t,x) - x_2(t,x)| + |a(y_1(t,x)) - a(y_2(t,x))|\right)^2 dx,
\end{aligned} \tag{6.43}$$

avec $L = \|a(y_1)\|^2_{L^\infty(0,T,L^\infty(\Omega))} + \|x_2\|^2_{L^\infty(0,T,L^\infty(\Omega))}$. On utilise dans (6.43) la relation $(|t_1| + |t_2|)^2 \leq 2(|t_1|^2 + |t_2|^2)$; $t_1, t_2 \geq 0$ pour obtenir,

$$\|a(y_1(t))x_1(t) - a(y_2(t))x_2(t)\|^2 \leq 2L \left(\|x_1(t) - x_2(t)\|^2 + \|y_1(t) - y_2(t)\|^2\right). \tag{6.44}$$

Puisque a est localement Lipschitz par l'hypothèse 3, et les fonctions x_1, x_2, y_1 et y_2 sont dans $\mathcal{B}_v(R) \times \mathcal{B}_u(R)$, on peut dire que la constante L ne dépend que de a et R. De (6.44), on conclut que,

$$\|a(y_1(t))x_1(t) - a(y_2(t))x_2(t)\| \leq L_{aR} \left(\|x_1(t) - x_2(t)\| + \|y_1(t) - y_2(t)\| \right), \tag{6.45}$$

avec L_{aR} ne dépendant que de a et R. D'où le résultat. □

Remarque 3. Dans nos résultats de convergence, nous ferons toujours l'hypothèse que les solutions discrètes et semi-discrètes des problèmes que nous étudions restent dans $\mathcal{B}_v(R) \times \mathcal{B}_u(R)$. Cette hypothèse nous permet simplement d'exploiter le caractère lipschitz local des fonctions a, b et f en utilisant la proposition 1. Cependant, si on suppose que les fonctions a, b et f sont globalement lipschitz, on n'a pas besoin de faire cette hypothèse et nos preuves restent valides.

Soit (v_h, u_h) la solution du problème discret (6.11)-(6.12). On définit pour la suite les notations suivantes.

$$\theta_v(t) = v_h(t) - P_h v(t), \qquad \rho_v(t) = P_h v(t) - v(t) \tag{6.46}$$

$$\theta_u(t) = u_h(t) - \mathcal{R}_h u(t), \qquad \rho_u(t) = \mathcal{R}_h u(t) - u(t). \tag{6.47}$$

Avec P_h la projection L^2 et $\mathcal{R}_h$ la projection elliptique définie par (6.29) et (6.30).

6.4.1 Étude de la convergence en espace

Dans cette section, nous étudions la convergence du problème semi-discret (6.11)-(6.12) obtenu en discrétisant le problème (6.8)-(6.9) par la méthode des éléments finis $\mathbb{P}_r$

Théorème 1 (convergence du problème semi-discret)**.** Soit (v, u) la solution de (6.8)-(6.9). On suppose que les hypothèses 1 à 3 sont vérifiées. On suppose aussi que la solution (v_h, u_h) de (6.11)-(6.12) existe, est unique et reste dans $\mathcal{B}_v(R) \times \mathcal{B}_u(R)$. Alors, il existe une constante $C > 0$ ne dépendant que de R, des données du problème (6.8)-(6.9) et de (v, u), telle que pour tout $t \leq T$, ,

$$\|v_h(t) - v(t)\| + \|u_h(t) - u(t)\| \leq C \left(\|\theta_v(0)\| + \|\theta_u(0)\| + Ch^r \right) \tag{6.48}$$

Dans toutes les preuves qui vont suivre, C désignera une constante positive ne dépendant pas du diamètre du maillage h, ni du pas de temps de discrétisation Δt. Cette constante ne sera pas la même tout le temps. Elle sera donc une constante permettant de majorer des expressions. Ceci sera dans le but de présenter lisiblement nos majorations.

Preuve. Soit $t < T$, par l'inégalité triangulaire on a,

$$\begin{aligned}\|v_h(t) - v(t)\| + \|u_h(t) - u(t)\| \leq \|v_h(t) - P_h v(t)\| + \|P_h v(t) - v(t)\| \\ + \|v_h(t) - \mathcal{R}_h u(t)\| + \|\mathcal{R}_h u(t) - u(t)\|,\end{aligned} \tag{6.49}$$

avec P_h et $\mathcal{R}_h$ les projections définies par (6.29) et (6.30).
En utilisant les notations données par (6.46), l'inégalité (6.49) devient,

$$\|v_h(t) - v(t)\| + \|u_h(t) - u(t)\| \leq \|\theta_v(t)\| + \|\theta_u(t)\| + \|\rho_v(t)\| + \|\rho_u(t)\|. \tag{6.50}$$

L'hypothèse 2 utilisée dans l'inégalité (6.50) nous permet d'écrire,

$$\|v_h(t) - v(t)\| + \|u_h(t) - u(t)\| \leq \|\theta_v(t)\| + \|\theta_u(t)\| + Ch^r. \tag{6.51}$$

Il reste donc à majorer $\|\theta_u(t)\|$ et $\|\theta_v(t)\|$. Soit $\varphi_h \in V_h$ on a,

$$\begin{aligned}(\partial_t \theta_u, \varphi_h) + (\sigma \nabla \theta_u, \nabla \varphi_h) &= (f(t, v_h, u_h), \varphi_h) - (\partial_t \mathcal{R}_h u, \varphi_h) - (\sigma \nabla \mathcal{R}_h u, \nabla \varphi_h) \\ &= (f(t, v_h, u_h), \varphi_h) - (\sigma \nabla u, \nabla \varphi_h) - (\partial_t \mathcal{R}_h u, \varphi_h) \\ &= (f(t, v_h, u_h) - f(t, v, u), \varphi_h) - (\partial_t \rho_u, \varphi_h).\end{aligned} \tag{6.52}$$

On prend $\varphi_h = \theta_u$ dans (6.52) et on y utilise respectivement l'inégalité de Cauchy-Schwarz et l'inégalité lipschitz de f (6.37) de la proposition 1 pour obtenir,

$$\begin{aligned}\frac{1}{2}\partial_t \|\theta_u\|^2 + (\sigma \nabla \theta_u, \nabla \theta_u) &\leq \|f(t, v_h, u_h) - f(t, v, u)\| \|\theta_u\| + \|\partial_t \rho_u\| \|\theta_u\| \\ &\leq L_f (\|u_h - u\| + \|v_h - v\|) \|\theta_u\| + \|\partial_t \rho_u\| \|\theta_u\|.\end{aligned} \tag{6.53}$$

Comme pour l'inégalité (6.49), on introduit dans (6.53) les termes $P_h v$ et $\mathcal{R}_h u$, puis on utilise l'inégalité triangulaire et les notations introduites en (6.46) et (6.47) pour obtenir,

$$\frac{1}{2}\partial_t \|\theta_u\|^2 + (\sigma \nabla \theta_u, \nabla \theta_u) \leq L_f (\|\theta_u\| + \|\theta_v\| + \|\rho_u\| + \|\rho_v\|) \|\theta_u\| + \|\partial_t \rho_u\| \|\theta_u\|. \tag{6.54}$$

En utilisant dans (6.54) l'inégalité de Young et la propriété de positivité $(\sigma \nabla \theta_u, \nabla \theta_u) \geq 0$, on obtient,

$$\frac{1}{2}\partial_t \|\theta_u\|^2 \leq \frac{L_f}{2}\left(\|\theta_u\|^2 + \|\theta_v\|^2 + \|\rho_u\|^2 + \|\rho_v\|^2 + \|\rho_{\partial_t u}\|^2\right) + \frac{2L_f + 1}{2}\|\theta_u\|^2, \tag{6.55}$$

avec $\rho_{\partial_t u} = \mathcal{R}_h \partial_t u - \partial_t u$. On utilise dans (6.55) les inégalités (6.31) et (6.32) pour obtenir,

$$\frac{1}{2}\partial_t \|\theta_u\|^2 \leq \frac{L_f}{2}\left(\|\theta_u\|^2 + \|\theta_v\|^2 + Ch^{2r}\right) + \frac{2L_f + 1}{2}\|\theta_u\|^2 \tag{6.56}$$

Par une multiplication par 2 de l'inégalité (6.56) et un arrangement des termes, on obtient,

$$\partial_t \|\theta_u\|^2 \leq C\left(\|\theta_u\|^2 + \|\theta_v\|^2 + h^{2r}\right). \tag{6.57}$$

Par ailleurs, on a,

$$\begin{aligned}(\partial_t\theta_v,\varphi_h) &= (\partial_t v_h,\varphi_h) - (\partial_t P_h v,\varphi_h)\\ &= (a(u_h)v_h + b(v_h,u_h),\varphi_h) - (\partial_t v,\varphi_h)\\ &= (a(u_h)v_h + b(v_h,u_h) - a(u)v - b(v,u),\varphi_h)\end{aligned} \tag{6.58}$$

Pour $\varphi_h = \theta_v$ et en utilisant respectivement l'inégalité de Cauchy-Schwarz et l'inégalité triangulaire dans (6.59), on a

$$\begin{aligned}\frac{1}{2}\partial_t \|\theta_v\|^2 &\leq \|a(u_h)v_h + b(v_h,u_h) - a(u)v - b(v,u)\| \|\theta_v\|,\\ &\leq \|a(u_h)v_h - a(u)v\| \|\theta_v\| + \|b(v_h,u_h) - b(v,u)\| \|\theta_v\|.\end{aligned} \tag{6.59}$$

En utilisant dans (6.59) les relations (6.36) et (6.38) de la proposition 1, on obtient,

$$\frac{1}{2}\partial_t \|\theta_v\|^2 \leq (L_{aR} + L_b)\left(\|v_h - v\| + \|u_h - u\|\right)\|\theta_v\|. \tag{6.60}$$

Comme pour l'inégalité (6.49), on introduit dans (6.60) les termes $P_h v$ et $\mathcal{R}_h u$, puis on utilise l'inégalité triangulaire et les notations introduites en (6.46) et (6.47) pour obtenir

$$\frac{1}{2}\partial_t \|\theta_v\|^2 \leq (L_{aR} + L_b)\left(\|\theta_v\| + \|\theta_u\| + \|\rho_v\| + \|\rho_u\|\right)\|\theta_v\|. \tag{6.61}$$

On utilise l'inégalité de Young dans (6.61) pour obtenir,

$$\frac{1}{2}\partial_t \|\theta_v\|^2 \leq \frac{L_{aR} + L_b}{2}\left(\|\theta_v\|^2 + \|\theta_u\|^2 + \|\rho_v\|^2 + \|\rho_u\|^2\right) + 2(L_{aR} + L_b)\|\theta_v\|^2. \tag{6.62}$$

On obtient par un arrangement des termes et les inégalités (6.32)-(6.31) de l'hypothèse 2 la majoration,

$$\partial_t \|\theta_v\|^2 \leq C\left(\|\theta_u\|^2 + \|\theta_v\|^2 + h^{2r}\right) \tag{6.63}$$

Des inégalités (6.63)-(6.57), on obtient,

$$\partial_t\left(\|\theta_v\|^2 + \|\theta_u\|^2\right) \leq C\left(\|\theta_u\|^2 + \|\theta_v\|^2 + h^{2r}\right) \tag{6.64}$$

Par une intégration de (6.64) dans l'intervalle $[0,t]$,

$$\begin{aligned}\|\theta_u(t)\|^2 + \|\theta_v(t)\|^2 \leq \|\theta_u(0)\|^2 + \|\theta_v(0)\|^2 + CTh^{2r}\\ + \int_0^t C\left(\|\theta_u(\tau)\|^2 + \|\theta_v(\tau)\|^2\right)d\tau.\end{aligned} \tag{6.65}$$

Par le lemme de Grönwall (voir [4] Chapitre 10, Lemme 10.1) on a

$$\|\theta_u(t)\|^2 + \|\theta_v(t)\|^2 \leq \left(\|\theta_u(0)\|^2 + \|\theta_v(0)\|^2 + h^{2r}\right)e^{CT}. \tag{6.66}$$

Il en découle donc l'inégalité recherchée. □

Dans la suite, $(t_n)_{0\leq n\leq m}$ désignera une suite de réels discrétisant l'intervalle $[0,T]$. On notera par g^n l'image en t_n d'une fonction $g \in, L^2(0,T,L^2(\Omega))$

6.4.2 Convergence de $\mathbb{P}_r + RL1 + FBE$

Dans cette section, on va s'intéresser aux fonctions : (v,u) solution du problème (6.8)-(6.9) et $(\overline{v}^n, \overline{u}^n)$ solution numérique approchant (v,u) par le schéma $\mathbb{P}_r + RL1 + FBE$ aux instants t_n. La solution numérique $(\overline{v}^n, \overline{u}^n)$ est définie par,

$$\left(\overline{v}^{n+1}, \psi_h\right) = (\overline{v}^n, \psi_h) + \Delta t \left(\varphi_1(\overline{\alpha}_n \Delta t)(\overline{\alpha}_n \overline{v}^n + \overline{\beta}_n), \psi_h\right) \qquad \forall \psi_h \in V_h, \tag{6.67}$$

$$\left(\frac{\overline{u}^{n+1} - \overline{u}^n}{\Delta t}, \varphi_h\right) + \left(\sigma \nabla \overline{u}^{n+1}, \nabla \varphi_h\right) = (f(t_n, \overline{v}^n, \overline{u}^n), \varphi_h) \qquad \forall \varphi_h \in V_h, \tag{6.68}$$

avec $\overline{\alpha}_n = a(\overline{u}^n)$ et $\overline{\beta}_n = b(\overline{v}^n, \overline{u}^n)$.

Définition 9 (Erreur de consistance de $RL1 + FBE$)**.** Soit (v,u) solution de (6.8)-(6.9). Les erreurs de consistances E_u^n et E_v^n en u et v respectivement du schéma $RL1 + FBE$ sont définies par,

$$\left(v^{n+1}, \psi\right) = (v^n, \psi) + \Delta t \left(\varphi_1(\alpha_n \Delta t)(\alpha_n v^n + \beta_n), \psi\right) + \Delta t \left(E_v^n, \psi\right) \quad \forall \psi \in L^2(\Omega), \tag{6.69}$$

$$\left(\frac{u^{n+1} - u^n}{\Delta t}, \varphi\right) + \left(\sigma \nabla u^{n+1}, \nabla \varphi\right) = (f(t_n, v^n, u^n), \varphi) + (E_u^n, \varphi) \quad \forall \varphi \in H^1(\Omega), \tag{6.70}$$

avec $\alpha_n = a(u^n) = a^n$, $\beta_n = b(v^n, u^n) = b^n$, $u^n = u(t_n)$, $v^n = v(t_n)$ et $t_n = n\Delta t$.

Théorème 2 (Consistance $RL1 + FBE$)**.** On suppose que les hypothèses 1 et 3 sont vérifiées. Les erreurs de consistances E_u^n et E_v^n pour le schéma $RL1 + FBE$ vérifient

$$\|E_u^n\|_{V'} + \|E_v^n\|_{V'} \leq C\Delta t \tag{6.71}$$

Avec C une constante dépendant uniquement des données du problème (6.8)-(6.9) et de la solution (v,u).

Preuve. On pose $\partial_1 u^{n+1} = \frac{u^{n+1}-u^n}{\Delta t}$. Soit $\varphi \in H^1(\Omega)$. Par les égalités (6.70) et (6.9) on a,

$$\begin{aligned} (E_u^n, \varphi) &= \left(\partial_1 u^{n+1} - \partial_t u^{n+1}, \varphi\right) + \left(f(t_{n+1}, v^{n+1}, u^{n+1}) - f(t_n, v^n, u^n), \varphi\right) \\ |\,(E_u^n, \varphi)\,| &\leq \left(\left\|\partial_1 u^{n+1} - \partial_t u^{n+1}\right\| + \left\|f(t_{n+1}, v^{n+1}, u^{n+1}) - f(t_n, v^n, u^n)\right\|\right) \|\varphi\| \end{aligned} \tag{6.72}$$

Or par les hypothèses de régularité 1, 3 sur (v,u) et f , on a

$$\begin{aligned}\|f(t_{n+1},v^{n+1},u^{n+1})-f(t_n,v^n,u^n)\| &\leq \|\int_{t_n}^{t_{n+1}}\partial_t f(\tau,v(\tau),u(\tau))d\tau\|\\ &\leq \int_{t_n}^{t_{n+1}}\|\partial_t f(\tau,v(\tau),u(\tau))\|\,d\tau\\ &\leq \Delta t\|\partial_t f(.,v,u)\|_{L^\infty(0,T,L^2(\Omega))}\leq C\Delta t \qquad (6.73)\end{aligned}$$

$$\|\partial_1 u^{n+1}-\partial_t u^{n+1}\| = \frac{1}{\Delta t}\|\int_{t_n}^{t_{n+1}}(\tau-t_n)\partial_{tt}u(\tau)d\tau\| \leq C\Delta t \qquad (6.74)$$

L'égalité dans la relation (6.74) se montre en utilisant une intégration par partie dans l'intégrale $\int_{t_n}^{t_{n+1}}(\tau-t_n)\partial_{tt}u(\tau)d\tau$.
En utilisant les relations (6.73)-(6.74) dans (6.72), on obtient,

$$\frac{|\,(E_u^n,\varphi)\,|}{\|\varphi\|}\leq C\Delta t \quad \text{ et donc } \quad \|E_u^n\|_{V'}\leq C\Delta t. \qquad (6.75)$$

Par ailleurs, de (6.69) on a pour $\psi\in L^2(\Omega)$,

$$\Delta t\,(E_v^n,\varphi) = \left(v^{n+1}-v^n-\Delta t\varphi_1(\alpha_n\Delta t)(\alpha_n v^n+\beta_n),\varphi\right). \qquad (6.76)$$

On pose,

$$z(\tau) = v^n+\tau\varphi_1(\alpha_n\tau)(\alpha_n v^n+\beta_n)\ presque\ partout\ \text{ dans } \Omega. \qquad (6.77)$$

Alors z est la solution du problème,

$$\partial_t z = \alpha_n z+\beta_n \quad , \quad z(0)=v^n\ presque\ partout \text{ dans } \Omega. \qquad (6.78)$$

Pour presque tout $x\in\Omega$, il existe par la formule de Taylor, τ_{1x}^n et τ_{2x}^n dans $]0,\Delta t[$ tels que,

$$v(t_{n+1},x)-z(\Delta t,x) = \frac{\Delta t^2}{2}\left(\partial_{tt}v(t_n+\tau_{1x}^n,x)-\partial_{tt}z(\tau_{2x}^n,x)\right). \qquad (6.79)$$

En effet, on a d'une part,

$$v(t_{n+1},x) = v(t_n,x)+\Delta t\partial_t v(t_n,x)+\frac{\Delta t^2}{2}\partial_{tt}v(t_n+\tau_{1x}^n,x). \qquad (6.80)$$

On utilise dans (6.80), l'expression de $\partial_t v$ donnée par (6.8) pour obtenir,

$$\begin{aligned}v(t_{n+1},x) = v(t_n,x)&+\Delta t\left(a(u(t_n,x))v(t_n,x)+b(v(t_n,x),u(t_n,x))\right)\\ &+\frac{\Delta t^2}{2}\partial_{tt}v(t_n+\tau_{1x}^n,x). \qquad (6.81)\end{aligned}$$

D'autre part,

$$z(\Delta t, x) = z(0, x) + \Delta t \partial_t z(0, x) + \frac{\Delta t^2}{2} \partial_{tt} z(\tau_{2x}^n, x). \tag{6.82}$$

$$\tag{6.83}$$

On utilise (6.78) dans (6.82) pour obtenir,

$$\begin{aligned} z(\Delta t, x) = v(t_n, x) + \Delta t \left(a(u(t_n, x)) v(t_n, x) + b(v(t_n, x), u(t_n, x)) \right) \\ + \frac{\Delta t^2}{2} \partial_{tt} z(\tau_{2x}^n, x). \end{aligned} \tag{6.84}$$

La soustraction de (6.84) à (6.81) donne alors l'égalité (6.79).
On utilise l'égalité (6.79) pour obtenir l'inégalité,

$$\left\| v^{n+1} - z(\Delta t) \right\| \leq C \Delta t^2, \tag{6.85}$$

avec $C = \frac{1}{2} \left(\|\partial_{tt} v\|_{L^\infty(0,T,L^\infty(\Omega))} + \|\partial_{tt} z\|_{L^\infty(0,T,L^\infty(\Omega))} \right) |\Omega|^{\frac{1}{2}}$. Notons que la fonction z définie par (6.77) ne dépend que de la solution (v, u) de (6.8)-(6.9). Il en est donc de même pour la constante C
On utilise dans (6.76), la fonction z définie en (6.77) pour obtenir,

$$(E_v^n, \varphi) = \frac{1}{\Delta t} \left(v^{n+1} - z(\Delta t), \varphi \right). \tag{6.86}$$

En appliquant sur (6.86), l'inégalité de Cauchy-Schwarz et (6.85) on a l'inégalité,

$$\begin{aligned} |\, (E_v^n, \varphi)\, | &\leq \frac{1}{\Delta t} \left\| v^{n+1} - z(\Delta t) \right\| \|\varphi\| \\ &\leq C \Delta t \, \|\varphi\| \end{aligned}$$

On obtient donc l'inégalité,

$$\frac{|\, (E_v^n, \varphi)\, |}{\|\varphi\|} \leq C \Delta t \quad \text{et donc} \quad \|E_v^n\|_{V'} \leq C \Delta t \tag{6.87}$$

De (6.75) et (6.87), on a le résultat recherché. □

Proposition 2. On suppose que l'hypothèse 3 est vérifiée. Soient $t_n = n\Delta t$, $(x_1, y_1), (x_2, y_2) \in \mathcal{B}_v(R) \times \mathcal{B}_u(R)$. On pose $\alpha_n^i = a(y_i^n)$, $\beta_n^i = b(x_i^n, y_i^n)$, $x_i^n = x_i(t_n, .)$ et $y_i^n = y_i(t_n, .)$, $i = 1, 2$. Alors,

$$\begin{aligned} \|\varphi_1(\alpha_n^1 \Delta t)(\alpha_n^1 x_1^n + \beta_n^1) - \varphi_1(\alpha_n^2 \Delta t)(\alpha_n^2 x_2^n + \beta_n^2)\| \\ \leq L_{ab\varphi} \left(\|x_1^n - x_2^n\| + \|y_1^n - y_2^n\| \right) \end{aligned} \tag{6.88}$$

où $L_{ab\varphi}$ est une constante positive ne dépendant que de R, des fonctions a, b, φ_1 et de la solution (v, u).

Preuve. Soient $(x_1, y_1), (x_2, y_2) \in \mathcal{B}_v(R) \times \mathcal{B}_u(R)$. Commençons par montrer que $t \in \mathbb{R} \mapsto \varphi_1(t) \in \mathbb{R}$ est localement Lipschitz. Pour $|t| \leq M$, M réel strictement positif,

$$\partial_t \varphi_1(t) = \int_0^1 (1-\tau) e^{t(1-\tau)} d\tau \leq e^M \int_0^1 |1-\tau| d\tau \leq e^M \tag{6.89}$$

On conclut d'après (6.89) que $\varphi_1 : \mathbb{R} \to \mathbb{R}$ est localement Lipschitz. On peut comme pour les fonctions a, b et f dans la preuve de la proposition 1 montrer que,

$$\|\varphi_1(y_1(t)) - \varphi_1(y_2(t))\| \leq L_{\varphi_1} \|y_1(t) - y_2(t)\|. \tag{6.90}$$

Par ailleurs, *pour presque tout* x dans Ω,

$$\begin{aligned}
&|\varphi_1(\alpha_n^1(x)\Delta t)(\alpha_n^1(x)x_1^n(x) + \beta_n^1(x)) - \varphi_1(\alpha_n^2(x)\Delta t)(\alpha_n^2(x)x_2^n(x) + \beta_n^2(x))| \\
&\qquad \leq |\varphi_1(\alpha_n^1(x)\Delta t)| \left(|\alpha_n^1(x)x_1^n(x) - \alpha_n^2(x)x_2^n(x)| + |\beta_n^1(x) - \beta_n^2(x)| \right) \\
&\qquad\quad + |\varphi_1(\alpha_n^1(x)\Delta t) - \varphi_1(\alpha_n^2(x)\Delta t)||\alpha_n^2(x)x_2^n(x) + \beta_n^2(x)| \\
&\qquad \leq |\varphi_1(\alpha_n^1(x)\Delta t)| \left(|\alpha_n^1(x)||x_1^n(x) - x_2^n(x)| + |\alpha_n^1(x) - \alpha_n^2(x)||x_2^n(x)| \right) \\
&\quad + |\varphi_1(\alpha_n^1(y)\Delta t)||\beta_n^1(x) - \beta_n^2(x)| + |\varphi_1(\alpha_n^1(y)\Delta t) - \varphi_1(\alpha_n^2(y)\Delta t)||\alpha_n^2(y)x_2^n(x) + \beta_n^2(x)| \\
&\qquad \leq L_{ab\varphi}(|x_1(t,x) - x_2(t,x)|^2 + |y_1(t,x) - y_2(t,x)|^2)^{\frac{1}{2}}.
\end{aligned} \tag{6.91}$$

La majoration (6.91) est justifiée d'une part par le caractère Lipschitz locale des fonctions a, b et φ_1. D'autre part, les fonctions a, b et φ_1 sont continues et donc localement bornées. Les majorations

$$|x_i^n(x)| \leq \|x_i\|_{L^\infty(0,T,L^\infty(\Omega))} \leq \|u\|_{L^\infty(0,T,L^\infty(\Omega))} + \|v\|_{L^\infty(0,T,L^\infty(\Omega))} + R = R_{vu} \tag{6.92}$$

$$|y_i^n(x)| \leq \|y_i\|_{L^\infty(0,T,L^\infty(\Omega))} \leq \|u\|_{L^\infty(0,T,L^\infty(\Omega))} + \|v\|_{L^\infty(0,T,L^\infty(\Omega))} + R = R_{vu} \tag{6.93}$$

nous permettent de dire que $|\alpha_n^i(x)|$ et $|\beta_n^i(x)|$ sont uniformément bornées. Il en est de même pour $|\varphi_1(\alpha_n^i(x)\Delta t)|$. Le résultat cherché est donc obtenu par une élévation au carré et une intégration dans Ω de (6.91). □

On rappelle que $(\overline{v}^{n+1}, \overline{u}^{n+1})$ est la solution du problème (6.67)-(6.68). On définit pour la suite les notations suivantes.

$$\overline{\theta}_v^{n+1} = \overline{v}^{n+1} - P_h v^{n+1}, \qquad \rho_v^{n+1} = P_h v^{n+1} - v^{n+1}. \tag{6.94}$$

$$\overline{\theta}_u^{n+1} = \overline{u}^{n+1} - \mathcal{R}_h u^{n+1}, \qquad \rho_u^{n+1} = \mathcal{R}_h u^{n+1} - u^{n+1}. \tag{6.95}$$

Avec P_h la projection L^2 et $\mathcal{R}_h$ la projection elliptique définie par (6.29) et (6.30).

Théorème 3 (Convergence de $\mathbb{P}_r + RL1 + FBE$)**.** On suppose que les hypothèses 1 à 3 sont vérifiées. On suppose aussi que la solution numérique $(\overline{v}^n, \overline{u}^n)$ définit par (6.67)-(6.68) existe, est unique et reste dans $\mathcal{B}_v(R) \times \mathcal{B}_u(R)$. Alors, il existe une constante C_s ne dépendant pas de h, une constante C strictement positive ne dépendant que de R, des données et de la solution (v, u) du problème (6.8)-(6.9), telle que pour tout $\Delta t < C_s$,

$$\left\|\overline{v}^{n+1} - v^{n+1}\right\| + \left\|\overline{u}^{n+1} - u^{n+1}\right\| \leq C\left(\left\|\overline{v}^0 - P_h v^0\right\| + \left\|\overline{u}^0 - \mathcal{R}_h v^0\right\| + h^r + \Delta t\right). \quad (6.96)$$

Preuve. Par l'inégalité triangulaire on a,

$$\left\|\overline{v}^{n+1} - v^{n+1}\right\| \leq \left\|\overline{v}^{n+1} - P_h v^{n+1}\right\| + \left\|P_h v^{n+1} - v^{n+1}\right\|, \quad (6.97)$$

avec P_h la projection définie en (6.30). On utilise les notations données en (6.94) pour obtenir,

$$\left\|\overline{v}^{n+1} - v^{n+1}\right\| \leq \left\|\overline{v}^{n+1} - P_h v^{n+1}\right\| + \left\|P_h v^{n+1} - v^{n+1}\right\| = \left\|\overline{\theta}_v^{n+1}\right\| + \left\|\rho_v^{n+1}\right\| \quad (6.98)$$

Soit $\psi_h \in V_h$ en utilisant (6.67) et (6.69),

$$\begin{aligned}
\left(\overline{\theta}_v^{n+1}, \psi_h\right) &= \left(\overline{v}^{n+1} - P_h v^{n+1}, \psi_h\right) \\
&= \left(\overline{v}^{n+1} - v^{n+1}, \psi_h\right) \\
&= \left(\overline{v}^n - v^n, \psi_h\right) + \Delta t\left(E_v^n, \psi_h\right) \\
&+ \Delta t\left(\varphi_1(\overline{\alpha}_n \Delta t)(\overline{\alpha}_n \overline{v}^n + \overline{\beta}_n) - \varphi_1(\alpha_n \Delta t)(\alpha_n v^n + \beta_n), \psi_h\right) \\
&= \left(\overline{v}^n - P_h v^n, \psi_h\right) + \Delta t\left(E_v^n, \psi_h\right) \\
&+ \Delta t\left(\varphi_1(\overline{\alpha}_n \Delta t)(\overline{\alpha}_n \overline{v}^n + \overline{\beta}_n) - \varphi_1(\alpha_n \Delta t)(\alpha_n v^n + \beta_n), \psi_h\right) \quad (6.99)
\end{aligned}$$

Dans l'égalité (6.99), on prend $\psi_h = \overline{\theta}_v^{n+1}$ et on utilise l'inégalité de Cauchy-Schwarz pour obtenir,

$$\begin{aligned}
\left\|\overline{\theta}_v^{n+1}\right\|^2 &\leq \left\|\overline{v}^n - P_h v^n\right\| \left\|\overline{\theta}_v^{n+1}\right\| + \Delta t \left\|E_v^n\right\|_{V'} \left\|\overline{\theta}_v^{n+1}\right\| \\
&+ \Delta t \left\|\varphi_1(\overline{\alpha}_n \Delta t)(\overline{\alpha}_n \overline{v}^n + \overline{\beta}_n) - \varphi_1(\alpha_n \Delta t)(\alpha_n v^n + \beta_n)\right\| \left\|\overline{\theta}_v^{n+1}\right\| \quad (6.100)
\end{aligned}$$

On utilise dans (6.100) l'inégalité (6.88) de la proposition 2 pour obtenir,

$$\begin{aligned}
\left\|\overline{\theta}_v^{n+1}\right\|^2 &\leq \left\|\overline{v}^n - P_h v^n\right\| \left\|\overline{\theta}_v^{n+1}\right\| + \Delta t \left\|E_v^n\right\|_{V'} \left\|\overline{\theta}_v^{n+1}\right\| \\
&+ \Delta t L_{ab\varphi}\left(\left\|\overline{v}^n - v^n\right\| + \left\|\overline{u}^n - u^n\right\|\right) \left\|\overline{\theta}_v^{n+1}\right\|. \quad (6.101)
\end{aligned}$$

On introduit dans (6.101) les termes $P_h v^n$ et $\mathcal{R}_h u^n$, puis on utilise l'inégalité triangulaire et les notations introduites en (6.94) et (6.95) pour obtenir,

$$\begin{aligned}\left\|\overline{\theta}_v^{n+1}\right\|^2 \leq & \left\|\overline{\theta}_v^{n}\right\| \left\|\overline{\theta}_v^{n+1}\right\| + \Delta t \left\|E_v^n\right\|_{V'} \left\|\overline{\theta}_v^{n+1}\right\| \\ & + \Delta t L_{ab\varphi} \left(\left\|\overline{\theta}_v^{n}\right\| + \left\|\rho_v^n\right\| + \left\|\overline{\theta}_u^{n}\right\| + \left\|\rho_u^n\right\|\right) \left\|\overline{\theta}_v^{n+1}\right\|\end{aligned} \tag{6.102}$$

Par l'inégalité de Young, l'inégalité (6.102) implique,

$$\begin{aligned}\left\|\overline{\theta}_v^{n+1}\right\|^2 \leq & \frac{1}{2}\left(\Delta t L_{ab\varphi} \left\|\overline{\theta}_u^{n}\right\|^2 + (1+\Delta t L_{ab\varphi}) \left\|\overline{\theta}_v^{n}\right\|^2\right) + \frac{\Delta t L_{ab\varphi}}{2}\left(\left\|\rho_v^n\right\|^2 + \left\|\rho_u^n\right\|^2\right) \\ & + \frac{\Delta t}{2}\left\|E_v^n\right\|_{V'}^2 + \frac{1}{2}\left\|\overline{\theta}_v^{n+1}\right\|^2 + \Delta t\left(2L_{ab\varphi} + \frac{1}{2}\right)\left\|\overline{\theta}_v^{n+1}\right\|^2\end{aligned} \tag{6.103}$$

Par un arrangement des termes, l'hypothèse 2 et l'inégalité (6.71) du théorème 2, l'inégalité (6.103) devient,

$$(1-\Delta t C)\left\|\overline{\theta}_v^{n+1}\right\|^2 \leq (1+\Delta t C)\left\|\overline{\theta}_v^{n}\right\|^2 + \Delta t C \left\|\overline{\theta}_u^{n}\right\|^2 + \Delta t C\left(h^{2r} + \Delta t^2\right) \tag{6.104}$$

On rappelle que,

$$\partial_1 u^{n+1} = \frac{u^{n+1} - u^n}{\Delta t}, \tag{6.105}$$

et on s'intéresse à présent à $\overline{\theta}_u^{n+1}$. Soit $\phi_h \in V_h$ on a

$$\begin{aligned}&\left(\partial_1 \overline{\theta}_u^{n+1}, \varphi_h\right) + \left(\sigma \nabla \overline{\theta}_u^{n+1}, \nabla \varphi_h\right) \\ &\quad = \left(\partial_1 \overline{u}^{n+1}, \varphi_h\right) + \left(\sigma \nabla \overline{u}^{n+1}, \nabla \varphi_h\right) - \left(\partial_1 \mathcal{R}_h u^{n+1}, \varphi_h\right) - \left(\sigma \nabla \mathcal{R}_h u^{n+1}, \nabla \varphi_h\right)\end{aligned} \tag{6.106}$$

On utilise dans l'égalité (6.106) la définition de la projection elliptique $\mathcal{R}_h$ donnée par (6.29) pour avoir,

$$\begin{aligned}&\left(\partial_1 \overline{\theta}_u^{n+1}, \varphi_h\right) + \left(\sigma \nabla \overline{\theta}_u^{n+1}, \nabla \varphi_h\right) \\ &\quad = \left(\partial_1 \overline{u}^{n+1}, \varphi_h\right) + \left(\sigma \nabla \overline{u}^{n+1}, \nabla \varphi_h\right) - \left(\partial_1 \mathcal{R}_h u^{n+1}, \varphi_h\right) - \left(\sigma \nabla u^{n+1}, \nabla \varphi_h\right)\end{aligned} \tag{6.107}$$

On utilise (6.68) et (6.9) dans (6.107) pour obtenir,

$$\begin{aligned}&\left(\partial_1 \overline{\theta}_u^{n+1}, \varphi_h\right) + \left(\sigma \nabla \overline{\theta}_u^{n+1}, \nabla \varphi_h\right) \\ &\quad = \left(f(t_n, \overline{v}^n, \overline{u}^n), \varphi_h\right) - \left(\partial_1 \mathcal{R}_h u^{n+1} - \partial_t u^{n+1}, \varphi_h\right) - \left(f(t_{n+1}, v^{n+1}, u^{n+1}), \varphi_h\right) \\ &\quad = \left(f(t_n, \overline{v}^n, \overline{u}^n) - f(t_n, P_h v^n, \mathcal{R}_h u^n), \varphi_h\right) + \left(f(t_n, P_h v^n, \mathcal{R}_h u^n) - f(t_n, v^n, u^n), \varphi_h\right) \\ &- \left(\partial_1 \rho_u^{n+1}, \varphi_h\right) - \left(\partial_1 u^{n+1} - \partial_t u^{n+1} + f(t_{n+1}, v^{n+1}, u^{n+1}) - f(t_n, v^n, u^n), \varphi_h\right).\end{aligned} \tag{6.108}$$

On utilise (6.70) et (6.9) dans (6.108) pour avoir,

$$\begin{aligned}&\left(\partial_1\overline{\theta}_u^{n+1},\varphi_h\right)+\left(\sigma\nabla\overline{\theta}_u^{n+1},\nabla\varphi_h\right)\\&\quad=\left(f(t_n,\overline{v}^n,\overline{u}^n)-f(t_n,P_hv^n,\mathcal{R}_hu^n),\varphi_h\right)+\left(f(t_n,P_hv^n,\mathcal{R}_hu^n)-f(t_n,v^n,u^n),\varphi_h\right).\\&\qquad-\left(\partial_1\rho_u^{n+1},\varphi_h\right)-\left(E_u^n,\varphi_h\right)\end{aligned}\tag{6.109}$$

Dans (6.109), on prend $\varphi_h=\overline{\theta}_u^{n+1}$ puis on utilise l'inégalité de Cauchy-Schwarz pour avoir,

$$\begin{aligned}&\frac{1}{2}\partial_1\|\overline{\theta}_u^{n+1}\|^2+\left(\sigma\nabla\overline{\theta}_u^{n+1},\nabla\overline{\theta}_u^{n+1}\right)\\&\quad\leq\left(\|f(t_n,\overline{v}^n,\overline{u}^n)-f(t_n,P_hv^n,\mathcal{R}_hu^n)\|+\|f(t_n,P_hv^n,\mathcal{R}_hu^n)-f(t_n,v^n,u^n)\|\right)\|\overline{\theta}_u^{n+1}\|\\&\qquad+\|E_u^n\|_{V'}\|\overline{\theta}_u^{n+1}\|+\|\partial_1\rho_u^{n+1}\|\|\overline{\theta}_u^{n+1}\|,\end{aligned}\tag{6.110}$$

où $\partial_1\left\|\overline{\theta}_u^{n+1}\right\|^2$ est une notation donnée par,

$$\partial_1\left\|\overline{\theta}_u^{n+1}\right\|^2=\partial_1\left(\overline{\theta}_u^{n+1},\overline{\theta}_u^{n+1}\right)=\left(\partial_1\overline{\theta}_u^{n+1},\overline{\theta}_u^{n+1}\right)+\left(\overline{\theta}_u^{n+1},\partial_1\overline{\theta}_u^{n+1}\right).$$

Dans (6.110), on utilise la relation (6.37) de la proposition 1 et les notations introduites dans (6.94)-(6.95) pour avoir,

$$\begin{aligned}\frac{1}{2}\partial_1\|\overline{\theta}_u^{n+1}\|^2+\left(\sigma\nabla\overline{\theta}_u^{n+1},\nabla\overline{\theta}_u^{n+1}\right)\leq&L_f\left(\|\overline{\theta}_v^n\|+\|\overline{\theta}_u^n\|+\|\rho_v^n\|+\|\rho_u^n\|\right)\|\overline{\theta}_u^{n+1}\|\\&+\|E_u^n\|_{V'}\|\overline{\theta}_u^{n+1}\|+\|\partial_1\rho_u^{n+1}\|\|\overline{\theta}_u^{n+1}\|,\end{aligned}\tag{6.111}$$

Or en utilisant la notation $\rho_{\partial_t u}=\mathcal{R}_h\partial_t u-\partial_t u$ et les hypothèses 2 et 1, on a,

$$\|\partial_1\rho_u^{n+1}\|=\frac{1}{\Delta t}\|\int_{t_n}^{t_{n+1}}\partial_t\rho_u d\tau\|\leq\frac{1}{\Delta t}\int_{t_n}^{t_{n+1}}\|\rho_{\partial_t u}\|d\tau\leq Ch^r.\tag{6.112}$$

En appliquant sur (6.111) la relation (6.112), le théorème de consistance 2, l'hypothèse 2 et l'inégalité de Young, on obtient,

$$\begin{aligned}\frac{1}{2}\partial_1\|\overline{\theta}_u^{n+1}\|^2+\left(\sigma\nabla\overline{\theta}_u^{n+1},\nabla\overline{\theta}_u^{n+1}\right)&\leq L_f\left(\|\theta_u^n\|+\|\theta_v^n\|+Ch^r+C\Delta t\right)\|\overline{\theta}_u^{n+1}\|\\&\leq\frac{L_f}{2}\left(\|\overline{\theta}_u^n\|^2+\|\overline{\theta}_v^n\|^2\right)+C\Delta t\|\overline{\theta}_u^{n+1}\|^2+Ch^{2r}+C\Delta t^2\end{aligned}\tag{6.113}$$

Par développement et un arrangement de termes, l'inégalité (6.113) devient,

$$(1-C\Delta t)\|\overline{\theta}_u^{n+1}\|^2\leq(1+\Delta tC)\|\overline{\theta}_u^n\|^2+\Delta tC\|\overline{\theta}_v^n\|^2+\Delta tC\left(h^{2r}+\Delta t^2\right).\tag{6.114}$$

Pour $0<\varepsilon<1-C\Delta t<1$, la combinaison de (6.104) et (6.114) permet d'avoir,

$$\|\overline{\theta}_u^{n+1}\|^2+\|\overline{\theta}_v^{n+1}\|^2\leq\frac{1+\Delta tC}{1-C\Delta t}\left(\|\overline{\theta}_u^n\|^2+\|\overline{\theta}_v^n\|^2\right)+\frac{\Delta tC}{\varepsilon}\left(h^{2r}+\Delta t^2\right)\tag{6.115}$$

On obtient par récurrence sur n,

$$\begin{aligned}\|\overline{\theta}_u^{n+1}\|^2+\|\overline{\theta}_v^{n+1}\|^2 &\leq \left(\frac{1+\Delta tC}{1-C\Delta t}\right)^{n+1}\left(\|\overline{\theta}_u^0\|^2+\|\overline{\theta}_v^0\|^2\right)\\ &\quad+\frac{\Delta tC}{\varepsilon}\left(h^{2r}+\Delta t^2\right)\left(\left(\frac{1+\Delta tC}{1-C\Delta t}\right)^{n}+\left(\frac{1+\Delta tC}{1-C\Delta t}\right)^{n-1}+\cdots+1\right)\\ &\leq \left(\frac{1+\Delta tC}{1-C\Delta t}\right)^{n+1}\left(\|\overline{\theta}_u^0\|^2+\|\overline{\theta}_v^0\|^2+\frac{Ct_{n+1}}{\varepsilon}\left(h^{2r}+\Delta t^2\right)\right) \quad (6.116)\end{aligned}$$

Remarquons pour conclure que pour $C\Delta t<1/2$,

$$\left(\frac{1+\Delta tC}{1-C\Delta t}\right)\leq(1+C\Delta t)^3 \quad (6.117)$$

En effet, l'inégalité (6.117) est équivalente à $0\leq C\Delta t\left(1-(C\Delta t+C^2\Delta t^2)\right)$. Puisque pour $C\Delta t<1/2$, on a $0\leq C\Delta t+C^2\Delta t^2<3/4<1$, l'inégalité (6.117) est alors vraie pour $C\Delta t<1/2$. Notons aussi que la condition $C\Delta t<1/2$ est équivalente à $1/2<1-C\Delta t<1$. On peut donc prendre $\varepsilon=1/2$ dans (6.116).
On prend $C_s=\frac{1}{2C}$ et $\varepsilon=1/2$. Alors, pour $\Delta t\leq C_s$, les inégalités (6.116) et (6.117) sont vérifiées. Si on utilise dans (6.116) les inégalités (6.117) et $(1+C\Delta t)^3\leq e^{3C\Delta t}$, on obtient,

$$\begin{aligned}\|\overline{\theta}_u^{n+1}\|^2+\|\overline{\theta}_v^{n+1}\|^2 &\leq e^{3Ct_{n+1}}\left(\|\overline{\theta}_u^0\|^2+\|\overline{\theta}_v^0\|^2+2Ct_{n+1}\left(h^{2r}+\Delta t^2\right)\right)\\ &\leq e^{3CT}\left(\|\overline{\theta}_u^0\|^2+\|\overline{\theta}_v^0\|^2+2CT\left(h^{2r}+\Delta t^2\right)\right) \quad (6.118)\end{aligned}$$

On a finalement,

$$\begin{aligned}\left\|\overline{v}^{n+1}-v^{n+1}\right\|+\left\|\overline{u}^{n+1}-u^{n+1}\right\| &\leq \left\|\overline{\theta}_u^{n+1}\right\|+\left\|\overline{\theta}_v^{n+1}\right\|+\left\|\rho_u^{n+1}\right\|+\left\|\rho_u^{n+1}\right\|\\ &\leq \left\|\overline{\theta}_u^{n+1}\right\|+\left\|\overline{\theta}_v^{n+1}\right\|+Ch^r\\ &\leq C\left(\left\|\overline{\theta}_u^0\right\|+\left\|\overline{\theta}_v^0\right\|+h^r+\Delta t\right)\text{ (en utilisant (6.118))}\end{aligned}$$

□

6.4.3 Convergence de $\mathbb{P}_r+RL2+SBDF2$

Dans cette section, on va s'intéresser aux fonctions : u,v solution du problème (6.8)-(6.9) et $(\overline{v}^n,\overline{u}^n)$ solution numérique approchant (v,u) par le schéma $\mathbb{P}_r+RL2+SBDF2$ aux instants t_n. La solution numérique $(\overline{v}^n,\overline{u}^n)$ est définie par,

$$\left(\overline{v}^{n+1},\psi_h\right)=(\overline{v}^n,\psi_h)+\Delta t\left(\varphi_1(\overline{\alpha}_n\Delta t)(\overline{\alpha}_n\overline{v}^n+\overline{\beta}_n),\psi_h\right)\ \forall\psi_h\in V_h, \quad (6.119)$$

$$\begin{aligned}&\left(\frac{\frac{3}{2}\overline{u}^{n+1}-2\overline{u}^n+\frac{1}{2}\overline{u}^{n-1}}{\Delta t},\varphi_h\right)+\left(\sigma\nabla\overline{u}^{n+1},\nabla\varphi_h\right) \quad (6.120)\\ &\qquad=\left(2f(t_n,\overline{v}^n,\overline{u}^n)-f(t_{n-1},\overline{v}^{n-1},\overline{u}^{n-1}),\varphi_h\right)\ \forall\varphi_h\in V_h,\end{aligned}$$

avec $\overline{\alpha}_n = \frac{3}{2}a(\overline{u}^n) - \frac{1}{2}a(\overline{u}^{n-1})$ et $\overline{\beta}_n = \frac{3}{2}b(\overline{v}^n, \overline{u}^n) - \frac{1}{2}b(\overline{v}^{n-1}, \overline{u}^{n-1})$

Définition 10 (Erreur de consistance de $RL2 + SBDF2$)**.** Soit (v, u) solution de (6.8)-(6.9). Les erreurs de consistances E_u^n et E_v^n en u et v respectivement du schéma $RL2 + SBDF2$ sont définis pour $\psi \in L^2(\Omega)$ et $\varphi \in H^1(\Omega)$ par,

$$\left(v^{n+1}, \psi\right) = (v^n, \psi) + \Delta t\,(\varphi_1(\alpha_n \Delta t)(\alpha_n v^n + \beta_n), \psi) + \Delta t\,(E_v^n, \psi)\,, \tag{6.121}$$

$$\left(\frac{\frac{3}{2}u^{n+1} - 2u^n + \frac{1}{2}u^{n-1}}{\Delta t}, \varphi\right) + \left(\sigma \nabla u^{n+1}, \nabla \varphi\right) \tag{6.122}$$
$$= \left(2f(t_n, v^n, u^n) - f(t_{n-1}, v^{n-1}, u^{n-1}), \varphi\right) + (E_u^n, \varphi)\,.$$

avec $\alpha_n = \frac{3}{2}a(u^n) - \frac{1}{2}a(u^{n-1})$ et $\beta_n = \frac{3}{2}b(v^n, u^n) - \frac{1}{2}b(v^{n-1}, u^{n-1})$, $u^n = u(t_n), v^n = v(t_n)$ et $t_n = n\Delta t$.

Théorème 4 (Consistance $RL2 + SBDF2$)**.** On suppose que les hypothèses 1 et 3 sont vérifiées. Les erreurs de consistances E_u^n et E_v^n pour le schéma $RL2 + SBDF2$ vérifient

$$\|E_u^n\|_{V'} + \|E_v^n\|_{V'} \leq C\Delta t^2. \tag{6.123}$$

Avec C une constante dépendant uniquement des données du problème (6.8)-(6.9) et de la solution (v, u).

Preuve. On pose $\partial_2 u^{n+1} = \dfrac{\frac{3}{2}u^{n+1} - 2u^n + \frac{1}{2}u^{n-1}}{\Delta t}$. Soit $\varphi \in H^1(\Omega)$. Par une soustraction entre les égalités (6.122) et (6.9) on a,

$$(E_u^n, \varphi) = \left(\partial_2 u^{n+1} - \partial_t u^{n+1}, \varphi\right) \tag{6.124}$$
$$+ \left(f(t_{n+1}, v^{n+1}, u^{n+1}) - (2f(t_n, v^n, u^n) - f(t_{n-1}, v^{n-1}, u^{n-1})), \varphi\right)$$

L'inégalité triangulaire et de Cauchy-Schwarz appliquée à (6.124) nous permet d'écrire,

$$|\,(E_u^n, \varphi)\,| \leq \left\|\partial_2 u^{n+1} - \partial_t u^{n+1}\right\| \|\varphi\| \tag{6.125}$$
$$+ \left\|f(t_{n+1}, v^{n+1}, u^{n+1}) - (2f(t_n, v^n, u^n) - f(t_{n-1}, v^{n-1}, u^{n-1}))\right\| \|\varphi\|\,.$$

Pour presque tout $x \in \Omega$, il existe par un développement de Taylor, τ_{1x}^n, τ_{2x}^n et τ_{3x}^n dans $]0, \Delta t[$ tels que,

$$\partial_2 u(t_{n+1}, x) - \partial_t u(t_{n+1}, x) \tag{6.126}$$
$$= \frac{\Delta t^2}{12}\left(3\partial_{ttt} u(t_n + \tau_{1x}^n, x) - \partial_{ttt} u(t_n - \tau_{2x}^n, x) + 6\partial_{ttt} u(t_n + \tau_{1x}^n, x)\right).$$

En effet, par la formule de Taylor il existe τ_{1x}^n et τ_{2x}^n dans $]0, \Delta t[$ tels que,

$$\begin{aligned}\partial_2 u(t_{n+1},x) &= \frac{3}{2\Delta t}\left(u(t_n,x)+\Delta t\partial_t u(t_n,x)+\frac{\Delta t^2}{2}\partial_{tt}u(t_n,x)+\frac{\Delta t^3}{6}\partial_{ttt}u(t_n+\tau_{1x}^n,x)\right)\\ &-\frac{2}{\Delta t}u(t_n,x)+\frac{1}{2\Delta t}\left(u(t_n,x)-\Delta t\partial_t u(t_n,x)+\frac{\Delta t^2}{2}\partial_{tt}u(t_n,x)-\frac{\Delta t^3}{6}\partial_{ttt}u(t_n-\tau_{2x}^n,x)\right),\\ &=\partial_t u(t_n,x)+\Delta t\partial_{tt}u(t_n,x)+\frac{\Delta t^2}{12}\left(3\partial_{ttt}u(t_n+\tau_{1x}^n,x)-\partial_{ttt}u(t_n-\tau_{2x}^n,x)\right). \quad (6.127)\end{aligned}$$

Par ailleurs, toujours par la formule de Taylor, il existe τ_{3x}^n dans $]0, \Delta t[$ tel que

$$\partial_t u(t_{n+1},x)=\partial_t u(t_n,x)+\Delta t\partial_{tt}u(t_n,x)+\frac{\Delta t^2}{2}\partial_{ttt}u(t_n+\tau_{3x}^n,x). \quad (6.128)$$

La soustraction de (6.128) à (6.127) donne (6.126).
De (6.126) on en déduit,

$$\left\|\partial_2 u^{n+1}-\partial_t u^{n+1}\right\|\leq C\Delta t^2, \quad (6.129)$$

où $C=\frac{5}{6}\left\|\partial_{ttt}u\right\|_{L^\infty(0,T,L^\infty(\Omega))}|\Omega|^{\frac{1}{2}}$.
Toujours par un développement de Taylor, pour presque tout $x\in\Omega$, il existe τ_{4x}^n et τ_{5x}^n dans $]0, \Delta t[$ tels que,

$$\begin{aligned}&f(t_{n+1},v(t_{n+1},x),u(t_{n+1},x))-\left(2f(t_n,v(t_n,x),u(t_n,x))-f(t_{n-1},v(t_{n-1},x),u(t_{n-1},x))\right),\\ &\quad=\frac{\Delta t^2}{2}\left(\partial_{tt}f(.,v,u)(t_n+\tau_{4x}^n,x)+\partial_{tt}f(.,v,u)(t_n-\tau_{5x}^n,x)\right). \quad (6.130)\end{aligned}$$

En effet, par le développement de Taylor, il existe $\tau_{4x}^n\in[0,\Delta t]$ tel que,

$$\begin{aligned}f(t_{n+1},v(t_{n+1},x),u(t_{n+1},x))&=f(t_n,v(t_n,x),u(t_n,x))+\Delta t\partial_t f(.,v,u)(t_n,x)\\ &+\frac{\Delta t^2}{2}\partial_{tt}f(.,v,u)(t_n+\tau_{4x}^n,x). \quad (6.131)\end{aligned}$$

Par ailleurs, il existe par le développement de Taylor $\tau_{5x}^n\in[0,\Delta t]$ tel que,

$$\begin{aligned}&2f(t_n,v(t_n,x),u(t_n,x))-f(t_{n-1},v(t_{n-1},x),u(t_{n-1},x))\\ &\qquad=2f(t_n,v(t_n,x),u(t_n,x)) \quad (6.132)\\ &-\left(f(t_n,v(t_n,x),u(t_n,x))-\Delta t\partial_t f(.,v,u)(t_n,x)+\frac{\Delta t^2}{2}\partial_{tt}f(.,v,u)(t_n-\tau_{5x}^n,x)\right).\end{aligned}$$

Une soustraction de (6.132) à (6.131) permet d'obtenir (6.130).
De (6.130) on déduit,

$$\left\|f(t_{n+1},v^{n+1},u^{n+1})-\left(2f(t_n,v^n,u^n)-f(t_{n-1},v^{n-1},u^{n-1})\right)\right\|\leq C\Delta t^2, \quad (6.133)$$

avec $C = \|\partial_{tt} f(.,v,u)\|_{L^\infty(0,T,L^\infty(\Omega))} |\Omega|$. En utilisant (6.129) et (6.133) dans (6.125) on obtient,

$$\frac{|\,(E_u^n,\varphi)\,|}{\|\varphi\|} \leq C\Delta t^2 \quad \text{et donc} \quad \|E_u^n\|_{V'} \leq C\Delta t^2. \tag{6.134}$$

Par ailleurs, de (6.121) on a pour $\varphi \in L^2(\Omega)$,

$$\Delta t\,(E_v^n,\varphi) = \left(v^{n+1} - v^n - \Delta t\varphi_1(\alpha_n\Delta t)(\alpha_n v^n + \beta_n), \varphi\right). \tag{6.135}$$

On pose,

$$z(\tau) = v^n + \tau\varphi_1(\alpha_n\tau)(\alpha_n v^n + \beta_n)\ presque\ partout\ \text{ dans }\ \Omega. \tag{6.136}$$

Alors z est la solution du problème,

$$\partial_t z = \alpha_n z + \beta_n \quad , \quad z(0) = v^n\ presque\ partout\ \text{ dans }\ \Omega. \tag{6.137}$$

Pour presque tout $x \in \Omega$, il existe par la formule de Taylor, $\rho_{ix}^n \in]0,\Delta t[\,;\ i = 1,\ldots,6$ tels que,

$$\begin{aligned} z(\Delta t,x) - v(t_{n+1},x) &= \frac{\Delta t^3}{6}\left(\partial_{ttt} z(\rho_{1x}^n,x) - \partial_{ttt} v(t_n+\rho_{6x}^n,x)\right) \\ &+ \frac{\Delta t^3}{2}\left(a(u)(t_n,x)v(t_n,x)\partial_t a(u)(t_n-\rho_{4x}^n,x) + \frac{1}{2}a(u)(t_n,x)\partial_t b(v,u)(t_n-\rho_{5x}^n,x)\right) \\ &+ \frac{\Delta t^3}{4}\left(-\partial_{tt} a(u)(t_n-\rho_{2x}^n,x)v(t_n,x) - \partial_{tt} b(v,u)(t_n-\rho_{3x}^n,x) + b(u)(t_n,x)\partial_t a(u)(t_n-\rho_{4x}^n,x)\right) \\ &+ \frac{\Delta t^3}{8}\left(\Delta t v(t_n,x)\partial_t a(u)(t_n-\rho_{4x}^n,x)^2 + \Delta t\partial_t a(u)(t_n-\rho_{4x}^n,x)\partial_t b(v,u)(t_n-\rho_{5x}^n,x)\right). \end{aligned} \tag{6.138}$$

Pour la preuve de (6.138) voir en annexe 8.
De (6.138) on déduit,

$$\left\|v^{n+1} - z(\Delta t)\right\| \leq C\Delta t^3 \tag{6.139}$$

où C est une constante qui dépend de la mesure de Ω et des sommes des normes $\|.\|_{L^\infty(0,T,L^\infty(\Omega))}$ de toutes les fonctions apparaissant dans (6.138). Plus précisément celles de $\partial_{ttt} z, \partial_{ttt} v, \partial_t a(u), \partial_t b(v,u), \partial_t a(u), \partial_{tt} a(u), \partial_{tt} b(v,u)$.
On utilise successivement (6.135), (6.136), l'inégalité de Cauchy-Schwarz et (6.139) pour avoir,

$$\begin{aligned} |\,(E_v^n,\varphi)\,| &= \frac{1}{\Delta t}|\left(v^{n+1} - v^n - \Delta t\varphi_1(\alpha_n\Delta t)(\alpha_n v^n + \beta_n), \varphi\right)| \\ &= \frac{1}{\Delta t}|\left(v^{n+1} - z(\Delta t), \varphi\right)| \\ &\leq \frac{1}{\Delta t}\left\|v^{n+1} - z(\Delta t)\right\| \|\varphi\| \\ &\leq C\Delta t^2 \|\varphi\|. \end{aligned}$$

On a par la suite ,

$$\frac{|\,(E_v^n,\varphi)\,|}{\|\varphi\|} \leq C\Delta t^2 \quad \text{et donc} \quad \|E_v^n\|_{V'} \leq C\Delta t^2. \tag{6.140}$$

Des inégalités (6.140) et (6.134) on obtient le résultat recherché. □

Proposition 3. On suppose que l'hypothèse 3 est vérifiée. Soient $t_n = n\Delta t$, $(x_1, y_1), (x_2, y_2) \in \mathcal{B}_v(R) \times \mathcal{B}_u(R)$. On pose $x_i^n = x_i(t_n, .)$, $y_i^n = y_i(t_n, .)$, $\alpha_n^i = \frac{3}{2}a(y_i^n) - \frac{1}{2}a(y_i^{n-1})$ et $\beta_n^i = \frac{3}{2}b(x_i^n, y_i^n) - \frac{1}{2}b(x_i^{n-1}, y_i^{n-1})$; $i = 1, 2$. Alors, il existe une constante $L_{ab\varphi_1}$ strictement positive ne dépendant que de R, des fonctions a, b, φ_1 et de la solution (v, u) telle que,

$$\begin{aligned}\|\varphi_1(\alpha_n^1\Delta t)(\alpha_n^1 x_1^n + \beta_n^1) - \varphi_1(\alpha_n^2\Delta t)(\alpha_n^2 x_2^n + \beta_n^2)\| \\ \leq L_{ab\varphi_1}\left(\|x_1^n - x_2^n\| + \|y_1^n - y_2^n\| + \left\|x_1^{n-1} - x_2^{n-1}\right\| + \left\|y_1^{n-1} - y_2^{n-1}\right\|\right).\end{aligned} \tag{6.141}$$

Preuve. La preuve se calque tout simplement sur celle de la proposition 2 tout en utilisant les mêmes arguments. □

Théorème 5 (Convergence de $\mathbb{P}_r + RL2 + SBDF2$)**.** On suppose que les hypothèses 1 à 3 sont vérifiées. On suppose aussi que la solution numérique $(\overline{v}^n, \overline{u}^n)$ définit par (6.119)-(6.120) existe, est unique et reste dans $\mathcal{B}_v(R) \times \mathcal{B}_u(R)$. Alors, il existe une constante C_s ne dépendant pas de h, une constante C strictement positive ne dépendant que de R, des données et de la solution (v, u) du problème (6.8)-(6.9), telle que pour tout $\Delta t < C_s$,

$$\begin{aligned}\left\|\overline{v}^{n+1} - v^{n+1}\right\| + \left\|\overline{u}^{n+1} - u^{n+1}\right\| \\ \leq C\left(\left\|\overline{v}^0 - P_h v^0\right\| + \left\|\overline{u}^0 - \mathcal{R}_h v^0\right\| + \left\|\overline{v}^1 - P_h v^1\right\| + \left\|\overline{u}^1 - \mathcal{R}_h v^1\right\| + h^r + \Delta t^2\right).\end{aligned} \tag{6.142}$$

Dans la preuve du théorème (5), nous aurons besoin du lemme de Grönwall discret que nous énonçons sans le démontrer. Pour plus de détails sur ce lemme, voir [4].

Lemme 1 (Grönwall discret)**.** *Soit k_n une suite de réels positifs et φ_n une suite telle que*

$$\begin{cases}\varphi_0 \leq g_0, \\ \varphi_m \leq g_0 + \sum_{s=0}^{m-1} p_s + \sum_{s=0}^{m-1} k_s\varphi_s, \qquad n \geq 0.\end{cases}$$

Si $g_0 \geq 0$ et $p_n \geq 0$ pour tout $n \geq 0$, alors

$$\varphi_n \leq \left(g_0 + \sum_{s=0}^{m-1} p_s\right) exp\left(\sum_{s=0}^{m-1} k_s\right).$$

Preuve du théorème. On garde les notations introduites en (6.94) et (6.95). Par l'inégalité triangulaire on a,

$$\left\|\overline{v}^{n+1}-v^{n+1}\right\| \leq \left\|\overline{v}^{n+1}-P_h v^{n+1}\right\| + \left\|P_h v^{n+1}-v^{n+1}\right\| = \left\|\overline{\theta}_v^{n+1}\right\| + \left\|\rho_v^{n+1}\right\| \tag{6.143}$$

$$\left\|\overline{u}^{n+1}-u^{n+1}\right\| \leq \left\|\overline{u}^{n+1}-P_h u^{n+1}\right\| + \left\|P_h u^{n+1}-u^{n+1}\right\| = \left\|\overline{\theta}_u^{n+1}\right\| + \left\|\rho_u^{n+1}\right\|. \tag{6.144}$$

Soit $\psi_h \in V_h$. Une soustraction entre les inégalités (6.119) et (6.121) nous donne,

$$\begin{aligned}
\left(\overline{\theta}_v^{n+1},\psi_h\right) &= \left(\overline{v}^{n+1}-P_h v^{n+1},\psi_h\right)\\
&= \left(\overline{v}^{n+1}-v^{n+1},\psi_h\right)\\
&= \left(\overline{v}^{n}-v^{n},\psi_h\right) + \Delta t\left(E_v^n,\psi_h\right)\\
&+ \Delta t\left((\varphi_1(\overline{\alpha}_n\Delta t)(\overline{\alpha}_n\overline{v}^n+\overline{\beta}_n)-\varphi_1(\alpha_n\Delta t)(\alpha_n v^n+\beta_n),\psi_h\right)\\
&= \left(\overline{v}^{n}-P_h v^{n},\psi_h\right) + \Delta t\left(E_v^n,\psi_h\right)\\
&+ \Delta t\left((\varphi_1(\overline{\alpha}_n\Delta t)(\overline{\alpha}_n\overline{v}^n+\overline{\beta}_n)-\varphi_1(\alpha_n\Delta t)(\alpha_n v^n+\beta_n),\psi_h\right).
\end{aligned} \tag{6.145}$$

Dans (6.145), on prend $\psi_h = \overline{\theta}_v^{n+1}$ et on utilise l'inégalité de Cauchy-Schwarz pour obtenir,

$$\begin{aligned}
\left\|\overline{\theta}_v^{n+1}\right\|^2 \leq \left\|\overline{\theta}_v^{n}\right\|\left\|\overline{\theta}_v^{n+1}\right\| + \Delta t\left\|E_v^n\right\|_{V'}\left\|\overline{\theta}_v^{n+1}\right\|\\
\Delta t\left\|\varphi_1(\overline{\alpha}_n\Delta t)(\overline{\alpha}_n\overline{v}^n+\overline{\beta}_n)-\varphi_1(\alpha_n\Delta t)(\alpha_n v^n+\beta_n)\right\|\left\|\overline{\theta}_v^{n+1}\right\|.
\end{aligned} \tag{6.146}$$

On applique à l'inégalité (6.146), la relation (6.141) de la proposition 3 pour avoir,

$$\begin{aligned}
\left\|\overline{\theta}_v^{n+1}\right\|^2 \leq \left\|\overline{\theta}_v^{n}\right\|\left\|\overline{\theta}_v^{n+1}\right\| + \Delta t\left\|E_v^n\right\|_{V'}\left\|\overline{\theta}_v^{n+1}\right\|\\
+ \Delta t L_{ab\varphi_1}\left(\left\|\overline{v}^n-v^n\right\| + \left\|\overline{u}^n-u^n\right\| + \left\|\overline{v}^{n-1}-v^{n-1}\right\| + \left\|\overline{u}^{n-1}-u^{n-1}\right\|\right)\left\|\overline{\theta}_v^{n+1}\right\|.
\end{aligned} \tag{6.147}$$

On introduit dans (6.147), les termes $P_h v^n$ et $\mathcal{R}_h u^n$, puis on utilise l'inégalité triangulaire et les notations introduites en (6.94) et (6.95) pour obtenir,

$$\begin{aligned}
\left\|\overline{\theta}_v^{n+1}\right\|^2 \leq \left\|\overline{\theta}_v^{n}\right\|\left\|\overline{\theta}_v^{n+1}\right\| + \Delta t\left\|E_v^n\right\|_{V'}\left\|\overline{\theta}_v^{n+1}\right\|\\
+ \Delta t L_{ab\varphi_1}\left(\left\|\overline{\theta}_v^{n}\right\| + \left\|\overline{\theta}_u^{n}\right\| + \left\|\overline{\theta}_v^{n-1}\right\| + \left\|\overline{\theta}_u^{n-1}\right\| + \left\|\rho_v^{n}\right\| + \left\|\rho_u^{n}\right\| + \left\|\rho_v^{n-1}\right\| + \left\|\rho_u^{n-1}\right\|\right)\left\|\overline{\theta}_v^{n+1}\right\|.
\end{aligned} \tag{6.148}$$

L'inégalité de Young appliquée à (6.148) permet d'obtenir,

$$\begin{aligned}
\left\|\overline{\theta}_v^{n+1}\right\|^2 \leq \frac{1}{2}\left\|\overline{\theta}_u^{n}\right\|^2 + \frac{\Delta t L_{ab\varphi}}{2}\left(\left\|\overline{\theta}_v^{n}\right\|^2 + \left\|\overline{\theta}_u^{n}\right\|^2 + \left\|\overline{\theta}_v^{n-1}\right\|^2 + \left\|\overline{\theta}_u^{n-1}\right\|^2\right)\\
+ \frac{\Delta t L_{ab\varphi}}{2}\left(\left\|\rho_v^{n}\right\|^2 + \left\|\rho_u^{n}\right\|^2 + \left\|\rho_v^{n-1}\right\|^2 + \left\|\rho_u^{n-1}\right\|^2\right)\\
+ \frac{\Delta t}{2}\left\|E_v^n\right\|_{V'}^2 + \frac{1}{2}\left\|\overline{\theta}_v^{n+1}\right\|^2 + \Delta t\left(4L_{ab\varphi}+\frac{1}{2}\right)\left\|\overline{\theta}_v^{n+1}\right\|^2.
\end{aligned} \tag{6.149}$$

Après avoir utilisé l'hypothèse 2 et l'inégalité (6.123) théorème 4 dans (6.149), un arrangement des termes nous permet d'avoir,

$$(1-\Delta tC)\left\|\overline{\theta}_v^{n+1}\right\|^2-\left\|\overline{\theta}_v^{n}\right\|^2\leq C\Delta t\left(\left\|\overline{\theta}_v^{n}\right\|^2+\left\|\overline{\theta}_u^{n}\right\|^2+\left\|\overline{\theta}_v^{n-1}\right\|^2+\left\|\overline{\theta}_u^{n-1}\right\|^2\right)+\Delta tC\left(h^{2r}+\Delta t^4\right). \tag{6.150}$$

On pose $\partial_2\overline{u}^{n+1}=\dfrac{\frac{3}{2}\overline{u}^{n+1}-2\overline{u}^{n}+\frac{1}{2}\overline{u}^{n-1}}{\Delta t}$ et on s'intéresse à présent à $\overline{\theta}_u^{n+1}$.
En utilisant la définition de la projection $\mathcal{R}_h$ introduite en (6.29) on a,

$$\begin{aligned}&\left(\partial_2\overline{\theta}_u^{n+1},\varphi_h\right)+\left(\sigma\nabla\overline{\theta}_u^{n+1},\nabla\varphi_h\right)\\&\quad=\left(\partial_2\overline{u}^{n+1},\varphi_h\right)+\left(\sigma\nabla\overline{u}^{n+1},\nabla\varphi_h\right)-\left(\partial_1\mathcal{R}_hu^{n+1},\varphi_h\right)-\left(\sigma\nabla\mathcal{R}_hu^{n+1},\nabla\varphi_h\right)\\&\quad=\left(\partial_2\overline{u}^{n+1},\varphi_h\right)+\left(\sigma\nabla\overline{u}^{n+1},\nabla\varphi_h\right)-\left(\partial_1\mathcal{R}_hu^{n+1},\varphi_h\right)-\left(\sigma\nabla u^{n+1},\nabla\varphi_h\right).\end{aligned} \tag{6.151}$$

On utilise (6.120) et (6.9) dans (6.151) pour avoir,

$$\begin{aligned}\left(\partial_2\overline{\theta}_u^{n+1},\varphi_h\right)+\left(\sigma\nabla\overline{\theta}_u^{n+1},\nabla\varphi_h\right)&=-\left(\partial_2\mathcal{R}_hu^{n+1}-\partial_tu^{n+1},\varphi_h\right)\\&+\left(2f(t_n,\overline{v}^n,\overline{u}^n)-f(t_{n-1},\overline{v}^{n-1},\overline{u}^{n-1})-f(t_{n+1},v^{n+1},u^{n+1}),\varphi_h\right).\end{aligned} \tag{6.152}$$

Dans le membre de droite de (6.152), on ajoute et on retranche les termes $f(t_n,P_hv^n,\mathcal{R}_hu^n)$, $f(t_n,v^n,u^n)$ et $f(t_{n-1},v^{n-1},u^{n-1})$ pour obtenir,

$$\begin{aligned}&\left(\partial_2\overline{\theta}_u^{n+1},\varphi_h\right)+\left(\sigma\nabla\overline{\theta}_u^{n+1},\nabla\varphi_h\right)\\&\quad=2\left(f(t_n,\overline{v}^n,\overline{u}^n)-f(t_n,P_hv^n,\mathcal{R}_hu^n),\varphi_h\right)+2\left(f(t_n,P_hv^n,\mathcal{R}_hu^n)-f(t_n,v^n,u^n),\varphi_h\right)\\&\qquad-\left(f(t_{n-1},\overline{v}^{n-1},\overline{u}^{n-1})-f(t_{n-1},v^{n-1},u^{n-1}),\varphi_h\right)-\left(\partial_2\rho_u^{n+1},\varphi_h\right)\\&\quad-\left(\partial_2u^{n+1}-\partial_tu^{n+1}-2f(t_n,v^n,u^n)+f(t_{n-1},v^{n-1},u^{n-1})+f(t_{n+1},v^{n+1},u^{n+1}),\varphi_h\right).\end{aligned} \tag{6.153}$$

On utilise l'égalité (6.124) dans (6.153) pour avoir,

$$\begin{aligned}&\left(\partial_2\overline{\theta}_u^{n+1},\varphi_h\right)+\left(\sigma\nabla\overline{\theta}_u^{n+1},\nabla\varphi_h\right)\\&\quad=2\left(f(t_n,\overline{v}^n,\overline{u}^n)-f(t_n,P_hv^n,\mathcal{R}_hu^n),\varphi_h\right)+2\left(f(t_n,P_hv^n,\mathcal{R}_hu^n)-f(t_n,v^n,u^n),\varphi_h\right)\\&\qquad-\left(f(t_{n-1},\overline{v}^{n-1},\overline{u}^{n-1})-f(t_{n-1},v^{n-1},u^{n-1}),\varphi_h\right)-\left(E_u^n,\varphi_h\right)-\left(\partial_2\rho_u^{n+1},\varphi_h\right).\end{aligned} \tag{6.154}$$

Dans (6.154), on prend $\varphi_h=\overline{\theta}_u^{n+1}$ puis on utilise l'inégalité de Cauchy-Schwarz pour avoir,

$$\begin{aligned}&\frac{1}{2}\partial_2\|\overline{\theta}_u^{n+1}\|^2+\left(\sigma\nabla\overline{\theta}_u^{n+1},\nabla\overline{\theta}_u^{n+1}\right)\\&\quad\leq2\left(\|f(t_n,\overline{v}^n,\overline{u}^n)-f(t_n,P_hv^n,\mathcal{R}_hu^n)\|+\|f(t_n,P_hv^n,\mathcal{R}_hu^n)-f(t_n,v^n,u^n)\|\right)\|\overline{\theta}_u^{n+1}\|\\&\qquad+\|f(t_{n-1},\overline{v}^{n-1},\overline{u}^{n-1})-f(t_{n-1},v^{n-1},u^{n-1})\|\|\overline{\theta}_u^{n+1}\|+\left(\|E_u^n\|_{V'}+\|\partial_2\rho_u^{n+1}\|\right)\|\overline{\theta}_u^{n+1}\|.\end{aligned} \tag{6.155}$$

Avec $\partial_2 \left\|\overline{\theta}_u^{n+1}\right\|^2$ donné par,

$$\partial_2 \left\|\overline{\theta}_u^{n+1}\right\|^2 = \partial_2 \left(\overline{\theta}_u^{n+1}, \overline{\theta}_u^{n+1}\right) = \left(\partial_2\overline{\theta}_u^{n+1}, \overline{\theta}_u^{n+1}\right) + \left(\overline{\theta}_u^{n+1}, \partial_2\overline{\theta}_u^{n+1}\right).$$

Dans l'inégalité (6.154), on applique (6.37) et l'inégalité triangulaire pour avoir,

$$\begin{aligned}
&\frac{1}{2}\partial_2\|\overline{\theta}_u^{n+1}\|^2 + \left(\sigma\nabla\overline{\theta}_u^{n+1}, \nabla\overline{\theta}_u^{n+1}\right) \\
&\quad\leq L_f\left(2(\|\overline{\theta}_v^n\| + \|\overline{\theta}_u^n\| + \|\rho_v^n\| + \|\rho_u^n\|) + \|\overline{v}^{n-1} - v^{n-1}\| + \|\overline{v}^{n-1} - v^{n-1}\|\right)\|\overline{\theta}_u^{n+1}\| \\
&\qquad\qquad + \left(\|\partial_2\rho_u^{n+1}\| + \|E_u^n\|_{V'}\right)\|\overline{\theta}_u^{n+1}\| \\
&\quad\leq L_f\left(2(\|\overline{\theta}_v^n\| + \|\overline{\theta}_u^n\| + \|\rho_v^n\| + \|\rho_u^n\|) + \|\overline{\theta}_v^{n-1}\| + \|\overline{\theta}_u^{n-1}\| + \|\rho_v^{n-1}\| + \|\rho_u^{n-1}\|\right)\|\overline{\theta}_u^{n+1}\| \\
&\qquad\qquad + \left(\|\partial_2\rho_u^{n+1}\| + \|E_u^n\|_{V'}\right)\|\overline{\theta}_u^{n+1}\|. \qquad (6.156)
\end{aligned}$$

On remarque que,

$$\partial_2\rho_u^{n+1} = \frac{1}{\Delta t}\left(\frac{3}{2}\rho_u^{n+1} - 2\rho_u^n + \frac{1}{2}\rho_u^{n-1}\right) = \frac{3}{2}\frac{\rho_u^{n+1} - \rho_u^n}{\Delta t} - \frac{1}{2}\frac{\rho_u^n - \rho_u^{n-1}}{\Delta t} \qquad (6.157)$$

En utilisant la notation de l'opérateur discret ∂_1 défini dans (6.105) et l'inégalité (6.112), (6.157) permet d'écrire,

$$\|\partial_2\rho_u^{n+1}\| = \|\frac{3}{2}\partial_1\rho_u^{n+1} - \frac{1}{2}\partial_1\rho_u^n\| \leq \frac{3}{2}\|\partial_1\rho_u^{n+1}\| + \frac{1}{2}\|\partial_1\rho_u^n\| \leq Ch^r. \qquad (6.158)$$

En utilisant dans (6.156), les inégalités (6.158) et la relation (6.123) du théorème de consistance 4, on a,

$$\begin{aligned}
&\frac{1}{2}\partial_2\|\overline{\theta}_u^{n+1}\|^2 + \left(\sigma\nabla\overline{\theta}_u^{n+1}, \nabla\overline{\theta}_u^{n+1}\right) \\
&\quad\leq L_f\left(2(\|\overline{\theta}_v^n\| + \|\overline{\theta}_u^n\|) + \|\overline{\theta}_v^{n-1}\| + \|\overline{\theta}_u^{n-1}\|\right)\|\overline{\theta}_u^{n+1}\| + C\left(\Delta t^2 + h^r\right)\|\overline{\theta}_u^{n+1}\| \qquad (6.159)
\end{aligned}$$

En appliquant l'inégalité de Young dans la relation (6.159), on obtient,

$$\begin{aligned}
\frac{1}{2}\partial_2\|\overline{\theta}_u^{n+1}\|^2 - C\|\overline{\theta}_u^{n+1}\|^2 \leq\ & C\left(\|\overline{\theta}_v^n\|^2 + \|\overline{\theta}_u^n\|^2 + \|\overline{\theta}_v^{n-1}\|^2 + \|\overline{\theta}_u^{n-1}\|^2\right) \\
& + C\left(\Delta t^4 + h^{2r}\right). \qquad (6.160)
\end{aligned}$$

L'écriture de (6.160) sans l'opérateur de dérivée discrète ∂_2 nous donne,

$$\begin{aligned}
\frac{1}{\Delta t}\left(\frac{3}{2}\overline{\theta}_u^{n+1} - 2\overline{\theta}_u^n + \frac{1}{2}\overline{\theta}_u^{n-1}, \overline{\theta}_u^{n+1}\right) - C\|\overline{\theta}_u^{n+1}\|^2 \leq\ & C\left(\|\overline{\theta}_v^n\|^2 + \|\overline{\theta}_u^n\|^2 + \|\overline{\theta}_v^{n-1}\|^2 + \|\overline{\theta}_u^{n-1}\|^2\right) \\
& + C\left(\Delta t^4 + h^{2r}\right). \qquad (6.161)
\end{aligned}$$

On multiplie (6.161) par $4\Delta t$ pour obtenir,

$$\left(6\overline{\theta}_u^{n+1}-8\overline{\theta}_u^{n}+2\overline{\theta}_u^{n-1},\overline{\theta}_u^{n+1}\right)-4C\Delta t\|\overline{\theta}_u^{n+1}\|^2\leq 4C\Delta t\left(\|\overline{\theta}_v^{n}\|^2+\|\overline{\theta}_u^{n}\|^2+\|\overline{\theta}_v^{n-1}\|^2+\|\overline{\theta}_u^{n-1}\|^2\right)+4C\Delta t\left(\Delta t^4+h^{2r}\right). \tag{6.162}$$

Pour continuer, nous nous inspirons d'une technique de récurrence utilisée dans [8]. Ainsi, on utilise dans (6.162) l'identité l'identité,

$$\left(6x^{n+1}-8x^n+2x^{n-1}\right)x^{n+1}=(x^{n+1})^2+(2x^{n+1}-x^n)^2-(x^n)^2-(2x^n-x^{n-1})^2+(x^{n+1}-2x^n+x^{n-1})^2,$$

pour avoir,

$$\begin{aligned}(1-C\Delta t)\left\|\overline{\theta}^{n+1}\right\|^2-\left\|\overline{\theta}^{n}\right\|^2+\left\|2\overline{\theta}^{n+1}-\overline{\theta}^{n}\right\|^2-\left\|2\overline{\theta}^{n}-\overline{\theta}^{n-1}\right\|^2 \\ \leq C\Delta t\left(\|\overline{\theta}_v^{n}\|^2+\|\overline{\theta}_u^{n}\|^2+\|\overline{\theta}_v^{n-1}\|^2+\|\overline{\theta}_u^{n-1}\|^2\right)+C\Delta t\left(\Delta t^4+h^{2r}\right)\end{aligned} \tag{6.163}$$

Par une combinaison de (6.163) et (6.150), on obtient sur $\left\|\overline{\theta}_v^{n+1}\right\|+\left\|\overline{\theta}_u^{n+1}\right\|$, l'inégalité suivante

$$\begin{aligned}(1-C\Delta t)\left(\left\|\overline{\theta}_v^{n+1}\right\|^2+\left\|\overline{\theta}_u^{n+1}\right\|^2\right)-\left(\left\|\overline{\theta}_v^{n}\right\|^2+\left\|\overline{\theta}_u^{n}\right\|^2\right)+\left\|2\overline{\theta}_u^{n+1}-\overline{\theta}_u^{n}\right\|^2-\left\|2\overline{\theta}_u^{n}-\overline{\theta}_u^{n-1}\right\|^2 \\ \leq C\Delta t\left(\|\overline{\theta}_v^{n}\|^2+\|\overline{\theta}_u^{n}\|^2+\|\overline{\theta}_v^{n-1}\|^2+\|\overline{\theta}_u^{n-1}\|^2\right)+C\Delta t\left(\Delta t^4+h^{2r}\right)\end{aligned} \tag{6.164}$$

On fait une somme sur l'inégalité (6.164) en faisant varier n entre 1 et m. Ceci nous donne,

$$\begin{aligned}(1-C\Delta t)\left(\left\|\overline{\theta}_v^{m+1}\right\|^2+\left\|\overline{\theta}_u^{m+1}\right\|^2\right)+\left\|2\overline{\theta}_u^{m+1}-\overline{\theta}_u^{m}\right\|^2-C\Delta t\sum_{n=1}^{m-1}\left(\left\|\overline{\theta}_v^{n}\right\|^2+\left\|\overline{\theta}_u^{n}\right\|^2\right) \\ \leq\left\|2\overline{\theta}_u^{1}-\overline{\theta}_u^{0}\right\|^2+\left\|\overline{\theta}_u^{1}\right\|^2+\left\|\overline{\theta}_u^{1}\right\|^2+Ct_m\left(\Delta t^4+h^{2r}\right) \\ +C\Delta t\sum_{n=1}^{m}\left(\left\|\overline{\theta}_v^{n}\right\|^2+\left\|\overline{\theta}_u^{n}\right\|^2+\left\|\overline{\theta}_v^{n-1}\right\|^2+\left\|\overline{\theta}_u^{n-1}\right\|^2\right)\end{aligned} \tag{6.165}$$

Par des majorations simples, et en choisissant convenablement la constante C et $\Delta t<\frac{1}{2C}$, (6.165) nous permet d'écrire,

$$\begin{aligned}\left(\left\|\overline{\theta}_v^{m+1}\right\|^2+\left\|\overline{\theta}_u^{m+1}\right\|^2\right)\leq\frac{7}{1-C\Delta t}\left(\left\|\overline{\theta}_v^{0}\right\|^2+\left\|\overline{\theta}_u^{0}\right\|^2+\left\|\overline{\theta}_v^{1}\right\|^2+\left\|\overline{\theta}_u^{1}\right\|^2\right) \\ +\frac{CT}{1-C\Delta t}\left(\Delta t^4+h^{2r}\right)+\sum_{n=1}^{m-1}\frac{C\Delta t}{1-C\Delta t}\left(\left\|\overline{\theta}_v^{n}\right\|^2+\left\|\overline{\theta}_u^{n}\right\|^2\right).\end{aligned} \tag{6.166}$$

On remarque pour la suite que, $C\Delta t < 1/2$ est équivalent à $1/2 < 1 - C\Delta t < 1$.
On prend $C_s = \frac{1}{2C}$. Pour $\Delta t < C_s$, on a $\frac{1}{1-C\Delta t} < 2$ et l'inégalité (6.166) nous permet d'écrire,

$$\left(\left\|\overline{\theta}_v^{m+1}\right\|^2 + \left\|\overline{\theta}_u^{m+1}\right\|^2\right) \leq 14\left(\left\|\overline{\theta}_v^{0}\right\|^2 + \left\|\overline{\theta}_u^{0}\right\|^2 + \left\|\overline{\theta}_v^{1}\right\|^2 + \left\|\overline{\theta}_u^{1}\right\|^2\right) + 2CT\left(\Delta t^4 + h^{2r}\right) + \sum_{n=1}^{m-1} 2CT\left(\left\|\overline{\theta}_v^{n}\right\|^2 + \left\|\overline{\theta}_u^{n}\right\|^2\right). \tag{6.167}$$

On utilise le Lemme de Grönwall discret 1 dans (6.167) pour avoir,

$$\left\|\overline{\theta}_v^{m+1}\right\|^2 + \left\|\overline{\theta}_u^{m+1}\right\|^2 \leq \left(14\left(\left\|\overline{\theta}_v^{0}\right\|^2 + \left\|\overline{\theta}_u^{0}\right\|^2 + \left\|\overline{\theta}_v^{1}\right\|^2 + \left\|\overline{\theta}_u^{1}\right\|^2\right) + 2CT\left(\Delta t^4 + h^{2r}\right)\right) e^{2CT} \tag{6.168}$$

On somme (6.143) et (6.144) pour obtenir,

$$\left\|\overline{v}^{n+1} - v^{n+1}\right\| + \left\|\overline{u}^{n+1} - u^{n+1}\right\| \leq \left\|\overline{\theta}_u^{n+1}\right\| + \left\|\overline{\theta}_v^{n+1}\right\| + \left\|\rho_u^{n+1}\right\| + \left\|\rho_u^{n+1}\right\|. \tag{6.169}$$

On utilise l'hypothèse 2 dans (6.169) pour avoir,

$$\left\|\overline{v}^{n+1} - v^{n+1}\right\| + \left\|\overline{u}^{n+1} - u^{n+1}\right\| \leq \left\|\overline{\theta}_u^{n+1}\right\| + \left\|\overline{\theta}_v^{n+1}\right\| + Ch^r. \tag{6.170}$$

On utilise (6.168) dans (6.170) pour obtenir,

$$\left\|\overline{v}^{n+1} - v^{n+1}\right\| + \left\|\overline{u}^{n+1} - u^{n+1}\right\| \leq C\left(\left\|\overline{\theta}_u^{0}\right\| + \left\|\overline{\theta}_v^{0}\right\| + \left\|\overline{\theta}_u^{1}\right\| + \left\|\overline{\theta}_v^{1}\right\| + h^r + \Delta t^2\right).$$

□

Bibliographie

[1] Georgios AKRIVIS, Michel CROUZEIX et Charalambos MAKRIDAKIS. „Implicit-explicit multistep finite element methods for nonlinear parabolic problems“. In : *Mathematics of Computation of the American Mathematical Society* 67.222 (1998), p. 457–477 (cf. p. 111).

[2] Georgios AKRIVIS, Michel CROUZEIX et Charalambos MAKRIDAKIS. „Implicit-explicit multistep methods for quasilinear parabolic equations“. In : *Numerische Mathematik* 82.4 (1999), p. 521–541 (cf. p. 111).

[3] Grégoire ALLAIRE. *Analyse numérique et optimisation : une introduction à la modélisation mathématique et à la simulation numérique*. Editions Ecole Polytechnique, 2005 (cf. p. 105, 106).

[4] JH BRANDTS. *A. Quarteroni, R. Sacco and F. Saleri. Numerical mathematics (Texts in applied mathematics ; 37). New York : Springer-Verlag, 2000 654 p., prijs $59.95 ISBN 0-387-98959-5l*. 2002 (cf. p. 116, 128).

[5] Susanne BRENNER et Ridgway SCOTT. *The mathematical theory of finite element methods*. T. 15. Springer Science & Business Media, 2007 (cf. p. 109, 112).

[6] Haïm BREZIS. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010 (cf. p. 104).

[7] Philippe G CIARLET. *The finite element method for elliptic problems*. SIAM, 2002 (cf. p. 109, 112).

[8] Marc ETHIER et Yves BOURGAULT. „Semi-implicit time-discretization schemes for the bidomain model“. In : *SIAM Journal on Numerical Analysis* 46.5 (2008), p. 2443–2468 (cf. p. 132).

[9] Lions JACQUES-LOUIS. *Quelques méthodes de résolution des problemes aux limites non linéaires*. T. 31. Dunod Paris, 1969 (cf. p. 104).

[10] Vidar THOMÉE. *Galerkin finite element methods for parabolic problems*. T. 1054. Springer, 1984 (cf. p. 109, 112).

7 Résultats Numériques.

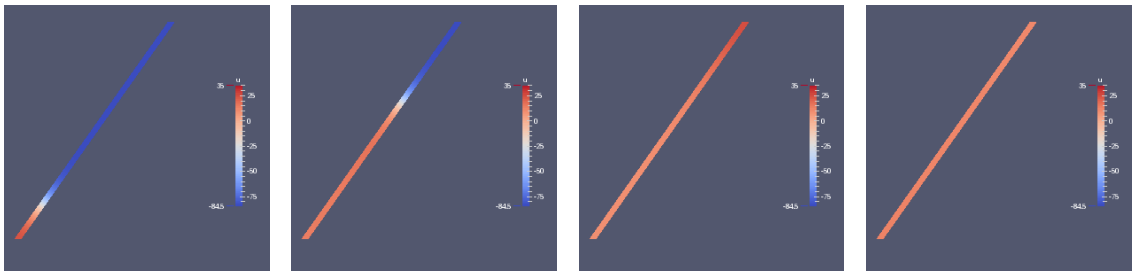
Contents

Toutes les simulations sont faites sur le modèle monodomaine, où le modèle ionique est le modèle de Beeler et Reuter [1]. Le problème mathématique qui caractérise le modèle monodomaine et que nous résolvons numériquement est celui qui a été décrit à la section 6.2 du Chapitre 6. Le problème discret que nous utilisons est celui qui a été donné par (6.27)-(6.28) du Chapitre 6. Les valeurs des constantes qui apparaissent dans le modèle monodomaine sont données par la table 7.1. Notre étude est particulièrement centrée sur le

TABLE 7.1: Paramètres utilisé dans le modèle monodomaine

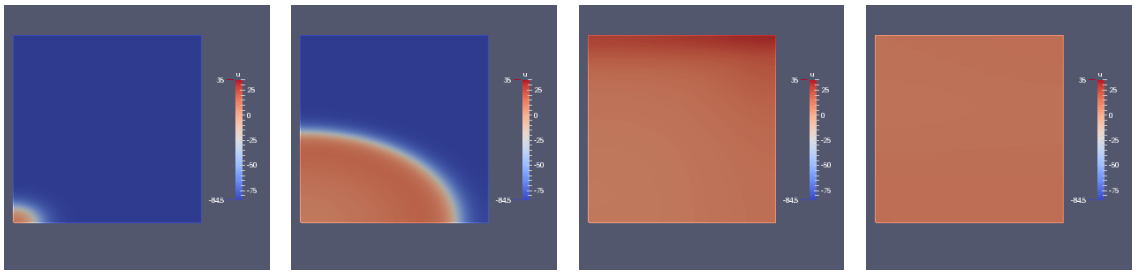
Paramètres du modèle	valeurs	unités
Conductivité C_m de la membrane	1	$[\mu F/cm^2]$
Aire de la membrane χ par unité de volume	2000	$[1/cm]$
Conductivité longitudinale σ_i^l du milieu intracellulaire	1.741	$[ms/cm]$
Conductivité transverse σ_i^t du milieu intracellulaire	0.1934	$[ms/cm]$

potentiel transmembranaire. L'unité de longueur est le centimètre. Sauf en cas d'indication contraire, le domaine Ω considéré est le segment $[0,1]$ pour les simulations en 1D, le carré $[0,1]\times[0,1]$ pour les simulations en 2D et le cube $[0,1]\times[0,1]\times[0,1]$ pour les simulations en 3D. Le temps final T est fixé à $30ms$, représentant un instant auquel le domaine considéré est totalement dépolarisé (voir figures 7.1 à 7.3). Cette phase correspondant au moment où la solution varie le plus rapidement est aussi la phase où l'accumulation des erreurs est plus forte. À chaque couple $(\Delta t, h)$ décrivant une discrétisation en temps et en espace (temps-



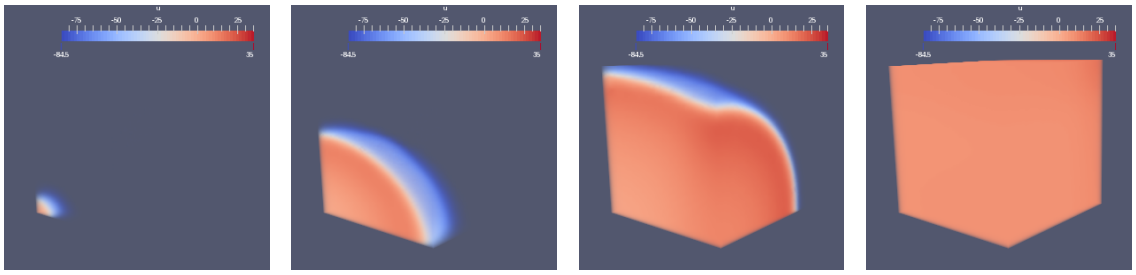
(a) t = 3ms **(b)** t = 7ms **(c)** t = 12ms **(d)** t=30ms

FIGURE 7.1: Propagation en 1D du potentiel transmembranaire pour le modèle monodomaine couplé au modèle ionique de Beeler Reuter.



(a) t = 3ms **(b)** t = 7ms **(c)** t = 12ms **(d)** t=30ms

FIGURE 7.2: Propagation en 2D du potentiel transmembranaire pour le modèle monodomaine couplé au modèle ionique de Beeler Reuter.



(a) t = 3ms **(b)** t = 12ms **(c)** t = 20ms **(d)** t=30ms

FIGURE 7.3: Propagation en 3D du potentiel transmembranaire pour le modèle monodomaine couplé au modèle ionique de Beeler Reuter.

espace), on associe une solution numérique $U_{\Delta t,h}$ calculée avec un schéma numérique d'intégration en temps et en espace. Les méthodes de discrétisations en espace seront les éléments finis de Lagrange $\mathbb{P}_1$ et $\mathbb{P}_2$. La solution exacte du problème couplé considéré n'étant pas accessible, nous allons la remplacer par une solution numérique U_{ref} dite de référence. Elle est calculée avec la combinaison de schémas numériques d'intégration espace-temps $\mathbb{P}_2 + RL_4 + SBDF4$ et une discrétisation en temps-espace $(\Delta t_{ref}, h_{ref})$. La solution de référence U_{ref} est donc la solution numérique $U_{\Delta t_{ref},h_{ref}}$, où Δt_{ref} et h_{ref} sont assez petits de telle sorte que l'erreur entre U_{ref} et la solution exacte du problème soit négligeable par rapport à l'erreur entre la solution numérique calculée et U_{ref}. On note par ndl_{ref} le nombre des degrés de libertés utilisés sur le maillage de référence pour le calcul de la solution de référence. De la même façon pour un maillage de diamètre h, on note ndl_h le nombre des degrés de libertés utilisés pour calculer une solution numérique $U_{\Delta t,h}$. Soient $(\phi_i)_{1\leq i\leq nd_{ref}}$ et $(\psi_i)_{1\leq i\leq nd_h}$ les bases fondamentales des espaces $V_{h_{ref}}$ et V_h décrites à la remarque 2 du chapitre 6. On note pour un instant $t_n = n\Delta t$, $U^n_{ref_i}$ et $U^n_{\Delta t,h_i}$ respectivement les degrés de liberté de U_{ref} et $U_{\Delta t,h}$ associés aux noeuds des degrés de liberté P_i et S_j à l'instant t_n. Pour tout $x \in \Omega$, on a

$$U_{ref}(t_n, x) = \sum_{i=1}^{ndl_{ref}} U_{ref_i}\phi_i(x) \quad \text{et} \quad U_{\Delta t,h}(t, x) = \sum_{i=1}^{ndl_h} U_{\Delta t,h_i}\psi_i(x) \tag{7.1}$$

On pose $U_{ref}(t_n, .) = U^n_{ref}$ et $U_{\Delta t,h}(t_n, .) = U^n_{\Delta t,h}$.
En espace, on considère la norme $L^2(\Omega)$ et la semi-norme $H^1(\Omega)$, puis on définit les normes relatives suivantes,

— L'erreur globale e_0 en norme L^2 et l'erreur globale e_1 en norme H^1. Ces deux erreurs permettent d'évaluer de deux façons différentes l'erreur totale commise lors de l'approximation en temps et en espace.

$$e_0 = \frac{\max\limits_n \|U^n_{\Delta t,h} - U^n_{ref}\|_{L^2(\Omega)}}{\max\limits_n \|U^n_{ref}\|_{L^2(\Omega)}} \quad , \quad e_1 = \frac{\max\limits_n \|\nabla(U^n_{\Delta t,h} - U^n_{ref})\|_{L^2(\Omega)}}{\max\limits_n \|\nabla U^n_{ref}\|_{L^2(\Omega)}}. \tag{7.2}$$

— L'erreur spatiale ρ^h_L en norme L^2 et l'erreur spatiale ρ^h_H en norme H^1. Ces deux erreurs permettent d'évaluer de deux façons différentes l'erreur commise lors de la discrétisation en espace. En effet, Δt_{ref} étant très petit, l'erreur en espace est négligeable devant l'erreur en temps.

$$\rho^h_L = \frac{\max\limits_n \|U^n_{\Delta t_{ref},h} - U^n_{ref}\|_{L^2(\Omega)}}{\max\limits_n \|U^n_{ref}\|_{L^2(\Omega)}} \quad , \quad \rho^h_H = \frac{\max\limits_n \|\nabla(U^n_{\Delta t_{ref},h} - U^n_{ref})\|_{L^2(\Omega)}}{\max\limits_n \|\nabla U^n_{ref}\|_{L^2(\Omega)}}. \tag{7.3}$$

— L'erreur temporelle $\theta^{\Delta t}_L$ en norme L^2 et l'erreur temporelle $\theta^{\Delta t}_H$ en norme H^1. Ces deux erreurs permettent d'évaluer de deux façons différentes l'erreur de discrétisation en temps. En effet, après avoir discrétisé le problème en espace, le problème obtenu

est un système d'équation aux dérivées ordinaires et permet d'évaluer la précision du schéma de discrétisation en temps.

$$\theta_L^{\Delta t} = \frac{\max\limits_n \|U^n_{\Delta t_{ref},h} - U^n_{\Delta t,h}\|_{L^2(\Omega)}}{\max\limits_n \|U^n_{ref}\|_{L^2(\Omega)}} \ , \ \theta_H^{\Delta t} = \frac{\max\limits_n \|\nabla(U^n_{\Delta t_{ref},h} - U^n_{\Delta t,h})\|_{L^2(\Omega)}}{\max\limits_n \|\nabla U^n_{ref}\|_{L^2(\Omega)}}. \tag{7.4}$$

Contrairement aux erreurs évaluées en norme $L^2(\Omega)$, les erreurs en semi-norme $H^1(\Omega)$ font intervenir des dérivées premières.
Nous travaillerons sur des maillages imbriqués voir figure (7.4) de façon à faciliter le calcul des normes définis en (7.2) et (7.3). En effet en considérant des maillages imbriqués, on

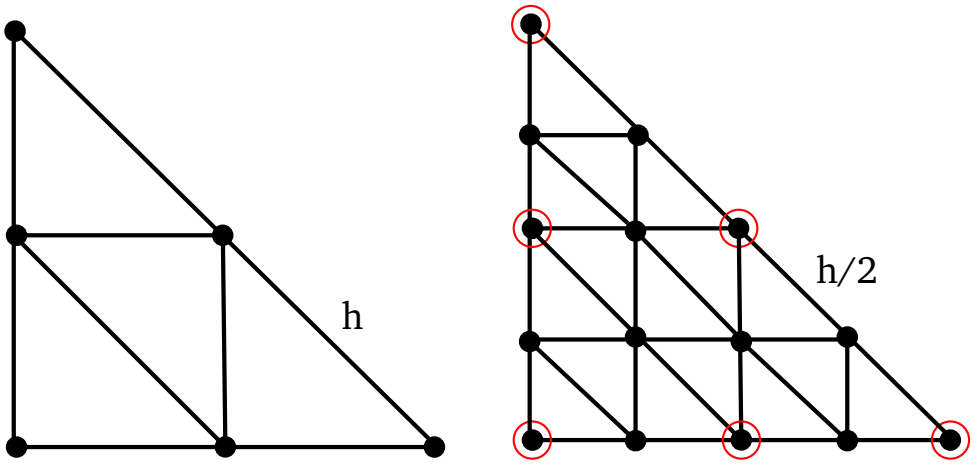


FIGURE 7.4: Deux maillages imbriqués : Le maillage de gauche est de diamètre h et le maillage de droite plus fin de diamètre h/2. Le maillage à droite contient tous les points du maillage à gauche. Ces points sont identifiés dans le maillage de droite par des cercles en rouges

aura toujours $V_h \subset V_{h_{ref}}$. Ainsi, on pourra écrire facilement la solution numérique $U^n_{\Delta t,h}$ à chaque instant t_n dans la base fondamentale $(\phi_i)_{1\leq i\leq nd_{ref}}$ de $V_{h_{ref}}$.
Les erreurs définies en (7.2) permettent de comparer globalement une solution numérique et la solution de référence. Les erreurs définies en (7.3) et (7.4) permettent d'avoir une majoration de (7.2) en fonction de h et Δt. En effet, en écrivant

$$U^n_{\Delta t,h} - U^n_{ref} = (U^n_{\Delta t,h} - U^n_{\Delta t_{ref},h}) + (U^n_{\Delta t_{ref},h} - U^n_{ref}),$$

on obtient par l'inégalité triangulaire,

$$e_0 \leq \rho^h_L + \theta^{\Delta t}_L, \tag{7.5}$$
$$e_1 \leq \rho^h_H + \theta^{\Delta t}_H. \tag{7.6}$$

Le principal objectif de cette partie est d'étudier la dépendance de e_0 et e_1 à Δt et h. Nous allons commencer par analyser numériquement comment varie $\theta^{\Delta t}_L$ et $\theta^{\Delta t}_H$ quand Δt décroît. Ceci nous permettra d'évaluer la précision et l'ordre de convergence du schéma en temps utilisé dans la résolution numérique.

Nous allons ensuite analyser comment varie ρ_L^h et ρ_H^h quand h décroît. Cette analyse va nous permettre d'évaluer la précision et l'ordre de convergence du schéma de discrétisation en espace utilisé dans la résolution numérique.
Une fois les erreurs temporelles et spatiales ρ_L^h, $\theta_L^{\Delta t}$, ρ_H^h et $\theta_H^{\Delta t}$ calculées, nous utiliserons (7.5) et (7.6) pour déduire une majoration en fonction du pas de temps Δt et du diamètre du maillage h des erreurs globales e_0 et e_1. Plus précisément, nous pourrons montrer numériquement que $e_0 \leq C(\Delta t^k + h^{r+1})$ (resp $e_1 \leq C(\Delta t^k + h^r)$) où k est l'ordre du schéma d'intégration en temps et r est l'ordre du schéma d'intégration en espace en semi-norme $H^1(\Omega)$, utilisés pour calculer $U_{\Delta t,h}$. Nous allons pour conclure, calculer l'ordre global de convergence pour les normes e_0 et e_1, et montrer sur un exemple comment il faut choisir Δt et h pour avoir une erreur inférieure à une valeur donnée. Dans toute la suite, nous utiliserons en 1D $(h_{ref}, \Delta t_{ref}) = (0.1/2^7, 0.1/2^7)$ et $(h_{ref}, \Delta t_{ref}) = (0.1/2^6, 0.1/2^6)$ en 2D.

7.1 Étude 1D

7.1.1 Convergence des schémas en temps, précision sur le problème semi-discret

Convergence en temps

L'étude des erreurs $\theta_L^{\Delta t}$ et $\theta_H^{\Delta t}$ définis en (7.4) consiste à évaluer numériquement leurs décroissances par rapport à celle de Δt, à maillage de diamètre h et schémas d'intégration en espace fixés. La pente de cette décroissance correspond à l'ordre de convergence du schéma en temps utilisé pour le problème semi-discrétisé. Pour un schéma en temps d'ordre k, on s'attend à observer,

$$\theta_L^{\Delta t} \leq C_1(h)\Delta t^k \qquad \text{et} \qquad \theta_H^{\Delta t} \leq C_2(h)\Delta t^k. \tag{7.7}$$

Où les constantes $C_1(h)$ et $C_2(h)$ sont des constantes qui peuvent dépendre de la méthode d'intégration en espace et du diamètre h du maillage utilisés. Cette dépendance sera étudiée numériquement dans cette section. On va d'ailleurs montrer (Numériquement) que pour une discrétisation assez fine en espace, C_1 et C_2 ne dépendent ni de h , ni de la méthode d'intégration en espace utilisée.
Commençons par l'étude numérique de la convergence du problème semi-discret. Il s'agit de : fixer h et une méthode d'intégration en espace, calculer numériquement les valeurs de $\theta_L^{\Delta t}$ (resp $\theta_H^{\Delta t}$) et analyser la pente de décroissance lorsque Δt décroît. Pour cette étude de convergence, nous fixons $h = 0.025$. Les erreurs spatiales $\theta_L^{\Delta t}$ et $\theta_H^{\Delta t}$ et les ordres de convergence correspondants à chaque méthode sont représentés dans la table (7.2)

ci-dessous.

TABLE 7.2: Erreurs $\theta_L^{\Delta t}$, $\theta_H^{\Delta t}$ en 1D et pente k associée pour les schémas $\mathbb{P}_1 + RL1 + FBE$ (au dessus à gauche), $\mathbb{P}_1 + RL2 + SBDF2$ (au dessus à droite), $\mathbb{P}_2 + RL2 + SBDF2$ (en bas à gauche) et $\mathbb{P}_2 + RL3 + SBDF3$ (en bas à droite) à maillage fixé de diamètre $h = 0.025$

$\mathbb{P}_1 + RL1 + FBE$				
Δt	$\theta_L^{\Delta t}$	k	$\theta_H^{\Delta t}$	k
0.1	6.52×10^{-3}	–	0.059	–
0.05	3.29×10^{-3}	0.98	0.029	1.00
0.025	1.66×10^{-3}	0.99	0.014	1.00
0.0125	8.32×10^{-4}	1.00	0.007 25	1.00
6.25×10^{-3}	4.17×10^{-4}	1.00	0.003 62	1.00

$\mathbb{P}_1 + RL2 + SBDF2$			
$\theta_L^{\Delta t}$	k	$\theta_H^{\Delta t}$	k
3.4×10^{-3}	–	0.03	–
8.1×10^{-4}	2.07	7.58×10^{-3}	1.98
1.98×10^{-4}	2.03	1.89×10^{-3}	2.00
4.87×10^{-5}	2.02	4.58×10^{-4}	2.04
1.2×10^{-5}	2.01	1.122×10^{-3}	2.03

$\mathbb{P}_2 + RL2 + SBDF2$				
Δt	$\theta_L^{\Delta t}$	k	$\theta_H^{\Delta t}$	k
0.1	3.58×10^{-3}	–	0.032	–
0.05	8.47×10^{-4}	2.08	8.02×10^{-3}	2.00
0.025	2.06×10^{-4}	2.03	1.93×10^{-3}	2.05
0.0125	5.08×10^{-5}	2.02	4.74×10^{-4}	2.02
6.25×10^{-3}	1.25×10^{-5}	2.01	1.17×10^{-4}	2.01

$\mathbb{P}_2 + RL3 + SBDF3$			
$\theta_L^{\Delta t}$	k	$\theta_H^{\Delta t}$	k
1.15×10^{-3}	–	0.012	–
1.31×10^{-4}	3.13	1.38×10^{-3}	3.30
1.69×10^{-5}	2.95	1.92×10^{-4}	2.84
2.12×10^{-6}	3.00	2.41×10^{-5}	2.99
2.62×10^{-7}	3.01	2.99×10^{-6}	3.01

Dans la table (7.2) on observe la convergence en temps à l'ordre 1, 2, 3 respectivement des schémas combinés RL_k et SBDFk, k= 1, 2 et 3, appliqués au problème monodomaine en 1D, où le modèle ionique de Beeler Reuter est utilisé.
Cette table nous montre qu'on a les même ordres de convergence pour les deux normes considérées. Pour les schémas d'ordre 1 et 2, les régimes de convergences sont atteints à partir du pas de temps $\Delta t = 0.05$ tandis-qu'à l'ordre 3, ce régime est atteint à partir de $\Delta t = 0.0125$. La combinaison des schémas $SBDF$ et RL au même ordre est donc cohérente dans la mesure où elle ne dégrade pas l'ordre de convergence de l'un ou de l'autre. La table (7.2) confirme donc les ordres attendus décrites par les inégalités 7.7. Il reste à étudier la dépendance des constantes $C_1(h)$ et $C_2(h)$ par rapport à h et au schéma d'intégration en espace utilisé, et d'en déduire une approximation des constantes C_1 et C_2.

Étude des constantes C_1 et C_2

Les valeurs de $\theta_L^{\Delta t}$, $\theta_H^{\Delta t}$ et les constantes associées sont évaluées numériquement pour divers h et représentées dans les tables 7.3–7.10. Dans toutes ces tables, le bloc supérieur représente l'erreur $\theta_L^{\Delta t}$ (resp $\theta_H^{\Delta t}$) et le bloc inférieur représente la constante $C_1(h)$ (resp $C_2(h)$). Les codes couleurs expriment la proximité entre les chiffres.
Dans les tables 7.3–7.8 où on a les schémas d'ordre 1 et 2, on peut voir que sur chaque colonne de la partie supérieure, les erreurs sont à deux chiffres significatifs près égales ou très proches. Ceci témoigne donc la dépendance très faible des erreurs $\theta_L^{\Delta t}$ (resp $\theta_H^{\Delta t}$)

par rapport à h. Dans la partie inférieure de chacune de ces tables, les valeurs de $C_1(h)$ (resp $C_2(h)$) sont aussi à deux chiffres près égales ou très proches. Jusqu'ici, d'après cette analyse, on peut dire que les constantes $C_1(h)$ et $C_2(h)$ ne dépendent pas de h de manière significative.
Par ailleurs, en regardant les valeurs dans la table 7.5 (resp 7.5) correspondant au cas où l'espace est intégré avec la méthode d'élément fini $\mathbb{P}_1$ et la table 7.7 (resp 7.8) où l'espace est intégré avec la méthode d'élément fini $\mathbb{P}_2$, on constate que toutes les valeurs sont égales à deux chiffres significatifs près. On peut donc conclure que $C_1(h)$ et $C_2(h)$ ne dépendent pas de manière significative de la méthode de discrétisation en espace utilisée.
Nous avons fait une analyse pour les schémas d'ordre 1 et 2. À présent, faisons la même analyse pour le schéma d'ordre 3. On peut observer dans le bloc supérieur de chacune des tables 7.9 et 7.10 que sur chaque colonne, les valeurs sont à 3 chiffres significatifs près égales ou très proches. Ceci confirme donc comme pour l'ordre 1 et 2 la dépendance très faible des erreurs $\theta_L^{\Delta t}$ et $\theta_H^{\Delta t}$ par rapport à h.

TABLE 7.3: Erreurs $\theta_L^{\Delta t}$ et constantes $C_1(h)$, en 1D du schéma $\mathbb{P}_1 + RL1 + FBE$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.00324422	0.00163048	0.000817399	0.00040925	0.00020476
0.025	0.00329762	0.00165983	0.000832856	0.00041719	0.00020879
0.0125	0.00331859	0.00167087	0.000838726	0.00042023	0.00021033
0.00625	0.00332487	0.00167416	0.000840333	0.00042102	0.00021073
constantes $C1(h) = \Delta t^{-1} \times \theta_L^{\Delta t}$					
0.05	0.0648845	0.065219	0.0653919	0.0654799	0.0655244
0.025	0.0659524	0.066393	0.0666284	0.0667504	0.0668127
0.0125	0.0663719	0.066835	0.0670981	0.0672368	0.0673081
0.00625	0.0664974	0.0669665	0.0672266	0.0673633	0.0674335

TABLE 7.4: Erreurs $\theta_H^{\Delta t}$ et constantes $C_2(h)$, en 1D du schéma $\mathbb{P}_1 + RL1 + FBE$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.0276843	0.0138179	0.00690353	0.00345048	0.00172492
0.025	0.0291832	0.0145416	0.00725928	0.0036269	0.00181278
0.0125	0.0296436	0.0147652	0.00735762	0.00367305	0.00183516
0.00625	0.0297375	0.0148125	0.00739464	0.00369607	0.00184792
constantes $C_2(h) = \Delta t^{-1} \times \theta_H^{\Delta t}$					
0.05	0.553686	0.552716	0.552283	0.552076	0.551975
0.025	0.583665	0.581666	0.580742	0.580303	0.58009
0.0125	0.592873	0.59061	0.58861	0.587688	0.58725
0.00625	0.594751	0.592502	0.591571	0.591371	0.591334

Dans le bloc inférieur de chacun de ces tableaux les valeurs de $C_1(h)$ et $C_2(h)$ sont aussi à 3 chiffres significatifs près égales ou très proches. Il vient donc comme pour l'ordre 1 et 2 que la dépendance de ces constantes au diamètre du maillage h est très faible. Elle est de plus en plus faible lorsque h tend vers 0. Nous allons pour la suite noter ces

TABLE 7.5: Erreurs $\theta_L^{\Delta t}$ et constantes $C_1(h)$, en 1D du schéma $\mathbb{P}_1 + RL2 + SBDF2$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.00074248	0.00018086 9	4.45046E-5	1.10236E-5	2.74851E-6
0.025	0.000809681	0.00019 8125	4.86892E-5	1.20573E-5	2.99972E-6
0.0125	0.000838179	0.000204549	5.03153E-5	1.24638E-5	3.10119E-6
0.00625	0.000846052	0.000206063	5.07196E-5	1.25687E-5	3.12795E-6
constantes $C_1(h) = \Delta t^{-2} \times \theta_L^{\Delta t}$					
0.05	0.296992	0.28939	0.28483	0.282204	0.281447
0.025	0.323872	0.316999	0.311611	0.308668	0.307171
0.0125	0.335272	0.327278	0.322018	0.319073	0.317562
0.00625	0.338421	0.329701	0.324605	0.32176	0.320302

TABLE 7.6: Erreurs $\theta_H^{\Delta t}$ et constantes $C_2(h)$, en 1D du schéma $\mathbb{P}_1 + RL2 + SBDF2$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.00711639	0.00172049	0.000412665	0.000101104	2.52468E-5
0.025	0.00758421	0.00189207	0.000458114	0.000112095	2.76967E-5
0.0125	0.00785129	0.00191358	0.000467223	0.000114966	2.84972E-5
0.00625	0.00798462	0.00193765	0.000472865	0.000116306	2.8827E-5
constantes $C_2(h) = \Delta t^{-2} \times \theta_H^{\Delta t}$					
0.05	2.84656	2.75278	2.64106	2.58825	2.58527
0.025	3.03368	3.02731	2.93193	2.86964	2.83614
0.0125	3.14051	3.06172	2.99023	2.94312	2.91811
0.00625	3.19385	3.10024	3.02634	2.97744	2.95189

TABLE 7.7: Erreurs $\theta_L^{\Delta t}$ et constantes $C_1(h)$, en 1D du schéma $\mathbb{P}_2 + RL2 + SBDF2$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.000845044	0.000206075	5.07008E-5	1.25612E-5	3.12569E-6
0.025	0.00084775	0.000206478	5.08056e-05	1.25865E-5	3.13184E-6
0.0125	0.000848762	0.000206593	5.08604E-5	1.26053E-5	3.13726E-6
0.00625	0.000848667	0.000206576	5.08536E-5	1.26045E-5	3.13721E-6
constantes $C_1(h) = \Delta t^{-2} \times \theta_L^{\Delta t}$					
0.05	0.338018	0.329720	0.324485	0.321566	0.320071
0.025	0.339100	0.330365	0.325156	0.322214	0.320700
0.0125	0.339505	0.330548	0.325507	0.322695	0.321255
0.00625	0.339467	0.330522	0.325463	0.322676	0.321250

TABLE 7.8: Erreurs $\theta_H^{\Delta t}$ et constantes $C_2(h)$, en 1D du schéma $\mathbb{P}_2 + RL2 + SBDF2$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.00760124	0.00185934	0.000451638	0.000110671	2.73638E-5
0.025	0.00802492	0.00193449	0.000474266	0.000117027	2.90548E-5
0.0125	0.00802401	0.00194317	0.000474305	0.00011668	2.89239E-5
0.00625	0.00805579	0.00195246	0.000476812	0.000116963	2.89603E-5
constantes $C_2(h) = \Delta t^{-2} \times \theta_H^{\Delta t}$					
0.05	3.04049	2.97494	2.89048	2.83317	2.80206
0.025	3.20997	3.09518	3.0353	2.9959	2.97521
0.0125	3.2096	3.10907	3.03555	2.987	2.9618
0.00625	3.22232	3.12394	3.05159	2.99425	2.96553

TABLE 7.9: Erreurs $\theta_L^{\Delta t}$ et constantes $C_1(h)$, en 1D du schéma $\mathbb{P}_2 + RL3 + SBDF3$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.000130589	1.69782E-5	2.12005E-6	2.62269E-7	3.25386E-8
0.025	0.000131621	1.69857E-5	2.12169E-6	2.62636E-7	3.25909E-8
0.0125	0.000131553	1.69969E-5	2.12159E-6	2.626E-7	3.25889E-8
0.00625	0.000131553	1.69980E-5	2.12189E-6	2.62606E-7	3.25891E-8
constantes $C_1(h) = \Delta t^{-3} \times \theta_L^{\Delta t}$					
0.05	1.04471	1.0866	1.08546	1.07426	1.06623
0.025	1.05296	1.08708	1.08631	1.07576	1.06794
0.0125	1.05242	1.0878	1.08625	1.07561	1.06787
0.00625	1.05242	1.08787	1.08641	1.07563	1.06788

TABLE 7.10: Erreurs $\theta_H^{\Delta t}$ et constantes $C_2(h)$, en 1D du schéma $\mathbb{P}_2 + RL3 + SBDF3$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.00134318	0.000188523	2.39029E-5	2.98064E-6	3.70819E-7
0.025	0.00138398	0.000192915	2.41636E-5	2.99125E-6	3.71623E-7
0.0125	0.00139098	0.000192520	2.43335E-5	3.00395E-6	3.72311E-7
0.00625	0.00139639	0.000192372	2.42901E-5	3.00722E-6	3.72533E-7
constantes $C_2(h) = \Delta t^{-2} \times \theta_H^{\Delta t}$					
0.05	10.7454	12.0655	12.2383	12.2087	12.151
0.025	11.0718	12.3465	12.3718	12.2522	12.1773
0.0125	11.1278	12.3213	12.4588	12.3042	12.1999
0.00625	11.1711	12.3118	12.4365	12.3176	12.2071

constantes sans le h (ie : C_1 et C_2 au lieu de $C_1(h)$ et $C_2(h)$). On peut se servir des tables 7.3–7.10 pour donner une approximation des constantes C_1 et C_2 pour chaque schéma. Plus précisément, on utilise : les tables 7.3 et 7.4 pour le schéma $RL1 + FBE$, les tables 7.5–7.8 pour le schéma $RL2 + SBDF2$ et les tables 7.9–7.10 pour le schéma $RL3 + SBDF3$. Nous nous sommes servis de ces tables pour donner dans la table 7.11, une approximation des constantes C_1 et C_2 pour chaque schéma.

TABLE 7.11: Valeurs approchées des constantes C_1 et C_2 pour les schémas $RL1 + FBE$, $RL2 + SBDF2$ et $RL3 + SBDF3$.

Constantes \ Schémas	$RL1 + FBE$	$RL2 + SBDF2$	$RL3 + SBDF3$
C_1	0.067	0.33	1.07
C_2	0.59	2.9	12.3

7.1.2 Étude de la convergence en espace

L'étude des erreurs ρ_L^h et ρ_H^h définis en (7.3) consiste à évaluer numériquement leurs décroissances par rapport à celle de h, à pas de temps Δt_{ref} et schémas d'intégration en temps fixés (ici $RL4 + SBDF4$). Ce pas de temps Δt_{ref} est très petit et la méthode d'intégration en temps utilisée est très précise (ordre 4). Par conséquent, dans l'erreur totale commise lors du calcul d'une solution sur un maillage grossier de diamètre h, l'erreur de discrétisation en temps est négligeable devant l'erreur de discrétisation en espace. La pente de décroissance par rapport à h de l'erreur totale dans ce cas (cas où Δt_{ref} est très petit et la méthode d'intégration en temps très précise) permet donc d'évaluer les constantes C_3 et C_4 tels que,

$$\rho_L^h \leq C_3 h^{r+1} \qquad \text{et} \qquad \rho_H^h \leq C_4 h^r. \tag{7.8}$$

où h désigne le diamètre du maillage sur lequel est calculée la solution numérique et p est un entier que nous allons déterminer numériquement dans cette section pour les méthodes d'élément finis $\mathbb{P}_1$ et $\mathbb{P}_2$.
Dans la table 7.12 (resp 7.13) à gauche, on observe que la pente de décroissance de ρ_L^h (resp ρ_H^h) est 1.99 (resp 1) donc presque 2. À droite, on observe que la pente est 3 (resp $1.97 \simeq 2$) pour ρ_L^h (resp ρ_H^h). Ces pentes correspondent donc aux ordres de convergence classiques des éléments finis $\mathbb{P}_1$ et $\mathbb{P}_2$ en normes L^2 et H^1.

TABLE 7.12: Erreurs ρ_L^h, constante C_3 en 1D et pente r associée pour les schémas $\mathbb{P}_1$ (à gauche) et $\mathbb{P}_2$ (à droite)

$\mathbb{P}_1$			
h	ρ_L^h	k	$h^{-2} \times \rho_L^h$
0.05	5.3×10^{-3}	–	2.12
0.025	1.46×10^{-3}	1.85	2.33
0.0125	3.76×10^{-4}	1.95	2.41
6.25×10^{-3}	9.47×10^{-5}	1.98	2.42
3.125×10^{-3}	2.37×10^{-5}	1.99	2.43

$\mathbb{P}_2$		
ρ_L^h	k	$h^{-3} \times \rho_L^h$
2.89×10^{-4}	–	2.31
2.84×10^{-5}	3.34	1.81
3.42×10^{-6}	3.05	1.75
4.25×10^{-7}	3.00	1.74
5.31×10^{-8}	3.00	1.74

TABLE 7.13: Erreurs ρ_H^h et C_4 en 1D et pente r associée pour les schémas $\mathbb{P}_1$ (à gauche) et $\mathbb{P}_2$ (à droite)

$\mathbb{P}_1$			
h	ρ_H^h	r	$h \times \rho_H^h$
0.05	0.145	–	2.90
0.025	0.072	1.01	2.88
0.0125	0.036	0.99	2.88
6.25×10^{-3}	0.018	1.00	2.88
3.125×10^{-3}	8.99×10^{-3}	1.00	2.88

$\mathbb{P}_2$		
ρ_H^h	r	$h^{-2} \times \rho_H^h$
0.026	–	10.67
6.47×10^{-3}	2.04	10.36
1.66×10^{-3}	1.95	10.65
4.25×10^{-4}	1.96	10.88
1.07×10^{-4}	1.97	11.05

Les tables 7.12 et 7.13 donnent aussi des approximations des constantes C_3 et C_4. Nous avons utilisé cette table pour déduire une approximation de ces constantes pour chaque méthode. Les résultats sont donnés dans la table 7.14.

TABLE 7.14: Valeurs approchées des constantes C_3 et C_4 pour les méthodes d'éléments finis Lagrange $\mathbb{P}_1$ et $\mathbb{P}_2$.

Constantes \ Schémas	$\mathbb{P}_1$	$\mathbb{P}_2$
C_3	2.42	1.74
C_4	2.88	11.05

7.1.3 Convergence globale

Dans cette section, nous allons commencer par exploiter les résultats des sous-sections 7.1.1 et 7.1.2 pour donner une majoration des erreurs globales e_0 et e_1 pour les schémas que nous avons précédemment étudiés. D'après les inégalité (7.5), (7.6), (7.7) et (7.8) et les constantes calculées dans les sous-sections 7.1.1 et 7.1.2 (tables 7.11 et 7.14), on a pour h et Δt assez petits, les inégalités pratiques suivantes.

— $P_1 + RL1 + FBE$

$$e_0 \leq 2.42 \times h^2 + 0.067 \times \Delta t. \tag{7.9}$$

$$e_1 \leq 2.88 \times h + 0.59 \times \Delta t. \tag{7.10}$$

— $P_1 + RL2 + SBDF2$

$$e_0 \leq 2.42 \times h^2 + 0.33 \times \Delta t^2. \tag{7.11}$$
$$e_1 \leq 2.88 \times h + 2.9 \times \Delta t^2. \tag{7.12}$$

— $P_2 + RL2 + SBDF2$

$$e_0 \leq 1.75 \times h^3 + 0.33 \times \Delta t^2. \tag{7.13}$$
$$e_1 \leq 11.05 \times h^2 + 2.9 \times \Delta t^2. \tag{7.14}$$

— $P_2 + RL3 + SBDF3$

$$e_0 \leq 1.75 \times h^3 + 1.07 \times \Delta t^3. \tag{7.15}$$
$$e_1 \leq 11.05 \times h^2 + 12.3 \times \Delta t^3. \tag{7.16}$$

Pour vérifier ces inégalités, nous calculons d'abord les erreurs e_0 (resp e_1) et la combinaison $C_1 h^{r+1} + C_2 \Delta t^k$ (resp $C_3 h^r + C_4 \Delta t^k$) pour divers $(h, \Delta t)$. Nous calculons ensuite les coefficients d'efficacité η_0 et η_1 définis par,

$$\eta_0 = \frac{e_0}{C_1 h^{r+1} + C_2 \Delta t^k}, \qquad \text{et} \qquad \eta_1 = \frac{e_1}{C_3 h^r + C_4 \Delta t^k}. \tag{7.17}$$

Les coefficients η_0 et η_1 nous permettent de dire pour chaque $(h, \Delta t)$, si l'une des inégalités (7.9)–(7.16) selon le schéma et la norme choisis, est vérifiée ou pas. En effet on pourra considérer trois situations : pour $l = 0$ ou 1,

1 le coefficient η_l peut être proche de la valeur 1 ($0.8 \leq \eta_l \leq 1.2$). Alors, on a presque l'égalité.

2 Le coefficient η_l peut être plus petit que 1 ($\eta_l \leq 1$). Alors, l'inégalité est vérifiée.

3 Le coefficient η_l n'est pas dans le cas 1 et 2 ($\eta_l > 1.2$). On va alors dire que l'inégalité n'est pas vérifiée

Les erreurs e_0 (resp e_1) ainsi que le coefficient d'efficacité η_0 (resp : η_1) correspondant pour divers $(h, \Delta t)$ sont présentés dans les tables 7.15–7.22. Le bloc supérieur de chacune de ces tables contient les erreurs, et le bloc inférieur les coefficients d'efficacité correspondants. Les coefficients d'efficacité strictement supérieurs à 1 et inférieur à 1.19 ($1 < \eta_l \leq 1.19$; $l = 0, 1.$) ont été coloriés en bleu et les coefficients supérieurs ou égaux à 2 ($\eta_l \geq 2$; $l = 0, 1.$) en rouge

Les coefficients d'efficacité donnés par les tables 7.15–7.22 sont inférieurs à 1 à partir d'un certain $(h, \Delta t)$. Cette infériorité à 1 nous permet de dire que, les inégalités (7.9)–(7.16) sont vraies à partir d'un certain $(h, \Delta t)$. Dans le cas où ces inégalités ne sont pas vérifiées, on est au moins proche de l'égalité (voir les chiffres en bleu des tables) à l'exception de quelques cas où les coefficients d'efficacité sont largement au dessus de 1 (ie. >2) (voir les chiffres rouge des tables 7.19 et 7.19). Les cas où on est largement au

dessus de 1 concernent donc uniquement les schémas $\mathbb{P}_2 + RL2 + SBDF2$ (table 7.19) et $\mathbb{P}_2 + RL2 + SBDF2$ (table 7.21), lorsque le pas de temps utilisé est grand et maillage utilisé est grossiers ($\Delta t, h \geq 0.05$).

Les coefficients d'efficacité dans la plupart des cas sont entre 0.8 et 1.15. Cette proximité à 1 des coefficients d'efficacité montrent que les majorations données par les formules (7.9)–(7.16) ne sont pas exagérées et peuvent donc servir à choisir de façon assez efficace $(h, \Delta t)$, pour le calcul d'une solution numérique à une précision donnée. En effet connaissant les constantes C_i $i = 1, \ldots, 4$ pour chaque schéma numérique, on sait que pour une discrétisation $(h, \Delta t)$, on a $e_0 \leq C_1 h^{r+1} + C_2 \Delta t^k$ (resp $e_1 \leq C_3 h^r + C_4 \Delta t^k$), avec r et k donné par une des inégalité (7.9)–(7.16) qui correspond au schéma numérique utilisé. L'efficacité vient du fait de la proximité en grandeur des quantités de part et d'autre de l'inégalité. Cependant, si on doit se servir des formules (7.9)–(7.16) pour choisir pour une

TABLE 7.15: Erreur globales e_0 et coefficients d'efficacité η_0 en 1D du schéma $\mathbb{P}_1 + RL1 + FBE$ quand Δt et h varient

h \ Δt	0.1	0.05	0.025	0.0125	0.00625	0.003125
0.1	0.0276312	0.0276314	0.0278362	0.027981	0.0280577	0.0280969
0.05	0.0083977	0.00625804	0.00557323	0.00538555	0.00533346	0.00531537
0.025	0.00674909	0.00375501	0.00238404	0.00177815	0.001566	0.00150072
0.0125	0.00660685	0.00337242	0.00175835	0.000988523	0.000623282	0.000462578
0.00625	0.00658251	0.00333674	0.00168728	0.000857043	0.00044878	0.000251103
0.003125	0.00657713	0.00332926	0.001678	0.000844	0.000424727	0.000215817
Coefficients d'efficacité η_0						
0.1	0.89421359	1.00295463	1.07579517	1.11756365	1.13968825	1.15106979
0.05	0.65864314	0.66574894	0.72145372	0.78193103	0.82449623	0.84918474
0.025	0.821807	0.77223856	0.74793412	0.75665957	0.81087379	0.87155899
0.0125	0.93341744	0.90458755	0.85642409	0.81317753	0.7821529	0.78736681
0.00625	0.96879549	0.96870691	0.95351873	0.91954444	0.87433759	0.82624132
0.03125	0.97821117	0.98684799	0.98785492	0.98010753	0.96009539	0.92621347

précision donnée le couple (h, Δt) qu'il faut, on doit tenir compte de certains paramètres pour que le choix se fasse efficacement. Les constantes C_1 (resp C_3) et C_2 (resp C_4) ne sont pas du même ordre de grandeur, le choix de $(h, \Delta t)$ doit être de sorte à équilibrer $C_1 h^{r+1}$ (resp $C_3 h^r$) et $C_2 \Delta t^k$ (resp $C_2 \Delta t^k$). Si cette équilibre n'est pas pris en compte, une des quantités dominera sur l'autre. La conséquence sera un surcoût en temps de calcul. On peut d'ailleurs observer sur les lignes et sur les colonnes des tables 7.15–7.22 qu'à partir d'une certaine valeur de Δt ou de h, les erreurs ou les quantités $C_1 h^{r+1} + C_2 \Delta t^k$ (resp $C_3 h^r + C_4 \Delta t^k$) ne décroissent plus considérablement. Utiliser donc un $(h, \Delta t)$ dans une telle zone revient à ajouter un surcoût en temps de calcul.

TABLE 7.16: Erreur globale e_1 et coefficients d'efficacité η_1 en 1D du schéma $\mathbb{P}_1 + RL1 + FBE$ quand Δt et h varient

h \ Δt	0.1	0.05	0.025	0.0125	0.00625	0.003125
0.1	0.259236	0.244038	0.242033	0.242437	0.242761	0.242951
0.05	0.15759	0.162965	0.157286	0.150327	0.147426	0.146318
0.025	0.0814979	0.077622	0.0797316	0.0811169	0.0747323	0.0728349
0.0125	0.0664884	0.0393442	0.0342905	0.0383267	0.0396659	0.0402663
0.00625	0.0616824	0.0326331	0.0193714	0.0151461	0.0172446	0.0190733
0.003125	0.0603452	0.0305529	0.0161509	0.00960858	0.00701202	0.00755329
Coefficients d'efficacité η_1						
0.1	0.74707781	0.76862362	0.79944839	0.82077698	0.83226398	0.83821369
0.05	0.77630542	0.93927954	0.99077795	0.9930768	0.99822937	1.00325177
0.025	0.62212137	0.76474877	0.91909625	1.0219452	0.98737969	0.98633804
0.0125	0.69987789	0.60067481	0.67567488	0.88361268	0.99945575	1.06401453
0.00625	0.80107013	0.68701263	0.59149313	0.59689064	0.79514006	0.96117417
0.03125	0.88742941	0.79358182	0.68003789	0.58678351	0.55267153	0.696557

TABLE 7.17: Erreur globale e_0 et coefficients d'efficacité η_0 en 1D du schéma $\mathbb{P}_1 + RL2 + SBDF2$ quand Δt et h varient

h \ Δt	0.1	0.05	0.025	0.0125	0.00625	0.003125
0.1	0.0269448	0.0278793	0.0280809	0.0281238	0.0281335	0.0281359
0.05	0.00401193	0.00478278	0.00517522	0.00527055	0.0052937	0.00529939
0.025	0.00241652	0.00107465	0.00132271	0.00142641	0.00145266	0.00145916
0.0125	0.00321306	0.000605959	0.000274878	0.000341297	0.000367514	0.000374207
0.00625	0.00349358	0.00076461	0.000154496	6.93762e-05	8.59576e-05	9.25207e-05
0.003125	0.00356497	0.000827452	0.000186609	3.8709e-05	1.73842e-05	2.15377e-05
Coefficients d'efficacité η_0						
0.1	0.98338686	1.115172	1.15085656	1.15974433	1.16194115	1.16249014
0.05	0.43372216	0.69821606	0.8280352	0.86402459	0.87318763	0.8754793
0.025	0.51278939	0.46471351	0.7723854	0.9129024	0.95256393	0.96274157
0.0125	0.89797185	0.51433967	0.47546054	0.79718076	0.9408238	0.98152656
0.00625	1.06041833	0.85476076	0.5245492	0.48000913	0.80310938	0.94741245
0.03125	1.10589713	1.00464043	0.83445423	0.52570322	0.48111965	0.80491296

TABLE 7.18: Erreur globale e_1 et coefficients d'efficacité η_1 en 1D du schéma $\mathbb{P}_1 + RL2 + SBDF2$ quand Δt et h varient

h \ Δt	0.1	0.05	0.025	0.0125	0.00625	0.003125
0.1	0.227899	0.239089	0.242295	0.242964	0.243108	0.243142
0.05	0.130031	0.141174	0.144601	0.145133	0.145168	0.14519
0.025	0.0499591	0.0638334	0.0709571	0.072082	0.0720528	0.0720509
0.0125	0.0406006	0.0226245	0.0317499	0.0358284	0.0361218	0.0361128
0.00625	0.0344617	0.0174452	0.00949161	0.0159599	0.0179473	0.0180326
0.003125	0.0328657	0.0110804	0.00875051	0.00343528	0.0079368	0.00897396
Coefficients d'efficacité η_1						
0.1	0.71892429	0.80978493	0.83604054	0.84229978	0.84379311	0.84416005
0.05	0.75162428	0.93338182	0.99169138	1.00470658	1.00731869	1.00806564
0.025	0.49464455	0.80546877	0.9613155	0.99487779	0.99916132	1.00031349
0.0125	0.62462462	0.52310983	0.83966678	0.98286237	1.00023592	1.00234482
0.00625	0.73322766	0.69089901	0.4790718	0.86488897	0.99083656	1.0002374
0.03125	0.86488684	0.68187077	0.80929572	0.36340171	0.87090488	0.99397895

TABLE 7.19: Erreur globale e_0 et coefficients d'efficacité η_0 en 1D du schéma $\mathbb{P}_2 + RL2 + SBDF2$ quand Δt et h varient

h \ Δt	0.1	0.05	0.025	0.0125	0.00625	0.003125
0.1	0.0110976	0.00808121	0.00742711	0.00727112	0.00723277	0.00722325
0.05	0.00346433	0.00077846	0.000304339	0.000282125	0.000286856	0.000288621
0.025	0.00358016	0.000841313	0.000200995	5.22696e-05	2.89789e-05	2.80891e-05
0.0125	0.00358839	0.000848285	0.000206229	5.05628e-05	1.25826e-05	4.3856e-06
0.00625	0.00358885	0.000848651	0.000206564	5.08332e-05	1.25864e-05	3.13211e-06
0.003125	0.00358887	0.000848675	0.000206586	5.08542e-05	1.26026e-05	3.13588e-06
Coefficients d'efficacité η_0						
0.1	2.19754455	3.13833398	3.79660575	4.03601323	4.10279144	4.11999065
0.05	0.98453428	0.74582994	0.71609176	1.04370907	1.23836988	1.30027031
0.025	1.07598262	0.98706267	0.86079229	0.66242703	0.72025361	0.91895349
0.0125	1.08626843	1.02399174	0.98391698	0.91965137	0.77153632	0.66042026
0.00625	1.08739191	1.02814446	0.99982575	0.97775521	0.94508102	0.85813584
0.03125	1.08751989	1.02869697	1.00138633	0.98524292	0.97362485	0.9572108

TABLE 7.20: Erreur globale e_1 et coefficients d'efficacité η_1 en 1D du schéma $\mathbb{P}_2 + RL2 + SBDF2$ quand Δt et h varient

h \ Δt	0.1	0.05	0.025	0.0125	0.00625	0.003125
0.1	0.0650066	0.049754	0.0506474	0.0510913	0.0512434	0.0512611
0.05	0.0303709	0.0168628	0.0230949	0.0268993	0.0268879	0.0267321
0.025	0.0323572	0.00836266	0.0037654	0.00423183	0.0059381	0.00657598
0.0125	0.0323029	0.00811868	0.00212927	0.000916554	0.000868813	0.00126012
0.00625	0.0323074	0.00807172	0.00196136	0.000517841	0.000225921	0.000195995
0.003125	0.0323078	0.00806761	0.00195393	0.000479096	0.000126649	5.58449e-05
Coefficients d'efficacité η_1						
0.1	0.46599713	0.42253928	0.4509507	0.46047642	0.46326626	0.46378249
0.05	0.53635143	0.48352115	0.78454013	0.95801608	0.96934273	0.96668682
0.025	0.90115788	0.59073978	0.43187384	0.57502534	0.84593983	0.94828951
0.0125	1.0513022	0.9044311	0.60164846	0.42049741	0.47222204	0.71806619
0.00625	1.09770981	1.05078082	0.87399182	0.58528318	0.4145948	0.42611314
0.03125	1.10993197	1.09645402	1.01745461	0.85394268	0.57258013	0.409931

TABLE 7.21: Erreur globale e_0 et coefficients d'efficacité η_0 en 1D du schéma $\mathbb{P}_2 + RL3 + SBDF3$ quand Δt et h varient

h \ Δt	0.1	0.05	0.025	0.0125	0.00625	0.003125
0.1	0.00835154	0.00731274	0.00722889	0.00722103	0.0072202	0.0072201
0.05	0.00110472	0.000314283	0.00029073	0.000289372	0.000289273	0.000289263
0.025	0.00115216	0.000133032	3.27795e-05	2.84945e-05	2.84073e-05	2.84079e-05
0.0125	0.00115845	0.000131307	1.72229e-05	4.04737e-06	3.43642e-06	3.42702e-06
0.00625	0.00115885	0.000131534	1.70025e-05	2.15507e-06	5.05203e-07	4.27011e-07
0.003125	0.00115887	0.000131552	1.69977e-05	2.12229e-06	2.67459e-07	6.34328e-08
Coefficients d'efficacité η_0						
0.1	2.96153901	3.88201194	4.09170101	4.12138075	4.12521568	4.1257007
0.05	0.85720272	0.89158298	1.23467958	1.31032422	1.32082097	1.32216382
0.025	1.04995717	0.82582407	0.74393191	0.96809758	1.0290673	1.03767844
0.0125	1.07922415	0.95732721	0.85529903	0.73484198	0.93401537	0.99316065
0.00625	1.08261243	0.98035328	0.99163648	0.85617859	0.733799	0.92848865
0.03125	1.08300547	0.9831988	1.01345091	0.99022508	0.85005848	0.73708074

TABLE 7.22: Erreur globale e_1 et coefficients d'efficacité η_1 en 1D du schéma $\mathbb{P}_2 + RL3 + SBDF3$ quand Δt et h varient

h \ Δt	0.1	0.05	0.025	0.0125	0.00625	0.003125
0.1	0.0505144	0.0505758	0.0511683	0.0512539	0.0512665	0.0512681
0.05	0.0216321	0.0254131	0.0267845	0.0267104	0.026695	0.0266928
0.025	0.0133023	0.0041224	0.0056609	0.0065501	0.00648517	0.00647738
0.0125	0.0129167	0.00173249	0.000934154	0.00139878	0.00165624	0.00166636
0.00625	0.0129223	0.00142219	0.00029977	0.000214158	0.000338219	0.000415693
0.003125	0.0129175	0.0014012	0.00020211	6.16333e-05	5.09531e-05	8.20389e-05
Coefficients d'efficacité η_1						
0.1	0.41135505	0.4514185	0.46225754	0.46373539	0.46393763	0.46396311
0.05	0.54181841	0.87143078	0.96287583	0.96605232	0.96622991	0.96624191
0.025	0.69260267	0.4882191	0.79748508	0.94514355	0.93862141	0.93784957
0.0125	0.9208744	0.53077762	0.4868555	0.79903347	0.95760218	0.96492061
0.00625	1.01497529	0.72223915	0.48053156	0.46999517	0.77815894	0.96220777
0.03125	1.04106977	0.85158106	0.67347551	0.46716668	0.45940943	0.75758519

7.2 Étude 2D

7.2.1 Convergence des schémas en temps, précision sur le problème semi-discret

Après avoir fait l'étude en 1D de $\theta_L^{\Delta t}$ et $\theta_H^{\Delta t}$, nous allons faire une étude semblable en 2D. Comme en 1D, nous allons commencer par étudier la convergence des schémas en temps.

Convergence en temps

Nous fixons le diamètre du maillage à $h = 0.025$. L'ordre de convergence d'un schéma d'intégration en temps est donné par la pente de décroissance des erreurs $\theta_L^{\Delta t}$ ou $\theta_H^{\Delta t}$ quand Δt décroît. Le principe d'étude est le même que celui du cas 1D et a été décrit en début de la section 7.1.1 . Les erreurs $\theta_L^{\Delta t}$ et $\theta_H^{\Delta t}$ pour divers Δt et les ordres de convergences correspondants à chaque méthode sont représentés dans la table (7.23) ci-dessous.

Dans ces tables, on observe les ordres 1, 2 et 3 respectivement pour les schémas combinés RL_k et SBDFk, k= 1, 2 et 3 appliqués au problème monodomaine en 2D, où le modèle ionique est le modèle de Beeler Reuter. Comme en 1D, les ordres de convergence pour les deux normes considérées sont les mêmes.

Le régime de convergence pour le schéma $RL1 + FBE$ contrairement au 1D n'est atteint qu'à partir d'un pas de temps très petit ($\Delta t \leq 0.0003125$ pour la norme $\theta_L^{\Delta t}$ et encore plus bas pour la norme $\theta_H^{\Delta t}$). C'est donc à partir du pas de temps de $\Delta t \leq 0.003125$

qu'on est certain qu'en raffinant Δt on améliore la solution avec le schéma $RL1 + FBE$. On peut d'ailleurs voir dans la table (7.23) (à gauche) que pour le schéma $RL1 + FBE$ l'erreur $\theta_H^{\Delta t}$ ne s'améliore presque pas quand $0.0125 \leq \Delta t \leq 0.05$. Cette faible précision n'était pas observée en 1D. Ceci montre donc que le problème en 2D est plus exigeant par rapport à la précision du schéma. Les schémas d'ordre élevé ne souffrent pas d'un tel problème.

Par rapport à la norme $\theta_L^{\Delta t}$, la norme $\theta_H^{\Delta t}$ est plus exigeante en terme de précision. Les schémas atteignent toujours leur régime de convergence plus rapidement en norme $\theta_L^{\Delta t}$, qu'en norme $\theta_H^{\Delta t}$. Par exemple, le régime de convergence pour le schéma $RL2 + SBDF2$ est atteint à $\Delta t \leq 0.025$ pour la norme $\theta_L^{\Delta t}$, tandis ce régime es atteint à $\Delta t \leq 0.0125$ pour la norme $\theta_H^{\Delta t}$. Le régime de convergence pour le schéma $RL3 + SBDF3$ est atteint à $\Delta t \leq 0.0125$ pour la norme $\theta_L^{\Delta t}$ tandis que ce régime est atteint à $\Delta t \leq 0.00625$ pour la norme $\theta_H^{\Delta t}$.

TABLE 7.23: Erreurs $\theta_L^{\Delta t}$, $\theta_H^{\Delta t}$ en 2D et pente k associée pour les schémas $\mathbb{P}_1 + RL1 + FBE$ (au dessus à gauche), $\mathbb{P}_1 + RL2 + SBDF2$ (au dessus à droite), $\mathbb{P}_2 + RL2 + SBDF2$ (en bas à gauche) et $\mathbb{P}_2 + RL3 + SBDF3$ (en bas à droite) à maillage fixé de diamètre $h = 0.025$

$\mathbb{P}_1 + RL1 + FBE$				
Δt	$\theta_L^{\Delta t}$	k	$\theta_H^{\Delta t}$	k
0.05	0.91	–	1.18	–
0.025	0.64	0.5	1.48	-0.33
0.0125	0.38	0.74	1.57	-0.07
6.25×10^{-3}	0.21	0.87	1.2	0.38
3.125×10^{-3}	0.11	0.95	0.71	0.75

$\mathbb{P}_1 + RL2 + SBDF2$			
$\theta_L^{\Delta t}$	k	$\theta_H^{\Delta t}$	k
0.23	–	1.34	–
0.061	1.9	0.46	1.55
0.015	2.00	0.115	1.99
3.8×10^{-3}	2.00	0.028	2.01
9.53×10^{-4}	2.00	7.1×10^{-3}	2.00

$\mathbb{P}_2 + RL2 + SBDF2$				
Δt	$\theta_L^{\Delta t}$	k	$\theta_H^{\Delta t}$	k
0.05	0.2554	–	1.431	–
0.025	0.0682	1.99	0.547	1.38
0.0125	0.0169	2.00	0.1378	1.98
6.25×10^{-3}	0.0042	2.00	0.0339	2.02
3.125×10^{-3}	0.001	2.00	0.0084	2.00

$\mathbb{P}_2 + RL3 + SBDF3$			
$\theta_L^{\Delta t}$	k	$\theta_H^{\Delta t}$	k
0.0144	–	0.1315	–
0.0024	2.56	0.0255	2.36
0.0003	2.80	0.0037	2.78
4.64×10^{-5}	2.91	0.0004	2.92
5.96×10^{-6}	2.96	6.22×10^{-5}	2.96

Nous avons montré numériquement par la table 7.23 que pour Δt suffisamment petit,

$$\theta_L^{\Delta t} \leq C_1(h)\Delta t^k \qquad \text{et} \qquad \theta_H^{\Delta t} \leq C_2(h)\Delta t^k. \tag{7.18}$$

Où $C_1(h)$ et $C_2(h)$ sont des constantes pouvant dépendre de h et de la méthode d'intégration en espace utilisée. Il reste donc à étudier la dépendance des constantes $C_1(h)$ et $C_2(h)$ par rapport à h et au schéma d'intégration en espace utilisé.

Étude des constantes C_1 et C_2

Avant de donner une approximation des constantes C_1 et C_2 de (7.18), nous aurons besoin d'étudier leur dépendance à h et à la méthode d'intégration en espace utilisée dans la simulation. Nous allons procéder exactement comme en 1D. Plus précisément, nous calculons les erreurs $\theta_L^{\Delta t}$, $\theta_H^{\Delta t}$ et les constantes $C_1(h)$, $C_2(h)$ de la formule (7.18) pour divers h. Les résultats de ces calculs sont représentés dans les tables 7.24–7.31. Le bloc supérieur de chacune des tables contient les erreurs $\theta_L^{\Delta t}$ (resp $\theta_H^{\Delta t}$) et le bloc inférieur les constantes $C_1(h)$ (resp $C_2(h)$). Les codes couleurs sont utilisés pour exprimer la proximité entre les chiffres.

À travers les codes couleurs, on peut constater que dans les blocs supérieurs des tables 7.24–7.29 et sur chaque colonne, les erreurs $\theta_L^{\Delta t}$ (resp $\theta_H^{\Delta t}$) pour les schémas d'ordre 1 et 2 sont à deux chiffres significatifs près égales, pour h et Δt assez petits. Pour le schéma $RL3 + SBDF3$ (table 7.30) l'égalité sur les colonnes entre les erreurs $\theta_L^{\Delta t}$ (resp $\theta_H^{\Delta t}$) est vraie à 3 trois chiffres (resp 1 chiffre) significatifs près. Vu l'ordre de grandeur des valeurs de ces erreurs, on peut conclure comme en 1D que leur variation par rapport à h est très faible. Dans la partie inférieure des tables, on observe que pour h et Δt assez petits,

TABLE 7.24: Erreurs $\theta_L^{\Delta t}$ et constantes $C_1(h)$, en 2D du schéma $\mathbb{P}_1 + RL1 + FBE$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.899368	0.610455	0.371937	0.206986	0.107829
0.025	0.919818	0.646416	0.386923	0.210873	0.108991
0.0125	0.918537	0.654260	0.398244	0.217383	0.112173
0.00625	0.917474	0.655420	0.401293	0.220076	0.113770
constantes $C_1(h) = \Delta t^{-1} \times \theta_L^{\Delta t}$					
0.05	17.9874	24.4182	29.755	33.1177	34.5053
0.025	18.3964	25.8566	30.9539	33.7396	34.8771
0.0125	18.3707	26.1704	31.8595	34.7813	35.8953
0.00625	18.3495	26.2168	32.1034	35.2122	36.4065

TABLE 7.25: Erreurs $\theta_H^{\Delta t}$ et constantes $C_2(h)$, en 2D du schéma $\mathbb{P}_1 + RL1 + FBE$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	1.26628	1.55108	1.40594	0.965045	0.559403
0.025	1.18056	1.48610	1.57049	1.20601	0.715316
0.0125	1.18049	1.46887	1.55361	1.24356	0.764465
0.00625	1.19039	1.49082	1.59029	1.28521	0.796513
constantes $C_2(h) = \Delta t^{-1} \times \theta_H^{\Delta t}$					
0.05	25.3255	62.0433	112.475	154.407	179.009
0.025	23.6111	59.4438	125.639	192.961	228.901
0.0125	23.6098	58.7546	124.289	198.970	244.629
0.0625	23.8079	59.6328	127.223	205.633	254.884

TABLE 7.26: Erreurs $\theta_L^{\Delta t}$ et constantes $C_1(h)$, en 2D du schéma $\mathbb{P}_1 + RL2 + SBDF2$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.246807	0.0702611	0.0178398	0.00447001	0.00111798
0.025	0.230432	0.0616598	0.0153816	0.00382676	0.00095375
0.0125	0.247801	0.0657873	0.0163627	0.00406517	0.00101263
0.00625	0.256453	0.0684750	0.0170326	0.00423165	0.00105408
constantes $C_1(h) = \Delta t^{-2} \times \theta_L^{\Delta t}$					
0.05	98.7229	112.418	114.174	114.432	114.481
0.025	92.1729	98.6557	98.4425	97.965	97.6647
0.0125	99.1202	105.26	104.721	104.068	103.693
0.00625	102.581	109.56	109.009	108.330	107.938

TABLE 7.27: Erreurs $\theta_H^{\Delta t}$ et constantes $C_2(h)$, en 2D du schéma $\mathbb{P}_1 + RL2 + SBDF2$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	1.15463	0.399095	0.103155	0.0258992	0.00648294
0.025	1.34881	0.460522	0.115211	0.0285895	0.00711618
0.0125	1.40574	0.519113	0.129952	0.0321141	0.00798129
0.00625	1.46082	0.554712	0.138740	0.0342816	0.00852093
constantes $C_2(h) = \Delta t^{-2} \times \theta_H^{\Delta t}$					
0.05	461.85	638.552	660.195	663.019	663.853
0.025	539.523	736.835	737.353	731.891	728.697
0.0125	562.295	830.581	831.694	822.121	817.285
0.00625	584.329	887.540	887.934	877.609	872.543

TABLE 7.28: Erreurs $\theta_L^{\Delta t}$ et constantes $C_1(h)$, en 2D du schéma $\mathbb{P}_2 + RL2 + SBDF2$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	96.1799	104.032	104.145	103.731	103.471
0.025	0.255464	0.0682161	0.0169762	0.00421759	0.00105060
0.0125	0.259284	0.0694015	0.0172860	0.00429321	0.00106927
0.00625	0.259647	0.0695418	0.0173225	0.00430228	0.00107153
constantes $C_1(h) = \Delta t^{-2} \times \theta_L^{\Delta t}$					
0.05	96.1799	104.032	104.145	103.731	103.471
0.025	102.186	109.146	108.648	107.970	107.581
0.0125	103.714	111.042	110.631	109.906	109.493
0.00625	103.858	111.267	110.864	110.138	109.724

TABLE 7.29: Erreurs $\theta_H^{\Delta t}$ et constantes $C_2(h)$, en 2D du schéma $\mathbb{P}_2 + RL2 + SBDF2$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	1.48139	0.525763	0.13258	0.0328799	0.00819222
0.025	1.43106	0.547053	0.137882	0.0339082	0.00842126
0.0125	1.47784	0.56560	0.144982	0.0356838	0.00884903
0.00625	1.64386	0.620948	0.158601	0.0390286	0.00967829
constantes $C_2(h) = \Delta t^{-2} \times \theta_H^{\Delta t}$					
0.05	592.555	841.221	848.510	841.726	838.883
0.025	572.425	875.284	882.444	868.050	862.337
0.0125	591.135	904.959	927.887	913.506	906.140
0.00625	657.544	993.517	1015.04	999.132	991.056

TABLE 7.30: Erreurs $\theta_L^{\Delta t}$ et constantes $C_1(h)$, en 2D du schéma $\mathbb{P}_2 + RL3 + SBDF3$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.0115763	0.0019115	0.000276252	3.70407e-05	4.78052e-06
0.025	0.0144338	0.00243764	0.000348820	4.64090e-05	5.96422e-06
0.0125	0.0138442	0.00243115	0.000352932	4.72271e-05	6.08511e-06
0.00625	0.0137331	0.00242488	0.000352751	4.72477e-05	6.09003e-06
constantes $C_1(h) = \Delta t^{-3} \times \theta_L^{\Delta t}$					
0.05	92.6106	122.336	141.441	151.719	156.648
0.025	115.470	156.009	178.596	190.091	195.436
0.0125	110.754	155.593	180.701	193.442	199.397
0.00625	109.865	155.192	180.608	193.526	199.558

TABLE 7.31: Erreurs $\theta_H^{\Delta t}$ et constantes $C_2(h)$, en 2D du schéma $\mathbb{P}_2 + RL3 + SBDF3$ quand h varie

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.108075	0.0190668	0.00269321	0.000356225	4.57979e-05
0.025	0.131558	0.0255386	0.00371047	0.000487444	6.22807e-05
0.0125	0.143689	0.0275393	0.00403901	0.000537812	6.88286e-05
0.00625	0.177523	0.0321080	0.00463722	0.000613770	7.85623e-05
constantes $C_2(h) = \Delta t^{-2} \times \theta_H^{\Delta t}$					
0.05	864.598	1220.28	1378.92	1459.10	1500.71
0.025	1052.46	1634.47	1899.76	1996.57	2040.81
0.0125	1149.51	1762.51	2067.97	2202.88	2255.38
0.00625	1420.18	2054.91	2374.26	2514.00	2574.33

les valeurs de $C_1(h)$ sont égales au moins à deux chiffres significatifs près, tandis que les valeurs des constantes $C_2(h)$ sont égales au moins à 1 chiffres significatif près. On peut conclure que la dépendance en h des constantes $C_1(h)$ est faible. En ce qui concerne la dépendance en h des constantes $C_2(h)$, l'interprétation peut être la même mais il faudra considérer des discrétisations en espace plus fines que celles qui sont dans les tables.

En se basant sur la faible dépendance à h des constantes observées dans les tables 7.24–7.29, on peut se servir des constantes obtenues pour h assez petit pour donner des approximations de C_1 et C_2. En utilisant les tables 7.24 et 7.25 pour le schéma $RL1+FBE$, 7.26–7.29 pour le schéma $RL2 + SBDF2$ et 7.30–7.31 pour le schéma $RL3 + SBDF3$, nous donnons ces approximations dans la table 7.32.

TABLE 7.32: Approximations des constantes C_1 et C_2 en 2D, pour les schémas $RL1 + FBE$, $RL2 + SBDF2$ et $RL3 + SBDF3$.

Constantes \ Schémas	$RL1 + FBE$	$RL2 + SBDF2$	$RL3 + SBDF3$
C_1	36.4	109.7	199.5
C_2	254.9	991	2574.3

7.2.2 Étude de la convergence en espace

L'étude des erreurs ρ_L^h et ρ_H^h se fait exactement comme en 1D. La méthode d'intégration en temps est toujours $RL4 + SBDF4$ et le pas de temps est Δt_{ref}. Les erreurs ρ_L^h (resp ρ_H^h) pour divers h et la pente de décroissance associées sont représentées dans la table 7.33 (resp 7.34), pour les méthodes éléments finis Lagrange $\mathbb{P}_1$ et $\mathbb{P}_2$. La pente de décroissance permet d'évaluer comme en 1D les constantes C_3 et C_4 telles que,

$$\rho_L^h \leq C_3 h^{r+1} \qquad \text{et} \qquad \rho_H^h \leq C_4 h^r. \tag{7.19}$$

où h désigne le diamètre assez petit d'un maillage et r la pente de décroissance en norme ρ_H^h .

Dans la table 7.33 à gauche, on observe que la pente de décroissance de ρ_L^h est 1.97 donc presque 2 quand h est assez petit. De même on observe une pente de 2 dans la table 7.34 (à droite) pour h assez petit. Ceci est cohérent avec les résultats classiques concernant les méthodes $\mathbb{P}_1$ et $\mathbb{P}_2$ en norme L^2 et H^1 respectivement. Cependant on observe aussi une pente en norme ρ_L^h au dessus de 3 en $\mathbb{P}_2$ (table 7.33 à droite) et une pente en norme ρ_H^h au dessus de 1 (table 7.34 à gauche). Ceci n'est sûrement pas une super convergence. En effet on remarque quand même que ces pentes décroissent vers les valeurs attendues au fur et à mesure que h devient petit. Ceci a un effet sur le calcul des constantes C_3 et C_4. Comme on peut voir dans les tables 7.33 et 7.34, ces constantes sont assez proches dans les cas où on observe bien la pente attendue. Dans le cas où on a plus que la pente attendue, les constantes ne sont pas assez proches. Mais on s'attend à ce qu'elle se stabilisent pour h

très petit. Nous allons prendre pour approximation des constantes C_3 et C_4, les valeurs que l'on a pour le plus petit h. Ces valeurs sont collectées dans la table 7.35.

TABLE 7.33: Erreurs ρ_L^h, constante C_3 en 2D et pente r associée pour les schémas $\mathbb{P}_1$ (à gauche) et $\mathbb{P}_2$ (à droite)

$\mathbb{P}_1$				$\mathbb{P}_2$		
h	ρ_L^h	r	$h^{-2} \times \rho_L^h$	ρ_L^h	r	$h^{-3} \times \rho_L^h$
0.05	0.4781	–	191.24	0.1075	–	860.68
0.025	0.2902	0.71	464.45	0.0168	2.67	1081.47
0.0125	0.099 02	1.55	633.77	0.0014	3.53	746.92
6.25×10^{-3}	0.026 68	1.89	683.03	0.0001	3.81	425.64
3.125×10^{-3}	0.0067	1.97	695.72	7.56×10^{-6}	3.77	247.95

TABLE 7.34: Erreurs ρ_H^h et C_4 en 2D et pente r associée pour les schémas $\mathbb{P}_1$ (à gauche) et $\mathbb{P}_2$ (à droite)

$\mathbb{P}_1$				$\mathbb{P}_2$		
h	ρ_H^h	r	$h \times \rho_H^h$	ρ_H^h	r	$h^{-2} \times \rho_H^h$
0.05	3.627	–	72.54	1.641	–	656.69
0.025	2.251	0.68	90.07	0.4888	1.74	782.22
0.0125	1.0139	1.15	81.11	0.1323	1.88	847.03
6.25×10^{-3}	0.3332	1.6	53.32	0.0328	2.01	840.48
3.125×10^{-3}	0.1247	1.41	39.93	0.0077	2.08	793.73

TABLE 7.35: Approximations des constantes C_3 et C_4 en 2D, pour les méthodes d'éléments finis Lagrange $\mathbb{P}_1$ et $\mathbb{P}_2$.

Constantes \ Schémas	$\mathbb{P}_1$	$\mathbb{P}_2$
C_3	695.7	247.9
C_4	39.9	793.7

7.2.3 Convergence globale

Comme dans le cas 1D, nous allons utiliser les résultats numériques obtenus dans les deux sous-sections précédentes pour majorer les erreurs e_0 et e_1 en fonction de h et Δt. En utilisant les inégalités (7.5), (7.6), (7.18), (7.19) et les constantes C_i, $i = 1, 2, 3, 4$ calculées précédemment pour chaque schéma numérique, on a approximativement les inégalités pratique suivantes,

— $P_1 + RL1 + FBE$

$$e_0 \leq 695.7 \times h^2 + 36.5 \times \Delta t. \tag{7.20}$$

$$e_1 \leq 39.9 \times h + 254.9 \times \Delta t. \tag{7.21}$$

— $P_1 + RL2 + SBDF2$

$$e_0 \leq 695.7 \times h^2 + 109.7 \times \Delta t^2. \tag{7.22}$$
$$e_1 \leq 39.9 \times h + 991 \times \Delta t^2. \tag{7.23}$$

— $P_2 + RL2 + SBDF2$

$$e_0 \leq 247.9 \times h^3 + 109.7 \times \Delta t^2. \tag{7.24}$$
$$e_1 \leq 793.7 \times h^2 + 991 \times \Delta t^2. \tag{7.25}$$

— $P_2 + RL3 + SBDF3$

$$e_0 \leq 247.9 \times h^3 + 199.5 \times \Delta t^3. \tag{7.26}$$
$$e_1 \leq 793.7 \times h^2 + 2574.3 \times \Delta t^3. \tag{7.27}$$

Pour vérifier ces majorations, nous aurons besoin comme en 1D, des valeurs des erreurs e_0 (resp e_1) et la combinaison $C_1 h^{r+1} + C_2 \Delta t^k$ (resp $C_3 h^r + C_4 \Delta t^k$). Nous aurons ensuite besoin des coefficients d'efficacité η_0 et η_1 définis par la formule (7.17).

Les erreurs e_0 (resp e_1) ainsi que le coefficient d'efficacité η_0 (resp : η_1) correspondant pour divers $(h, \Delta t)$ sont présentés dans les tables 7.36–7.43. Le bloc supérieur de chacune de ces tables contient les erreurs, et le bloc inférieur les coefficients d'efficacité correspondants. Les coefficients d'efficacité strictement supérieurs à 1 et inférieur à 1.19 ($1 < \eta_l \leq 1.19\,;\, l = 0, 1$) ont été coloriés en bleu et les coefficients supérieurs ou égaux à 2 ($\eta_l \geq 2\,;\, l = 0, 1$) en rouge. On rappelle qu'avoir un coefficient d'efficacité compris entre 1 et 1.19 veut dire que l'on a presque l'égalité entre l'erreur e_0 (resp : e_1) et la valeur de $C_1 h^{r+1} + C_2 \Delta t^k$ (resp $C_3 h^r + C_4 \Delta t^k$) tandis qu'avoir une erreur supérieur à 1.2 veut dire que la valeur de l'erreur e_0 (resp : e_1) est très grande devant celle de de $C_1 h^{r+1} + C_2 \Delta t^k$ (resp $C_3 h^r + C_4 \Delta t^k$).

Les coefficients d'efficacité donnés par les tables 7.36–7.43 sont inférieurs à 1 à partir d'un certain $(h, \Delta t)$. Ceci veut dire que les inégalités (7.20)–(7.27) sont vraies lorsque Δt et h deviennent assez petits.

En dehors des cas (7.24) et (7.25) représentant respectivement les majorations des erreurs e_0 et e_1 pour le schéma $P_2 + RL2 + SBDF2$, les majorations que l'on a obtenues en 1D sont meilleures par rapport à celles obtenues en 2D. En effet, dans le cas 1D, les coefficients d'efficacité étaient en majorité comprise entre 0.8 et 1.15. Mais comme on peut voir dans les tables 7.36–7.43, ces coefficients sont comprises entre 0.2 et 0.8 dans le cas 2D.
Le cas 2D du schéma $P_2 + RL2 + SBDF2$ qui fait l'exception, a la majorité des coefficients d'efficacité comprises entre 0.8 et 1.05 à partir de $(h, \Delta t) = 0.0125, 0.025)$ (voir table 7.40 et 7.41). Ceci nous permet de dire que pour le schéma $P_2 + RL2 + SBDF2$, les inégalités

(7.24) et (7.25) peuvent nous permettre de choisir efficacement des valeurs de h et Δt pour obtenir une erreur e_0 ou e_1 donnée.

Exemple pratique : si on considère l'inégalité (7.24) et on veut choisir efficacement h et Δt pour obtenir une erreur e_0 de l'ordre de 10^{-2}. Il suffit d'évaluer le terme de droite ($247.9\times h^3+109.7\times\Delta t^2$) de l'inégalité (7.24) pour divers (h,Δt), et de prendre le plus grand $h \leq 0.0125$ et le plus grand $\Delta t \leq 0.025$ pour lesquelles la formule $247.9 \times h^3 + 109.7 \times \Delta t^2$ donne une valeur proche de 10^{-2}. Quand nous faisons ces calculs, on constate que pour $h = \Delta t = 0.0125$, on obtient $247.9 \times h^3 + 109.7 \times \Delta t^2 = 0.01762$. La table 7.40 nous donne à peu près la même valeur $e_0 = 0.0159$ pour $h = \Delta t = 0.0125$. On aurait pu utiliser $h = 0.003125$ et $\Delta t = 0.0125$ pour obtenir le même résultat (voir table 7.40) mais avec un temps de calcul plus élevé (ie : non efficace).

TABLE 7.36: Erreurs globale e_0 et coefficients d'efficacité η_0 en 2D du schéma $\mathbb{P}_1 + RL1 + FBE$ quand Δt et h varient

h \ Δt	0.1	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.493725	0.257386	0.0927938	0.08394	0.138295	0.163507
0.025	0.535516	0.33132	0.156179	0.0521047	0.0474184	0.0807897
0.0125	0.55481	0.368588	0.215843	0.106076	0.0375798	0.0119149
0.00625	0.56022	0.379091	0.232648	0.128783	0.0638316	0.027753
0.003125	0.561605	0.381805	0.237016	0.134692	0.0705342	0.0351401
Coefficients d'efficacité η_0						
0.05	0.09178324	0.07231467	0.03502644	0.03825453	0.07031651	0.08823907
0.025	0.13142102	0.14693905	0.11613441	0.05855694	0.07159521	0.14727529
0.0125	0.14800052	0.19110665	0.21188018	0.18817706	0.11177707	0.0535614
0.00625	0.15276606	0.20522736	0.24824372	0.26708724	0.25063867	0.19693345
0.03125	0.15399965	0.20900277	0.25852701	0.29167121	0.30105003	0.29151276

TABLE 7.37: Erreur globale e_1 et coefficients d'efficacité η_1 en 2D du schéma $\mathbb{P}_1 + RL1 + FBE$ quand Δt et h varient

h \ Δt	0.1	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	4.61968	3.74388	1.56133	1.86457	2.57514	2.86192
0.025	4.71564	4.37836	2.84233	0.963101	1.14121	1.73583
0.0125	4.74246	4.56099	3.59311	2.19999	0.900566	0.320174
0.00625	4.74987	4.60349	3.776	2.56689	1.43915	0.66357
0.003125	4.7518	4.61385	3.81995	2.65296	1.5741	0.825651
Coefficients d'efficacité η_1						
0.05	0.18052853	0.29147161	0.2412345	0.56742848	1.5211637	3.19299352
0.025	0.18481879	0.34286464	0.44429195	0.29992051	0.70529414	2.11300061
0.0125	0.18600629	0.35769008	0.56329513	0.68911535	0.5630792	0.39882317
0.00625	0.18633109	0.36115552	0.59240121	0.80522087	0.90246743	0.83141519
0.03125	0.86547591	0.36200149	0.59940627	0.83252589	0.98781646	1.0360095

TABLE 7.38: Erreur globale e_0 et coefficients d'efficacité η_0 en 2D du schéma $\mathbb{P}_1 + RL2 + SBDF2$ quand Δt et h varient

h \ Δt	0.1	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.0915063	0.124436	0.171105	0.183267	0.186335	0.187105
0.025	0.156563	0.0393167	0.0942098	0.10775	0.111042	0.111857
0.0125	0.228289	0.0497679	0.015846	0.0319073	0.0358167	0.0367828
0.00625	0.249392	0.0779415	0.0129637	0.00434457	0.00861443	0.00967744
0.003125	0.254727	0.0853666	0.0205026	0.00325425	0.00111314	0.00218607
Coefficients d'efficacité η_0						
0.05	0.03226313	0.06180084	0.09464754	0.10434296	0.10687195	0.10751176
0.025	0.10220768	0.05544885	0.1871563	0.23840968	0.2528868	0.25662115
0.0125	0.18934097	0.1299582	0.08939127	0.25354696	0.31699482	0.33507627
0.00625	0.22184431	0.25857609	0.1354077	0.09803524	0.2738135	0.34259978
0.03125	0.23077405	0.30374822	0.27207492	0.13596442	0.10047206	0.27794101

TABLE 7.39: Erreur globale e_1 et coefficients d'efficacité η_1 en 2D du schéma $\mathbb{P}_1 + RL2 + SBDF2$ quand Δt et h varient

h \ Δt	0.1	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	1.54111	2.43807	2.95461	3.07326	3.10212	3.10927
0.025	2.89589	1.01893	1.97593	2.1781	2.22875	2.24117
0.0125	3.75278	1.17372	0.455893	0.774018	0.853256	0.872881
0.00625	3.9593	1.74201	0.338399	0.171046	0.244567	0.265563
0.003125	4.0078	1.88439	0.488395	0.103554	0.0767675	0.0861545
Coefficients d'efficacité η_1						
0.05	0.12945065	0.54512465	1.13014009	1.42952715	1.52534952	1.5510074
0.025	0.2654953	0.29321727	1.22206726	1.8901478	2.15086515	2.22519813
0.0125	0.36054089	0.39436203	0.4077299	1.18424939	1.58756839	1.71682414
0.00625	0.38971886	0.63883016	0.38952403	0.42315207	0.84893765	1.02513106
0.03125	0.3993946	0.72415612	0.65638975	0.3704559	0.46981783	0.64119639

TABLE 7.40: Erreur globale e_0 et coefficients d'efficacité η_0 en 2D du schéma $\mathbb{P}_2 + RL2 + SBDF2$ quand Δt et h varient

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.147941	0.0511166	0.0930705	0.103965	0.106682
0.025	0.242158	0.0527974	0.00515607	0.0129805	0.0159108
0.0125	0.258204	0.0680889	0.0159612	0.003022	0.000581524
0.00625	0.259573	0.0694539	0.0172333	0.0042132	0.000983618
0.003125	0.259666	0.0695468	0.0173199	0.00429763	0.00106636
Coefficients d'efficacité η_0					
0.05	0.48467505	0.51347664	1.93380669	2.94746696	3.32769889
0.025	0.87068533	0.72888403	0.24536287	1.59102247	3.21772877
0.0125	0.9398321	0.98612866	0.90561028	0.63363198	0.37385982
0.00625	0.94481509	1.00589789	0.97778676	0.88339452	0.6323647
0.03125	0.94679614	1.014244	1.0100133	1.0011438	0.98842286

TABLE 7.41: Erreur globale e_1 et coefficients d'efficacité η_1 en 2D du schéma $\mathbb{P}_2 + RL2 + SBDF2$ quand Δt et h varient

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	1.88452	1.27604	1.55183	1.62011	1.63643
0.025	1.75434	0.74208	0.433273	0.463308	0.482682
0.0125	1.78352	0.667237	0.21917	0.143204	0.126137
0.00625	1.78572	0.666361	0.173634	0.0584871	0.0423914
0.003125	1.78587	0.66629	0.169702	0.0429283	0.0147
Coefficients d'efficacité η_1					
0.05	0.42232506	0.4900072	0.72529189	0.80066285	0.82050103
0.025	0.58995439	0.66520701	0.66551798	0.86616069	0.95417113
0.0125	0.68556135	0.89752134	0.78586362	0.87985947	0.94325943
0.00625	0.71186432	1.0245612	0.93424208	0.83885354	1.0418278
0.03125	0.71858682	1.0624467	1.04369905	0.92390711	0.84334156

TABLE 7.42: Erreur globale e_0 et coefficients d'efficacité η_0 en 2D du schéma $\mathbb{P}_2 + RL3 + SBDF3$ quand Δt et h varient

h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	0.0417968	0.0393042	0.0388739	0.0388102	0.0388015
0.025	0.010829	0.00709898	0.00643199	0.00633457	0.00632152
0.0125	0.00473488	0.00131521	0.000658407	0.000560376	0.000547117
0.00625	0.00417407	0.000805229	0.000150346	$5.200\,44 \times 10^{-5}$	$3.878\,52 \times 10^{-5}$
0.003125	0.00413766	0.000770837	0.000116184	$1.754\,85 \times 10^{-5}$	$4.204\,02 \times 10^{-6}$
Coefficients d'efficacité η_0					
0.05	0.74737237	1.15245733	1.23892387	1.25048129	1.25192016
0.025	0.37586417	1.01549932	1.50876242	1.61508003	1.62945511
0.0125	0.18625362	0.36519713	0.75347264	1.05157913	1.11595039
0.00625	0.16697602	0.25339994	0.33397605	0.47614356	0.58227293
0.03125	0.16587092	0.24668758	0.29250019	0.31185691	0.30791914

TABLE 7.43: Erreur globale e_1 et coefficients d'efficacité η_1 en 2D du schéma $\mathbb{P}_2 + RL3 + SBDF3$ quand Δt et h varient

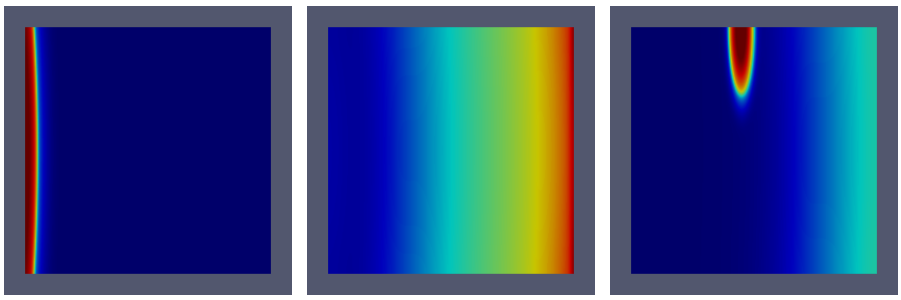
h \ Δt	0.05	0.025	0.0125	0.006 25	0.003 125
0.05	1.00125	0.936139	0.925351	0.923824	0.923621
0.025	0.285572	0.190839	0.17481	0.17253	0.172228
0.0125	0.135046	0.0410937	0.0289424	0.0280243	0.027925
0.00625	0.122881	0.023254	0.0073216	0.00670047	0.00666552
0.003125	0.122053	0.0217997	0.00354104	0.0016564	0.00161813
Coefficients d'efficacité η_1					
0.05	0.43418635	0.4624111	0.46516929	0.46543101	0.4654577
0.025	0.34917405	0.35585307	0.34885919	0.34735883	0.34713515
0.0125	0.30292744	0.25020662	0.22428397	0.22483451	0.22503068
0.00625	0.34831063	0.32647576	0.20319806	0.211823	0.21444628
0.03125	0.37037556	0.45440267	0.27710032	0.1976736	0.20666987

7.3 Précision sur des fronts de spirales

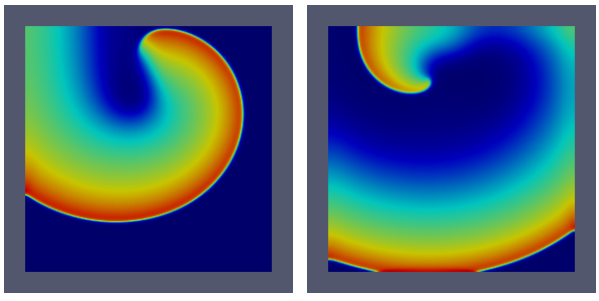
Nous allons évaluer l'intérêt d'utiliser des schémas d'ordre élevés. Pour que cette étude soit pertinente, nous la ferons sur des spirales. Pour rendre possible l'observabilité des spirales, nous travaillerons dans le cas 2D sur le carré $[0,4] \times [0,4]$ et dans le cas 3D sur le cube $[0,4] \times [0,4] \times [0,4]$. Pour pouvoir observer plusieurs tours de spirales, nous fixerons le temps final T à $600ms$. L'intérêt d'utiliser des spirales pour une telle étude est la possibilité d'avoir à tout instant une zone du domaine activée et très variable. Avoir un tel comportement favorise l'accumulation des erreurs au cours du temps de simulation. L'étude de la variation des erreur globales pour $T = 600ms$ nécessite un calcul assez coûteux. Pour cette raison, nous ferons une étude quantitative en 2D, mais en 3D nous ferons une étude qualitative. Pour le 2D, nous allons évaluer l'erreur des spirales calculées par rapport à une spirale de référence calculée par le schéma $\mathbb{P}_2 + RL4 + SBDF4$ et sur une discrétisation, $h_{ref} = \Delta t = 0.1/2^4$. Nous allons nous intéresser aux erreurs globales e_0 et e_1. Nous commençons par décrire brièvement comment nous construisons nos spirales.

7.3.1 Construction d'une spirale

Nous faisons partir à $t_{s_1} = 1ms$ un front d'onde du potentiel membranaire, instancié en stimulant pendant $2ms$ une petite partie du domaine du côté $\{x = 0\}$ (voir 7.5a pour le 2D et 7.5a pour le 3D). Ce front se propage pendant plusieurs dizaines de millisecondes



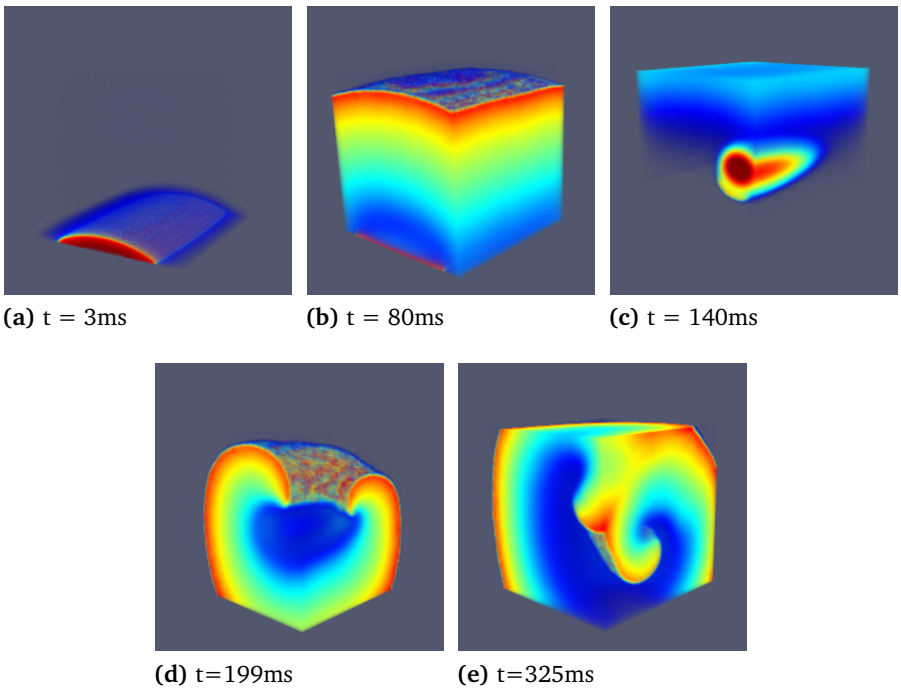
(a) t = 1ms **(b)** t = 90ms **(c)** t = 130ms



(d) t=196ms **(e)** t=333ms

FIGURE 7.5: Illustration des étapes de construction d'une spirale en 2D.

puis à $t_{s_2} = 130ms$, nous initialisons un autre front en stimulant une zone du domaine



(a) t = 3ms **(b)** t = 80ms **(c)** t = 140ms

(d) t=199ms **(e)** t=325ms

FIGURE 7.6: Illustration des étapes de construction d'une spirale en 3D.

limitée par une demie ellipse (voir 7.5c pour le 2D et 7.6c pour le 3D). La propagation de ce dernier front produit une spirale qui tourne tout au long du temps restant. Le temps final $T = 600ms$ nous permet d'observer au moins quatre tours. Les erreurs que l'on peut observer pour chaque schéma peuvent donc nous donner des indications sur l'ordre des erreurs que l'on aurait si l'on utilisait ces schémas pour faire une simulation de plusieurs battements de cœur.

7.3.2 Erreurs e_0 et e_1 sur les fronts de spirales en 2D

Les calculs pour des simulations de front de spirales sont très chers en temps de calcul. On ne saurait donc utiliser des discrétisations très fines pour de telles simulations. C'est dans l'idée de diminuer ce temps de calcul que nous avons calculé en 2D des erreurs sur des pas de temps Δt supérieurs ou égaux à $0.025ms$ et des maillages de diamètres h supérieurs ou égaux à 0.0125. Nous avons choisi des discrétisations vérifiant $h = \Delta t/2$ car c'est celle qui donnait au mieux les pentes de convergences attendues. Cependant on serait arrivé aux mêmes conclusions en faisant le choix $h = \Delta t/2$ ou $h = 2\Delta t$. Les erreurs globales e_0 et e_1 sur les spirales pour les schémas $\mathbb{P}_r + RL_k + SBDF_k$, $(r, k) \in \{(1, 1); (1, 2); (2, 2); (2, 3)\}$ ont été calculées et représentées dans les tables 7.44 et 7.45 respectivement. Dans la table 7.44 on remarque que pour une discrétisation assez grossière $(h, \Delta t) = (0.05, 0.1)$, les schémas $\mathbb{P}_1 + RL_1 + FBE$ et $\mathbb{P}_1 + RL_2 + SBDF2$ ont à peu près la même précision qui est au dessus de 10%. On observe dans la même table et pour le même schéma que pour être en dessous de 10% d'erreur, il faut utiliser des discrétisations beaucoup plus fines contrairement aux schémas $\mathbb{P}_2 + RL_2 + SBDF_2$ et $\mathbb{P}_2 + RL_3 + SBDF_3$, pour lesquels on

a des erreurs inférieures à 10% dès la discrétisation $(h, \Delta t) = (0.05, 0.1)$. Pour obtenir des erreurs de l'ordre de 2% pour des discrétisations que nous avons choisies, on est obligé d'utiliser un schéma d'ordre au moins 2 en temps et en espace. Pour observer une réduction considérable des erreur, on est obligé d'utiliser un schéma d'ordre global 3. Il est intéressant d'observer aussi que dès la discrétisation $(h,\ dt) = (0.05, 0.1)$, le schéma $\mathbb{P}_2{+}RL_3{+}SBDF_3$ produit une erreur proche de 1%. Ceci fait de celui-ci un schéma efficace pour simuler à un faible coût plusieurs battements de cœur tout en gardant une bonne précision. Dans la table 7.45, on observe une erreur e_1 pour le schéma $\mathbb{P}_1 + RL_1 + FBE$

TABLE 7.44: Erreur e_0 en 2D sur un front de spirale

		$\mathbb{P}_1$		$\mathbb{P}_2$	
h	Δt	$RL1 + FBE$	$RL2 + SBDF2$	$RL2 + SBDF2$	$RL3 + SBDF3$
0.05	0.1	0.141	0.113	0.071	0.013
0.025	0.05	0.097	0.028	0.020	9.72E-3
0.0125	0.025	0.070	0.022	5.96E-3	1.35E-3

qui ne varie presque pas et reste autour de 100% pour les discrétisations assez grossières $(h, \Delta t)$ que nous avons choisies. Ceci décrit donc quantitativement la mauvaise qualité des résultats que l'on peut obtenir en utilisant ce schéma pour des simulations de plusieurs potentiels d'actions consécutifs. Ceci peut s'améliorer très légèrement en utilisant le schéma $\mathbb{P}_1 + RL_2 + SBDF_2$, mais reste très insatisfaisant. Pour obtenir des erreurs inférieur à 50%, on est obligé d'utiliser les schémas $\mathbb{P}_2 + RL_2 + SBDF_2$ ou $\mathbb{P}_2 + RL_3 + SBDF_3$. Quand on observe l'ordre de grandeur des erreurs e_0 et e_1, on remarque que les erreurs e_1 sont au moins 10 fois plus grandes que les erreurs e_0. Ceci peut s'expliquer par le fait que la norme e_1 dépend des dérivées en espace qui en général réduisent l'ordre des schémas d'intégration en espace par rapport à la norme L^2.

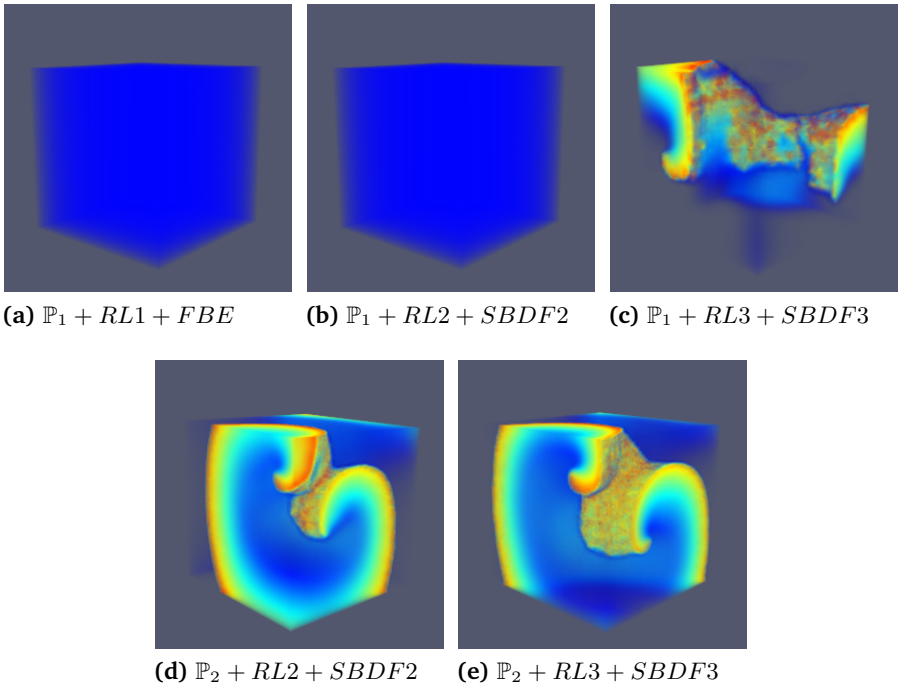
TABLE 7.45: Erreur e_1 en 2D sur un front de spirale

		$\mathbb{P}_1$		$\mathbb{P}_2$	
h	Δt	$RL1 + FBE$	$RL2 + SBDF2$	$RL2 + SBDF2$	$RL3 + SBDF3$
0.05	0.1	0.883	0.948	1.02	0.453
0.025	0.05	0.981	0.817	0.679	0.400
0.0125	0.025	1.02	0.745	0.253	0.065

7.3.3 Étude qualitative des erreurs sur les fronts de spirales en 3D

Comme nous l'avons précédemment souligné, on ne saurait faire une étude quantitative en 3D comme nous l'avons fait en 2D. Nous nous proposons donc d'observer visuellement les différences entre les spirales selon l'ordre du schéma et la finesse de la discrétisation utilisée. Après $600ms$ de simulation, on va s'intéresser à la position du front de spirale à $t = 600ms$. Nous avons utilisé le principe décrit en section 7.3 pour construire des fronts de spirales pour diverses discrétisations. Nous avons représenté les résultats obtenus sur les figures 7.7-7.8. Sur les figures 7.7a et 7.7b, on constate que le front de

spirale n'a pas pu se propager. Ceci s'explique d'une part du fait que la discrétisation est très grossière, et d'autre part de la faible précision des schémas $\mathbb{P}_1 + RL1 + FBE$ et $\mathbb{P}_1 + RL2 + SBDF2$ pour cette discrétisation. On peut faiblement améliorer la situation en augmentant l'ordre du schéma d'intégration en temps. Notamment en utilisant le schéma $RL3 + SBDF3$, on peut voir dans la figure 7.7c un front de spirale même si celui-ci est très loin d'avoir la bonne vitesse de propagation. Comme on peut observer sur les figures 7.7d et 7.7e on peut apporter une amélioration considérable en utilisant l'ordre élevé.

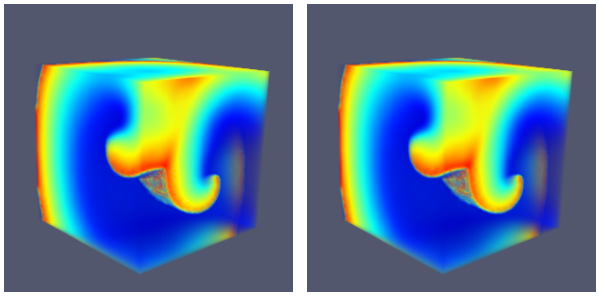


(a) $\mathbb{P}_1 + RL1 + FBE$ **(b)** $\mathbb{P}_1 + RL2 + SBDF2$ **(c)** $\mathbb{P}_1 + RL3 + SBDF3$

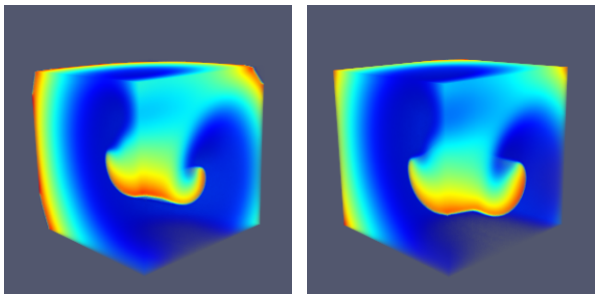
(d) $\mathbb{P}_2 + RL2 + SBDF2$ **(e)** $\mathbb{P}_2 + RL3 + SBDF3$

FIGURE 7.7: Spirale en 3D à l'instant $t = 600ms$ pour un maillage de diamètre $h = 0.1$, un pas de temps $\Delta t = 0.1$. Les schémas considérés sont : $\mathbb{P}_1 + RL1 + FBE$, $\mathbb{P}_1 + RL2 + SBDF2$, $\mathbb{P}_1 + RL2 + SBDF2$, $\mathbb{P}_2 + RL2 + SBDF2$ et $\mathbb{P}_2 + RL3 + SBDF3$

Si on raffine le maillage par exemple en prenant $h = 0.05$ tout en gardant le même pas de temps comme sur la figure 7.8, on obtient un front plus avancé par rapport à celui précédemment observé. Les fronts de spirales obtenus avec de l'ordre élevé (voir figure 7.8c et 7.8d) avancent toujours plus rapidement par rapport à ceux calculés avec du $\mathbb{P}_1$ (voir figure 7.8a et 7.8b). Il y a une légère différence entre le $\mathbb{P}_2 + RL2 + SBDF2$ et $\mathbb{P}_2 + RL3 + SBDF3$, mais qui visuellement n'est pas très importante. Ce qui n'était pas le cas pour la discrétisation $h = \Delta t = 0.1$ (voir 7.7d et 7.7e), où cette différence était visuellement observable.



(a) $\mathbb{P}_1 + RL1 + FBE$ **(b)** $\mathbb{P}_1 + RL2 + SBDF2$



(c) $\mathbb{P}_2 + RL2 + SBDF2$ **(d)** $\mathbb{P}_2 + RL3 + SBDF3$

FIGURE 7.8: Spirale en 3D à l'instant $t = 600ms$ pour un maillage de diamètre $h = 0.05$, un pas de temps $\Delta t = 0.1$. Les schémas considérés sont : $\mathbb{P}_1 + RL1 + FBE$, $\mathbb{P}_1 + RL2 + SBDF2$, $\mathbb{P}_1 + RL2 + SBDF2$, $\mathbb{P}_2 + RL2 + SBDF2$ et $\mathbb{P}_2 + RL3 + SBDF3$

Bibliographie

[1] G.W. BEELER et H. REUTER. „Reconstruction of the Action Potential of Ventricular Myocardial Fibres“. English. In : *J. Physiol.* 268 (1977), p. 177–210 (cf. p. 137).

8 Conclusions

Cette thèse avait pour objectif principal le développement des schémas numériques d'ordre élevé en temps, adaptés à la résolution des équations différentielles rencontrées dans les modèles mathématiques en cardiologie. Ceci était dans le but d'améliorer la qualité des résultats obtenus par des simulations numériques en cardiologie, tout en gardant un coût acceptable. Ces simulations qui étaient jusqu'ici faites par des schémas d'ordre 1 ou 2, sont désormais possibles par ce travail, d'être faites à moindre coût par des nouveaux schémas d'ordre élevé adaptés aux modèles physiologiques. Ces nouveaux schémas permettront de faire des simulations réalistes en des durées acceptables. L'étude bibliographique sur les méthodes d'intégration numérique en temps existantes nous a d'abord permis d'observer leurs limites et ensuite apporter des améliorations par rapport à celles-ci. Nous nous sommes intéressés pour commencer à la résolution des modèles de la membrane qui posent une grande difficulté de résolution numérique à cause de la non-linéarité et la de raideur de ses équations. Après avoir proposé deux nouvelles familles de schémas à savoir les schémas IEAB et RL, nous avons fait une étude théorique de leurs propriétés numériques. Cette étude nous a permis d'énoncer des résultats de convergence que nous avons ensuite vérifiés numériquement sur des cas tests choisis dans le contexte de la cardiologie. Le schéma EAB a été établi comme étant un cas particulier du schéma IEAB. Une étude de la stabilité au sens de Dahlquist nous a permis de construire les domaines de stabilités des schémas que nous avons proposés. Ces domaines nous ont permis d'observer que les schémas EAB et RL étaient capables de faire face à la raideur des équations de la membrane cellulaire. Cette observation a été confirmée par la grandeur des pas de temps critiques qui sont du même ordre que ceux des schémas implicites CN et BDF. Sous un critère prenant en compte le coût et la précision, nous avons fait une comparaison entre les schémas EAB, RL et quelques schémas classiques utilisés pour les problèmes de cardiologie. Nous avons trouvé que le schéma RL3 offrait une bonne alternative pour faire des simulations à des pas de temps assez grands, tout en restant précis. Par les schémas que nous avons proposés et considérés, nous avons illustré le calcul sans dégradation de la précision du schéma utilisé, des temps d'activation, de récupération et la durée du potentiel d'action. Suite à cela nous avons aussi observé que les pentes de convergences sur ces valeurs physiologiques pouvaient indiquer l'ordre du schéma numérique utilisé. Les schémas que nous avons proposés entrent dans la classe des schémas exponentiels multi-pas et peuvent être utilisés dans d'autres domaines que l'électrophysiologie cardiaque.

Après avoir fait ce grand travail sur les équations différentielles ordinaires, nous avons montré comment il faut utiliser les schémas que nous avons introduits pour la

résolution des EDP en cardiologie, en particulier sur le modèle monodomaine couplé au modèle ionique de Beeler et Reuter. Pour cela nous avons eu besoin de la méthode des éléments finis de Lagrange $\mathbb{P}_1$ et $\mathbb{P}_2$ pour l'intégration en espace et le schéma SBDF pour l'intégration en temps. Nous avons énoncé et prouvé des résultats sur la convergence des schémas que nous avons nommés $\mathbb{P}_r + RL1 + FBE$ et $\mathbb{P}_r + RL2 + SBDF2$. Une étude numérique sur les constantes apparaissant dans les résultat de convergence a été faite en 1D et 2D pour les combinaisons de schémas $\mathbb{P}_1 + RL1 + FBE$, $\mathbb{P}_1 + RL2 + SBDF2$, $\mathbb{P}_2 + RL2 + SBDF2$ et $\mathbb{P}_1 + RL3 + SBDF2$. Cette étude nous a permis de vérifier que l'on obtenait numériquement les ordre attendus pour chaque schéma. Elle s'est achevée par l'évaluation de l'intérêt des schémas d'ordre élevé sur des fronts de spirales, pour des pas de temps assez grands ($0.1 \leq \Delta t \leq 0.025$). L'étude en 2D s'est faite quantitativement et a montrée que les schémas d'ordre élevé offraient des meilleurs résultats en terme de précision. L'étude en 3D a été plus qualitative et a montré une grande différence entre les spirales calculées par des schémas d'ordre élevé et des schémas obtenus, par combinaison des méthodes éléments finis Lagrange $\mathbb{P}_1$ et des schémas en temps d'ordre 1 et 2.

Rendu à la fin de cette thèse, nos contributions scientifiques ont été à la fois théoriques et pratiques. Nous avons en effet :

— Développé une nouvelle famille de schéma d'intégration en temps de type exponentiel multi-pas, adaptés pour les problèmes de la cardiologie. Nous les avons nommés IEAB. Nous avons proposé un théorème permettant de prouver la zero stabilité sans être obligé de prouver le caractère Lipschitz de la fonction génératrice de la méthode.

— Nous avons proposé une méthode permettant de construire les schémas Rush-Larsen à l'ordre arbitraire. De cette méthode, nous avons proposé l'ordre 3 et 4 de ce type de schéma, qui sont à notre connaissance, les schémas exponentiels multi-pas d'ordre élevé les plus simples que l'on puisse trouver dans la littérature actuelle.

— Nous avons fait une étude de l'efficacité des nouveaux schémas et ceux qui existaient déjà dans le domaine, pour sélectionner le schéma le plus avantageux en terme de coût, précision et facilité de mise en œuvre, pour les simulations de l'équation de la cellule membranaire. Nous avons trouvé que le schéma RL3 était le meilleur.

— Nous avons aussi fait une étude théorique de la convergence des schémas combiné $\mathbb{P}_r + RL1 + FBE$ et $\mathbb{P}_r + RL2 + SBDF2$. Nous avons calculé des constantes pouvant nous indiquer quel pas de temps et quel diamètre de maillage il faut prendre pour obtenir une précision voulue. Nous avons aussi par l'étude de précision sur les spirales montré l'intérêt d'utiliser les schémas d'ordre élevé.

Comme perspectives, nous comptons combiner nos schémas avec les méthodes de volumes finis à la place des éléments finis que nous avons utilisés dans ce travail, pour résoudre les EDP rencontrées en cardiologie. Nous avons combiné pour la résolution temporelle les schémas RL et SBDF. On aimerait dans le futur expérimenter le schéma RL pour les EDP

rencontrées en cardiologie, sans avoir à utiliser le schéma SBDF. Ceci nous permettra de comparer les deux approches.

Annexe

Calcul de l'erreur de consistance pour le schéma RL2

Soit $z(\tau)$ la fonction introduite en (6.136). pour presque tout $x \in \Omega$, il existe d'après la formule de Taylor $\rho_{1x}^n \in [0, \Delta t]$ tel que,

$$\begin{aligned} z(\Delta t, x) &= z(0,x) + \Delta t \partial_t z(0,x) + \frac{\Delta t^2}{2} \partial_{tt} z(0,x) + \partial_{ttt} z(\rho_{1x}^n, x) \\ &= v(t_n, x) + \Delta t \partial_t z(0,x) + \frac{\Delta t^2}{2} \partial_{tt} z(0,x) + \partial_{ttt} z(\rho_{1x}^n, x). \end{aligned} \tag{8.1}$$

Pour faciliter la lecture, on va omettre dans toute la suite la variable x mais on gardera le x sur les ρ^n pour exprimer le fait que toutes les fonctions sont aussi évaluées en x et que les constantes ρ^n dépendent de ce x. On notera aussi $\partial_t a(u), \partial_{tt} a(u), \partial_{ttt} a(u)$ et $\partial_t b(v,u), \partial_{tt} b(v,u), \partial_{ttt} b(v,u)$ tous simplement $\partial_t a, \partial_{tt} a, \partial_{ttt} a$ et $\partial_t b, \partial_{tt} b, \partial_{ttt} b$ respectivement
L'équation (8.1) s'écrira donc,

$$z(\Delta t) = v^n + \Delta t \partial_t z(0) + \frac{\Delta t^2}{2} \partial_{tt} z(0) + \partial_{ttt} z(\rho_{1x}^n). \tag{8.2}$$

D'après l'équation différentielle (6.137) décrivant z, on a,

$$\begin{aligned} \partial_t z(0) &= (\frac{3}{2} a^n - \frac{1}{2} a^{n-1}) v^n + \frac{3}{2} b^n - \frac{1}{2} b^{n-1} \\ &= \frac{3}{2}(a^n v^n + b^n) v^n - \frac{1}{2}(a^{n-1} v^n + b^{n-1}) \end{aligned} \tag{8.3}$$

Par un développement de Taylor, il existe $\rho_{2x}^n, \rho_{3x}^n \in [0, \Delta t]$ tels que,

$$a^{n-1} = a^n - \Delta t \partial_t a^n + \frac{\Delta t^2}{2} \partial_{tt} a(t_n - \rho_{2x}^n) \tag{8.4}$$

$$a^{n-1} = b^n - \Delta t \partial_t b^n + \frac{\Delta t^2}{2} \partial_{tt} b(t_n - \rho_{2x}^n) \tag{8.5}$$

On utilise (8.4) et(8.5) dans (8.3) pour obtenir,

$$\partial_t z(0) = a^n v^n + b^n + \frac{\Delta t}{2}(\partial_t a^n v^n + \partial_t b^n) - \frac{\Delta t^2}{4}(\partial_{tt} a(t_n - \rho_{2x}^n) v^n + \partial_{tt} b(t_n - \rho_{3x}^n)) \tag{8.6}$$

On a alors d'après (8.6),

$$\Delta t\partial_t z(0) = \Delta t(a^n v^n + b^n) + \frac{\Delta t^2}{2}(\partial_t a^n v^n + \partial_t b^n) \\ - \frac{\Delta t^3}{4}(\partial_{tt} a(t_n - \rho_{2x}^n)v^n + \partial_{tt} b(t_n - \rho_{3x}^n)). \quad (8.7)$$

Une fois de plus d'après l'équation différentielle ordinaire (6.137) décrivant z, on a,

$$\partial_{tt} z(0) = \alpha_n \partial_t z(0) = \alpha_n^2 v^n + \alpha_n \beta_n, \\ = \left(\frac{3}{2}a^n - \frac{1}{2}a^{n-1}\right)^2 v^n + \left(\frac{3}{2}a^n - \frac{1}{2}a^{n-1}\right)\left(\frac{3}{2}b^n - \frac{1}{2}b^{n-1}\right). \quad (8.8)$$

Or par un développement de Taylor il existe $\rho_{4x}^n \in [0, \Delta t]$ tel que,

$$\frac{3}{2}a^n - \frac{1}{2}a^{n-1} = \frac{3}{2}a^n - \left(\frac{1}{2}a^n - \frac{\Delta t}{2}\partial_t a(t_n - \rho_{4x}^n)\right) = a^n + \frac{\Delta t}{2}\partial_t a(t_n - \rho_{4x}^n). \quad (8.9)$$

Par le même procédé que dans (8.9), il existe $\rho_{5x}^n \in [0, \Delta t]$ tel que,

$$\frac{3}{2}b^n - \frac{1}{2}b^{n-1} = b^n + \frac{\Delta t}{2}\partial_t b(t_n - \rho_{4x}^n). \quad (8.10)$$

On déduit d'après (8.9) que,

$$\left(\frac{3}{2}a^n - \frac{1}{2}a^{n-1}\right)^2 v^n = (a^n)^2 v^n + \Delta t a^n v^n \partial_t a(t_n - \rho_{4x}^n) + \frac{\Delta t^2}{4}\partial_t a(t_n - \rho_{4x}^n)^2 v^n. \quad (8.11)$$

D'après (8.9) et (8.10) on déduit que,

$$\left(\frac{3}{2}a^n - \frac{1}{2}a^{n-1}\right)\left(\frac{3}{2}b^n - \frac{1}{2}b^{n-1}\right) = a^n b^n + \frac{\Delta t}{2}\left(a^n \partial_t b(t_n - \rho_{5x}^n) + b^n \partial_t a(t_n - \rho_{4x}^n)\right) \\ + \frac{\Delta t^2}{4}\partial_t a(t_n - \rho_{4x}^n)\partial_t b(t_n - \rho_{5x}^n) \quad (8.12)$$

Les égalités (8.11) et (8.12) dans (8.8) nous permet d'écrire,

$$\frac{\Delta t^2}{2}\partial_{tt} z(0) = \frac{\Delta t^2}{2}((a^n)^2 v^n + a^n b^n) \\ + \frac{\Delta t^3}{2}\left(a^n v^n \partial_t a(t_n - \rho_{4x}^n) + \frac{1}{2}\left(a^n \partial_t b(t_n - \rho_{5x}^n) + b^n \partial_t a(t_n - \rho_{4x}^n)\right)\right) \\ + \frac{\Delta t^3}{2}\left(\frac{\Delta t}{4}\partial_t a(t_n - \rho_{4x}^n)^2 v^n + \frac{\Delta t}{4}\partial_t a(t_n - \rho_{4x}^n)\partial_t b(t_n - \rho_{5x}^n)\right) \quad (8.13)$$

Les égalités (8.7) et (8.13) dans (8.2) nous permettent d'écrire,

$$\begin{aligned} z(\Delta t) = v^n + \Delta t(a^n v^n + b^n) + \frac{\Delta t^2}{2}\left(\partial_t a^n v^n + \partial_t b^n + (a^n)^2 v^n + a^n b^n\right) \\ + \frac{\Delta t^3}{2}\left(a^n v^n \partial_t a(t_n - \rho_{4x}^n) + \frac{1}{2}\left(a^n \partial_t b(t_n - \rho_{5x}^n) + b^n \partial_t a(t_n - \rho_{4x}^n)\right)\right) \\ - \frac{\Delta t^3}{4}\left(\partial_{tt} a(t_n - \rho_{2x}^n) v^n + \partial_{tt} b(t_n - \rho_{3x}^n) - \frac{2}{3}\partial_{ttt} z(\rho_{1x}^n)\right) \\ + \frac{\Delta t^3}{8}\left(\Delta t \partial_t a(t_n - \rho_{4x}^n)^2 v^n + \Delta t \partial_t a(t_n - \rho_{4x}^n)\partial_t b(t_n - \rho_{5x}^n)\right). \end{aligned} \tag{8.14}$$

Par un développement de Taylor, il existe $\rho_{6x}^n \in [0, \Delta t]$ tel que,

$$\begin{aligned} v^{n+1} &= v^n + \Delta t \partial_t v^n + \frac{\Delta t^2}{2}\partial_{tt} v^n + \frac{\Delta t^3}{6}\partial_{ttt} v(t_n + \rho_{6x}^n) \\ &= v^n + \Delta t(a^n v^n + b^n) + \frac{\Delta t^2}{2}\left(v^n \partial_t a^n + \partial_t b^n + (a^n)^2 v^n + a^n b^n\right) \\ &\quad + \frac{\Delta t^3}{6}\partial_{ttt} v(t_n + \rho_{6x}^n) \end{aligned} \tag{8.15}$$

la soustraction de (8.14) à (8.13) pour obtenir,

$$\begin{aligned} z(\Delta t) - v(t_{n+1}) = \frac{\Delta t^3}{6}\left(\partial_{ttt} z(\rho_{1x}^n) - \partial_{ttt} v(t_n + \rho_{6x}^n)\right) \\ + \frac{\Delta t^3}{2}\left(a^n v^n \partial_t a(t_n - \rho_{4x}^n) + \frac{1}{2}\left((a^n \partial_t b(t_n - \rho_{5x}^n) + b^n \partial_t a(t_n - \rho_{4x}^n)\right)\right) \\ + \frac{\Delta t^3}{4}\left(-\partial_{tt} a(t_n - \rho_{2x}^n) v^n - \partial_{tt} b(t_n - \rho_{3x}^n)\right) \\ + \frac{\Delta t^3}{8}\left(\Delta t \partial_t a(t_n - \rho_{4x}^n)^2 v^n + \Delta t \partial_t a(t_n - \rho_{4x}^n)\partial_t b(t_n - \rho_{5x}^n)\right) \end{aligned} \tag{8.16}$$

Bibliographie

[1] http://www.who.int/mediacentre/factsheets/fs317/fr/.

[2] L ABIA et JM SANZ-SERNA. „The spectral accuracy of a fully-discrete scheme for a nonlinear third order equation“. In : *Computing* 44.3 (1990), p. 187–196.

[3] Georgios AKRIVIS et Michel CROUZEIX. „Linearly implicit methods for nonlinear parabolic equations“. In : *Mathematics of computation* 73.246 (2004), p. 613–635.

[4] Georgios AKRIVIS, Michel CROUZEIX et Charalambos MAKRIDAKIS. „Implicit-explicit multistep finite element methods for nonlinear parabolic problems“. In : *Mathematics of Computation of the American Mathematical Society* 67.222 (1998), p. 457–477.

[5] Georgios AKRIVIS, Michel CROUZEIX et Charalambos MAKRIDAKIS. „Implicit-explicit multistep methods for quasilinear parabolic equations“. In : *Numerische Mathematik* 82.4 (1999), p. 521–541.

[6] Georgios AKRIVIS et Yiorgos-Sokratis SMYRLIS. „Implicit–explicit BDF methods for the Kuramoto–Sivashinsky equation“. In : *Applied Numerical Mathematics* 51.2-3 (2004), p. 151–169.

[7] Rubin R. ALIEV et Alexander V. PANFILOV. „A simple two-variable model of cardiac excitation“. In : *Chaos, Solitons & Fractals* 7.3 (1996), p. 293–301.

[8] Grégoire ALLAIRE. *Analyse numérique et optimisation : une introduction à la modélisation mathématique et à la simulation numérique*. Editions Ecole Polytechnique, 2005.

[9] Uri M ASCHER, Steven J RUUTH et Brian TR WETTON. „Implicit-explicit methods for time-dependent partial differential equations“. In : *SIAM Journal on Numerical Analysis* 32.3 (1995), p. 797–823.

[10] Ezio BARTOCCI, Elizabeth M CHERRY, James GLIMM et al. „Toward real-time simulation of cardiac dynamics“. In : *Proceedings of the 9th International Conference on Computational Methods in Systems Biology*. ACM. 2011, p. 103–112.

[11] G.W. BEELER et H. REUTER. „Reconstruction of the Action Potential of Ventricular Myocardial Fibres“. English. In : *J. Physiol.* 268 (1977), p. 177–210.

[12] Omer BERENFELD et José JALIFE. „Purkinje-muscle reentry as a mechanism of polymorphic ventricular arrhythmias in a 3-dimensional model of the ventricles“. In : *Circulation Research* 82.10 (1998), p. 1063–1077.

[13] Miguel O BERNABEU, Pras PATHMANATHAN, Joe PITT-FRANCIS et David KAY. „Stimulus protocol determines the most computationally efficient preconditioner for the bidomain equations“. In : *IEEE Transactions on Biomedical Engineering* 57.12 (2010), p. 2806–2815.

[14] Martin J BISHOP, Patrick M BOYLE, Gernot PLANK, Donald G WELSH et Edward J VIGMOND. „Modeling the role of the coronary vasculature during external field stimulation“. In : *IEEE Transactions on Biomedical Engineering* 57.10 (2010), p. 2335–2345.

[15] Martin J BISHOP, David J GAVAGHAN, Natalia A TRAYANOVA et Blanca RODRIGUEZ. „Photon scattering effects in optical mapping of propagation and arrhythmogenesis in the heart“. In : *Journal of electrocardiology* 40.6 (2007), S75–S80.

[16] Yves BOURGAULT, Yves COUDIERE et Charles PIERRE. „Existence and uniqueness of the solution for the bidomain model used in cardiac electrophysiology“. In : *Nonlinear analysis : Real world applications* 10.1 (2009), p. 458–482.

[17] Franck BOYER et Pierre FABRIE. *Mathematical tools for the study of the incompressible Navier-Stokes equations and related models*. T. 183. Springer Science & Business Media, 2012.

[18] Groupe de Travail des Doctorants BRACHET MATTHIEU. *Introduction à l'approximation des équations au dérivées partielles-Les différences finis*. Institut Elie Cartan de Lorraine.

[19] JH BRANDTS. *A. Quarteroni, R. Sacco and F. Saleri. Numerical mathematics (Texts in applied mathematics ; 37). New York : Springer-Verlag, 2000 654 p., prijs $59.95 ISBN 0-387-98959-5l*. 2002.

[20] Susanne BRENNER et Ridgway SCOTT. *The mathematical theory of finite element methods*. T. 15. Springer Science & Business Media, 2007.

[21] Haïm BREZIS. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.

[22] M. P. CALVO et C. PALENCIA. „A class of explicit multistep exponential integrators for semilinear problems“. In : *Numer. Math.* 102.3 (2006), p. 367–381.

[23] J. CERTAINE. „The solution of ordinary differential equations with large time constants“. In : *Mathematical methods for digital computers*. Wiley, New York, 1960, p. 128–132.

[24] Elizabeth M CHERRY, Henry S GREENSIDE et Craig S HENRIQUEZ. „Efficient simulation of three-dimensional anisotropic cardiac tissue using an adaptive mesh refinement method“. In : *Chaos : An Interdisciplinary Journal of Nonlinear Science* 13.3 (2003), p. 853–865.

[25] Philippe G CIARLET. *The finite element method for elliptic problems*. SIAM, 2002.

[26] Colleen E CLANCY et Yoram RUDY. „Na+ channel mutation that causes both Brugada and long-QT syndrome phenotypes“. In : *Circulation* 105.10 (2002), p. 1208–1213.

[27] Yves COUDIÈRE et Rodolphe TURPAULT. „Very high order finite volume methods for cardiac electrophysiology“. In : *Computers & Mathematics with Applications* (2017).

[28] Richard COURANT. „Variational methods for the solution of problems of equilibrium and vibrations“. In : *Lecture Notes in Pure and Applied Mathematics* (1994), p. 1–1.

[29] Michel CROUZEIX. „Une méthode multipas implicite-explicite pour l'approximation des équations d'évolution paraboliques“. In : *Numerische Mathematik* 35.3 (1980), p. 257–276.

[30] D DIFRANCESCO et Denis NOBLE. „A model of cardiac electrical activity incorporating ionic pumps and concentration changes“. In : *Philosophical Transactions of the Royal Society of London B : Biological Sciences* 307.1133 (1985), p. 353–398.

[31] Karima DJABELLA. „Modélisation de l'activité électrique du coeur et de sa régulation par le système nerveux autonome“. Thèse de doct. Université Paris Sud-Paris XI, 2008.

[32] W EINTHOVEN. „The string galvanometer and the human electrocardiogram“. In : *KNAW Proceedings*. T. 6. 1903, p. 107–115.

[33] Marc ETHIER et Yves BOURGAULT. „Semi-implicit time-discretization schemes for the bidomain model“. In : *SIAM Journal on Numerical Analysis* 46.5 (2008), p. 2443–2468.

[34] Richard FITZHUGH. „Impulses and physiological states in theoretical models of nerve membrane“. In : *Biophysical journal* 1.6 (1961), p. 445–466.

[35] L. GERARDO-GIORDA, M. PEREGO et A. VENEZIANI. „Optimized Schwarz coupling of bidomain and monodomain models in electrocardiology“. In : *M2AN* (2010).

[36] E. HAIRER, S. P. NORSETT et G. WANNER. *Solving ordinary differential equations. I*. Second. T. 8. Springer Series in Computational Mathematics. Nonstiff problems. Springer-Verlag, Berlin, 1993, p. xvi+528.

[37] E. HAIRER et G. WANNER. *Solving ordinary differential equations. II*. T. 14. Springer Series in Computational Mathematics. Stiff and differential-algebraic problems, Second revised edition, paperback. Springer-Verlag, Berlin, 2010, p. xvi+614.

[38] M. HOCHBRUCK et A. OSTERMANN. „Explicit Exponential Runge-Kutta Methods for Semilinear Parabolic Problems.“ In : *SIAM J. Numerical Analysis* 43.3 (2005), p. 1069–1090.

[39] M. HOCHBRUCK et A. OSTERMANN. „Exponential integrators“. In : *Acta Numer.* 19 (2010), p. 209–286.

[40] Marlis HOCHBRUCK. „A short course on exponential integrators“. In : *Matrix functions and matrix equations*. T. 19. Ser. Contemp. Appl. Math. CAM. Higher Ed. Press, Beijing, 2015, p. 28–49.

[41] Marlis HOCHBRUCK et Alexander OSTERMANN. „Exponential multistep methods of Adams-type“. In : *BIT* 51.4 (2011), p. 889–908.

[42] Alan L HODGKIN et Andrew F HUXLEY. „A quantitative description of membrane current and its application to conduction and excitation in nerve“. In : *The Journal of physiology* 117.4 (1952), p. 500.

[43] Allan L HODGKIN et Andrew F HUXLEY. „Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo“. In : *The Journal of physiology* 116.4 (1952), p. 449.

[44] Vincent JACQUEMET et Craig S HENRIQUEZ. „Finite volume stiffness matrix for solving anisotropic cardiac propagation in 2-D and 3-D unstructured meshes“. In : *IEEE transactions on biomedical engineering* 52.8 (2005), p. 1490–1492.

[45] Lions JACQUES-LOUIS. *Quelques méthodes de résolution des problemes aux limites non linéaires*. T. 31. Dunod Paris, 1969.

[46] Franck JEDRZEJEWSKI. *Introduction aux méthodes numériques*. Springer Science & Business Media, 2006.

[47] Claes JOHNSON. *Numerical solution of partial differential equations by the finite element method*. Courier Corporation, 2012.

[48] James P KEENER et James SNEYD. *Mathematical physiology*. T. 1. Springer, 2009.

[49] James KEENER et James SNEYD. *Mathematical Physiology, Interdisciplinary Applied Mathematics 8*. 1998.

[50] Rikkert H KELDERMANN, Martyn P NASH, Hanneke GELDERBLOM, Vicky Y WANG et Alexander V PANFILOV. „Electro-Mechanical wavebreak in a model of the human left ventricle“. In : *American Journal of Physiology-Heart and Circulatory Physiology* (2010).

[51] John KIM et Parviz MOIN. „Application of a fractional-step method to incompressible Navier-Stokes equations“. In : *Journal of computational physics* 59.2 (1985), p. 308–323.

[52] Peter D LAX et Robert D RICHTMYER. „Survey of the stability of linear finite difference equations“. In : *Communications on pure and applied mathematics* 9.2 (1956), p. 267–293.

[53] Pierre-Louis LIONS et Bertrand MERCIER. „Splitting algorithms for the sum of two nonlinear operators“. In : *SIAM Journal on Numerical Analysis* 16.6 (1979), p. 964–979.

[54] Ching-hsing LUO et Yoram RUDY. „A dynamic model of the cardiac ventricular action potential. I. Simulations of ionic currents and concentration changes.“ In : *Circulation research* 74.6 (1994), p. 1071–1096.

[55] Ching-hsing LUO et Yoram RUDY. „A model of the ventricular cardiac action potential. Depolarization, repolarization, and their interaction.“ In : *Circulation research* 68.6 (1991), p. 1501–1526.

[56] Jaakko MALMIVUO et Robert PLONSEY. *Bioelectromagnetism*. `http://www.bem.fi/book/`.

[57] Walter T MILLER et David B GESELOWITZ. „Simulation studies of the electrocardiogram. I. The normal heart.“ In : *Circulation Research* 43.2 (1978), p. 301–315.

[58] Gary R MIRAMS, Christopher J ARTHURS, Miguel O BERNABEU et al. „Chaste : an open source C++ library for computational physiology and biology“. In : *PLoS computational biology* 9.3 (2013), e1002970.

[59] Gary R MIRAMS, Yi CUI, Anna SHER et al. „Simulation of multiple ion channel block provides improved early prediction of compounds' clinical torsadogenic risk“. In : *Cardiovascular research* 91.1 (2011), p. 53–61.

[60] Cleve MOLER et Charles VAN LOAN. „Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later“. In : *SIAM review* 45.1 (2003), p. 3–49.

[61] Jinichi NAGUMO, Suguru ARIMOTO et Shuji YOSHIZAWA. „An active pulse transmission line simulating nerve axon“. In : *Proceedings of the IRE* 50.10 (1962), p. 2061–2070.

[62] Steven A NIEDERER, Eric KERFOOT, Alan P BENSON et al. „Verification of cardiac tissue electrophysiology simulators using an N-version benchmark“. In : *Phil. Trans. R. Soc. A* 369.1954 (2011), p. 4331–4351.

[63] Denis NOBLE. „A modification of the Hodgkin—Huxley equations applicable to Purkinje fibre action and pacemaker potentials“. In : *The Journal of Physiology* 160.2 (1962), p. 317.

[64] Sundnes J. Artebrant R. Skavhaug O et Tveito A. „A second-order Algorithm for solving Dynamic cell Membrane Equations“. In : *IEEE Transactions On Biomedical Engineering* 26 (oct. 2009), p. 2546–2548.

[65] M. PEREGO et A. VENEZIANI. „An efficient generalization of the Rush-Larsen method for solving electro-physiology membrane equations“. In : *ETNA* 35 (2009), p. 234–256.

[66] C. PIERRE. „Modélisation et simulation de l'activité électrique du coeur dans le thorax, analyse numérique et méthodes de volumes finis“. Thèse de doct. Math´ematiques et Applications, sept. 2005.

[67] Charles PIERRE. „Modélisation et simulation de l'activité électrique du coeur dans le thorax, analyse numérique et méthodes de volumes finis". Thèse de doct. Université de Nantes, 2005.

[68] Bernardo RUDY et Linda E IVERSON. *Ion channels*. T. 207. Gulf Professional Publishing, 1997.

[69] S RUSH et H LARSEN. „A practical algorithm for solving dynamic membrane equations." In : *IEEE Trans Biomed Eng* 25.4 (juil. 1978), p. 389–92.

[70] Y. SAAD. *Iterative methods for sparse linear systems*. Second. Philadelphia, PA : Society for Industrial et Applied Mathematics, 2003, p. xviii+528.

[71] Hasan I SALEHEEN et Kwong T NG. „A new three-dimensional finite-difference bidomain formulation for inhomogeneous anisotropic cardiac tissues". In : *IEEE Transactions on Biomedical engineering* 45.1 (1998), p. 15–25.

[72] Robin M SHAW et Yoram RUDY. „Electrophysiologic effects of acute myocardial ischemia : a theoretical study of altered cell excitability and action potential duration". In : *Cardiovascular Research* 35.2 (1997), p. 256–272.

[73] Eric Goncalvès da SILVA. „Méthodes et Analyse Numériques". In : (2007).

[74] J. SUNDNES, G. T. LINES, X. CAI et al. *Computing the Electrical Activity in the Heart*. Springer, 2006.

[75] Joakim SUNDNES, Glenn Terje LINES, Xing CAI et al. *Computing the electrical activity in the heart*. T. 1. Springer Science & Business Media, 2007.

[76] P SYVERT et NØRSETT. „An A-stable modification of the Adams-Bashforth methods". In : *Conference on the Numerical Solution of Differential Equations*. T. 109. Lecture Notes in Mathematics. Springer, Berlin, 1969, p. 214–219.

[77] KHWJ TEN TUSSCHER, D NOBLE, PJ NOBLE et Alexander V PANFILOV. „A model for human ventricular tissue". In : *American Journal of Physiology-Heart and Circulatory Physiology* 286.4 (2004), H1573–H1589.

[78] Vidar THOMÉE. *Galerkin finite element methods for parabolic problems*. T. 1054. Springer, 1984.

[79] Mayya TOKMAN, John LOFFELD et Paul TRANQUILLI. „New Adaptive Exponential Propagation Iterative Methods of Runge–Kutta Type". In : *SIAM Journal on Scientific Computing* 34.5 (2012), A2650–A2669.

[80] Mark L TREW, Bruce H SMAILL, David P BULLIVANT, Peter J HUNTER et Andrew J PULLAN. „A generalized finite difference method for modeling cardiac electrical activation on arbitrary, irregular computational meshes". In : *Mathematical biosciences* 198.2 (2005), p. 169–189.

[81] Mark TREW, Ian LE GRICE, Bruce SMAILL et Andrew PULLAN. „A finite volume method for modeling discontinuous electrical activation in cardiac tissue". In : *Annals of biomedical engineering* 33.5 (2005), p. 590–602.

[82] Leslie TUNG. „A bi-domain model for describing ischemic myocardial dc potentials." Thèse de doct. Massachusetts Institute of Technology, 1978.

[83] Balth VAN DER POL. „LXXXVIII. On "relaxation-oscillations"". In : *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (1926), p. 978–992.

[84] Richard S VARGA. *Matrix iterative analysis*. T. 27. Springer Science & Business Media, 2009.

[85] Edward J VIGMOND, Matt HUGHES, G PLANK et L Joshua LEON. „Computational tools for modeling electrical activity in cardiac tissue". In : *Journal of electrocardiology* 36 (2003), p. 69–74.

[86] L. Vu Thai et A. Ostermann. „Explicit exponential Runge-Kutta methods of high order for parabolic problems“. In : *J. Comput. Appl. Math.* 256 (2014), p. 168–179.

[87] Jinglin Zeng, Kenneth R Laurita, David S Rosenbaum et Yoram Rudy. „Two components of the delayed rectifier K+ current in ventricular myocytes of the guinea pig type“. In : *Circulation Research* 77.1 (1995), p. 140–152.

[88] H Zhang, AV Holden, I Kodama et al. „Mathematical models of action potentials in the periphery and center of the rabbit sinoatrial node“. In : *American Journal of Physiology-Heart and Circulatory Physiology* 279.1 (2000), H397–H421.

[89] Hao Zhuang. „Exponential Time Integration for Transient Analysis of Large-Scale Circuits“. In : (2016).

Colophon

This thesis was typeset with LaTeX 2$_\varepsilon$. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at `http://cleanthesis.der-ric.de/`.