



RAPPORT DE STAGE

Compression de maillage et problèmes d'évolution

Intégration temporelle et multirésolution adaptative pour les EDP en temps.

Étudiant : Alexandre EDELINE
École : ENSTA Paris - Institut Polytechnique de Paris
Période : du 14/04/2025 au 15/09/2025

Laboratoire : CMAP - École Polytechnique
Maîtres de stages : Marc MASSOT et Christian TENAUD
Tuteur académique : Patrick CIARLET

6 août 2025

Remerciements

Je tiens à remercier...

Résumé

Résumé

Résumé du rapport de stage en français (150-300 mots). Présenter brièvement le contexte, les objectifs, la méthodologie, les principaux résultats et conclusions.

Mots-clés : Schémas Numériques, Simulation des EDP d'Évolution, Multirésolution Adaptative, Méthodes ImEx, Advection-Diffusion-Réaction, Analyse d'erreur numérique, Analyse de stabilité

English abstract of the internship report (150-300 words). Briefly present the context, objectives, methodology, main results and conclusions.

Keywords : Numerical Schemes, Evolution PDE Simulation, Adaptive Multiresolution, ImEx Methods, Advection-Diffusion-Reaction, Numerical Error Analysis, Stability Analysis

Table des matières

Remerciements	2
Résumé	3
Liste des figures	3
Liste des tableaux	4
1 Présentation du laboratoire	5
1.0.1 Historique et activités	5
1.0.2 La recherche au CMAP	5
1.0.3 L'équipe HPC@Math et l'environnement de travail	6
2 Description du travail objectifs et état de l'art	7
2.1 Présentation du sujet et problématique générale	7
2.2 Quelques notions techniques	8
2.2.1 Intégrations des EDOs	8
2.2.2 Les équations d'advection-diffusion-réaction	11
2.2.3 Simulation des EDPs d'évolution	16
2.2.4 Analyse de schéma numériques	16
2.2.5 La Multirésolution Adaptative	17
2.3 Objectifs	20
2.4 Méthode de travail et outils	21
3 Contribution	22
3.1 Étude de méthodes ImEx sur une équation de diffusion-réaction	23
3.1.1 L'équation de Nagumo	24
3.1.2 Les méthodes ImEx	26
3.1.3 Analyse de stabilité	29
3.1.4 Étude de la convergence	32
3.1.5 Conclusion	32
3.2 Impact de la multirésolution adaptative sur l'erreur d'une méthode des lignes	33
3.2.1 Cadre de l'étude	33
3.2.2 Les équations équivalentes	36
3.2.3 Comparaison	37
3.2.4 Conclusion	38
3.2.5 Complément expérimental	38
3.2.6 Conclusion	38
4 Conclusion	39

Bibliographie	41
.1 Annexe A : Titre de l'annexe	42
.2 Annexe B : Titre de l'annexe	42

Table des figures

2.1	Illustration du comportement attendu de l'erreur d'un schéma d'ordre deux dont le seuil d'instabilité serait $\Delta t > 10^{-1}$.	9
2.2	Exemple de maillage adapté par multirésolution adaptative grâce au logiciel Samurai.	13
2.3	Exemple de grille dyadique	17
3.1	Profils des ondes solutions de l'équation de Nagumo pour différents ratios k/D à $kD = 1$ fixé (c'est à dire à vitesse fixée). L'augmentation du ratio k/D accentue le gradient spatial.	24
3.2	Plage de valeurs du terme de réaction non-linéaire et de sa différentielle pour deux coefficients de réactions : $k = 1$ et $k = 10$.	25
3.3	Pour différents couples D et k , diagrammes de stabilité des méthodes ImEx et de référence sur l'équation de Nagumo.	30
3.4	Pour $k = 500$ et $D = 500$: diagrammes de stabilité des méthodes ImEx et de référence sur l'équation de Nagumo.	33

Liste des tableaux

Chapitre 1

Présentation du laboratoire

1.0.1 Historique et activités

Le Centre de Mathématiques Appliquées de l'École Polytechnique¹ (CMAP) a été créé en 1974 lors du déménagement de l'École Polytechnique vers Palaiseau. Cette création répond au besoin émergent de mathématiques appliquées face au développement des méthodes de conception et de simulation par calcul numérique dans de nombreuses applications industrielles de l'époque (nucléaire, aéronautique, recherche pétrolière, spatial, automobile). Le laboratoire fut fondé grâce à l'impulsion de trois professeurs : Laurent SCHWARTZ, Jacques-Louis LIONS et Jacques NEVEU. Jean-Claude NÉDÉLEC en fut le premier directeur, et la première équipe de chercheurs associés comprenait PA. RAVIART, P. CIARLET, R. GLOWINSKI, R. TEMAM, J.M. THOMAS et J.L. LIONS. Les premières recherches se concentraient principalement sur l'analyse numérique des équations aux dérivées partielles. Le CMAP s'est diversifié au fil des décennies, intégrant notamment les probabilités dès 1976, puis le traitement d'images dans les années 1990 et les mathématiques financières à partir de 1997. Le laboratoire a formé plus de 230 docteurs depuis sa création et a donné naissance à plusieurs startups spécialisées dans les applications industrielles des mathématiques appliquées.

1.0.2 La recherche au CMAP

Le CMAP comprend trois pôles de recherche : le pôle analyse, le pôle probabilités et le pôle décision et données. Chaque pôle accueille en son sein plusieurs équipes :

1. Analyse

- ◇ EDP pour la physique.
- ◇ Mécanique, Matériaux, Optimisation de Formes.
- ◇ HPC@Maths (calcul haute performance).
- ◇ PLATON (quantification des incertitudes en calcul scientifique), avec l'INRIA.

2. Probabilités

- ◇ Mathématiques financières.
- ◇ Population, système particules en interaction.
- ◇ ASCII (interactions stochastiques coopératives), avec l'INRIA.

1. <https://cmap.ip-paris.fr>

- ◇ MERGE (évolution, reproduction, croissance et émergence), avec l'INRIA.

3. Décision et données

- ◇ Statistiques, apprentissage, simulation, image.
- ◇ RandOpt (optimisation aléatoire).
- ◇ Tropical (algèbre $(\max, +)$), avec l'INRIA.

J'ai intégré l'équipe **HPC@Maths pole analyse**. De nombreuses équipes sont partagées entre le CMAP et l'INRIA ce qui démontre l'aspect appliqué du laboratoire.

1.0.3 L'équipe HPC@Math et l'environnement de travail

.i L'équipe HPC@Math L'équipe HPC@Math² travaille à l'interface des mathématiques de la physique (mécanique des fluides, thermodynamique) et de l'informatique pour développer des méthodes numériques complètes (schéma, analyse d'erreur, implémentation) pour la simulation des EDP. L'équipe se centre sur les problèmes multi-échelles; les EDPs cibles qui typiquement étudiées sont les équations d'advection-réaction-diffusion qui représente de manière générale le couplage entre la mécanique des fluides, la thermodynamique et la chimie (typiquement un problème de combustion). Tout cela se fait dans le contexte HPC (high performance computing). Le HPC désigne l'usage optimal des ressources informatiques disponibles cela peut être développer une simulation efficace sur une petite machine comme des schéma hautement parallélisable dans des paradigmes de calculs hybrides ou dans des contextes hexascale (échelle hexaflopique)³. Ainsi l'application des méthodes développées est au cur des réflexions de l'équipe.

.ii Environnement de travail

2. <https://initiative-hpc-maths.gitlab.labos.polytechnique.fr/site/index.html>

3. Plateformes de calculs ayant une capacité de calcul théorique de 10^{16} opérations par seconde (hexaflops).

Chapitre 2

Description du travail objectifs et état de l'art

Cette partie présente les objectifs du stage et les méthodes employées. Elle introduit également le lecteur au sujet, à ses problématiques et comprend un préambule mathématique présentant un bref état de l'art et les notions élémentaires des différents domaines convoqués.

2.1 Présentation du sujet et problématique générale

Ce travail participe à l'élaboration de méthodes numériques pour l'approximation des équations aux dérivées partielles d'évolution. En particulier les équations d'advection-diffusion-réaction (présentation en 2.2.2). Elles décrivent par exemple les systèmes physiques couplant mécanique des fluides, thermodynamique et réactions chimiques¹. Ces équations sont difficiles à simuler du fait de leur caractère multi-échelle². Pour gérer les différentes échelles spatiales, des méthodes de compression de maillage sont souvent mises en oeuvre. La méthode de compression utilisée et étudiée ici est la multirésolution adaptative [10]. Les différentes échelles temporelles³ sont usuellement gérées par force brute ou par séparation d'opérateurs. Pour pallier le problème de la large gamme d'échelles temporelles rencontrées, une approche hybride est ici étudiée : les méthodes implicites-explicites (ImEx) [2]. Ce travail vise donc principalement à comprendre comment la multirésolution adaptative interagit avec les différentes méthodes d'intégration temporelle. Il s'intéresse aux questions suivantes :

- ◇ Comment les effets de la compression de maillage par multirésolution adaptative (MRA) sur les solutions numériques dépendent-ils du problème étudié et de la méthode numérique sur lesquelles elle se greffent ?
- ◇ Comment évoluent les propriétés des méthodes ImEx selon les caractéristiques des opérateurs des équations de diffusion-réaction⁴ ?

1. Typiquement des problèmes de combustion.

2. Une réaction chimique a des temps et distances typiques généralement plusieurs ordres de grandeur plus faibles que les temps et distances typiques de la mécanique des fluides.

3. En termes techniques, les différents termes des équations étudiées ont des raideurs très différentes.

4. Même si l'objectif est bien les équations d'advection-diffusion-réaction, l'étude s'est concentrée par simplicité sur l'interaction entre phénomènes de diffusion et de réactions.

2.2 Quelques notions techniques

2.2.1 Intégrations des EDOs

Les techniques d'approximation d'EDPs d'évolution comportent souvent une étape nécessitant la résolution d'une équation différentielle ordinaire (EDO⁵), c'est à dire une équation différentielle ne faisant intervenir qu'une seule variable différenciée (ici le temps). Nous commençons donc cette section par rappeler quelques notions d'analyse et de simulation des EDOs⁶.

Définition 2.2.1 (Équation différentielle ordinaire). Une équation différentielle ordinaire est une équation de la forme :

$$\begin{aligned} u' &= A(u, t) \quad u : t \in \mathbb{R}^+ \mapsto u(t) \in \mathbb{R}^d \\ u(0) &= u_0 \in \mathbb{R}^d. \end{aligned} \quad (2.1)$$

A Schémas explicites et implicites.

L'approximation des EDO se fait grâce à des schémas numériques; formellement un schéma numérique est un élément de $(\mathbb{R}^d)^{\mathbb{N}}$. Ceux-ci se divisent en deux catégories, les schémas explicites et les schémas implicites. Seuls les schémas à un pas sont ici présentés et non pas les schémas multi-pas. Ce choix est fait en raison de la barrière de Dahlquist⁷. Donnée un pas de discrétisation temporel Δt , on note u^n l'approximation de la solution d'une EDO au pas de temps n , c'est à dire au temps $t^n = n\Delta t$. L'objectif est d'avoir $u^n \approx u(t = n\Delta t)$.

Définition 2.2.2 (Schéma explicite). Un schéma numérique est dit explicite si le pas de temps $n + 1$ est obtenu seulement grâce au pas de temps n , usuellement formulé sous la forme :

$$u^{n+1} = u^n + f(u^n, \Delta t). \quad (2.2)$$

Définition 2.2.3 (Schéma implicite). Un schéma numérique est dit implicite si le pas de temps $n + 1$ est obtenu au moins en partie grâce au pas de temps $n + 1$, souvent écrit comme :

$$u^{n+1} = u^n + f(u^{n+1}, \Delta t). \quad (2.3)$$

Ainsi, une itération d'un schéma implicite nécessite l'inversion d'un système linéaire ou non linéaire.

De fait une itération implicite est souvent plus coûteuse qu'une itération d'un schéma explicite⁸. Cependant pour des raisons de stabilité (voir C) les méthodes explicites peuvent nécessiter des pas de temps bien plus fin, et donc bien plus d'itérations. Le choix entre méthode explicite et implicite dépend de bien des facteurs (du problème, du niveau de précision voulu, de la difficulté d'implémentation etc...) c'est un enjeu central de la simulation numérique.

5. On utilisera aussi le terme *système dynamique*, même si en toute rigueur ce concept est un peu plus large.

6. Pour nos besoins nous nous restreignons aux EDOs du premier ordre.

7. https://fr.wikipedia.org/wiki/Methode_lineaire_a_pas_multiples#Premiere_et_deuxieme_limites_de_Dahlquist

8. En particulier si la dimension de la solution d est grande.

B Ordre des schémas numériques

C Stabilité des schémas numériques

Un schéma numérique d'ordre p converge vers la solution exacte de l'EDO avec une erreur qui décroît asymptotiquement en Δt^p lorsque le pas de temps diminue. Cependant, cette convergence n'est garantie que si le schéma reste stable. L'instabilité d'un schéma désigne la divergence de la solution numérique : au-delà d'un pas de temps critique Δt_0 , la norme de la solution discrète $\|u^n\|$ tend vers l'infini⁹. Cette instabilité peut s'interpréter de deux manières complémentaires : d'un point de vue mathématique, le schéma se comporte comme une suite géométrique de raison $|r| > 1$; d'un point de vue physique, le schéma introduit artificiellement de l'énergie dans le système à chaque itération. La contrainte de stabilité impose donc la contrainte $\Delta t < \Delta t_0$. Lorsque ce seuil Δt_0 est très restrictif, c'est à dire très faible, la résolution de l'EDO nécessite un nombre d'itérations $T_{\text{final}}/\Delta t$ important, ce qui augmente considérablement le coût calculatoire. Les schémas explicites sont généralement plus prompts aux instabilités que les méthodes implicites.



FIGURE 2.1 – Illustration du comportement attendu de l'erreur d'un schéma d'ordre deux dont le seuil d'instabilité serait $\Delta t > 10^{-1}$.

Définition 2.2.4 (Stabilité d'un schéma numérique). Un schéma numérique $(u^n)_{n \in \mathbb{N}} \in (\mathbb{R}^d)^{\mathbb{N}}$ est stable si et seulement si :

$$\|u^{n+1}\| \leq \|u^n\|. \quad (2.4)$$

Cette condition est souvent vérifiée à la condition que le pas de discrétisation Δt n'excède pas un seuil de stabilité Δ_0 fonction de l'ODE et le schéma d'intégration.

9. Phénomène communément appelé "explosion" de la solution numérique.

La stabilité d'une méthode d'intégration d'EDO dépend entre autres de l'opérateur intervenant dans l'équation (le A dans l'équation 2.1). Un opérateur tendant à poser des problèmes de stabilité est dit raide.

Définition 2.2.5 (Problème raide). Un système dynamique, est dit raide si les méthodes explicites ne sont pas adaptées à sa résolution. En termes plus mathématiques le système :

$$\frac{du}{dt} = A(u, t), \quad u(t) \in \mathbb{R}^d, \forall t \geq 0. \quad (2.5)$$

est dit raide si la jacobienne de A , J_A possède des valeurs propres négatives de grande amplitudes devant les autres valeurs propres. Si tel est le cas, plusieurs relaxations sont mises en jeu mais chacune avec des temps caractéristiques d'ordres de grandeur différents. Pour les méthodes explicites, si le pas de temps n'est pas assez petit, la relaxation rapide est mal résolue et impose un gradient fort trop longtemps (le gradient devrait s'atténuer, mais à une échelle trop rapide pour être captée pour le pas de temps du schéma) ce qui déstabilise la méthode.

En simplifiant, si un opérateur est raide, il impose une condition de stabilité très restrictive aux méthodes explicites et force à choisir des méthodes implicites¹⁰.

Exemple 2.2.6 (Équation de Dahlquist). Pour saisir de manière plus intuitive le concept de raideur, prenons le cas de l'équation de Dahlquist¹¹ :

$$\begin{aligned} \frac{du}{dt} &= -\lambda u, \quad \lambda > 0 \\ u(t=0) &= u_0 \end{aligned} \quad (2.6)$$

La solution analytique est : $u(t) = u_0 e^{-\lambda t}$. Ainsi passé quelque $1/\lambda$ la dynamique du système est au point mort. En pratique la dynamique digne d'intérêt du système se concentre donc entre $t = 0$ et $t = \frac{10}{\lambda}$. Au delà, $u(t > \frac{10}{\lambda}) = o(u_0)$, la dynamique est terminée. Ainsi, si l'on souhaite simuler le comportement d'un tel système, il faut choisir des pas de temps petits devant $|\lambda|^{-1}$. Lorsque λ est de grande amplitude cela devient très contraignant... Si l'on souhaite utiliser des méthodes explicites, c'est encore pire car la raideur du système n'est plus une simple contrainte de précision mais de stabilité. En effet si l'on cherche à approximer le système par un schéma d'Euler explicite, alors : $U^{n+1} = U^n(1 - \lambda \Delta t)$ alors la contrainte de stabilité est $\Delta t \lambda < 1/2$ ce qui est contraignant si λ est grand. Si $\lambda = 10^5$ alors il faut avoir $\Delta t \approx 10^{-5}$ donc pour simuler le système entre $t = 0$ et $t = 1$ il faut cent mille points! A l'inverse si l'on choisit un schéma d'Euler implicite : $u^{n+1} = u^n - \lambda \Delta t u^{n+1}$, alors la condition de stabilité devient : $\|(1 + \lambda \Delta t)^{-1}\| \leq 1$ ce qui est toujours vrai, quelque soit $-\lambda \in \mathbb{R}^-$, la raideur du système n'est pas un problème pour la méthode implicite. On comprend mieux la définition précédente *Un système dynamique, est dit raide si les méthodes explicites ne sont pas adaptées à sa résolution.*

Il existe différents types de stabilité comme la A-stabilité (méthode stable indépendamment de la raideur du problème), la L-stabilité (schéma amortissant les hautes fréquences), par souci de concision cette partie s'achève ici mais ces notions sont développées par exemple dans [9].

10. La réalité est plus nuancée, nous le verrons.

11. C'est le cas le plus simple d'une valeur propre négative

2.2.2 Les équations d'advection-diffusion-réaction

Le contexte physique naturel des équations d'advection, diffusion, réaction est le suivant : des particules sont placées dans un milieu fluide où elles **diffusent**. Ce milieu fluide est en mouvement, il déplace les particules, les **advecte**. Enfin les particules **réagissent** entre elles et ces réactions modifient les grandeurs thermodynamiques (température, pression) et *in fine* les propriétés du milieu fluide. Les équations d'advection, diffusion, réaction modélisent donc ces trois phénomènes et leurs couplages respectifs.

A Les trois opérateurs

A.i Advection L'advection désigne le transport d'une quantité par un flot. L'opérateur d'advection le plus simple est l'opérateur de transport $c \frac{\partial}{\partial x}$:

$$\frac{\partial u}{\partial t} = c \frac{\partial u}{\partial x} \quad (2.7)$$

De manière générale l'opérateur d'advection d'une quantité u par un flot \underline{a} s'écrit $\underline{a} \cdot \underline{\nabla} u$. Par exemple dans les équations de Navier-Stokes, le terme $\underline{v} \cdot \underline{\nabla} \underline{v}$ modélise la vitesse \underline{v} transportée par elle-même. Une version simplifiée de ce phénomène est l'équation bien connue de Burgers.

Les opérateurs d'advections sont généralement à valeurs propres imaginaires¹². Ainsi ils sont peu raides mais résonnants. Les méthodes explicites sont souvent les plus adaptées pour les traiter.

A.ii Diffusion La diffusion désigne l'*éparpillement* de particules au sein d'un milieu fluide¹³. Ce phénomène est la limite macroscopique du déplacement microscopique des particules dû à l'agitation thermique. L'opérateur de diffusion le plus classique est celui de l'équation de la chaleur :

$$\frac{\partial u}{\partial t} = D \Delta u. \quad (2.8)$$

Le spectre de cet opérateur est \mathbb{R}^- , il est donc infiniment raide. Lorsqu'il est discrétisé seul une partie de sa raideur est captée, en pratique la raideur de l'opérateur augmente de manière quadratique avec la finesse de la discrétisation spatiale.

Cet opérateur est donc moyennement raide (comparé aux opérateurs de réaction). Ainsi on pourrait penser qu'une méthode implicite est adéquate. Cependant ce n'est pas toujours le cas. En effet le coefficient de diffusion est généralement fonction de la température, et donc l'opérateur $D(T)\Delta(\cdot)$ varie souvent en temps et en espace. Ainsi il faut inverser à chaque itération l'opérateur implicite. Cette lourde tâche est rendue plus ardue par le fait que l'opérateur de diffusion couple tout l'espace, c'est-à-dire qu'il est non local.¹⁴, il faut inverser une matrice de taille $d \gg 1$ dont la structure peut être très hétérogène (car le coefficient de diffusion varie dans l'espace). Aujourd'hui il semble plus pertinent d'utiliser des méthodes explicites stabilisées qui parviennent à gérer la raideur moyenne

12. Par abus, s'il s'agit d'un opérateur non-linéaire on lui associera les valeurs propres de sa Jacobienne.

13. En théorie de l'information cela décrit la tendance de l'entropie augmenter et l'information à se moyenner, se flouter.

14. Si l'opérateur de diffusion eût été local il aurait pu être inversé en résolvant plusieurs petits systèmes et ce, potentiellement en parallèle. Cela serait bien moins coûteux qu'inverser un grand système. Par exemple inverser une matrice pleine de taille 10^6 demande environ 10^{18} opérations, alors qu'inverser 100 systèmes de taille 10^4 n'en demande $100 \times 10^{12} = 10^{14}$ soit dix mille fois moins. Si ces résolutions sont parallélisées l'accélération est d'un million. Malheureusement comme l'opérateur de diffusion couple tout l'espace ce n'est pas possible en dimension deux ou trois.

comme les méthodes ROK2 et ROK4[1]. Sans entrer dans les détails, la clé de ces méthodes consiste à prendre une méthode numérique explicite standard et lui ajouter des étages de sorte à ce que la fonction d'amplification résultante soit la fonction d'amplification de la méthode standard multipliée par un polynôme minimal sur \mathbb{R}^- comme un polynôme de Tchebichev.

A.iii Réaction Les phénomènes de réaction sont en général bien adaptés aux méthodes implicites car ils sont locaux et extrêmement raides. En effet, les temps typiques d'une réaction chimique¹⁵ sont de l'ordre de la nano-seconde. De fait, les réactions chimiques sont très difficiles à simuler par des méthodes explicites. En revanche les méthodes implicites sont très efficaces dans ce contexte. En effet l'inversion de l'opérateur implicite peut être décomposé en plusieurs résolutions de petits systèmes ce qui est moins coûteux et parallélisable. Cela est permis grâce à la localité des réactions chimiques (à chaque pas de temps les particules au sein d'une cellule ne réagissent qu'avec les autres particules de la même cellule). En pratique cela signifie qu'il est possible de mettre en oeuvre une petite méthode implicite par cellule plutôt qu'une gargantuesque méthode globale; ce qui revient à inverser un opérateur de petite dimension en chaque cellule et non inverser un énorme système.

B Difficultés mathématiques intrinsèques

La simulations des équations d'advection-réaction-diffusion se heurte à deux difficultés majeures, **le couplage des trois opérateurs** mentionnés précédemment et **le caractère multi-échelles des solutions**.

B.i Première difficulté : le couplage des opérateurs Le développement précédent montre que résoudre chaque phénomène individuellement, n'est pas insurmontable. Toutefois, les résoudre tous en même temps, c'est-à-dire les coupler, est en pratique très difficile. En effet, lorsque ces trois opérateurs sont couplés, il en résulte un unique opérateur qui doit être traité par une méthode numérique. C'est là que surgissent les difficultés : si la méthode est explicite (éventuellement stabilisée), la raideur de la réaction impose des pas de temps extrêmement restrictifs, à l'inverse si la méthode est implicite, la non-localité de la diffusion demande l'inversion d'un système de taille déraisonnable. Cette approche naïve, monolithique, n'est donc pas adaptée. Il faut par conséquent, trouver d'autres stratégies.

B.ii Seconde difficulté : le caractère multi-échelles des solutions Les solutions étudiées sont souvent multi-échelles, en temps et en espace. Cela signifie que certaines zones spatio-temporelles requièrent une finesse d'approximation élevée pour pouvoir reproduire fidèlement le comportement physique, tandis qu'en d'autres zones une approximation grossière est suffisante. Prenons l'exemple d'un incendie. Au début le foyer est très restreint et seule cette zone doit être maillée finement, car partout ailleurs *il ne se passe rien*. Petit à petit l'incendie se propage et la zone à mailler finement augmente. Un autre exemple de phénomène multi-échelle est la détonation, il faut mailler finement, au foyer de l'explosion et le front de l'onde de choc. Mais la zone non atteinte par l'explosion, qui n'a pas encore reçu le choc, pourrait être maillée très grossièrement. Dans ces conditions, il est imaginable

15. En réalité une réaction chimique aussi simple en apparence qu'une combustion H_2/O_2 fait intervenir une dizaine de composés et réactions intermédiaires, dont les temps typiques sont très faibles.)

qu'un maillage naïf mène à ce qu'en certains instants, 90% du domaine soit maillé avec un pas d'espace 100 fois plus fin que nécessaire, or en trois dimension mailler 100 fois trop finement multiplie par un million le nombre de cellules. Il y a alors une grande inefficacité computationnelle.

C Les stratégies de simulation

Pour surmonter ces difficultés, il est d'usage d'utiliser de concert des stratégies d'adaptation de maillage et d'intégration en temps spécifiques.

C.i L'adaptation de maillage Pour tirer parti du caractère multi-échelle des solutions, il est courant de recourir à des méthodes d'adaptation de maillage. C'est-à-dire d'employer une résolution variable pour la grille de calcul selon les différentes zones du domaine. L'adaptation doit se mettre à jour au fil des itérations pour suivre la dynamique de la solution car la répartition de la complexité physique évolue tout au long de la simulation. La méthode d'adaptation de maillage sur laquelle le stage se concentre est la multirésolution-adaptative par transformée d'ondelettes introduite par Ami Harten dans les années 1990 [10]. Cette méthode est très étudiée par l'équipe du CMAP et a donné lieu au développement du logiciel Samurai¹⁶.



FIGURE 2.2 – Exemple de maillage adapté par multirésolution adaptative grâce au logiciel [Samurai](#).

La multi-résolution adaptative La multirésolution adaptative se base sur une compression de la solution par transformée d'ondelettes¹⁷. Les détails mathématiques sont donnés plus tard et s'appuient notamment sur [14] mais pour résumer, la compression se fait de la manière suivante : une transformée multi-échelle sur une base d'ondelettes représente la solution sur différentes échelles d'espace¹⁸. Cela quantifie l'information portée par chaque niveau de résolution. Enfin la compression consiste à ignorer les échelles contenant moins d'information qu'un seuil de compression ε

16. <https://github.com/hpc-maths/samurai>

17. C'est le même procédé qui est à l'oeuvre dans la compression d'image jpg.

18. Par exemple les échelles $\Delta x, \Delta x/2, \Delta x/4, \dots, \Delta x/2^n$.

fixé par l'utilisateur.

Autres méthodes Il existe d'autres stratégies pour raffiner la grille de calcul que la multirésolution adaptative. La plus classique est l'adaptation basée sur la magnitude des gradients de la solution. Une grande magnitude des gradients révèle une complexité locale de la solution, appelant une fine résolution du maillage. Cette approche est simple à mettre en oeuvre et part de la même heuristique que la multirésolution adaptative (détecter où se trouve la complexité de la solution), cependant elle est moins systématique (pas de quantification intrinsèque de l'information perdue) et plus difficile à analyser. Elle est cependant utilisée dans des logiciels industriels comme Ansys¹⁹.

C.ii Les techniques d'intégration Comme expliqué précédemment, la simulation de chaque phénomène intervenant dans les problèmes d'advection-diffusion-réaction est aisée individuellement mais très difficile conjointement. Des stratégies pour intégrer les trois opérateurs en même temps sont donc nécessaires. Elles reposent en réalité sur le fait d'intégrer chaque opérateur... séparément (ou presque).

La séparation d'opérateurs La séparation d'opérateurs (en Anglais : *operator splitting*) consiste à intégrer successivement chaque opérateur. L'intégration d'un opérateur A , dans une équation comme :

$$\frac{\partial u}{\partial t} = Au. \quad (2.9)$$

Peut s'écrire formellement avec la notion d'exponentielle de matrice²⁰ :

$$u(\Delta t) = e^{\Delta t A} u_0. \quad (2.10)$$

Si deux opérateurs A et B interviennent dans l'EDP cela donne :

$$\frac{\partial u}{\partial t} = (A + B)u. \quad (2.11)$$

$$u(\Delta t) = e^{\Delta t(A+B)} u_0 \quad (2.12)$$

Il est alors tentant d'écrire :

$$u(\Delta t) \approx e^{\Delta t(B)} e^{\Delta t(A)} u_0. \quad (2.13)$$

Ce n'est malheureusement qu'une approximation car $e^{\Delta t(A+B)} = e^{\Delta t(B)} e^{\Delta t(A)}$ n'est vrai que si les opérateurs A et B commutent. Cependant c'est vrai à l'ordre $O(\Delta t)$. Il est donc possible de simuler le

19. <https://www.ansys.com/fr-fr/blog/how-to-accelerate-ansys-fluent-simulations-with-adaptive-meshing>

20. Ou de manière plus rigoureuse et plus générale avec la notion de semi-groupe.

problème en séparant les opérateurs de la manière suivante :

$$\text{Étape 1 : simuler } v = e^{\Delta t A} u^n, \quad (2.14)$$

$$\text{Étape 2 : simuler } u^{n+1} = e^{\Delta t B} v, \quad (2.15)$$

$$(2.16)$$

Cet algorithme correspond au schéma de splitting de Lie, introduisant une erreur de l'ordre Δt . Il existe un autre schéma : le splitting de Strang qui est précis à l'ordre Δt^2 grâce au développement $e^{\Delta t(A+B)} = e^{\Delta t B} e^{\Delta t A} e^{\Delta t B}$. Ces méthodes de séparation d'opérateurs, très simples à mettre en oeuvre, permettent de traiter les opérateurs indépendamment les uns à la suite des autres ; chacun avec une méthode numérique adaptée. Malheureusement elles sont aveugles à certains couplages entre les différents phénomènes modélisés et il est difficile de monter au-delà de l'ordre deux en temps. La promesse des approches ImEx présentées par la suite est de combler ces lacunes.

Une étude extensive de l'usage du splitting, pour les équations d'advection-diffusion-réaction couplées à la multirésolution adaptative a été réalisée dans la thèse de Max Duarte, préparée à Centrale Paris sous la direction de Marc Massot [6]. Cela a montré que les techniques d'opérateurs sont très efficaces mais qu'en pratique, une perte de l'ordre formel peut avoir lieu.

Les méthodes ImEx Ces méthodes sont détaillées en 3.1.2. Ces méthodes ImEx²¹ [13] [12] sont très proches de la séparation d'opérateurs lors de l'implémentation. Toutefois, elles apportent plus de cohérence mathématique et facilitent la montée en ordre. Une méthode Runge et Kutta est une technique d'intégration en temps dont chaque itération se décompose en plusieurs étapes (aussi appelées étages). Chaque étage engendre une nouvelle approximation, puis en fin d'itération une combinaison linéaire des ces approximations intermédiaires donne la solution au pas de temps suivant. Cette approche permet de monter en ordre efficacement. Dans les méthodes Runge et Kutta ImEx à chaque étage, l'approximation associée est obtenue en ajoutant d'une part des contributions explicites des opérateurs pour lesquels les méthodes explicites sont adaptées, et d'autre part des contributions implicites des opérateurs pour lesquels les méthodes implicites sont adaptées. Ainsi, un traitement différent est appliqué à chaque opérateur tout en conservant une approche globalement cohérente²².

D Conclusion

Cette introduction a mis en évidence la complexité intrinsèque des équations d'advection-diffusion-réaction, qui réside dans le couplage de trois phénomènes physiques aux propriétés mathématiques antagonistes. L'advection, peu raide, la diffusion, moyennement raide et non-locale, et la réaction, extrêmement raide mais locale, ne peuvent être traitées efficacement par une approche monolithique classique. Les deux défis principaux identifiés : le couplage des opérateurs et le caractère multi-échelles des solutions, nécessitent des stratégies numériques spécifiques. D'une part, l'adaptation de maillage, notamment par multi-résolution adaptative, permet de concentrer les ressources de calcul

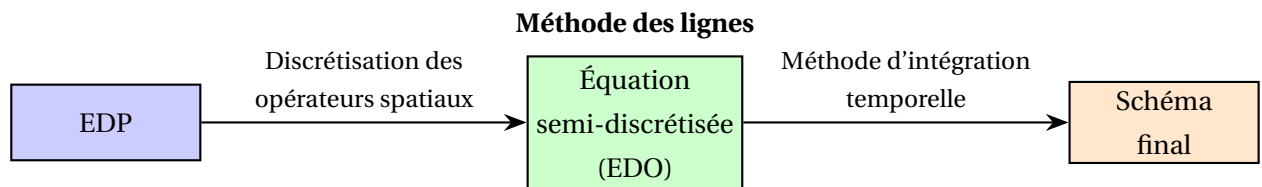
21. Les méthodes présentées ici sont les méthodes ImEx Runge et Kutta (RK-ImEx). Il existe également des approches ImEx couplées espace-temps[15].

22. Cependant les méthodes de splitting offrent le luxe de choisir chaque technique de résolution numérique indépendamment des autres, ici ce n'est plus le cas des relations d'ordre doivent être respectées ; plus de cohérence mais plus de contrainte.

là où l'information physique est la plus riche. D'autre part, des techniques d'intégration découplées, séparation d'opérateurs ou des méthodes ImEx, exploitant les propriétés spécifiques de chaque phénomène et éviter l'écueil du solver monolithique.

2.2.3 Simulation des EDPs d'évolution

Définition 2.2.7 (Méthode des lignes). Une méthode des lignes est une famille de méthodes numériques pour approximer les EDP d'évolutions. Elle consiste à discrétiser les opérateurs spatiaux de l'équation afin d'obtenir une équation semi-discrétisée en espace, puis à utiliser une technique d'intégration en temps, pour obtenir la discrétisation complète de l'équation.



Définition 2.2.8 (Méthodes d'intégration espace temps).

Dans la suite de notre étude nous allons utiliser la paradigme des volumes finis. Les volumes finis sont particulièrement adaptés au lois de conservations. Les volumes finis discrétisent la valeur moyenne sur les mailles, alors que les différences finies discrétisent la valeur au nuds du maillage et les éléments finis discrétisent l'espace fonctionnel lui même.

Définition 2.2.9 (Volumes finis). Donné un maillage $(C_j)_{j \in J}$ d'un domaine Ω , la discrétisation par volume fini approxime les quantités :

$$U_j = \frac{1}{|C_j|} \int_{C_j} u(x) d\Omega. \quad (2.17)$$

2.2.4 Analyse de schéma numériques

Stabilité... Convergence...

Définition 2.2.10 (Procédure de Cauchy-Kovalevskaya).

Définition 2.2.11 (Équation modifiée).

2.2.5 La Multirésolution Adaptative

La multi-résolution adaptative (MRA) est une méthode très efficace pour les problèmes multi-échelles. L'objectif est de concentrer les efforts computationnels là où ils sont nécessaires. Concrètement cela consiste à augmenter la résolution de la grille de calcul où la solution est complexe et la diminuer où la solution est simple à décrire. La MRA est donc une méthode de HPC (*high performance computing*) puisqu'elle vise à optimiser l'allocation des ressources de calcul.

Cette partie introduit le lecteur à cette méthode en présentant d'abord le concept mathématique de transformée multi-échelle (ou transformée en ondelette) qui est à la base de la MRA. Puis il est expliqué comment la transformée multi-échelle permet d'adapter le maillage pour optimiser la charge computationnelle. Une fois ces prérequis établis, l'algorithme typique de mise en oeuvre de la multi-résolution adaptative est décrit. Sensuit alors naturellement une présentation des différentes implémentations de la MRA, avec une attention particulière sur celle développée au CMAP au travers du logiciel Samurai. Enfin l'impact de la multi-résolution sur la qualité des solutions numériques est abordé.

A La transformée multi-échelle

Cette partie présente la transformée multi-échelle discrète. La transformée multi-échelle continue en simulation numérique, c'est bien sûr la version discrète qui est utile. Elle se veut avant tout introductive et omet ou simplifie certaines notions; plus de détails sont donnés en [14].

A.i Définition mathématique Les explications sont développées en dimensions un à des fins pédagogiques, la plupart des concepts s'entendent naturellement aux dimensions supérieures. De plus, la discrétisation de l'espace se fait selon une grille dyadique, d'autre choix pourraient être fait mais c'est un choix simple, naturel et standard. Il faut détailler cette notion.

Définition 2.2.12 (Grille dyadique). Une grille dyadique ou discrétisation dyadique d'un intervalle $I \subset \mathbb{R}$ est une série de partitions de I indexées par des entiers $j \in J \subset \mathbb{N}^*$. La discrétisation de niveau j correspond à une partitions de I en 2^j intervalles (voir fig. 2.3). Ainsi à chaque changement de niveau, du niveau j vers le niveau $j+1$, la résolution de la discrétisation est doublée. Les cases de cette partition dyadique sont indexées par deux entiers : j le niveau de résolution de la grille et k l'index de la case au sein de ce niveau. En particulier les cases $2k$ et $2k+1$ du niveau $j+1$ correspondent à la case k du niveau j .

Niveau $j-1$	$k=0$				$k=1$			
Niveau j	$k=0$		$k=1$		$k=2$		$k=3$	
Niveau $j+1$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$

FIGURE 2.3 – Exemple de grille dyadique

Dans ce qui suit il est supposé sans perte de généralité que la discrétisation se fait sur l'intervalle $[0, 1]$, ainsi le niveau j correspond à des cellules de tailles $1/2^j$ et la cellule k du niveau j est centrée en $x_k^j = \frac{k+(k+1)}{2} \frac{1}{2^j} = \frac{2k+1}{2^j}$. La notion d'ondelette se définit de la manière suivante :

Définition 2.2.13 (Ondelette). Une ondelette est une fonction $\Phi \in L^2(\mathbb{R})$ à support compact de moyenne nulle. Pour qu'une ondelette soit pertinente dans le cas de la transformée en multi-échelle il est requis que la famille $(x \mapsto \Phi(2^j k - x))_{(j,k) \in \mathbb{Z} \times \mathbb{Z}}$ forme une base de $L^2(\mathbb{R})$. En effet la transformée en ondelette sera une projection sur cette base, un peu comme la transformée de Fourier est une projection sur les fonctions trigonométriques.

Alors la transformée en ondelette discrète peut être définie :

Définition 2.2.14 (Transformée en ondelette discrète - *Discrete Wavelet Transform, DWT*). Donnée une fonction f , le coefficient γ_k^j de sa DWT sur la cellule k au niveau de résolution j est :

$$\gamma_k^j = \frac{1}{N_j} \int_{\mathbb{R}} \Phi(2^j \cdot k - t) f(t) dt, \quad (2.18)$$

Où N_j est un coefficient normalisation dépendant du niveau j .

Contrairement à une transformée de Fourier, les coefficients ne dépendent pas d'une mais de deux variables. En effet, la transformée en ondelette est plus riche d'informations. Là où la transformée de Fourier ne donne qu'une information sur le contenu fréquentiel d'un signal, la transformée en ondelette donne une information sur le contenu en fréquence et sur la localisation de ce contenu fréquentiel.

A.ii La notion de détails La multi-résolution adaptative se sert de la transformée multi-échelle pour adapter le maillage, c'est à dire compresser l'information (voir B). Cela requiert l'introduction de la notion de détail. Ce concept permet de ne pas utiliser la transformée en ondelette pour quantifier le contenu absolu porté par une échelle particulière, mais plutôt à comprendre en quoi ce contenu s'éloigne de ce que les échelles supérieures pourraient laisser supposer, en quoi il est *inattendu*. Pour résumer à partir d'un niveau de résolution j , on définit un **prédicteur** polynomial, qui tente d'inférer l'allure de la fonction au niveau $j + 1$. Puis le **détail** ne cherche pas naïvement à quantifier et localiser l'information contenue aux échelles du niveau $j + 1$ mais plutôt, à quantifier l'écart à la prédiction polynomiale.

Le prédicteur Donné un point central (x_0, y_0) , et $2s$ voisins $(x_{-s}, y_{-s}), (x_{-s+1}, y_{-s+1}) \dots (x_{s-1}, y_{s-1}), (x_s, y_s)$, un prédicteur polynomial ponctuel cherche le polynôme P de degré $2s$ passant par ces $2s + 1$ points. Cela permet d'inférer des valeurs pour y en tout point x . Pour trouver $P(X) = \sum_{k=0}^{2s} a_k X^k$ revient à résoudre le système linéaire :

$$\forall j \in \{-s, \dots, 0, \dots, s\} : y_j = \sum_{k=0}^{2s} a_k x_j^k \quad (2.19)$$

Ce stage se focalise sur les volumes finis, donc ce n'est pas un prédicteur ponctuel, adapté au différences finies (voir ??), qui est utilisé mais un prédicteur sur la valeur moyenne. Il ne cherche à imposer les valeurs en chaque point mais à fixer la valeur moyenne sur chaque cellule, cela ajoute peu de complexité puisqu'il suffit d'ajouter une intégration lors de l'établissement du système linéaire pour travailler sur les valeurs moyennes. Pour l'usage que souhaité ici, il s'agit en d'évaluer la solution sur la cellule k du niveau de résolution j (ce qui correspond à la cellule de centre $x_k^j = (k + 1/2)2^{-j}$ au

niveau de résolution supérieure $j + 1$, il faut donc appliquer le correcteur linéaire centré sur x_k^j en $x_{\pm} = \pm 2^{-(j+1)}$. En pratique cela revient à faire une combinaison linéaire des $2s$ voisins qui vient corriger la valeur en x_k^j . Le prédicteur dépend donc du nombre de voisins pris de part et d'autre, ce nombre noté s et appelé le *stencil* du prédicteur. Plus s est grand plus l'opération de prédiction est précise (mais peut éventuellement devenir bruitée²³) et plus elle est coûteuse. Le coût exact n'est pas évident à estimer puisque qu'une combinaison linéaire quelques termes se fait en $O(1)$ sur les machines modernes, toutefois quelques subtilités détaillé en B interviennent. Ainsi les valeurs usuelles du stencil sont généralement $s = 1$ ou $s = 2$

Les détails À présent que le prédicteur à été décrit le concept de détail peut enfin être abordé. On suppose que l'on dispose d'une fonction \tilde{u}^j qui soit une approximation de la fonction u , au niveau de résolution j . Comme vu précédemment, le prédicteur permet d'obtenir une approximation de u au niveau $j + 1$ grâce à \tilde{u}^j . Cette prédiction est noté \hat{u}^{j+1} . Les détails à la résolution $j + 1$ sont alors définis comme les coefficient de la transformée multi-échelle de la différence entre cette prédiction et la vraie fonction : $u - \hat{u}^{j+1}$. Alors les coefficients de détails n'encode que ce qui n'était pas prédictible par l'interpolateur polynomial. Ce concept de détail est essentiel. En pratique on ne réalise pas l'opération comme expliqué plus haut puisque à prédicteur et ondelette fixés Φ , il existe une *ondelette duale* Ψ qui permet directement d'obtenir les détails pour la DWT sur Φ en réalisant une DWT sur Ψ ce qui accélère considérablement les calculs.

Une autre façon de voir la notion de détail est la suivante : pour un niveau de résolution j on note $V_j = \text{Vect}\left((\Phi_k^j)_k\right)$, c'est à dire l'ensemble de fonction représentables par les ondelettes de niveau j . Pour les ondelettes classiques²⁴ la relation suivante est vérifiée $V_0 \subset V_1 \subset V_2 \subset \dots \subset V_N$. Et bien alors l'espace des détails, celui accessible par l'ondelette duale est Q_{j+1} le supplémentaire de V_j dans V_{j+1} , en d'autre terme, il représente toutes les informations, les *détails* contenues dans le niveau d'approximation V_{j+1} qui n'étaient pas prise en compte par V_j (à l'échelle j ce n'était que des détails). Les coefficients de la décomposition par l'ondelette duale sont Ainsi grâce à l'ondelette duale, il est possible de calculer les coefficients de détails d_k^j qui à chaque montée en résolution n'encode que l'information qui n'était pas contenue dans la décomposition en ondelette au niveau précédent.

Les deux visions ne sont pas rigoureusement les mêmes, la première représente ce qui est réalisé en pratique lors de la MRA, la seconde est la vision standard de la théorie des ondelettes. Toutefois les deux approches ont la même motivation, ne calculer et ne mettre en valeur à chaque niveau que ce qui est nouveau, ce qui n'était pas contenu dans les niveaux précédents.

Intuition sur les détails Lorsque l'on s'intéresse aux coefficients de détails d_k^j , l'indice j de *dilatation* fixe l'échelle analysée, c'est à dire la longueur d'onde analysée. Par exemple si $j = 5$, les coefficients d_k^5 donne une information sur l'information portée par les longueurs d'onde de l'ordre de $2^{-5} = 1/32$. La variable k précise l'indice de la cellule analysée. Par exemple $\|d_1^5\| > \|d_5^5\|$ signifie que l'information portée par l'échelle $1/32$ est plus importante au voisinage de la case 10 qu'au voisinage de la case 5. De même si $\|d_7^j\| > \|d_{14}^{j+1}\|$ cela signifie qu'au voisinage de $x = \frac{7}{2^j}$ les longueurs d'ondes $\frac{1}{2^j}$ sont plus présentes que les longueurs d'ondes $\frac{1}{2^{j+1}}$. Pour se fixer les idées, c'est comme si la transformée de Fourier n'avait qu'une vision globale du contenu en fréquence, quelle ne voyait que la

23. Ce n'est pas détaillé ici mais le lecteur se référera à la théorie de l'interpollation...

24. C'est en tout cas vrai pour les ondelettes de Haar[14].

moyenne sur le domaine de la transformée en ondelette pour chaque longueur d'onde.

$$\|TF[f](\omega = 2^j)\|^2 \sim \left\| \sum_k d_k^j \right\|^2. \quad (2.20)$$

B L'adaptation

B.i La compression Pour compresser une fonction, le processus est très simple, il suffit de fixer un seuil de compression $\varepsilon > 0$, de calculer les coefficients de détails de la fonction, puis d'omettre (en pratique de retirer de la mémoire) les coefficient dont la norme est inférieure à ε . En pratique le coefficient de compression dépend du niveau étudié puisque le volume des cellules mise en jeu chute avec le niveau j (un détail de 10^{-2} à moins d'importance s'il porte sur des cellules de taille 10 que sur des cellules de taille 10^{-5}). En pratique l'algorithme serait le suivant :

1. Calculer les coefficients d_k^j .
2. Pour chaque niveau j et chaque coefficient d_k^j : Si $d_k^j < \varepsilon_j = 2^{-j}\varepsilon$, alors $d_k^j \leftarrow 0$.

Pour reconstruire au niveau de résolution souhaité : utiliser les détails jusqu'au niveau j le plus fin conservé puis utiliser le prédicteur pour interpoler jusqu'au niveau désiré. Le vocabulaire est heureusement choisit, pour compresser on omet les détails négligeables l'on conserve les détails importants.

B.ii L'adaptation

C Algorithmes de simulation numérique

C.i

C.ii

D Impact sur les solutions

D.i

D.ii

E Implémentations de la multi-résolution

E.i Méthodes usuelles

E.ii Le logiciel Samurai

2.3 Objectifs

Mon premier objectif a été de comparer les performances des méthodes ImEx avec une méthode de séparation d'opérateur car sur le papier ces deux méthodes sont très proches. J'ai mis en lumière leurs différents domaines de stabilité et réalisé des tests de convergence. Cette étude a été réalisée

sur une équation simple mais très représentative des équations de réaction-diffusion, l'équation de Nagumo.

Mon second objectif était d'explorer l'impact théorique de la multi-résolution adaptative (MRA) sur un schéma numérique classique. J'ai mené une analyse d'erreur sur une équation de diffusion résolue par une méthode des lignes d'ordre deux, et j'ai découvert que la multirésolution-adaptative peut dégrader l'ordre de convergence d'une méthode. J'ai par la suite tenté (sans succès) de mettre en évidence ce résultat expérimentalement grâce au code Samurai. J'ai choisi de travailler sur la diffusion pour compléter le travail réalisé par l'équipe en [3] qui se centrait sur l'advection.

2.4 Méthode de travail et outils

Chapitre 3

Contribution

Cet partie vise à présenter mes travaux. Rappelons qu'il se sont concentrés sur la résolutions des EDPs d'advection-diffusion-réaction grâce aux méthodes modernes de compression de maillage (multirésolution adaptative) et de découplage des opérateurs (ImEx et splitting).

Dans un premier temps, je présente l'analyse d'une méthode ImEx appliquée à l'équation de Nagumo (une équation de diffusion-réaction). J'étudie la stabilité des méthodes de manière générale puis dans le cas spécifique de cette équation. Par la suite, j'ai mené une comparaison expérimentale avec une méthode concurrente : la séparation d'opérateur.

Dans un second temps je présente une étude de l'impact de la multirésolution adaptative sur la convergence d'une méthode des lignes sur un problème de diffusion. Cette étude comprend un pan théorique (obtention de l'équation équivalente du schéma avec multirésolution adaptative) et un pan expérimental (étude de convergence grâce au logiciel Samurai).

3.1 Étude de méthodes ImEx sur une équation de diffusion-réaction

Pour cette première étude nous présentons d'abord l'équation de Nagumo et expliquons en quoi elle est une excellente équation test pour éprouver les méthodes de résolution des équations d'advection-diffusion-réaction. Dans un second temps nous présentons les méthodes ImEx utilisées. Par la suite, nous évaluons la stabilité de ces méthodes ImEx dans le contexte de l'équation de Nagumo. Cela permet de valider la pertinence *a priori* de ces ImEx sur l'équation, et nous mène naturellement à une étude de convergence expérimentale où les méthodes ImEx présentées sont confrontées à une méthode de séparation d'opérateur classique (splitting de Strang).

3.1.1 L'équation de Nagumo

L'équation de Nagumo (ou FitzHugh-Nagumo) est issue de modèles de transmission de l'information nerveuse [7]. Nous utilisons la forme spatiale de l'équation [11] avec un terme de réaction cubique pour ajouter de la non-linéarité :

$$\partial_t u = \underbrace{D \partial_{xx} u}_{\text{diffusion}} - \underbrace{ku(1-u^2)}_{\text{réaction}}. \quad (3.1)$$

A Solutions Analytiques

Cette équation admet des solutions propagatives sous la forme¹ :

$$u(x - ct) = \frac{e^{-\sqrt{\frac{k}{2D}}((x-x_0)-ct)}}{1 + e^{-\sqrt{\frac{k}{2D}}((x-x_0)-ct)}} \quad (3.2)$$

Avec : $c = \sqrt{\frac{kD}{2}}$ et x_0 le point de départ de l'onde. Ainsi, le produit kD fixe la vitesse et le ratio $\frac{k}{D}$ la magnitude du gradient d'espace.

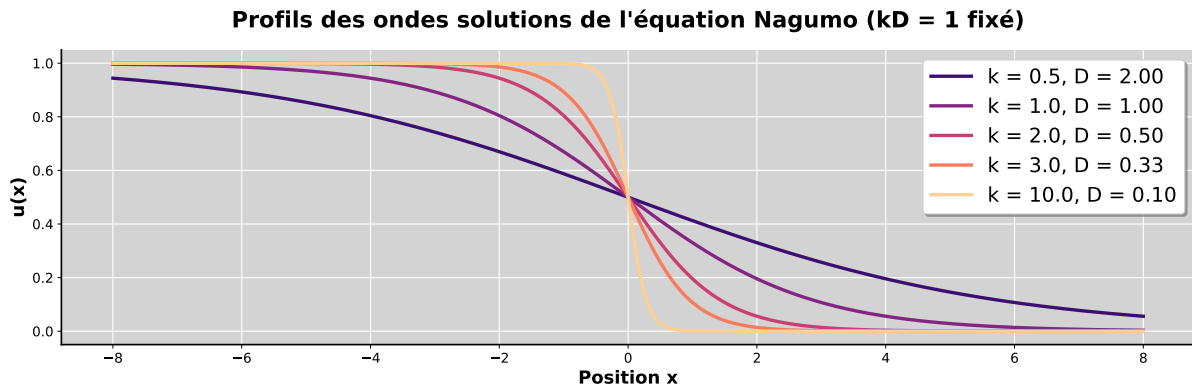


FIGURE 3.1 – Profils des ondes solutions de l'équation de Nagumo pour différents ratios k/D à $kD = 1$ fixé (c'est à dire à vitesse fixée). L'augmentation du ratio k/D accentue le gradient spatial.

B Analyse des opérateurs

Le terme de diffusion est non-local et, discrétisé à l'ordre deux par n points et un pas Δx , les valeurs propres associées sont $\{\frac{2D}{\Delta x^2}(\cos \frac{p\pi}{n+1} - 1) \mid p \in \{1, \dots, n\}\}$ [4], ainsi la raideur du terme de diffusion croît linéairement avec le coefficient de diffusion D et de manière quadratique avec la finesse du maillage $1/\Delta x$. En effet les valeurs propres sont négatives et :

$$\max_p \left| \cos \frac{p\pi}{n+1} - 1 \right| \sim 2, \quad (3.3)$$

$$\min_p \left| \cos \frac{p\pi}{n+1} - 1 \right| \sim \frac{1}{2} \left(\frac{\pi}{n+1} \right)^2. \quad (3.4)$$

1. C'est une sigmoïde, qui se propage à vitesse $\sqrt{\frac{kD}{2}}$.

Et donc :

$$\frac{\max_p |1 - \cos \frac{p\pi}{n+1}|}{\min_p |1 - \cos \frac{p\pi}{n+1}|} \approx n^2. \quad (3.5)$$

Concernant le terme de réaction, en choisissant un état initial correspondant à 3.2, la solution reste entre 0 et 1. Ainsi le terme de réaction est local, et ses valeurs propres sont comprises entre $-k$ et $2k$. En fonction de la valeur de u , la réaction se comporte comme une relaxation de temps caractéristique $\tau \sim \frac{1}{k}$ ou comme une explosion de temps caractéristique $\tau \sim \frac{1}{2k}$. Pour les valeurs étudiées, $k \leq 20$, ainsi la réaction reste peu raide. En effet :

$$R(u) = ku(1 - u^2), \quad (3.6)$$

$$R'(u) = k(1 - 3u^2). \quad (3.7)$$

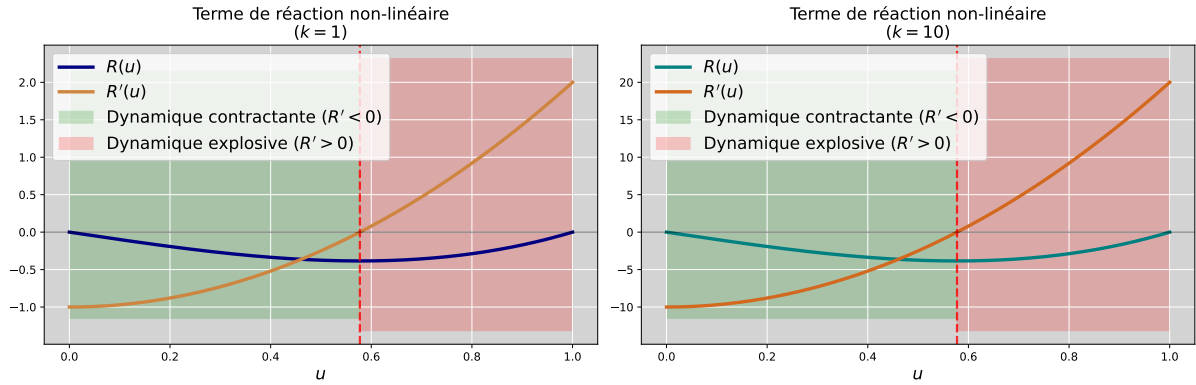


FIGURE 3.2 – Plage de valeurs du terme de réaction non-linéaire et de sa différentielle pour deux coefficients de réactions : $k = 1$ et $k = 10$.

C Conclusion sur l'équation de Nagumo

Ainsi l'équation de Nagumo, présente un terme de réaction², et un terme de diffusion. Cette équation fait émerger un front d'onde³ et dispose de deux paramètres k et D simple pour modifier les propriétés de la solution. Cela en fait donc un modèle-test de choix pour étudier le comportement de diverses méthodes dédiées aux équations d'advection-réaction-diffusion.

2. à noter qu'il n'est pas raide, comparé aux termes de réaction rencontrés en combustion.

3. Cela permet de tester le comportement de la multi-résolution adaptative.

3.1.2 Les méthodes ImEx

La classe de méthodes ImEx étudiée sont les méthodes de Runge et Kutta additives (RK-ImEx ou RK-additive). Ces méthodes consistent à sommer plusieurs méthodes de Runge et Kutta appliquées chacune à un opérateur différent. L'objectif est d'employer des RK explicites (RKE) et des RK implicites (RKI), en adéquation avec les besoin de chaque opérateur.

A Un exemple

Pour introduire aux méthodes de Runge et Kutta additives, commençons par un exemple simple et usons d'une méthode RK-ImEx d'ordre un, résultant de la somme de deux méthodes RK à un étages (RK1). Nous notons cette méthode ImEx111 [2]. Les méthodes RK1 servant de briques élémentaires à la RK111 sont : un schéma d'Euler explicite et un schéma d'Euler implicite. Supposons que l'on cherche à approcher une équation d'évolution faisant intervenir deux opérateurs : A^E se prêtant à des méthodes explicites⁴ et A^I se prêtant aux méthodes implicites⁵. L'équation cible serait de la forme :

$$\partial_t u = A^E u + A^I u. \quad (3.8)$$

A.i Résolution par approche monolithique Rappelons d'abord comme le problème serait résolu en n'utilisant qu'une seule RK1 pour tout le problème (approche monolithique).

Euler explicite En résolvant avec Euler explicite, le schéma s'écrit :

$$u^{n+1} = u^n + \Delta t(A^E + A^I)u^n. \quad (3.9)$$

Mais si l'opérateur A^I est très raide, la stabilité risque d'imposer un pas de temps très restrictif risquant de rendre la méthode non viable.

Euler implicite En résolvant avec Euler implicite, le schéma s'écrit :

$$u^{n+1} = (Id - \Delta t(A^E + A^I))^{-1} u^n. \quad (3.10)$$

Mais si l'opérateur A^E rend l'inversion coûteuse; par exemple s'il est non-local (impliquant la résolution d'un gros système au lieu de plusieurs petits systèmes), ou s'il est non linéaire (nécessite d'être réinverser à chaque pas de temps); alors cette méthode ne sera pas viable non plus.

A.ii Résolution par une méthode ImEx : une Runge et Kutta Additive Mettons en oeuvre la méthode ImEx111. L'approximation au pas de temps $n + 1$ s'écrit en sommant une contribution issue de la méthode Euler explicite (RKE1) et une contribution issue de la méthode Euler implicite (RKI1) :

$$u^{n+1} = u^n + \Delta t \left(\underbrace{k_1}_{\text{RKE1}} + \underbrace{k'_1}_{\text{RKI1}} \right) \quad (3.11)$$

4. Par exemple, un opérateur peu raide mais non local.

5. Par exemple un opérateur raide mais local.

La contribution RKE1 s'écrit :

$$k_1 = A^E u^n. \quad (3.12)$$

La contribution RKI1 s'écrit :

$$k'_1 = A^I u^{n+1} \quad (3.13)$$

Ainsi :

$$\begin{aligned} u^{n+1} &= u^{n+1} = u^n + \Delta t (A^E u^n + A^I u^{n+1}), \\ \text{donc : } u^{n+1} - \Delta t A^I u^{n+1} &= u^n + \Delta t A^E u^n, \\ \text{et donc : } u^{n+1} &= (Id - \Delta t A^I)^{-1} \circ (Id + \Delta t A^E) u^n. \end{aligned} \quad (3.14)$$

Ainsi dans cette méthode seul l'opérateur $Id - \Delta t A^I$ doit être inversé. Ce qui était bien l'objectif. Les opérateurs ont été découplés lors de la résolution.

B Cadre mathématique général

Pour construire des méthodes plus complexes et d'ordres supérieurs introduisons le formalisme de [2] pour traiter les méthodes RK-additives. Ici, nous travaillons uniquement sur méthodes ImEx pour deux opérateurs mais théoriquement, il est possible de construire des méthodes ImEx pour traiter autant d'opérateurs que l'on le souhaite [12].

B.i Notation Une méthode ImEx additive est construite à partir d'une méthode implicite à s étages (une méthode DIRK et si possible SDIRK) et d'une méthode explicite à $s+1$ étages⁶. Pour uniformiser, le tableau de Butcher de la méthode implicite est complété par une ligne et une colonne de zéros afin que les deux méthodes s'écrivent comme si elles avaient le même nombre d'étages. Les tableaux de Butcher des deux méthodes s'écrivent alors :

Méthode RKE, $s+1$ étages :

$$\text{RKE : } \begin{array}{c|c} \tilde{c} & \tilde{A} \\ \hline & \tilde{b}^T \end{array} = \begin{array}{c|cccccc} 0 & 0 & 0 & 0 & \cdots & 0 \\ \tilde{c}_1 & \tilde{a}_{10} & 0 & 0 & \cdots & 0 \\ \tilde{c}_2 & \tilde{a}_{20} & \tilde{a}_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \tilde{c}_s & \tilde{a}_{s0} & \tilde{a}_{s1} & \tilde{a}_{s2} & \cdots & 0 \\ \hline & \tilde{b}_0 & \tilde{b}_1 & \tilde{b}_2 & \cdots & \tilde{b}_s \end{array} \quad (3.15)$$

6. Au besoin, la méthode explicite peut être à s étages, qui est un cas particulier d'une méthode à $s+1$ étages.

Méthode RKI (DIRK) s étages :

$$\text{RKI: } \begin{array}{c|c} c & A \\ \hline & b^T \end{array} = \begin{array}{c|cccccc} 0 & 0 & 0 & 0 & \cdots & 0 \\ c_1 & 0 & a_{11} & 0 & \cdots & 0 \\ c_2 & 0 & a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ c_s & 0 & a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline & 0 & b_1 & b_2 & \cdots & b_s \end{array} \quad (3.16)$$

où les coefficients \tilde{a}_{ij} , \tilde{b}_i , \tilde{c}_i définissent la méthode explicite et les coefficients a_{ij} , b_i , c_i définissent la méthode implicite DIRK.

B.ii Schéma général d'une méthode RK-additive Une étape de la méthode RK-additive appliquée au système $\frac{du}{dt} = A^E u + A^I u$ s'écrit :

Calcul des étages : En initialisant $u_0 = u^n$,

$$u_i = u^n + \Delta t \sum_{j=0}^{i-1} \tilde{a}_{ij} A^E u_j + \Delta t \sum_{j=1}^i a_{ij} A^I u_j, \quad i = 0, 1, \dots, s \quad (3.17)$$

Soit en mettant en lumière le caractère implicite de la méthode sur A^I :

$$(Id - \Delta t a_{ii} A^I) u_i = u^n + \Delta t \sum_{j=0}^{i-1} \tilde{a}_{ij} A^E u_j + a_{ij} A^I u_j, \quad i = 0, 1, \dots, s \quad (3.18)$$

Solution à l'étape suivante :

$$u^{n+1} = u^n + \Delta t \sum_{i=0}^s \tilde{b}_i A^E u_i + \Delta t \sum_{i=1}^s b_i A^I u_i \quad (3.19)$$

Cette formulation générale permet de construire des méthodes d'ordre élevé.

B.iii Ordre de convergence L'ordre d'une méthode RK-additive est bien sûr borné par l'ordre le plus faible des méthodes RK individuelles qu'elle convoque. Naturellement, cette borne n'est pas nécessairement atteinte pour toutes les méthodes RK additionnées. Des conditions d'ordre liant les coefficients des méthodes individuelles entre eux doivent être respectées. Le nombre de ces conditions augmente (très) rapidement avec l'ordre de la méthode et le nombre d'opérateurs à résoudre [12], le lecteur motivé se référera par exemple à [8].

3.1.3 Analyse de stabilité

Nous étudions la viabilité des RK-ImEx sur l'équation de Nagumo. Pour cela nous étudions sa stabilité. Dans un premier temps, une étude générale de la stabilité de des RK-ImEx est menée. Dans un second temps, l'étude se centre sur l'application à l'équation de Nagumo.

A Étude de stabilité générale des RK-ImEx

Lorsque l'on use d'une méthode ImEx, les deux (ou plus) opérateurs sont découplés, c'est bien là l'intérêt. Cependant cela complique légèrement l'analyse usuelle de stabilisé. En effet la fonction de stabilité attend alors deux variables, le coefficient spectral associé à l'opérateur traité explicitement et le coefficient spectral associé à l'opérateur traité implicitement. Ainsi, pour chaque couple de valeur propre, la fonction de stabilité prend une valeur différente et comme les coefficients spectraux sont complexes, on ne peut plus visualiser d'un simple coup d'il le domaine de stabilité (comme en ... AFAIRE), puisque celui-ci est de dimension quatre⁷.

A.i Calcul des fonction d'amplification Afin d'étudier la stabilité linéaire des méthodes, les fonctions d'amplifications ont été numériquement.

B Étude de stabilité appliquée à l'équation de Nagumo

Nous allons particulariser la démarche suivante en la centrant sur l'équation de Nagumo. Cela va nous permettre de comprendre comme se comportent les méthodes ImEx sur ce problème particulier.

B.i Valeurs propres mises en jeu Comme expliqué en 3.1.1 l'équation présente deux opérateurs :

- ◊ La diffusion dont le spectre s'étend de $\frac{-1}{L^2}$ à $\frac{-1}{\Delta x^2}$ (où L est la taille du domaine discrétisé).
- ◊ La réaction dont le spectre balaie continûment $-k$ jusqu'à $2k$

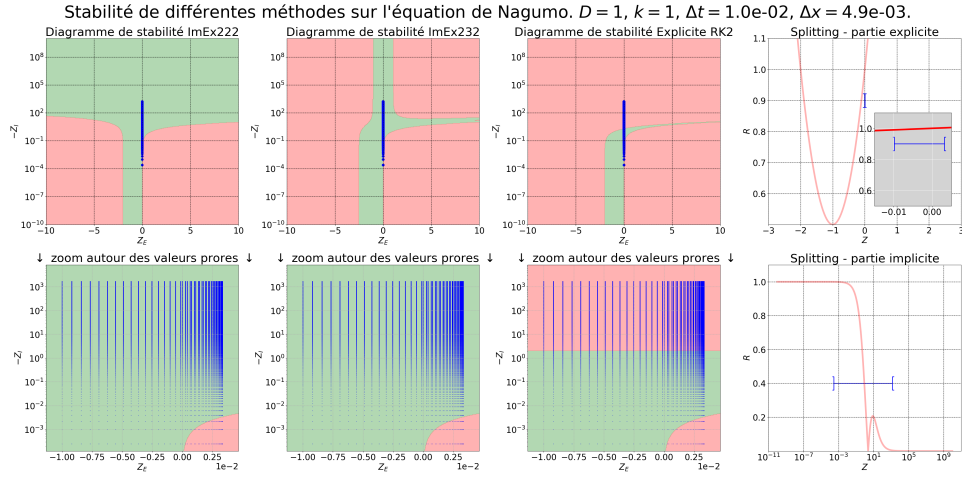
Pour restreindre l'analyse de stabilité il faut donc tracer le diagramme de stabilité des méthodes étudiées en prenant $Z_I \in \mathbb{R}^-$ et $Z_E \in [-k; 2k] \subset \mathbb{R}$ ce qui nous donne un espace à deux dimensions. Il faut ensuite placer des couples (Z_E, Z_I) correspondant. Lorsque l'on réalise ce travail nous trouvons les diagrammes suivant :

B.ii Résultats Le lecteur est invité à prendre un peu de temps pour comprendre la logique de ces graphiques car ils sont très éclairants. Ces diagrammes permettent d'analyser respectivement la stabilité de la méthode ImEx222, de la méthode ImEx232, ainsi qu'à titre de comparaison, la stabilité d'une méthode RKE d'ordre 2⁸ et d'une méthode de splitting⁹. Chaque colonne représente l'analyse d'une méthode différente. La première ligne présente le domaine de stabilité en fonction des indices spectraux $Z_E \in \mathbb{R}$ et $Z_I \in \mathbb{R}^-$. Les points bleus représentent les couples d'indices spectraux intervenant dans la résolution de l'équation de Nagumo pour les paramètres d'équation choisis (D et k) et les paramètres de discrétisation retenus (Δt et Δx). La seconde ligne n'est qu'un zoom de la première

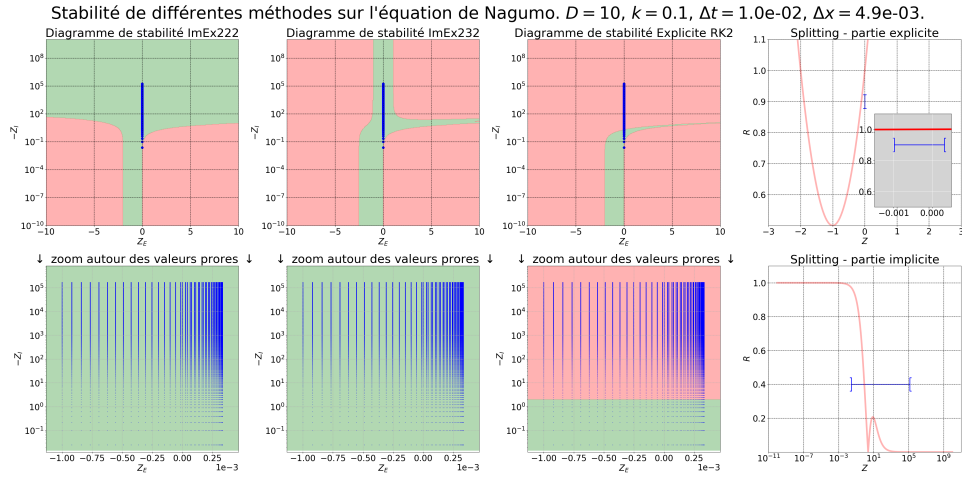
7. En effet la fonction de stabilité $\mathbb{R}\mathbb{C} \times \mathbb{C} \rightarrow \mathbb{R}$ et $\dim \mathbb{C} \times \mathbb{C} = 4$.

8. Celle apparaissant dans ImEx222.

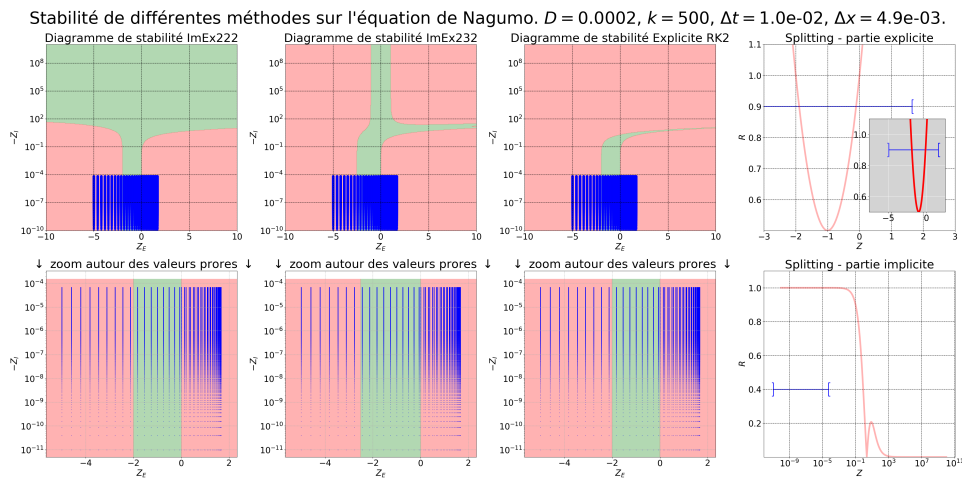
9. Où l'on utilise les méthodes implicites et explicites de la méthode ImEx222 mais dans un contexte de splitting de Strang.



(a) Cas standard
 $D=1, k=1, dt=1.0e-03, dx=2.4e-03$



(b) Cas diffusion plus raide, réaction moins raide
 $D=10, k=0.1, dt=1.0e-02, dx=4.9e-03$



(c) Cas diffusion moins raide, réaction plus raide
 $D=2e-4, k=500, dt=1.0e-02, dx=4.9e-03$

FIGURE 3.3 – Pour différents couples D et k , diagrammes de stabilité des méthodes ImEx et de référence sur l'équation de Nagumo.

autour de ces indices spectraux. La dernière colonne (splitting) présente une disposition différente, puisque les opérateurs sont totalement découplés. La première ligne correspond à la fonction de stabilité de la méthode explicite (avec un zoom autour des indices spectraux de la réaction). la seconde ligne représente la fonction de stabilité de la méthode implicite. Dans les deux cas, l'intervalle tracé en bleu représente la plage de valeurs d'indices spectraux balayés par chaque opérateur.

B.iii Analyse

Analyse générale Analysons les domaines de stabilité des figures en fig. 3.3, pour l'instant nous ignorons les marqueurs bleus sur les figures.

- ◊ **Méthode RKE2** : En troisième colonne, le diagramme de stabilité d'une méthode explicite naïve RKE2, sert de référence. Le domaine de stabilité accepte des valeurs propres négatives de magnitude deux, ce qui est résultat classique des méthodes Runge et Kutta explicites d'ordre deux. Ainsi domaine de stabilité s'étend jusqu'à -2 selon l'axe portant Z_E tant que la valeur propre Z_I est négligeable. De même le domaine de stabilité s'étend jusqu'à -2 selon Z_I tant que la valeur propre Z_E est négligeable. Enfin il y a une zone intermédiaire quand Z_E et Z_I sont tous les deux de l'ordre de l'unité¹⁰, où la raideur résultante est $Z_E + Z_I$.
- ◊ **Méthode ImEx232** : En observant la seconde colonne, nous constatons que la méthode ImEx232 maintient un domaine de stabilité restreint (jusqu'à -2) selon l'axe Z_E , mais selon l'axe Z_I , le domaine de stabilité s'est étendu considérablement. C'est logique puisque la valeur propre Z_E est explicitée, son domaine pris seul n'a évolué, et la valeur propre Z_I peut être très raide (très négative) puisque la méthode explicite l'opérateur lié à Z_I .
- ◊ **Méthode ImEx222** : Passant à la première colonne, le domaine de stabilité ImEx222 ressemble beaucoup à celui de l'ImEx232. Seulement, le domaine de stabilité s'élargit considérablement selon Z_E , pourvu que Z_I soit assez grand. Cette propriété est remarquable, cela signifie que la méthode traite couple les raideurs dans son traitement. Plus précisément, plus l'opérateur implicite est raide, plus l'opérateur explicite peut être raide¹¹.

Analyse selon les paramètres de l'équation k et D Analysons grâce aux graphiques fig. 3.3 la disposition des valeurs couples de valeurs propres mis en jeu par l'équation de Nagumo selon les paramètres k et D . Les paramètres de simulation : Δt et Δx sont fixés. Les jeux de valeurs choisis sont $(k, D) = (1, 1)$, $(k, D) = (0.1, 10)$, $(k, D) = (500, 2 \cdot 10^{-4})$. Le produit kD est maintenu égal à un, ainsi la vitesse de propagation est toujours la même. Ces couples de valeurs propres Z_E, Z_I mis en jeu par les opérateurs de l'équation sont tracés en bleus¹².

- ◊ **Cas standard**, $(k, D) = (1, 1)$ - fig. 3.3a :

Dans ce cas, la raideur de la diffusion (Z_I) déstabilise la méthode RKE2 (on voit que de nombreux couples de v.p. entrent dans les zones rouges quand Z_I augmente). Pour ces valeurs de $(\Delta x, \Delta t)$ cette méthode n'est donc pas viable. C'est tout à fait normal, les méthodes imposent des pas de temps très restrictifs sur les problèmes de diffusion. En revanche, les méthodes

10. Attention à l'échelle logarithmique.

11. Cette analyse est partiellement erronée, nous verrons pourquoi au prochain paragraphe.

12. Pour les Z_I le spectre est discret, pour Z_E , le spectre est continu, il a donc fallu échantillonner le long de l'axe Z_E

ImEx sont tout à fait stable puisque, comme constaté précédemment, le domaine de stabilité s'étend infiniment quand $Z_I \rightarrow -\infty$. Le point notable est que certains couples de valeurs propres tombent malgré tout dans une zone instable (en bas à gauche). Mais cela n'est pas un problème car il s'agit de couples de valeurs propres où la valeur propre de l'opérateur de réaction (Z_E) est positive. Donc la méthode n'est pas instable au sens où elle reflète simplement la dynamique explosive de la réaction. D'ailleurs si on se penche sur le graphique de la partie explicite du splitting, on constate qu'il y a une zone (correspondant à Z_E positive) où la fonction d'amplification est d'amplitude supérieure à un, le splitting reproduit donc fidèlement la dynamique de la réaction. Ce qui peut être un problème est l'inverse, pour les méthodes ImEx, il y a des couples de valeurs propres où Z_E est positif et où la fonction d'amplification est d'amplitude inférieure à un. Cela pourrait être un frein pour reproduire fidèlement la dynamique explosive de la réaction dans les zones concernées¹⁴.

♦ **Cas diffusion raide, réaction peu raide**, $(k, D) = (0.1, 10)$ - fig. 3.3b :

Ici, $D = 10$ donc toutes les valeurs propres liées à la diffusion sont multipliées par 10 par rapport au cas précédent. De fait la méthode RK2E de référence présente des instabilités pour encore plus de couples de valeurs propres n'est pas viable. Concernant les méthodes ImEx222 et ImEx232 elles sont stables, et cette fois-ci toutes les valeurs propres liées à la dynamique explosive de la réaction sont amorties.

♦ **Cas diffusion peu raide, réaction très raide**, $(k, D) = (500, 2 \cdot 10^{-4})$ - fig. 3.3c

Dans ce cas de figure, $k = 500$. La grande valeur du coefficient de réaction rend cette dernière très raide. Cela a pour effet de dilater selon l'axe des abscisses les indices spectraux puisque $Z_E \in [-500\Delta t, +1000\Delta t]$ alors que dans le cas $k = 1$: $Z_E \in [-\Delta t, 2\Delta t]$. Ici la méthode explicite au sein des ImEx n'est plus stable pour la réaction, ainsi toutes les méthodes deviennent instables. Le splitting également devient instable car il utilise aussi la méthode RK2E pour la réaction. Le fait que la méthode explicite de l'ImEx soit instable pour l'opérateur explicite peu sembler un obstacle infranchissable, cependant ce n'est pas si simple. Pour illustrer ce point, étendons l'analyse avec le cas spécial en fig. 3.4, dans ce cas la réaction est toujours raide $k = 500$ mais la diffusion est également très raide car $D = 500$ ¹⁵. Alors la méthode ImEx222 devient stable, comme vu en B.iii, plus l'opérateur traité implicitement est raide, plus la méthode permet à l'opérateur traité explicitement d'être raide. C'est un cas remarquable où le couplage intervenant au sein de la méthode ImEx la rend plus stable que le splitting!

Analyse selon les paramètres de simulation Δt et Δx

3.1.4 Étude de la convergence

3.1.5 Conclusion

13. Dans cette section, nous identifions valeurs propres λ et indices spectraux $z = \lambda\Delta t$ puisque le pas de temps Δt est maintenu constant. Cette identification permet de discuter directement en termes de raideur des opérateurs.

14. Il n'est pas évident d'avoir *a priori* la bonne intuition car peut être que la diffusion calme en quelque sorte le caractère explosif de la réaction et qu'alors une fonction d'amplification d'amplitude < 1 est normal... Restons prudent sur cette analyse.

15. Jusqu'ici, la vitesse de propagation était la même dans tous les scénarios puisque kD était maintenu constant. Dans le scénario présenté ici, ce n'est plus le cas

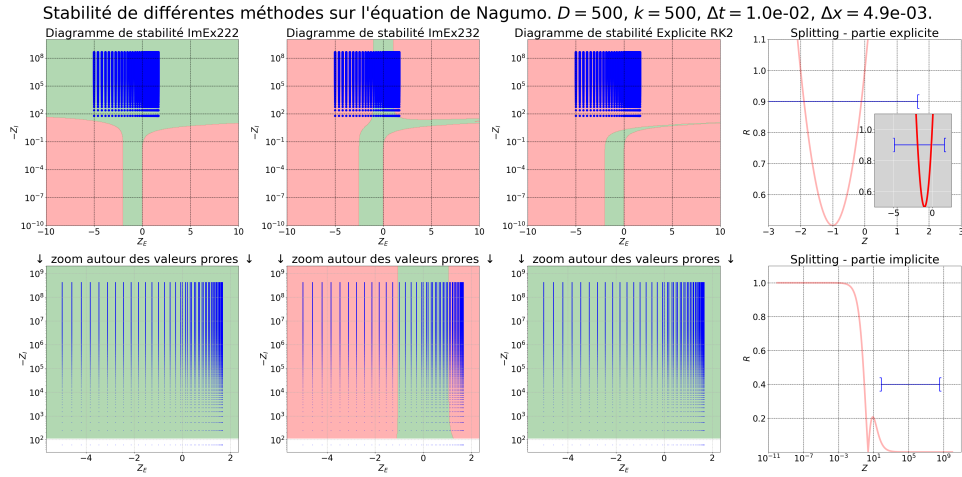


FIGURE 3.4 – Pour $k = 500$ et $D = 500$: diagrammes de stabilité des méthodes ImEx et de référence sur l'équation de Nagumo.

3.2 Impact de la multirésolution adaptative sur l'erreur d'une méthode des lignes

La multirésolution adaptative (MRA) a démontré une grande efficacité expérimentalement. Cependant, son impact sur la qualité des solutions obtenues n'est pas encore totalement compris mathématiquement. En [3], une étude de l'erreur introduite par la multirésolution adaptative a été menée. Cette étude se concentre sur les équations d'advection résolues par des schémas de type *one-step* [5] et compare l'équation équivalente des schémas avec MRA et celle des schémas sans MRA pour mettre en lumière les différences introduites par la multirésolution. La présente étude se place dans la continuité de cette démarche¹⁶ et suit un cheminement similaire pour déterminer, grâce aux équations équivalentes, l'impact de la MRA sur une équation de diffusion résolue par une méthode des lignes d'ordre deux. La différence est donc double : d'une part la nature de l'opérateur étudié est différent (diffusion et non advection), d'autre part le type de schéma utilisés est également différent (méthode des lignes, ne couplant pas les erreurs en espace et en temps contre les schémas *one-step* traitant d'un coup d'un seul les erreurs en temps et en espace.). L'objectif est d'évaluer l'impact de la MRA dans des contextes variés. D'abord, le problème cible est présenté ainsi que le schéma de référence. Par la suite, les équations équivalentes du schéma de référence et du schéma avec multirésolution adaptative sont évaluées puis analysées. Enfin, les résultats théoriques obtenus sont éprouvés expérimentalement.

3.2.1 Cadre de l'étude

A Problème cible

Nous cherchons à résoudre le problème de diffusion suivant :

$$\partial_t u = D \partial_{xx} u. \quad (3.20)$$

16. À un niveau plus modeste.

Nous ignorons les problématiques de conditions de bords.

B Méthode des lignes utilisée

Pour résoudre cette équation aux dérivées partielles, nous utilisons une méthode des lignes. D'abord un schéma volume fini pour la discrétisation spatiale menant à l'équation semi-discrétisée suivante :

$$\partial_t U(t) = \frac{D}{\Delta x} \left(\frac{U_{k+1} - 2U_k + U_{k-1}}{\Delta x} \right) \quad (3.21)$$

Puis une méthode de Runge und Kutta explicite d'ordre deux sur l'opérateur linéaire donne :

$$\begin{aligned} U_k^{n+1} = & U_k^n \\ & + D \frac{\Delta t}{\Delta x} \left(\frac{U_{k+1} - 2U_k + U_{k-1}}{\Delta x} \right) \\ & + D^2 \frac{\Delta t^2}{\Delta x^2} \left(\frac{U_{k+2} - 4U_{k+1} + 6U_k - 4U_{k-1} + U_{k-2}}{\Delta x^2} \right). \end{aligned} \quad (3.22)$$

Cela s'écrit sous la forme conservative suivante :

$$u_k^{n+1} = u_k^n + \lambda \left(\Phi_{k+1/2}^n - \Phi_{k-1/2}^n \right) \quad (3.23)$$

Avec :

$$\lambda = D \frac{\Delta t}{\Delta x^2} \quad (3.24)$$

$$\Phi_{k+1/2}^n = u_{k+1}^n - u_k^n + \frac{1}{2} \lambda (u_{k+2}^n - 3u_{k+1}^n + 3u_k^n - u_{k-1}^n), \quad (3.25)$$

$$\Phi_{k-1/2}^n = u_k^n - u_{k-1}^n + \frac{1}{2} \lambda (u_{k+1}^n - 3u_k^n + 3u_{k-1}^n - u_{k-2}^n). \quad (3.26)$$

C multirésolution adaptative

La multirésolution adaptative consiste à compresser le maillage, puis à effectuer les calculs sur le maillage compressé. Le schéma classique est le suivant :

1. Partir d'un état compressé au pas de temps n .
2. Calculer la solution au pas de temps $n + 1$
3. Compresser de nouveau selon un seuil de compression ε grâce à une transformée multiéchelle.

Lors de la compression, la transformée multiéchelle représente la solution sur plusieurs niveaux de détails, du plus global, au plus local. Plus le niveau est profond, c'est à dire plus il est local, moins les détails associés portent d'information. L'opération de compression est réalisée en supprimant en chaque cellules, les niveaux dont la valeur des détails passent sous un certain seuil¹⁷ [postelApprox]

Ce seuil ε n'est pas l'unique juge lors la compression, des heuristiques reposant sur la quantité d'information des détails de niveau supérieur sont utilisées pour ne pas seuiller systématiquement.

17. Typiquement $2^{\Delta l} \varepsilon$ où Δl

L'objectif est en quelque sorte d'anticiper le besoin de détails¹⁸. La plus connue est l'heuristique d'Ami Harten [10].

Plusieurs stratégies existent pour réaliser le calcul d'un pas de temps à l'autre. Généralement, on estime les quantités au temps $n + 1$ aux niveaux courants, à partir des quantités au niveau courant au temps n . Ensuite une opération de reconstruction-prédiction détermine le niveau de finesse requis de la solution au temps $n + 1$. Il est également possible de calculer les quantités du temps $n + 1$ au niveau courant, à partir des quantités au temps n **reconstruites à un niveau plus fin**. Bien que cela ait une faible efficacité computationnelle, cela réduirait les erreurs liées à la multirésolution selon la qualité du prédicteur employé comme discuté en [3]. Ici nous allons étudier théoriquement les erreurs dans un contexte similaire. Nous nous plaçons sur une cellule à un niveau de détail fixé, les flux sont calculés à partir de quantités reconstruites à un niveau de détails Δl plus fin. Le raisonnement et les ressources de calcul formel de [3] ont été d'une aide précieuse.

Calcul du flux au travers de Δl niveaux :

Lorsque l'on applique le procédé de multirésolution, étant donné une cellule à un niveau de détail donné l , on cherche à faire évoluer la valeur à l'étape n vers la valeur à l'étape $n + 1$. Pour ce faire, il faut évaluer les flux à partir des cellules voisines. Dès lors plusieurs choix s'offrent à nous. Où bien on utilise les cellules voisines à leurs niveaux courants, où bien on use de l'opérateur de reconstruction afin d'estimer les cellules voisines à des niveaux plus fins.

Dans un premier temps le stencil est choisi égal à 1. L'opérateur de prédiction d'un niveau à l'autre s'écrit alors :

$$\hat{u}_{2k}^{l+1} = +\frac{1}{8}u_{k-1}^l + u_k^l - \frac{1}{8}u_{k+1}^l, \quad (3.27)$$

$$\hat{u}_{2k+1}^{l+1} = -\frac{1}{8}u_{k-1}^l + u_k^l + \frac{1}{8}u_{k+1}^l. \quad (3.28)$$

Puis en notant $\hat{u}_{(\cdot)}^{l+\Delta l}$ cet opérateur de prédiction itéré au travers de Δl niveaux¹⁹ :

$$\begin{bmatrix} \hat{u}_{2^{\Delta l}k-2}^{(l+\Delta l)} \\ \hat{u}_{2^{\Delta l}k-1}^{(l+\Delta l)} \\ \hat{u}_{2^{\Delta l}k}^{(l+\Delta l)} \\ \hat{u}_{2^{\Delta l}k+1}^{(l+\Delta l)} \end{bmatrix} = \underbrace{\begin{bmatrix} +1/8 & 1 & -1/8 & 0 \\ -1/8 & 1 & +1/8 & 0 \\ 0 & +1/8 & 1 & -1/8 \\ 0 & -1/8 & 1 & +1/8 \end{bmatrix}}_{\text{Matrice de passage } P \text{ pour } s=1.}^{\Delta l} \begin{bmatrix} u_{k-2}^l \\ u_{k-1}^l \\ u_k^l \\ u_{k+1}^l \end{bmatrix} \quad (3.29)$$

En particulier, si la cellule étudiée est au niveau courant l alors on choisira d'aller approximer le flux au niveau le plus fin, c'est à dire avec $\bar{\Delta l} = \bar{l} - l$. Dès lors les flux approximés au niveau fins sont :

$$\hat{\Phi}_{k-1/2} = \hat{u}_{2^{\bar{\Delta l}}k}^{l+\bar{\Delta l}} - \hat{u}_{2^{\bar{\Delta l}}k-1}^{l+\bar{\Delta l}} + \frac{1}{2}\lambda \left(\hat{u}_{2^{\bar{\Delta l}}k+1}^{l+\bar{\Delta l}} - 3\hat{u}_{2^{\bar{\Delta l}}k}^{l+\bar{\Delta l}} + 3\hat{u}_{2^{\bar{\Delta l}}k-1}^{l+\bar{\Delta l}} - \hat{u}_{2^{\bar{\Delta l}}k-2}^{l+\bar{\Delta l}} \right), \quad (3.30)$$

$$\hat{\Phi}_{k+1/2} = \hat{u}_{2^{\bar{\Delta l}}(k+1)}^{l+\bar{\Delta l}} - \hat{u}_{2^{\bar{\Delta l}}(k+1)-1}^{l+\bar{\Delta l}} + \frac{1}{2}\lambda \left(\hat{u}_{2^{\bar{\Delta l}}(k+1)+1}^{l+\bar{\Delta l}} - 3\hat{u}_{2^{\bar{\Delta l}}(k+1)}^{l+\bar{\Delta l}} + 3\hat{u}_{2^{\bar{\Delta l}}(k+1)-1}^{l+\bar{\Delta l}} - \hat{u}_{2^{\bar{\Delta l}}(k+1)-2}^{l+\bar{\Delta l}} \right) \quad (3.31)$$

18. Même si la quantité d'information laisse entendre que certains détails pourraient être ignorés, l'intuition physique pose son veto et force certains détails à être conservés par précaution, par exemple si un front d'onde arrive.

19. Au sens où l'on applique le prédicteur à des données déjà issues d'une prédiction.

Cela s'écrit sous la forme matricielle suivante (utile pour utiliser les outils de calcul formel).

$$\hat{\Phi}_{k-1/2} = \begin{bmatrix} -\frac{\lambda}{2} & (\frac{3}{2}\lambda - 1) & (1 - \frac{3}{2}\lambda) & \frac{\lambda}{2} \end{bmatrix} \begin{bmatrix} +1/8 & 1 & -1/8 & 0 \\ -1/8 & 1 & +1/8 & 0 \\ 0 & +1/8 & 1 & -1/8 \\ 0 & -1/8 & 1 & +1/8 \end{bmatrix} \overline{\Delta l} \begin{bmatrix} u_{k-2}^l \\ u_{k-1}^l \\ u_k^l \\ u_{k+1}^l \end{bmatrix} \quad (3.32)$$

$$\hat{\Phi}_{k+1/2} = \begin{bmatrix} -\frac{\lambda}{2} & (\frac{3}{2}\lambda - 1) & (1 - \frac{3}{2}\lambda) & \frac{\lambda}{2} \end{bmatrix} \begin{bmatrix} +1/8 & 1 & -1/8 & 0 \\ -1/8 & 1 & +1/8 & 0 \\ 0 & +1/8 & 1 & -1/8 \\ 0 & -1/8 & 1 & +1/8 \end{bmatrix} \overline{\Delta l} \begin{bmatrix} u_{k-1}^l \\ u_k^l \\ u_{k+1}^l \\ u_{k+2}^l \end{bmatrix}. \quad (3.33)$$

Attention le schéma final est légèrement différent car il fait ici intervenir deux pas d'espace : Δx le pas au niveau le plus fin et $\tilde{\Delta x} = 2^{\Delta l} \Delta x$ le pas du niveau courant. Ainsi le schéma final est :

$$u_k^{n+1} = u_k^n + \frac{\lambda}{2^{\Delta l}} \left(\hat{\Phi}_{k+1/2}^n - \hat{\Phi}_{k-1/2}^n \right) \quad (3.34)$$

3.2.2 Les équations équivalentes

A Équation équivalente du schéma sans MRA

Le calcul de l'équation équivalente sans MRA donne :

$$\frac{\partial u}{\partial t} = +D \frac{\partial^2 u}{\partial x^2} + \Delta x^2 \frac{D}{12} \frac{\partial^4 u}{\partial x^4} - \Delta t^2 \frac{D^3}{6} \frac{\partial^6 u}{\partial x^6} - \Delta t^3 \frac{D^4}{24} \frac{\partial^8 u}{\partial x^8}. \quad (3.35)$$

Le schéma de base est donc bien d'ordre deux en espace et en temps. En supposant une relation de stabilité du type $\lambda = \frac{D \Delta t}{\Delta x^2}$:

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2} + \Delta x^2 \frac{D}{12} \frac{\partial^4 u}{\partial x^4} - \Delta x^4 \frac{D \lambda^2}{6} \frac{\partial^6 u}{\partial x^6} - \Delta x^6 \frac{D \lambda^3}{24} \frac{\partial^8 u}{\partial x^8}. \quad (3.36)$$

B Équation équivalente du schéma avec MRA

Dans le cas avec MRA, le logiciel de calcul formel donne l'équation équivalente :

$$\begin{aligned} \frac{\partial u}{\partial t} = & D \frac{\partial^2 u}{\partial x^2} \\ & - \frac{\Delta t}{2} D^2 (2^{2\Delta l} - 1) \frac{\partial^4 u}{\partial x^4} - \Delta t^2 \frac{D^3}{6} \frac{\partial^6 u}{\partial x^6} - \Delta t^3 \frac{D^4}{24} \frac{\partial^8 u}{\partial x^8} \\ & + 2^{2\Delta l} \frac{D \Delta x^2}{12} \frac{\partial^4 u}{\partial x^4} - 2^{2\Delta l} \frac{D \Delta l \Delta x^2}{4} \frac{\partial^4 u}{\partial x^4} \end{aligned} \quad (3.37)$$

Nous constatons que formellement le schéma est formellement d'ordre un en temps. Cela suggère donc que théoriquement, la multirésolution devrait faire perdre l'ordre de convergence temporelle de la méthode des lignes. Cependant en pratique la contrainte de stabilité impose une relation type

CFL $\lambda = \frac{D\Delta t}{\Delta x^2} < 1/2$. Cela masque la perte en ordre²⁰ puisque cela mène à l'équation équivalente :

$$\begin{aligned} \frac{\partial u}{\partial t} = & D \frac{\partial^2 u}{\partial x^2} \\ & + \Delta x^2 \left(\frac{2^{2\Delta l} D \lambda \frac{\partial^4 u}{\partial x^4}}{2} - \frac{2^{2\Delta l} D \Delta l \frac{\partial^4 u}{\partial x^4}}{4} - \frac{D \lambda \frac{\partial^4 u}{\partial x^4}}{2} + \frac{2^{2\Delta l} D \frac{\partial^4 u}{\partial x^4}}{12} \right) \\ & - \Delta x^6 \frac{D \lambda^3 \frac{\partial^8 u}{\partial x^8}}{24} - \Delta x^4 \frac{D \lambda^2 \frac{\partial^6 u}{\partial x^6}}{6} \end{aligned} \quad (3.38)$$

3.2.3 Comparaison

Un ordre de précision a été perdu en temps. Nous essayons à présent de comprendre quel mécanisme mène à cette baisse de performances. Ma démarche consiste à comparer les équations équivalentes avec et sans multirésolution, **avant** d'appliquer la procédure de Cauchy-Kovaleskaya.

.i Sans MRA L'équation modifiée sans multirésolution, avant procédure de Cauchy-Kovaleskaya est :

$$\begin{aligned} \frac{\partial u}{\partial t} = & D \frac{\partial^2 u}{\partial x^2} \\ & + \frac{1}{2} \underbrace{\left(D^2 \frac{\partial^4 u}{\partial x^4} - \frac{\partial^2 u}{\partial t^2} \right)}_{\substack{\text{Se compense par} \\ \text{la procédure de} \\ \text{Cauchy-Kovaleskaya}}} \Delta t + \frac{D}{12} \frac{\partial^4 u}{\partial x^4} \Delta x^2 - \frac{1}{24} \frac{\partial^4 u}{\partial t^4} \Delta t^3 - \frac{1}{6} \frac{\partial^3 u}{\partial t^3} \Delta t^2. \end{aligned} \quad (3.39)$$

La méthode est bien d'ordre un, car à l'ordre un : $\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2}$ et donc le terme $D^2 \frac{\partial^4 u}{\partial x^4} - \frac{\partial^2 u}{\partial t^2}$ se compense au cours de la procédure de Cauchy-Kovaleskaya.

.ii Avec MRA L'équation modifiée avec multirésolution, avant procédure de Cauchy-Kovaleskaya est :

$$\begin{aligned} \frac{\partial u}{\partial t} = & D \frac{\partial^2 u}{\partial x^2} \\ & + \frac{\Delta t}{2} \underbrace{\left(2^{2\Delta l} D^2 \frac{\partial^4 u}{\partial x^4} - \frac{\partial^2 u}{\partial t^2} \right)}_{\text{Ne se compensent plus}} - \frac{\Delta t^3}{24} \frac{\partial^4 u}{\partial t^4} - \frac{\Delta t^2}{6} \frac{\partial^3 u}{\partial t^3} + \frac{\Delta x^2}{12} (1 - 3\Delta l) 2^{2\Delta l} D \frac{\partial^4 u}{\partial x^4} \end{aligned} \quad (3.40)$$

Dans ce cas le terme en facteur du Δt ne se s'annule plus. En effet le terme $D^2 \frac{\partial^4 u}{\partial x^4}$ est devenu au cours de la reconstruction $2^{2\Delta l} D^2 \frac{\partial^4 u}{\partial x^4}$. En conséquence, la méthode perd un ordre de convergence temporel.

Ce mécanisme s'explique de la manière suivante : dans l'équation équivalente, le terme $\frac{\partial^2 u}{\partial t^2}$ apparaît indépendamment de la discrétisation spatiale²¹. La méthode des lignes initiale crée un terme *sur mesure* pour le compenser en approximant le terme spatial $D^2 \frac{\partial^4 u}{\partial x^4}$. Cependant au cours du processus de compression-reconstruction, cette approximation est entachée d'un facteur $2^{2\Delta l}$. En d'autres

20. Cela a été vérifié expérimentalement

21. Il émerge de la différence $u_k^{n+1} - u_k^n$ à k fixé.

termes le terme spatial construit pour compenser un terme temporel a été modifié par la multi-résolution, alors que le terme temporel lui n'est pas affecté par la multirésolution. Ainsi, les deux termes ne se compensent plus et l'ordre est perdu.

3.2.4 Conclusion

Il a été ici mis en lumière que la multirésolution-adaptative appliquée à méthode des lignes très simple peut théoriquement mener à un couplage des erreurs espace-temps polluant l'ordre initial de la méthode. En particulier l'étape de reconstruction-reconstruction altère des termes spatiaux qui ne compensent plus certaines erreurs temporelles.

3.2.5 Complément expérimental

3.2.6 Conclusion

Chapitre 4

Conclusion

Bibliographie

- [1] Assyr ABDULLE. “Fourth Order Chebyshev Methods with Recurrence Relation”. In : *SIAM Journal on Scientific Computing* 23.6 (2002), p. 2041-2054. DOI : [10.1137/S1064827500379549](https://doi.org/10.1137/S1064827500379549).
- [2] Uri M. ASCHER, Steven J. RUUTH et Raymond J. SPITERI. “Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations”. In : *Applied Numerical Mathematics* 25.2 (1997). Special Issue on Time Integration, p. 151-167. ISSN : 0168-9274. DOI : [https://doi.org/10.1016/S0168-9274\(97\)00056-1](https://doi.org/10.1016/S0168-9274(97)00056-1). URL : <https://www.sciencedirect.com/science/article/pii/S0168927497000561>.
- [3] BELLOTI et al. “Modified equation and error analyses on adaptative meshes for the resolution of evolutionary PDEs with Finite Volume schemes”. In : (2025).
- [4] M. BOUCHET. *Le laplacien discret 1D*. Notes de cours. Agrégation externe de mathématiques 2019-2020, Leçons 144, 155, 222, 226, 233. ENS Rennes, 2020. URL : <https://perso.eleves.ens-rennes.fr/~mbouc892/lapdisc1d.pdf>.
- [5] V. DARU et C. TENAUD. “High order one-step monotonicity-preserving schemes for unsteady compressible flow calculations”. In : *Journal of Computational Physics* 193.2 (2004), p. 563-594. ISSN : 0021-9991. DOI : <https://doi.org/10.1016/j.jcp.2003.08.023>. URL : <https://www.sciencedirect.com/science/article/pii/S0021999103004327>.
- [6] Max Pedro DUARTE. “Méthodes numériques adaptives pour la simulation de la dynamique de fronts de réaction multi-échelle en temps et en espace”. 2011ECAP0057. Thèse de doct. 2011. URL : <http://www.theses.fr/2011ECAP0057/document>.
- [7] Richard FITZHUGH. “Impulses and Physiological States in Theoretical Models of Nerve Membrane”. In : *Biophysical Journal* 1.6 (1961), p. 445-466. ISSN : 0006-3495. DOI : [https://doi.org/10.1016/S0006-3495\(61\)86902-6](https://doi.org/10.1016/S0006-3495(61)86902-6). URL : <https://www.sciencedirect.com/science/article/pii/S0006349561869026>.
- [8] E. HAIRER. “Order conditions for numerical methods for partitioned ordinary differential equations”. In : *Numerische Mathematik* 36.4 (1981), p. 431-445. ISSN : 0945-3245. DOI : [10.1007/BF01395956](https://doi.org/10.1007/BF01395956). URL : <https://doi.org/10.1007/BF01395956>.
- [9] Ernst HAIRER, Syvert P. NØRSETT et Gerhard WANNER. *Solving Ordinary Differential Equations I: Nonstiff Problems*. 2^e éd. T. 8. Springer Series in Computational Mathematics. Springer Berlin, Heidelberg, 1993, p. XV, 528. ISBN : 978-3-540-56670-0. DOI : [10.1007/978-3-540-78862-1](https://doi.org/10.1007/978-3-540-78862-1). URL : <https://doi.org/10.1007/978-3-540-78862-1>.

- [10] Ami HARTEN. "Adaptive Multiresolution Schemes for Shock Computations". In : *Journal of Computational Physics* 115.2 (1994), p. 319-338. ISSN : 0021-9991. DOI : <https://doi.org/10.1006/jcph.1994.1199>. URL : <https://www.sciencedirect.com/science/article/pii/S0021999184711995>.
- [11] James KEENER et James SNEYD. *Mathematical Physiology*. 1^{re} éd. Interdisciplinary Applied Mathematics. New York, NY : Springer, 1998, p. 281. ISBN : 978-0-387-22706-1. DOI : [10.1007/b98841](https://doi.org/10.1007/b98841).
- [12] Christopher A. KENNEDY et Mark H. CARPENTER. "Additive RungeKutta schemes for convectiondiffusionreaction equations". In : *Applied Numerical Mathematics* 44.1 (2003), p. 139-181. ISSN : 0168-9274. DOI : [https://doi.org/10.1016/S0168-9274\(02\)00138-1](https://doi.org/10.1016/S0168-9274(02)00138-1). URL : <https://www.sciencedirect.com/science/article/pii/S0168927402001381>.
- [13] L. PARESCHI et G. RUSSO. *Implicit-explicit Runge-Kutta schemes and applications to hyperbolic systems with relaxation*. 2010. arXiv : [1009.2757 \[math.NA\]](https://arxiv.org/abs/1009.2757). URL : <https://arxiv.org/abs/1009.2757>.
- [14] Marie POSTEL. "Approximations multiéchelles". Polycopié, Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie.
- [15] Louis REBOUL. "Development and analysis of efficient multi-scale numerical methods, with applications to plasma discharge simulations relying on multi-fluid models". 2024IPPAX134. Thèse de doct. 2024. URL : <http://www.theses.fr/2024IPPAX134/document>.

Annexes

.1 Annexe A : Titre de l'annexe

Contenu de la première annexe.

.2 Annexe B : Titre de l'annexe

Contenu de la deuxième annexe.