

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES 1

ÉCOLE DOCTORALE N° 601
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *Mathématiques et Interactions*

Par

Léopold TRÉMANT

Méthodes d'analyse asymptotique et d'approximation numérique de modèles dissipatifs multi-échelles

EDOs à variété centrale et modèles cinétiques

Thèse présentée et soutenue à « Lieu », le « date »

Unité de recherche : « voir liste sur le site de votre école doctorale »

Thèse N° : « si pertinent »

Rapporteurs avant soutenance :

Prénom NOM	Fonction et établissement d'exercice
Prénom NOM	Fonction et établissement d'exercice
Prénom NOM	Fonction et établissement d'exercice

Composition du Jury :

Attention, en cas d'absence d'un des membres du Jury le jour de la soutenance, la composition du jury doit être revue pour s'assurer qu'elle est conforme et devra être répercutée sur la couverture de thèse

Président :	Prénom NOM	Fonction et établissement d'exercice (à préciser après la soutenance)
Examineurs :	Prénom NOM	Fonction et établissement d'exercice
	Prénom NOM	Fonction et établissement d'exercice
	Prénom NOM	Fonction et établissement d'exercice
	Prénom NOM	Fonction et établissement d'exercice
	Prénom NOM	Fonction et établissement d'exercice
Dir. de thèse :	Prénom NOM	Fonction et établissement d'exercice
Co-dir. de thèse :	Prénom NOM	Fonction et établissement d'exercice (si pertinent)

Invité(s) :

Prénom NOM	Fonction et établissement d'exercice
------------	--------------------------------------

ACKNOWLEDGEMENT

Je tiens à remercier

I would like to thank. my parents..

J'adresse également toute ma reconnaissance à

....

TABLE OF CONTENTS

Introduction	7
Application de méthodes numériques générales	7
Application directe de schémas standards	9
Méthodes implicites	12
Méthodes de splitting	15
Méthodes générales pour systèmes raides	18
Exposition de l'état de l'art	18
Méthodes avec traitement spécifique de la partie raide	18
Notions de convergence	18
Contribution personnelle	19
 1 La moyennisation en bref	 21
1.1 Présentation d'une méthode	21
1.1.1 L'équation homologique	21
1.1.2 Définition d'une décomposition approchée	21
1.2 Contexte d'un problème autonome	21
1.2.1 Un résultat géométrique	21
1.2.2 Cas d'un opérateur linéaire	22
1.3 Aspect numérique	22
1.3.1 Définition d'un nouveau problème	22
1.3.2 Convergence uniforme	22
 2 Convergence uniforme pour un problème dissipatif	 23
2.1 Construction et résultat de précision uniforme	23
2.1.1 Hypothèses et définitions	23
2.1.2 Construction d'un développement asymptotique	23
2.1.3 Problème micro-macro et précision uniforme	24
2.2 Preuves des théorèmes	24
2.2.1 Bonne définition des morphismes, et leurs bornes	24

TABLE OF CONTENTS

2.2.2	Caractère bien posé du problème micro-macro	24
2.2.3	Précision uniforme avec les schémas exponentiels	24
2.3	Extension partielle à des EDP discrétisées	24
2.3.1	Équation de télégraphe	25
2.3.2	Relaxation hyperbolique	25
2.4	Résultats numériques	25
2.4.1	Applications directes	25
2.4.2	EDP discrétisées	26
2.4.3	Commentaires sur extensions directes	26
3	Discussion d'extension des résultats	27
3.1	Coût de calcul, erreurs d'arrondis, derivative-free, pullback	27
3.2	Autour de l'équation de télégraphe	27
A	Un développement double-échelle	29
B	Présentation de schémas numériques	31
C	Autour de notions géométriques	33
C.1	Algèbre de Lie	33
C.2	Géométrie de la moyennisation stroboscopique	33
	Bibliography	35

INTRODUCTION

Contextualisation

Modèles cinétiques, dynamique des populations

Annonce de l'équation

$$\begin{cases} \partial_t x^\varepsilon = a(x^\varepsilon, z^\varepsilon), & x^\varepsilon(0) = x_0, \\ \partial_t z^\varepsilon = -\frac{1}{\varepsilon} \Lambda z^\varepsilon + b(x^\varepsilon, z^\varepsilon), & z^\varepsilon(0) = z_0. \end{cases} \quad (1a)$$

(1b)

Passage en notation u^ε ,

$$\partial_t u^\varepsilon = -\frac{1}{\varepsilon} A u^\varepsilon + f(u^\varepsilon). \quad (2)$$

Présentation de ce qu'on entend par régime raide ou non-raide

Comportement de la solution

En un temps assez court (de l'ordre de ε), on rejoint un système de la forme

$$\begin{cases} \partial_t x^\varepsilon = a(x^\varepsilon, \varepsilon h^\varepsilon(x^\varepsilon)) \\ z^\varepsilon = \varepsilon h^\varepsilon(x^\varepsilon) \end{cases} \quad (3a)$$

(3b)

Théorème de variété centrale

Figure de simulation de modèle jouet

Application de méthodes numériques générales

On cherche à résoudre le problème (2) numériquement. Dans un premier temps, il faut présenter comment les méthodes de résolution standards se comportent lorsqu'on les adapte au problème en question. Par « méthodes standards », on entend des méthodes enseignées en études supérieures, et qui peuvent être trouvées dans les livres

Hairer Wanner 1
& 2 + GNI

de référence **REF**. Il s'agit de méthodes de résolution pour des équations différentielles ordinaires quelconques, qui ne prennent pas en compte la structure spécifique du problème.

Il est difficile de traiter le continu d'un point de vue numérique, donc on commence d'abord par *discrétiser* l'intervalle de temps en un nombre arbitraire $N \in \mathbb{N}^*$ d'intervalles.

Dessin avec la définition des t_n (voir commentaire)

Notion d'ordre et de solution

Pour des raisons pratiques, les mathématiciens aiment diviser l'intervalle de temps de manière uniforme comme sur le schéma ci-dessus. Néanmoins, une autre approche qui semble plus efficace serait de placer les points de discrétisation (t_n) de manière uniforme le long de la courbe solution

Figure pour le pas de temps adaptatif avec l'interprétation de distance uniforme sur la courbe (préciser que c'est une simplification, vu qu'avec un système affine on peut prendre le pas qu'on veut)

Dans les faits, on ne placerait pas les points avec une répartition uniforme sur la courbe, mais avec une répartition uniforme sur la courbe de sa dérivée à un certain ordre lié à l'ordre du schéma.

définition ordre
avant ??

Dans la suite de cette section, on restreint notre étude du comportement de certaines méthodes appliquées au problème particulier

$$\partial_t z(t) = -\frac{1}{\varepsilon} z(t) + \cos(t), \quad z(0) = 1. \quad (4)$$

Ce problème est un cas particulier de (1) avec $a(x, z) = 1$ et $b(x, z) = \cos(x)$. Une condition initiale $x(0) = 0$ permet d'assimiler x à la variable de temps t pour obtenir le problème ci-dessus. On en connaît la solution exacte

$$z(t) = e^{-t/\varepsilon} \left(z(0) - \frac{\varepsilon}{1 + \varepsilon^2} \right) + \frac{\varepsilon}{1 + \varepsilon^2} (\cos(t) + \varepsilon \sin(t)).$$

Dire que c'est comme résoudre un problème non-raide en temps long

Application directe de schémas standards

On cherche à résoudre le problème (2) numériquement. En termes simples, cela veut dire trouver une méthode générale pour obtenir la solution du problème. On va montrer que les méthodes « standards » sont inefficaces pour résoudre même le problème simplifié

$$\partial_t z = -\frac{1}{\varepsilon} z \quad (5)$$

avec une donnée initiale $z(0) \in \mathbb{R}$ quelconque. Par « méthodes standards », on entend des méthodes qui peuvent être trouvées dans le livre de référence **REF**. Il s'agit de méthodes de résolution pour des équations différentielles ordinaires quelconques, qui ne prennent pas en compte la structure spécifique du problème qui nous intéresse.

Il est difficile de traiter le continu d'un point de vue numérique, donc on commence d'abord par *discrétiser* l'intervalle de temps en un nombre arbitraire $N \in \mathbb{N}^*$ d'intervalles.

Dessin avec la définition des t_n (voir commentaire)

Pour le moment, on choisit de se restreindre à une discrétisation uniforme, c'est-à-dire que l'intervalle de temps $[0, T]$ est divisé en N intervalles de taille égale. De manière équivalente, on définit les points de séparation $(t_n)_{0 \leq n \leq N}$ avec

$$t_n = \frac{n}{N} T,$$

ou encore

$$t_{n+1} = t_n + \Delta t \quad \text{avec} \quad t_0 = 0 \quad \text{et} \quad \Delta t = \frac{T}{N}.$$

L'idée est maintenant d'obtenir une approximation de $u_n \approx u^\varepsilon(t_n)$ en chaque point. La méthode la plus directe est la méthode d'Euler : On a accès à la valeur initiale u_0 , et on peut ensuite calculer le prochain point par projection

Graphique : Tracer la droite qui part de z_n avec la pente $-z_n/\varepsilon$ pour placer le point z_{n+1} .

On obtient ainsi la *suite d'Euler* du problème (5) :

$$z_{n+1} = \left(1 - \frac{\Delta t}{\varepsilon}\right) z_n, \quad z_0 = z(0). \quad (6)$$

D'autres méthodes de projection existent, notamment les méthodes de Runge-Kutta qui impliquent des points intermédiaires, ou les méthodes multi-points qui utilisent les

valeurs en $n, n+1, \dots, n+s$ pour calculer le point en $n+s+1$. Le raisonnement qui suit s'applique (à une constante près) à toutes ces méthodes de calcul tant qu'elles sont *explicites*. Le schéma numérique est dit *stable* si la norme de la solution calculée ne croît pas,¹ i.e. si

$$\left| 1 - \frac{\Delta t}{\varepsilon} \right| \leq 1.$$

Cette condition est généralement interprétée comme une restriction sur le pas de temps Δt . Néanmoins, nous sommes aussi intéressés par le comportement du schéma par rapport à la variable de raideur ε . En considérant Δt fixé, il est clair qu'il faut prendre Δt plus petit que ε pour avoir une solution décroissante. Le coût de calcul dans la limite $\varepsilon \rightarrow 0$ évolue ainsi au moins en $\mathcal{O}(1/\varepsilon)$.

Vérifions que cette condition se remarque aussi sur l'erreur. On pose τ_n l'erreur dite de troncature du schéma entre t_n et t_{n+1} ,

$$\tau_n = z(t_{n+1}) - z(t_n) - \Delta t \partial_t z(t_n).$$

On applique la formule de Taylor avec reste intégral entre t_n et t_{n+1} , d'où

$$\tau_n = \int_{t_n}^{t_{n+1}} (t_{n+1} - t) \partial_t^2 z(t) dt,$$

et on peut borner cette erreur à l'aide d'une inégalité de Hölder,

$$|\tau_n| \leq \Delta t \int_{t_n}^{t_{n+1}} |\partial_t^2 z(t)| dt.$$

On remarque ensuite

$$z_{n+1} - z(t_{n+1}) = \left(1 - \frac{\Delta t}{\varepsilon} \right) (z_n - z(t_n)) - \tau_n$$

d'où, de proche en proche, et comme $z_0 = z(0)$,

$$z_{n+1} - z(t_{n+1}) = - \sum_{k=0}^n \left(1 - \frac{\Delta t}{\varepsilon} \right)^k \tau_{n-k}.$$

1. Cette définition atteint vite sa limite lorsqu'on considère des problèmes à solution croissante, par exemple $\partial_t x = x$. Heureusement, ce n'est pas le cas ici.

Encore par inégalité de Hölder (discrète, cette fois-ci), on obtient

$$|z_{n+1} - z(t_{n+1})| \leq \sum_{k=0}^n |\tau_k| \leq \Delta t \|\partial_t^2 z\|_{L^1}$$

avec la norme L^1 définie de manière canonique, c'est-à-dire $\|u\|_{L^1} \leq \int_0^T |u(t)| dt$.

En fait, cette dernière inégalité est vraie quel que soit le problème résolu (i.e. même si on modifie l'équation sur z), tant que la condition de stabilité (à adapter au problème) est vérifiée. En particulier, cela veut dire qu'on obtient la même borne d'erreur si on utilise la méthode d'Euler sur le problème plus complexe (2).

Dans le cas simplifié $\partial_t z = -z/\varepsilon$, on connaît la solution exacte du problème, et on peut calculer $\|\partial_t^2 z\|_{L^1} = \frac{1}{\varepsilon}(1 - e^{-T/\varepsilon})|z_0|$. Ainsi, la borne d'erreur devient pour tout indice n entre 0 et N ,

$$|z_n - z(t_n)| \leq \frac{\Delta t}{\varepsilon} |z_0|.$$

Il apparaît très clairement que le paramètre à ajuster pour modifier l'erreur est $\Delta t/\varepsilon$, pas juste Δt . On aurait pu espérer qu'en fixant ce rapport avec $\Delta t = \varepsilon$, l'erreur tende vers zéro dans la limite $\varepsilon \rightarrow 0$, mais ce n'est même pas le cas ici.

On aurait ainsi l'impression que pour obtenir une erreur de taille ε , il faudrait choisir $\Delta t = \varepsilon \delta$. Heureusement on peut exploiter la borne

$$|z_n - z(t_n)| \leq \sum_{k \geq 0} |\tau_k|$$

pour essayer de répartir presque uniformément l'erreur entre les τ_n en modulant le pas de temps à chaque itération. Ainsi, on remplacerait le schéma (6) par

$$z_{n+1} = z_n - \frac{\Delta t_n}{\varepsilon} z_n$$

avec Δt_n qui varie en fonction de l'itération. La formule (7) reste valide en remplaçant Δt par Δt_n , et c'est ça qu'on souhaite exploiter. Notamment, on peut séparer l'intervalle de temps en deux parties : pendant la phase transitoire (pas de temps Δt_0) et après la phase transitoire (pas de temps Δt_1). Ici, on détermine la fin de la phase transitoire au temps t_f tel que $\partial_t^2 z(t_f) = \mathcal{O}(1)$, soit $t_f = -\varepsilon \log(\varepsilon)$ en supposant $\varepsilon < 1$.

En posant $z_0 = 1$ pour simplifier, supposons qu'on souhaite avoir une erreur de $\delta/2$

à la fin de la période transitoire. Cela se traduit en

$$\Delta t_0 \int_0^{t_f} |\partial_t^2 z(t)| dt \leq \Delta t_0 \frac{1-\varepsilon}{\varepsilon} \leq \frac{\delta}{2}.$$

Ainsi on choisit $\Delta t_0 = \frac{\varepsilon \delta}{2(1-\varepsilon)}$. L'erreur après la phase transitoire est

$$\Delta t_1 \int_{t_f}^T |\partial_t^2 z(t)| dt \leq \Delta t_1,$$

donc si on souhaite avoir une erreur $\delta/2$ lors de cette seconde phase, on peut poser $\Delta t_1 = \min(\varepsilon, \delta/2)$ de sorte à être certain de respecter la condition de stabilité. Le coût associé à la phase transitoire est

$$\frac{t_f}{\Delta t_0} = \frac{2(1-\varepsilon) \log(1/\varepsilon)}{\delta},$$

et celui associé à la période qui suit est $\frac{T - \varepsilon \log(1/\varepsilon)}{\min(\varepsilon, \delta/2)}$. Ce coût est largement préférable au coût en $1/(\varepsilon \delta)$ qu'on aurait trouvé avec un pas de temps uniforme, mais il reste en $\mathcal{O}(1/\varepsilon)$ quand $\varepsilon \rightarrow 0$ à δ fixé à cause de la condition de stabilité.

Parler de méthodes de Chebichev (grand domaine de stabilité) + des estimations d'erreur

Méthodes implicites

Un moyen généralement invoqué pour stabiliser les problèmes raides comme celui-ci est l'utilisation de schémas *implicites*. Je n'ai pas étudié ces méthodes en détail, mais elles méritent d'être mentionnées et leur non-étude demande justification. On présente ici rapidement la méthode d'Euler implicite, ainsi que d'autres méthodes adaptées aux problèmes raides compilées dans **REF**. On discute enfin de leurs limites.

Hairer et al 2

Pour la méthode d'Euler implicite, au lieu de faire une projection directe (dans la méthode explicite), on cherche le point qui, en temps rétrograde, aurait donné le point précédent.

Schéma avec différentes lignes de niveau et les tangentes qui partent de différents u_{n+1} possibles

Cette méthode de projection rétrograde « directe » est appelée méthode d'Euler im-

plicite, et le schéma associé dans le cas du problème (5) est

$$z_{n+1} = z_n - \frac{\Delta t}{\varepsilon} z_{n+1},$$

ou encore

$$z_{n+1} = \frac{\varepsilon}{\Delta t + \varepsilon} z_n.$$

En comparant les figures **REF**, il est clair que cette méthode est en général bien plus coûteuse que les méthodes explicites de projection directe. Néanmoins, elle présente quelques propriétés sympathiques que nous devons présenter. Ce schéma est clairement stable pour tout Δt et tout ε , puisque le facteur multiplicatif est toujours inférieur à 1. Cependant, cela n'indique rien sur l'erreur associée au schéma. Calculons-la rapidement.

Labels

On pose τ_n l'erreur dite de troncature du schéma entre t_n et t_{n+1} ,

$$\tau_n = z(t_{n+1}) - z(t_n) - \Delta t \partial_t z(t_{n+1}).$$

En général on borne cette erreur par une application successive d'intégration par parties, inégalité triangulaire et enfin inégalité de Hölder, pour trouver

$$|\tau_n| \leq \int_0^{\Delta t} s \left| \partial_t^2 z(t_n + s) \right| ds \leq \Delta t \int_{t_n}^{t_{n+1}} \left| \partial_t^2 z(t) \right| dt. \quad (7)$$

En revanche cette approche n'exploite pas le caractère décroissant de $t \mapsto z(t)$, et d'ailleurs cette dernière borne serait aussi valide pour le schéma d'Euler explicite. Cette caractéristique peut être exploitée en s'intéressant directement à la définition de τ_n . En effet dans notre cas particulier, on a $z(t) = e^{-t/\varepsilon} z_0$ donc

$$\tau_n = \left(e^{-\Delta t/\varepsilon} - 1 + \frac{\Delta t}{\varepsilon} e^{-\Delta t/\varepsilon} \right) e^{-n\Delta t/\varepsilon} z_0,$$

ce qui peut se borner très grossièrement,

$$|\tau_n| \leq e^{-n\Delta t/\varepsilon} |z_0| \quad (8)$$

On remarque maintenant que l'erreur globale vérifie

$$\left(1 + \frac{\Delta t}{\varepsilon} \right) (z_{n+1} - z(t_{n+1})) = z_n - z(t_n) - \tau_n$$

d'où de proche en proche,

$$|z_{n+1} - z(t_{n+1})| \leq \sum_{k=0}^n \left(1 + \frac{\Delta t}{\varepsilon}\right)^{-k-1} |\tau_{n-k}|.$$

On applique brutalement une inégalité de Hölder sur la somme pour obtenir finalement

$$|z_{n+1} - z(t_{n+1})| \leq \left(1 + \frac{\Delta t}{\varepsilon}\right)^{-1} \sum_{k=0}^n |\tau_k|.$$

En exploitant (7) et (8), on obtient enfin

$$|z_{n+1} - z(t_{n+1})| \leq \frac{\varepsilon}{\varepsilon + \Delta t} \min \left(\Delta t \|\partial_t^2 z\|_{L^1}, \frac{|z_0|}{1 - e^{-\Delta t/\varepsilon}} \right).$$

Ici la norme L^1 est définie de manière canonique, c'est-à-dire $\|u\|_{L^1} \leq \int_0^T |u(t)| dt$. En l'occurrence, on connaît exactement $z(t) = e^{-t/\varepsilon} z_0$, donc on peut calculer $\|\partial_t^2 z\|_{L^1} = \frac{1}{\varepsilon}(1 - e^{-T/\varepsilon})z_0$, ce qui donne finalement pour tout n entre 0 et N ,

$$|z_n - z(t_n)| \leq \frac{\Delta t}{\varepsilon + \Delta t} |z_0| \min \left(1, \frac{\varepsilon/\Delta t}{1 - e^{-\Delta t/\varepsilon}} \right).$$

Grâce au premier argument du min, on est assuré que l'erreur est de la forme $C\Delta t$ pour tout $\Delta t < \Delta t_0$ avec C et Δt_0 des constantes indépendantes de ε . On remarque là une différence par rapport au comportement du schéma explicite, qui demandait $\Delta t \ll \varepsilon$. Le deuxième argument du min nous assure que l'erreur tend vers zéro quand ε tend vers zéro. On dit que ce schéma est L-stable.

détails

Tests avec Radau ?

Cependant, le coût de calcul à chaque pas du schéma est difficile à connaître *a priori*, puisqu'il y a un système de grande taille à inverser. Ce coût augmente avec la dimension et la non-linéarité du problème, et l'implémentation des méthodes demande un investissement certain. En conséquence, ces méthodes sont disponibles en « boîte noire » dans la boîte à outils de résolution d'équations différentielles `solve_ivp` de `scipy` en Python. Même pour les étudiants en analyse numérique, la résolution de tels systèmes est rarement (voire jamais) abordée. Il serait intéressant d'avoir accès à des méthodes tout aussi performantes mais plus facilement abordables, quitte à perdre en généralité.

Méthodes de splitting

Une autre approche courante est de séparer le problème en deux parties, une raide et une non-raide. Pour la bonne étude de la méthode, on écrit (4) sous la forme (2), i.e.

$$\partial_t u = -\frac{1}{\varepsilon} Au + f(u)$$

avec $A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ et $f(u) = \begin{pmatrix} 1 \\ \cos(x) \end{pmatrix}$. Le problème (4) se sépare en

$$\begin{cases} \partial_t u^{(1)} = -\frac{1}{\varepsilon} Au^{(1)}, \\ \partial_t u^{(2)} = f(u^{(2)}). \end{cases}$$

On note φ_t , $\varphi_t^{(1)}$ et $\varphi_t^{(2)}$ les t -flots associés aux problèmes en u , $u^{(1)}$ et $u^{(2)}$ respectivement. On remarque qu'il est simple de calculer $\varphi^{(1)}$ de manière exacte, et simple de calculer $\varphi^{(2)}$ de manière numérique. Cependant, elles sont mélangées dans φ , ce qui rend le flot du problème d'origine difficile à calculer. Ainsi, on est en droit de se poser la question : Est-il possible d'obtenir φ à partir de $\varphi^{(1)}$ et de $\varphi^{(2)}$?

La réponse à cette question est « partiellement », et on appelle cette approche le *splitting*. Le plus couramment utilisé est le splitting de Strang, qui s'écrit généralement

$$\varphi_t = \varphi_{t/2}^{(1)} \circ \varphi_t^{(2)} \circ \varphi_{t/2}^{(1)} + \mathcal{O}(t^3).$$

Il est obtenu par symétrie à partir du splitting de Lie $\Phi_t = \varphi_t^{(1)} \circ \varphi_t^{(2)}$, qui a une erreur $\varphi_t = \Phi_t + \mathcal{O}(t^2)$. Attention, Φ_t n'est pas un t -flot au même sens que φ_t , c'est la solution particulière d'une EDO dans l'espace des morphismes. Dans le cas où f est linéaire, on dispose de la formule de Baker-Campbell-Hausdorff (BCH),

REF BCH

$$\log(\Phi_t) = t \left(\frac{-1}{\varepsilon} A + f \right) - \frac{t^2}{2\varepsilon} [A, f] + \frac{t^3}{12\varepsilon^2} ([A, [A, f]] + \varepsilon [f, [A, f]]) + \dots$$

avec $[f, g] = fg - gf$ le commutateur de champs linéaires. Cette formule n'est valide que formellement, c'est-à-dire qu'elle ne converge pas forcément et que l'important est le terme général qui la génère. On peut la comparer à un développement de Taylor qu'on a poussé à un ordre « infini ». L'objet obtenu est une série entière en t , mais

celle-ci peut avoir un rayon de convergence nul. Néanmoins, les termes de la série ont du sens et on peut en prendre les premiers termes pour construire des approximations, bien qu'il faille être prudent avec la régularité de la fonction pour obtenir des résultats rigoureux.

Paragraphe utile ?

On remarque l'équivalence des propriétés suivantes :

- Le morphisme Φ_t est un flot ;
- Φ_t et φ_t coïncident en tout t ;
- Les champs A et f commutent.

Le morphisme Φ_t est un flot si et seulement si cette expression est linéaire en t , or ce n'est pas le cas en général. Le seul moyen que ce soit le cas est que A et f commutent. Ce résultat est aussi valide dans le cas non-linéaire, avec le commutateur

$$[f, g] = \partial_u f \cdot g - \partial_u g \cdot f$$

qui définit une algèbre de Lie sur les champs de vecteurs. Ceci correspond à une manière plus géométrique de considérer les champs de vecteurs qui revient à les associer à l'opérateur de transport $\mathcal{D}_f(g) = \partial_u g \cdot f$. Ainsi, $[f, g] = \mathcal{D}_g(f) - \mathcal{D}_f(g)$.

On définit l'erreur de splitting

$$\text{err}_{\text{Lie}} = \log(\Phi_t) - t \left(\frac{-1}{\varepsilon} A + f \right)$$

qui permet d'obtenir l'erreur de troncature du schéma par un passage à l'exponentielle. Il apparaît à partir de la formule de BCH que cette erreur est de taille t^2/ε , pas indépendante de ε . Néanmoins, on a déjà vu avec Euler implicite que l'accumulation d'erreur pouvait devenir indépendante de ε , grâce aux propriétés de décroissance de z . C'est le cas ici, et on peut borner l'erreur *indépendamment* de ε .

On cherche ensuite à améliorer la convergence du schéma : il serait agréable de pouvoir profiter d'une convergence à un ordre plus élevé. La méthode généralement considérée est le splitting de Strang :

$$\varphi_t \approx \varphi_{t/2}^{(1)} \circ \varphi_t^{(2)} \circ \varphi_{t/2}^{(1)}.$$

Cette méthode a l'avantage d'être d'ordre 2, et de présenter des propriétés géométriques sympathiques de par sa symétrie (elle génère d'ailleurs la méthode de Störmer-Verlett, voir Annexe **REF**). Néanmoins, il n'est pas clair qu'elle présente un bon comportement

Storm-Verl

lorsque la raideur augmente, i.e. lorsque ε diminue. En effet, un calcul de l'erreur de troncature donne

$$\varphi_t - \varphi_{t/2}^{(1)} \circ \varphi_t^{(2)} \circ \varphi_{t/2}^{(1)} = \frac{t^3}{12\varepsilon^2} ([A, [A, f]] + \varepsilon [[A, f], f]) + \mathcal{O}(t^4).$$

Ainsi, même si un $1/\varepsilon$ est compensé par la décroissance rapide de z , l'erreur ne sera pas indépendante de ε . On peut vérifier ce résultat de manière numérique.

Figure de convergence en fonction de Δt et ε , avec Lie en haut et Strang en bas.
Séparer x et z ?

Dans cette figure, on observe que le comportement de la solution est le bon attendu pour $\Delta t \ll \varepsilon$. Néanmoins, lorsqu'on trace l'erreur en fonction de ε , on voit qu'à Δt fixé, il y a toujours un seuil à partir duquel une réduction de ε entraîne une augmentation de l'erreur. Cette augmentation, de tendance prédite en $1/\varepsilon$, entraîne une *réduction d'ordre*, c'est-à-dire qu'il n'y a pas de constante d'erreur C telle que

$$\sup_{\varepsilon} \text{err} \leq C \Delta t^2.$$

Annexe B : Stormer-Verlett est un cas particulier de Strang mélangé à Euler explicite. En effet avec $\dot{q} = v$ et $\dot{v} = F(q)$,

$$\begin{aligned} v_{n+1/2} &= v_n + \frac{\Delta t}{2} F(q_n) \\ q_{n+1} &= q_n + \Delta t v_{n+1/2} \\ v_{n+1} &= v_{n+1/2} + \frac{\Delta t}{2} F(q_{n+1}). \end{aligned}$$

Ce qui revient à séparer le système en $\partial_t(q, v) = (0, F(q))$ et $\partial_t(q, v) = (v, 0)$.

En particulier, on dit le schéma préserve le comportement asymptotique, ce qui se traduit en *asymptotic preserving* (AP) en anglais. Cette notion reviendra dans la prochaine section.

Rq : On peut obtenir une approximation en tout point $t \in [0, T]$ par interpolation. Le sujet de cette thèse n'est pas l'étude de schémas numériques ou celle de méthodes d'interpolation, donc nous éviterons de fournir des détails superflus sur ces sujets. On évaluera la qualité d'une approximation en fonction de la valeur aux points.

Illustration rapide sur méthode d'Euler

Démonstration d'erreur en $\Delta t/\varepsilon$

Distinction entre convergence et stabilité

Méthodes générales pour systèmes raides

Bouquin de Hairer, Lubich, Wanner

Présentation rapide de méthodes stiff (implicites)

Discussion du coût du pas de temps adaptatif

Splitting + simulations

Foreshadowing : ordre des opérations pour AP et splitting exact pour formes normales

Exposition de l'état de l'art

Cette section observe le comportement de méthodes numériques usuelles lorsqu'on les applique au système.

Méthodes avec traitement spécifique de la partie raide

IMEX-BDF, convergence théorique

expRK, convergence théorique

Rq : Lawson et IMEX-RK, pas particulièrement étudiées

Test des méthodes proches et loin de l'équilibre

Notions de convergence

Def : AP (faire un diagramme)

Def : UA restreint

Def : UA tout court (trouver un diagramme)

Introduction de la norme des schémas exponentiels (notion de norme relative)

Contribution personnelle

Résultat de convergence uniforme avec IMEX-BDF ou expRK

Explication de la méthode à détailler (sans prendre le crédit) :

- Lien avec moyennisation
- Formes normales, splitting quasi-exact
- Conservation exacte du défaut

Ouverture au cinétique

Annonce plan

LA MOYENNISATION EN BREF

Lors de ma troisième année de thèse, j'ai eu l'occasion d'un peu plus me pencher sur les méthodes de moyennisation, et notamment de rédiger un mini-article compilant certains résultats du sujet. Une partie est ici, l'autre sera présentée en Annexe _____

quelle annexe ?

1.1 Présentation d'une méthode

1.1.1 L'équation homologique

Dérivation de l'équation homologique

Distinction entre averaging standard et stroboscopique

1.1.2 Définition d'une décomposition approchée

Relation de récurrence et résultat sur les bornes

Discussion autour de la méthode pour les résultats (boules imbriquées, estimations de Cauchy...)

1.2 Contexte d'un problème autonome

$$\partial_t y = \frac{1}{\varepsilon} G(y) + K(y) \quad (1.1)$$

On décompose

$$y^\varepsilon(t) = \Omega_{t/\varepsilon}^\varepsilon \circ \Psi_t^\varepsilon \circ (\Omega_0^\varepsilon)^{-1} \quad (1.2)$$

1.2.1 Un résultat géométrique

Ω est un flot, et commute avec Ψ .

1.2.2 Cas d’un opérateur linéaire

Averaging standard : crochets de Lie

Formes normales

1.3 Aspect numérique

Si on ne connaît pas le défaut, il est clair qu’on peut calculer numériquement

$$\partial_t v^{[n]} = F^{[n]}(v^{[n]}) \quad (1.3)$$

et alors

$$u^\varepsilon(t) = \Phi_{t/\varepsilon}^{[n]}(v^{[n]}(t)) + \mathcal{O}(\varepsilon^{n+1}) \quad (1.4)$$

Mais en fait si on montre ça, on peut montrer mieux (en supposant un peu de régularité). Section basée principalement sur l’article avec Gilles.

1.3.1 Définition d’un nouveau problème

Micro-macro ou pullback

Raideur “retardée”

1.3.2 Convergence uniforme

Résultat de convergence uniforme

Présentation de schémas intégraux et composition de schémas

CONVERGENCE UNIFORME POUR UN PROBLÈME DISSIPATIF

Ce chapitre reprend un article à paraître dans *Mathematics of Computation*, intitulé

A uniformly accurate numerical method for a class of dissipative systems

Dans cet article, on ...

2.1 Construction et résultat de précision uniforme

2.1.1 Hypothèses et définitions

Hyp : A diagonale avec des vap entières

Hyp : f analytique autour de $(x,0)$

Justification de l'hypothèse avec le théorème de variété centrale

Introduction des ensembles \mathcal{K}_ρ

Def : R + borne sur f et ses dérivées

Def : normes

2.1.2 Construction d'un développement asymptotique

Équation homologique

Définition de F et du projecteur

Relation de récurrence avec condition de fermeture

Def : défaut η

Thm : Bonne définition des morphismes et bornes

2.1.3 Problème micro-macro et précision uniforme

Obtention du problème micro-macro

Thm : problème micro-macro bien posé avec dérivées bornées

Rq : calcul de la donnée initiale

Présentation des schémas exponentiels

Def : norme associée aux schémas

Thm : précision uniforme :D

Rq : adaptation du résultat aux schémas IMEX-BDF

2.2 Preuves des théorèmes

2.2.1 Bonne définition des morphismes, et leurs bornes

Définition des morphismes périodiques

Filtrage des morphismes périodiques

Prop : hypothèses de UA périodique vérifiées

Thm de bonnes propriétés sur les morphismes périodiques

Principe du maximum pour conclure la preuve

2.2.2 Caractère bien posé du problème micro-macro

Caractère bien posé sur v : $v(0)$ est gentil et $v(t)$ aussi par Gronwall. $\Omega(v)$ est okay

Borner le terme linéaire L , puis Gronwall avec la source pour w bien posé

2.2.3 Précision uniforme avec les schémas exponentiels

Séparation entre v et w

Partie v bornée classique, passage à norme modifiée grâce à Ω

Partie w application directe des bornes de Hochbruck et Ostermann

2.3 Extension partielle à des EDP discrétisées

Présentation rapide du système

Disclaimer comme quoi on les regarde post-discrétisation

2.3.1 Équation de télégraphe

Transformation en Fourier

Tentative directe d'obtenir la variable en $-1/\varepsilon \rightarrow$ échec

Relaxation pour obtenir la variable z , montrer que ça va mieux

Développement à l'ordre zéro, montrer que ça se passe okay grâce à $\exp -tA/\varepsilon$

Passage à l'ordre 1, observation comme quoi $-A/\varepsilon + F^{[1]}$ est mal posé

Relaxation dans le changement de variable, montrer que tout se passe mieux

Eq : $\eta^{[1]}$

Prop : si les fréquences sont bornées, on peut faire du num tranquille et l'erreur est indépendante de ε

Rq : les fréquences bornées c'est plus une CFL qu'un souci avec ε

2.3.2 Relaxation hyperbolique

Présentation du système + condition de stabilité

Rq : relaxation indépendante de l'espace

Volumes finis avec Upwind, avec la variable z naturelle

Développements ordre 1 parachutés avec relaxation (justif de télégraphe)

Rq : passage au continu

Commentaire sur le coût de résolution (indépendant de ε malgré la relaxation)

Eq : défauts $\eta^{[0]}$ et $\eta^{[1]}$

2.4 Résultats numériques

2.4.1 Applications directes

Problème jouet lentement oscillant :

Exposition, dev à l'ordre 1 puis ordre 2

Rq : on peut retrouver la variété centrale

Rappel pb micro-macro, et commentaire sur le besoin de calcul explicite de L pour l'arrondi

Problème inspiré de l'hyperbolique :

Exposition, passage aux variables x, z

Eq : dev à l'ordre 1, ordre 2 trop lourd

Choix de la fonction g et de la donnée initiale

Résultats :

Figures : err. sup. sur ε à gauche pour syst. original (ERK2), miMa 1 (ERK2) et miMa 2 (ERK3), et err. en fonction de ε à droite (miMa 2 ERK3), pour différents Δt

=> Mise en avant et description du phénomène de réduction d'ordre

=> Observation de l'ordre de convergence grâce au problème micro-macro, ordres 1 et 2

Ajouter simus avec IMEX-BDF

2.4.2 EDP discrétisées

Télégraphe : rappel de l'équation + donnée initiale

Résultats : réduction d'ordre, discussion sur le côté "erreur uniforme"

Relaxation hyperbolique : rappel de l'équation + donnée initiale

Résultats avec disclaimer qu'on regarde pas l'erreur en espace

Figures : même format que pour les EDOs mais on s'arrête à miMa 1 (forcément)

2.4.3 Commentaires sur extensions directes

Dans la discussion

DISCUSSION D'EXTENSION DES RÉSULTATS

3.1 Coût de calcul, erreurs d'arrondis, derivative-free, pullback

Coût de calcul avec les non-linéarités

Difficulté du calcul de la relaxation pour certains schémas

Gain d'ordre avec $z(0) = 0$ (figure avec err sup sur ε)

Donner une clé pour le gain d'ordre : $v_z(0) = \mathcal{O}(\varepsilon)$

3.2 Autour de l'équation de télégraphe

UN DÉVELOPPEMENT DOUBLE-ÉCHELLE

Résultats du stage et du début de thèse

PRÉSENTATION DE SCHÉMAS NUMÉRIQUES

Méthodes composées
Tableaux de Butcher

AUTOUR DE NOTIONS GÉOMÉTRIQUES

C.1 Algèbre de Lie

C.2 Géométrie de la moyennisation stroboscopique

Sources primaires

ASTAIR, Mickey, *La pendule et le fou*, Paris : Firgoult, 1895.

BICON, Pierre, *De la memoire*, Paris : Serigraph, 1901.

DENYS, Frédéric et Hector GELLEE, « Voyages en Anatolie », in : *Revue des colonies* 13 (1905), p. 45-85.

DOULE, Eloise et Marie BADY, « Les castes dans les Indes Britanniques », in : *Revue des pays lointains* 15.7 (1887), p. 223-258.

HISTRION, Pierre et Jean-Marie GASTON, « Les phrases du début à la fin », in : *Revue internationale de linguistique* 50 (1925), p. 52-98.

KOHLER, Hans, *L'alsace, un paradis financier*, Paris : Hopla, 1903.

LAURENT, Jean, *Constructions et ouvrages d'art*, Paris : Les arts et métiers, 1925.

SBIRRE, François et Jacques HELOI, « Etudes épidémiologiques sur la paresse ambiante », in : *Revue du bon vieux temps* 23.4 (1906), p. 53-102.

SIMON, Zoé, *As-tu pensé à fermer le gaz ?*, Paris : Editions de l'oubli, 1903.

WALTHER, Gusty, *Controverses sur le soleil*, Paris : Editeur local, 1897.

— *Dialogues sur la pluie et le beau temps*, Paris : Editeur local, 1896.

Sources secondaires

AIMARRE, Jean, *Crise de foie*, Paris : Editions du Pénitent, 1999.

ASPHANE, Pierre et Jean-Marie VIGOUT, « De novembre à décembre », in : *Revue alsacienne du temps qu'il fait* 45 (2001), p. 23-89.

DESFOIS, Marie, *Les glaçons*, Paris : Grazmick, 1998.

DROCHER, Gérard, *Partir et revenir*, Paris : Novrebond, 2006.

FALCON, Alphonse et Gabriel SERVAIS, « Les avatars de la biologie », in : *Biomolle* 22 (1985), p. 65-97.

GEORGES, Christine et Estelle DUPUYS, « La blondeur décolorée. Histoire de la stupidité », in : *Revue des brunes* 12 (1974), p. 125-184.

HENRY, Michel et Philippe KHRON, « Le destin et l'avenir », in : *Histoire et partage* 2 (1956), p. 21-35.

HOCHET, Eric, *Histoire d'un lac*, Paris : Van Dame, 2003.

KANTROP, Julie, *Le loup et ses proies*, Paris : Nuit Bleue, 2005.

Titre : titre (en français).....

Mot clés : de 3 à 6 mots clefs

Résumé : Eius populus ab incunabulis primis ad usque pueritiae tempus extremum, quod annis circumcluditur fere trecentis, circummurana pertulit bella, deinde aetatem ingressus adultam post multiplices bellorum aerumnas Alpes transcendit et fretum, in iuvenem erectus et virum ex omni plaga quam orbis ambit inensus, reportavit laureas et triumphos, iamque vergens in senium et nomine solo aliquotiens vincens ad tranquilliora vitae discessit. Hoc immaturo interitu ipse quoque sui pertaesus excessit e vita aetatis nono anno atque vicensimo cum quadriennio imperasset. natus apud Tuscos in Massa Vaternensi, patre Constantio Constantini fratre imperatoris, matreque Galla. Thalassius vero

ea tempestate praefectus praetorio praesens ipse quoque adrogantis ingenii, considerans incitationem eius ad multorum augeri discrimina, non maturitate vel consiliis mitigabat, ut aliquotiens celsae potestates iras principum molliverunt, sed adversando iurgandoque cum parum congrueret, eum ad rabiem potius evibrabat, Augustum actus eius exaggerando creberrime docens, idque, incertum qua mente, ne lateret adfectans. quibus mox Caesar acrius efferatus, velut contumaciae quoddam vexillum altius erigens, sine respectu salutis alienae vel suae ad vertenda opposita instar rapidi fluminis irrevocabili impetu ferebatur. Hae duae provinciae bello quondam piratico catervis mixtae praedonum.

Title: titre (en anglais).....

Keywords: de 3 à 6 mots clefs

Abstract: Eius populus ab incunabulis primis ad usque pueritiae tempus extremum, quod annis circumcluditur fere trecentis, circummurana pertulit bella, deinde aetatem ingressus adultam post multiplices bellorum aerumnas Alpes transcendit et fretum, in iuvenem erectus et virum ex omni plaga quam orbis ambit inensus, reportavit laureas et triumphos, iamque vergens in senium et nomine solo aliquotiens vincens ad tranquilliora vitae discessit. Hoc immaturo interitu ipse quoque sui pertaesus excessit e vita aetatis nono anno atque vicensimo cum quadriennio imperasset. natus apud Tuscos in Massa Vaternensi, patre Constantio Constantini fratre imperatoris, matreque Galla. Thalassius vero

ea tempestate praefectus praetorio praesens ipse quoque adrogantis ingenii, considerans incitationem eius ad multorum augeri discrimina, non maturitate vel consiliis mitigabat, ut aliquotiens celsae potestates iras principum molliverunt, sed adversando iurgandoque cum parum congrueret, eum ad rabiem potius evibrabat, Augustum actus eius exaggerando creberrime docens, idque, incertum qua mente, ne lateret adfectans. quibus mox Caesar acrius efferatus, velut contumaciae quoddam vexillum altius erigens, sine respectu salutis alienae vel suae ad vertenda opposita instar rapidi fluminis irrevocabili impetu ferebatur. Hae duae provinciae bello quondam piratico catervis mixtae praedonum.