Développement et analyse de méthodes adaptatives pour les équations de transport

Martin Campos Pinto

UNIVERSITE PIERRE & MARIE CURIE
LA SCIENCE A PARIS

École doctorale Sciences Mathématiques de Paris Centre
UFR 921

Développement et analyse de schémas adaptatifs pour les équations de transport

THÈSE

présentée et soutenue publiquement le 18 novembre 2005

pour l'obtention du

Doctorat de l'université Pierre et Marie Curie – Paris 6
(spécialité mathématiques appliquées)

par

Martin Campos Pinto

Composition du jury

Albert Cohen	*Directeur de thèse*
Wolfgang Dahmen	
Bruno Després	
Yvon Maday	*Président du jury*
Sylvie Mas Gallic	
Valerie Perrier	*Rapporteur*
Eric Sonnendrücker	*Rapporteur*

Laboratoire Jacques-Louis Lions — UMR 7598

Mis en page avec la classe thloria.

Remerciements

En premier lieu, je voudrais remercier Albert Cohen d'avoir accepté de diriger ma thèse, et de l'avoir fait avec autant de disponibilité, d'attention et d'exigence. Grâce à la qualité de son regard mathématique, travailler sous sa direction aura été à la fois un réel honneur et un très grand plaisir. J'ose donc espérer qu'il trouvera dans ces quelques pages une forme de remerciement.

Je voudrais également remercier Eric Sonnendrücker de la façon la plus forte pour m'avoir invité aussi tôt à présenter mes travaux dans le cadre des ateliers de l'irma à Strasbourg, et pour avoir lancé mes recherches sur le semi-lagrangien en m'invitant notamment au cemracs de Luminy à l'été 2003. Pour avoir accepté, enfin, de relire ma thèse et d'en être rapporteur.

Je voudrais aussi remercier Valérie Perrier d'avoir accepté d'être rapporteur, pour l'extrême finesse et la précision de sa relecture.

Je voudrais remercier Wolfgang Dahmen pour les recherches passionnantes menées en sa compagnie, et l'intérêt qu'il porte à mon travail en m'acceptant dans son équipe à Aix-la-Chapelle.

Je voudrais remercier Yvon Maday de m'avoir accueilli dès le premier jour comme un membre à part entière du laboratoire Jacques-Louis Lions, car son optimisme et son énergie ont été pour moi une forme d'encouragement continue.

Je voudrais remercier Bruno Després et Sylvie Mas-Gallic pour l'intérêt porté à ma thèse en acceptant d'en être les juges.

Je voudrais remercier Pencho Petrushev et Ronald DeVore pour l'extrême gentillesse avec laquelle ils m'ont accepté comme collaborateur.

Je voudrais remercier Simon Masnou, François Murat et Luc Tartar pour le temps important passé à répondre à mes questions de novice au sujet des fonctions de variation et de courbure totale bornée.

Je voudrais remercier Marie Postel et Sidi Mahmoud Kaber pour l'accueil chaleureux qu'ils m'ont réservé dans leur équipe 'multi-échelle' dès mon arrivée au laboratoire, et pour le temps précieux passé à me transmettre leur savoir.

Je voudrais remercier Michel Mehrenberger pour la patience avec laquelle il m'a supporté lors de séances de travail souvent tardives, et l'acuité toujours vive de ses critiques.

Je voudrais remercier l'ensemble des membres du laboratoire - thésards et permanents - pour ces années passées en si bonne compagnie, et tout particulièrement ceux dont la bonne humeur et l'écoute attentive m'ont permis d'avancer dans les terres arides de la recherche.

Je voudrais remercier mes amis, pour leur présence véritable.

Je voudrais remercier mes parents, pour leur amour qui vit en moi.

Je voudrais remercier ma femme Clémentine, pour mon amour qui vit en elle.

Je voudrais remercier mes enfants à venir, pour le jour où ils seront fiers de leur papa.

A Clémentine.

Résumé

Les résultats présentés dans cette thèse portent sur l'approximation adaptative de deux problèmes de transport non-linéaire : le système de Vlasov-Poisson et les lois de conservation scalaires. Pour le premier, et dans une approche semi-lagrangienne, on a proposé un schéma adaptatif original à base d'éléments finis hiérarchiques où l'évolution des maillages est réalisée par une étape de prédiction très simple suivie d'une étape de correction plus classique. En introduisant la notion de courbure totale pour étendre la semi-norme $W^{2,1}(\mathbb{R}^2)$ aux fonctions affines par morceaux, on a alors établi une estimation d'erreur a priori prouvant la convergence de ce schéma en distance L^∞, et donné des éléments de preuve concernant sa complexité optimale. Les lois de conservations scalaire ne pouvant être approchées en distance L^∞, on a considéré leur analyse en distance uniforme de Hausdorff, moins répandue bien que plus géométrique. Après avoir montré que les solutions de ces équations étaient stables vis-à-vis de cette distance, on a établi un résultat d'approximation adaptative d'ordre élevé.

Abstract

This thesis focuses on adaptive approximation of two nonlinear transport problems, namely the Vlasov-Poisson system and the scalar conservation laws. In a semi-lagrangian approach, we propose a new adaptive scheme for the first one, in which the mesh is, at each time step, first predicted in a very simple way, then corrected by a classical algorithm. In order to extend the $W^{2,1}(\mathbb{R}^2)$ semi-norm to piecewise affine functions, the notion of total curvature is introduced and employed in a rigorous analysis to obtain a priori error estimates that establish the convergence of this scheme in the L^∞ metric, while a partial result of complexity is proposed. As scalar conservation laws may not be approximated in the same metric, we consider the uniform Hausdorff distance which appears as a natural substitute for the L^∞ one, and show that the solutions are stable with respect to this distance. Equipped with this new result, we prove a high order adaptive approximation theorem for these equations.

Table des matières

There's more to life than sitting around in the sun in your underwear playing the clarinet.

(Woody Allen)

Introduction

Les travaux réalisés au cours de cette thèse ont eu pour principal objet l'approximation adaptative de deux problèmes de transport non-linéaire, à savoir le système de Vlasov-Poisson et les lois de conservation scalaires.

Lors d'une simulation numérique, la précision des calculs dépend principalement, une fois fixé le schéma général de résolution, de deux ingrédients a priori indépendants : la finesse de la discrétisation et la régularité des solutions. Or, dans de nombreux problèmes rencontrés en calcul scientifique (on pense notamment aux phénomènes d'ondes de choc et de tourbillons en hydrodynamique ou en acoustique), il arrive que cette régularité soit fortement non-uniforme. Dans une onde de choc, par exemple, la solution est discontinue à l'endroit du choc mais généralement régulière ailleurs. Il est donc naturel de vouloir que la finesse de la discrétisation employée par le calcul soit également non-uniforme, de façon à être en "bonne adéquation" avec la régularité locale des solutions. On évoquera à ce propos une spécificité du transport non-linéaire, dans lequel les singularités sont susceptibles de d'évoluer de façon complexe, leur position, leur forme ou leur nature même pouvant changer rapidement au cours d'une simulation. Pour résumer, on dira donc que *les méthodes adaptatives étudiées dans cette thèse se distinguent par leur souci d'exploiter au mieux les ressources de calcul pour approcher avec une précision donnée des solutions pouvant présenter des singularités isolées et changeantes.*

On attend donc de ces méthodes qu'elles génèrent des maillages adaptés aux solutions, et ceci implique qu'elles soient capables de faire évoluer ces maillages au cours du temps. Une fois encore, insistons sur le fait que des singularités sont susceptibles d'apparaître ou de disparaître entre le début et la fin d'une simulation, et que par conséquent l'évolution d'un maillage ne saurait se réduire à un simple déplacement de ses mailles. Le terme "adaptatif" prend ainsi un sens double : d'une part, il fait référence à la propriété qu'*à un instant donné*, la discrétisation est censée être adaptée à la solution, et d'autre part au fait que *d'un pas de temps à l'autre*, le schéma modifie le maillage d'une façon qui lui est propre et qui n'utilise que les informations disponibles à cet instant de la simulation.

D'après le sens commun, qui veut que la capacité d'adaptation soit un signe d'intelligence, ces méthodes de calcul seraient donc *intelligentes* ? En un sens, oui, car on permet au domaine dans lequel on cherche la solution de varier en fonction des résultats observés. Cet aspect est bien illustré par le jeu suivant : en n'utilisant que des questions

dont la réponse est "oui" ou "non", trouvez la valeur d'un nombre n compris entre 1 et 100. Une méthode de recherche linéaire, consistant à poser une suite de questions sans tenir compte des réponses, ressemblerait à : "est-ce que n est égal à 1 ?" - "non" ; "est-ce que n est égal à 2 ?" - "non" ; *etc.* , et serait évidemment fastidieuse. Spontanément, on cherchera plutôt par dichotomie en demandant : "est-ce que n est entre 1 et 50 ?" - "non" ; "est-ce que n est entre 51 et 75 ?" - "non" ; *etc.*

Pour autant, notre devise ne sera pas "pourquoi faire simple quand on peut faire compliqué ?". En particulier, on soulignera la *simplicité relative d'une méthode de calcul employant un maillage fixé à l'avance*. D'un point de vue numérique d'abord, il est bon d'avoir une idée des difficultés que peut poser la gestion dynamique d'une discrétisation, car elles sont loin d'être négligeables. La conception du schéma de résolution, d'une part, et sa mise en œuvre informatique d'autre part, sont souvent délicates, et les codes doivent faire appel à des structures de données complexes pour matérialiser ces maillages dynamiques. Il faut savoir également que les opérations élémentaires effectuées par un programme sont d'autant plus lentes que les structures de données employées sont évoluées, de sorte que le gain offert par l'adaptativité est souvent moindre en pratique qu'il n'apparaît sur le papier. D'un point de vue théorique ensuite, il est remarquable qu'un très grand nombre de méthodes adaptatives ont été conçues depuis les années 1970, mais qu'on ne dispose d'une analyse rigoureuse que pour très peu d'entre elles. Dans ce domaine, pourtant, on peux citer de nombreux travaux, depuis la mise au point à la fin des années 1970 par Babuška et Rheinboldt [4] d'estimations d'erreur a posteriori pour des techniques d'éléments finis adaptatifs, aux méthodes d'ondelettes proposées par Maday, Perrier, Ravel et Bertoluzza [49, 9] pour approcher des problèmes de transport, en passant par le formalisme du raffinement adaptatif de maillages de Berger, Oliger et Colella [7, 6]. Dans le cadre plus spécifique des ondes de chocs propagées par des lois de conservation, on peut également citer les travaux de Harten [39, 40], ceux de Dahmen, Gottschlich-Müller et Müller [26], de Bertoluzza et Maday [8], et enfin les estimations d'erreur établies par Cohen, Kaber, Müller et Postel [23] pour un schéma de volumes finis à base d'ondelettes.

Toutefois, ces estimations d'erreur sont rares, et plus rares encore sont les cas (à quelques exceptions près, dont un schéma de Lucier [48]), où l'on sait *prouver* qu'une méthode adaptative est plus efficace qu'une méthode uniforme.

Il devient donc crucial, dans ces conditions, de justifier le recours à l'adaptativité de façon rigoureuse. On tentera de le faire par l'intermédiaire des deux questions suivantes :

- est-on capable de régler les différents paramètres d'une simulation pour garantir à l'avance une précision donnée ?
- une fois fixés ces paramètres, quel sera l'ordre de complexité des calculs, notamment en termes de place mémoire et de temps d'exécution ?

Depuis une vingtaine d'années, ce type de questions a reçu des réponses très satisfaisantes au sein du cadre offert par la *théorie de l'approximation*. Un problème d'approximation dans un espace de Banach $\mathcal{X}$ y est considéré par le biais d'un système de complexité croissante : on commence par se donner une suite dense de parties $(\Sigma_N)_{N\geq 1}$ de $\mathcal{X}$, telle que chaque Σ_N contient des fonctions déterminées par $\mathcal{O}(N)$ paramètres.

Pour approcher des fonctions continues sur un intervalle I, par exemple, on peut considérer des polynômes de degré N, ou bien des fonctions polynomiales de degré fixé sur une subdivision de I en N intervalles. Ces ensembles Σ_N représentent un cadre, une approche du problème. Dans un deuxième temps, on considère l'approximation proprement dite d'une fonction f de $\mathcal{X}$, et comme solutions admissibles, on ne retient que des suites d'approximations successives f_N choisies dans les ensembles Σ_N. On peut alors distinguer deux sortes d'approches importantes. La première correspond au cas où ces ensembles sont des *espaces vectoriels* tels les polynômes de degré N, et où les approximants f_N peuvent être simplement obtenus par des projections linéaires. La deuxième correspond au cas où, à l'image des fonctions polynomiales de degré fixé sur N morceaux, les ensembles Σ_N ne se réduisent plus à *un* espace vectoriel de dimension $\mathcal{O}(N)$, mais en contiennent *plusieurs*. Dans cette approche *non-linéaire*, le choix d'un approximant peut encore se faire par une projection linéaire, mais il demande qu'on détermine au préalable sur quel espace vectoriel cette projection devra être effectuée. Ce *choix préalable*, dans le cadre de l'approximation polynomiale par morceaux, correspond précisément au choix de la subdivision de I en N intervalles, et c'est en ce sens qu'une méthode adaptative est non-linéaire.

L'analyse d'une méthode d'approximation donnée

$$A : (N, f) \to f_N \in \Sigma_N$$

consiste alors à étudier le comportement de la suite $\|f - f_N\|_{\mathcal{X}}$. Les Σ_N finissant par remplir l'espace $\mathcal{X}$ tout entier, on peut s'attendre à ce que qu'un algorithme "raisonnable" fasse converger la suite f_N vers f. Inversement, il est naturel que la complexité des approximants tende vers l'infini pour que l'erreur puisse tendre vers 0. On peut donc interpréter la qualité d'une méthode d'approximation sous la forme d'un *compromis* entre la précision des approximations d'une part, et la complexité des approximants d'autre part. Une façon de mesurer la "qualité" d'un tel compromis est alors de chercher à savoir, pour une méthode d'approximation A donnée, *s'il en existe une autre qui soit sensiblement plus efficace*. Autrement dit, s'il est possible de réaliser un meilleur compromis, pour une même suite Σ_N. Répondre à cette question, c'est établir le caractère *optimal* d'une méthode d'approximation, indépendamment de ses performances absolues.

Parmi les résultats les plus significatifs de la théorie de l'approximation non-linéaire, figurent ainsi des théorèmes permettant de caractériser des ordres d'"approximabilité" par des propriétés classiques de régularité comme l'appartenance à des espaces de Sobolev ou de Besov. On citera en particulier le résultat suivant, dû à DeVore, Petrushev et Popov (en renvoyant à [54, 33, 30, 29] pour plus de détails) : si une fonction f définie sur un intervalle I appartient à L^p, $p > 0$, et si l'on désigne par $\Sigma_N = \Sigma_{r,N}$ l'ensemble des fonctions polynomiales de degré inférieur ou égal à un entier r fixé sur une subdivision de I en N intervalles, il est équivalent de dire que

$$\begin{gathered}\text{“}f \text{ peut être approchée dans } L^p \text{ par une suite de fonctions } f_N \in \Sigma_N \\ \text{avec une erreur } \|f - f_N\|_{L^p} \text{ de l'ordre de } N^{-s}\text{”},\end{gathered} \tag{1}$$

pour un réel strictement positif s vérifiant $s < r + 1$, ou bien que

$$\text{“}f \text{ possède } s \text{ dérivées dans } L^q\text{”}, \tag{2}$$

où q est égal à p lorsque les N intervalles sont choisis de façon uniforme, mais vaut $(1/p+s)^{-1} < p$ s'ils sont libres de s'adapter à f. En d'autres termes, la différence entre les approches linéaire et non-linéaire se traduit par des façons différentes de mesurer la régularité de la fonction f, pour un même ordre d'approximation. On a l'habitude de représenter ces différentes régularités dans le diagramme de la figure 1 où sont indiqués en ordonnées le nombre de dérivées s et en abscisses l'inverse $1/p$ de l'exposant de l'espace L^p dans lequel ces dérivées sont mesurées. Le théorème de caractérisation ci-dessus est alors résumé par les deux demi-droites en pointillés issues du point $(0, 1/p)$ qui représente l'espace L^p dans lequel on mesure la qualité des approximations : les ordres d'approximation accessibles par une méthode uniforme correspondent aux espaces situés sur la demi-droite verticale issue de L^p, tandis que les ordres accessibles par une méthode adaptative correspondent aux espaces situés sur la demi-droite de pente 1 (ou de pente d en dimension supérieure). On pourra d'ailleurs y reconnaître la demi-droite correspondant à l'injection critique des espaces de Sobolev dans l'espace L^p.

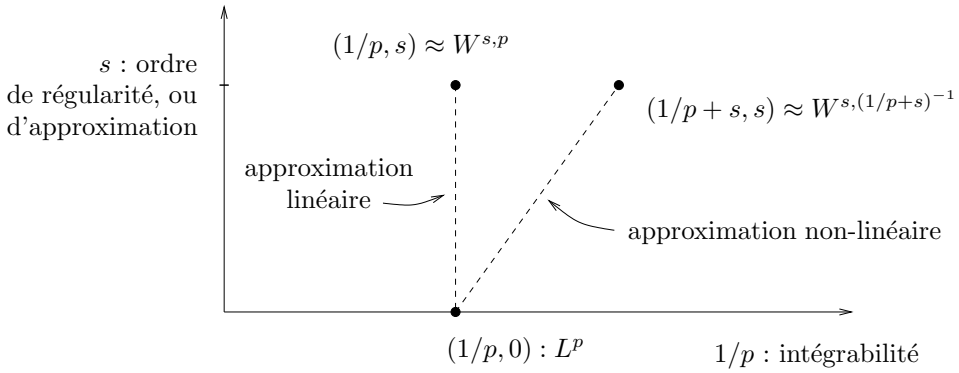


FIG. 1 – les espaces caractérisant l'approximation uniforme (linéaire) dans L^p sont situés sur la demi-droite verticale, tandis que ceux qui caractérisent l'approximation adaptative (non-linéaire) sont situés sur la demi-droite de pente 1.

Bien entendu, l'équivalence (1)-(2) est "abstraite", au sens où elle ne nous apprend pas en général de quelle façon f doit être approchée. En particulier, elle ne nous dit pas *comment choisir la subdivision de I en N intervalles*. C'est néanmoins un résultat remarquable qui met en évidence la supériorité potentielle des méthodes adaptatives sur leurs "petites sœurs" uniformes. Ainsi, la fonction $f(x) = \sqrt{x}$ possède une dérivée intégrable sur l'intervalle $[0, 1]$: si on observe la figure 1, on voit que l'espace $W^{1,1}$ est situé sur la demi-droite de pente 1 issue de L^∞, ce qui signifie qu'il existe une suite f_N de fonctions constantes sur N morceaux *bien choisis* dans $[0, 1]$ telle que la suite $\|f - f_N\|_{L^\infty}$ tend vers 0 comme $1/N$. En revanche, f n'est pas lipschitzienne sur $[0, 1]$, de sorte qu'aucune suite de fonctions f_N constantes sur une subdivision de $[0, 1]$ en N intervalles *uniformes* ne pourra converger vers f à cette vitesse dans L^∞.

Dans le domaine qui nous intéresse, à savoir celui de l'analyse numérique des équations aux dérivées partielles, ces résultats de caractérisation nous poussent à réinterpréter la régularité des solutions exactes, pour lesquelles on dispose souvent de résultats théoriques, comme une information a priori sur l'*objectif à atteindre* en ce qui concerne la

vitesse de convergence d'une méthode numérique. A titre d'exemple, on peut rappeler le problème posé par l'approximation des lois de conservation scalaires

$$\partial_t u(t,x) + \partial_x \cdot [f(u(t,x))] = 0, \quad t > 0, \quad x \in \mathbb{R} \tag{3}$$

associées à une condition initiale $u(0,\cdot) = u_0$. Depuis les travaux de Kružkov [44], on sait que la variation totale $|\cdot|_{BV}$ des solutions n'augmente pas au cours du temps : si u_0 appartient à l'espace BV (situé au point $(1,1)$ sur le diagramme 1), $u(t,\cdot)$ y reste pour tout t. Par contre, les solutions perdent généralement leur continuité après un temps fini, et ceci quelle que soit la régularité de la donnée initiale. Les espaces situés à la verticale de L^∞ étant tous inclus dans l'espace $\mathcal{C}$ des fonctions continues, on voit donc qu'il est inutile de chercher à approcher $u(t,\cdot)$ en norme L^∞ par une méthode uniforme utilisant des polynômes par morceaux (ce qui peut sembler assez évident). Quant aux approximations en norme L^1, elles se limitent a priori à l'ordre 1 car les espaces $W^{s,1}$ sont également contenus dans $\mathcal{C}$ lorsque $s \geq 1$. Intuitivement, on sent bien qu'une méthode adaptative ne devrait pas être gênée de la même façon par la présence de discontinuités. Et de fait, DeVore et Lucier ont démontré en 1990 [31, 32] que les espaces caractérisant l'approximation non-linéaire dans L^1 sont effectivement laissés stables par les lois de conservation scalaires, et ceci sans limitation d'ordre. En d'autres termes, dans une approche adaptative, tout se passe "comme si" les solutions de (3) conservaient une régularité arbitraire de type $W^{s,1}$ au cours du temps, au sens où il est a priori possible de les approcher dans L^1 avec une vitesse de convergence arbitraire, pour peu que la solution initiale soit suffisamment régulière et la méthode d'ordre suffisamment élevé.

Dans ce contexte de recherche active de méthodes d'approximation non-linéaire associées à une analyse accessible, la notion d'*adaptativité multi-échelle* peut constituer un cadre particulièrement agréable. Pour proposer au lecteur un aperçu ludique de cette approche, considérons un employé d'une société de communications sans fil devant installer des antennes relais le long d'une route de montagne. En raison du relief perturbé et des phénomènes de réverbération, notre employé sait à l'avance qu'il devra faire varier la distance entre deux antennes successives de façon que le signal circule correctement (il sait donc qu'il va devoir travailler de façon *adaptative*) mais il n'a aucun moyen a priori de savoir précisément à quelle distance placer les antennes. Pour les besoins de notre exemple, supposons que la pose d'une antenne prend un temps non négligeable, et qu'on ne peut être sûr de la communication entre deux antennes qu'à partir du moment où elles sont bien installées toutes les deux. Deux solutions (au moins) s'offrent alors : la première consiste à démonter la dernière antenne installée tant que le signal est bien reçu, et à la remonter un peu plus loin, jusqu'à ce que le signal ne passe plus. L'antenne est alors fixée définitivement au dernier endroit où la communication était bien établie, et on recommence de la même façon pour la suivante. Une deuxième solution consiste à choisir une distance L relativement grande, et à installer la première antenne à cette distance L du relais initial. Si le signal est bien reçu, on ne la démonte pas et on passe à la suivante. Si le signal ne passe pas, on ne la démonte pas non plus mais on en installe une nouvelle à distance $L/2$ du relais initial. Et ainsi de suite, de sorte qu'à chaque fois que la communication n'est pas établie entre deux antennes, on en rajoute une à mi-distance des deux.

Dans la première solution, on utilise un nombre minimal d'antennes, ce qui sera intéressant si leur coût est élevé, mais au prix de multiples installations. Quant à la deuxième, elle peut sembler plus rapide, quoique cela dépende en réalité de plusieurs paramètres (comme le temps de pose d'une antenne, la distance entre les antennes, la vitesse de la voiture, *etc.*) mais son intérêt principal à nos yeux est qu'elle simplifie grandement le travail d'installation, en ne laissant finalement qu'une seule décision au libre arbitre de l'employé, à savoir la distance L maximale entre deux antennes (n'oublions pas que dans les cas qui nous intéressent, ces choix seront fait en réalité par un algorithme de calcul qu'on souhaite le plus simple possible). Concernant le réseau obtenu, enfin, il est clair que l'ensemble des configurations possibles est bien plus "structuré" avec cette seconde méthode. En particulier, les antennes ne peuvent être installées qu'à distance $k2^{-\ell}L$ de l'antenne initiale, où k et ℓ sont entiers. Si l'on associe à chaque antenne α l'indice $\ell(\alpha)$ pour lequel la fraction $k2^{-\ell}$ est irréductible, ce qui advient lorsque l'entier k est impair, on peut alors observer qu'à l'instant où elle est ajoutée au réseau, cette antenne α se trouve toujours à distance $2^{-\ell(\alpha)}L$ de l'antenne la plus proche, et en particulier à distance $2^{-\ell(\alpha)}L$ d'une antenne β d'indice $\ell(\beta) = \ell(\alpha) - 1$ déjà installée. Un tel réseau peut donc être muni d'une structure arborescente en niveaux qu'illustre la figure 2, où les arcs en pointillés relient deux antennes lorsque la pose de l'une précède toujours celle de l'autre. Dans nos schémas adaptatifs, on utilisera des maillages construits selon ce principe où les raffinements reproduisent un même motif de façon locale et à différentes *échelles de résolution dyadiques*, comme l'illustre la figure 3.

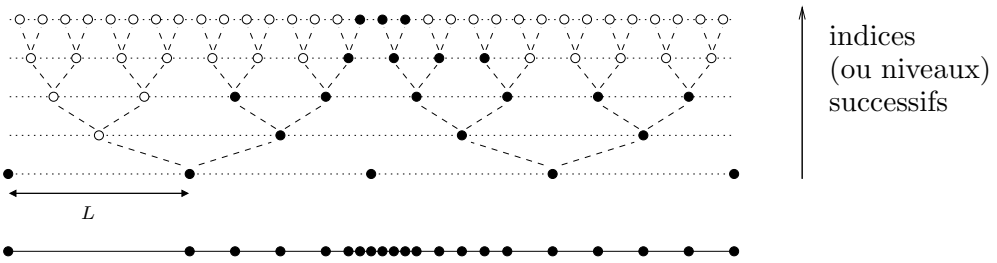


FIG. 2 – représentation multi-échelle d'un réseau de points dyadiques.

En résumé, on insistera sur les aspects suivants d'une discrétisation multi-échelle :

- premièrement, la résolution spatiale n'y varie pas de façon continue, mais un peu à la façon des systèmes quantiques, en *niveaux successifs distincts* (les échelles).
- deuxièmement, ces différents niveaux sont reliés entre eux par une *structure hiérarchique locale* qu'on peut décrire comme un graphe reliant tout élément d'un niveau donné à quelques éléments proches dans les niveaux adjacents.

Dans le contexte d'un problème d'approximation, on pourra interpréter l'approche multi-échelle comme l'abandon d'une certaine "liberté totale" dans le choix des approximants, pour ne garder que ceux ayant précisément une structure multi-échelle. On ne saurait à ce propos passer sous silence le rôle majeur qu'ont joué les ondelettes dans ce domaine depuis la fin des années 1980, aussi bien d'un point de vue théorique que dans leurs applications en approximation non-linéaire et en analyse numérique. Parmi les ouvrages ou articles de référence, citons ceux de Meyer [53], Mallat [50, 51], Daubechies [27], Cohen, Daubechies et Fauveau [22], Dahmen [25], DeVore [29] et Cohen [21].

En ce qui nous concerne, on utilisera plutôt des éléments finis multi-échelles, dont la description en termes d'ondelettes peut d'ailleurs être faite dans le cadre discret de multirésolution proposé par Harten [39], et dans ce cas la structure hiérarchique va correspondre à des règles de modifications locales des maillages. Ainsi, on raffinera des mailles en les remplaçant par leurs "filles" dans une arborescence multi-échelle (*i.e.* leurs successeurs immédiats), et inversement, le dé-raffinement consistera à remplacer un groupe de mailles par leur "mère" (*i.e.* leur prédécesseur commun). Ceci implique que les maillages aux différentes échelles sont emboîtés, comme sur la figure 3.

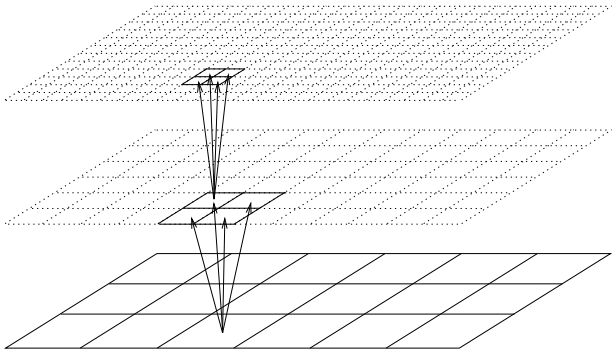


FIG. 3 – structure hiérarchique d'une discrétisation multi-échelle dyadique.

Comme on l'a écrit plus haut, les résultats obtenus au cours de cette thèse concernent deux problèmes de transport non-linéaire : le système de Vlasov-Poisson et les lois de conservation scalaires. Notre plan est donc le suivant : dans une première partie, on propose au lecteur une sorte de "bagage méthodologique minimal" destiné à lui permettre d'aborder la présentation de nos différents travaux avec une vision assez précise du cadre théorique dans lequel ils s'inscrivent et des principaux outils utiles à leur analyse. Ainsi, le chapitre 1 détaille certaines techniques propres à l'étude d'une méthode d'approximation polynomiale par morceaux, et précise dans quelle mesure des hypothèses classiques de régularité peuvent caractériser des vitesses de convergence pour ce type d'approximation non-linéaire. On présente ensuite dans le chapitre 2 une classe d'éléments finis adaptatifs affines par morceaux en dimension deux, basés sur des maillages multi-échelles dyadiques. Ces maillages, qu'on peut obtenir par de simples découpages récursifs des mailles comme celui représenté sur la figure 3, se prêtent particulièrement bien à une gestion dynamique en arbres par un code de calcul. En outre, on peut établir des résultats de complexité concernant leur utilisation à des fins d'approximation adaptative de type affine par morceaux lorsque les fonctions à approcher sont dans l'espace $W^{2,1}(\mathbb{R}^2)$. Dans la suite, on voudra utiliser ces résultats dans un contexte où les solutions numériques transportées par notre schéma sont affines par morceaux. Leurs dérivées secondes étant des mesures de Dirac concentrées sur les arêtes, ces fonctions ne sont plus $W^{2,1}(\mathbb{R}^2)$, mais leurs dérivées secondes étant des mesures de Radon, elles ont ce qu'on appellera une *courbure totale* bornée (tout au moins localement). L'étude de ce nouvel espace $BC(\mathbb{R}^2)$ (pour *bounded curvature*, ainsi désigné par analogie avec l'espace BV des fonctions à variations bornées) fait l'objet

du chapitre 3, où l'on étend les résultats précédents à ce nouveau contexte.

Muni de la classe de discrétisations multi-échelles introduite au chapitre 2, on présente dans la deuxième partie un schéma adaptatif original de type transport-projection dans lequel l'évolution des maillages de calcul se fait par prévision et correction. Ce schéma a été développé à partir d'un travail effectué avec Michel Mehrenberger au CEMRACS 2003 sous la direction conjointe d'Albert Cohen et Eric Sonnendrücker au cours duquel on a développé un code de calcul adaptatif (voir [17]) pour approcher les solutions du système de Vlasov-Poisson. Ce système d'équations, qu'on présente au chapitre 4, modélise l'évolution d'un plasma peu dense de particules chargées soumises essentiellement à leur propre champ électrique. En pratique, le recours à des méthodes adaptatives se justifie dans ce genre de problèmes par la taille des domaines de calculs, les simulations se faisant dans un espace des phases à six dimensions, et par le fait que malgré la régularité théorique des solutions, des structures fines et bien localisées peuvent apparaître et se propager de façon complexe. Pour mettre en avant les principales qualités de notre schéma, et en particulier sa gestion dynamique très simple des maillages adaptatifs, on se place au chapitre 5 dans un cadre abstrait présentant dans ses grandes lignes les propriétés du système de Vlasov-Poisson et de sa discrétisation en temps, telles qu'elles seront établies au chapitre 6. On considère donc un problème de transport non-linéaire suffisamment régulier pour que les solutions préservent leur caractère lipschitzien, et on suppose connu un schéma de transport numérique approchant correctement les trajectoires caractéristiques associées au transport exact. Notre schéma adaptatif utilise alors de façon intensive les propriétés algorithmiques élémentaires des partitions multi-échelles dyadiques, ce qui lui permet d'être rapide et relativement simple à programmer, aussi bien lors de la "prévision" du maillage d'un pas de temps à l'autre, étape ainsi désignée dans la mesure où elle précède le transport de la solution numérique, que lors de leur "correction". La deuxième qualité de notre schéma est que *son analyse est accessible*. En particulier, on montre que les erreurs de projection réalisées sur ces maillages "prédits" sont contrôlées en norme L^∞ par un paramètre de tolérance choisi à l'avance par l'utilisateur. Pour étudier le caractère optimal de ce schéma, on donne alors plusieurs arguments permettant d'estimer la taille des maillages produits. Plus précisément, on montre que l'ordre de complexité des maillages est préservé par l'étape de prévision, et que la correction devrait permettre d'établir un résultat de complexité, sous réserve d'une propriété de décroissance de la courbure par les interpolations qu'on est encore incapable de démontrer. Dans le chapitre 7, on décrit dans ses grandes lignes le code de calcul YODA écrit avec Michel Mehrenberger, et on présente des résultats de simulations numériques qui mettent en évidence un aspect optimal des maillages produits par notre méthode. L'essentiel de ces travaux a été soumis sous la forme d'un article [18] à la revue *SIAM Journal on Numerical Analysis*.

La troisième partie est consacrée à l'analyse des lois de conservation scalaires (3) en distance de Hausdorff. Définie de façon géométrique entre les graphes des fonctions, cette distance permet de considérer l'approximation uniforme d'une fonction discontinue comme un problème a priori bien posé, contournant ainsi l'impossibilité évoquée plus haut d'approcher les solutions des lois de conservation scalaires en norme L^∞. Dans un travail effectué en collaboration avec Albert Cohen, Wolfgang Dahmen et Ronald DeVore, on a pu démontrer la *stabilité* dans cette distance des lois de conservation

scalaires vis-à-vis des perturbations de la donnée initiale u_0 ou du flux f. Après avoir rappelé dans le chapitre 8 de quelle façon Lax [45] a pu proposer une description semi-explicite des solutions de (3) en termes de trajectoires caractéristiques, on présente dans le chapitre 9 nos résultats de stabilité en distance de Hausdorff, qui peuvent être vus comme une extension des propriétés de stabilité L^1 démontrées par Kružkov, et ont fait l'objet d'un article [15] publié dans le *Journal of Hyperbolic Differential Equations*. Un peu plus haut, on a cité un théorème de DeVore et Lucier selon lequel les espaces caractérisant l'approximation adaptative dans L^1 au moyen de polynômes par morceaux étaient préservés par ces équations. En utilisant nos théorèmes de stabilité, on présente dans le chapitre 10 une généralisation de ces résultats (réalisée en collaboration avec Albert Cohen et Pencho Petrushev) à la distance de Hausdorff. Cette distance étant *uniforme*, elle ne permet pas qu'une "bonne approximation" s'écarte localement de la solution exacte comme cela se produit avec des oscillations de Gibbs. En ce sens, les résultats que nous avons obtenus, et qui ont fait l'objet d'un deuxième article [16] publié dans le *Journal of Hyperbolic Differential Equations*, améliorent sensiblement le théorème de DeVore et Lucier.

Constantes.

Lorsqu'on ne cherche pas à connaître leur valeur, on désignera les "constantes" par la lettre C, en s'efforçant de préciser les variables dont elles dépendent. Souvent, en effet, ces constantes ne sont pas absolues : ainsi, dans l'énoncé

$$\text{“}f \text{ étant lipschitzienne, on a } |f(y) - f(x)| \leq C|x - y|\text{”},$$

la constante C dépend évidemment de f, mais on insiste ici sur le fait qu'elle est indépendante de x et de y. D'autre part, la valeur de C pourra varier d'une fois sur l'autre, y compris au sein d'une même équation. On pourra écrire par exemple

$$\Delta t \leq C \quad \Longrightarrow \quad (1 + C\Delta t\,)^2 \leq 1 + C\Delta t\,,$$

car la quantité $2C + C^2\Delta t$ est inférieure à une constante que l'on peut désigner à nouveau par la lettre C.

Domaines d'intégration.

Lorsqu'une intégrale est donnée sans bornes, on prendra par défaut le domaine d'intégration maximal sur lequel l'intégrale a un sens. Ainsi, on écrira indifféremment $\int f$, $\int f(x)\,\mathrm{d}x$ ou $\int_\Omega f(x)\,\mathrm{d}x$ lorsque f est une fonction définie - éventuellement par le contexte - sur un domaine Ω. Il en sera de même pour les normes ou les semi-normes écrites sans précision du domaine.

Première partie

Quelques outils mathématiques

Chapitre 1

Eléments d'approximation non-linéaire

Dans ce chapitre, on présente de façon détaillée le cadre théorique dans le quel on se placera pour envisager les problèmes d'approximation rencontrés dans la suite. En particulier, on précise en quel sens une méthode d'approximation donnée pourra être qualifiée d'optimale. En considérant l'exemple élémentaire de l'approximation dans L^∞ d'une fonction continue par des fonctions constantes par morceaux, on décrit certaines techniques représentatives du *développement* et de l'*analyse* d'une méthode d'approximation, en prenant soin de distinguer les approches uniforme, adaptative "libre" et multi-échelle. Ce chapitre sera également l'occasion de rappeler de quelle façon des hypothèses classiques de régularité peuvent caractériser des ordres d'approximation par des polynômes par morceaux dans des espaces L^p, avec $0 < p \leq \infty$.

1.1 Problématique

On considère donc ici l'approximation d'une fonction f appartenant à un espace de Banach $\mathcal{X}$ par des *approximants* f_N, N=1, 2, ... choisis dans une suite dense $(\Sigma_N)_{N\geq 1}$ de parties de $\mathcal{X}$. L'idée étant de remplacer f - dont la structure est a priori d'une complexité arbitraire - par des représentations calculables, les ensembles Σ_N se distinguent par le fait que leurs éléments sont des fonctions déterminées par $\mathcal{O}(N)$ paramètres dans un système donné. En dimension 1, par exemple, on peut considérer les ensembles $\Sigma_N := \Pi_N$ composés des polynômes de degré inférieur où égal à N pour approcher les fonctions continues sur un intervalle I. Ou bien fixer un degré maximal r et définir Σ_N comme l'ensemble des fonctions $f_N := \sum_{0\leq k\leq N-1} p_k \chi_{I_k}$ coïncidant avec des polynômes $p_k \in \Pi_r$ sur une subdivision uniforme de I en N intervalles I_k de longueur $|I|/N$. Mais on peut également décider de définir Σ_N comme l'ensemble des $f_N := \sum_{0\leq k\leq N-1} p_k \chi_{I_k}$ pour une partition $(I_k)_{k=0,\ldots,N-1}$ *arbitraire* de I en N intervalles. Dans tous les cas, on peut vérifier que les éléments de Σ_N sont bien déterminés par $\mathcal{O}(N)$ paramètres. Plus précisément, Σ_N est un sous-espace de $\mathcal{C}(I)$ dans les deux premiers cas, de dimension $N+1$ et $N(r+1)$ respectivement. Dans le troisième, Σ_N contient beaucoup plus de fonctions, mais il suffit toujours de $N(r+2)$ paramètres pour retrouver la position des intervalles I_k et la valeur des N polynômes p_k. Comme on va le voir au cours de ce chapitre, cette façon de considérer un problème d'approximation en se fixant une suite d'ensembles approximants Σ_N de complexités N n'est réellement

intéressante que dans une approche *non-linéaire* où ces ensembles ne se réduisent pas à des espaces vectoriels de dimension N (ce qui correspond à une approche linéaire), mais contiennent "beaucoup" d'espaces de dimension N. Dans le contexte de l'approximation polynomiale par morceaux, ces différents espaces correspondent aux différentes façons de subdiviser I en N intervalles, et c'est cette "liberté de choix" qui permet à une méthode d'être adaptative.

Pour étudier la qualité des approximations, on définit l'erreur optimale d'approximation de f par les éléments de Σ_N comme

$$\sigma_N(f) := \inf_{g \in \Sigma_N} \|f - g\|_{\mathcal{X}}, \tag{1.1}$$

qu'on peut voir comme la distance entre f et Σ_N. Le fait que les Σ_N forment une suite dense de $\mathcal{X}$ entraîne que $\sigma_N(f)$ tend toujours vers 0. La question est donc : à quelle vitesse ? Est-on capable, par exemple, de caractériser les fonctions f pour lesquelles la suite $\sigma_N(f)$ tend vers 0 comme N^{-s} lorsque s est un réel strictement positif ? A ce sujet, observons que l'ensemble $\mathcal{A}^s(\mathcal{X}) \subset \mathcal{X}$ défini comme

$$f \in \mathcal{A}^s(\mathcal{X}) \iff \sigma_N(f) \leq C(f) N^{-s} \tag{1.2}$$

est un sous-espace de $\mathcal{X}$, souvent appelé *espace d'approximation d'ordre s*, qu'on peut munir de la norme $\|\cdot\|_{\mathcal{A}^s(\mathcal{X})} := \|\cdot\|_{\mathcal{X}} + |\cdot|_{\mathcal{A}^s(\mathcal{X})}$ où

$$|f|_{\mathcal{A}^s(\mathcal{X})} := \sup_{N \geq 1} N^s \sigma_N(f). \tag{1.3}$$

1.1.1 Méthodes optimales d'approximation

On peut alors proposer des algorithmes pratiques construisant des approximations successives de f choisies parmi les ensembles Σ_N, $N \geq 1$. En particulier, on s'attachera à ce que nos *méthodes d'approximation*

$$A\colon (N, f) \to f_N \in \Sigma_N \tag{1.4}$$

soient *explicitement calculables*, et à ce que le calcul de chaque approximation f_N se fasse en un nombre d'opérations élémentaires de l'ordre de N. Lorsque f appartient à un espace $\mathcal{A}^s(\mathcal{X})$, il est alors naturel de se demander si la suite $\|f - f_N\|_{\mathcal{X}}$ tend vers 0, et lorsque c'est le cas, si elle converge à la même vitesse s que les erreurs optimales $\sigma_N(f)$. En particulier, on dira que la méthode (1.4) est *optimale* si elle réalise l'erreur de meilleure approximation à une constante multiplicative près. Autrement dit, si

$$\|f - f_N\|_{\mathcal{X}} \leq C \sigma_N(f) \tag{1.5}$$

avec une constante C indépendante de f et N. Une telle méthode converge bien entendu avec une vitesse optimale, *i.e.*

$$\|f - f_N\|_{\mathcal{X}} \leq C |f|_{\mathcal{A}^s(\mathcal{X})} N^{-s}. \tag{1.6}$$

Remarque 1.1 *En toute rigueur, on ne devrait parler d'approximation "optimale" que lorsque f_N réalise exactement l'erreur*

$$\|f - f_N\|_{\mathcal{X}} = \sigma_N(f). \tag{1.7}$$

Toutefois, compte tenu de la complexité des situations qu'on rencontrera, il ne sera en général pas réaliste d'espérer approcher f de cette façon. On se contentera donc de chercher des méthodes réalisant (1.5) *ou* (1.6)*, autrement dit qui se "comportent comme" une approximation optimale. Et on mettra en avant le fait qu'une telle méthode saisit, pour l'essentiel, la complexité de la fonction cible f.*

Au risque de nous répéter, insistons sur le fait que les notions de méthode optimale, ou de meilleure erreur d'approximation n'ont de sens que pour *une* suite donnée d'ensembles Σ_N. Ainsi le lemme élémentaire suivant, selon lequel n'importe quelle suite bornée de projections linéaires $A_N : \mathcal{X} \to \Sigma_N$ (ce qui implique que les Σ_N sont des espaces vectoriels) réalise une approximation optimale, illustre à sa façon la grande simplicité de l'approche linéaire.

Lemme 1.2 *Si les A_N sont des projections linéaires sur les espaces Σ_N et vérifient*

$$\|A\| := \sup_{N\geq 1} \sup_{\|g\|_{\mathcal{X}}\leq 1} \|A_N g\|_{\mathcal{X}} \leq C, \tag{1.8}$$

alors l'approximation $(N, f) \to A_N f$ est optimale au sens où

$$\|f - A_N f\|_{\mathcal{X}} \leq (1 + \|A\|)\sigma_N(f). \tag{1.9}$$

Preuve. Soit N un entier fixé. Pour tout $g_N \in \Sigma_N$, on a

$$\|f - A_N f\|_{\mathcal{X}} \leq \|f - g_N\|_{\mathcal{X}} + \|A_N(g_N - f)\|_{\mathcal{X}} \leq (1 + \|A\|)\|g_N - f\|_{\mathcal{X}}. \tag{1.10}$$

Comme g_N est une fonction arbitraire de Σ_N, on en déduit facilement (1.9). □

1.1.2 Approximation d'un problème de transport

Dans le cadre de cette thèse, les fonctions qu'on cherche à approcher ne sont pas données de façon explicite, comme une image qu'on souhaite compresser, mais de façon implicite, comme solutions d'une équation de transport

$$\partial_t f(t,x) + F(t,x) \cdot \nabla_x f(t,x) = 0, \quad t > 0, \quad x \in \mathbb{R}^d \tag{1.11}$$

associée à une condition initiale

$$f(0,x) = f_0(x). \tag{1.12}$$

Dans les équations qu'on étudiera, le terme de force $F : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ dépendra de la solution elle-même, de sorte que le problème aux valeurs initiales (1.11)-(1.12), appelé aussi problème de Cauchy , est *non-linéaire*. Pour résoudre numériquement ce problème, on s'est principalement intéressé à des schémas de type "transport-projection" associés à une discrétisation uniforme en temps. A chaque instant $t_n := n\Delta t$, $n \in \mathbb{N}$, la solution $f(t_n, \cdot)$ y est approchée par une solution numérique f^n donnée par

$$f^n := \begin{cases} P_0 f_0 & \text{pour } n = 0 \\ P_n \mathcal{T} f^{n-1} & \text{pour } n \geq 1, \end{cases} \tag{1.13}$$

où $\mathcal{T}$ et P_n désignent respectivement un opérateur de transport associé à l'équation (1.11) et une projection associée à une discrétisation spatiale. Dans une approche linéaire, par exemple, on pourra définir P_n comme une projection linéaire sur l'espace vectoriel Σ_N. Bien entendu, dans ce cas les projections ne dépendent plus de l'instant n, et on n'a plus affaire à un schéma adaptatif, du moins pas au sens où la discrétisation s'adapte à la solution d'un pas de temps à l'autre. Notre démarche correspond plutôt à une approche non-linéaire, où les ensembles Σ_N contiennent un grand nombre d'espaces de dimension $\mathcal{O}(N)$, et où P_n est défini à chaque pas de temps comme une projection linéaire sur un de ces espaces *judicieusement choisi*. Dans le cadre d'une approximation polynomiale par morceaux sur un intervalle I, ceci correspondra au choix d'une subdivision de I en N intervalles *appropriés à la solution transportée* $\mathcal{T}f^{n-1}$.

On choisit alors d'analyser la précision d'un schéma de type (1.13) en décomposant l'erreur numérique au pas de temps n suivant

$$\|f(t_n) - f^n\|_{\mathcal{X}} \leq \|f(t_n) - \mathcal{T}f(t_{n-1})\|_{\mathcal{X}} + \|\mathcal{T}f(t_{n-1}) - \mathcal{T}f^{n-1}\|_{\mathcal{X}} + \|(I - P_n)\mathcal{T}f^{n-1}\|_{\mathcal{X}}, \tag{1.14}$$

faisant ainsi apparaître trois termes importants. Le premier terme $\|f(t_n) - \mathcal{T}f(t_{n-1})\|_{\mathcal{X}}$ peut être vu comme l'erreur de discrétisation en temps associée à l'opérateur de transport $\mathcal{T}$. Son étude est a priori indépendante de la discrétisation choisie en espace, et ne fait intervenir que les propriétés de l'équation (1.11) et de la donnée initiale (1.12). Le deuxième terme $\|\mathcal{T}f(t_{n-1}) - \mathcal{T}f^{n-1}\|_{\mathcal{X}}$ exprime une certaine *régularité* de l'opérateur $\mathcal{T}$, liée à la façon dont les solutions de (1.11)-(1.12) réagissent par rapport à des petites perturbations de la donnée initiale. L'erreur de projection $\|(I - P_n)\mathcal{T}f^{n-1}\|_{\mathcal{X}}$, enfin, correspond à l'erreur d'approximation spatiale qui nous intéresse en premier lieu. C'est principalement de son analyse que sont issues les stratégies de prédiction de maillages qu'on proposera aux chapitres 5 et 6.

Remarque 1.3 *Dans la discussion ci-dessus, on a évoqué la régularité de l'opérateur $\mathcal{T}$ qui est un opérateur approché. En décomposant le terme $\|f(t_n) - \mathcal{T}f^{n-1}\|_{\mathcal{X}}$ comme $\|f(t_n) - \mathcal{T}_{\text{ex}}f^{n-1}\|_{\mathcal{X}} + \|\mathcal{T}_{\text{ex}}f^{n-1} - \mathcal{T}f^{n-1}\|_{\mathcal{X}}$, où $\mathcal{T}_{\text{ex}}$ désigne l'opérateur d'évolution* exact *associé à l'équation* (1.11)*, il est également possible de ne faire apparaître que la stabilité de l'équation elle-même vis-à-vis de la solution initiale. Mais on doit alors estimer l'erreur de discrétisation en temps sur la solution numérique f^{n-1}, parfois moins régulière que la solution exacte.*

Une fois que l'on aura estimé l'erreur numérique à chaque pas de temps n en fonction de la précédente, on utilisera le lemme élémentaire de Gronwall ci-dessous pour majorer ces erreurs de façon plus globale (intégrée en temps, si l'on peut dire).

Lemme 1.4 (de Gronwall, discret) *Si u_n est une suite positive qui vérifie*

$$u_{n+1} \leq (1 + c\Delta t\,)u_n + a$$

pour tout $n \geq 0$ et pour un réel $a \geq 0$, alors elle est majorée par

$$u_n \leq e^{cn\Delta t}\,(a/(c\Delta t\,) + u_0). \tag{1.15}$$

En pratique, la quantité $n\Delta t$ sera toujours inférieure au "temps maximal" T de la simulation, aussi $e^{cn\Delta t}$ pourra-t-elle être considérée comme une constante indépendante du pas de temps Δt.

Preuve. Par récurrence, on obtient

$$u_n \leq \sum_{n'=0}^{n-1} (1 + c\Delta t\,)^{n'} a + (1 + c\Delta t\,)^n u_0 = \left(\frac{(1 + c\Delta t\,)^n - 1}{c\Delta t} \right) a + (1 + c\Delta t\,)^n u_0.$$

La majoration par x du logarithme de $1 + x$ entraînant $n \ln(1 + c\Delta t\,) \leq cn\Delta t$, on en déduit que $(1 + c\Delta t\,)^n \leq e^{cn\Delta t}$, d'où (1.15). □

Ce lemme peut d'ailleurs être vu comme une version discrète du lemme suivant :

Lemme 1.5 (de Gronwall) *Si ψ est une fonction positive de L^1 qui vérifie*

$$\psi(t) \leq \psi(0) + c \int_0^t \psi(s)\,\mathrm{d}s$$

pour tout $t > 0$, avec $c \geq 1$ (par commodité), alors elle est majorée par

$$\psi(t) \leq c\psi(0)e^{ct},$$

également pour tout $t > 0$.

Preuve. Pour commencer, observons qu'une fonction $\varphi > 0$ telle que $\varphi'(t) \leq c\varphi(t)$ vérifie $(\ln \varphi(t))' \leq c$, d'où $\ln \varphi(t) \leq \ln \varphi(0) + ct$ et finalement

$$\varphi(t) \leq \varphi(0)e^{ct}.$$

Posons alors

$$\varphi_\varepsilon(t) := \varepsilon + \psi(0) + c \int_0^t \psi(s)\,\mathrm{d}s$$

pour un $\varepsilon > 0$ arbitraire. Dans la mesure où $c \geq 1$, on a

$$\varphi'_\varepsilon(t) = \psi(t) \leq \psi(0) + c \int_0^t \psi(s)\,\mathrm{d}s \leq c\varphi_\varepsilon(t). \tag{1.16}$$

On déduit alors de notre première observation que $\varphi_\varepsilon(t) \leq \varphi_\varepsilon(0)e^{ct} = (\psi(0) + \varepsilon)e^{ct}$, et en utilisant une deuxième fois l'inégalité (1.16), que

$$\psi(t) \leq c\varphi_\varepsilon(t) \leq c(\psi(0) + \varepsilon)e^{ct}.$$

Le lemme s'en déduit alors en faisant tendre ε vers 0. □

1.2 Approximation par des constantes par morceaux

Pour décrire plus en détails certaines techniques d'approximation non-linéaire qu'on utilisera dans la suite, intéressons-nous à présent à l'approximation d'une fonction *connue* f de $\mathcal{X} := \mathcal{C}([0, 1])$ par des fonctions constantes par morceaux

$$f_N(x) = \sum_{k=0}^{N-1} c_k \chi_{I_k}(x), \tag{1.17}$$

sur une partition $(I_k)_{k=0,\ldots,N-1}$ de $[0, 1]$ en N intervalles. La précision des approximations sera mesurée en norme L^∞.

Remarque 1.6 *On pourra objecter que les approximants* (1.17) *ne sont pas continus. C'est vrai, mais ce n'est pas gênant car ils sont bornés, de sorte que les erreurs* $\|f - f_N\|_{L^\infty}$ *sont bien définies, et ils sont denses dans* $\mathcal{C}([0,1])$*. On peut à ce sujet voir l'espace* $\mathcal{C}([0,1])$ *comme un cadre fonctionnel où l'approximation dans* L^∞ *par des fonctions de type polynômes par morceaux est "bien posée", notamment en ce qui concerne la* stabilité *des résultats. En particulier, on peut tout à fait approcher une fonction* f *discontinue par des fonctions* f_N *constantes par morceaux avec une grande précision dans* L^∞*, mais il faut pour cela que les discontinuités des fonctions* f_N *coïncident exactement avec celles de* f*. Cette extrême sensibilité de l'erreur d'approximation vis-à-vis de la position des intervalles fait qu'il sera impossible d'obtenir des résultats intéressants pour des méthodes d'approximation utilisables en pratique. On aura l'occasion de reprendre cette discussion au chapitre 9.*

1.2.1 Approche linéaire uniforme

La façon la plus simple de choisir les fonctions approximantes (1.17) est de fixer une fois pour toutes les partitions $(I_k)_{k=0,\dots,N-1}$. N'ayant a priori aucune raison de privilégier telle ou telle zone de l'intervalle $[0,1]$, on peut par exemple prendre pour chaque entier N la partition uniforme de pas $1/N$. On pose alors $I_k := [k/N, (k+1)/N]$, et on approche f par

$$f_N := \sum_{k=0}^{N-1} c_k(f)\chi_{I_k} \text{ avec } c_k(f) := N\int_{I_k} f. \tag{1.18}$$

Comme f est continue, il existe dans chaque intervalle I_k un x_k tel que $c_k(f) = f(x_k)$. Si f est lipschitzienne, on en déduit que

$$\|f - f_N\|_{L^\infty(I_k)} = \sup_{x\in I_k} |f(x) - f(x_k)| \leq |f|_{W^{1,\infty}(I_k)} N^{-1}, \tag{1.19}$$

d'où

$$\|f - f_N\|_{L^\infty} \leq |f|_{W^{1,\infty}} N^{-1}. \tag{1.20}$$

On retrouve ainsi la notion évoquée dans l'introduction de "compromis" entre la précision des approximations et les ressources qu'on est prêt à leur accorder. L'inégalité (1.20) exprime en effet

- une propriété de *précision a priori* vérifiée par l'algorithme (1.18), qui approche une fonction f lipschitzienne avec une erreur inférieure à $|f|_{W^{1,\infty}} N^{-1}$.
- une information sur la *complexité* de l'algorithme, au sens où le nombre N_ε de morceaux nécessaires à l'approximation de f avec une précision ε peut être choisi de l'ordre de $|f|_{W^{1,\infty}} \varepsilon^{-1}$.

Pour parler d'approximation optimale, il faut voir chaque fonction f_N comme un élément de l'ensemble

$$\Sigma_N := \left\{ g = \sum_{k=0}^{N-1} c_k \chi_{I_k} : c_0, \dots, c_{N-1} \in \mathbb{R} \right\}. \tag{1.21}$$

Bien évidemment, l'inégalité (1.20) entraîne qu'il existe pour chaque fonction $f \in W^{1,\infty}$ une constante $C(f)$ telle que

$$\sigma_N(f) \leq C(f) N^{-1}, \tag{1.22}$$

autrement dit, que les fonctions lipschitziennes peuvent être approchées à une vitesse au moins linéaire par des éléments de Σ_N. Pour reprendre les espaces d'approximation $\mathcal{A}^s(\mathcal{X})$ introduits au début de la section 1.1, nous sommes en train de dire que l'espace $\mathcal{A}^1(\mathcal{C}([0,1]))$ - relatif aux approximants de la forme (1.18) - contient les fonctions de $W^{1,\infty}$.

Il est facile de voir que la réciproque est vraie, autrement dit que $\mathcal{A}^1(\mathcal{C}([0,1]))$ ne contient *que* les fonctions de $W^{1,\infty}$. Le principal argument consiste à observer qu'une fonction g constante sur deux intervalles consécutifs $[a,b[$ et $]b,c]$ vérifie pour toute fonction f continue

$$\begin{aligned}|g(a)-g(c)| \leq |g(a)-g(b^-)| + |g(b^-)-f(b^-)| + |f(b^-)-f(b^+)| + |f(b^+)-g(b^+)| \\ + |g(b^+)-g(c)| = |g(b^-)-f(b^-)| + |f(b^+)-g(b^+)| \leq 2\|f-g\|_{L^\infty}. \quad (1.23)\end{aligned}$$

Considérons alors f dans $\mathcal{A}^1(\mathcal{C}([0,1]))$, et désignons par $g_N \in \Sigma_N$ une fonction réalisant l'erreur optimale $\|f-g_N\|_{L^\infty} \leq \sigma_N(f)$, éventuellement à une constante multiplicative près. Pour x et y distincts dans $[0,1]$, on peut toujours choisir $N \geq 1$ tel que

$$\frac{1}{2N} \leq |x-y| \leq \frac{1}{N}, \quad (1.24)$$

de sorte que x et y appartiennent soit au même intervalle $I_k = [k/N,(k+1)/N]$, soit à deux intervalles consécutifs I_k et I_{k+1}. Dans un cas comme dans l'autre, on peut utiliser l'argument ci-dessus pour voir que

$$\begin{aligned}|f(x)-f(y)| &\leq |f(x)-g_N(x)| + |g_N(x)-g_N(y)| + |g_N(y)-f(y)| \\ &\leq 4\|f-g_N\|_{L^\infty} \leq C\sigma_N(f) \leq C(f)N^{-1} \quad (1.25)\end{aligned}$$

et on déduit de (1.24) que f est lipschitzienne. On voit donc que $\mathcal{A}^1(\mathcal{C}([0,1]))$ coïncide exactement avec $W^{1,\infty}$, et les mêmes arguments nous montreraient que $\mathcal{A}^s(\mathcal{C}([0,1]))$, lorsque s est strictement compris entre 0 et 1, coïncide avec l'espace de Hölder

$$\mathcal{C}^s := \{f \in L^\infty : |f(x)-f(y)| \leq C(f)|x-y|^s\}. \quad (1.26)$$

On montre de cette façon que l'approximation (1.18) est optimale au sens (1.5) lorsque f est une fonction de $\mathcal{C}^s$ ou de $W^{1,\infty}$ (ce qui n'est pas surprenant, au vu du lemme 1.2), et en particulier qu'elle converge avec une *vitesse optimale*, au sens où aucune suite de fonctions g_N appartenant aux Σ_N ne converge vers f avec une vitesse supérieure à celle des f_N données par (1.18). A ce sujet, on peut se demander quelles fonctions peuvent être approchées avec une vitesse plus grande que 1. Autrement dit, quelle régularité permet de caractériser les espaces $\mathcal{A}^s(\mathcal{C}([0,1]))$ lorsque $s > 1$? Si l'on reprend l'argument utilisé en (1.25), on peut voir que les seules fonctions de $\mathcal{A}^s(\mathcal{C}([0,1]))$, lorsque $s > 1$, sont les constantes ! En théorie de l'approximation, on parle de *saturation* pour désigner ce genre de phénomène, et il faut y voir le fait que les fonctions constantes par morceaux sont en quelque sorte trop "rigides" pour être sensibles à des ordres plus élevés de régularité. Pour en tirer parti, il faut alors avoir recours à des méthodes d'ordre également plus élevé, utilisant par exemple des fonctions approximantes affines ou paraboliques par morceaux.

1.2.2 Approche adaptative libre

Pour obtenir de meilleurs résultats, tout en conservant la structure constante par morceaux (1.17) des fonctions approximantes, le principe de l'approche adaptative consiste à utiliser des partitions plus générales. Plus précisément, à *choisir les N intervalles I_k, $k = 0, \dots, N-1$ en fonction de f.*

On peut alors se demander : et les coefficients $c_k(f)$? Pourquoi ne pas essayer aussi d'améliorer les approximations en choisissant par exemple d'interpoler f en des points x_k choisis d'une façon *particulièrement astucieuse* dans les intervalles I_k ? Dans la section précédente, la seule propriété que nous avons utilisée dans l'analyse était que les x_k appartenaient aux I_k. Avec des partitions uniformes, on peut donc voir qu'un choix quelconque permet d'écrire (1.20), ce qui correspond à une approximation optimale. Mais est-ce le cas pour des partitions générales ? La réponse est oui, et pour le voir il suffit d'appliquer le lemme 1.2, à n'importe quelle projection

$$P_{\{x_0,\dots,x_{N-1}\}} \colon f \to \sum_{k=0}^{N-1} f(x_k)\chi_{I_k} \tag{1.27}$$

associée à une distribution arbitraire des points x_k dans les intervalles I_k. Dans la mesure où $P_{\{x_0,\dots,x_{N-1}\}}$ fait clairement décroître la norme L^∞, on en déduit que

$$\|f - P_{\{x_0,\dots,x_{N-1}\}} f\|_{L^\infty} \leq 2 \inf \left\{ \|f - g\|_{L^\infty} : g = \sum_{k=0}^{N-1} c_k \chi_{I_k}, c_k \in \mathbb{R} \right\}, \tag{1.28}$$

autrement dit que n'importe quelle projection $P_{\{x_0,\dots,x_{N-1}\}} f$ réalise l'erreur optimale à une constante multiplicative près, une fois qu'on a choisi les intervalles $I_0, \dots, I_{N-1}$.

Pour choisir ces intervalles, donc, on peut s'inspirer de l'estimation suivante, valable pour $f \in W^{1,1}(I_k)$ et $x_k \in I_k$:

$$\|f - f(x_k)\|_{L^\infty(I_k)} = \sup_{x \in I_k} \left| \int_{x_k}^{x} f'(y)\, \mathrm{d}y \right| \leq \|f'\|_{L^1(I_k)}. \tag{1.29}$$

En choisissant des intervalles I_k de façon à répartir équitablement ces quantités $\|f'\|_{L^1(I_k)}$, *i.e.*

$$\|f'\|_{L^1(I_k)} = \|f'\|_{L^1} N^{-1} \quad \text{pour } k = 0, \dots N-1, \tag{1.30}$$

ce qui est toujours possible lorsque $f \in W^{1,1}$, on aura

$$\|f - f_N\|_{L^\infty} \leq |f|_{W^{1,1}} N^{-1}. \tag{1.31}$$

Il est important de distinguer cette estimation de l'inégalité (1.20) valable pour une méthode uniforme. Dans les deux cas, les approximations convergent vers f en $\mathcal{O}(N^{-1})$, mais à la différence de (1.20), l'estimation ci-dessus est vérifiée par des fonctions bien moins régulières. La fonction $x \to x^s$, par exemple, n'appartient qu'à $\mathcal{C}^s([0,1])$ lorsque $0 < s < 1$, d'où l'on déduit que sa vitesse optimale de convergence est en $\mathcal{O}(N^{-s})$ avec une méthode uniforme. Comme elle est dans $W^{1,1}([0,1])$, elle sera en revanche approchée en $\mathcal{O}(N^{-1})$ par la méthode adaptative décrite ci-dessus. L'avantage de (1.20) sur (1.31) ne se limite d'ailleurs pas à cette situation. Ainsi, une fonction lipschitzienne

dont les variations sont bien localisées aura une semi-norme $W^{1,1}$ bien plus petite que sa semi-norme $W^{1,\infty}$.

Dans la construction précédente, on s'est inspiré de l'inégalité (1.29) dans laquelle les quantités $\|f'\|_{L^1(I_k)}$ jouent le rôle d'indicateurs d'erreur a priori, et on a ensuite cherché à équilibrer ces indicateurs par un bon choix des intervalles. On peut alors se demander s'il ne serait pas plus efficace de vouloir équilibrer directement les erreurs d'interpolation, de façon à avoir

$$\|f - f_N\|_{L^\infty(I_k)} = \varepsilon \quad \text{pour } k = 0, \dots N-1. \tag{1.32}$$

Comme on ne sait pas a priori quelle valeur prendre pour ε, on peut commencer par construire une partition vérifiant (1.32) pour un ε arbitraire, et compter ensuite le nombre d'intervalles obtenus. On pose alors $y_0 := 0$, et pour tout i entier,

$$y_{i+1} := \sup\{y \in]y_i, 1] : |f(y) - f(y_i)| \leq \varepsilon\}. \tag{1.33}$$

Comme $|f(\cdot) - f(y_i)|$ est continue, on peut écrire

$$y_{i+1} < 1 \implies \|f - f(y_i)\|_{L^\infty([y_i, y_{i+1}])} = \varepsilon, \tag{1.34}$$

de sorte que si $f \in W^{1,1}$, on a pour tout j tel que $y_j < 1$, en utilisant l'estimation locale (1.29),

$$j\varepsilon = \sum_{i=0}^{j-1} \|f - f(y_i)\|_{L^\infty([y_i, y_{i+1}])} \leq \sum_{i=0}^{j-1} |f|_{W^{1,1}([y_i, y_{i+1}])} \leq |f|_{W^{1,1}}. \tag{1.35}$$

En particulier, ceci implique $j \leq |f|_{W^{1,1}}\varepsilon^{-1}$, et on en déduit qu'on a construit

$$N(\varepsilon) = \lfloor |f|_{W^{1,1}}\varepsilon^{-1} \rfloor + 1 \leq C|f|_{W^{1,1}}\varepsilon^{-1}$$

intervalles $I_i := [y_i, y_{i+1}]$ de cette façon, de sorte qu'on retrouve la même vitesse de convergence qu'en équilibrant les quantités $\|f'\|_{L^1(I_k)}$.

Dans un calcul pratique, toutefois, il est assez rare qu'on puisse construire des partitions qui réalisent des équilibrages exacts tels que (1.30) ou (1.34). Il convient donc d'étudier la qualité de ces approximations lorsque ces équilibrages sont *approchés*, ce qui revient à établir une forme de stabilité pour ces approximations dans l'esprit de la remarque 1.6. On peut ainsi observer que si les intervalles I_k vérifient

$$c_1 \leq \|f'\|_{L^1(I_k)} \leq c_2 \quad \text{pour } k = 0, \dots N-1, \tag{1.36}$$

on aura en sommant sur k :

$$Nc_1 \leq \sum_{k=0}^{N-1} \|f'\|_{L^1(I_k)} = |f|_{W^{1,1}} \tag{1.37}$$

et par conséquent

$$\|f - f_N\|_{L^\infty} \leq c_2 \leq \frac{c_2}{c_1}|f|_{W^{1,1}}N^{-1}. \tag{1.38}$$

De même, si les intervalles I_k vérifient

$$c\varepsilon \leq \|f - f_\pi\|_{L^\infty(I_k)} \leq \varepsilon \quad \text{pour } k = 0, \dots N-1, \tag{1.39}$$

on peut déduire de (1.29) que

$$cN\varepsilon \leq \sum_{k=0}^{N-1} \|f - f_N\|_{L^\infty(I_k)} \leq \sum_{k=0}^{N-1} \|f'\|_{L^1(I_k)} = |f|_{W^{1,1}}, \tag{1.40}$$

d'où

$$\|f - f_N\|_{L^\infty} \leq \varepsilon \leq \frac{1}{c}|f|_{W^{1,1}} N^{-1}. \tag{1.41}$$

On retrouve ainsi des vitesses de convergence en $\mathcal{O}(N^{-1})$ pour toute fonction $f \in W^{1,1}$. D'une certaine façon, on peut voir les inégalités de gauche de (1.36) et (1.39) comme une garantie que les intervalles I_k ont tous une "efficacité minimale" dans la perspective d'un partage de la semi-norme $|f|_{W^{1,1}}$.

1.2.3 Approche adaptative multi-échelle

Dans l'introduction, on a longuement évoqué l'intérêt des structures multi-échelles, et notamment des maillages dyadiques, dans un contexte de gestion dynamique rapide de la discrétisation par un algorithme "simple". En termes plus publicitaires, on pourrait ainsi dire que suivre une démarche multi-échelle, c'est faire le choix de la simplicité... En contre-partie, il est assez clair que les erreurs optimales $\sigma_N(f)$ seront plus grandes une fois qu'on aura limité les ensembles Σ_N aux fonctions constantes sur des intervalles dyadiques de la forme

$$I_{\ell,k} := [2^{-\ell}k, 2^{-\ell}(k+1)], \quad \text{avec} \quad \ell,\ k \in \mathbb{N} \quad \text{et} \quad 0 \leq k \leq 2^\ell - 1. \tag{1.42}$$

La question est donc : que perd-on, en termes de vitesses de convergence des approximations optimales ? Soit : quel est le prix de la simplicité ?

Pour choisir une partition dyadique adaptée à une fonction $f \in W^{1,1}$, l'approche la plus naturelle consiste sans doute à se donner une erreur maximale ε, et à appliquer un algorithme de découpage adaptatif partant de la maille racine $I_{0,0} = [0,1]$. Le principe de cet algorithme est de raffiner une partition "progressive" initialement réduite à $I_{0,0}$, en testant chacun de ses intervalles $I_{\ell,k}$. On peut demander par exemple que l'indicateur local y vérifie

$$\|f'\|_{L^1(I_{\ell,k})} \leq \varepsilon. \tag{1.43}$$

Si c'est le cas, alors $I_{\ell,k}$ est un "bon" intervalle, et on le laisse passer. Dans le cas contraire, on le découpe en deux, ne laissant derrière lui que ses enfants $I_{\ell+1,2k}$ et $I_{\ell+1,2k+1}$. En appliquant ce procédé jusqu'à ce qu'il ne reste plus que des bons intervalles, ce qui arrivera toujours en un nombre fini de subdivisions lorsque f est dans $W^{1,1}$, on obtient une partition de $[0,1]$ en intervalles I_k, $k = 0, \dots, N(\varepsilon) - 1$ pour laquelle l'estimation

$$\|f - f_{N(\varepsilon)}\|_{L^\infty} \leq \varepsilon \tag{1.44}$$

est évidente. Il est moins évident, par contre, d'estimer le nombre d'intervalles $N(\varepsilon)$ construits par cet algorithme. En particulier, on ne peut pas écrire ici l'inégalité de

gauche dans (1.36), car la fonction f peut très bien être constante sur certains intervalles I_k, voire sur la plupart d'entre eux : si $f(x) = (1-2x)\chi_{x\leq 1/2}(x)$, par exemple, l'intervalle $[1/2, 1]$ fera toujours partie de la partition dyadique construite par cet algorithme, du moins lorsque $\varepsilon < 1$. Et plus généralement, si la "masse" $\|f'\|_{L^1}$ se concentre sur une zone arbitrairement petite, il nous faudra effectuer un nombre fini mais arbitrairement grand de raffinements dyadiques avant d'obtenir une partition de $[0, 1]$ en bons intervalles. On est donc fortement pénalisé, dans cette situation, par la rigidité des partitions dyadiques "simples".

Pour pouvoir majorer le nombres d'intervalles $N(\varepsilon)$ obtenus de cette façon, il apparaît donc indispensable de prévenir tout phénomène de concentration arbitraire de la masse $\|f'\|_{L^1}$. Une façon naturelle de le faire serait de demander à f' d'être "un peu plus régulière" que L^1, ce qu'on peut voir comme le supplément à payer pour pouvoir adopter une approche multi-échelle. La question est alors : de quelle régularité a-t-on besoin pour établir une approximation d'ordre 1 ? Doit-on aller jusqu'à supposer que f est lipschitzienne ? Voici une réponse donnée par DeVore dans [28], et on aura l'occasion de donner d'autres réponses à cette question très importante à la fin des chapitres 2 et 3 : on peut estimer $N(\varepsilon)$ dès que f' appartient à l'espace $L \log L$, qui contient les fonctions g de L^1 telles que

$$\|g\|_{L\log L} := \int |g(x)|(1 + \log|g(x)|)\,\mathrm{d}x < \infty \tag{1.45}$$

et vérifie $L^p \subsetneq L\log L \subsetneq L^1$ pour tout $p > 1$. Plus précisément, on peut montrer que

$$N(\varepsilon) \leq C\|f'\|_{L\log L}\varepsilon^{-1}, \tag{1.46}$$

de sorte que le "supplément de régularité" est finalement très raisonnable. Pour établir cette estimation, on peut suivre un raisonnement semblable à (1.36)-(1.38), mais en utilisant la fonction maximale de Hardy-Littlewood $M(f')$ à la place de f'. Cette fonction est définie par

$$M(f')(x) := \sup_{J\ni x} |J|^{-1} \int_J |f'(y)|\,\mathrm{d}y, \tag{1.47}$$

où la borne supérieure est prise sur tous les intervalles de $[0, 1]$ contenant x, et elle vérifie (voir [60] et [5])

$$C\|f'\|_{L\log L} \leq \|M(f')\|_{L^1} \leq C'\|f'\|_{L\log L}. \tag{1.48}$$

Si I_k est un "bon" intervalle dyadique construit par l'algorithme de découpage adaptatif, on sait par construction que son parent J_k est un "mauvais" intervalle dont le découpage en deux a produit I_k. En particulier, cet intervalle vérifie

$$\varepsilon < \int_{J_k} |f'(y)|\,\mathrm{d}y. \tag{1.49}$$

Comme I_k est un sous-intervalle de J_k pour lequel on a $|J_k| = 2|I_k|$, on aura $(2|I_k|)^{-1}\varepsilon < M(f')(x)$ pour tout $x \in I_k$. On en déduit que

$$\varepsilon/2 < \|M(f')\|_{L^1(I_k)}, \tag{1.50}$$

et en utilisant le fait que les I_k sont (d'intérieur) disjoints, on conclut facilement que

$$N(\varepsilon)\varepsilon/2 \leq \sum_{k=0}^{N(\varepsilon)-1} \|M(f')\|_{L^1(I_k)} \leq \|M(f')\|_{L^1} \leq C'\|f'\|_{L\log L} \tag{1.51}$$

en utilisant (1.48).

1.3 Approximation polynomiale par morceaux

Au chapitre 10, on aura besoin d'utiliser les théorèmes de caractérisation présentés dans l'introduction à propos de l'approximation polynomiale par morceaux d'ordre élevé sur l'intervalle $[0,1]$. Pour les décrire de façon cohérente avec les notations qu'on utilisera alors, faisons dès maintenant le choix de ne travailler qu'avec des puissances de 2 pour l'ordre de complexité N, qu'on notera 2^n en prenant garde de ne pas confondre cet exposant avec le numéro des pas de temps utilisé dans l'écriture des solutions de schémas numériques. On désigne alors par $\Sigma_n = \Sigma_{n,r}$ (plutôt que par Σ_{2^n}, pour simplifier les notations) l'ensemble des fonctions polynomiales de degré inférieur ou égal à r sur une subdivision de $[0,1]$ en 2^n intervalles arbitraires, autrement dit les fonctions qui s'écrivent

$$S_n(x) = \sum_{k=0}^{2^n-1} p_k(x)\chi_{I_k}(x), \quad p_k \in \Pi_r, \quad k = 0,\dots,2^n-1, \tag{1.52}$$

en utilisant la lettre S_n pour ne pas confondre ces fonctions avec des solutions numériques f^n. De même, on désignera pour $0 < p \leq \infty$

$$\sigma_n(f)_p := \inf_{S\in\Sigma_n} \|f - S\|_{L^p} \tag{1.53}$$

(plutôt que $\sigma_{2^n}(f)_p$) l'erreur optimale d'approximation de f dans L^p par des éléments de Σ_n.

Pour caractériser les fonctions pouvant être approchées dans L^p avec une vitesse 2^{-ns}, on a introduit dans la section 1.1 les espaces d'approximation $\mathcal{A}^s(L^p)$. Les théorèmes de caractérisation font en réalité intervenir de légères modifications de ces espaces, obtenus en faisant varier un troisième paramètre q : on définit ainsi l'*espace d'approximation* $\mathcal{A}^s_q(L^p)$ comme l'ensemble des fonctions $f \in L^p$ pour lesquelles la quantité

$$\|f\|_{\mathcal{A}^s_q(L^p)} := \begin{cases} \left(\sum_{n=-1}^{\infty} [2^{ns}\sigma_n(f)_p]^q\right)^{1/q} & \text{lorsque } q < \infty \\ \sup_{n\geq -1} 2^{ns}\sigma_n(f)_p & \text{lorsque } q = \infty \end{cases} \tag{1.54}$$

est finie. On suppose ici que l'ensemble Σ_{-1} ne contient que la fonction nulle, de sorte que $\sigma_{-1}(f)_p$ vaut $\|f\|_{L^p}$, et $\|\cdot\|_{\mathcal{A}^s_q(L^p)}$ est une norme pour laquelle $\mathcal{A}^s_q(L^p)$ est un espace de Banach.

Remarque 1.7 *De même que les erreurs optimales σ_n, ces espaces sont toujours définis pour une* méthode d'approximation donnée*, et dépendent donc du degré r choisi pour les fonctions* (1.52) *de Σ_n. Pour ne pas alourdir les notations, on suivra toutefois le choix de DeVore [29] en ne rappelant pas cette dépendance de façon explicite.*

On peut également définir ces espaces à partir des erreurs optimales $\sigma_N(f)_p$ réalisées sur les ensembles Σ_N de complexités entières, dont les $\Sigma_n := \Sigma_{2^n}$ sont une suite extraite, en posant

$$\|f\|_{\tilde{\mathcal{A}}^s_q(L^p)} := \begin{cases} \left(\sum_{N=-1}^{\infty} [N^s \sigma_N(f)_p]^q \frac{1}{N}\right)^{1/q} & \text{lorsque } q < \infty \\ \sup_{N \geq -1} N^s \sigma_N(f)_p & \text{lorsque } q = \infty. \end{cases} \tag{1.55}$$

Lorsque les Σ_N sont emboîtés, la suite $\sigma_N(f)_p$ est décroissante et on a

$$C(s)[2^{ns}\sigma_n(f)_p]^q \leq \sum_{N=2^n}^{2^{n+1}-1} [N^s \sigma_N(f)_p]^q \frac{1}{N} \leq C'(s)[2^{ns}\sigma_n(f)_p]^q \tag{1.56}$$

pour tout n, de sorte que ces deux normes sont équivalentes, et les espaces sont les mêmes. Clairement, $\mathcal{A}^s_\infty(L^p)$ correspond à l'espace $\mathcal{A}^s(\mathcal{X})$ introduit en (1.2), et pour $q < \infty$, les $\mathcal{A}^s_q(L^p)$ en sont une légère variation dans la mesure où l'on a

$$\mathcal{A}^{s+\eta}_\infty(L^p) \subset \mathcal{A}^s_q(L^p) \subset \mathcal{A}^s_\infty(L^p) \tag{1.57}$$

pour tout q et tout $\eta > 0$.

Remarque 1.8 *Si les fonctions $S_n \in \Sigma_n$, $n = -1, 0, \ldots$ forment une suite d'approximations de f optimale à une constante multiplicative près*

$$\|f - S_n\|_{L^p} \leq C\sigma_n(f)_p, \tag{1.58}$$

on obtient une norme équivalente à (1.54) en y remplaçant les termes $\sigma_n(f)_p$ par les différences $\|S_{n+1} - S_n\|_{L^p}$. En effet, on a d'une part

$$\|S_{n+1} - S_n\|_{L^p} \leq C(\sigma_{n+1}(f)_p + \sigma_n(f)_p)$$

et d'autre part, S_n tend vers f dans L^p, de sorte que $\|f - S_n\|_{L^p}$ est majoré par $\sum_{n' \geq n} \|S_{n'+1} - S_{n'}\|_{L^p}$. D'après l'inégalité discrète de Hardy, ceci entraîne

$$\sum_{n=-1}^{\infty} \left(2^{ns}\|f - S_n\|_{L^p}\right)^q \leq C \sum_{n=-1}^{\infty} \left(2^{ns}\|S_{n+1} - S_n\|_{L^p}\right)^q$$

pour tout $s > 0$ et tout $q > 0$, et de façon immédiate, ceci entraîne également

$$\sup_{n \geq -1} 2^{ns}\|f - S_n\|_{L^p} \leq \sum_{n=-1}^{\infty} 2^{ns}\|S_{n+1} - S_n\|_{L^p}.$$

1.3.1 Caractérisation de l'approximabilité dans L^p

Dans l'introduction, on a évoqué le fait que les espaces d'approximations pouvaient être identifiés avec des espaces de régularité. On peut maintenant être plus précis, et écrire que pour tout couple (s, p), il existe une valeur de q pour laquelle l'espace $\mathcal{A}^s_q(L^p)$ coïncide avec un espace de Besov. Prenons donc un instant pour présenter ces espaces.

Spontanément, on pourra voir les espaces de Besov comme un moyen (parmi d'autres)

de combler le vide laissé par les espaces de Sobolev $W^{s,p}$ lorsque qu'on souhaite parler de fonctions ayant "s dérivées dans L^p" avec s non entier. Pour des valeurs de s strictement comprises entre 0 et 1, on a déjà présenté les espaces de Hölder $\mathcal{C}^s$ qui contiennent les fonctions $f \in L^\infty$ pour lesquelles

$$\sup_{x\in[0,1-h]} |f(x+h) - f(x)| \leq Ch^s, \qquad h \in]0,1[, \tag{1.59}$$

et qu'on peut munir de la norme $\|f\|_{\mathcal{C}^s} = \|f\|_{L^\infty} + |f|_{\mathcal{C}^s}$ en désignant par $|f|_{\mathcal{C}^s}$ la plus petite constante possible apparaissant dans (1.59). A leur façon, ces espaces comblent le vide laissé entre L^∞ et $W^{1,\infty}$, puisque on retrouve ces deux espaces lorsqu'on prend s respectivement égal à 0 et 1 dans (1.59). De la même façon, on peut donner un sens à l'assertion "f possède s dérivées dans L^p" lorsque $0 < s < 1$ en considérant les espaces $\mathrm{Lip}(s, L^p)$ qui contiennent les fonctions $f \in L^p$ pour lesquelles la semi-norme

$$|f|_{\mathrm{Lip}(s,L^p)} := \sup_{0<h<1} h^{-s} \|f(\cdot + h) - f\|_{L^p([0,1-h])} \tag{1.60}$$

est finie (espace qu'on peut munir, à nouveau, de la norme $\|f\|_{\mathrm{Lip}(s,L^p)} = \|f\|_{L^p} + |f|_{\mathrm{Lip}(s,L^p)}$). Clairement, $\mathrm{Lip}(1, L^\infty)$ est l'espace $W^{1,infty}$ des fonctions lipschitziennes, mais remarquons qu'on ne retrouve pas toujours $W^{1,p}$ en prenant $s = 1$ dans cette définition. Ainsi $\mathrm{Lip}(1, L^1)$ coïncide-t-il avec l'espace BV composé des fonctions de L^1 dont la *variation totale*

$$|f|_{BV} := \sup_{\substack{\varphi\in\mathcal{C}_c^\infty(]0,1[) \\ \|\varphi\|_{L^\infty}\leq 1}} \langle f, \varphi' \rangle \tag{1.61}$$

est bornée. Lorsque f' est dans L^1, la variation totale de f correspond à $|f|_{BV} = \|f'\|_{L^1}$, mais de façon plus générale, BV contient les fonctions dont la dérivée est une mesure de Radon μ, leur variation totale étant alors égale à la masse totale $|\mu|$ de cette mesure. On peut alors montrer qu'il s'injecte continuement dans L^∞, et que

$$|f|_{BV} = \sup\{\sum_i |f(x_i) - f(x_{i-1})| : 0 < x_0 < \ldots < x_m < 1, m \in \mathbb{N}\}. \tag{1.62}$$

En particulier, BV contient les fonctions constantes par morceaux qui sont discontinues, et n'est donc pas égal à $W^{1,1}$.

Les espaces de Besov peuvent se définir suivant le même principe. Bien entendu, il ne suffit pas pour cela de considérer la définition (1.60) pour des valeurs de s strictement supérieures à 1, car les seules fonctions qu'on obtient de cette façon sont les constantes. Pour représenter les dérivées de f d'ordre élevé, on utilise l'opérateur de différences finies d'ordre r donné par

$$\Delta_h^r := \Delta_h^1 \Delta_h^{r-1} = (\Delta_h^1)^r \quad \text{et} \quad \Delta_h^1 : f \to f(\cdot + h) - f, \tag{1.63}$$

à partir duquel on peut remplacer le module de continuité

$$\omega(f,t)_p := \|f(\cdot + h) - f\|_{L^p([0,1-h])} \tag{1.64}$$

par le *module de régularité* d'ordre k

$$\omega_k(f,t)_p := \sup_{0<h\leq t} \|\Delta_h^k f\|_{L^p}([0, 1-kh]), \qquad t \in]0, 1/k[\tag{1.65}$$

où la borne supérieure permet que $\omega_k(f,\cdot)_p$ soit une fonction croissante. On définit alors l'espace de Besov $B_q^{s,p}$ comme l'ensemble des fonctions pour lesquelles la semi-norme

$$|f|_{B_q^{s,p}} := \begin{cases} \left(\int_0^{1/k} \left(t^{-s} \omega_k(f,t)_p \right)^q \frac{\mathrm{d}t}{t} \right)^{1/q} & \text{lorsque } q < \infty \\ \sup_{t \in]0,1/k[} \; t^{-s} \omega_k(f,t)_p & \text{lorsque } q = \infty \end{cases} \tag{1.66}$$

est finie, k étant un entier strictement supérieur à la partie entière de s

$$k \geq \lfloor s \rfloor + 1, \tag{1.67}$$

les normes définies pour différentes valeurs de k étant alors toutes équivalentes (pour une présentation plus complète de ces espaces, le lecteur pourra consulter l'ouvrage de référence [30]). L'indice q joue ici un rôle semblable à celui des normes (1.54), c'est-à-dire que les différences induites par deux valeurs différentes de cet indice sont de second ordre par rapport aux deux premiers indices s et p (à la façon d'un ordre lexicographique). On a ainsi $B_q^{s_2,p} \subset B_{q'}^{s_1,p}$ pour tout $s_1 < s_2$ et $B_q^{s,p_2} \subset B_{q'}^{s,p_1}$ pour tout $p_1 < p_2$, quels que soient q et q'. D'autre part, la contrainte de q-intégrabilité pour la mesure $\frac{\mathrm{d}t}{t}$ est d'autant plus forte que q est petit, d'où $B_{q_1}^{s,p} \subset B_{q_2}^{s,p}$ lorsque $q_1 < q_2 \leq \infty$. En revanche, ce "troisième indice" joue un rôle important dans l'établissement des théorèmes de caractérisation, où il doit être finement réglé en fonction des paramètres s et p.

On observera que les espaces $\mathrm{Lip}(s, L^p)$ généralisent bien les espaces de Hölder $\mathcal{C}^s$, au sens où $\mathrm{Lip}(s, L^\infty) = \mathcal{C}^s$ pour $0 < s < 1$, et qu'à leur tour les Besov vérifient $B_\infty^{s,p} = \mathrm{Lip}(s, L^p)$ pour $0 < s < 1$. En raison de la contrainte (1.67), $\mathrm{Lip}(1, L^p)$ ne coïncide en revanche plus avec $B_\infty^{1,p}$, qui est un peu plus grand. Pour la même raison l'espace $B_q^{s,p}$ ne coïncide en général pas avec l'espace de Sobolev $W^{s,p}$ lorsque s est entier (alors que la définition classique des espaces de Sobolev fractionnaires correspond à $W^{s,p} = B_p^{s,p}$, mais nous n'utiliserons pas ces espaces fractionnaires). On a par exemple lorsque $s = 1$, $W^{1,1} \subsetneq \mathrm{Lip}(1, L^1) = BV \subsetneq B_\infty^{1,1}$. Citons tout de même l'exception des espaces de Hilbert $H^s = W^{s,2}$ qui coïncident avec $B_2^{s,2}$ pour tout s, mais gardons surtout en tête qu'il est toujours possible d'"ordonner" les espaces de Sobolev et de Besov grâce aux relations

$$B_q^{s',p} \subset W^{s,p} \subset B_\infty^{s,p} \quad \text{pour} \quad s' > s \in \mathbb{N} \quad \text{et} \quad 0 < p,\, q \leq \infty. \tag{1.68}$$

On a alors le théorème suivant :

Théorème 1.1 (DeVore, Petrushev, Popov (88)) *Lorsque p est fini, les espaces $\mathcal{A}_q^s(L^p)$ associés à l'approximation polynomiale par morceaux de degré r sur des partitions libres sont caractérisés par*

$$\mathcal{A}_q^s(L^p) = B_q^{s,q} \quad \textit{avec} \quad 1/q = s + 1/p \tag{1.69}$$

pour tout $s < r + 1$, et les normes de ces espaces sont équivalentes.

Le lecteur pourra trouver une preuve de ce théorème dans les ouvrages [30] ou [29]. Indiquons tout de même que cette preuve repose d'une part sur des arguments d'interpolation fonctionnelle entre espaces de Besov [33], et d'autre part sur les estimations suivantes établies par Petrushev [54] :

$$\sigma_n(f)_p \leq C 2^{-sn} \|f\|_{B_q^{s,q}} \tag{1.70}$$

et

$$S \in \Sigma_n \implies \|S\|_{B_q^{s,q}} \leq C 2^{sn} \|S\|_{L^p}. \tag{1.71}$$

Ce type d'inégalités est central en approximation non-linéaire. (1.70) est ce qu'on appelle une estimation de Jackson, qui exprime un ordre de convergence des erreurs optimales sous des hypothèses de régularité. C'est une estimation "directe", qui permet d'établir des inclusions dans le sens $B_q^{s,q} \subset \mathcal{A}_q^s(L^p)$. Quant à (1.71), il s'agit d'une estimation de Bernstein, souvent qualifiée d'estimation "inverse" dans la mesure où elle permet de remonter, par le biais d'une suite d'approximations optimales, à la régularité des fonctions "bien approchées". En d'autres termes, elle fournit un argument essentiel pour établir l'inclusion $\mathcal{A}_q^s(L^p) \subset B_q^{s,q}$.

Comme on l'a annoncé plus haut, ce théorème ne permet pas de caractériser *tous* les espaces d'approximation $\mathcal{A}_q^s(L^p)$, mais *un d'entre eux* pour chaque couple (s, p). On peut néanmoins en déduire

$$\mathcal{A}_\infty^{s_2}(L^p) \subset W^{s,q} \subset \mathcal{A}_\infty^{s_1}(L^p) \quad \text{pour } 1/q = s + 1/p \text{ et tout } s_1 < s < s_2. \tag{1.72}$$

D'autre part, la limitation sur l'ordre s ne doit pas nous surprendre : elle correspond à la rigidité d'une méthode d'approximation utilisant des polynômes d'ordre peu élevé, et au phénomène de saturation qui en résulte, déjà mentionné dans la section 1.2.1.

1.3.2 Caractérisation de l'approximabilité dans L^∞

On considère maintenant l'approximation dans L^∞ des fonctions continues sur l'intervalle $[0, 1]$, et pour des raisons qui apparaîtront clairement dans quelques lignes, on commence par redéfinir $\Sigma_n = \Sigma_{n,r}$ comme l'ensemble des fonctions *continues* qui sont polynomiales de degré inférieur ou égal à r sur moins de 2^n intervalles. Ce type d'approximation a également été étudiée par Petrushev dans [54], où les estimations de Jackson et de Bernstein

$$\sigma_n(f)_\infty \leq C 2^{-sn} \|f'\|_{B_q^{s-1,q}} \tag{1.73}$$

et

$$S \in \Sigma_n \implies \|S'\|_{B_q^{s-1,q}} \leq C 2^{sn} \|S\|_{L^\infty}, \tag{1.74}$$

sont établies pour $1 < s < r + 1$ et $q = 1/s$. A partir des ces inégalités, il est alors possible d'identifier les espaces $\mathcal{A}_q^s(L^\infty)$ avec

$$\tilde{B}^s := \{f \in W^{1,1}(\mathbb{R}) : f' \in B_q^{s-1,q}, q = 1/s\} \tag{1.75}$$

muni de la norme

$$\|f\|_{\tilde{B}^s} := \|f\|_{L^\infty} + \|f'\|_{B_q^{s-1,q}}. \tag{1.76}$$

Cet espace ressemble à l'espace de Besov $B_q^{s,q}$, mais on observera qu'il est sensiblement plus petit que ce dernier, qui contient des fonctions discontinues lorsque $q < 1$.

Théorème 1.2 *Les espaces $\mathcal{A}_q^s(L^\infty)$ associés à l'approximation polynomiale par morceaux de degré r sont caractérisés par*

$$\mathcal{A}_q^s(L^\infty) = \tilde{B}^s \quad \textit{avec} \quad 1/q = s \tag{1.77}$$

lorsque $1 < s < r + 1$, et leurs normes sont équivalentes.

Preuve. On peut établir ce résultat de façon directe à partir du théorème 1.1. Pour une fonction f de $\mathcal{A}_q^s(L^\infty)$, on désigne par $S_n \in \Sigma_n$, $n = 0, 1, \ldots$ une suite approchant f de façon quasi-optimale. On considère alors les fonctions discontinues $T_n := S_n'$ polynomiales de degré inférieur ou égal à $r-1$ sur 2^n morceaux, comme des approximations de f'. Observons qu'un polynôme S de degré r vérifie toujours

$$\|S'\|_{L^1([a,b])} \leq C\|S\|_{L^\infty([a,b])} \tag{1.78}$$

avec une constante C pouvant dépendre de r mais pas de l'intervalle $[a,b]$ par un argument de changement d'échelle. Comme $T_n - T_{n-1}$ est une fonction polynomiale sur moins de $\frac{3}{2}2^n$ intervalles I_k, on a

$$\|T_n - T_{n-1}\|_{L^1} \leq \sum_k \|T_n - T_{n-1}\|_{L^1(I_k)} \leq C2^n\|S_n - S_{n-1}\|_{L^\infty}. \tag{1.79}$$

En utilisant la remarque 1.8 (et en rappelant que la norme $\|\cdot\|_{\mathcal{A}_q^s(L^\infty)}$ est définie en (1.54)), ceci entraîne que

$$\sum_{n=-1}^{\infty} [2^{n(s-1)}\|T_n - T_{n-1}\|_{L^1}]^q \leq C\|f\|^q_{\mathcal{A}_q^s(L^\infty)}, \tag{1.80}$$

d'où l'on déduit que la suite T_n converge dans L^1 vers une fonction qui est forcément f'. Il s'ensuit que

$$\|f'\|_{\mathcal{A}_q^{s-1}(L^1)} \leq C\|f\|_{\mathcal{A}_q^s(L^\infty)}, \tag{1.81}$$

et d'après le théorème 1.1 avec $p = 1$,

$$\|f'\|_{B_q^{s-1,q}} \leq C\|f\|_{\mathcal{A}_q^s(L^\infty)}. \tag{1.82}$$

Comme il est clair que $\|f\|_{L^\infty} \leq \|f\|_{\mathcal{A}_q^s(L^\infty)}$, on en déduit que

$$\|f\|_{\tilde{B}^s} \leq C\|f\|_{\mathcal{A}_q^s(L^\infty)}. \tag{1.83}$$

Dans l'autre sens, considérons que f appartient à $\tilde{B}^s$. f' appartient donc à $B_q^{s-1,q}$ avec $1/q = 1 + (s-1)$, ce qui entraîne d'après le théorème 1.1 que f' est dans $\mathcal{A}_q^{s-1}(L^1)$ avec $\|f'\|_{\mathcal{A}_q^{s-1}(L^1)} \leq C\|f'\|_{B_q^{s-1,q}}$. Il existe alors une suite de fonctions discontinues T_n, $n = -1, 0, \ldots$ avec $T_{-1} = 0$, qui sont polynomiales de degré au plus $r-1$ sur 2^n morceaux et approchent f' de façon quasi-optimale dans L^1, de sorte que

$$\sum_{n=-1}^{\infty} \|f' - T_n\|^q_{L^1} \leq C\|f'\|^q_{B_q^{s-1,q}}. \tag{1.84}$$

Comme il existe pour chaque n une partition de I en 2^n intervalles J_k tels que $\|f' - T_n\|_{L^1(J_k)} \leq 2^{-n}\|f' - T_n\|_{L^1}$, et que chaque T_n est constitué de 2^n morceaux polynomiaux, il est toujours possible de construire une partition de I en 2^{n+1} intervalles I_k sur lesquels T_n est un polynôme et tels que

$$\|f' - T_n\|_{L^1(I_k)} \leq 2^{-n}\|f' - T_n\|_{L^1}. \tag{1.85}$$

Sur chacun de ces intervalles $I_k = [a_k, b_k]$, on pose

$$P_{n+1}(x) := f(a_k) + \int_{a_k}^{x} T_n(s)\,\mathrm{d}s \tag{1.86}$$

qu'on modifie ensuite en

$$S_{n+1}(x) := P_{n+1}(x) + (f(b_k) - P_{n+1}(b_k))\frac{x - a_k}{b_k - a_k}. \tag{1.87}$$

On peut vérifier que la fonction S_{n+1} ainsi obtenue appartient à Σ_{n+1} : elle est bien polynomiale de degré au plus r sur les intervalles I_k qui sont au nombre de 2^{n+1}, et elle est continue. Sur chaque I_k, on a clairement

$$|f(x) - P_{n+1}(x)| \leq \|f' - T_n\|_{L^1(I_k)} \leq 2^{-n}\|f' - T_n\|_{L^1} \tag{1.88}$$

d'après (1.85), et de même

$$|f(x) - P_{n+1}(x)|\frac{x - b_k}{a_k - b_k} \leq 2^{-n}\|f' - T_n\|_{L^1}. \tag{1.89}$$

On en déduit que

$$\|u - S_{n+1}\|_{L^\infty} \leq 2^{-n+1}\|f' - T_n\|_{L^1}, \tag{1.90}$$

ce qui entraîne

$$\|f\|^q_{\mathcal{A}^s_q(L^\infty)} \leq C(\|f\|^q_{L^\infty} + \|f'\|^q_{\mathcal{A}^{s-1}_q(L^1)}). \tag{1.91}$$

En utilisant à nouveau le théorème 1.1 pour $p = 1$, on trouve alors finalement

$$\|f\|^q_{\mathcal{A}^s_q(L^\infty)} \leq C\|f\|_{\tilde{B}^s}, \tag{1.92}$$

ce qui termine cette preuve. □

A nouveau, on pourra signaler que ce théorème entraîne les inclusions suivantes

$$\mathcal{A}^{s_2}_\infty(L^\infty) \subset W^{s,q} \subset \mathcal{A}^{s_1}_\infty(L^\infty) \quad \text{pour } 1/q = s \text{ et tout } s_1 < s < s_2. \tag{1.93}$$

Chapitre 2

Discrétisations adaptatives multi-échelles de type $\mathcal{P}^1$

On introduit dans ce chapitre une classe $\mathcal{M}(\mathbb{R}^d)$ de maillages dyadiques multi-échelles de $\mathbb{R}^d$, auxquels on associe en dimension 2 des éléments finis conformes de type $\mathcal{P}^1$, autrement dit affines par morceaux. La raison pour laquelle on introduit ces discrétisations adaptatives est que la structure géométrique des mailles dyadiques nous permettra de proposer dans la deuxième partie un schéma alliant *une gestion algorithmique très simple* des maillages de calcul à *une analyse accessible* de leurs propriétés, notamment en ce qui concerne leur caractère optimal dans l'approximation adaptative des solutions.

En prévision de cette analyse, on montrera que la semi-norme $|\cdot|_{W^{2,1}}$ constitue un bon indicateur *a priori* de l'erreur d'interpolation locale mesurée dans L^∞, au sens où

1. pour toute fonction f appartenant à $W^{2,1}(\mathbb{R}^2)$, on indiquera comment construire un maillage dyadique M par un algorithme de découpage récursif guidé par les semi-normes *locales* $|f|_{W^{2,1}(\alpha)}$.
2. le maillage ainsi obtenu permettra (via son espace d'éléments finis associé) d'approcher f dans L^∞ avec une précision ε fixée à l'avance
3. la complexité de ce maillage sera de l'ordre de ε^{-1} dès que f appartient à $W^{2,p}$ avec $p > 1$.

En particulier, cet indicateur nous donnera un moyen pratique d'approcher une fonction f de façon *adaptative* par des approximants f_M affines par morceaux, avec une vitesse de convergence en $\|f - f_M\|_{L^\infty(\mathbb{R}^2)} \leq C\#(M)^{-1}$. Au regard des résultats de caractérisation donnés par la théorie de l'approximation non-linéaire, cette vitesse peut être considérée comme quasiment optimale lorsque f est dans un espace $W^{2,p}(\mathbb{R}^2)$ avec p proche de 1. En comparaison, l'interpolation affine par morceaux sur un maillage uniforme n'atteint cette vitesse que lorsque f appartient à l'espace $W^{2,\infty}(\mathbb{R}^2)$.

2.1 Partitions adaptatives dyadiques

Définition 2.1 (partitions dyadiques) *On dira qu'une partition M de $\mathbb{R}^d$ est* dyadique *si elle ne contient que des cellules de la forme*

$$\alpha_{\ell,k} = \Pi_{1\leq i\leq d}[2^{-\ell}k_i, 2^{-\ell}(k_i+1)[, \quad \textit{avec } \ell \in \mathbb{N} \textit{ et } k \in \mathbb{Z}^d. \tag{2.1}$$

L'entier ℓ est appelé le niveau *de la cellule $\alpha_{\ell,k}$, et on ne considérera dans la suite que des niveaux supérieurs à un $\ell_0 > 0$ fixé.*

Pour $\ell \geq \ell_0$, on peut donc définir la partition dyadique *uniforme* de niveau ℓ par

$$\mathbb{D}_\ell(\mathbb{R}^d) := \{\alpha_{\ell,k} = \Pi_{1\leq i\leq d}[2^{-\ell}k_i, 2^{-\ell}(k_i+1)[: k \in \mathbb{Z}^d\}, \tag{2.2}$$

et poser $\mathbb{D}_\ell(\mathbb{R}^d) = \emptyset$ lorsque $\ell < \ell_0$. Avec cette convention, on écrira $\mathbb{D}(\mathbb{R}^d) := \cup_\ell \mathbb{D}_\ell(\mathbb{R}^d)$ l'ensemble de toutes les cellules dyadiques, et on désignera $\ell(\alpha) \geq \ell_0$ le niveau d'une cellule dyadique donnée.

2.1.1 Structure multi-échelle des cellules dyadiques

Les niveaux successifs $\mathbb{D}_\ell(\mathbb{R}^d)$, $\ell \geq \ell_0$ formant une suite de partitions emboîtées, on peut les munir d'une structure multi-échelle en définissant pour chaque cellule $\alpha \in \mathbb{D}(\mathbb{R}^d) \setminus \mathbb{D}_{\ell_0}(\mathbb{R}^d)$ son unique *parente* $\mathcal{P}(\alpha)$ définie par

$$\mathcal{P}(\alpha) \in \mathbb{D}_{\ell(\alpha)-1}(\mathbb{R}^d) \quad \text{et} \quad \alpha \subset \mathcal{P}(\alpha) \tag{2.3}$$

et ses 2^d *filles* par $\mathcal{F}(\alpha)$ définies par

$$\mathcal{F}(\alpha) := \{\beta \in \mathbb{D}_{\ell(\alpha)+1}(\mathbb{R}^d) : \beta \subset \alpha\}. \tag{2.4}$$

A partir de ces relations, on dira qu'un ensemble $\Lambda \subset \mathbb{D}(\mathbb{R}^d)$ est un *arbre* s'il vérifie

$$\mathbb{D}_{\ell_0}(\mathbb{R}^d) \subset \Lambda \quad \text{et} \quad \alpha \in \Lambda \implies \mathcal{P}(\alpha) \in \Lambda \tag{2.5}$$

(à l'exception, bien entendu, des cellules α de niveau ℓ_0 qui n'ont pas de parente), et qu'il est un *arbre-partition* si

$$\alpha \in \Lambda \implies \mathcal{F}(\alpha) \subset \Lambda \quad \text{ou} \quad \mathcal{F}(\alpha) \cap \Lambda = \emptyset. \tag{2.6}$$

On peut alors vérifier que les *feuilles internes* de Λ, définies par

$$\partial\Lambda := \{\alpha \in \Lambda : \mathcal{F}(\alpha) \cap \Lambda = \emptyset\}, \tag{2.7}$$

forment toujours une partition dyadique de $\mathbb{R}^d$.

Remarque 2.2 *Inversement, on peut observer que si M est une partition dyadique, le plus petit arbre $\Lambda = \Lambda(M)$ contenant les cellules de M est un arbre-partition, qu'il vérifie $\partial\Lambda(M) = M$, et qu'il est le seul à avoir cette propriété.*

Remarque 2.3 *Quitte à ajouter un niveau supplémentaire $\ell_0 - 1$, et à remplacer* (2.5) *par*

$$\mathbb{D}_{\ell_0-1}(\mathbb{R}^d) \subset \Lambda \quad \textit{et} \quad \alpha \in \Lambda \implies \mathcal{P}(\alpha) \in \Lambda, \tag{2.8}$$

on peut voir de façon équivalente une partition dyadique comme les feuilles internes d'un arbre-partition vérifiant (2.5)-(2.6) *ou comme les feuilles externes*

$$\partial^{ext}\Lambda := \{\alpha \in \mathbb{D}(\mathbb{R}^d) \setminus \Lambda : \mathcal{P}(\alpha) \in \Lambda\} \tag{2.9}$$

d'un arbre Λ vérifiant simplement (2.8).

2.1.2 L'algorithme de découpage dyadique récursif

D'après la remarque 2.2, une partition dyadique peut toujours être obtenue par un processus récursif de découpages dyadiques dans lequel, à partir des cellules les plus grosses du niveau ℓ_0, on découpe en quatre chaque cellule suivant un objectif particulier. On rencontrera par exemple dans la suite la situation où on souhaite construire la partition dyadique M contenant le moins de cellules possible et telle que la semi-norme $|f|_{W^{2,1}(\alpha)}$ est inférieure à un $\varepsilon > 0$ fixé pour tout $\alpha \in M$. On dira donc qu'une cellule α (générale) est *de bonne qualité* si $|f|_{W^{2,1}(\alpha)} \leq \varepsilon$, et *de mauvaise qualité* dans le cas contraire. Pour construire notre partition M, il nous suffit de raffiner récursivement les cellules de mauvaise qualité à partir de la partition "racine" $\mathbb{D}_{\ell_0}(\mathbb{R}^d)$, comme l'illustre la figure 2.1 ci-dessous. On peut écrire cet algorithme sous la forme suivante.

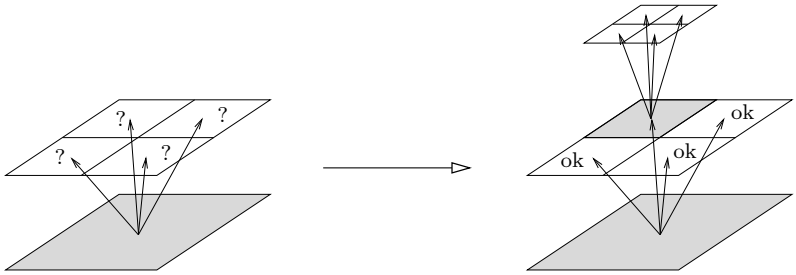


FIG. 2.1 – construction d'une partition dyadique par découpage récursif des mailles.

Algorithme 2.4 (découpage dyadique guidé par la semi-norme $W^{2,1}$)

- *Poser* $\Lambda_{\ell_0} := \mathbb{D}_{\ell_0}(\mathbb{R}^d)$.
- *Pour* $\ell \geq \ell_0$, *calculer*

  $$\Lambda_{\ell+1} := \Lambda_\ell \cup \{\alpha \in \mathcal{F}(\beta) : \beta \in \Lambda_\ell \text{ et } |f|_{W^{2,1}(\beta)} > \varepsilon\}$$

  jusqu'à ce que $\Lambda_{L+1} = \Lambda_L$, *et prendre* $\tilde{M} = \partial\Lambda_L$.

Remarque 2.5 *Lors du raffinement conditionnel de l'arbre Λ_ℓ, dans la deuxième étape de l'algorithme, seules les cellules de niveau ℓ sont susceptibles d'être raffinées.*

Si cet algorithme converge, autrement dit s'il existe bien un niveau maximal $L = L(f)$ pour lequel $\Lambda_{L+1} = \Lambda_L$, il est clair que la partition M constituée des feuilles de Λ_L vérifiera

$$|f|_{W^{2,1}(\alpha)} \leq \varepsilon \quad \text{pour toute cellule } \alpha \in M, \tag{2.10}$$

tandis que les cellules internes $\beta \in \Lambda_L \setminus M$ représentent les cellules de mauvaise qualité qu'il a fallu subdiviser pour obtenir M. On s'intéressera bientôt (voir en particulier les théorèmes 2.1 et 3.2) à la complexité de M, autrement dit au nombre minimal de cellules dont une partition dyadique a besoin pour satisfaire (2.10).

2.1.3 Partitions dyadiques graduées

Dans nos applications, on aura besoin que la résolution des maillages ne varie pas de façon brutale d'une cellule à une autre. La notion que nous utiliserons est la suivante.

Définition 2.6 (partitions dyadiques graduées) *Une partition dyadique M sera dite* graduée *si deux cellules voisines y sont toujours de niveaux voisins,* i.e.

$$\alpha \in M,\ \beta \in M,\ \bar{\alpha} \cap \bar{\beta} \neq \emptyset \implies |\ell(\alpha) - \ell(\beta)| \leq 1. \tag{2.11}$$

On notera dans la suite $\mathcal{M}(\mathbb{R}^d)$ l'ensemble des partitions dyadiques graduées de $\mathbb{R}^d$, qu'on appellera plus simplement **maillages dyadiques**.

Remarque 2.7 *En notant, pour un niveau donné ℓ,*

$$\mathcal{V}_\ell(\alpha) := \{\beta \in \mathbb{D}_\ell(\mathbb{R}^d) : \alpha \not\subset \beta, \beta \not\subset \alpha, \bar{\alpha} \cap \bar{\beta} \neq \emptyset\} \tag{2.12}$$

l'ensemble de toutes les voisines de niveau ℓ *d'une cellule α, on peut observer que la propriété* (2.11) *est équivalente à*

$$\alpha \in \Lambda(M) \implies \mathcal{V}_{\ell(\alpha)-1}(\alpha) \subset \Lambda(M), \tag{2.13}$$

où $\Lambda(M)$ désigne l'unique arbre tel que $\partial\Lambda(M) = M$.

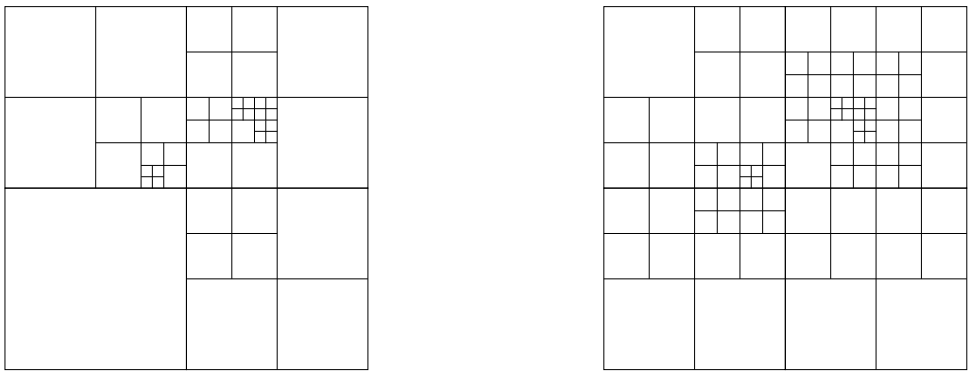


FIG. 2.2 – une partition dyadique $\tilde{M}$ et son plus petit raffinement gradué M.

Une façon d'obtenir des maillages dyadiques est de raffiner en cascade une partition dyadique là où elle n'est pas graduée : à chaque fois que deux cellules α et β sont voisines et que leurs niveaux vérifient $\ell(\alpha) \geq \ell(\beta) + 2$, on raffine β en la remplaçant par ses filles. En appliquant ce procédé à une partition dyadique $\tilde{M}$ de façon systématique (ou bien en suivant l'algorithme 2.8 ci-dessous) on construit un maillage dyadique M indépendant de l'ordre des raffinements, et qui a la propriété d'être, parmi tous les raffinements gradués de $\tilde{M}$, celui qui a le moins de cellules. On l'appellera le *plus petit raffinement gradué* de $\tilde{M}$, et on peut proposer l'algorithme suivant pour le construire de façon qui nous semble efficace a priori (l'efficacité réelle d'un tel algorithme dépendant en pratique de la structure de donnée choisie pour matérialiser les partitions dyadiques).

Algorithme 2.8 (plus petit raffinement gradué d'une partition dyadique)

- *Poser* $\Lambda_{\ell_0} := \Lambda(\tilde{M})$.
- *Pour* $\ell \geq \ell_0$, *calculer*
  $\Lambda_{\ell+1} := \Lambda_\ell \cup \{\alpha \in \mathcal{F}(\beta) : \beta \in \Lambda_\ell, \ell(\beta) = \ell \text{ et } \mathcal{V}_{\ell+2}(\beta) \cap \Lambda_\ell \neq \emptyset\}$.
- *Prendre* $M := \partial(\Lambda_{\ell(\tilde{M})-1})$.

En particulier, on observera que chaque Λ_ℓ est un arbre-partition au sens de (2.5)-(2.6), dans lequel les cellules de niveau inférieur ou égal à $\ell+1$ forment un arbre gradué. On en déduit que l'algorithme converge en moins de $\ell(\tilde{M})-\ell_0-1$ étapes, au sens où $\Lambda_{\ell(\tilde{M})}$, si on le calculait, serait toujours égal à $\Lambda_{\ell(\tilde{M})-1}$.

Une propriété très intéressante est que l'ordre de complexité est préservé par l'algorithme 2.8. Plus précisément, et quitte à définir le *cardinal réduit* d'une partition dyadique

$$\mathrm{card}_{\ell_0}(M) := \#\big(\{\alpha \in M : \ell(\alpha) > \ell_0\}\big) \leq \#(M) \tag{2.14}$$

de façon à pouvoir considérer des partitions "finies" sur des domaines non bornés, on peut montrer le résultat suivant (dont on trouvera une version pour un domaine borné dans la section 2.1.4).

Proposition 2.9 *Si le cardinal réduit* (2.14) *de la partition dyadique $\tilde{M}$ est fini, son plus petit raffinement gradué M vérifie*

$$\mathrm{card}_{\ell_0}(\tilde{M}) \leq \mathrm{card}_{\ell_0}(M) \leq \frac{6^d}{2^d-1}\mathrm{card}_{\ell_0}(\tilde{M}), \tag{2.15}$$

cette constante n'étant sans doute pas optimale.

Preuve. L'inégalité de gauche est évidente. Pour établir l'inégalité de droite, on peut commencer par observer que l'arbre $\tilde{\Lambda} = \tilde{\Lambda}(\tilde{M})$ dont les feuilles forment la partition dyadique $\tilde{M}$ vérifie

$$\mathrm{card}_{\ell_0}(\tilde{M}) \leq \mathrm{card}_{\ell_0}(\tilde{\Lambda}) \leq \frac{2^d}{2^d-1}\mathrm{card}_{\ell_0}(\tilde{M}). \tag{2.16}$$

Ici encore, l'inégalité de gauche est triviale, et on peut montrer celle de droite par le raisonnement suivant : $\tilde{\Lambda}$ peut être obtenu par r raffinements depuis la racine $\mathbb{D}_{\ell_0}(\mathbb{R}^d)$, dont le cardinal réduit est nul. Comme chaque raffinement ajoute respectivement 2^d et 2^d-1 cellules à l'arbre et à ses feuilles, on a $\mathrm{card}_{\ell_0}(\tilde{\Lambda}) = r2^d$ et $\mathrm{card}_{\ell_0}(\tilde{M}) = r(2^d-1)$, d'où l'on déduit (2.16). On peut d'autre part écrire $\tilde{\Lambda}$ comme

$$\tilde{\Lambda} = \mathbb{D}_{\ell_0}(\mathbb{R}^d) \cup \left(\bigcup\nolimits_{\alpha\in\tilde{\Lambda}\setminus\partial\tilde{\Lambda}} \mathcal{F}(\alpha)\right), \tag{2.17}$$

et vérifier que l'ensemble

$$\Lambda := \mathbb{D}_{\ell_0}(\mathbb{R}^d) \cup \left(\bigcup\nolimits_{\alpha\in\tilde{\Lambda}\setminus\partial\tilde{\Lambda}} \left(\mathcal{F}(\alpha) \cup \bigcup\nolimits_{\beta\in\mathcal{V}_{\ell(\alpha)}(\alpha)} \mathcal{F}(\beta)\right)\right) \tag{2.18}$$

est un arbre gradué qui contient $\tilde{\Lambda}$, de sorte que $\partial\Lambda$ est un raffinement gradué de $\tilde{M}$. Comme la réunion (2.17) est disjointe, et que pour toute cellule α, on a

$$\#\left(\mathcal{F}(\alpha) \cup \bigcup\nolimits_{\beta\in\mathcal{V}_{\ell(\alpha)}(\alpha)} \mathcal{F}(\beta)\right) = \Big(1+\#\big(\mathcal{V}_{\ell(\alpha)}(\alpha)\big)\Big)\#\big(\mathcal{F}(\alpha)\big) = 3^d\#\big(\mathcal{F}(\alpha)\big), \tag{2.19}$$

on en déduit que

$$\mathrm{card}_{\ell_0}(\Lambda) \leq 3^d\mathrm{card}_{\ell_0}(\tilde{\Lambda}). \tag{2.20}$$

En utilisant les deux inégalités de (2.16), on obtient alors

$$\mathrm{card}_{\ell_0}(\partial\Lambda) \leq \mathrm{card}_{\ell_0}(\Lambda) \leq 3^d \mathrm{card}_{\ell_0}(\tilde{\Lambda}) \leq \frac{6^d}{2^d-1}\mathrm{card}_{\ell_0}(\tilde{M}). \tag{2.21}$$

Dans la mesure où le plus petit raffinement gradué M vérifie $\mathrm{card}_{\ell_0}(M) \leq \mathrm{card}_{\ell_0}(\partial\Lambda)$, on en déduit l'inégalité de droite de (2.15). □

Pour finir, on énoncera deux courtes propositions vérifiées par les maillages dyadiques, qui nous serviront dans la suite.

Proposition 2.10 (cellules voisines dans une partition graduée) *En désignant par $\mathcal{V}_M(\alpha) := \{\beta \in M : \bar{\beta} \cap \bar{\alpha} \neq \emptyset\}$ les voisines d'une cellule α dans une partition dyadique M, on a*

$$\sup_{\alpha\in M} \#\big(\mathcal{V}_M(\alpha)\big) \leq 3^d \cdot 2^{d-1} \tag{2.22}$$

lorsque la partition dyadique M est graduée, autrement dit lorsque M est un maillage dyadique au sens de la définition 2.6.

Preuve. Si α appartient à une partition graduée M, ses voisines dans M sont de niveau supérieur ou égal à $\ell(\alpha) - 1$, et il est clair que leur nombre est maximal lorsque leur niveau est minimal donc égal à $\ell(\alpha) - 1$. α possède alors $3^d - 1$ voisines β de même niveau *dans l'arbre $\Lambda(M)$ associé à M*, et chacune de ces voisines β possède exactement 2^d filles dont la moitié au plus est en contact avec α (et appartient à M). Comme on prend en compte, de cette façon, toutes les voisines de α dans M, (2.22) est établi. □

Proposition 2.11 *Si m et m' sont deux points appartenant respectivement à deux cellules α et α' d'une même partition dyadique graduée $M \in \mathcal{M}(\mathbb{R}^d)$, et si $\ell(\alpha)-\ell(\alpha') \geq 1$, alors*

$$\big|m - m'\big| \geq 2^{-\ell(\alpha')} - 2^{-\ell(\alpha)+1}, \tag{2.23}$$

où $|m| := \max_i |m_i|$ désigne la norme ℓ^∞ de $\mathbb{R}^d$.

Preuve. A nouveau, on peut observer que le cas critique, autrement dit la distance entre m et m' est minimale, s'obtient lorsque la graduation (ou la *gradation…*) est "saturée" autour de la cellule la plus grande, en l'occurrence il s'agit de α'. La graduation des cellules impose alors que des cellules de niveaux intermédiaires $\ell(\alpha')-1$, $\ell(\alpha')-2$, …, $\ell(\alpha)+1$ s'intercalent entre α' et α, de sorte que la distance minimale entre les points m et m' est au moins égale à $2^{-(\ell(\alpha')-1)} + 2^{-(\ell(\alpha')-2)} + \cdots + 2^{-(\ell(\alpha)+1)} = 2^{-\ell(\alpha')} - 2^{-\ell(\alpha)+1}$, ce qui apparaît assez clairement sur la figure 2.3. □

2.1.4 Maillages dyadiques sur des domaines bornés

Comme on ne souhaite généralement mailler que des domaines bornés, on pourrait adapter les constructions de ce chapitre à un pavé Ω de $\mathbb{R}^d$, ou à une réunion finie de pavés, en définissant une partition (respectivement, un maillage) dyadique de Ω comme

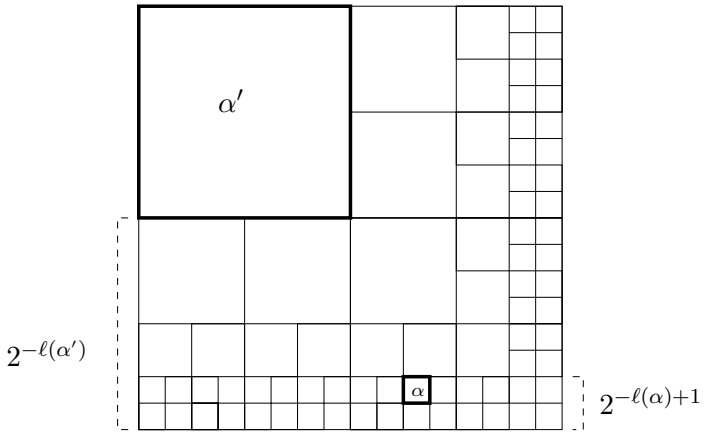


FIG. 2.3 – graduation saturée autour d'une cellule α'.

"l'intersection" avec Ω d'une partition (respectivement, d'un maillage) dyadique de $\mathbb{R}^d$, au sens de

$$M_\Omega = \{\alpha \in M : \alpha \cap \Omega \neq \emptyset\}.$$

On vérifie alors que M_Ω est un maillage dyadique de Ω si et seulement si la partition M est Ω-*graduée* au sens où la condition (2.11) est au moins vérifiée pour les cellules α et β qui intersectent Ω.

Pour écrire une version "domaine borné" de la proposition 2.9, on peut voir le cardinal de M_Ω comme une sorte de "cardinal réduit" de M

$$\mathrm{card}_\Omega(M) := \#(M_\Omega) = \#\big(\{\alpha \in M : \alpha \cap \Omega \neq \emptyset\}\big), \tag{2.24}$$

et vérifier que le plus petit raffinement Ω-gradué M d'une partition dyadique $\tilde{M}$ de $\mathbb{R}^d$ vérifie

$$\mathrm{card}_\Omega(\tilde{M}) \leq \mathrm{card}_\Omega(M) \leq C\mathrm{card}_\Omega(\tilde{M}) \tag{2.25}$$

pour une constante C dépendant uniquement de la dimension.

2.2 Eléments finis $\mathcal{P}^1$ conformes associés aux maillages adaptatifs dyadiques en dimension deux

En dimension $d = 2$, on montre à présent comment associer à chaque maillage dyadique $M \in \mathcal{M}(\mathbb{R}^2)$ une discrétisation par éléments finis de type $\mathcal{P}^1$. On désignera respectivement par V_M et P_M l'espace fonctionnel et l'interpolation affine par morceaux ainsi associés au maillage dyadique M.

2.2.1 Triangulations conformes

Grâce à la propriété de graduation, il est possible d'associer à M (dont les mailles sont carrées) une triangulation conforme de $\mathbb{R}^2$ qu'on notera $\mathcal{K}(M)$. On rappelle à cette occasion qu'une triangulation d'un domaine $\Omega \subset \mathbb{R}^2$ est dite conforme lorsque toute arête d'un de ses éléments est soit une arête d'un autre élément, soit une partie du

bord $\partial\Omega$.

Pour cela, on commençe par construire une triangulation $\tilde{\mathcal{K}}(M)$ *non conforme* en découpant chaque maille carrée α de M en deux triangles selon la règle suivante : si α est une fille "supérieure gauche" ou "inférieure droite" de sa cellule parente $\mathcal{P}(\alpha)$, on la divise en ses moitiés triangulaires "supérieure droite" et "inférieure gauche". Si par contre α est une fille "supérieure droite" ou "inférieure gauche" de sa parente, on la divise en ses moitiés "supérieure gauche" et "inférieure droite". Comme on peut l'observer sur la figure 2.4 (au milieu), cette règle revient à découper les filles d'une "super-cellule" $\mathcal{P}(\alpha)$ en suivant ses deux diagonales. Quant aux cellules de niveau initial ℓ_0, qui n'ont pas de parente, on peut les subdiviser de façon arbitraire.

La triangulation obtenue $\tilde{\mathcal{K}}(M)$, on le voit bien, contient des triangles non conformes provenant de cellules adjacentes de niveaux différents. Plus précisément, on peut observer que si α partage une de ses arêtes γ avec deux cellules β et λ de niveau supérieur $\ell(\beta) = \ell(\lambda) = \ell(\alpha) + 1$, le découpage de α, β et λ produit entre autres trois triangles α_t, β_t et λ_t qui se partagent l'arête γ et sont par conséquent non conformes. Comme M est gradué, cette situation est la seule qui puisse produire un défaut de conformité, et il suffit de réunir les triangles β_t et λ_t pour la faire disparaître, comme on peut le voir sur la figure 2.4 (à droite). En répétant ce procédé là où c'est nécessaire, on obtient bien une triangulation conforme $\mathcal{K}(M)$ qui vérifie clairement

$$\mathrm{card}_{\ell_0}(\mathcal{K}(M)) \leq 2\mathrm{card}_{\ell_0}(M) \tag{2.26}$$

en désignant par $\mathrm{card}_{\ell_0}(\mathcal{K}(M))$ le nombre de triangles issus de cellules de niveau $\ell > \ell_0$.

On pourra observer sur la figure 2.5 que la structure de M est, autant géométriquement qu'algorithmiquement, plus simple à manipuler que celle de sa triangulation associée $\mathcal{K}(M)$.

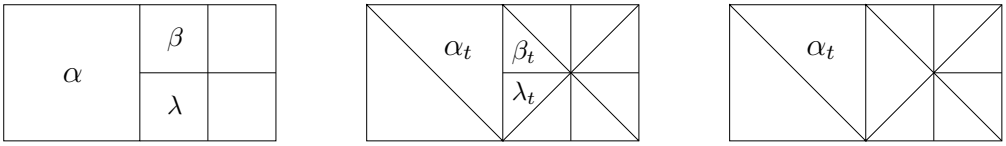


FIG. 2.4 – modifications locales des mailles pour déduire $\tilde{\mathcal{K}}(M)$ et $\mathcal{K}(M)$ de M.

2.2.2 Contrôle des erreurs d'interpolation par la semi-norme $W^{2,1}$

Pour contrôler les erreurs locales d'interpolation sur les triangles rectangles isocèles de $\mathcal{K}(M)$, on peut énoncer l'estimation suivante, dont la forme est très classique (voir par exemple [20]) :

Lemme 2.12 *Si K est un triangle rectangle isocèle, l'interpolation affine par morceaux P_K vérifie*

$$\|f - P_K f\|_{L^\infty(K)} \leq C|f|_{W^{2,1}(K)} \tag{2.27}$$

pour toute fonction $f \in W^{2,1}(K)$, avec une constante C absolue.

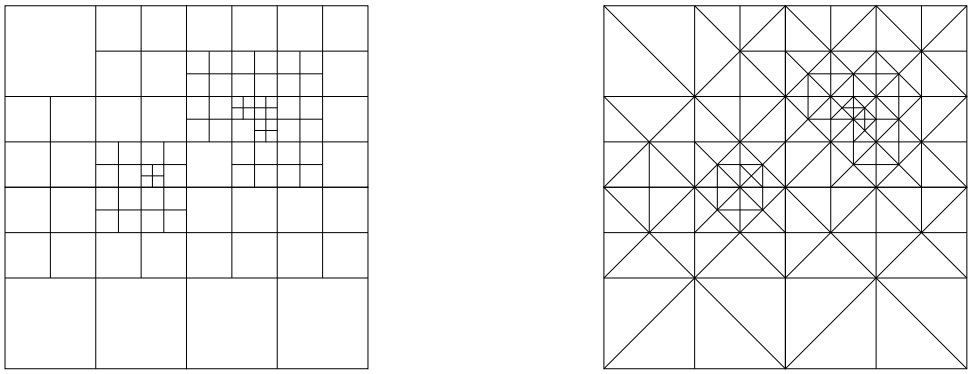


FIG. 2.5 – un maillage dyadique $M \in \mathcal{M}(\mathbb{R}^d)$ et sa triangulation conforme $\mathcal{K}(M)$.

Preuve. On peut raisonner en trois temps. (i) en considérant d'abord le triangle $\bar{K}$ de sommets $(0,0)$, $(1,0)$ et $(0,1)$, montrons que $W^{2,1}(\bar{K})$ s'injecte continuement dans $L^\infty(\bar{K})$. Pour le voir, calculons pour une fonction $g \in W^{2,1}(\bar{K})$ et pour $(x,y) \in \bar{K}$

$$\begin{aligned} g(x,y) &= g(x,0) + \int_0^y \partial_y g(x,v)\,\mathrm{d}v = g(x,0) + \int_0^y \partial_y g(0,v)\,\mathrm{d}v + \int_0^y \int_0^x \partial^2_{xy} g(u,v)\,\mathrm{d}u\,\mathrm{d}v \\ &= g(0,0) + \int_0^x \partial_x g(u,0)\,\mathrm{d}u + \int_0^y \partial_y g(0,v)\,\mathrm{d}v + \int_0^y \int_0^x \partial^2_{xy} g(u,v)\,\mathrm{d}u\,\mathrm{d}v. \quad (2.28) \end{aligned}$$

En écrivant

$$g(0,0) = (x,0) - \int_0^x \partial_x g(u,0)\,\mathrm{d}u,$$

on trouve d'autre part que $|g(0,0)| \le |g(x,0)| + \int_0^1 |\partial_x g(u,0)|\,\mathrm{d}u$, d'où

$$|g(0,0)| \le \|g(\cdot,0)\|_{W^{1,1}([0,1])}, \quad (2.29)$$

de sorte que l'égalité (2.28) nous donne

$$\|g\|_{L^\infty(\bar{K})} \le \|g(\cdot,0)\|_{W^{1,1}([0,1])} + \|\partial_y g(0,\cdot)\|_{L^1([0,1])} + \|g\|_{W^{2,1}(\bar{K})}. \quad (2.30)$$

Pour majorer le terme $\|g(\cdot,0)\|_{W^{1,1}([0,1])}$, on peut introduire la paramétrisation

$$\Psi\colon (u,s) \in [0,1] \times [0,1/2] \to (u(1-s),s) \in \bar{K}, \quad (2.31)$$

dont le jacobien vérifie $J(\Psi) = 1 - s \in [1/2,1]$. On dérive alors $Ag(u(1-s),s)$ par rapport à s, A étant respectivement égal à I où à ∂_x, ce qui nous donne

$$Ag(u,0) = Ag(u(1-s),s) - \int_0^s \left[\partial_y Ag(u(1-t),t) - u\partial_x Ag(u(1-t),t)\right] \mathrm{d}t, \quad (2.32)$$

et on en déduit

$$|Ag(u,0)| \le |Ag(u(1-s),s)| + \int_0^{1/2} \left(|\partial_y Ag(u(1-t),t)| + |\partial_x Ag(u(1-t),t)|\right) \mathrm{d}t. \quad (2.33)$$

En intégrant cette inégalité par rapport à s et à u, on trouve alors

$$\frac{1}{2}\int_0^1 |Ag(u,0)|\,\mathrm{d}u \le \int_0^1\int_0^{1/2} \big(|Ag(u(1-s),s)| + |\partial_y Ag(u(1-s),s)| + |\partial_x Ag(u(1-s),s)|\big)\,\mathrm{d}s\,\mathrm{d}u$$
$$\le \big(J(\Psi)\big)^{-1}\iint_{\bar K}\big(|Ag(x,y)| + |\partial_y Ag(x,y)| + |\partial_x Ag(x,y)|\big)\,\mathrm{d}x\,\mathrm{d}y. \quad (2.34)$$

En utilisant $\big(J(\Psi)\big)^{-1} \le 2$, on trouve donc

$$\|g(\cdot,0)\|_{W^{1,1}([0,1])} \le C\|g\|_{W^{2,1}(\bar K)}, \quad (2.35)$$

et un argument symétrique nous permettrait d'écrire une estimation semblable pour $\|\partial_y g(0,\cdot)\|_{L^1([0,1])}$. On déduit alors de (2.30) l'injection désirée, *i.e.*

$$\|g\|_{L^\infty(\bar K)} \le C\|g\|_{W^{2,1}(\bar K)}. \quad (2.36)$$

(ii) ensuite, la norme $W^{2,1}(\bar K)$ d'une fonction g qui s'annule aux 3 sommets de $\bar K$ est contrôlée par sa semi-norme :

$$\|g\|_{W^{2,1}(\bar K)} \le C|g|_{W^{2,1}(\bar K)}. \quad (2.37)$$

Pour s'en convaincre, raisonnons par l'absurde en considérant une suite g_n de fonctions de $W^{2,1}(\bar K)$ qui s'annulent aux sommets de $\bar K$ et vérifient

$$\|g_n\|_{W^{1,1}(\bar K)} > n|g_n|_{W^{2,1}(\bar K)}. \quad (2.38)$$

Quitte à normaliser les g_n dans $W^{1,1}$, on peut écrire $\|g_n\|_{W^{1,1}(\bar K)} = 1$ et $|g_n|_{W^{2,1}(\bar K)} \to 0$. On en déduit que la suite g_n est bornée dans $W^{2,1}$, et comme $W^{2,1}$ s'injecte de façon compacte dans $W^{1,1}$, qu'il en existe une sous-suite $g_{n'}$ convergeant vers une fonction $\bar g$ dans $W^{1,1}$. La sous-suite $\partial^2_{xx} g_{n'}$ converge alors vers $\partial^2_{xx}\bar g$ au sens des distributions, et comme $\|\partial^2_{xx} g_{n'}\|_{L^1(\bar K)} \le |g_{n'}|_{W^{2,1}(\bar K)} \to 0$, on trouve que la limite $\partial^2_{xx}\bar g$ est nulle. Il en va de même pour $\partial^2_{xy}\bar g$ et $\partial^2_{yy}\bar g$, de sorte que $\bar g$ est affine. On en déduit que $|\bar g - g_{n'}|_{W^{2,1}(\bar K)} = |g_{n'}|_{W^{2,1}(\bar K)} \to 0$, et donc que la suite $g_{n'}$ converge vers $\bar g$ dans $W^{2,1}$. D'après l'injection (2.36), $g_{n'}$ converge également vers $\bar g$ de façon uniforme, d'où l'on déduit que $\bar g$ s'annule aux 3 sommets de $\bar K$. Comme elle est affine, elle doit être nulle, et l'hypothèse $\|g_{n'}\|_{W^{1,1}} = 1$ devient absurde. Il existe donc une constante C pour laquelle (2.37) est vérifiée.
(iii) en appliquant les inégalités (2.36) et (2.37) à $g = f - P_{\bar K} f$, qui s'annule bien aux sommets de $\bar K$, on voit que l'inégalité (2.27) est bien vérifiée lorsque K est le triangle de référence $\bar K$. Si K est obtenu en faisant tourner $\bar K$ d'un angle θ, on peut se ramener à $\bar K$ en considérant la fonction $\bar f = f \circ \phi$ où $\phi(x,y) = (x\cos\theta - y\sin\theta, y\cos\theta + x\sin\theta)$ est la rotation d'angle θ qui vérifie $K = \phi(\bar K)$. D'une part, la norme L^∞ n'est pas modifiée par la rotation, de sorte que $\|f - P_K f\|_{L^\infty(K)} = \|\bar f - P_{\bar K}\bar f\|_{L^\infty(\bar K)}$. D'autre part, on peut calculer

$$\partial^2_{xx}\bar f = \partial^2_{xx} f(\phi)\cos^2\theta + 2\partial^2_{xy} f(\phi)\cos\theta\sin\theta + \partial^2_{yy} f(\phi)\sin^2\theta \quad (2.39)$$
$$\partial^2_{xy}\bar f = -\partial^2_{xx} f(\phi)\cos\theta\sin\theta + \partial^2_{xy} f(\phi)(\cos^2\theta - \sin^2\theta) + \partial^2_{yy} f(\phi)\cos\theta\sin\theta \quad (2.40)$$
$$\partial^2_{yy}\bar f = \partial^2_{xx} f(\phi)\sin^2\theta - 2\partial^2_{xy} f(\phi)\cos\theta\sin\theta + \partial^2_{yy} f(\phi)\cos^2\theta, \quad (2.41)$$

d'où, en utilisant le fait que ϕ préserve la mesure,

$$|\bar{f}|_{W^{2,1}(\bar{K})} \leq 6 \iint_{\bar{K}} \big(|\partial^2_{xx} f| + |\partial^2_{xy} f| + |\partial^2_{yy} f|\big)(\phi(x,y))\,\mathrm{d}x\,\mathrm{d}y = 6|f|_{W^{2,1}(K)}. \quad (2.42)$$

On en déduit que (2.27) est encore vérifiée (avec une constante absolue) pour les triangles tournés $K = \phi(\bar{K})$. Il reste à considérer le cas d'un triangle K_λ qui se déduit de $K = \phi(\bar{K})$ par un changement d'échelle $K_\lambda = \lambda K$. Pour $f_\lambda \in W^{2,1}(K_\lambda)$, la fonction $f = f_\lambda(\lambda x, \lambda y)$ est définie sur K, et elle vérifie clairement

$$|f|_{W^{2,1}(K)} = \iint_K \lambda^2 \big(|\partial^2_{xx} f| + |\partial^2_{xy} f| + |\partial^2_{yy} f|\big)(\lambda x, \lambda y)\,\mathrm{d}x\,\mathrm{d}y = |f_\lambda|_{W^{2,1}(K_\lambda)}. \quad (2.43)$$

La norme L^∞ étant également préservée par ce changement d'échelle, on en déduit finalement que (2.27) est bien vérifiée pour tous les triangles rectangles isocèles, et ceci avec une constante absolue. □

2.2.3 Adaptation de maillages dyadiques par la semi-norme $W^{2,1}$

Commençons par déduire du lemme 2.12, qui exprime un contrôle de l'erreur d'interpolation sur les triangles, l'estimation suivante qui ne fait intervenir que les cellules dyadiques (carrées).

Proposition 2.13 *Pour tout maillage dyadique $M \in \mathcal{M}(\mathbb{R}^2)$, l'interpolation P_M vérifie*

$$\|f - P_M f\|_{L^\infty(\mathbb{R}^2)} \leq C \sup_{\alpha \in M} |f|_{W^{2,1}(\alpha)} \quad (2.44)$$

avec une constante absolue.

Preuve. D'après la construction de la triangulation $\mathcal{K}(M)$, une cellule $\alpha \in M$ intersecte toujours deux triangles rectangles isocèles de $\mathcal{K}(M)$, qui eux-mêmes n'intersectent jamais que α et éventuellement une de ses voisines dans M. On en déduit que

$$\|f - P_M f\|_{L^\infty(\alpha)} \leq \sum_{\beta \in \mathcal{V}_M(\alpha)} |f|_{W^{2,1}(\beta)}, \quad (2.45)$$

et on sait d'après la proposition 2.10 que les voisines de α sont toujours en nombre borné dans le maillage dyadique M (en vertu de la graduation des niveaux). On en déduit donc bien (2.44). □

Cette estimation nous apprend donc que l'erreur d'interpolation sur un maillage M sera de l'ordre de ε dès que

$$\sup_{\alpha \in M} |f|_{W^{2,1}(\alpha)} \leq \varepsilon, \quad (2.46)$$

ce qu'on peut voir comme une propriété d'*ε-adéquation* (ou d'*ε-adaptation*) entre le maillage M et la fonction f. Pour obtenir un tel maillage, on peut utiliser l'algorithme 2.4, qui construit la plus petite partition dyadique $\tilde{M}$ vérifiant

$$\sup_{\alpha \in \tilde{M}} |f|_{W^{2,1}(\alpha)} \leq \varepsilon,$$

le caractère minimal de $\tilde{M}$ étant traduit par la propriété que les "cellules internes" β de $\Lambda(\tilde{M})\setminus\tilde{M}$, autrement dit les cellules qu'il a fallu découper pour obtenir $\tilde{M}$, vérifient

$$|f|_{W^{2,1}(\beta)} > \varepsilon. \tag{2.47}$$

Observons toutefois que cette partition $\tilde{M}$ n'a *a priori* aucune raison d'être graduée, la semi-norme $W^{2,1}$ de f pouvant par exemple se concentrer sur une petite zone. Ce n'est donc pas un maillage dyadique au sens de la définition 2.6, mais son plus petit raffinement gradué, donné par l'algorithme 2.8, en est un, et ses propriétés sont essentiellement les mêmes que celles de $\tilde{M}$.

Leur complexités sont en effet du même ordre, en vertu de la proposition 2.9 (et de sa version "domaine borné", dans la section 2.1.4). D'autre part, cette nouvelle partition M vérifie bien (2.46). On se permettra d'insister sur le fait que cette propriété est une conséquence de la monotonie

$$\beta \subset \alpha \implies |f|_{W^{2,1}(\beta)} \leq |f|_{W^{2,1}(\alpha)}, \tag{2.48}$$

grâce à laquelle on sait a priori que le raffinement d'une cellule α de "bonne qualité", *i.e.* telle que $|f|_{W^{2,1}(\alpha)} \leq \varepsilon$, produit des cellules β qui sont encore de bonne qualité. Dans le cas présent, la monotonie (2.48) est évidente, mais on rencontrera au chapitre 5 un algorithme de découpages dyadiques construit sur un critère de qualité différent, et pour lequel la monotonie sera bien moins immédiate.

On pourra donc considérer l'algorithme suivant pour construire le plus petit maillage dyadique ε-adapté à une fonction f de $W^{2,1}$ au sens de l'inégalité (2.46).

Algorithme 2.14 (maillage ε-adaptaté au sens de la semi-norme $W^{2,1}$)

- *Poser* $\Lambda_{\ell_0} := \mathbb{D}_{\ell_0}(\mathbb{R}^2)$.
- *Pour* $\ell \geq \ell_0$, *calculer*

  $$\Lambda_{\ell+1} := \Lambda_\ell \cup \{\alpha \in \mathcal{F}(\beta) : \beta \in \Lambda_\ell, \text{ et } |f|_{W^{2,1}(\beta)} > \varepsilon\}$$

  jusqu'à ce que $\Lambda_{L+1} = \Lambda_L$, *et prendre* $\tilde{M} = \partial\Lambda_L$.
- *Définir* $\boldsymbol{A}_\varepsilon(f)$ *comme le plus petit raffinement gradué (algorithme 2.8) de* $\tilde{M}$.

Remarque 2.15 *Lors du raffinement conditionnel de l'arbre Λ_ℓ, dans la deuxième étape de l'algorithme, seules les cellules de niveau ℓ sont susceptibles d'être raffinées.*

Remarque 2.16 *On introduira dans la suite (voir en particulier les sections 3.3.2 et 6.2.1) d'autres indicateurs d'erreur, et on rapellera pour chacun d'entre eux l'algorithme d'ε-adaptation correspondant. Par souci de simplicité, on n'utilisera toutefois qu'une seule notation ($\boldsymbol{A}_\varepsilon$), laissant au contexte le soin de déterminer de quel algorithme il s'agit.*

Les propriétés de cet algorithme sont résumées par le théorème suivant.

Théorème 2.1 (précision et complexité des maillages adaptés) *Soit Ω un pavé borné de $\mathbb{R}^2$ et $f \in W^{2,p}(\mathbb{R}^2)$ avec $p > 1$. Le maillage dyadique $\boldsymbol{A}_\varepsilon(f)$ construit par l'algorithme 2.14 vérifie*

$$\|f - P_{\boldsymbol{A}_\varepsilon(f)} f\|_{L^\infty(\mathbb{R}^2)} \leq C\varepsilon \tag{2.49}$$

avec une constante C absolue, et

$$\mathrm{card}_\Omega(\boldsymbol{A}_\varepsilon(f)) := \#\{\alpha \in \boldsymbol{A}_\varepsilon(f) : \alpha \cap \Omega \neq \emptyset\} \leq C(\Omega, p, f)\varepsilon^{-1} \tag{2.50}$$

avec $C(\Omega, p, f) = C|\Omega|^{1-\frac{1}{p}}|f|_{W^{2,p}(\mathbb{R}^2)}$.

Remarque 2.17 *On donnera au chapitre suivant une version de ce théorème pour des fonctions f plus générales. En particulier, cela nous permettra de montrer que - quitte à modifier légèrement l'algorithme 2.4 -, les estimations de précision* (2.49) *et de complexité* (2.50) *(à un terme logarithmique près) sont encore vérifiées lorsque f appartient à un espace V_M associée à un maillage dyadique M (arbitraire). On observera qu'une telle fonction f, continue et affine par morceaux sur la triangulation conforme $\mathcal{K}(M)$ associée à M, n'est pas dans $W^{2,1}$.*

Preuve. L'estimation d'erreur (2.49) se déduit immédiatement de la discussion ci-dessus. Plus précisément de la proposition 2.13 et du fait que le maillage $\boldsymbol{A}_\varepsilon(f)$ vérifie (2.46). Quant à la complexité de $\boldsymbol{A}_\varepsilon(f)$, on a vu qu'elle était du même ordre de grandeur (*i.e.* égale à une constante multiplicative près) que celle de la partition dyadique $\tilde{M}$ construite par l'algorithme 2.4 (et dont $\boldsymbol{A}_\varepsilon(f)$ est le plus petit raffinement gradué).

On proposera deux méthodes pour évaluer la complexité de cette partition $\tilde{M}$. La première consiste à utiliser la fonction maximale de Hardy-Littlewood déjà évoquée dans la section 1.2.3, en posant cette fois

$$M(D^2 f)(x) := \sup_{A \ni x} |A|^{-1} \int_A |\partial^2_{xx} f(y)| + |\partial^2_{xy} f(y)| + |\partial^2_{yy} f(y)| \, \mathrm{d}y \tag{2.51}$$

et en désignant par $|A|$ la surface du domaine A. Si $f \in W^{2,p}$ avec $p > 1$, on sait que la fonction $M(D^2 f)$ appartient à L^p : on peut alors reprendre l'argument selon lequel une cellule α de niveau $\ell > \ell_0$ dans $\tilde{M}$ vérifie $|f|_{W^{2,1}(\mathcal{P}(\alpha))} > \varepsilon$, et dans la mesure où $|\mathcal{P}(\alpha)| = 4|\alpha|$, la fonction maximale vérifie

$$(4|\alpha|)^{-1}\varepsilon < (4|\alpha|)^{-1}|f|_{W^{2,1}(\mathcal{P}(\alpha))} \leq M(D^2 f)(x), \quad x \in \alpha. \tag{2.52}$$

On en déduit que $\frac{1}{4}\varepsilon \leq \|M(D^2 f)\|_{L^1(\alpha)}$ et sur tout pavé Ω borné,

$$\varepsilon \, \mathrm{card}_\Omega(\tilde{M}) \leq 4\|M(D^2 f)\|_{L^1(\Omega)} \leq C\|M(D^2 f)\|_{L^p(\Omega)} \tag{2.53}$$

avec $C = 4|\Omega|^{1-\frac{1}{p}}$.

On peut également évaluer la complexité de $\tilde{M}$ de façon plus directe. Considérons pour cela les cellules internes β de $\Lambda(\tilde{M}) \setminus \tilde{M}$, pour lesquelles on a (2.47). Si f est dans $W^{2,p}$ avec $p > 1$, l'inégalité de Hölder s'applique et nous donne alors

$$\varepsilon < |f|_{W^{2,1}(\beta)} \leq |\beta|^{1-\frac{1}{p}}|f|_{W^{2,p}(\beta)}, \tag{2.54}$$

en écrivant $|\beta| = 2^{-2\ell(\beta)}$ la surface de β. Désignons par

$$\Lambda_\Omega(\ell) := \{\beta \in \mathbb{D}_\ell(\mathbb{R}^2) : |f|_{W^{2,1}(\beta)} > \varepsilon \text{ et } \beta \cap \Omega \neq \emptyset\}$$

les cellules d'un niveau $\ell \geq \ell_0$ qui sont dans $\Lambda(\tilde{M}) \backslash \tilde{M}$ et intersectent Ω. Clairement, les cellules d'un même $\Lambda_\Omega(\ell)$ sont deux à deux disjointes. On peut donc élever l'inégalité précédente à la puissance p et la sommer sur chaque ensemble $\Lambda_\Omega(\ell)$ avec $\ell > \ell_0$, pour trouver

$$N_\Omega(\ell) \leq 2^{-2\ell(p-1)} |f|^p_{W^{2,p}(\mathbb{R}^2)} \varepsilon^{-p} \tag{2.55}$$

avec $N_\Omega(\ell) := \#\big(\Lambda_\Omega(\ell)\big)$. Comme le nombre total N_Ω de cellules de $\Lambda(\tilde{M}) \setminus \tilde{M}$ qui intersectent Ω vérifie $N_\Omega = \sum_{\ell \geq \ell_0} N_\Omega(\ell)$, on pourrait sommer directement l'inégalité précédente sur les niveaux ℓ et voir que $N_\Omega \leq C(\Omega, p, f)\varepsilon^{-p}$, mais cela ne nous permet pas de retrouver la décroissance (2.50) en ε^{-1} (on en est d'ailleurs d'autant plus loin que p est grand, autrement dit que l'hypothèse est forte !). Pour y parvenir, il faut utiliser le fait - évident - que $N_\Omega(\ell)$ est également inférieur à $C(\Omega)2^{2\ell}$, argument qu'on vient d'utiliser sans le dire pour le niveau initial ℓ_0. On obtient alors

$$N_\Omega \leq C(\Omega, p, f) \sum_{\ell \geq \ell_0} \min\big(2^{-2\ell(p-1)}\varepsilon^{-p}, 2^{2\ell}\big), \tag{2.56}$$

et désignant par $\bar{\ell}$ un niveau auquel $2^{-2\ell(p-1)}\varepsilon^{-p}$ et $2^{2\ell}$ sont équivalents (au sens où le rapport de ces deux quantités est respectivement minoré et majoré par deux constantes absolues), on trouve

$$N_\Omega \leq C(\Omega, p, f) \left(\sum_{\ell=\ell_0}^{\bar{\ell}} 2^{2\ell} + \sum_{\ell=\bar{\ell}}^{\infty} 2^{-2\ell(p-1)}\varepsilon^{-p} \right) \leq C(\Omega, p, f) 2^{2\bar{\ell}} \leq C(\Omega, p, f)\varepsilon^{-1} \tag{2.57}$$

lorsque $\bar{\ell} \geq \ell_0$, et

$$N_\Omega \leq C(\Omega, p, f) \sum_{\ell=\ell_0}^{\infty} 2^{-2\ell(p-1)}\varepsilon^{-p} \leq C(\Omega, p, f)\varepsilon^{-p} \tag{2.58}$$

dans le cas contraire. Mais on a alors $2^{-2\ell_0(p-1)}\varepsilon^{-p} \leq C2^{2\ell_0}$, et on en déduit que $\varepsilon^{-p} \leq C\varepsilon^{-1}$. Autrement dit, on a bien

$$N_\Omega \leq C(\Omega, p, f)\varepsilon^{-1} \tag{2.59}$$

dans tous les cas. On en déduit (2.50) avec les arguments évoqués ci-dessus. □

Chapitre 3

Courbure totale des fonctions définies sur le plan

Dans le chapitre précédent, on a montré que la semi-norme $W^{2,1}$ était un bon indicateur a priori pour les erreurs locales d'interpolation affine par morceaux en deux dimensions. Le théorème 2.1 ne nous satisfait toutefois pas entièrement, car dans le cadre du schéma adaptatif qu'on présentera au chapitre 6, on souhaitera l'appliquer à des solutions numériques qui seront affines par morceaux et n'appartiendront donc pas à l'espace $W^{2,1}(\mathbb{R}^2)$. Dans ce chapitre, on se propose donc d'étendre les estimations d'erreur basées sur la semi-norme $W^{2,1}$ aux fonctions dont les dérivées secondes sont des mesures de Radon, qu'on appellera *fonctions de courbure totale bornée* par analogie avec les fonctions de variation totale bornée.

3.1 Fonctions de courbure totale bornée

3.1.1 Définition de la courbure totale

Que valent donc les dérivées secondes d'une fonction continue et affine par morceaux sur une triangulation arbitraire ? Si l'on pense au fait que le gradient de f est constant sur chaque triangle, on peut imaginer que les dérivées secondes sont des distributions de Dirac portées par les arêtes de la triangulation.

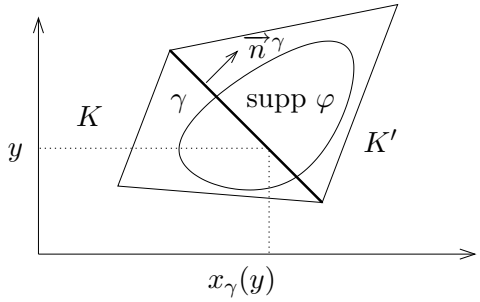


FIG. 3.1 – illustration du calcul (3.1)-(3.2).

Pour le vérifier, considérons le cas où f est affine sur deux triangles K et K' partageant

une arête γ, et désignons par ω l'intérieur de $K \cup K'$. Pour une fonction $\varphi \in \mathcal{C}_c^\infty(\omega)$, on a

$$\langle \partial_{xx}^2 f, \varphi \rangle = -\iint_\omega \partial_x f \partial_x \varphi = -\partial_x f_{|K'} \iint_{K'} \partial_x \varphi - \partial_x f_{|K} \iint_K \partial_x \varphi. \tag{3.1}$$

Si γ est horizontale, ces deux intégrales sont nulles. Sinon, on peut supposer, quitte à intervertir K et K', que le vecteur normal unitaire $\overrightarrow{n}^\gamma = (n_x^\gamma, n_y^\gamma)$ dirigé de K vers K' vérifie $n_x^\gamma > 0$, autrement dit que K' est à droite de K. On calcule alors, en paramétrant l'arête γ par $y \to (x_\gamma(y), y)$,

$$\begin{aligned} &- \partial_x f_{|K'} \iint_{K'} \partial_x \varphi - \partial_x f_{|K} \iint_K \partial_x \varphi \\ &\qquad = \big(\partial_x f_{|K'} - \partial_x f_{|K}\big) \int_y \varphi(x_\gamma(y), y)\, \mathrm{d}y = \big(\partial_x f_{|K'} - \partial_x f_{|K}\big) n_x^\gamma \int_\gamma \varphi, \end{aligned} \tag{3.2}$$

de sorte qu'en notant $[\partial_x f]_\gamma = \partial_x f_{|K'} - \partial_x f_{|K}$ le saut de $\partial_x f$ de part et d'autre de γ, on a dans tous les cas

$$\partial_{xx}^2 f_{|\omega} = [\partial_x f]_\gamma n_x^\gamma \delta_\gamma. \tag{3.3}$$

Par un calcul similaire, on trouverait

$$\partial_{xy}^2 f_{|\omega} = [\partial_x f]_\gamma n_y^\gamma \delta_\gamma = [\partial_y f]_\gamma n_x^\gamma \delta_\gamma, \tag{3.4}$$

et

$$\partial_{yy}^2 f_{|\omega} = [\partial_y f]_\gamma n_y^\gamma \delta_\gamma, \tag{3.5}$$

en observant que l'égalité

$$[\partial_x f]_\gamma n_y^\gamma = [\partial_y f]_\gamma n_x^\gamma \tag{3.6}$$

est une simple conséquence de la continuité de f le long de l'arête γ.

Dans la mesure où ce calcul peut facilement s'étendre à un ouvert contenant plusieurs arêtes, on en déduira que les dérivées secondes de f s'écrivent effectivement comme une somme de distributions de Dirac concentrées sur les arêtes de la triangulation. Ces dérivées ne sont donc pas dans L^1, mais leur *masse totale* étant finie, on pourra considérer l'espace des fonctions dont les dérivées secondes sont des mesures de Radon sur $\mathbb{R}^2$, autrement dit des mesures de Borel μ dont la masse totale

$$|\mu|(\omega) := \sup \Big\{ \sum\nolimits_I |\mu(\omega_i)| : \{\omega_i\}_I \text{ forme une partition de } \omega \Big\}$$

est finie sur tout compact ω. On trouvera dans l'ouvrage [1] de Ambrosio, Fusco et Pallara une présentation détaillée des mesures de Radon et de leurs principales propriétés. Rappelons toutefois que lorsque ω est un ouvert, les mesures de Radon sur ω forment le dual de $\mathcal{C}(\omega)$, la masse totale de μ étant alors caractérisée par la relation

$$|\mu|(\omega) = \sup_{\substack{\varphi \in \mathcal{C}_c^\infty(\omega) \\ \|\varphi\|_{L^\infty} \leq 1}} \langle \mu, \varphi \rangle,$$

en rappelant que le produit scalaire entre μ et $\varphi \in \mathcal{C}_c^\infty(\omega)$ correspond à l'intégrale de φ par rapport à la mesure μ, *i.e.*

$$\langle \mu, \varphi \rangle = \int_\omega \varphi \, \mathrm{d}\mu.$$

Par analogie avec l'espace BV des fonctions de variation totale bornée, dont les dérivées premières sont des mesures de Radon, on parlera de fonctions de courbure totale bornée.

Définition 3.1 (courbure totale) *On dira d'une fonction f de $W^{1,1}_{\text{loc}}(\mathbb{R}^2)$ qu'elle est de courbure totale localement bornée si ses dérivées secondes sont des mesures de Radon. On désigne alors sa courbure totale sur une partie Borélienne $\omega \subset \mathbb{R}^2$ par*

$$|f|_{BC(\omega)} = \mu_f(\omega), \tag{3.7}$$

où μ_f désigne la mesure positive $|\partial^2_{xx} f| + |\partial^2_{xy} f| + |\partial^2_{yy} f|$, et on note $BC(\omega)$ l'espace des fonctions de courbure totale bornée sur ω.

3.1.2 Courbure discrète des fonctions affines par morceaux

Lorsque f est affine par morceaux sur une triangulation arbitraire $\mathcal{K}$, on peut donner une formule explicite pour calculer sa courbure totale sur un domaine régulier ω, ouvert ou fermé. On voit en effet d'après (3.3)-(3.5) que la courbure totale de f sur une arête γ associée à $\mathcal{K}$ vaut

$$|f|_{BC(\gamma)} = \big(|\partial^2_{xx} f| + |\partial^2_{xy} f| + |\partial^2_{yy} f|\big)(\gamma) = |\gamma|_{\mathcal{H}^1}\Big(\big|[\partial_x f]_\gamma\big|\big(|n^\gamma_x| + |n^\gamma_y|\big) + \big|[\partial_y f]_\gamma\, n^\gamma_y\big|\Big), \tag{3.8}$$

où $|\cdot|_{\mathcal{H}^1}$ désigne la mesure de Hausdorff uni-dimensionnelle. Sur le domaine ω, on aura alors

$$|f|_{BC(\omega)} = \sum_\gamma |f|_{BC(\gamma \cap \omega)}$$

en sommant sur toutes les arêtes de la triangulation $\mathcal{K}$. On peut observer que l'expression (3.8) résulte d'une pondération particulière des dérivées secondes qui n'est pas isotrope. En effet, si f désigne par exemple la fonction en "feuille pliée" $f(x,y) = x\chi_{x>0}(x,y)$, sa courbure totale est concentrée sur la droite $\Delta = \{x = 0\}$, où l'on a $[\partial_x f]_\Delta = 1$ et $[\partial_y f]_\Delta = 0$. D'après (3.8), on peut donc calculer sur le segment $\gamma = [(0,0),(0,1)]$, pour lequel $\overrightarrow{n}^\gamma = (1,0)$:

$$|f|_{BC(\gamma)} = 1.$$

Si on fait ensuite tourner f et γ d'un angle $\pi/4$, on trouvera $[\partial_x \tilde{f}]_{\tilde{\Delta}} = [\partial_y \tilde{f}]_{\tilde{\Delta}} = 1/\sqrt{2}$ et $\overrightarrow{n}^{\tilde{\gamma}} = (1/\sqrt{2}, 1/\sqrt{2})$ de sorte que

$$|\tilde{f}|_{BC(\tilde{\gamma})} = \frac{3}{2}.$$

Dans le chapitre 6, on étudiera la façon dont un opérateur de transport peut faire évoluer la régularité d'une fonction affine par morceaux. On utilisera alors la semi-norme suivante, plus géométrique, pour mesurer la courbure des fonctions affines par morceaux.

Définition 3.2 (courbure discrète) *Pour une fonction f continue et affine par morceaux sur une triangulation arbitraire $\mathcal{K}$, on définit sa courbure discrète sur un domaine régulier ω de $\mathbb{R}^2$ par*

$$|f|_{\star(\omega)} := \sum_\gamma |\gamma \cap \omega|_{\mathcal{H}^1}\ \|[Df]_\gamma\|_2, \tag{3.9}$$

où la somme parcourt les arêtes de la triangulation $\mathcal{K}$ et où l'on désigne par $\|[Df]_\gamma\|_2$ la norme euclidienne du vecteur $[Df]_\gamma = ([\partial_x f]_\gamma, [\partial_y f]_\gamma)$.

De façon assez évidente, cette semi-norme est cette fois bien invariante par rotation. Et comme on s'y attend, elle est localement équivalente à la courbure totale (3.7) des fonctions affines par morceaux.

Proposition 3.3 *Pour toute fonction f continue et affine par morceaux sur une triangulation K, et pour toute arête γ de $\mathcal{K}$, on a*

$$|f|_{\star(\gamma)} \leq |f|_{BC(\gamma)} \leq \frac{3}{2}|f|_{\star(\gamma)}. \tag{3.10}$$

Preuve. Commençons par observer que le saut $[Df]$ (comme il n'y a pas d'ambiguïté, on ne précise plus l'arête γ) est essentiellement scalaire : comme f est continue, son gradient le long de γ ne "saute" pas, et on peut déduire de (3.6) :

$$[\partial_x f]n_y = [\partial_y f]n_x \tag{3.11}$$

que

$$[Df] = [\partial_{\overrightarrow{n}} f]\,\overrightarrow{n} = [\partial_x f\, n_x + \partial_y f\, n_y]\,\overrightarrow{n}. \tag{3.12}$$

On a donc

$$|f|_{\star(\gamma)} = |\gamma|_{\mathcal{H}^1}\big|[\partial_x f]n_x + [\partial_y f]n_y\big|, \tag{3.13}$$

de sorte que l'inégalité de gauche de (3.10) se déduit facilement de (3.8). D'un autre côté, on peut supposer quitte à commuter x et y que $n_x \neq 0$. On obtient alors en utilisant successivement (3.11), $|n_x n_y| \leq 1/2$, $n_x^2 + n_y^2 = 1$ et à nouveau (3.11) :

$$\begin{aligned}(|\gamma|_{\mathcal{H}^1})^{-1}|f|_{BC(\gamma)} &= \big|[\partial_x f]\big|\left(|n_x| + |n_y| + n_y^2|n_x|^{-1}\right) = \big|[\partial_x f]\big||n_x|^{-1}(1 + |n_x n_y|) \\ &\leq 3/2\big|[\partial_x f]\big||n_x|^{-1} = 3/2\big|[\partial_x f]\big|\big|n_x + n_y^2 n_x^{-1}\big| = 3/2(|\gamma|_{\mathcal{H}^1})^{-1}|f|_{\star(\gamma)}\end{aligned}$$

ce qui établit l'inégalité de droite dans (3.10). □

3.1.3 Calculs explicites

Lorsque f est affine sur une triangulation *structurée*, on peut facilement exprimer sa courbure totale (ou sa courbure discrète) en fonction de ses valeurs aux sommets des triangles.

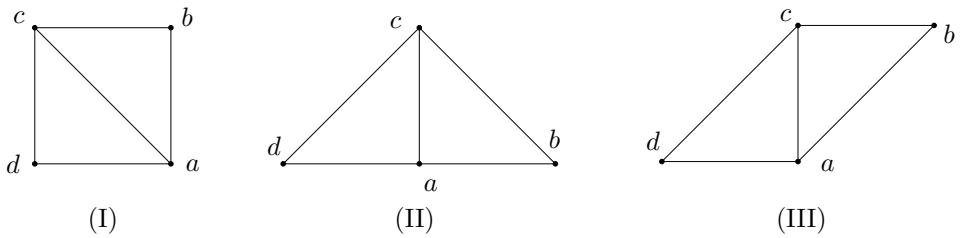


FIG. 3.2 – configurations des arêtes dans une triangulation structurée.

Sur la figure 3.2 ci-dessus, on a représenté le cas où f est affine sur deux triangles rectangles isocèles $[abc]$ et $[cda]$ dont les petits côtés sont parallèles aux axes x et y.

On peut observer qu'en toute généralité, ces triangles sont ceux qu'on obtiendra en découpant les mailles carrées d'un maillage dyadique, mais qu'avec la règle de découpage décrite dans la section 2.2.1, on ne rencontrera que les deux premières configurations. En utilisant les relations (3.8) et (3.13), on trouve

– dans la configuration (I) :

$$|f|_{BC([ac])} = 3|f(b) - f(a) - f(c) + f(d)| \tag{3.14}$$

et

$$|f|_{\star([ac])} = 2|f(b) - f(a) - f(c) + f(d)|, \tag{3.15}$$

– dans la configuration (II) :

$$|f|_{BC([ac])} = |f|_{\star([ac])} = |f(b) - 2f(c) + f(d)|, \tag{3.16}$$

– et dans la configuration (III) :

$$|f|_{BC([ac])} = |f|_{\star([ac])} = |f(b) - f(a) - f(c) + f(d)|. \tag{3.17}$$

Lorsque la géométrie des triangles est arbitraire, il est encore possible d'exprimer la courbure de f (affine par morceaux) à partir de ses valeurs nodales, mais le calcul n'est pas aussi simple. Pour le mener à bien, considérons que les triangles $[abc]$ et $[cda]$ ont leurs sommets orientés dans le sens direct comme sur la figure 3.3, et désignons respectivement par $\overrightarrow{n}$ et $\overrightarrow{n}^{\perp} = (-n_y, n_x)$ le vecteur normal à l'arête $[ac]$ (orienté de d vers b) et son vecteur tourné de $\pi/2$. On calcule alors que sur $[abc]$, le vecteur gradient de f vaut

$$(Df)_{|[abc]} = \frac{1}{2\big|[abc]\big|}\left[f(a)\overrightarrow{bc}^{\perp} + f(b)\overrightarrow{ca}^{\perp} + f(c)\overrightarrow{ab}^{\perp}\right] \tag{3.18}$$

en désignant par $\big|[abc]\big|$ la surface du triangle $[abc]$, et l'on a une formule semblable pour $(Df)_{|[cda]}$. La courbure discrète (3.9) de f sur l'arête $[ac]$, en utilisant les notations de la figure 3.3, s'écrit alors

$$[Df]_{[ac]} = f(a)\left(\frac{\overrightarrow{bc}^{\perp}}{2|[abc]|} + \frac{\overrightarrow{dc}^{\perp}}{2|[cda]|}\right) + f(b)\frac{\overrightarrow{ca}^{\perp}}{2|[abc]|} + f(c)\left(\frac{\overrightarrow{ab}^{\perp}}{2|[abc]|} + \frac{\overrightarrow{ad}^{\perp}}{2|[cda]|}\right) + f(d)\frac{\overrightarrow{ca}^{\perp}}{2|[cda]|}. \tag{3.19}$$

En observant que $2|[abc]| = \|\overrightarrow{ac}\|_2(\overrightarrow{n}, \overrightarrow{ab})$ et $2|[cda]| = -\|\overrightarrow{ac}\|_2(\overrightarrow{n}, \overrightarrow{ad})$ on calcule ensuite

$$\|\overrightarrow{ac}\|_2\left(\frac{\overrightarrow{ab}}{2|[abc]|} + \frac{\overrightarrow{ad}}{2|[cda]|}\right) = \overrightarrow{n}^{\perp}\left(\frac{(\overrightarrow{n}^{\perp}, \overrightarrow{ab})}{(\overrightarrow{n}, \overrightarrow{ab})} - \frac{(\overrightarrow{n}^{\perp}, \overrightarrow{ad})}{(\overrightarrow{n}, \overrightarrow{ad})}\right), \tag{3.20}$$

de sorte que l'on a $\|\overrightarrow{ac}\|_2[Df]_{[ac]} = \overrightarrow{n}S$ avec

$$S = [f(b)-f(a)]\frac{\|\overrightarrow{ac}\|_2}{(\overrightarrow{n}, \overrightarrow{ab})} - [f(d)-f(a)]\frac{\|\overrightarrow{ac}\|_2}{(\overrightarrow{n}, \overrightarrow{ad})} - [f(c)-f(a)]\left(\frac{(\overrightarrow{n}^{\perp}, \overrightarrow{ab})}{(\overrightarrow{n}, \overrightarrow{ab})} - \frac{(\overrightarrow{n}^{\perp}, \overrightarrow{ad})}{(\overrightarrow{n}, \overrightarrow{ad})}\right). \tag{3.21}$$

On retrouve ainsi une propriété visible dans (3.12) selon laquelle le saut du vecteur gradient de part et d'autre d'une arête est toujours perpendiculaire à cette arête. En particulier, on trouve que la courbure discrète (3.9) concentrée sur l'arête $[ac]$ vaut $|f|_{\star([ac])} = |S|$, où S est donné par (3.21) et se calcule explicitement à partir des données géométriques des triangles.

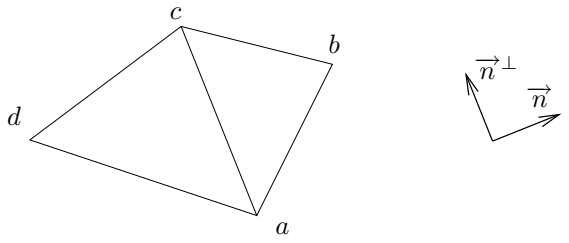


FIG. 3.3 – notations correspondant aux calculs (3.18)-(3.21)

3.2 Propriétés des fonctions de $BC(\mathbb{R}^2)$

3.2.1 Continuité des fonctions de $BC(\mathbb{R}^2)$

Pour donner un sens à l'interpolation $\mathcal{P}^1$ des fonctions de $BC(\mathbb{R}^2)$, et pour voir si cet espace fournit un bon cadre pour l'approximation dans $L^\infty(\mathbb{R}^2)$, une question naturelle que l'on peut se poser est de savoir si les fonctions de $BC(\mathbb{R}^2)$ sont continues. C'est bien sûr le cas des fonctions de $W^{2,1}(\mathbb{R}^2)$, ce qu'on peut voir soit comme une conséquence de la densité de $\mathcal{C}_c^\infty$ dans $W^{2,1}$ et de l'injection continue (2.36) (si ψ_n est une suite de $\mathcal{C}_c^\infty$ qui converge vers f dans $W^{2,1}$, on a $\|f-\psi_n\|_{L^\infty} \leq C\|f-\psi_n\|_{W^{2,1}}$ et la continuité de f se déduit de celle des ψ_n), soit plus simplement en faisant tendre x et y vers 0 dans l'égalité (2.28).

Pour se donner une idée de la réponse, on peut penser à plusieurs arguments. Les mesures de Radon étant "presque" des fonctions de L^1, on peut par exemple voir les fonctions de $BC(\mathbb{R}^2)$ comme des fonctions qui seraient "presque" dans $W^{2,1}(\mathbb{R}^2)$, or les fonctions de $W^{2,1}(\mathbb{R}^2)$ sont continues. Malheureusement, il existe un exemple célèbre où ce "raisonnement" ne fonctionne pas : en dimension 1, les fonctions de $W^{1,1}(\mathbb{R})$ sont continues, mais ce n'est pas le cas des fonctions de $BV(\mathbb{R})$. La fonction de Heaviside $\chi_{\mathbb{R}_+}$, par exemple, a pour dérivée la masse de Dirac δ_0 qui est une mesure de Radon.

Pour être plus précis, on rappellera que la continuité des fonctions de $W^{1,1}(\mathbb{R})$ découle de la relation

$$f(x) = f(0) + \int_0^x f'(u)\,\mathrm{d}u, \tag{3.22}$$

à partir de laquelle on peut faire tendre x vers 0. Si l'on considère maintenant les fonctions de $BV(\mathbb{R})$, on peut continuer à écrire une relation de type (3.22), à condition de préciser le sens du terme "$\int_0^x f'$". Si μ désigne la mesure de Radon f', on pourra par exemple vérifier que

$$f(x) = f(x') + \int_{]x',x]} \mathrm{d}\mu = f(x') + \mu(]x',x]) \quad \text{lorsque} \quad x' \leq x, \tag{3.23}$$

en observant que l'exclusion ou l'inclusion des bornes x et x' n'est pas figée, mais correspond au choix des valeurs prises par f aux points de discontinuité qui sont toujours isolés pour une fonction de $BV(\mathbb{R})$ (l'écriture (3.23) correspondant au cas où f est toujours continue à droite). Pour établir (3.23) de façon rigoureuse, on peut

dériver la fonction

$$g(x) = \begin{cases} \mu(]0,x]) \text{ lorsque } x \geq 0 \\ -\mu(]x,0]) \text{ lorsque } x < 0 \end{cases}$$

en calculant pour tout $\varphi \in \mathcal{C}_c^\infty(\mathbb{R})$:

$$\langle g', \varphi \rangle = -\langle g, \varphi' \rangle = -\int_{\mathbb{R}_+} \left(\int_{]0,x]} \mathrm{d}\mu \right) \varphi'(x)\,\mathrm{d}x + \int_{\mathbb{R}_-} \left(\int_{]x,0]} \mathrm{d}\mu \right) \varphi'(x)\,\mathrm{d}x.$$

D'après le théorème de Fubini, on a

$$\int_{\mathbb{R}_+} \left(\int_{]0,x]} \mathrm{d}\mu \right) \varphi'(x)\,\mathrm{d}x = \int_{\mathbb{R}} \int_{\mathbb{R}} \chi_{\{0<u\leq x\}} \varphi'(x)\,\mathrm{d}x\,\mathrm{d}\mu(u),$$

et on voit que lorsque u est négatif, l'intégrale $\int_{\mathbb{R}} \chi_{\{0<u\leq x\}} \varphi'(x)\,\mathrm{d}x$ est nulle. Dans le cas contraire, elle vaut

$$\int_{\mathbb{R}} \chi_{\{0<u\leq x\}} \varphi'(x)\,\mathrm{d}x = \int_u^\infty \varphi'(x)\,\mathrm{d}x = -\varphi(u),$$

le même calcul nous donnant $\int_{\mathbb{R}} \chi_{\{x<u\leq 0\}} \varphi'(x)\,\mathrm{d}x = \chi_{u\leq 0}\varphi(u)$ lorsque $x \leq 0$. On en déduit que

$$\langle g', \varphi \rangle = \int_{\mathbb{R}_+} \varphi(u)\,\mathrm{d}\mu(u) + \int_{\mathbb{R}_-} \varphi(u)\,\mathrm{d}\mu(u) = \int_{\mathbb{R}} \varphi\,\mathrm{d}\mu = \langle \mu, \varphi \rangle,$$

ce qui signifie que μ est bien la dérivée de g et justifie l'égalité (3.23). L'espace $BV(\mathbb{R})$ s'injecte donc encore dans $L^\infty(\mathbb{R})$ de façon continue. L'existence de fonctions discontinues à variations bornées nous indique alors que les fonctions de $\mathcal{C}_c^\infty(\mathbb{R})$ n'ont aucune chance d'être denses dans $BV(\mathbb{R})$.

Voici un autre "raisonnement" naïf (et faux) auquel on pourrait penser : comme une fonction dont les variations sont bornées est forcément bornée, et que les dérivées premières de $f \in BC(\mathbb{R}^2)$ sont dans $BV(\mathbb{R}^2)$, on peut imaginer que f est lipschitzienne, donc continue. Ce raisonnement, qui est tout à fait correct en dimension 1, ne l'est plus en dimension 2 car la variation totale $|g|_{BV(\mathbb{R}^2)} = |\partial_x g|(\mathbb{R}^2) + |\partial_y g|(\mathbb{R}^2)$ d'une fonction g s'y écrit, d'après la formule géométrique de la co-aire, comme

$$|g|_{BV(\mathbb{R}^2)} = \int_{\lambda\in\mathbb{R}} p(\Omega_g(\lambda))\,\mathrm{d}\lambda \tag{3.24}$$

où $\Omega_g(\lambda) := \{x \in \mathbb{R}^2 : \lambda \leq g(x)\}$ désigne le domaine de $\mathbb{R}^2$ sur lequel g est supérieure à λ (le bord $\partial\Omega_g(\lambda)$ étant ce qu'on appelle la ligne de niveau associée à la valeur λ), et $p(\Omega_g(\lambda)) = |\chi_{\Omega_g(\lambda)}|_{BV(\mathbb{R}^2)}$ son périmètre essentiel. A la vue de (3.24), il est assez clair qu'une fonction de $\mathbb{R}^2$ peut être de variations bornées sans pour autant être elle-même bornée : ainsi la fonction en "tour de Babel" $g = \sum_{n\geq 1} \chi_{B(0,1/n^2)}$, qui est égale à $n \geq 1$ sur l'anneau $\frac{1}{(n+1)^2} \leq \|(x,y)\|_2 \leq \frac{1}{n^2}$ et tend très clairement vers $+\infty$ en 0, vérifie

$$|g|_{BV(\mathbb{R}^2)} = \sum_{n\geq 1} \int_{n-1}^n p(\Omega_g(\lambda))\,\mathrm{d}\lambda = \sum_{n\geq 1} \frac{2\pi}{n^2} < \infty.$$

Ce que nous montre la discussion précédente, c'est que la continuité des fonctions de courbure totale (localement) bornée n'a rien d'évident. Elle est toutefois avérée, et on a la proposition suivante (dont la preuve nous a été très aimablement suggérée par Luc Tartar).

Proposition 3.4 *Les fonctions de l'espace $BC(\mathbb{R}^2)$ sont continues.*

Preuve. Pour démontrer ce résultat, on introduit les espaces de Lorentz $L^{p,q}$, définis pour $1 \leq p \leq \infty$ et $1 \leq q \leq \infty$ comme l'ensemble des fonctions f pour lesquelles la quantité

$$\|f\|_{L^{p,q}(\mathbb{R}^2)} := \begin{cases} \left(\int_{\lambda>0} \left(\left|\Omega_{|f|}(\lambda)\right|^{\frac{1}{p}} \lambda \right)^q \frac{\mathrm{d}\lambda}{\lambda} \right)^{\frac{1}{q}} & \text{lorsque } q < \infty \\ \sup_{\lambda>0} \left|\Omega_{|f|}(\lambda)\right|^{\frac{1}{p}} \lambda & \text{lorsque } q = \infty \end{cases}$$

est finie, où $\left|\Omega_{|f|}(\lambda)\right|$ désigne la mesure de Lebesgue de $\Omega_{|f|}(\lambda) = \{x \in \mathbb{R}^2 : \lambda \leq |f(x)|\}$. On appelle *fonction de répartition* de f la fonction $\lambda \to \left|\Omega_{|f|}(\lambda)\right|$, et *$L^p$-faible* l'espace $L^{p,\infty}$ qui contient les fonctions pour lesquelles $\left|\Omega_{|f|}(\lambda)\right|$ décroît comme C/λ^p. On observera que ces espaces sont croissants par rapport à q (l'espace $L^{p,1}$ étant plus petit que l'espace $L^{p,\infty}$), et que $L^{p,q}$ coïncide avec L^p lorsque $q = p$, d'après

$$\begin{aligned} \|f\|^p_{L^p(\mathbb{R}^2)} = \int_{\mathbb{R}^2} |f(x)|^p \,\mathrm{d}x = \int_{\mathbb{R}^2} \int_0^{|f(x)|^p} \mathrm{d}\nu \,\mathrm{d}x = \int_{\mathbb{R}_+} \int_{\mathbb{R}^2} \chi_{\{0 \leq \nu \leq |f(x)|^p\}} \,\mathrm{d}x \,\mathrm{d}\nu \\ = \int_{\mathbb{R}_+} \left|\Omega_{|f|}(\nu^{\frac{1}{p}})\right| \mathrm{d}\nu = p \int_{\mathbb{R}_+} \left|\Omega_{|f|}(\lambda)\right| \lambda^{p-1} \,\mathrm{d}\lambda = p \|f\|^p_{L^{p,p}(\mathbb{R}^2)}. \end{aligned}$$

On peut faire apparaître ces espaces de façon assez naturelle en utilisant la formule (3.24) avec l'inégalité iso-périmétrique appliquée à $\Omega_g(\lambda)$, selon laquelle on a

$$\left|\Omega_g(\lambda)\right|^{\frac{1}{2}} \leq C p(\Omega_g(\lambda)) \tag{3.25}$$

pour une constante absolue. Si g est une fonction positive de $BV(\mathbb{R}^2)$, on en déduit que

$$\|g\|_{L^{2,1}(\mathbb{R}^2)} = \int_{\lambda>0} \left|\Omega_g(\lambda)\right|^{\frac{1}{2}} \mathrm{d}\lambda \leq C \int_{\lambda>0} p(\Omega_g(\lambda)) \,\mathrm{d}\lambda \leq C |g|_{BV(\mathbb{R}^2)}.$$

Dans le cas général, on peut décomposer $g = g^+ - g^-$ en ses parties positive $g^+ = \max(g, 0)$ et négative $g^- = \max(-g, 0)$, qui vérifient

$$\Omega_{g^+}(\lambda) = \begin{cases} \Omega_g(\lambda) & \text{pour } \lambda > 0 \\ \mathbb{R}^2 & \text{pour } \lambda \leq 0 \end{cases} \quad \text{et} \quad \Omega_{g^-}(\lambda) = \begin{cases} \mathbb{R}^2 \setminus \Omega_g(-\lambda) & \text{pour } \lambda > 0 \\ \mathbb{R}^2 & \text{pour } \lambda \leq 0, \end{cases}$$

de sorte que

$$p(\Omega_{g^+}(\lambda)) = \chi_{\{\lambda>0\}} p(\Omega_g(\lambda)) \quad \text{et} \quad p(\Omega_{g^-}(\lambda)) = \chi_{\{\lambda>0\}} p(\Omega_g(-\lambda)).$$

En appliquant (3.25) respectivement à g^+ et g^-, on obtient alors

$$\|g^\pm\|_{L^{2,1}(\mathbb{R}^2)} = \int_{\lambda>0} \left|\Omega_{g^\pm}(\lambda)\right|^{\frac{1}{2}} \mathrm{d}\lambda \leq C \int_{\lambda>0} p(\Omega_{g^\pm}(\lambda)) \,\mathrm{d}\lambda \leq C \int_{\mathbb{R}_\pm} p(\Omega_g(\lambda)) \,\mathrm{d}\lambda,$$

d'où

$$\|g\|_{L^{2,1}(\mathbb{R}^2)} \leq \|g^+\|_{L^{2,1}} + \|g^-\|_{L^{2,1}} \leq C \int_{\mathbb{R}} p(\Omega_g(\lambda))\,\mathrm{d}\lambda \leq C|g|_{BV(\mathbb{R}^2)}. \tag{3.26}$$

Ceci nous montre qu'en dimension 2, l'espace $BV(\mathbb{R}^2)$ s'injecte continuement dans $L^{2,1}$. La deuxième propriété des espaces de Lorentz qu'on va utiliser est la dualité entre $L^{2,1}$ et $L^{2,\infty}$, au sens où l'on a

$$\iint |gh(x,y)|\,\mathrm{d}x\,\mathrm{d}y \leq C\|g\|_{L^{2,1}(\mathbb{R}^2)}\|h\|_{L^{2,\infty}(\mathbb{R}^2)} \tag{3.27}$$

avec une constante absolue, inégalité qu'on peut établir en utilisant les réarrangements radiaux décroissants de g et h (qu'on supposera positives). Rappelons que le réarrangement décroissant g^* de g est l'unique fonction positive décroissante définie sur $\mathbb{R}_+$ qui vérifie $|\Omega_g(\lambda)| = |\Omega_{g^*}(\lambda)|$, autrement dit telle que

$$g^*(t) = \lambda \quad \text{où } \lambda \text{ vérifie} \quad |\Omega_g(\lambda)| = t > 0. \tag{3.28}$$

Il est alors facile de voir que $\iint_{\mathbb{R}^2} g(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}_+} g^*(t)\,\mathrm{d}t$, et une propriété intéressante des réarrangements (voir [30]) est l'inégalité suivante

$$\iint_{\mathbb{R}^2} gh(x,y)\,\mathrm{d}x\,\mathrm{d}y \leq \int_{\mathbb{R}_+} g^*h^*(t)\,\mathrm{d}t \tag{3.29}$$

(on peut se convaincre de cette inégalité en considérant que pour $a \leq b$ et $c \leq d$, on a toujours $ad + bc \leq ac + bd$). Dans la mesure où (3.28) entraîne que $s \to g^*(s^2)$ est la réciproque (éventuellement discontinue) de la fonction décroissante $\lambda \to |\Omega_g(\lambda)|^{1/2}$, on a

$$\int_{\mathbb{R}_+} g^*(t)\frac{\mathrm{d}t}{\sqrt{t}} = C\int_{\mathbb{R}_+} g^*(s^2)\,\mathrm{d}s = C\int_{\lambda>0} |\Omega_g(\lambda)|^{\frac{1}{2}}\,\mathrm{d}\lambda = C\|g\|_{L^{2,1}(\mathbb{R}^2)}.$$

D'autre part, la fonction h étant dans $L^{2,\infty}(\mathbb{R}^2)$, on a $|\Omega_h(\lambda)| \leq \|h\|_{L^{2,\infty}(\mathbb{R}^2)}/\lambda$, d'où l'on déduit en utilisant (3.28) que $h^*(t) \leq \|h\|_{L^{2,\infty}(\mathbb{R}^2)}/\sqrt{t}$. On a donc

$$\int_{\mathbb{R}_+} g^*h^*(t)\,\mathrm{d}t \leq \|h\|_{L^{2,\infty}(\mathbb{R}^2)}\int_{\mathbb{R}_+} g^*(t)\frac{\mathrm{d}t}{\sqrt{t}} \leq C\|h\|_{L^{2,\infty}(\mathbb{R}^2)}\|g\|_{L^{2,1}(\mathbb{R}^2)},$$

ce qui entraîne bien (3.27) en utilisant (3.29).

Si maintenant f est une fonction de $BC(\mathbb{R}^2)$, ses dérivées $\partial_x f$ et $\partial_y f$ qui sont dans $BV(\mathbb{R}^2)$ sont également dans $L^{2,1}(\mathbb{R}^2)$ d'après (3.26), et il se trouve que les fonctions de $\mathcal{C}_c^\infty(\mathbb{R}^2)$ sont denses dans cet espace. On peut donc approcher $\partial_x f$ et $\partial_y f$ dans $L^{2,1}(\mathbb{R}^2)$ par deux suites ψ_n^x et ψ_n^y de fonctions de $\mathcal{C}_c^\infty(\mathbb{R}^2)$, et poser

$$\psi_n(x,y) = f(0,0) + \int_0^x \psi_n^x(u,0)\,\mathrm{d}u + \int_0^y \psi_n^y(x,v)\,\mathrm{d}v$$

pour tout $(x,y) \in \mathbb{R}^2$, de façon que les dérivées de ψ_n soient bien ψ_n^x et ψ_n^y. On observe alors que la solution élémentaire du Laplacien $\Delta\phi = \delta_0$ dans $\mathbb{R}^2$ est $\phi = \log r$ (en

coordonnées polaires) dont les dérivées $\partial_x \phi = \frac{x}{r^2} = \frac{\cos\theta}{r}$ et $\partial_y \phi = \frac{\sin\theta}{r}$ appartiennent à $L^{2,\infty}(\mathbb{R}^2)$, puisque

$$|\Omega_{|\partial_x \phi|}(\lambda)| = |\{(x,y) : r \leq |\cos\theta|\lambda^{-1}\}| = \int_0^{2\pi} \int_0^{\frac{|\cos\theta|}{\lambda}} r \,\mathrm{d}r \,\mathrm{d}\theta \leq \frac{\pi}{\lambda^2},$$

et de même pour $|\Omega_{|\partial_y \phi|}(\lambda)|$. On peut alors utiliser l'inégalité (3.27) pour écrire

$$\begin{aligned} |\psi_m(x,y) - \psi_n(x,y)| &= |\langle \psi_m - \psi_n, \delta_{(x,y)} \rangle| = |\langle \psi_m - \psi_n, \Delta \phi_{(x,y)} \rangle| \\ &= |\langle \nabla(\psi_m - \psi_n), \nabla \phi_{(x,y)} \rangle| \leq C \|\nabla(\psi_m - \psi_n)\|_{L^{2,1}}. \end{aligned}$$

La suite ψ_n étant de Cauchy dans L^∞, elle converge donc vers une fonction ψ qui est continue. Pour voir enfin que cette fonction coïncide avec f, on peut utiliser le fait que $L^{2,1}(\mathbb{R}^2)$ s'injecte dans $L^2(\mathbb{R}^2)$ pour dire que la convergence de $\nabla \psi_n$ vers ∇f dans $L^{2,1}(\mathbb{R}^2)$ est aussi une convergence au sens des distributions. $\nabla \psi_n$ tendant d'autre part vers $\nabla \psi$, on en déduit que la distribution $\psi - f$ est constante, et comme elle s'annule en $(0,0)$, elle est nulle en tout point. □

3.2.2 Stabilité des interpolations $\mathcal{P}^1$

D'après la proposition 3.4, l'espace $BC(\mathbb{R}^2)$ constitue un cadre agréable pour étudier les propriétés d'approximation des éléments finis $\mathcal{P}^1$ adaptatifs. En particulier, l'interpolation P_M associée à un maillage dyadique est bien définie pour une fonction f de courbure totale bornée, et la fonction affine par morceaux $P_M f$ appartient toujours à $BC(\mathbb{R}^2)$, ou du moins à l'espace $BC_{\mathrm{loc}}(\mathbb{R}^2)$. Dans l'analyse des schémas basés sur ces discrétisations adaptatives, on aura besoin d'une propriété un peu plus précise, à savoir que les opérateurs d'interpolation soient stables vis à vis de la courbure totale, ce qu'exprime la proposition suivante.

Proposition 3.5 *L'interpolation $\mathcal{P}^1$ associée à un maillage dyadique M vérifie*

$$\sup_{\alpha \in M} |P_M f|_{BC(\alpha)} \leq C \sup_{\alpha \in M} |f|_{BC(\alpha)}$$

pour toute fonction $f \in BC(\mathbb{R}^2)$, avec une constante C absolue.

Cette inégalité peut se voir comme une conséquence du lemme 3.6 ci-dessous. Pour établir ce lemme, on utilisera l'argument de régularisation suivant : si l'on convole une fonction $f \in BC(\mathbb{R}^2)$ avec un noyau de régularisation

$$\varphi_\varepsilon(x,y) := \frac{1}{\varepsilon^2} \varphi\left(\frac{x}{\varepsilon}, \frac{y}{\varepsilon}\right)$$

construit à partir d'une fonction φ positive, continue, de support inclus dans la boule $B(0,1)$ et de masse

$$\iint_{B(0,1)} \varphi = \iint_{B(0,\varepsilon)} \varphi_\varepsilon = 1,$$

on obtient une fonction

$$f_\varepsilon := f * \varphi_\varepsilon \tag{3.30}$$

dont les dérivées secondes sont continues. Plus précisément, en désignant par μ_{xx} la mesure de Radon $\partial^2_{xx} f$, on peut calculer (en s'appuyant notamment sur la description faite dans [1] du produit de convolution entre une fonction continue et une mesure de Radon) que la dérivée $\partial^2_{xx} f_\varepsilon$ vaut

$$\partial^2_{xx} f_\varepsilon(x,y) = (\mu_{xx} * \varphi_\varepsilon)(x,y) = \iint \varphi_\varepsilon(x-x', y-y')\,\mathrm{d}\mu_{xx}(x',y'), \tag{3.31}$$

pour tout $(x,y) \in \mathbb{R}^2$, avec des expressions identiques pour les dérivées $\partial^2_{xy} f_\varepsilon$ et $\partial^2_{yy} f_\varepsilon$. En ce qui concerne le comportement de cette régularisation lorsque ε tend vers 0, on voit sans peine que la continuité de f entraîne une convergence uniforme de f_ε vers f.

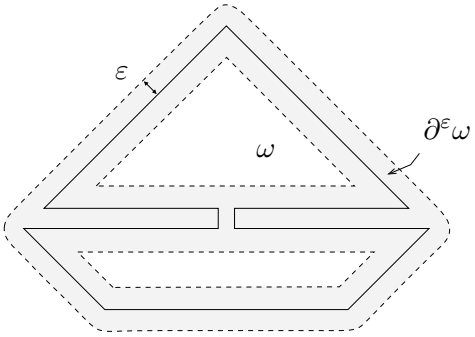


FIG. 3.4 – un domaine polygonal ω (délimité par la courbe en traits pleins) et sa frontière d'épaisseur 2ε, notée $\partial^\varepsilon\omega$ (délimitée par les courbes en pointillés).

D'autre part, si l'on désigne par

$$\partial^\varepsilon\omega := \{(x,y) \in \mathbb{R}^2 : B\big((x,y),\varepsilon\big) \cap \partial\omega \neq \emptyset\} = \omega + \varepsilon \setminus \omega - \varepsilon$$

la *frontière d'épaisseur* 2ε d'un domaine $\omega \subset \mathbb{R}^2$ telle qu'on peut la représenter sur la figure 3.4, on aura

$$\left| \iint_\omega \partial^2_{xx} f_\varepsilon(x,y)\,\mathrm{d}x\,\mathrm{d}y - \mu_{xx}(\omega \setminus \partial^\varepsilon\omega) \right| \leq |\mu_{xx}|(\partial^\varepsilon\omega) \tag{3.32}$$

et

$$\iint_\omega |\partial^2_{xx} f_\varepsilon(x,y)|\,\mathrm{d}x\,\mathrm{d}y \leq |\mu_{xx}|(\omega \cup \partial^\varepsilon\omega), \tag{3.33}$$

avec des relations semblables pour $\partial^2_{xy} f_\varepsilon$ et $\partial^2_{yy} f_\varepsilon$. Pour s'en convaincre, utilisons (3.31) pour calculer

$$\begin{aligned}\iint_\omega \partial^2_{xx} f_\varepsilon(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \iint_\omega \left[\iint_{\mathbb{R}^2} \varphi_\varepsilon(x-x', y-y')\,\mathrm{d}\mu_{xx}(x',y') \right] \mathrm{d}x\,\mathrm{d}y \\ &= \iint_{\mathbb{R}^2} A(x',y')\,\mathrm{d}\mu_{xx}(x',y')\end{aligned}$$

avec $A(x',y') := \iint_\omega \varphi_\varepsilon(x-x', y-y')\,\mathrm{d}x\,\mathrm{d}y$. On décompose alors le plan en trois parties : si (x',y') appartient à $\omega \setminus \partial^\varepsilon\omega$, la boule $B\big((x',y'),\varepsilon\big)$ est incluse dans ω et on en déduit que $A(x',y')$ est égal à 1. On a donc

$$\iint_{\omega \setminus \partial^\varepsilon\omega} A(x',y')\,\mathrm{d}\mu_{xx}(x',y') = \mu_{xx}(\omega \setminus \partial^\varepsilon\omega).$$

Si par contre (x', y') appartient à $\omega^{\mathsf{c}} \setminus \partial^{\varepsilon}\omega$, la boule $B\big((x', y'), \varepsilon\big)$ n'intersecte pas ω : $A(x', y')$ est alors nul et clairement,

$$\iint_{\omega^{\mathsf{c}}\setminus\partial^{\varepsilon}\omega} A(x', y')\,\mathrm{d}\mu_{xx}(x', y') = 0.$$

Enfin dans le cas où $(x', y') \in \partial^{\varepsilon}\omega$, on peut toujours écrire que $|A(x', y')| \leq 1$, ce qui entraîne

$$\left|\iint_{\partial^{\varepsilon}\omega} A(x', y')\,\mathrm{d}\mu_{xx}(x', y')\right| \leq |\mu_{xx}|(\partial^{\varepsilon}\omega),$$

et l'inégalité (3.32) se déduit facilement des relations précédentes. L'inégalité (3.33) s'obtient de la même façon, en observant que l'égalité (3.31) nous permet également d'écrire

$$|\partial^2_{xx} f_{\varepsilon}(x, y)| \leq \iint \varphi_{\varepsilon}(x - x', y - y')\,\mathrm{d}|\mu_{xx}|(x', y').$$

Venons-en donc à notre résultat de stabilité.

Lemme 3.6 *Soit K et K' une paire de triangles conformes* fermés *partageant une arête $\gamma = K \cap K'$. En désignant par ω l'*intérieur *de $K \cup K'$, l'interpolation affine sur la paire $\{K, K'\}$ vérifie*

$$|P_{\{K,K'\}} f|_{BC(\omega)} = |P_{\{K,K'\}} f|_{BC(\gamma)} \leq C|f|_{BC(\omega)} \tag{3.34}$$

pour toute fonction $f \in BC(\omega)$, avec une constante C qui dépend uniquement des angles entre les différentes arêtes des triangles.

Remarque 3.7 *En particulier, ce lemme nous apprend que l'interpolation affine par morceaux sur la triangulation conforme $\mathcal{K}(M)$ associée à un maillage dyadique $M \in \mathcal{M}(\mathbb{R}^2)$ vérifie*

$$|P_M f|_{BC(\mathbb{R}^2)} \leq C|f|_{BC(\mathbb{R}^2)}$$

avec une constante C absolue.

Preuve. Le fait que la courbure totale de $P_{\{K,K'\}} f$ sur ω soit concentrée sur l'arête γ est une conséquence directe des calculs (3.1)-(3.5). D'autre part, on sait d'après la proposition 3.3 que

$$|P_{\{K,K'\}} f|_{BC(\gamma)} \leq \frac{3}{2}|P_{\{K,K'\}} f|_{\star(\gamma)},$$

où le terme de droite est ce qu'on a appelé la courbure discrète (3.9) d'une fonction affine par morceaux sur une de ses arêtes, et qu'on a calculé explicitement en (3.18)-(3.21) avec $K = [abc]$ et $K' = [cda]$. Comme la courbure discrète est invariante par rotation, on peut reprendre les notations de ce calcul (qui sont celles de la figure 3.3) en considérant que l'arête $\gamma = [ac]$ est verticale, ce qui nous donnera $\overrightarrow{n} = \overrightarrow{e_x}$ et $\overrightarrow{n}^{\perp} = \overrightarrow{e_y}$ comme indiqué sur la figure 3.5. La courbure discrète de la fonction interpolée vaut alors

$$|P_{\{K,K'\}} f|_{\star([ac])} = |S| \tag{3.35}$$

où S est le scalaire donné par (3.21), ici égal à

$$S = \|\overrightarrow{ac}\|_2 \left(\frac{[f(b) - f(a)]}{(\overrightarrow{e_x}, \overrightarrow{ab})} - \frac{[f(d) - f(a)]}{(\overrightarrow{e_x}, \overrightarrow{ad})}\right) - [f(c) - f(a)] \left(\frac{(\overrightarrow{e_y}, \overrightarrow{ab})}{(\overrightarrow{e_x}, \overrightarrow{ab})} - \frac{(\overrightarrow{e_y}, \overrightarrow{ad})}{(\overrightarrow{e_x}, \overrightarrow{ad})}\right). \tag{3.36}$$

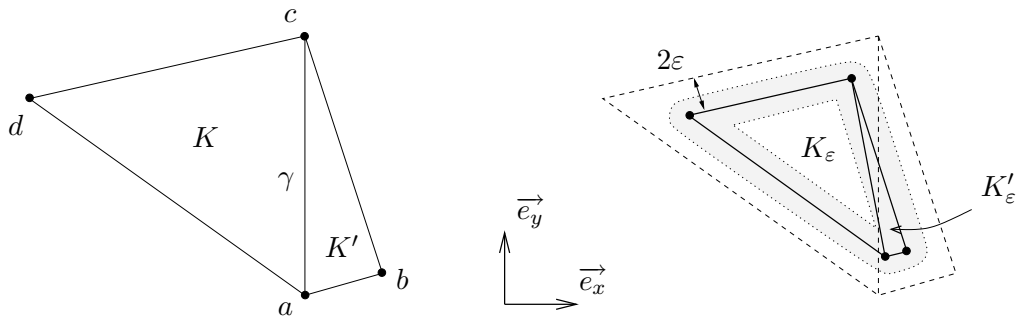


FIG. 3.5 – illustration de la preuve du lemme 3.6.

Pour exprimer cette quantité comme une intégrale double des dérivées secondes de f, commençons par considérer que f est de classe $\mathcal{C}^2$ (on appliquera ensuite notre calcul à une régularisation f_ε de f lorsque celle-ci n'est que dans BC). On peut alors écrire les différences $[f(b)-f(a)]$, $[f(c)-f(a)]$ et $[f(d)-f(a)]$ comme les intégrales des gradients correspondants le long des segments $[ab]$, $[ac]$ et $[ad]$, ce qui nous donnera

$$S = \|\overrightarrow{ac}\|_2 \left[\int_0^1 \partial_x f(a+s\overrightarrow{ab})\,\mathrm{d}s + \int_0^1 \partial_y f(a+s\overrightarrow{ab})\,\mathrm{d}s \frac{(\overrightarrow{e_y},\overrightarrow{ab})}{(\overrightarrow{e_x},\overrightarrow{ab})} - \int_0^1 \partial_x f(a+s\overrightarrow{ad})\,\mathrm{d}s \right.$$
$$\left. - \int_0^1 \partial_y f(a+s\overrightarrow{ad})\,\mathrm{d}s \frac{(\overrightarrow{e_y},\overrightarrow{ad})}{(\overrightarrow{e_x},\overrightarrow{ad})} + \int_0^1 \partial_y f(a+s\overrightarrow{ac})\,\mathrm{d}s \left(\frac{(\overrightarrow{e_y},\overrightarrow{ad})}{(\overrightarrow{e_x},\overrightarrow{ad})} - \frac{(\overrightarrow{e_y},\overrightarrow{ab})}{(\overrightarrow{e_x},\overrightarrow{ab})} \right) \right].$$

En utilisant le vecteur gradient $Df = (\partial_x f, \partial_y f)$, on peut regrouper ces termes en

$$S = \|\overrightarrow{ac}\|_2 \int_0^1 \left[\left([Df(a+s\overrightarrow{ab}) - Df(a+s\overrightarrow{ac})], \frac{\overrightarrow{ab}}{(\overrightarrow{e_x},\overrightarrow{ab})} \right) - \right.$$
$$\left. \left([Df(a+s\overrightarrow{ad}) - Df(a+s\overrightarrow{ac})], \frac{\overrightarrow{ad}}{(\overrightarrow{e_x},\overrightarrow{ad})} \right) \right] \mathrm{d}s,$$

et en désignant par $\mathcal{H}f = \left(\begin{smallmatrix} \partial^2_{xx} f & \partial^2_{xy} f \\ \partial^2_{xy} f & \partial^2_{yy} f \end{smallmatrix}\right)$ la matrice Hessienne de f, on peut à nouveau écrire les différences du vecteur gradient comme des intégrales le long des segments appropriés. On trouve alors

$$S = \|\overrightarrow{ac}\|_2 \iint_{[0,1]^2} \left[\left([\mathcal{H}f(a+s\overrightarrow{ac}+st\overrightarrow{cb})\overrightarrow{cb}], \frac{\overrightarrow{ab}}{(\overrightarrow{e_x},\overrightarrow{ab})} \right) \right.$$
$$\left. - \left([\mathcal{H}f(a+s\overrightarrow{ac}+st\overrightarrow{cd})\overrightarrow{cd}], \frac{\overrightarrow{ad}}{(\overrightarrow{e_x},\overrightarrow{ad})} \right) \right] s\,\mathrm{d}t\,\mathrm{d}s.$$

On calcule ensuite que les bijections $\Psi_{[abc]}\colon (s,t) \in [0,1]^2 \to a + s\overrightarrow{ac} + st\overrightarrow{cb} \in [abc]$ et $\Psi_{[cda]}\colon (s,t) \in [0,1]^2 \to a + s\overrightarrow{ac} + st\overrightarrow{cd} \in [cda]$ ont respectivement pour Jacobiens

$$|J(\Psi_{[abc]})| = |\det(\overrightarrow{ac} + t\overrightarrow{cb}, s\overrightarrow{cb})| = s|\det(\overrightarrow{ac},\overrightarrow{cb})| = s\|\overrightarrow{ac}\|_2(\overrightarrow{e_x},\overrightarrow{cb})$$

et $|J(\Psi_{[cda]})| = -s\|\overrightarrow{ac}\|_2(\overrightarrow{e_x}, \overrightarrow{cd})$, ce qui nous donne l'expression désirée pour le scalaire S, à savoir

$$S = \iint_{[abc]} \left(\frac{\overrightarrow{ab}}{(\overrightarrow{e_x}, \overrightarrow{ab})}\right)^t \mathcal{H}f(x,y) \left(\frac{\overrightarrow{cb}}{(\overrightarrow{e_x}, \overrightarrow{cb})}\right) \mathrm{d}x\,\mathrm{d}y + \iint_{[cda]} \left(\frac{\overrightarrow{ad}}{(\overrightarrow{e_x}, \overrightarrow{ad})}\right)^t \mathcal{H}f(x,y) \left(\frac{\overrightarrow{cd}}{(\overrightarrow{e_x}, \overrightarrow{cd})}\right) \mathrm{d}x\,\mathrm{d}y. \quad (3.37)$$

On peut en effet observer que dans la base $(\overrightarrow{e_x}, \overrightarrow{e_y})$, les coordonnées des vecteurs $\overrightarrow{ab}/(\overrightarrow{e_x}, \overrightarrow{ab})$, $\overrightarrow{cb}/(\overrightarrow{e_x}, \overrightarrow{cb})$, *etc.* ne dépendent que des angles entre les différentes arêtes de K et K'. On déduit alors de l'égalité ci-dessus qu'il existe une constante C dépendant uniquement de ces angles, telle que

$$|S| \leq C|f|_{W^{2,1}(\omega)},$$

en rappelant que ω désigne ici l'intérieur de $K \cup K'$. Pour pouvoir écrire une inégalité semblable lorsque f n'est plus de classe $\mathcal{C}^2$ mais appartient à $BC(\omega)$, utilisons sa fonction régularisée f_ε définie par (3.30), et considérons le domaine

$$\omega_\varepsilon := \omega \setminus \partial^{2\varepsilon}\omega$$

obtenu en s'écartant à distance 2ε du bord de ω. Pour des valeurs suffisamment petites de ε, ω_ε se compose de deux triangles d'intérieurs disjoints qu'on désigne par K_ε et K'_ε, et qu'on a représentés dans la partie droite de la figure 3.5 (la zone grisée correspondant à la frontière $\partial^\varepsilon(\omega_\varepsilon)$). Si l'on interpole cette fonction f_ε sur les triangles K_ε et K'_ε, le calcul ci-dessus s'applique et l'on a

$$|S_\varepsilon| \leq C|f_\varepsilon|_{W^{2,1}(\omega_\varepsilon)},$$

en désignant par S_ε la quantité correspondant à (3.36), obtenue en remplaçant f par f_ε et les différents sommets de K et K' par ceux de K_ε et K'_ε. Compte tenu de la continuité (uniforme) de f, de la convergence uniforme de f_ε vers f et de la convergence des sommets de K_ε et K'_ε vers ceux de K et K', il n'est pas difficile de voir que S_ε tendra vers S lorsque $\varepsilon \to 0$. D'un autre côté, le domaine $\omega_\varepsilon \cup \partial^\varepsilon(\omega_\varepsilon)$ est toujours fortement inclus dans l'ouvert ω. L'inégalité (3.33) *appliquée à* ω_ε nous donne alors

$$|f_\varepsilon|_{W^{2,1}(\omega_\varepsilon)} \leq |f|_{BC\left(\omega_\varepsilon \cup \partial^\varepsilon(\omega_\varepsilon)\right)} \leq |f|_{BC(\omega)}$$

car $\omega_\varepsilon \cup \partial^\varepsilon(\omega_\varepsilon)$ est toujours inclus dans ω, d'où l'on déduit que

$$|S| \leq C|f|_{BC(\omega)}$$

par passage à la limite, pour une constante C dépendant uniquement des angles entre les différentes arêtes. Comme les courbures totales et discrètes de l'interpolée $P_{\{K,K\}}f$ vérifient

$$\frac{2}{3}|P_{\{K,K\}}f|_{BC(\gamma)} \leq |P_{\{K,K\}}f|_{\star(\gamma)} = |S|$$

d'après (3.10) et (3.35), on en déduit finalement l'inégalité (3.34). □

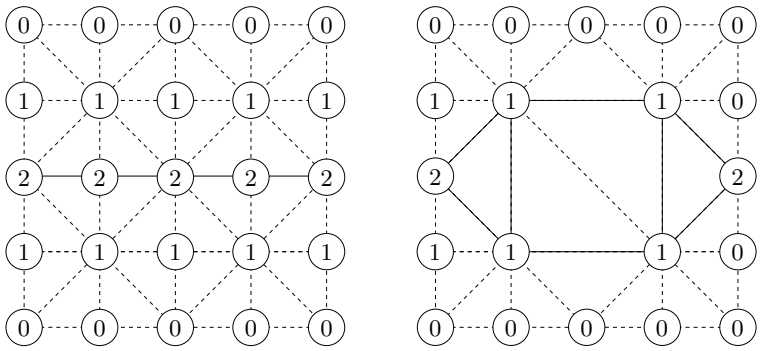


FIG. 3.6 – pour la fonction affine par morceaux dont les valeurs nodales sont indiquées à gauche, l'interpolation sur la triangulation de droite fait augmenter la courbure totale (les arêtes γ en traits pleins correspondent à une courbure totale $|\cdot|_{BC(\gamma)}$ égale à 2, les autres à une courbure nulle).

3.2.3 Un résultat de décroissance vérifié par les interpolations $\mathcal{P}^1$

Dans les chapitre 5 et 6, où l'on propose un schéma semi-lagrangien adaptatif de la forme (1.13) on s'intéressera à la façon dont la courbure totale des solutions numériques évolue au cours des itérations, notamment dans le but d'estimer a priori le nombre de mailles adaptatives générées par notre schéma. Dans cette perspective, il est essentiel de contrôler a priori l'accroissement éventuel de la courbure par les projections P_n, qui seront des interpolations affines par morceaux dans notre schéma. Le lemme 3.6 ci-dessus (qui exprime la stabilité de ces interpolations) *ne nous permet malheureusement pas* de majorer la courbure totale d'une solution f^N de façon indépendante du pas de temps Δt lorsque le nombre d'itérations N correspond à un temps de simulation fixé $T = N\Delta t$. Pour s'en convaincre, on peut considérer le cas où le transport est une simple translation qui ne modifie pas la courbure. Les solutions du schéma (1.13) vérifient alors

$$|f^N|_{BC} \leq C|f^{N-1}|_{BC} \leq \cdots \leq C^N|f^0|_{BC}$$

et ne sont a priori pas majorées par une constante indépendante de Δt, à moins que la constante C ne soit plus petite que 1. Il serait donc intéressant d'établir une propriété de *décroissance* de la courbure par des interpolations affines par morceaux, autrement dit une inégalité de type

$$|P_{\mathcal{K}} f|_{BC} \leq |f|_{BC}.$$

Cette décroissance, très facile à démontrer en dimension 1, n'est a priori pas vérifiée en dimension 2. On peut le voir sur l'exemple représenté figure 3.6, où la fontion f affine par morceaux sur la triangulation de gauche est interpolée sur une triangulation déraffinée $\mathcal{K}$ à droite. Les valeurs nodales étant indiquées aux sommets des triangles, on a calculé à partir des formules (3.14)-(3.16) les courbures totales de f et $P_{\mathcal{K}} f$ correspondant aux arêtes de leurs triangulations respectives. On vérifie donc que sur cet exemple,

$$|P_{\mathcal{K}} f|_{BC} > |f|_{BC},$$

et on trouverait la même chose avec les courbures discrètes $|\cdot|_\star$.

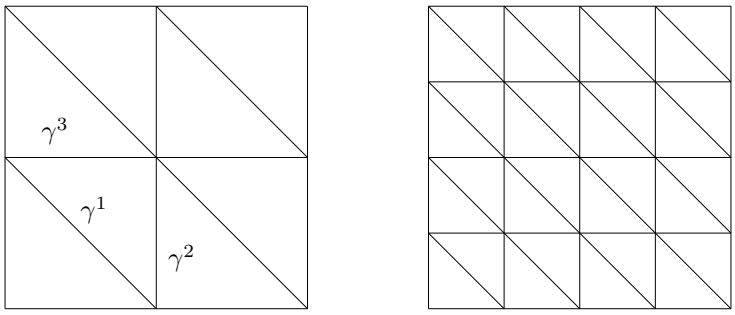


FIG. 3.7 – sur ces triangulations, l'interpolation $\mathcal{P}^1$ fait décroître la courbure (3.39).

Toutefois, si l'on accepte de se restreindre à des triangulations particulières $\mathcal{K}$, on peut proposer une semi-norme équivalente à la courbure totale $|\cdot|_{BC(\omega)}$ qui est diminuée par les interpolations $P_\mathcal{K}$. Plus précisément, on a établi le résultat suivant (dont un résumé est présenté dans [14]).

Théorème 3.1 *Pour $\ell \in \mathbb{N}$, désignons par $\mathcal{K}_\ell$ les triangulations uniformes (représentées sur la figure 3.7) obtenues en découpant tous les carrés dyadiques de niveau ℓ par la même diagonale. Les interpolations affines par morceaux P_ℓ associées à ces triangulations vérifient*

$$|P_\ell f|_{\star\star(\mathbb{R}^2)} \leq |f|_{\star\star(\mathbb{R}^2)} \tag{3.38}$$

pour toute fonction $f \in BC(\mathbb{R}^2)$, où la courbure $|\cdot|_{\star\star}$ est définie par

$$|f|_{\star\star(\mathbb{R}^2)} := \big(|\partial^2_{xx} f - \partial^2_{xy} f| + |\partial^2_{xy} f - \partial^2_{yy} f| + 2|\partial^2_{xy} f|\big)(\mathbb{R}^2). \tag{3.39}$$

et coïncide avec la courbure discrète $|\cdot|_\star$ de f lorsque celle-ci est précisément continue et affine par morceaux sur une triangulation $\mathcal{K}_\ell$. Lorsque f appartient à $W^{2,1}(\mathbb{R}^2)$, on a de plus la caractérisation suivante :

$$|f|_{\star\star(\mathbb{R}^2)} = \lim_{\ell \to \infty} |P_\ell f|_{\star(\mathbb{R}^2)}. \tag{3.40}$$

Remarque 3.8 *Avec les arguments de la preuve ci-dessous, il est également possible de construire pour* toute *triangulation conforme $\mathcal{K}$ une semi-norme $|\cdot|_{\star\star(\mathcal{K})}$ dépendante de $\mathcal{K}$ qui sera diminuée par les interpolations affines par morceaux P_ℓ associées aux raffinements* uniformes *de $\mathcal{K}$.*

Preuve. On désignera par Γ^1_ℓ, Γ^2_ℓ et Γ^3_ℓ les arêtes de $\mathcal{K}_\ell$ correspondant aux trois orientations possibles, telles qu'on les a représentées sur la figure 3.7.

Pour vérifier que la courbure $|f|_{\star\star(\mathbb{R}^2)}$ coïncide bien avec la courbure discrète $|f|_{\star(\mathbb{R}^2)}$ lorsque f est continue et affine par morceaux sur une triangulation $\mathcal{K}_\ell$, on peut reprendre le calcul (3.1)-(3.5) mené au début de ce chapitre. Il nous apprend en effet que les dérivées secondes d'une telle fonction f s'écrivent comme des masses de Dirac concentrées sur les arêtes de $\mathcal{K}_\ell$, et les expressions (3.3), (3.4) et (3.5) nous permettent

d'en calculer les amplitudes. Dans la mesure où le vecteur normal $\overrightarrow{n}^\gamma$ vaut respectivement $(\overrightarrow{e_x}+\overrightarrow{e_y})/\sqrt{2}$, $\overrightarrow{e_x}$ ou $\overrightarrow{e_y}$ suivant que l'arête γ est de type 1, 2 ou 3, on obtient de cette façon :

– lorsque γ est une arête de type 1,

$$|f|_{\star\star(\gamma)} = 2|\partial^2_{xy} f|(\gamma) = |\gamma|_{\mathcal{H}^1}\sqrt{2}\big|[\partial_x f]_\gamma\big| = |\gamma|_{\mathcal{H}^1}\|[Df]_\gamma\|_2 = |f|_{\star(\gamma)},$$

la troisième égalité provenant du fait que $\big|[\partial_x f]_\gamma\big| = \big|[\partial_x f]_\gamma\big| = \|[Df]_\gamma\|_2/\sqrt{2}$ se déduit de(3.6),

– lorsque γ est une arête de type 2,

$$|f|_{\star\star(\gamma)} = |\partial^2_{xx} f - \partial^2_{xy} f|(\gamma) = |\gamma|_{\mathcal{H}^1}\big|[\partial_x f]_\gamma\big| = |\gamma|_{\mathcal{H}^1}\|[Df]_\gamma\|_2 = |f|_{\star(\gamma)}$$

car (3.6) entraîne cette fois $[\partial_y f]_\gamma = 0$,

– et lorsque γ est une arête de type 3,

$$|f|_{\star\star(\gamma)} = |\partial^2_{xy} f - \partial^2_{yy} f|(\gamma) = |\gamma|_{\mathcal{H}^1}\big|[\partial_y f]_\gamma\big| = |\gamma|_{\mathcal{H}^1}\|[Df]_\gamma\|_2 = |f|_{\star(\gamma)}$$

dans la mesure où (3.6) entraîne alors $[\partial_x f]_\gamma = 0$.

En résumé, on trouve bien

$$|f|_{\star\star(\mathbb{R}^2)} = |f|_{\star(\mathbb{R}^2)} \tag{3.41}$$

lorsque f est continue et affine par morceaux sur une triangulation de type $\mathcal{K}_\ell$.

Si f est à présent une fonction de $W^{2,1}$, on sait que la courbure discrète de son interpolée sur $\mathcal{K}_\ell$ peut s'écrire

– soit comme une intégrale double des dérivées secondes de f, en reprenant les calculs effectués dans la preuve du lemme 3.6,

– soit comme une formule de quadrature sur les valeurs nodales de f, en reprenant les expressions de la section 3.1.3.

Désignons donc par $\omega(\gamma)$ le quadrilatère formé par la réunion des triangles situés de part et d'autre d'une arête γ, et calculons d'après (3.35) et (3.37) que

$$|P_\ell f|_{\star(\gamma)} = \begin{cases} \left|\iint_{\omega(\gamma)} 2\partial^2_{xy} f\right| & \text{si } \gamma \in \Gamma^1_\ell, \\ \left|\iint_{\omega(\gamma)} \partial^2_{xx} f - \partial^2_{xy} f\right| & \text{si } \gamma \in \Gamma^2_\ell, \\ \left|\iint_{\omega(\gamma)} \partial^2_{xy} f - \partial^2_{yy} f\right| & \text{si } \gamma \in \Gamma^3_\ell. \end{cases} \tag{3.42}$$

Chaque famille de quadrilatères $\{\omega(\gamma) : \gamma \in \Gamma^1_\ell\}$ réalisant une partition de $\mathbb{R}^2$ d'autant plus fine que ℓ est grand, on peut alors facilement imaginer que les sommes $\sum_{\gamma\in\Gamma^i_\ell} |P_\ell f|_{\star(\gamma)}$ vont converger vers les quantités $\iint_{\mathbb{R}^2} 2|\partial^2_{xy} f|$, $\iint_{\mathbb{R}^2} |\partial^2_{xy} f - \partial^2_{yy} f|$ et $\iint_{\mathbb{R}^2} |\partial^2_{xx} f - \partial^2_{xy} f|$ suivant que i vaut 1, 2 ou 3. Et l'on en déduira la propriété (3.40). Pour l'établir en toute rigueur, on pourra considérer les fonctions

$$\varphi^i_\ell := \sum_{\gamma\in\Gamma^i_\ell} c_i(\gamma)\chi_{\omega(\gamma)}$$

définies par

$$c_i(\gamma) = \begin{cases} |\omega(\gamma)|^{-1} \left| \iint_{\omega(\gamma)} 2\partial^2_{xy} f \right| & \text{pour } i = 1, \\ |\omega(\gamma)|^{-1} \left| \iint_{\omega(\gamma)} \partial^2_{xx} f - \partial^2_{xy} f \right| & \text{pour } i = 2, \\ |\omega(\gamma)|^{-1} \left| \iint_{\omega(\gamma)} \partial^2_{xy} f - \partial^2_{yy} f \right| & \text{pour } i = 3, \end{cases}$$

de sorte que $|P_\ell f|_{\star(\mathbb{R}^2)} = \|\varphi^1_\ell\|_{L^1} + \|\varphi^2_\ell\|_{L^1} + \|\varphi^3_\ell\|_{L^1}$. On observera alors que lorsque ℓ tend vers l'infini, les suites φ^i_ℓ convergent respectivement vers $2|\partial^2_{xy} f|$, $|\partial^2_{xx} f - \partial^2_{xy} f|$ et $|\partial^2_{xy} f - \partial^2_{yy} f|$ presque partout au titre de moyennes locales de ces fonctions de L^1. Et comme il est clair que l'on a

$$\|\varphi^1_\ell\|_{L^1(\mathbb{R}^2)} \leq \|2\partial^2_{xy} f\|_{L^1(\mathbb{R}^2)}$$

avec des expressions semblables pour φ^2_ℓ et φ^3_ℓ, le lemme de Fatou nous assure que

$$\|2\partial^2_{xy} f\|_{L^1(\mathbb{R}^2)} \leq \liminf_{\ell\to\infty} \|\varphi^1_\ell\|_{L^1(\mathbb{R}^2)} \leq \limsup_{\ell\to\infty} \|\varphi^1_\ell\|_{L^1(\mathbb{R}^2)} \leq \|2\partial^2_{xy} f\|_{L^1(\mathbb{R}^2)}$$

avec des expressions semblables pour φ^2_ℓ et φ^3_ℓ. On en déduit donc la limite annoncée

$$\begin{aligned} |f|_{\star\star(\mathbb{R}^2)} &= \|2\partial^2_{xy} f\|_{L^1(\mathbb{R}^2)} + \|\partial^2_{xx} f - \partial^2_{xy}\|_{L^1(\mathbb{R}^2)} + \|\partial^2_{xy} f - \partial^2_{yy} f\|_{L^1(\mathbb{R}^2)} \\ &= \lim_{\ell\to\infty} \|\varphi^1_\ell\|_{L^1(\mathbb{R}^2)} + \|\varphi^2_\ell\|_{L^1(\mathbb{R}^2)} + \|\varphi^3_\ell\|_{L^1(\mathbb{R}^2)} = \lim_{\ell\to\infty} |P_\ell f|_{\star(\mathbb{R}^2)} \end{aligned} \quad (3.43)$$

lorsque $f \in W^{2,1}(\mathbb{R}^2)$, soit (3.40).

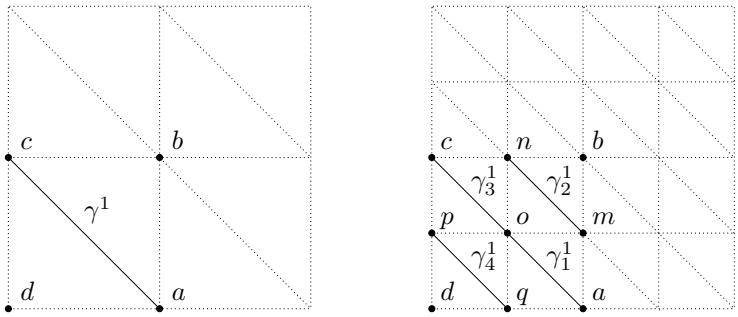


FIG. 3.8 – propagation des courbures discrètes au cours du raffinement.

Pour établir la propriété (3.38), on va alors montrer que les courbures discrètes $|P_\ell f|_{\star(\mathbb{R}^2)}$ forment une suite *croissante* de ℓ, et l'on va utiliser pour ceci l'expression (3.15) de la courbure discrète en fonction des valeurs nodales. Avec les notations de la figure 3.8, (3.15) s'écrit

$$|P_\ell f|_{\star(\gamma^1)} = 2|f(b) - f(a) - f(c) + f(d)|, \quad (3.44)$$

et une inégalité triangulaire nous donne

$$\begin{aligned}|P_\ell f|_{\star(\gamma^1)} &= 2|f(b) - f(a) - f(c) + f(d)| \\ &\leq 2\big(|f(m) - f(a) - f(o) + f(q)| + |f(b) - f(m) - f(n) + f(o)| \\ &\qquad + |f(n) - f(o) - f(c) + f(p)| + |f(o) - f(q) - f(p) + f(d)|\big) \\ &= \sum_{j=1}^{4} |P_{\ell+1} f|_{\star(\gamma_j^1)}.\end{aligned}$$

On obtient donc ainsi

$$\sum_{\gamma \in \Gamma_\ell^1} |P_\ell f|_{\star(\gamma)} \leq \sum_{\gamma \in \Gamma_{\ell+1}^1} |P_{\ell+1} f|_{\star(\gamma)},$$

et dans la mesure où ce calcul s'applique également aux arêtes de Γ_ℓ^2 et Γ_ℓ^3, on en déduit la monotonie anoncée

$$|P_\ell f|_{\star(\mathbb{R}^2)} \leq |P_{\ell+1} f|_{\star(\mathbb{R}^2)} \quad \text{pour tout } \ell \in \mathbb{N}. \tag{3.45}$$

En utilisant successivement l'égalité (3.41) (appliquée à $P_\ell f$), la limite (3.43) et la monotonie (3.45) ci-dessus, on obtient alors

$$|P_\ell f|_{\star\star(\mathbb{R}^2)} = |P_\ell f|_{\star(\mathbb{R}^2)} \leq |P_{\ell+1} f|_{\star(\mathbb{R}^2)} \leq \cdots \leq |f|_{\star\star(\mathbb{R}^2)}$$

lorsque f est une fonction de $W^{2,1}(\mathbb{R}^2)$.

Dans le cas général où f appartient à $BC(\mathbb{R}^2)$, il est toujours possible d'écrire des relations semblables à (3.42) par un argument de régularisation, et la monotonie (3.45) est toujours établie. Il n'est pas clair, en revanche, que l'on puisse encore exprimer la courbure $|f|_{\star\star}$ comme limite (3.43) des courbures discrètes $|P_\ell f|_\star$. Cela n'est toutefois pas nécessaire : d'après (3.33), on peut en effet observer que la régularisation f_ε de f définie par (3.30) vérifie $|f_\varepsilon|_{\star\star(\mathbb{R}^2)} \leq |f|_{\star\star(\mathbb{R}^2)}$, de sorte que l'on a

$$|P_\ell f_\varepsilon|_{\star\star(\mathbb{R}^2)} \leq |f_\varepsilon|_{\star\star(\mathbb{R}^2)} \leq |f|_{\star\star(\mathbb{R}^2)}.$$

Il ne nous reste alors plus qu'à observer que f étant continue, f_ε converge uniformément vers f lorsque ε tend vers 0. Les courbures discrètes $|P_\ell f_\varepsilon|_{\star(\mathbb{R}^2)}$ s'écrivant d'après (3.44) comme une fonction continue des valeurs nodales de f_ε, on en déduit que

$$|P_\ell f_\varepsilon|_{\star\star(\mathbb{R}^2)} = |P_\ell f_\varepsilon|_{\star(\mathbb{R}^2)} \to |P_\ell f|_{\star(\mathbb{R}^2)} = |P_\ell f|_{\star\star(\mathbb{R}^2)}$$

(en utilisant à nouveau que les courbures $|\cdot|_{\star(\mathbb{R}^2)}$ et $|\cdot|_{\star\star(\mathbb{R}^2)}$ coïncident pour des fonctions affines par morceaux), et ceci établit finalement la propriété (3.38) lorsque f est une fonction de $BC(\mathbb{R}^2)$. □

3.3 Application au contrôle des éléments finis adaptatifs

3.3.1 Estimation a priori des erreurs d'interpolation

En ce qui concerne la convergence du schéma qu'on présentera au chapitre 5, la propriété essentielle de la courbure totale est qu'elle permet de contrôler les erreurs d'interpolation de la même façon que la semi-norme $W^{2,1}$ (voir le lemme 2.12).

Lemme 3.9 *Si K est un triangle rectangle isocèle, qu'on peut considérer ouvert, l'interpolation affine par morceaux P_K vérifie*

$$\|f - P_K f\|_{L^\infty(K)} \leq C|f|_{BC(K)} \tag{3.46}$$

pour toute fonction f de courbure totale bornée sur K, avec une constante C absolue.

Remarque 3.10 *Dans la mesure où f et $P_K f$ sont continues, l'erreur d'interpolation dans L^∞ ne "voit" pas le bord du triangle. La précision sur le fait que K peut être choisi sans son bord signifie donc simplement que la courbure de f concentrée sur le bord du triangle ne contribue pas à l'erreur d'interpolation. C'est très clair lorsque f est une fonction continue et affine sur le triangle, auquel cas $P_K f = f$ sur K, indépendamment du comportement de f hors de K.*

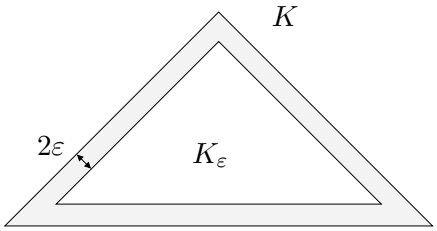


FIG. 3.9 – le triangle K_ε obtenu en s'éloignant des bords de K.

Preuve. On peut appliquer le lemme (2.27) à une régularisation de f. Plus précisément, on peut considérer la fonction f_ε de classe $\mathcal{C}^2$ donnée par (3.30), et observer son interpolation affine $P_{K_\varepsilon} f_\varepsilon$ sur le triangle

$$K_\varepsilon := K \setminus \partial^{2\varepsilon} K = \big\{(x,y) \in K : B\big((x,y), 2\varepsilon\big) \cap \partial K = \emptyset\big\}$$

obtenu en s'éloignant de 2ε du bord de K comme indiqué sur la figure 3.9. Le lemme (2.27) s'applique, et nous apprend que

$$\|(I - P_{K_\varepsilon}) f_\varepsilon\|_{L^\infty(K_\varepsilon)} \leq |f_\varepsilon|_{W^{2,1}(K_\varepsilon)}. \tag{3.47}$$

En ce qui concerne les semi-normes $|f_\varepsilon|_{W^{2,1}(K_\varepsilon)}$, l'inégalité (3.33) s'applique et nous permet d'écrire

$$|f_\varepsilon|_{W^{2,1}(K_\varepsilon)} \leq |f|_{CB(K)} \tag{3.48}$$

en utilisant le fait que l'ensemble $K_\varepsilon \cup \partial^\varepsilon K_\varepsilon$ est toujours fortement inclus dans le triangle K. D'autre part, on peut vérifier que la convergence uniforme de f_ε vers f, la stabilité des interpolations affines et l'uniforme continuité de f sur K entraînent

$$\|(I - P_{K_\varepsilon}) f_\varepsilon\|_{L^\infty(K_\varepsilon)} \to \|(I - P_K) f\|_{L^\infty(K)} \quad \text{lorsque } \varepsilon \to 0 \tag{3.49}$$

Pour nous en convaincre, écrivons

$$\begin{aligned} &\big|\|(I - P_{K_\varepsilon}) f_\varepsilon\|_{L^\infty(K_\varepsilon)} - \|(I - P_K) f\|_{L^\infty(K_\varepsilon)}\big| \\ &\qquad \leq \|f - f_\varepsilon\|_{L^\infty(K_\varepsilon)} + \|(P_K - P_{K_\varepsilon}) f\|_{L^\infty(K_\varepsilon)} + \|P_{K_\varepsilon}(f - f_\varepsilon)\|_{L^\infty(K_\varepsilon)}, \end{aligned}$$

et observons que : (i) la continuité de $(I - P_K)f$ sur K entraîne

$$\|(I - P_K)f\|_{L^\infty(K_\varepsilon)} \to \|(I - P_K)f\|_{L^\infty(K)},$$

(ii) la stabilité des interpolations et la convergence uniforme de f_ε vers f entraîne

$$\|f - f_\varepsilon\|_{L^\infty(K_\varepsilon)} + \|P_{K_\varepsilon}(f - f_\varepsilon)\|_{L^\infty(K_\varepsilon)} \leq 2\|f - f_\varepsilon\|_{L^\infty(K)} \to 0 \ \text{ lorsque } \varepsilon \to 0$$

(iii) l'uniforme continuité de f et la convergence des sommets du triangle K_ε vers ceux de K entraîne

$$\|(P_K - P_{K_\varepsilon})f\|_{L^\infty(K_\varepsilon)} \to 0 \ \text{ lorsque } \varepsilon \to 0.$$

L'inégalité (3.46) se déduit alors de (3.47), (3.48) et (3.49). □

3.3.2 Adaptation de maillages dyadiques par la courbure totale

On peut à présent reformuler les résultats de la section 2.2.3 dans le cadre de l'espace $BC(\mathbb{R}^2)$. Dans la suite, on n'appliquera ces résultats "généralisés" que pour des fonctions appartenant aux espaces V_M introduits dans la section 2.2, autrement dit continues et affines par morceaux sur des triangulations adaptatives structurées. On sera alors attentifs au fait que le résultat de complexité énoncé par le théorème 3.2 *s'applique également à ces fonctions.*

Commençons donc par déduire du lemme 3.9, qui exprime un contrôle de l'erreur d'interpolation sur les triangles, l'estimation suivante qui ne fait intervenir que les cellules dyadiques (carrées).

Proposition 3.11 *Pour tout maillage dyadique $M \in \mathcal{M}(\mathbb{R}^2)$, l'interpolation P_M vérifie*

$$\|f - P_M f\|_{L^\infty(\mathbb{R}^2)} \leq C \sup_{\alpha \in M} |f|_{BC(\alpha)} \tag{3.50}$$

avec une constante absolue.

Preuve. Les arguments sont les mêmes que pour la proposition 2.13 : tout d'abord, on peut déduire du lemme 3.9 ci-dessus que

$$\|f - P_M f\|_{L^\infty(\alpha)} \leq \sum_{\beta \in \mathcal{V}_M(\alpha)} |f|_{BC(\beta)},$$

car une cellule $\alpha \in M$ intersecte toujours deux triangles rectangles isocèles de $\mathcal{K}(M)$, qui eux-mêmes n'intersectent jamais que α et éventuellement une de ses voisines dans M. L'estimation (3.50) découle alors de la proposition 2.10. □

De même, on peut proposer la version suivante de l'algorithme 2.14 pour une fonction f appartenant à $BC(\mathbb{R}^2)$.

Algorithme 3.12 (maillage ε-adaptaté au sens de la courbure totale)

- *Poser* $\Lambda_{\ell_0} := \mathbb{D}_{\ell_0}(\mathbb{R}^2)$.
- *Pour* $\ell \geq \ell_0$, *calculer*

  $$\Lambda_{\ell+1} := \Lambda_\ell \cup \{\alpha \in \mathcal{F}(\beta) : \beta \in \Lambda_\ell \text{ et } |f|_{BC(\beta)} > \varepsilon\}$$

  jusqu'à ce que $\Lambda_{L+1} = \Lambda_L$, *et prendre* $\tilde{M} = \partial\Lambda_L$.
- *Définir* $\boldsymbol{A}_\varepsilon(f)$ *comme le plus petit raffinement gradué (algorithme 2.8) de* $\tilde{M}$.

Remarque 3.13 *Lors du raffinement conditionnel de l'arbre* Λ_ℓ, *dans la deuxième étape de l'algorithme, seules les cellules de niveau* ℓ *sont susceptibles d'être raffinées.*

Notre théorème est alors le suivant.

Théorème 3.2 *Soit* $f \in BC(\mathbb{R}^2)$. *Le maillage dyadique* $\boldsymbol{A}_\varepsilon(f)$ *donné par l'algorithme 3.12 vérifie*

$$\sup_{\alpha \in \boldsymbol{A}_\varepsilon} |f|_{BC(\alpha)} \leq \varepsilon \tag{3.51}$$

de sorte que l'erreur de projection affine par morceaux y vérifie $\|f - P_{\boldsymbol{A}_\varepsilon(f)} f\|_{L^\infty(\mathbb{R}^2)} \leq C\varepsilon$ *avec une constante absolue.*
D'autre part, s'il existe un réel $s > 0$ *et une constante* c_f *pour lesquels* f *vérifie*

$$|f|_{BC(\alpha)} \leq c_f |\alpha|^s, \tag{3.52}$$

alors on peut majorer la complexité de $\boldsymbol{A}_\varepsilon(f)$ *(sur un domaine borné* Ω*) par*

$$\mathrm{card}_\Omega \left(\boldsymbol{A}_\varepsilon(f)\right) \leq C(\varepsilon/c_f)^{-\frac{1}{s}} \tag{3.53}$$

et par

$$\mathrm{card}_\Omega \left(\boldsymbol{A}_\varepsilon(f)\right) \leq C|f|_{BC(\mathbb{R}^2)} \frac{|\log[\varepsilon/c_f]|}{\varepsilon}. \tag{3.54}$$

Remarque 3.14 (interprétation de l'hypothèse (3.52)**)** *On a déjà évoqué, dans la section 1.2.3, la nécessité d'empêcher tout phénomène de "concentration de l'indicateur d'erreur" pour pouvoir estimer la complexité d'un maillage obtenu par découpages dyadiques. Dans le théorème 2.1, c'est l'hypothèse de régularité* $W^{2,p}$, $p > 1$, *faite sur* f *qui jouait ce rôle, car l'inégalité de Hölder nous permettait d'écrire*

$$|f|_{W^{2,1}(\alpha)} \leq |\alpha|^{1-\frac{1}{p}} |f|_{W^{2,p}(\alpha)},$$

la preuve utilisant ensuite la p-sommabilité des termes $\{|f|_{W^{2,p}(\alpha)} : \ell(\alpha) = \ell\}$ *à niveau* ℓ *fixé, et le fait que* $1 - \frac{1}{p}$ *est un exposant strictement positif. La propriété* (3.52)*, qui empêche également une concentration arbitraire de l'indicateur d'erreur, peut donc être lue comme une hypothèse de même nature, et qui sera vérifiée pour les fonctions* f *d'un espace* V_M *associé à un maillage dyadique* M.

Remarque 3.15 *Vis-à-vis de la seule variable* ε *(autrement dit pour une fonction* f *fixée), la première estimation* (3.53) *est meilleure que la deuxième lorsque* $s \geq 1$. *On sera toutefois attentifs au fait que la "constante"* c_f *n'apparaît qu'en logarithme dans la deuxième estimation, ce qui peut être utile si l'on cherche à établir une estimation dépendante de* f.

Preuve. L'estimation (3.51) et son corollaire se démontrent exactement de la même façon que (2.49) dans le théorème 2.1, la proposition 3.11 jouant ici le rôle de la proposition 2.13 dans le cas où la fonction f est dans $W^{2,1}$.

En ce qui concerne la première estimation de complexité, on laisse au lecteur le soin de vérifier que l'argument (direct) utilisé dans la preuve du théorème 2.1 peut être repris à l'identique en introduisant un $p > \frac{1}{s}$, et que l'estimation qui en résulte est précisément (3.53).

Pour montrer (3.54), commençons par déduire de (3.52) que les cellules de $\Lambda(\tilde{M}) \setminus \tilde{M}$, qui vérifient (2.47) par construction, vérifient également

$$\ell(\alpha) \leq Cs^{-1} |\log[\varepsilon/c_f]|. \tag{3.55}$$

Pour estimer la complexité de $\tilde{M}$, on peut alors décomposer cette partition en plusieurs groupes de cellules, dont chacun aura d'une certaine façon "dépensé" une quantité au moins égale à ε de la courbure totale $|f|_{BC(\mathbb{R}^2)}$. La borne (3.55) nous permettra alors de majorer la complexité de chacun de ces groupes, et finalement celle de $\tilde{M}$. Définissons donc, pour toute cellule dyadique $\alpha \in \mathbb{D}(\mathbb{R}^2)$ et pour i entier compris entre 0 et $\ell(\alpha)$, les ensembles

$$S^i(\alpha) := \mathcal{F}\big((\mathcal{P})^{i+1}(\alpha)\big) = \{\beta \in \mathbb{D}_{\ell(\alpha)-i}(\mathbb{R}^2) : \mathcal{P}(\beta) = \mathcal{P}(\cdots(\mathcal{P}(\alpha))) = (\mathcal{P})^{i+1}(\alpha)\}.$$

Ainsi, $S^0(\alpha)$ désigne les "sœurs" de α, (y compris α), $S^1(\alpha)$ ses "tantes" (y compris $\mathcal{P}(\alpha)$), et ainsi de suite, de sorte que la hiérarchie

$$S(\alpha) := \bigcup_{i=0}^{\ell(\alpha)} S_i(\alpha) \tag{3.56}$$

correspond aux cellules créées par un algorithme de découpage récursif pour arriver jusqu'à α (ou l'une quelconque de ses sœurs). En utilisant (3.55), on voit alors que

$$\#\big(S(\alpha)\big) \leq 4\ell(\alpha) \leq C|\log[\varepsilon/c_f]| \text{ lorsque } \alpha \in \tilde{M}, \tag{3.57}$$

et il ne nous reste plus qu'à compter le nombre de "hiérarchies" distinctes de type (3.56) dans l'arbre $\Lambda(\tilde{M})$. Pour résoudre les redondances liées au fait que les différentes sœurs d'un même groupe S^0 possèdent la même hiérarchie, on désigne par

$$\mathcal{S}_\varepsilon(f) := \{\alpha \in \tilde{M} : S^0(\alpha) \subset \tilde{M}\}/\mathcal{R}, \quad \alpha\mathcal{R}\beta \iff \alpha \in S^0(\beta),$$

les cellules dont toutes les sœurs sont dans $\tilde{M}$ modulo l'appartenance au même groupe de sœurs S^0. Pour deux éléments distincts $\dot{\alpha}$ et $\dot{\beta}$ de $\mathcal{S}_\varepsilon(f)$, les cellules parentes $\mathcal{P}(\alpha)$ et $\mathcal{P}(\beta)$ sont alors *d'intérieur disjoints*. En particulier, il existe une partition dyadique M' contenant l'ensemble des $\mathcal{P}(\alpha)$ telles que $\dot{\alpha} \in \mathcal{S}_\varepsilon(f)$. En utilisant le fait que $\mathcal{P}(\alpha)$ vérifie $\varepsilon < |f|_{BC(\mathcal{P}(\alpha))}$, on en déduit que

$$\#\big(\mathcal{S}_\varepsilon(f)\big) \leq \#(\tilde{M}) \leq \sum_{\beta \in \tilde{M}} |f|_{BC(\beta)} \varepsilon^{-1} \leq |f|_{BC(\mathbb{R}^2)} \varepsilon^{-1}. \tag{3.58}$$

Finalement, on peut observer que les groupes $S(\alpha)$ recouvrent $\tilde{M}$ lorsque les classes $\dot{\alpha}$ parcourent $\mathcal{S}_\varepsilon(f)$. D'après (3.57) et (3.58), on en déduit que

$$\#(\tilde{M}) \leq \#\big(\mathcal{S}_\varepsilon(f)\big) \sup_{\alpha\in\tilde{M}} \#\big(S(\alpha)\big) \leq C(\Omega)|f|_{BC(\mathbb{R}^2)} \frac{|\log[\varepsilon/c_f]|}{\varepsilon},$$

ce qui prouve (3.54) en utilisant la proposition 2.9. □

Dans la remarque 3.14 ci-dessus, on a annoncé que la propriété (3.52) était toujours vérifiée par une fonction d'un espace V_M. Concluons donc ce chapitre par une proposition en ce sens.

Proposition 3.16 *Soit $M \in \mathcal{M}(\mathbb{R}^2)$ un maillage dyadique et f une fonction appartenant à V_M. En désignant par $\ell(M) = \sup\{\ell(\beta) : \beta \in M\}$ le niveau maximum des cellules de M, on a*

$$|f|_{BC(\alpha)} \leq C|f|_{W^{1,\infty}} 2^{\ell(M)}|\alpha|$$

pour toute maille dyadique $\alpha \in \mathbb{D}(\mathbb{R}^2)$, avec une constante C absolue.

Preuve. D'après la proposition 3.3 qui établit l'équivalence des courbures totales et discrètes, et en utilisant le fait qu'une fonction de V_M est toujours lipschitzienne, on peut majorer

$$|f|_{BC(\alpha)} \leq C|f|_{\star(\alpha)} \leq C \sum_\gamma |\gamma \cap \alpha|_{\mathcal{H}^1} |f|_{W^{1,\infty}},$$

la somme parcourant les arêtes γ de la triangulation conforme $\mathcal{K}(M)$ associée à M. Dans ces conditions, on observera que la quantité $\sum_\gamma |\gamma \cap \alpha|_{\mathcal{H}^1}$ ne peut qu'augmenter lorsqu'on raffine le maillage M, et que lorsque toutes les mailles de M sont de niveau $\ell(M)$, elle est de l'ordre de $|\alpha|2^{\ell(M)}$, ce qui établit cette proposition. □

Deuxième partie

Etude d'un schéma adaptatif semi-lagrangien pour l'équation de Vlasov

Chapitre 4

Ce qu'il convient de savoir sur l'équation de Vlasov-Poisson

L'équation, ou plutôt le système d'équations auquel on s'intéresse dans cette partie a été introduit par Vlasov en 1948 (voir [64]) pour décrire en termes statistiques l'évolution d'un nuage de particules chargées (des ions et des électrons) sous l'effet d'un champ électromagnétique. On prend ici le temps de présenter cette équation, de rappeler d'où elle vient, et d'écrire les propriétés importantes vérifiées par ses solutions notamment en vue de l'analyse de schéma qu'on mènera au chapitre 6.

4.1 Présentation de l'équation

Dans le modèle de Vlasov, l'état de chaque espèce $\mathcal{E}$ présente dans le plasma est représenté à l'instant t par ce qu'on appelle une *fonction de distribution*, (ou *densité*) $f_{\mathcal{E}}(t)$ définie dans l'espace des phases (positions, vitesses) $\mathbb{R}^d \times \mathbb{R}^d$. La quantité de particules de l'espèce $\mathcal{E}$ situées à l'instant t dans un domaine $\omega_x \subset \mathbb{R}^d$ et de vitesses $v \in \omega_v \subset \mathbb{R}^d$ vaut ainsi

$$Q_{\mathcal{E}}(t,\omega) = \iint_{\omega} f_{\mathcal{E}}(t,x,v)\,\mathrm{d}x\,\mathrm{d}v. \tag{4.1}$$

Bien que certains modèles prennent en compte un grand nombre de particules, on ne considérera ici que deux espèces :

- des ions positifs *relativement lourds*, qu'on supposera répartis de façon uniforme et constante. En particulier, on supposera que leur fonction de distribution vérifie $f_p(t,x,v) = f_p(v)$, et qu'elle est normalisée $\int f_p(v)\,\mathrm{d}v = 1$.
- des électrons *plus légers*, dont la fonction de distribution sera notée f : l'inconnue de notre problème.

On se placera d'autre part en dimension 1, où l'équation de Vlasov-Poisson s'écrit sous la forme suivante

$$\partial_t f(t,x,v) + v \cdot \partial_x f(t,x,v) + E(t,x) \cdot \partial_v f(t,x,v) = 0, \tag{4.2}$$

$$\partial_x E(t,x) = \int f(t,x,v)\,\mathrm{d}v - 1, \tag{4.3}$$

E représentant le champ électrique (normalisé). Enfin, on considérera le problème de Cauchy obtenu en ajoutant à (4.2)-(4.3) une condition initiale

$$f(0,\cdot,\cdot) = f_0 \tag{4.4}$$

régulière et à support compact dans l'espace des phases $\mathbb{R} \times \mathbb{R}$.

4.1.1 Interprétation physique

Depuis les années 1870, et notamment grâce à la théorie unificatrice de Maxwell, on sait qu'un système isolé de particules chargées $\{q_i \in \mathbb{R}, x_i(t) \in \mathbb{R}^3\}_i$, engendre un champ électromagnétique dont il subit simultanément l'influence. Plus précisément, on peut introduire les densités de charge électrique et de courant

$$\rho(t,x) = \sum_i q_i \delta_{\{x_i(t)\}}(x) \quad \text{et} \quad \jmath(t,x) = \sum_i v_i(t) q_i \delta_{\{x_i(t)\}}(x) \tag{4.5}$$

associées à ce système. Le champ électromagnétique $(E,B)(t,x)$ engendré par les particules vérifie alors les *lois de Maxwell* suivantes

$$\nabla \cdot E = \rho/\varepsilon_0 \tag{4.6}$$
$$\nabla \wedge E = -\partial_t B \tag{4.7}$$
$$\nabla \cdot B = 0 \tag{4.8}$$
$$\nabla \wedge B = \mu_0(\jmath + \varepsilon_0 \partial_t E) \tag{4.9}$$

où ε_0 et μ_0 désignent respectivement les constantes de permittivité et de perméabilité diamagnétique du vide. D'autre part, la *force de Lorentz* subie par chaque particule i vérifie

$$F_i(t) = q_i[E(x_i(t)) + v_i(t) \wedge B(t, x_i(t)], \tag{4.10}$$

où $v_i(t) = x_i'(t)$ désigne la vitesse de la particule. D'après la loi fondamentale de la dynamique $m_i \dot{v}_i = F_i$, ceci entraîne que la trajectoire de la particule dans l'espace des phases est solution de l'équation différentielle ordinaire

$$\dot{x}_i(t) = v_i(t), \qquad \dot{v}_i(t) = (q_i/m_i)[E(x_i(t)) + v_i(t) \wedge B(t, x_i(t)]. \tag{4.11}$$

L'*approximation de Vlasov-Poisson* considère le cas où les effets du champ magnétique sont négligés, du moins en ce qui concerne le champ auto-induit, car il n'est pas très difficile d'étendre l'analyse de (4.2) à celle de

$$\partial_t f(t,x,v) + v \cdot \partial_x f(t,x,v) + (E(t,x) + E_{\text{ext}}(t,x) + v \wedge B_{\text{ext}}(t,x)) \cdot \partial_v f(t,x,v) = 0, \tag{4.12}$$

où $(E_{\text{ext}}, B_{\text{ext}})$ est un champ électromagnétique imposé par un système extérieur indépendant. Le couplage entre le champ E et l'état du plasma se réduit alors à l'équation (4.6), qu'on a l'habitude d'appeler de Maxwell-Gauss dans le système complet (4.6)-(4.9). Comme E est alors de rotationnel nul, il dérive d'un potentiel électrique ϕ au sens où

$$E = -\nabla\phi, \tag{4.13}$$

et l'équation (4.6) est équivalente à l'équation de Poisson

$$\Delta\phi = -\rho/\varepsilon_0. \tag{4.14}$$

En utilisant des constantes normalisées, on peut déduire le modèle de Vlasov de ce qui précède en écrivant que la densité d'électrons f doit être conservée ponctuellement le long des trajectoires correspondant à (4.11). Plus précisément, on appelle *trajectoire caractéristique du problème* toute solution

$$t \to (X(t), V(t)) = (X(t; s, x, v), V(t; s, x, v)) \tag{4.15}$$

du système différentiel ordinaire

$$\partial_t X(t) = V(t), \qquad \partial_t V(t) = E(t, X(t)), \qquad (X, V)(s) = (x, v). \tag{4.16}$$

La densité d'électrons f doit alors vérifier

$$\partial_t f(t, X(t; 0, x, v), V(t; 0, x, v)) = 0 \tag{4.17}$$

pour tout couple (x, v) de l'espace des phases $\mathbb{R}^6$, ce qui nous conduit au système tri-dimensionnel de Vlasov-Poisson :

$$\partial_t f(t, x, v) + v \cdot \nabla_x f(t, x, v) + E(t, x) \cdot \nabla_v f(t, x, v) = 0, \qquad x, v \in \mathbb{R}^3 \tag{4.18}$$

$$E(t, x) = -\nabla_x \phi(t, x), \qquad -\Delta_x \phi(t, x) = \rho(x, t). \tag{4.19}$$

4.1.2 Propriétés de transport

Il est intéressant d'observer que l'équation (4.17), qui exprime le transport des valeurs ponctuelles de f le long des trajectoires caractéristiques, entraîne également une propriété de transport local des mesures de charges associées à f, autrement dit

$$\iint_{\omega} f(0, x, v)\, \mathrm{d}x\, \mathrm{d}v = \iint_{\mathcal{A}_t(\omega)} f(t, x, v)\, \mathrm{d}x\, \mathrm{d}v, \qquad \text{pour tout } \omega \subset \Omega, \tag{4.20}$$

où

$$\mathcal{A}_t : (x, v) \to (X, V)(t; s, x, v) \tag{4.21}$$

désigne le difféomorphisme associé aux trajectoires (4.16). On peut voir cette propriété comme une conséquence de la divergence nulle, dans l'espace des phases, du champ d'advection $(v, E(t, x))$. Premièrement, parce que cela permet d'écrire l'équation (4.2) sous forme conservative, *i.e.*

$$\partial_t f(t, x, v) + \nabla_{x,v} \cdot [(v, E(t, x)) f(t, x, v)] = 0. \tag{4.22}$$

Ensuite, parce qu'en introduisant le champ à divergence nulle $\Phi(t, x, v) = (1, v, E(t, x))$ défini sur le domaine

$$\mathcal{D}_{\tau,\omega} = \{(t, x, v) : t \in [0, \tau], \mathcal{A}_t^{-1}(x, v) \in \omega\} \subset [0, \tau] \times \Omega, \tag{4.23}$$

on voit d'après la formule de Stokes que

$$\iint_{\mathcal{A}_\tau(\omega)} \mathrm{d}x\, \mathrm{d}v - \iint_{\omega} \mathrm{d}x\, \mathrm{d}v = \iint_{\partial \mathcal{D}_{\tau,\omega}} \Phi \cdot n\, \mathrm{d}\sigma = \iiint_{\mathcal{D}_{\tau,\omega}} \nabla \cdot \Phi\, \mathrm{d}t\, \mathrm{d}x\, \mathrm{d}v = 0, \tag{4.24}$$

la première égalité venant du fait que la frontière de $\mathcal{D}_{\tau,\omega}$ suit les lignes de champ de Φ en dehors des "faces" $\mathcal{A}_\tau(\omega)$ et ω. Or (4.24) signifie précisément que le difféomorphisme $\mathcal{A}_t$ préserve la mesure de Lebesgue, donc que son Jacobien est égal à un, et la propriété (4.20) se déduit alors immédiatement de (4.17).

4.2 Existence et unicité des solutions

Pour décrire notre schéma et montrer sa convergence, on supposera que la donnée initiale est lipschitzienne et de courbure totale bornée, autrement dit qu'elle appartient à l'espace $W^{1,\infty} \cap W^{2,1}(\mathbb{R}^2)$.

4.2.1 Solutions faibles

Au début des années 1970, Arsen'ev [3] a démontré l'existence globale en temps de solutions du problème tri-dimensionnel (4.18)-(4.19) pour des conditions très faibles de régularité, à savoir lorsque les solutions initiales sont dans $L^1 \cap L^2(\mathbb{R}^3 \times \mathbb{R}^3)$ et sont d'énergie finie

$$\iint f_0(x,v)|v|^2 \,\mathrm{d}x\,\mathrm{d}v + \int |E_0(x)|^2 \,\mathrm{d}x < \infty. \tag{4.25}$$

Pour ces solutions, l'unicité n'a été démontrée que bien plus tard, lorsque Lions et Perthame [47] ont prouvé que sous des hypothèses relativement peu contraignantes sur la donnée initiale, les solutions ont des moments d'ordres élevés en vitesse et par conséquent des propriétés de régularité d'où l'on peut déduire l'unicité.

4.2.2 Solutions classiques en dimension 1

Définition 4.1 *On dira que le couple (f, E) est une solution classique du problème* (4.2)-(4.3) *associé à la donnée initiale f_0 si*

1. *f est continue sur $[0,\infty[\times\mathbb{R}\times\mathbb{R}$,*
2. *E est lipschitzienne sur $[0,\infty[\times\mathbb{R}$,*
3. *l'équation de Vlasov* (4.2) *est vérifiée au sens des distributions.*

Compte tenu du caractère lipschitzien de E, les trajectoires caractéristiques peuvent être définies en tout point (x,v) de l'espace des phases, aussi le point 3 est-il équivalent à

3'. f est constante sur les trajectoires caractéristiques définies par (4.16).

Remarque 4.2 *Lorsque f n'est que continue, les dérivées apparaissant dans l'équation* (4.2) *doivent être prises en un sens faible. On les appelle néanmoins solutions classiques (ou fortes), car les trajectoires caractéristiques sont bien définies et l'équation* (4.17) *est vérifiée en un sens classique.*

Un des tous premiers résultats d'existence est dû à Iordanskii (voir [43]) qui a démontré au début des années 1960 l'existence globale en temps et l'unicité d'une solution classique sous quelques conditions. La donnée initiale f_0 doit bien sûr être continue et elle doit également vérifier les hypothèses suivantes d'intégrabilité

$$\rho_0(x) = \int f_0(x,v)\,\mathrm{d}v - 1 < \infty \quad \text{et} \quad \int v^2\theta(v)\,\mathrm{d}v < \infty, \tag{4.26}$$

où θ est une fonction décroissante de $|v|$ qui domine f_0 et f_p. Le champ électrique, enfin, doit satisfaire la condition limite

$$\lim_{x\to-\infty} E(t,x) = 0, \qquad t > 0, \tag{4.27}$$

condition "nécessaire" dans la mesure où seule la divergence de E apparaît dans l'équation. On pourra remarquer que les hypothèses (4.26) sont assez naturelles car elles permettent de donner un sens (au moins à l'instant initial) aux densités de courant et d'énergie cinétique, deux quantités physiques fondamentales respectivement définies par

$$\jmath(t,x) = \int v[f(t,x,v) - f_p(v)]\,\mathrm{d}v \quad \text{et} \quad \varepsilon_c(t,x) = \int v^2[f(t,x,v) + f_p(v)]\,\mathrm{d}v. \tag{4.28}$$

Enfin, Iordanskii montre que f et E vérifient une équation supplémentaire, à savoir la loi d'Ampère

$$\partial_t E(t,x) = \int v[f_p(v) - f(t,x,v)]\,\mathrm{d}v. \tag{4.29}$$

En 1980, Cooper et Klimas [24] ont étendu ces résultats à des conditions aux limites plus générales que (4.27), en particulier dans le cas d'un problème périodique où $x \in \mathbb{T} := \mathbb{R}/\mathbb{Z}$. *C'est dans ce cadre qu'on se placera pour décrire notre schéma, et les solutions initiales qu'on considérera seront à supports compacts.* Leur résultat est le suivant.

Théorème 4.1 *Si f_0 est continue sur $\mathbb{T} \times \mathbb{R}$ (donc 1-périodique en x), si elle vérifie*

$$\rho_0(x) = \int f_0(x,v)\,\mathrm{d}v - 1 < \infty \quad \text{et} \quad \int |v|\theta(v)\,\mathrm{d}v < \infty, \tag{4.30}$$

où θ est définie comme en (4.26)*, et si la charge est nulle sur une période,* i.e. *si*

$$\int_0^1 \rho_0(x)\,\mathrm{d}x = \int_0^1 \int f_0(x,v)\,\mathrm{d}v\,\mathrm{d}x - 1 = 0, \tag{4.31}$$

alors il existe une unique solution classique de (4.2)-(4.3) *vérifiant $\int_0^1 E(0,x)\,\mathrm{d}x = 0$, et cette solution est 1-périodique en x.*

On pourra observer que la continuité du champ $E(t,\cdot)$ en 0 est équivalente à la propriété de conservation globale de masse vérifiée par f

$$\iint f(t,x,v)\,\mathrm{d}x\,\mathrm{d}v = \iint f_0(x,v)\,\mathrm{d}x\,\mathrm{d}v = 1. \tag{4.32}$$

En désignant par $-G(x,y)$ la fonction de Green associée à l'équation de Poisson (4.3) uni-dimensionnelle, soit pour tout $y \in [0,1]$, la solution de

$$\partial^2_{xx} G(\cdot,y) = \delta(\cdot - y) \quad \text{sur} \quad [0,1] \tag{4.33}$$

avec conditions aux bords périodiques $G(0,y) = G(1,y)$, on peut voir que E vérifie

$$E(t,x) = \int K(x,y)\left(\int f(t,y,v)\,\mathrm{d}v - 1\right)\mathrm{d}y \tag{4.34}$$

où

$$K(x,y) = \partial_x G(x,y) = \begin{cases} y-1 & \text{si } \ 0 \le x < y \\ y & \text{si } \ y \le x \le 1. \end{cases} \tag{4.35}$$

4.3 Régularité des solutions classiques

Pour établir la précision d'un schéma de discrétisation en temps du système de Vlasov-Poisson dans la section 6.1.2, on aura besoin d'utiliser la régularité des solutions f et E. On sait (et l'on peut citer sur ce sujet l'article [55] de Raviart) que lorsque la solution initiale f_0 appartient à un espace de Sobolev $W^{m,p}$, c'est également le cas de la solution $f(t)$ pour tout $t < \infty$. Plus précisément, la régularité dont on aura besoin est résumée dans la proposition suivante. On en donne ici une preuve complète (inspirée notamment des techniques exposées dans le livre [34] de Glassey), car elle est assez simple dans ce contexte uni-dimensionnel, et qu'on s'en est inspiré pour établir des estimations d'erreur dans l'étude du schéma numérique.

Proposition 4.3 *On désigne ici par Ω l'espace des phases $\mathbb{T} \times \mathbb{R}$ périodisé en x. Si f_0 appartient à $W^{1,\infty}(\Omega)$ et vérifie les hypothèses* (4.30)-(4.31) *du théorème de Cooper et Klimas, alors pour tout temps final $T < \infty$, la solution possède un support borné en vitesse*

$$\Sigma_v(f(t)) := \sup\{|v| : \exists x, f(t,x,v) > 0\} \leq \Sigma_v(f_0) + 2T, \qquad t \leq T \tag{4.36}$$

et satisfait les estimations de régularité suivantes :

$$\begin{array}{lcl} \|f\|_{L^\infty([0,T];W^{1,\infty}(\Omega))} & \leq & C(f_0,T) \\ \|\partial_t f\|_{L^\infty([0,T];L^\infty(\Omega))} & \leq & C(f_0,T) \\ \|E\|_{L^\infty([0,T];W^{2,\infty}([0,1]))} & \leq & C(f_0,T) \\ \|\partial_t E\|_{L^\infty([0,T];W^{1,\infty}([0,1]))} & \leq & C(f_0,T) \\ \|\partial^2_{tt} E\|_{L^\infty([0,T];L^\infty([0,1]))} & \leq & C(f_0,T). \end{array} \tag{4.37}$$

Preuve. Pour commencer, montrons que les bornes

$$\|E\|_{L^\infty([0,T];W^{1,\infty}([0,1]))} \leq C(T) \tag{4.38}$$

et

$$\|\partial_t E\|_{L^\infty([0,T];L^\infty([0,1]))} \leq C(T) \tag{4.39}$$

s'obtiennent dès que f_0 est continue (elles sont d'ailleurs prouvées dans [24]). La conservation de f le long des trajectoires caractéristiques (4.15) entraînant d'une part un "principe du maximum" en temps

$$0 \leq f \leq \|f_0\|_{L^\infty(\Omega)}, \tag{4.40}$$

d'autre part une propagation bornée en vitesse

$$\Sigma_v(f(t)) - \Sigma_v(f_0) \leq \sup_{(x,v)\in\Omega} \int_0^T |\partial_t V(\tau;0,x,v)|\,\mathrm{d}\tau \leq T\|E\|_{L^\infty([0,T];L^\infty([0,1]))}, \tag{4.41}$$

on obtient en utilisant successivement (4.34), (4.40) et (4.32) :

$$\|E(t)\|_{L^\infty([0,1])} \leq \|K\|_{L^\infty} \left(\iint |f(t,x,v)|\,\mathrm{d}x\,\mathrm{d}v + 1 \right) \leq 2. \tag{4.42}$$

Grâce à (4.41), cette dernière inégalité entraîne (4.36). On a d'autre part

$$\|\partial_x E(t,\cdot)\|_{L^\infty([0,1]))} \leq \Sigma_v(f(t))\|f_0\|_{L^\infty(\Omega)} + 1 \tag{4.43}$$

en utilisant l'équation de Poisson (4.3), et

$$\|\partial_t E(t,\cdot)\|_{L^\infty([0,1]))} \leq \Sigma_v(f(t))^2\|f_0\|_{L^\infty(\Omega)} + \int v f_p(v)\,\mathrm{d}v$$

en utilisant la loi d'Ampère (4.29), ce qui établit respectivement (4.38) et (4.39). Si on suppose ensuite $f_0 \in W^{1,\infty}(\Omega)$, on peut écrire

$$|f(t,x,v) - f(t,\tilde{x},\tilde{v})| = |f_0(X_0(t),V_0(t)) - f_0(\tilde{X}_0(t),\tilde{V}_0(t))| \leq |f_0|_{W^{1,\infty}}\big(|e_x(t)| + |e_v(t)|\big)$$

en désignant pour $0 \leq s \leq t \leq T$

$$\left\{\begin{array}{l}(X_0,V_0)(s) := (X,V)(t-s;t,x,v)\\(\tilde{X}_0,\tilde{V}_0)(s) := (X,V)(t-s;t,\tilde{x},\tilde{v})\end{array}\right. \text{ et } \left\{\begin{array}{l}e_x(s) := X_0(s) - \tilde{X}_0(s)\\e_v(s) := V_0(s) - \tilde{V}_0(s)\end{array}\right. . \tag{4.44}$$

D'après la définition (4.16) des trajectoires caractéristiques, ces quantités vérifient $e_x'(s) = e_v(s)$ et $e_v'(s) = E(t-s,X_0(s)) - E(t-s,\tilde{X}_0(s))$. En utilisant (4.38), on obtient donc

$$|e_x'(s)| + |e_v'(s)| \leq C(T)\big(|e_x(s)| + |e_v(s)|\big). \tag{4.45}$$

On en déduit que la fonction $\psi(s) := |e_x(s)| + |e_v(s)|$ vérifie

$$\psi(t) = \psi(0) + \Big|\int_0^t e_x'(s)\,\mathrm{d}s\Big| + \Big|\int_0^t e_v'(s)\,\mathrm{d}s\Big| \leq \psi(0) + C(T)\int_0^t \psi(s)\,\mathrm{d}s, \tag{4.46}$$

et le lemme de Gronwall (1.5) nous permet d'écrire

$$\big(|e_x(t)| + |e_v(t)|\big) \leq C(T)\big(|e_x(0)| + |e_v(0)|\big) \leq C(T)\big(|x-\tilde{x}| + |v-\tilde{v}|\big). \tag{4.47}$$

Ceci nous montre que $f(t)$ est bien uniformément lipschitzienne sur $[0,T]$, et par conséquent que la norme $\|f\|_{L^\infty([0,T];W^{1,\infty}(\Omega))}$ est bien finie pour tout temps T. La borne

$$\|\partial_t f(t)\|_{L^\infty(\Omega)} \leq Q(T)\|\partial_x f(t)\|_{L^\infty(\Omega)} + \|E\|_{L^\infty([0,1])}\|\partial_v f(t)\|_{L^\infty(\Omega)}$$

provient alors directement de l'équation de Vlasov (4.2). Voyons maintenant le champ électrique : en dérivant l'équation de Poisson (4.3) par rapport à x et t, on obtient respectivement

$$\|\partial_{xx}^2 E\|_{L^\infty([0,T];L^\infty([0,1]))} \leq Q(T)\|\partial_x f\|_{L^\infty([0,T];L^\infty(\Omega))} \tag{4.48}$$

$$\|\partial_{tx}^2 E(t)\|_{L^\infty([0,T];L^\infty([0,1]))} \leq Q(T)\|\partial_t f\|_{L^\infty([0,T];L^\infty(\Omega))}, \tag{4.49}$$

et on majore facilement $\|\partial_{tt}^2 E\|_{L^\infty([0,T];L^\infty([0,1]))}$ en dérivant l'équation d'Ampère (4.29) par rapport à t. □

Chapitre 5

Le schéma adaptatif semi-lagrangien dans un cadre abstrait

Afin de mettre en avant notre stratégie d'adaptation dynamique des maillages de calculs, on considère dans ce chapitre un problème de transport "abstrait" de type Vlasov, qu'on suppose discrétisé en temps par une méthode lagrangienne dont les propriétés s'inspirent de celles qu'on établira au chapitre suivant pour un schéma de time-splitting appliqué au système de Vlasov-Poisson.

On commence donc par décrire ces hypothèses - numérotées de (HT.1) à (HT.5) - avant de présenter la façon dont notre schéma fait évoluer les maillages d'un pas de temps à l'autre. Cette évolution se fait en deux étapes : lors d'une première étape, on *prédit* le maillage sur lequel sera approchée la prochaine solution numérique. On calcule alors cette solution, et dans une deuxième étape, on *corrige* le maillage prédit pour équilibrer au mieux la courbure totale de la solution obtenue. D'un point de vue algorithmique, et dans la mesure où notre schéma met en œuvre les éléments finis multi-échelles introduits au chapitre 2, il est possible de décrire cette stratégie d'adaptation de maillage en ne faisant intervenir que des *maillages dyadiques* composées de cellules carrées.

La principale nouveauté de notre approche réside alors sans doute dans l'opérateur de prédiction de maillage $\boldsymbol{T}[\mathcal{A}]$ qu'on présente dans la section 5.2, et qui repose essentiellement sur un calcul des trajectoires caractéristiques associées au flot $\mathcal{A}$. Cet opérateur, qui *associe à un maillage dyadique un autre maillage dyadique*, ne transporte évidemment pas les mailles de façon exacte, mais par simplicité, nous en parlerons comme d'un opérateur de "transport de maillages". Enfin, et sous réserve de vérifier les hypothèses (HT.1)-(HT.5) faites sur le problème abstrait et sur sa discrétisation en temps, on établit un résultat de précision uniforme *a priori* pour ce schéma adaptatif, ainsi qu'une analyse partielle de sa complexité.

5.1 Hypothèses de travail

On s'intéresse donc à une équation de transport de type Vlasov

$$\partial_t f(t,x,v) + F(t,x,v)\cdot\nabla_{x,v} f(t,x,v) = 0, \quad t>0, \quad (x,v)\in\mathbb{R}^2 \tag{5.1}$$

associée à une condition initiale

$$f(0,\cdot) = f_0 \in W^{1,\infty},$$

où la non-linéarité vient du fait que le terme de force $F : \mathbb{R}_+ \times \mathbb{R}^2 \to \mathbb{R}^2$ est couplé avec f. On supposera néanmoins que la solution f existe de façon unique sur tout intervalle de temps $[0,T]$, et qu'elle est uniformément lipschitzienne

$$f \in L^\infty([0,T]; W^{1,\infty}(\mathbb{R}^2)). \tag{5.2}$$

On supposera aussi qu'entre les instants $t_n := n\Delta t$ et t_{n+1}, f est transportée le long de trajectoires caractéristiques $t \to (X(t),V(t)) = (X(t;t_n,x,v),V(t;t_n,x,v))$ associées au champ $F(t,x)$ de la même façon qu'en (4.15)-(4.16). L'opérateur de transport exact $\mathcal{T}^{\text{exact}} = \mathcal{T}^{\text{exact}}_{\Delta t} : f(t_n) \to f(t_{n+1})$ peut alors s'écrire sous la forme

$$\mathcal{T}^{\text{exact}} f(t_n) = f(t_n) \circ \big(\mathcal{A}^{\text{exact}}[f(t_n)]\big)^{-1}, \tag{5.3}$$

où

$$\mathcal{A}^{\text{exact}}[f(t_n)] : (x,v) \to (X(t_{n+1};t_n,x,v), V(t_{n+1};t_n,x,v))$$

désigne le difféomorphisme de $\mathbb{R}^2$ dans lui-même qui représente le flot caractéristique entre t_n et t_{n+1}.

D'autre part, on suppose connue une discrétisation en temps matérialisée par un opérateur de transport approché $\mathcal{T} : W^{1,\infty}(\mathbb{R}^2) \to W^{1,\infty}(\mathbb{R}^2)$ vérifiant

$$\mathcal{T} f(t_n) = f(t_n) \circ \big(\mathcal{A}[f(t_n)]\big)^{-1} \approx f(t_{n+1}), \tag{5.4}$$

où $\mathcal{A}[f(t_n)] = \mathcal{A}_{\Delta t}[f(t_n)]$ désigne un difféomorphisme de $\mathbb{R}^2$ dans lui-même qui approche le flot caractéristique exact $\mathcal{A}^{\text{exact}}[f(t_n)]$. On peut signaler que dans le chapitre 6, ce transport approché sera donné par (6.6)-(6.7).

Remarque 5.1 (à propos de la non-linéarité des opérateurs de transport) *L'équation* (5.1) *étant non-linéaire, il en est de même pour les opérateurs de transport* $\mathcal{T}^{\text{exact}}$ *et* $\mathcal{T}$*. En ce qui concerne les flots caractéristiques (exacts ou approchés), ceci se manifeste par une dépendance vis-à-vis d'une "donnée de départ" lipschitzienne :* $\mathcal{A}[g]$*, par exemple, correspond au déplacement des trajectoires "issues" de* $g \in W^{1,\infty}$ *pendant une durée* Δt*, (indépendamment, d'ailleurs, de l'instant initial). On notera* $\mathcal{A}[\cdot]$ *l'opérateur de déplacement approché défini par* (5.4)*, la lettre* $\mathcal{A}$ *seule désignant un difféomorphisme quelconque de* $\mathbb{R}^2$*, éventuellement donné par* $\mathcal{A}[g]$*. Parallèlement, on pourra désigner par* $\mathcal{T}[g] : \tilde{g} \to \tilde{g} \circ \big(\mathcal{A}[g]\big)^{-1}$ *l'opérateur de transport* linéraire *associé à une donnée* g*, en observant que* $\mathcal{T}$ *vérifie* $\mathcal{T} g = \mathcal{T}[g]g$*.*

5.1.1 Précision et régularité du transport approché

Notre première hypothèse porte sur l'ordre de la discrétisation en temps. On suppose ainsi qu'il existe un $\sigma > 0$ pour lequel

$$\|f(t_{n+1}) - \mathcal{T} f(t_n)\|_{L^\infty} \leq c_1 \Delta t^{\,\sigma+1}. \tag{HT.1}$$

En utilisant (5.2), on pourra remarquer qu'il suffit pour cela que les déplacements approchés vérifient

$$\sup_{x\in\mathbb{R}^2} \left| \big(\mathcal{A}[f(t_n)]\big)^{-1}(x) - \big(\mathcal{A}^{\text{exact}}[f(t_n)]\big)^{-1}(x) \right| \leq C\Delta t^{\,\sigma+1}, \tag{5.5}$$

où $|x| := \max_i |x_i|$ désigne la norme ℓ^∞ de $\mathbb{R}^2$.

On supposera également que les déplacements $\mathcal{A}[g]$ et $\big(\mathcal{A}[g]\big)^{-1}$ sont uniformément lipschitziens. Plus précisément, qu'il existe une constante c_2 telle que

$$\big|\mathcal{A}[g](x) - \mathcal{A}[g](x')\big| \leq c_2 |x - x'| \tag{HT.2}$$

et une constante $c_3 < 2$ telle que

$$\Big|\big(\mathcal{A}[g]\big)^{-1}(x) - \big(\mathcal{A}[g]\big)^{-1}(x')\Big| \leq c_3 |x - x'| \tag{HT.3}$$

pour toute fonction $g \in W^{1,\infty}$.

Remarque 5.2 *Dans le chapitre suivant, cette hypothèse sera vérifiée dès lors que le pas de temps Δt sera inférieur à une constante qui dépendra uniquement de la donnée initiale f_0 et du temps final $T = N\Delta t$. En particulier, Δt ne sera pas soumis à une condition de type Courant-Friedrichs-Lewy, difficilement compatible avec la présence de mailles arbitrairement fines.*

Enfin, le transport approché vérifiera une propriété de "Δt-stabilité" vis-à-vis des perturbations de densités

$$\|[\mathcal{T}[g_1] - \mathcal{T}[g_2]]g_3\|_{L^\infty} \leq c_4 \Delta t\, |g_3|_{W^{1,\infty}} \|g_1 - g_2\|_{L^\infty}. \tag{HT.4}$$

Remarque 5.3 *Les hypothèses (HT.1) et (HT.4) correspondent à une précision* globale *d'ordre σ pour le schéma de discrétisation en temps*

$$f_\tau^0 := f_0 \quad \textit{et} \quad f_\tau^{n+1} := \mathcal{T} f_\tau^n.$$

On a en effet après N pas de temps

$$\|f(N\Delta t) - f_\tau^N\|_{L^\infty} \leq C\Delta t^{\,\sigma}, \tag{5.6}$$

la constante C ne dépendant que de c_1, c_4, du temps final $N\Delta t$ et de f_0. Pour le voir, on peut "fixer la linéarité" des opérateurs de transport en écrivant $\tilde{\mathcal{T}}^n = \mathcal{T}[f(t_n)]$ et $\mathcal{T}_\tau^n = \mathcal{T}[f_\tau^n]$, ce qui nous permet de décomposer $e_{n+1}^\tau := \|f(t_{n+1}) - f_\tau^{n+1}\|_{L^\infty}$ suivant

$$\begin{aligned} e_{n+1}^\tau &\leq \|f(t_{n+1}) - \mathcal{T} f(t_n)\|_{L^\infty} + \|\big(\tilde{\mathcal{T}}^n - \mathcal{T}_\tau^n\big) f(t_n)\|_{L^\infty} + \|\mathcal{T}_\tau^n \big(f(t_n) - f_\tau^n\big)\|_{L^\infty} \\ &\leq c_1 \Delta t^{\,\sigma+1} + (1 + c_4 \Delta t\, |f(t_n)|_{W^{1,\infty}}) \|f(t_n) - f_\tau^n\|_{L^\infty} \\ &\leq C\Delta t^{\,\sigma+1} + (1 + C\Delta t)\, e_n^\tau. \end{aligned}$$

Pour la deuxième inégalité, on a utilisé les hypothèses (HT.1), (HT.4) et le fait qu'un transport ne fait jamais augmenter la norme L^∞. Quant à la troisième, elle vient de l'hypothèse (5.2)*. L'estimation* (5.6) *découle alors du lemme de Gronwall discret 1.4.*

5.1.2 Stabilité locale vis-à-vis de l'indicateur d'erreur

On énonce à présent une hypothèse qui sera fondamentale dans l'analyse d'erreur du schéma, et qui exprime à la fois une propriété de l'opérateur de transport approché $\mathcal{T}$ et des discrétisations adaptatives.

Dans la première partie, on a montré que la semi-norme $W^{2,1}$ d'une fonction g donnée (ou sa courbure totale, si elle n'appartient qu'à l'espace $BC(\mathbb{R}^2)$) pouvait être vue comme un bon *indicateur local* de l'erreur d'interpolation affine par morceaux $\|g - P_M g\|_{L^\infty}$ associée à un maillage dyadique M. Par "bon indicateur", on entend ici

- d'une part, le fait que cet indicateur *contrôle localement les erreurs d'interpolation*, au sens des propositions 2.13 et 3.11,
- d'autre part, le fait qu'un maillage dyadique obtenu par un équilibrage de ces indicateurs aura une complexité optimale ou quasi-optimale, au sens des théorèmes 2.1 et 3.2.

La propriété de compatibilité qu'on exige alors du transport approché est une *stabilité locale* vis-à-vis de ces indicateurs d'erreur. Pour écrire cette propriété, on va désigner par $\mathcal{E}(g, \alpha)$ l'indicateur d'erreur (6.15) qu'on sera amené à considérer au chapitre suivant. Cet indicateur sera défini pour une fonction $g \in W^{1,\infty} \cap BC(\mathbb{R}^2)$ et une cellule dyadique $\alpha \in \mathbb{D}(\mathbb{R}^2)$, et de même que la courbure totale $|g|_{BC(\alpha)}$ (et la semi-norme $|g|_{W^{2,1}(\alpha)}$), il vérifiera les propriétés suivantes :

1. $\mathcal{E}$ contrôle les erreurs locales de projection au sens où

$$\|g - P_M g\|_{L^\infty(\alpha)} \leq C \sum_{\beta \in \mathcal{V}_M(\alpha)} \mathcal{E}(g, \beta) \text{ pour tout } \alpha \in M \tag{5.7}$$

   pour une constante C indépendante du maillage dyadique $M \in \mathcal{M}(\mathbb{R}^2)$ et de la fonction g (l'ensemble $\mathcal{V}_M(\alpha)$ désignant les cellules voisines de α dans le maillage M).

2. pour toute fonction g, l'indicateur a priori $\mathcal{E}(g, \cdot)$ est sous-additif

$$\sum_{\beta \in \mathcal{F}(\alpha)} \mathcal{E}(g, \beta) \leq \mathcal{E}(g, \alpha), \tag{5.8}$$

   de sorte qu'en désignant par

$$\mathcal{E}_{\ell_0}(g) := \sum_{\ell(\alpha) = \ell_0} \mathcal{E}(g, \alpha) \tag{5.9}$$

   la somme de ses valeurs prises au niveau le plus bas, on voit que

$$\sum_{\alpha \in M} \mathcal{E}(g, \alpha) \leq \mathcal{E}_{\ell_0}(g) \tag{5.10}$$

   pour toute partition dyadique M, graduée ou non.

3. lorsque α est une cellule de M, P_M est stable par rapport à $\mathcal{E}(\cdot, \alpha)$, au sens où

$$\mathcal{E}(P_M g, \alpha) \leq C \mathcal{E}(g, \alpha) \tag{5.11}$$

   avec une constante C indépendante du maillage $M \in \mathcal{M}(\mathbb{R}^2)$ et de la fonction g.

Pour formuler la stabilité du transport $\mathcal{T}$ vis-à-vis de cet indicateur d'erreur, on utilise la notion de *domaine d'influence dans M d'une cellule α vis-à-vis d'un flot $\mathcal{A}$* (qui est un difféomorphisme de $\mathbb{R}^2$ dans lui-même), défini comme l'ensemble des cellules de M dont l'image par le flot $\mathcal{A}$ intersecte α :

$$\mathcal{I}_{M,\mathcal{A}}(\alpha) := \{\beta \in M : \mathcal{A}(\beta) \cap \alpha \neq \emptyset\}. \tag{5.12}$$

Remarque 5.4 *Dans la mesure où α sera dans la suite une cellule potentielle d'un maillage* prédit *pour une solution à venir, on pourra voir $\mathcal{I}_{M,\mathcal{A}}$ comme un domaine d'influence* dans le passé.

La propriété minimale de compatibilité qu'on fera sur le transport approché $\mathcal{T} : g \to g \circ (\mathcal{A}[g])^{-1}$ est alors qu'il existe une constante absolue C pour laquelle on a

$$\mathcal{E}(\mathcal{T}g, \alpha) \leq C \sum_{\beta \in \mathcal{I}_{M,\mathcal{A}[g]}(\alpha)} \mathcal{E}(g, \beta) \tag{HT.5}$$

pour toute cellule dyadique α, tout maillage $M \in \mathcal{M}(\mathbb{R}^2)$ et toute fonction g. Bien entendu, cette hypothèse n'a de sens que si l'indicateur d'erreur $\mathcal{E}$ est bien défini lorsque les solutions transportées $\mathcal{T}g$ appartiennent à $W^{1,\infty} \cap BC$, autrement dit lorsque $\mathcal{T}$ préserve cette régularité.

A la fin de ce chapitre, on verra que cette stabilité (HT.5) permettra, sous réserve d'une propriété importante (cf. (5.28)) qui devra être vérifiée par les maillages prédits, d'établir *une estimation d'erreur a priori* pour notre schéma adaptatif. Pour pouvoir écrire un résultat de *complexité*, on aura besoin que le transport approché vérifie une stabilité "lipschitzienne" vis-à-vis de cet indicateur d'erreur, à savoir qu'il existe une constante absolue C telle que

$$\mathcal{E}(\mathcal{T}g, \alpha) \leq (1 + C\Delta t\,) \sum_{\beta \in \mathcal{I}_{M,\mathcal{A}[g]}(\alpha)} \mathcal{E}(g, \beta) \tag{HT.5$'$},$$

ce qui à nouveau, est une propriété raisonnable dans la mesure où Δt représente précisément le pas de temps sur lequel $\mathcal{T}$ transporte les solutions. Cette stabilité "améliorée" sera également vérifiée au chapitre suivant.

5.2 Gestion dynamique des maillages

Dans la section 3.3.2, on a décrit un algorithme de découpage récursif $\boldsymbol{A}_\varepsilon$ qui à une fonction donnée g associait le plus petit maillage dyadique sur lequel les courbures totales $|g|_{BC(\alpha)}$ étaient uniformément bornées par un paramètre arbitraire ε. En utilisant le même principe, on décrit à présent un algorithme de *prédiction* du maillage de calcul.

5.2.1 Transport des maillages dyadiques

L'idée mise en œuvre par notre opérateur $\boldsymbol{T}[\mathcal{A}]$ consiste à transporter le long du flot $\mathcal{A}$ ce qu'on pourrait appeler la "carte de résolution locale" d'un maillage M, autrement dit l'application

$$(x, v) \to \max\{\ell(\alpha) : \alpha \in M, (x, v) \in \bar{\alpha}\},$$

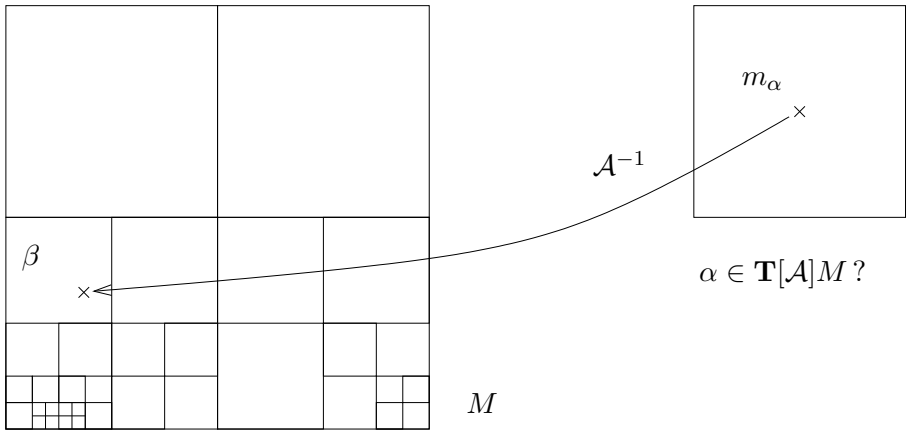


FIG. 5.1 – "transport" du maillage par comparaison des niveaux $\ell(\alpha)$ et $\ell^*_{M,\mathcal{A}}(\alpha)$.

qu'on peut voir comme une surface constante sur les mailles de M (le max n'étant utile que lorsque (x, v) est situé sur le bord d'une maille).

Pour un maillage dyadique M donné, appelons donc *niveau rétrograde* d'une cellule α (relativement au difféomorphisme $\mathcal{A}$) l'entier

$$\ell^*_{M,\mathcal{A}}(\alpha) := \max\{\ell(\beta) : \beta \in M, \mathcal{A}^{-1}(m_\alpha) \in \bar{\beta}\} \tag{5.13}$$

où $m_\alpha \in \mathbb{R}^2$ désigne le centre de la maille dyadique α. Observons que la plupart du temps, le point $\mathcal{A}^{-1}(m_\alpha)$ est à l'intérieur d'une maille de M et à nouveau, le max est inutile.

Notre algorithme de prédiction de maillage pourra donc s'intituler

Algorithme 5.5 (prédiction par transport des résolutions locales)

ou bien, de façon équivalente,

Algorithme 5.5 (découpage dyadique guidé par les niveaux rétrogrades)

- *Poser* $\Lambda_{\ell_0} := \mathbb{D}_{\ell_0}(\mathbb{R}^2)$.
- *Pour* $\ell \geq \ell_0$, *calculer*
  $\Lambda_{\ell+1} := \Lambda_\ell \cup \{\alpha \in \mathcal{F}(\beta) : \beta \in \Lambda_\ell \text{ et } \ell^*_{M,\mathcal{A}}(\beta) > \ell(\beta)\}$
  jusqu'à ce que $\Lambda_{L+1} = \Lambda_L$, *et prendre* $\tilde{M} = \partial\Lambda_L$.
- *Définir* $\boldsymbol{T}[\mathcal{A}]M$ *comme le plus petit raffinement gradué (algorithme* (2.8)*) de* $\tilde{M}$.

Remarque 5.6 *De la même façon que dans les algorithmes précédents, seules les cellules de niveau ℓ sont susceptibles d'être raffinées lors du raffinement conditionnel de l'arbre Λ_ℓ, dans la deuxième étape de l'algorithme.*

En d'autres termes, le "critère de qualité" utilisé ici pour savoir si une maille doit ou non être subdivisée est que son niveau doit toujours être plus grand que celui de la cellule à laquelle appartient l'antécédent de son centre par $\mathcal{A}$. Toutefois, pour que

le maillage dyadique $\boldsymbol{T}[\mathcal{A}]M$ construit de cette façon vérifie bien $\ell^*_{M,\mathcal{A},}(\alpha) \leq \ell(\alpha)$ sur chacune de ses mailles, il est nécessaire que cette propriété (clairement vraie pour la partition $\tilde{M}$) soit préservée par le raffinement, autrement dit qu'elle vérifie une forme de monotonie semblable à (2.48). Le lemme suivant montre qu'on peut établir cette monotonie sous la condition de stabilité (HT.3).

Lemme 5.7 *Soit $M \in \mathcal{M}(\mathbb{R}^2)$, $\mathcal{A}$ un difféomorphisme de $\mathbb{R}^2$ et β une cellule dyadique vérifiant $\ell^*_{M,\mathcal{A}}(\beta) \leq \ell(\beta)$. Si $\mathcal{A}^{-1}$ est lipschitzien de constante $c_{\mathcal{A}} < 2$, alors*

$$\ell^*_{M,\mathcal{A}}(\alpha) \leq \ell(\alpha) \quad \textit{dès lors que} \quad \alpha \subset \beta. \tag{5.14}$$

Preuve. Comme il n'y a rien à montrer lorsque α et β sont de même niveau, on peut considérer que $\ell(\beta) < \ell(\alpha)$. On a alors

$$\big|m_\alpha - m_\beta\big| \leq (2^{-\ell(\beta)} - 2^{-\ell(\alpha)})/2, \tag{5.15}$$

l'hypothèse sur $\mathcal{A}$ nous permettant d'écrire

$$\big|\mathcal{A}^{-1}(m_\alpha) - \mathcal{A}^{-1}(m_\beta)\big| \leq c_{\mathcal{A}}\big|m_\alpha - m_\beta\big| < 2^{-\ell(\beta)} - 2^{-\ell(\alpha)}. \tag{5.16}$$

Soient maintenant α^* et β^* deux cellules de M contenant respectivement $\mathcal{A}^{-1}(\alpha)$ et $\mathcal{A}^{-1}(\beta)$. On aura démontré le lemme si on établit que $\ell(\alpha^*) \leq \ell(\alpha)$. L'hypothèse $\ell^*_{M,\mathcal{A}}(\beta) \leq \ell(\beta)$ entraînant $\ell(\beta^*) \leq \ell(\beta)$, on voit qu'on peut se restreindre au cas où $\ell(\beta^*) < \ell(\alpha^*)$. On utilise alors le fait que M est gradué en invoquant la proposition 2.11, selon laquelle

$$2^{-\ell(\beta^*)} - 2^{-\ell(\alpha^*)+1} \leq \big|\mathcal{A}^{-1}(m_\alpha) - \mathcal{A}^{-1}(m_\beta)\big|. \tag{5.17}$$

En utilisant (5.16) et le fait que $\ell(\beta^*) \leq \ell(\beta)$, ceci entraîne que

$$2^{-\ell(\beta^*)} - 2^{-\ell(\alpha^*)+1} < 2^{-\ell(\beta)} - 2^{-\ell(\alpha)} \leq 2^{-\ell(\beta^*)} - 2^{-\ell(\alpha)}, \tag{5.18}$$

d'où l'on déduit $\ell(\alpha^*) < 1 + \ell(\alpha)$. L'inégalité souhaitée $\ell(\alpha^*) \leq \ell(\alpha)$ provient alors du fait que les niveaux sont des entiers. □

5.2.2 Propriétés des maillages transportés

Les propriétés essentielles de notre algorithme $\boldsymbol{T}[\mathcal{A}]$ sont exprimées par les deux théorèmes compémentaires suivants :

- le premier exprime le fait que l'ordre de complexité des maillages est préservé par l'algorithme, ce qui est un résultat de complexité essentiel.
- le deuxième garantit que les domaines d'influence (5.12) associés aux cellules de $\boldsymbol{T}[\mathcal{A}]M$ ont un cardinal uniformément borné. Associé à la stabilité (HT.5) du transport vis-à-vis des indicateurs d'erreur a priori, cet argument sera crucial dans l'analyse d'erreur.

Théorème 5.1 (complexité des maillages transportés) *Si $\mathcal{A}$ et $\mathcal{A}^{-1}$ sont tous les deux lipschitziens, alors il existe une constante C pour laquelle*

$$\#\big(\boldsymbol{T}[\mathcal{A}]M\big) \leq C\#(M), \qquad M \in \mathcal{M}(\mathbb{R}^2). \tag{5.19}$$

Preuve. En utilisant la proposition 2.9, on voit qu'on peut se ramener à la partition dyadique *non graduée* $\tilde{\boldsymbol{T}}[\mathcal{A}]M$, résultat du découpage adaptatif brut (non raffiné :-) dans l'algorithme 5.5. Et on observera que les cellules α de $\tilde{\boldsymbol{T}}[\mathcal{A}]M$, si elles ne sont pas de niveau égal à ℓ_0, ont une parente $\tilde{\alpha} := \mathcal{P}(\alpha)$ qui a dû être découpée au cours de la construction de $\tilde{\boldsymbol{T}}[\mathcal{A}]M$. Il existe donc une cellule $\tilde{\beta} \in M$ contenant $\mathcal{A}^{-1}(m_{\tilde{\alpha}})$ qui vérifie

$$\ell(\tilde{\beta}) \geq \ell(\tilde{\alpha}) - 1 = \ell(\alpha). \tag{5.20}$$

Pour compter les cellules de $\tilde{\boldsymbol{T}}[\mathcal{A}]M$, on peut alors introduire pour chaque cellule $\beta \in M$ un domaine d'influence "dans le futur"

$$\mathcal{J}_{M,\mathcal{A}}(\beta) := \{\alpha \in \tilde{\boldsymbol{T}}[\mathcal{A}]M : m_\alpha \in \mathcal{A}(\beta)\} \tag{5.21}$$

contenant les cellules de $\tilde{\boldsymbol{T}}[\mathcal{A}]M$ dont le centre est contenu dans $\mathcal{A}(\beta)$. Comme $\tilde{\boldsymbol{T}}[\mathcal{A}]M = \cup_{\beta\in M}\mathcal{J}_{M,\mathcal{A}}(\beta)$, on aura prouvé (5.19) si l'on trouve une constante C pour laquelle

$$\sup_{\beta\in M} \#\big(\mathcal{J}_{M,\mathcal{A}}(\beta)\big) \leq C. \tag{5.22}$$

D'après la construction de $\tilde{\boldsymbol{T}}[\mathcal{A}]M$, on voit sans peine que les cellules α de $\mathcal{J}_{M,\mathcal{A}}(\beta)$ vérifient $\ell(\beta) \leq \ell(\alpha)$. On peut en réalité écrire une inégalité inverse, à savoir

$$\ell(\alpha) \leq \ell(\beta) + 1 \text{ pour toute cellule } \alpha \text{ de } \mathcal{J}_{M,\mathcal{A}}(\beta). \tag{5.23}$$

Pour le voir, souvenons nous qu'il existe dans le maillage M une cellule $\tilde{\beta}$ dont le niveau est supérieur à celui de α, et qui contient le point $\mathcal{A}^{-1}(m_{\tilde{\alpha}})$. Comme $\tilde{\alpha}$ désigne ici la cellule parente de α, le fait que $\mathcal{A}^{-1}$ soit lipschitzien de constante $c_3 < 2$, d'après (HT.3), entraîne

$$\big|\mathcal{A}^{-1}(m_{\tilde{\alpha}}) - \mathcal{A}^{-1}(m_\alpha)\big| \leq c_3|m_{\tilde{\alpha}} - m_\alpha| \leq c_3 2^{-\ell(\alpha)}. \tag{5.24}$$

D'autre part, m_α est par hypothèse un point de β. On peut donc appliquer la proposition 2.11, selon laquelle les niveaux de β et $\tilde{\beta}$ ne sauraient être arbitrairement distants. Plus précisément, on peut supposer que $\ell(\beta) \leq \ell(\tilde{\beta})-1$, car dans le cas contraire (5.23) se déduit immédiatement de (5.20). La proposition 2.11 nous dit alors que

$$2^{-\ell(\beta)} - 2^{-\ell(\tilde{\beta})+1} \leq \big|\mathcal{A}^{-1}(m_{\tilde{\alpha}}) - \mathcal{A}^{-1}(m_\alpha)\big| \leq c_3 2^{-\ell(\alpha)}, \tag{5.25}$$

d'où l'on déduit $2^{\ell(\beta)} \leq c_3 2^{-\ell(\alpha)} < 2 \cdot 2^{-\ell(\alpha)}$ d'après (5.20), et finalement (5.23). Pour établir (5.22), il suffit alors de montrer que les cellules de $\mathcal{J}_{M,\mathcal{A}}(\beta)$ intersectent toutes une boule de rayon $C2^{-\ell(\beta)}$. Ce qui est le cas, puisqu'on a

$$\big|m_\alpha - \mathcal{A}(m_\beta)\big| \leq C|\mathcal{A}^{-1}(m_\alpha) - m_\beta| \leq C2^{-\ell(\beta)} \tag{5.26}$$

en utilisant le fait que $\mathcal{A}$ est lipschitzien. □

Théorème 5.2 (stabilité du transport de maillages dyadiques) *Si $\mathcal{A}$ vérifie l'hypothèse du lemme 5.7, alors il existe deux constantes c_6 et c_7 dépendant uniquement de $c_\mathcal{A}$ pour lesquelles*

$$\sup_{\beta\in\mathcal{I}_{M,\mathcal{A}}(\alpha)} \ell(\beta) \leq \ell(\alpha) + c_6, \qquad M \in \mathcal{M}(\mathbb{R}^2), \qquad \alpha \in \boldsymbol{T}[\mathcal{A}]M, \tag{5.27}$$

$$\#\big(\mathcal{I}_{M,\mathcal{A}}(\alpha)\big) \leq c_7, \qquad M \in \mathcal{M}(\mathbb{R}^2), \qquad \alpha \in \boldsymbol{T}[\mathcal{A}]M \tag{5.28}$$

Preuve. Soit M un maillage dyadique et α une cellule fixée de $\boldsymbol{T}[\mathcal{A}]M$. Si $\eta \in M$ contient $\mathcal{A}^{-1}(m_\alpha)$, on a par construction

$$\ell(\eta) \leq \ell(\alpha). \tag{5.29}$$

Si d'autre part $\beta \in \mathcal{I}_{M,\mathcal{A}}(\alpha)$, il existe un point $m \in \alpha$, tel que $\mathcal{A}^{-1}(m) \in \beta$. L'hypothèse faite sur $\mathcal{A}$ entraîne alors que

$$\left|\mathcal{A}^{-1}(m_\alpha) - \mathcal{A}^{-1}(m)\right| \leq c_{\mathcal{A}} |m_\alpha - m| \leq c_{\mathcal{A}}\, 2^{-\ell(\alpha)-1}, \tag{5.30}$$

ce qui signifie que β intersecte la boule B^*_α de rayon $r^*_\alpha = c_{\mathcal{A}}\, 2^{-\ell(\alpha)-1} \leq c_{\mathcal{A}}\, 2^{-\ell(\eta)-1}$(il s'agit d'une boule "carrée", pour la distance ℓ^∞ de $\mathbb{R}^2$) dont le centre $\mathcal{A}^{-1}(m_\alpha)$ appartient à η. Comme β et η font toutes deux parties du même maillage gradué M, on peut voir sur la figure 5.2 que l'hypothèse $c_{\mathcal{A}}\, 2^{-\ell(\eta)-1} < 2^{-\ell(\eta)}$ va entraîner d'une part, que le niveau de β ne peut pas être arbitrairement supérieur à celui de η, et d'autre part que le nombre de cellules β est effectivement borné de façon indépendante de α. Plus précisément, la proposition 2.11 s'applique : si $\delta := \ell(\beta) - \ell(\eta) \geq 1$, alors

$$2^{-\ell(\eta)}[1 - 2^{1-\delta}] \leq \left|\mathcal{A}^{-1}(m_\alpha) - \mathcal{A}^{-1}(m)\right|. \tag{5.31}$$

On en déduit en utilisant (5.30) que $1 - 2^{1-\delta} \leq c_{\mathcal{A}}/2$, ce qui implique δ est inférieur à $2 - \ln_2(2 - c_{\mathcal{A}})$. En utilisant (5.29), ceci implique déjà (5.27) avec $c_6 = 2 - \ln_2(2 - c_{\mathcal{A}})$. Revenant au nombre maximal de cellules β, on peut alors vérifier qu'il est inférieur à $c_7 = 9\, 2^{2\delta}$ (en considérant (i) que ce nombre est maximal dans le cas où la graduation de M est saturée autour de η, comme sur la figure 2.3, (ii) qu'en raison de $r^*_\alpha < 2^{-\ell(\eta)}$, B^*_α intersecte au plus 9 "super-cellules" de niveau $\ell(\eta)$, et (iii) que chacune de ces "super-cellules" contient au plus $2^{2\delta}$ cellules de niveau $\ell(\eta) + \delta$). □

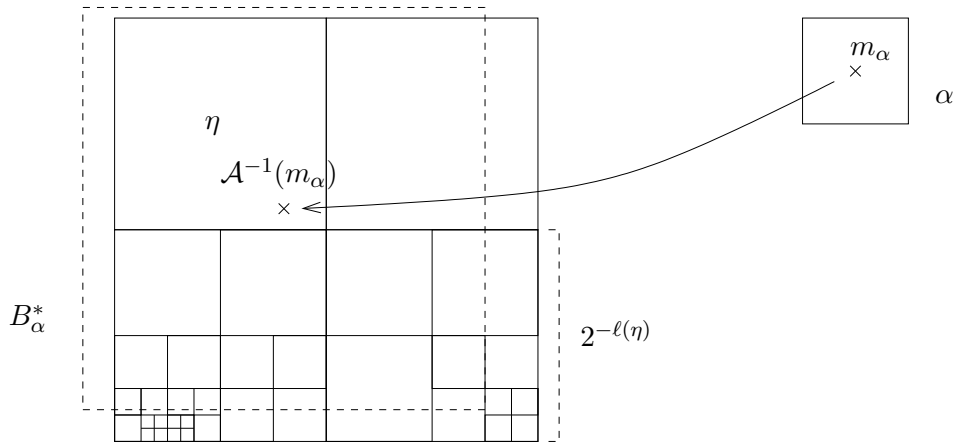


FIG. 5.2 – les cellules de $\mathcal{I}_{M,\mathcal{A}}(\alpha)$ intersectent toutes la "boule" B^*_α.

5.3 Le schéma adaptatif de prédiction et correction

On désigne par $(M^n, f^n) \in \mathcal{M}(\mathbb{R}^2) \times V_{M^n}$ la solution numérique adaptative au pas de temps n. Le schéma est alors déterminé par deux paramètres scalaires : un pas de temps constant Δt, et une *tolérance* $\varepsilon > 0$ qui représente à chaque pas de temps l'erreur L^∞ autorisée pour les approximations adaptatives.

5.3.1 Description formelle du schéma

La *solution numérique initiale* est obtenue en projetant la donnée initiale exacte f_0 sur le plus petit maillage dyadique qui lui est ε-adapté au sens de l'indicateur d'erreur $\mathcal{E}$. Dans l'esprit des algorithmes d'ε-adaptation proposés dans les sections 2.2.3 et 3.3.2, on utilisera donc ici l'algorithme $\boldsymbol{A}_\varepsilon$ sous la forme 6.3 du chapitre suivant, pour poser

$$M^0 := \boldsymbol{A}_\varepsilon(f_0), \qquad f^0 := P_{M^0} f_0. \tag{5.32}$$

Au pas de temps n, en désignant par $\mathcal{A}^n := \mathcal{A}[f^n]$ le déplacement correspondant au transport approché (5.4), on calcule alors (M^{n+1}, f^{n+1}) par une première étape où le maillage de calcul est *prédit* de façon très simple pour pouvoir appliquer le schéma de transport $\mathcal{T}$, et une deuxième étape où ce maillage est *corrigé*. En d'autres termes, notre schéma s'écrit

$$\tilde{M}^{n+1} := \boldsymbol{T}[\mathcal{A}^n] M^n, \qquad \tilde{f}^{n+1} := P_{\tilde{M}^{n+1}} \mathcal{T} f^n, \tag{5.33a}$$

$$M^{n+1} := \boldsymbol{A}_\varepsilon(\tilde{f}^{n+1}), \qquad f^{n+1} := P_{M^{n+1}} \tilde{f}^{n+1}. \tag{5.33b}$$

En particulier, on observera qu'*il n'est jamais nécessaire de réappliquer l'opérateur de transport $\mathcal{T}$ sur un maillage corrigé*. La raison étant que notre algorithme de prédiction $\boldsymbol{T}[\mathcal{A}^n]$, comme on va bientôt le voir et malgré sa relative simplicité, génère des maillages suffisamment précis.

5.3.2 Analyse intuitive

A chaque pas de temps, la solution numérique est donc obtenue par un schéma de transport-projection

$$f^{n+1} = P_{M^{n+1}} P_{\tilde{M}^{n+1}} \mathcal{T} f^n, \tag{5.34}$$

tandis que les maillages dyadiques sur lesquels repose la discrétisation adaptative sont obtenus par un schéma de transport-correction. On peut en donner une vision synthétique en introduisant la fonctionnelle

$$\bar{\mathcal{E}}(g, M) := \sup_{\alpha \in M} \mathcal{E}(g, \alpha) \tag{5.35}$$

qui mesure l'adéquation globale entre g et M au sens où l'on a

$$\|g - P_M g\|_{L^\infty} \leq C \bar{\mathcal{E}}(g, M) \tag{5.36}$$

d'après (5.7) et (2.22). L'algorithme $\boldsymbol{A}_\varepsilon$ permet alors d'écrire

$$\bar{\mathcal{E}}(g, \boldsymbol{A}_\varepsilon(g)) \leq \varepsilon, \tag{5.37}$$

et sous réserve que le transport approché (5.4) vérifie certaines des hypothèses énoncées dans la section 5.1, on peut montrer que $\boldsymbol{T}[\mathcal{A}^n]$ préserve l'ordre d'adéquation au cours du transport, *i.e.*

$$\bar{\mathcal{E}}(\mathcal{T} f^n, \boldsymbol{T}[\mathcal{A}^n] M^n) \leq C \bar{\mathcal{E}}(f^n, M^n) \tag{5.38}$$

mais ne l'empêche en général pas d'augmenter, la constante C pouvant être supérieure à 1. Il faut donc voir le pas (5.33a) comme une méthode simple pour faire évoluer le maillage entre deux instants t_n et t_{n+1}, *avec une précision acceptable sur quelques*

itérations, mais insuffisante à long terme (en particulier, la résolution maximale n'augmente pas lorsqu'on passe de M^n à $\boldsymbol{T}[\mathcal{A}^n]M^n$, alors que la régularité des solutions peut se dégrader fortement au cours de la simulation). Le pas (5.33b) permet alors de ramener l'adéquation du maillage en-dessous du paramètre de tolérance ε.

Quant à la complexité de notre schéma, son analyse tient en deux arguments : le premier est exprimé par le théorème 5.1, selon laquelle notre algorithme de transport dyadique $\mathbf{T}[\mathcal{A}^n]$ préserve l'ordre de complexité des maillages mais ne l'empêche pas d'augmenter. A nouveau, on dispose donc *d'une stabilité acceptable sur quelques itérations, mais insuffisante à long terme.* En particulier, la taille des maillages transportés peut augmenter de façon exponentielle avec le nombre d'itérations, sans aucun rapport avec le pas de temps ou avec les propriétés d'approximation de la solution exacte f. Le deuxième argument consiste alors à écrire un résultat semblable au théorème 3.2 où l'on estime la complexité de $\boldsymbol{A}_\varepsilon(g)$ à partir de la régularité de g, et à estimer cette régularité au cours du schéma.

5.4 Propriétés du schéma adaptatif semi-lagrangien

5.4.1 Analyse d'erreur

Notre principal résultat est le suivant.

Théorème 5.3 (Précision des solutions) *Si le transport approché* (5.4) *satisfait les hypothèses (HT.1)-(HT.5) énoncées dans la section 5.1, l'erreur numérique* $e_n := \|f(t_n) - f^n\|_{L^\infty}$ *associée au schéma* (5.32)-(5.33b) *vérifie*

$$e_N \leq C(\Delta t^{\,\sigma} + \varepsilon \Delta t^{\,-1} + \varepsilon). \tag{5.39}$$

D'autre part, la constante C *de* (5.39) *ne dépend que des* c_i*, du temps final* $T = N\Delta t$ *et de la solution initiale* f_0*.*

Preuve. D'après les estimations (5.36) et (5.37), l'erreur initiale $e_0 := \|(I - P_{M^0})f_0\|_{L^\infty}$ se majore sans peine par $C\varepsilon$. La forme (5.34) du schéma nous permet alors de décomposer l'erreur e_{n+1} en

$$e_{n+1} \leq \|f(t_{n+1}) - \mathcal{T} f(t_n)\|_{L^\infty} + \|\mathcal{T} f(t_n) - \mathcal{T} f^n\|_{L^\infty} + \|(\tilde{P} - I)\mathcal{T} f^n\|_{L^\infty} + \|(P - I)\tilde{f}^{n+1}\|_{L^\infty}, \tag{5.40}$$

où P et $\tilde{P}$ désignent respectivement $P_{M^{n+1}}$ et $P_{\tilde{M}^{n+1}}$. Le premier terme est en quelque sorte "résolu", puisqu'il s'agit de l'erreur en temps majorée par l'hypothèse (HT.1). Le deuxieme terme se majore comme dans la remarque 5.3, en exploitant conjointement la régularité lipschitzienne (5.2) de la solution exacte et la stabilité (HT.4) du transport $\mathcal{T}$ relativement aux perturbations de densité. On a ainsi en posant $\tilde{\mathcal{T}}^n := \mathcal{T}[f(t_n)]$ et $\mathcal{T}^n := \mathcal{T}[f^n]$ - tous deux linéaires,

$$\|\mathcal{T} f(t_n) - \mathcal{T} f^n\|_{L^\infty} \leq \|\mathcal{T}_n\big(f(t_n) - f^n\big)\|_{L^\infty} + \|\big(\tilde{\mathcal{T}}^n - \mathcal{T}^n\big) f(t_n)\|_{L^\infty} \leq (1 + C\Delta t\,)e_n,$$

en utilisant (HT.4), (5.2), et le fait que le transport diminue toujours la norme L^∞.

Restent donc les erreurs de projection $\|(P - I)\tilde{f}^{n+1}\|_{L^\infty}$ et $\|(\tilde{P} - I)\mathcal{T} f^n\|_{L^\infty}$. D'après (5.35)-(5.36), elles seront contrôlées par ε si l'on est capable de garantir respectivement

A1. que $\mathcal{E}(\tilde{f}^{n+1}, \alpha)$ est au plus de l'ordre de ε pour toute cellule $\alpha \in M^{n+1}$,

A2. que $\mathcal{E}(\mathcal{T}f^n, \alpha)$ est au plus de l'ordre de ε pour toute cellule $\alpha \in \tilde{M}^{n+1}$.

A nouveau, on insistera sur la différence entre les assertions A1 et A2. La première correspond à un problème d'adaptation *statique*, dont la solution naturelle consiste à prendre $M^{n+1} = \boldsymbol{A}_\varepsilon(\tilde{f}^{n+1})$. La seconde, quant à elle, correspond à un problème *dynamique* pour lequel la solution précédente ne s'applique plus, dans la mesure où ne souhaitant pas calculer les valeurs de la solution transportée $\mathcal{T}f^n$ sur un maillage uniformément fin, on ne dispose pas en pratique de ses valeurs à ce stade du calcul. On peut alors montrer que la construction de $\tilde{M}^{n+1}$ par l'algorithme de prédiction $\boldsymbol{T}[\mathcal{A}^n]$ fournit une réponse satisfaisante. La démonstration se fait en trois étapes.

1. Adéquation du maillage de départ. En utilisant (5.11), (5.32) et (3.51), on voit que (M^0, f^0) vérifie

$$\bar{\mathcal{E}}(f^0, M^0) \leq C\bar{\mathcal{E}}(f_0, M^0) \leq C\varepsilon, \tag{5.41}$$

et pour $n \geq 1$, on obtient de la même façon grâce à (5.33b) que

$$\bar{\mathcal{E}}(f^n, M^n) \leq C\bar{\mathcal{E}}(\tilde{f}^n, M^n) \leq C\varepsilon. \tag{5.42}$$

On en déduit que les maillages M_n sont toujours bien adaptés aux solutions numériques f_n, au sens où l'on a

$$\bar{\mathcal{E}}(f^n, M^n) \leq C\varepsilon \tag{5.43}$$

avec une constante absolue.

2. Evolution des indicateurs d'erreurs. Comme $\mathcal{T}f^n = f^n \circ \big(\mathcal{A}^n\big)^{-1}$, l'hypothèse de compatiblité (HT.5) nous permet d'écrire

$$\mathcal{E}(\mathcal{T}f^n, \alpha) \leq C \sum_{\beta \in \mathcal{I}_{M^n, \mathcal{A}^n}(\alpha)} \mathcal{E}(f^n, \beta) \tag{5.44}$$

pour toute cellule dyadique α (on ne fait ici que recopier une hypothèse, mais c'est en réalité une étape importante de l'analyse du schéma).

3. Contrôle des domaines d'influence. La dernière étape consiste à exploiter la propriéte (5.28) des maillages transportés. En vertu de l'hypothèse (HT.3), en effet, le théorème 5.2 s'applique et toute cellule α appartenant à $\tilde{M}^{n+1} = \boldsymbol{T}[\mathcal{A}^n]M^n$ vérifie

$$\mathcal{E}(\mathcal{T}f^n, \alpha) \leq C\#\big(\mathcal{I}_{M^n, \mathcal{A}^n}(\alpha)\big)\bar{\mathcal{E}}(f^n, M^n) \leq C\varepsilon \tag{5.45}$$

pour tout entier n, en utilisant (5.44), (5.28) et (5.43). On en déduit donc bien

$$\bar{\mathcal{E}}(\mathcal{T}f^n, \tilde{M}^{n+1}) \leq C\varepsilon, \tag{5.46}$$

ce qui répond au problème A2.

Au cours de cette analyse d'erreur, on a successivement vérifié que (i) l'erreur de discrétisation en temps $\|f(t_{n+1}) - \mathcal{T}f(t^n)\|_{L^\infty}$ était de l'ordre de $\Delta t^{\,\sigma+1}$, (ii) l'erreur de couplage $\|\mathcal{T}f(t_n) - \mathcal{T}f^n\|_{L^\infty}$ était inférieure à $(1 + C\Delta t\,)e_n$, (iii) l'erreur de projection "statique" $\|(P_{M^{n+1}} - I)\tilde{f}^{n+1}\|_{L^\infty}$ était de l'ordre de ε, (iv) ainsi que l'erreur

de projection "dynamique" $\|(P_{\tilde{M}^{n+1}} - I)\mathcal{T} f^n\|_{L^\infty}$. Compte tenu de la décomposition (5.40), on en déduit

$$e_{n+1} \leq C(\Delta t^{\,\sigma+1} + \varepsilon) + (1 + C\Delta t\,)e_n, \tag{5.47}$$

et finalement l'estimation (5.39) en utilisant le lemme de Gronwall 1.4. □

5.4.2 Analyse de complexité

Théorème 5.4 (Complexité des maillages) *Si les solutions numériques intermédiaires $\tilde{f}^n$ font décroître les indicateurs d'erreur d'une façon semblable à* (3.52)*, autrement dit s'il existe un $s > 0$ pour lequel on a*

$$\mathcal{E}(\tilde{f}^n, \alpha) \leq c_s|\alpha|^s, \qquad n \geq 1 \tag{5.48}$$

alors les maillages prédits $\tilde{M}^n$ et corrigés M^n générés par le schéma (5.32)-(5.33b) *vérifient*

$$\max\{\mathrm{card}_{\ell_0}(M^n), \mathrm{card}_{\ell_0}(\tilde{M}^n)\} \leq C\mathcal{E}_{\ell_0}(\tilde{f}^n)|\log[\varepsilon/c_s]|\varepsilon^{-1}, \tag{5.49}$$

où la quantité $\mathcal{E}_{\ell_0}(\tilde{f}^n)$ définie en (5.9) *est l'analogue de la courbure totale sur le domaine $\mathbb{R}^2$ tout entier.*

Remarque 5.8 *On rapprochera bien évidemment l'hypothèse* (5.48) *de l'hypothèse* (3.52) *faite dans le théorème 3.2 et commentée dans la remarque 3.14.*

Preuve. Ce théorème peut être démontré de la même façon que le théorème 3.2 en ce qui concerne la complexité des maillages corrigés M^n, et en utilisant le théorème 5.1 en ce qui concerne les maillages prédits $\tilde{M}^n$. □

Chapitre 6

Application au système de Vlasov-Poisson

Dans ce chapitre, on décrit en détails comment appliquer notre schéma adaptatif semi-lagrangien au système uni-dimensionnel de Vlasov-Poisson (4.2)-(4.3) périodique en x lorsque les données initiales sont dans $W^{1,\infty} \cap BC(\mathbb{T} \times \mathbb{R})$ avec $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. Dans un premier temps, on présente une discrétisation en temps classique (voir [19]) de ce système basée sur une décomposition des trajectoires dans les directions alternées x et v, décomposition dont l'avantage sera la grande simplicité analytique des trajectoires approchées. Après avoir établi une nouvelle estimation d'erreur pour cette discrétisation en temps, on précise la forme du schéma de calcul adaptatif. Par rapport à la présentation qu'on en a fait au chapitre précédent, on observera que notre schéma a gagné en technicité. On peut voir deux raisons à cela. La première est que la décomposition du transport sur les directions alternées s'accompagne nécessairement d'une décomposition similaire du schéma. Mais la principale source de difficultés est que plusieurs propriétés essentielles du transport numérique reposent sur une régularité du champ électrique qui ne pourra à son tour être établie que par une majoration du support des solutions numériques.

Les "hypothèses de travail" (HT.1)-(HT.5) énoncées au chapitre précédent pourront alors être établies, et au terme d'une analyse faisant intervenir plusieurs estimations des quantités numériques, on sera en mesure d'établir une borne a priori de l'erreur L^∞ réalisée par notre schéma. On discutera également de son caractère optimal, en donnant plusieurs arguments susceptibles de mener à un résultat de complexité.

6.1 Décomposition du transport sur les directions alternées

$\Omega := \mathbb{T} \times \mathbb{R}$ désignera l'espace des phases dans tout ce chapitre.

6.1.1 L'opérateur de transport approché de Cheng et Knorr

Dans un article [19] publié en 1976, Cheng et Knorr proposent d'approcher le transport exact $\mathcal{T}^{\text{exact}} : f(t_n) \to f(t_{n+1})$ associé aux solutions de l'équation de Vlasov-Poisson (4.2) par une combinaison $\mathcal{T} = \mathcal{T}_x \mathcal{T}_v \mathcal{T}_x$ de transports dirigés suivant x et v,

dite de "time-splitting". Ces opérateurs sont définis par

$$\mathcal{T}_x g = g \circ \big(\mathcal{A}_x\big)^{-1} \quad \text{et} \quad \mathcal{T}_v g = g \circ \big(\mathcal{A}_v[g]\big)^{-1} \tag{6.1}$$

(dans tout ce chapitre, g désignera une fonction définie sur Ω, donc 1-périodique en x), avec

$$\mathcal{A}_x \colon (x,v) \to (x + v\Delta t\,/2, v) \tag{6.2}$$

et

$$\mathcal{A}_v[g] \colon (x,v) \to (x, v + \Delta t\, \tilde{E}[g](x)). \tag{6.3}$$

Ici, $\tilde{E}[g]$ désigne le champ électrique (4.34) associé à la densité g, défini par

$$\tilde{E}[g](x) = \int K(x,y) \left(\int g(y,v)\,\mathrm{d}v - 1 \right) \mathrm{d}y \tag{6.4}$$

et qui vérifie en particulier

$$(\tilde{E}[g])'(x) = \int g(x,v)\,\mathrm{d}v - 1. \tag{6.5}$$

Le déplacement approché $\mathcal{A}[g] \colon (x,v) \to (\tilde{x},\tilde{v})$ correspondant à l'opérateur de transport

$$\mathcal{T} = \mathcal{T}_x \mathcal{T}_v \mathcal{T}_x \colon g \to g \circ \big(\mathcal{A}[g]\big)^{-1} \tag{6.6}$$

vérifie alors

$$\left\{ \begin{array}{l} \tilde{x} = x + v\Delta t \, + (\Delta t^{\,2}/2)\tilde{E}[\mathcal{T}_x g](x + v\Delta t\,/2) \\ \tilde{v} = v + \Delta t\, \tilde{E}[\mathcal{T}_x g](x + v\Delta t\,/2). \end{array} \right. \tag{6.7}$$

Si g est telle que $\iint g(x,v)\,\mathrm{d}x\,\mathrm{d}v = 1$, on peut vérifier que $\tilde{E}[g]$ est une fonction 1-périodique, de sorte que $\mathcal{A}[g]$ est bien un difféomorphisme de Ω dans lui-même. Son inverse $\big(\mathcal{A}[g]\big)^{-1} \colon (\tilde{x},\tilde{v}) \to (x,v)$ vérifie alors

$$\left\{ \begin{array}{l} x = \tilde{x} - \tilde{v}\Delta t \, + (\Delta t^{\,2}/2)\tilde{E}[\mathcal{T}_x g](\tilde{x} - \tilde{v}\Delta t\,/2) \\ v = \tilde{v} - \Delta t\, \tilde{E}[\mathcal{T}_x g](\tilde{x} - \tilde{v}\Delta t\,/2). \end{array} \right. \tag{6.8}$$

6.1.2 Erreur de discrétisation en temps

Lorsque la solution exacte est lipschitzienne, on peut montrer que cette approximation est globalement précise à l'ordre 2 (ce qui correspond à l'hypothèse (HT.1) avec $\sigma = 2$ dans la section 5.1).

Proposition 6.1 *Si f_0 appartient à $W^{1,\infty}(\Omega)$, alors le transport approché défini par* (6.6)- (6.8) *vérifie*

$$\|f(t_{n+1}) - \mathcal{T} f(t_n)\|_{L^\infty} \leq C\Delta t^{\,3},$$

avec une constante qui ne dépend que de l'instant final $T = N\Delta t$ et de la solution initiale f_0.

Remarque 6.2 *A notre connaissance, ce résultat est original.*

Preuve. Fixons l'entier n et un point $(\tilde{x},\tilde{v})$ dans Ω. Comme f est constante le long des trajectoires caractéristiques, nous pouvons désigner par $(X,V)(s) = (X,V)(s;t_{n+1},\tilde{x},\tilde{v})$ la solution de (4.16) vérifiant

$$(X(t_{n+1}),V(t_{n+1})) = (\tilde{x},\tilde{v})$$

et voir que la solution exacte au temps t_{n+1} satisfait

$$f(t_{n+1},\tilde{x},\tilde{v}) = f(t_n,X(t_n),V(t_n)),$$

tandis que la solution approchée vaut

$$\mathcal{T}f(t_n)(\tilde{x},\tilde{v}) = f(t_n,X^n,V^n)$$

si l'on note

$$(X^n,V^n) \text{ avec } \begin{cases} X^n = \tilde{x} - \tilde{v}\Delta t + (\Delta t^2/2)\tilde{E}[\mathcal{T}_x f(t_n)](\tilde{x} - \tilde{v}\Delta t/2) \\ X^n = \tilde{v} - \Delta t\,\tilde{E}[\mathcal{T}_x f(t_n)](\tilde{x} - \tilde{v}\Delta t/2). \end{cases}$$

On observera que sur le pas de temps $[t_n,t_{n+1}]$, $(X(t_n),V(t_n))$ et (X^n,V^n) sont respectivement points de départs des trajectoires *exacte* et *approchée* qui arrivent en $(X(t_{n+1}),V(t_{n+1})) = (\tilde{x},\tilde{v})$.

D'après la proposition 4.3, $f(t_n)$ est lipschitzienne dès que f_0 l'est, aussi a-t-on

$$\begin{aligned} \|f(t_{n+1}) - \mathcal{T}f(t_n))\|_{L^\infty(\Omega)} &\le |f(t_n)|_{W^{1,\infty}(\Omega)} \left(|X(t_n) - X^n| + |V(t_n) - V^n|\right) \\ &\le C(T) \max\left(|X(t_n) - X^n|, |V(t_n) - V^n|\right). \end{aligned}$$

Il nous faut donc montrer que

$$\max\left(|X^n - X(t_n)|, |V^n - V(t_n)|\right) \le C(T)\Delta t^3. \tag{6.9}$$

En utilisant conjointement les estimations de régularité données par la proposition 4.3 et l'équation différentielle (4.16) vérifiée par les trajectoires caractéristiques, on voit que le champ exact suivant les trajectoires caractéristiques $E_X(t) := E(t,X(t))$ vérifie (en désignant par concision la norme $\|\cdot\|_{L^\infty([0,T],L^\infty([0,1]))}$ par $\|\cdot\|_\infty$)

$$\|E_X\|_{L^\infty([0,T])} \le C(T) \tag{6.10}$$

$$\begin{aligned} \|\dot{E}_X\|_{L^\infty([0,T])} &\le \|\partial_t E\|_\infty + \|V\|_{L^\infty([0,T])}\|\partial_x E\|_\infty \\ &\le \|\partial_t E\|_\infty + \sup_{0\le t\le T} \Sigma_v(f(t))\|\partial_x E\|_\infty \le C(T) \end{aligned} \tag{6.11}$$

$$\begin{aligned} \|\ddot{E}_X\|_{L^\infty([0,T])} &\le \|\partial_{tt}^2 E\|_\infty + 2\|V\|_{L^\infty([0,T])}\|\partial_{tx}^2 E\|_\infty \\ &\quad + \|V^2\|_{L^\infty([0,T])}\|\partial_{xx}^2 E\|_\infty + \|E\|_\infty\|\partial_x E\|_\infty \le C(T). \end{aligned} \tag{6.12}$$

On décompose alors

$$\begin{aligned} X^n - X(t_n) = X(t_{n+1}) - X(t_n) - \tilde{v}\Delta t + \Delta t^2/2\ \tilde{E}[\mathcal{T}_x f(t_n)](\tilde{x} - \tilde{v}\Delta t/2) \\ = \mathcal{E}_1 + \frac{\Delta t^2}{2}(\mathcal{E}_2 + \mathcal{E}_3), \end{aligned}$$

où l'on a défini

$$\begin{aligned}\mathcal{E}_1 &:= X(t_{n+1}) - X(t_n) - \tilde{v}\Delta t \, + \Delta t^{\,2}/2 \; E_X(t_{n+1/2}) \\ \mathcal{E}_2 &:= E(t_{n+1/2}, \tilde{x} - \tilde{v}\Delta t\,/2) - E_X(t_{n+1/2}) \\ \mathcal{E}_3 &:= \tilde{E}[\mathcal{T}_x f(t_n)](\tilde{x} - \tilde{v}\Delta t\,/2) - E(t_{n+1/2}, \tilde{x} - \tilde{v}\Delta t\,/2)\end{aligned}$$

et $t_{n+1/2} = (n+1/2)\Delta t$. De même, on peut écrire

$$V^n - V(t_n) = V(t_{n+1}) - V(t_n) - \Delta t \;\; \tilde{E}[\mathcal{T}_x f(t_n)](\tilde{x} - \tilde{v}\Delta t\,/2) = \mathcal{E}_4 - \Delta t\,(\mathcal{E}_2 + \mathcal{E}_3)$$

avec

$$\mathcal{E}_4 := V(t_{n+1}) - V(t_n) - \Delta t\, E_X(t_{n+1/2}).$$

Il nous reste alors à montrer

$$|\mathcal{E}_1| \le C(T)\Delta t^{\,3}, \quad |\mathcal{E}_2| \le C(T)\Delta t^{\,2}, \quad |\mathcal{E}_3| \le C(T)\Delta t^{\,2} \quad \text{et} \quad |\mathcal{E}_4| \le C(T)\Delta t^{\,3}.$$

En s'inspirant de l'équation (4.16) vérifiée par les trajectoires exactes, on calcule que

$$\begin{aligned}\mathcal{E}_1 &= \int_{t_n}^{t_{n+1}} (V(t) - \tilde{v})\,\mathrm{d}t + \Delta t^{\,2}/2 \; E_X(t_{n+1/2}) \\ &= \int_{t_n}^{t_{n+1}} (V(t) - V(t_{n+1}))\,\mathrm{d}t + \int_{t_n}^{t_{n+1}} \int_{t}^{t_{n+1}} E_X(t_{n+1/2})\,\mathrm{d}s\,\mathrm{d}t \\ &= \int_{t_n}^{t_{n+1}} \int_{t}^{t_{n+1}} (-E_X(s) + E_X(t_{n+1/2}))\,\mathrm{d}s\,\mathrm{d}t.\end{aligned}$$

Grâce à (6.11), on voit ensuite que

$$|E_X(t_{n+1/2}) - E_X(s)| \le |\dot{E}_X|_{L^\infty([0,T])} |t_{n+1/2} - s| \le C(T)\Delta t\,,$$

d'où l'on déduit $|\mathcal{E}_1| \le C(T)\Delta t^{\,3}$. Pour le deuxième terme, on calcule

$$\begin{aligned}|\mathcal{E}_2| &= \left|E(t_{n+1/2}, \tilde{x} - \tilde{v}\Delta t\,/2) - E(t_{n+1/2}, X(t_{n+1/2}))\right| \\ &\le \|\partial_x E(t_{n+1/2})\|_{L^\infty([0,1])} |X(t_{n+1/2}) - \tilde{x} + \tilde{v}\Delta t\,/2| \\ &\le C(T) |X(t_{n+1/2}) - X(t_{n+1}) + \tilde{v}\Delta t\,/2| \\ &\le C(T) \int_{t_{n+1/2}}^{t_{n+1}} |\tilde{v} - V(t)|\,\mathrm{d}t \\ &\le C(T) \int_{t_{n+1/2}}^{t_{n+1}} |V(t_{n+1}) - V(t)|\,\mathrm{d}t \\ &\le C(T) \|E_X\|_{L^\infty([0,T])} \Delta t^{\,2} \le C(T)\Delta t^{\,2},\end{aligned}$$

cette dernière inégalité provenant de (6.10). Considérant maintenant le troisième terme, on peut déduire des expressions respectives (4.34) et (6.4) des champs E et $\tilde{E}$ que

$$\begin{aligned}|\mathcal{E}_3| &= \left| \int K(\tilde{x} - \tilde{v}\Delta t\,/2, y) \left[\int \left[[\mathcal{T}_x f(t_n)](y,v) - f(t_{n+1/2}, y, v) \right] \mathrm{d}v \right] \mathrm{d}y \right| \\ &= \left| \int K(\tilde{x} - \tilde{v}\Delta t\,/2, y) \left[\int \left[f(t_n, y - v\Delta t\,/2, v) - f(t_{n+1/2}, y, v) \right] \mathrm{d}v \right] \mathrm{d}y \right| \\ &\le \int \left| \int [f(t_n, y - v\Delta t\,/2, v) - f(t_{n+1/2}, y, v)]\,\mathrm{d}v \right| \mathrm{d}y = \int \left| \int A(y,v)\,\mathrm{d}v \right| \mathrm{d}y\end{aligned} \tag{6.13}$$

avec $A(y,v) := f(t_n, y - v\Delta t\,/2, v) - f(t_{n+1/2}, y, v)$, l'inégalité ci-dessus venant du fait que $\|K\|_{L^\infty([0,1]^2)} \leq 1$. En posant $t_s := t_n + \Delta t\,/2 - s$ et $y_s(v) := y - vs$, on observe alors que

$$\begin{aligned} A(y,v) &= \int_0^{\Delta t/2} \frac{\mathrm{d}}{\mathrm{d}s} f(t_s, y_s(v), v)\ \mathrm{d}s \\ &= \int_0^{\Delta t/2} -(\partial_t f + v\partial_x f)(t_s, y_s(v), v)\,\mathrm{d}s \\ &= \int_0^{\Delta t/2} B(s,y,v)\,\mathrm{d}s, \end{aligned}$$

où $B(s,y,v) = -E(t_s, y_s(v))\partial_v f(t_s, y_s(v), v)$, cette dernière égalité venant de l'équation de Vlasov (4.2). Plutôt que de majorer directement B par une constante, ce qui nous donnerait $|\mathcal{E}_3| \leq C(T)\Delta t$ qui est insuffisant, on intègre par parties

$$\begin{aligned} \int s\partial_x E(t_s, y_s(v)) f(t_s, y_s(v), v)\,\mathrm{d}v &= -\int \frac{\mathrm{d}}{\mathrm{d}v}[\partial_x E(t_s, y_s(v))] f(t_s, y_s(v), v)\,\mathrm{d}v \\ &= \int E(t_s, y_s(v)) \frac{\mathrm{d}}{\mathrm{d}v}[f(t_s, y_s(v), v)]\,\mathrm{d}v \\ &= \int E(t_s, y_s(v))\,[(-s\partial_x f + \partial_v f)(t_s, y_s(v), v)]\ \mathrm{d}v, \end{aligned}$$

d'où l'on déduit

$$\int B(s,y,v)\,\mathrm{d}v = -s\int [\partial_x E(t_s, y_s(v)) f(t_s, y_s(v), v) + E(t_s, y_s(v))\partial_x f(t_s, y_s(v), v)]\,\mathrm{d}v.$$

D'après la proposition 4.3, on obtient alors

$$\begin{aligned} \left|\int A(y,v)\,\mathrm{d}v\right| = \left|\int\int_0^{\Delta t/2} B(s,y,v)\,\mathrm{d}s\,\mathrm{d}v\right| &\leq \Delta t \sup_{|s|\leq\Delta t}\left|\int B(s,y,v)\,\mathrm{d}v\right| \\ &\leq \Delta t^{\,2} \sup_{0\leq t\leq T} \Sigma_v(f(t))C(T) \leq C(T)\Delta t^{\,2}, \end{aligned}$$

ce qui implique $|\mathcal{E}_3| \leq C(T)\Delta t^{\,2}$ grâce à (6.13). Pour le quatrième terme, enfin, on calcule

$$\begin{aligned} \mathcal{E}_4 &= V(t_{n+1}) - V(t_{n+1/2}) + V(t_{n+1/2}) - V(t_n) - \Delta t\, E_X(t_{n+1/2}) \\ &= \int_0^{\Delta t/2} [E_X(t_{n+1} - t) + E_X(t_n + t)]\,\mathrm{d}t - \Delta t\, E_X(t_{n+1/2}) \\ &= \int_0^{\Delta t/2} [E_X(t_{n+1} - t) - E_X(t_{n+1/2}) + E_X(t_n + t) - E_X(t_{n+1/2})]\,\mathrm{d}t \\ &= \int_0^{\Delta t/2}\int_t^{\Delta t/2} [\dot{E}_X(t_{n+1} - s) - \dot{E}_X(t_n + s)]\,\mathrm{d}s\,\mathrm{d}t, \end{aligned}$$

ce qui nous donne

$$|\mathcal{E}_4| \leq \Delta t^{\,3}\|\ddot{E}_X\|_{L^\infty([0,T])} \leq C(T)\Delta t^{\,3}$$

en utilisant (6.12), et termine la preuve. □

6.1.3 Décomposition formelle du schéma

Le calcul des trajectoires faisant appel aux solutions "intermédiaires" $\mathcal{T}_x f_n$, on a suivi le choix fait par Besse [10] et Sonnendrücker, Roche, Bertrand et Ghizzo [59] de décomposer notre schéma suivant les directions alternées x et v. En respectant l'esprit de (5.32)-(5.33), un schéma adaptatif basé sur le transport 6.6 prendra donc la forme suivante (qu'on précisera plus bas) :

$$M_1^n := \boldsymbol{T}[\mathcal{A}_x]M^n \qquad \text{et} \qquad f_1^n := P_{M_1^n}\mathcal{T}_x f^n \tag{6.14a}$$
$$M_2^n := \boldsymbol{T}[\mathcal{A}_v(f_1^n)]M_1^n \qquad \text{et} \qquad f_2^n := P_{M_2^n}\mathcal{T}_v f_1^n \tag{6.14b}$$
$$M_3^n := \boldsymbol{T}[\mathcal{A}_x]M_2^n \qquad \text{et} \qquad f_3^n := P_{M_3^n}\mathcal{T}_x f_2^n \tag{6.14c}$$
$$M^{n+1} := \boldsymbol{A}_\varepsilon(f_3^n) \qquad \text{et} \qquad f^{n+1} := P_{M^{n+1}} f_3^n. \tag{6.14d}$$

6.2 Description complète du schéma adaptatif

6.2.1 Un indicateur d'erreur basé sur la courbure totale

On rappelle que notre schéma utilise comme discrétisations adaptatives les éléments finis $\mathcal{P}^1$ multi-échelles introduits au chapitre 2, et que ces derniers sont associés aux maillages dyadiques de la classe $\mathcal{M}(\mathbb{R}^2)$. Pour manipuler ces maillages dyadiques, on a montré dans la section 3.3.1 qu'une fonctionnelle basée sur la courbure totale $|g|_{BC(\alpha)}$ était un bon indicateur a priori pour les erreurs locales d'interpolation affine par morceaux. Toutefois, et notamment pour pouvoir établir une propriété de stabilité du transport approché vis-à-vis des indicateurs d'erreur (correspondant à l'hypothèse (HT.5) dans la section 5.1), on utilisera la fonctionnelle

$$\mathcal{E}(g,\alpha) := |g|_{BC(\alpha)} + \Delta t\, 2^{-2\ell(\alpha)}|g|_{W^{1,\infty}(\alpha)} \tag{6.15}$$

à la place de $|g|_{BC(\alpha)}$, et on redéfinit l'algorithme d'adaptation de maillages dyadiques de la façon suivante.

Algorithme 6.3 (maillage ε-adaptaté au sens de la fonctionnelle $\mathcal{E}$)

- *Poser* $\Lambda_{\ell_0} := \mathbb{D}_{\ell_0}(\mathbb{R}^2)$.
- *Pour* $\ell \geq \ell_0$, *calculer*
  $$\Lambda_{\ell+1} := \Lambda_\ell \cup \{\alpha \in \mathcal{F}(\beta) : \beta \in \Lambda_\ell \text{ et } \mathcal{E}(g,\beta) > \varepsilon\}$$
  jusqu'à ce que $\Lambda_{L+1} = \Lambda_L$, *et prendre* $\tilde{M} = \partial\Lambda_L$.
- *Définir* $\boldsymbol{A}_\varepsilon(g)$ *comme le plus petit raffinement gradué (algorithme* (2.8)*) de* $\tilde{M}$.

Remarque 6.4 *De la même façon que dans les algorithmes précédents, seules les cellules de niveau ℓ sont susceptibles d'être raffinées lors du raffinement conditionnel de l'arbre Λ_ℓ, dans la deuxième étape de l'algorithme.*

6.2.2 Forme exacte du schéma

Pour écrire notre schéma sous sa forme définitive, on introduit les objets suivants (dont la définition est donnée plus bas)

– une approximation E^n du champ électrique $\tilde{E}[\mathcal{T}_x f^n]$, construite de sorte que le transport

$$\mathcal{T}_v^n \colon g \to g \circ (\mathcal{A}_v^n)^{-1} \text{ associé à } \mathcal{A}_v^n(x,v) := (x, v + \Delta t\, E^n(x)) \tag{6.16}$$

préserve la structure périodique et affine par morceaux des fonctions.
– un opérateur T_n de troncature en vitesse.

On pose alors

$$M^0 := \boldsymbol{A}_\varepsilon(f_0) \qquad \text{et} \qquad f^0 := P_{M^0} f_0, \tag{6.17a}$$

puis pour $n \geq 0$,

$$M_1^n := \boldsymbol{T}[\mathcal{A}_x] M^n \qquad \text{et} \qquad f_1^n := P_{M_1^n} \mathcal{T}_x f^n \tag{6.17b}$$
$$M_2^n := \boldsymbol{T}[\mathcal{A}_v^n] M_1^n \qquad \text{et} \qquad f_2^n := P_{M_2^n} \mathsf{T}_{n+1} \mathcal{T}_v^n f_1^n \tag{6.17c}$$
$$M_3^n := \boldsymbol{A}_\varepsilon(f_2^n) \qquad \text{et} \qquad f_3^n := P_{M_3^n} f_2^n \tag{6.17d}$$
$$M^{n+1} := \boldsymbol{T}[\mathcal{A}_x] M_3^n \qquad \text{et} \qquad f^{n+1} := P_{M^{n+1}} \mathcal{T}_x f_3^n. \tag{6.17e}$$

On pourra enfin désigner par $\mathbb{S}_{\Delta t,\varepsilon}$ le schéma associé

$$\mathbb{S}_{\Delta t,\varepsilon} \colon f^n \to f^{n+1} = P_{M^{n+1}} \mathcal{T}_x P_{M_3^n} P_{M_2^n} \mathsf{T}_{n+1} \mathcal{T}_v^n P_{M_1^n} \mathcal{T}_x f^n. \tag{6.18}$$

6.2.3 E^n : périodisation $\mathcal{P}^1$ du champ électrique

Dans notre schéma, on utilise le flot $\mathcal{A}_v^n$ donné par (6.16) au lieu de $\mathcal{A}_v[f_1^n]$ donné par (6.3), car ce dernier présente deux inconvénients. Le premier est dû au fait que $\tilde{E}[f_1^n]$ n'étant pas affine par morceaux, l'opérateur $\mathcal{T}_v[f_1^n]$ ne préserve pas la structure affine par morceaux des fonctions. Le deuxième inconvénient est une conséquence du manque de conservativité des projections $\mathcal{P}^1$. En effet, comme $\iint f_1^n \,\mathrm{d}x\,\mathrm{d}v$ est a priori différent de $\iint f_0 \,\mathrm{d}x\,\mathrm{d}v$, on ne saurait garantir que la condition (4.31) est satisfaite par f_1^n, par conséquent ni $\tilde{E}[f_1^n]$, ni $\mathcal{A}_v[f_1^n]$ n'ont de raison d'être périodiques. On commence donc par périodiser le champ $\tilde{E}[f_1^n]$ en posant

$$\tilde{E}^n(x) := \tilde{E}[f_1^n](\{x\}) + \{x\}\big(\tilde{E}[f_1^n](0) - \tilde{E}[f_1^n](1)\big), \tag{6.19}$$

$\{x\}$ désignant la *partie fractionnaire* de x, et on définit ensuite E^n comme l'interpolation affine par morceaux de $\tilde{E}^n$ sur les noeuds

$$\Gamma_x^n := \Gamma_x(M_1^n), \tag{6.20}$$

où l'ensemble

$$\Gamma_x(M) := \bigcup\nolimits_{\alpha \in M} \partial(\alpha_x) \tag{6.21}$$

n'est rien d'autre que la projection des noeuds du maillage dyadique $M \in \mathcal{M}(\Omega)$ sur l'axe des x.

En utilisant l'expression (6.4), on peut alors voir qu'une régularité lipschitzienne des solutions numériques entraînera une régularité $W^{2,\infty}$ du champ associé $\tilde{E}[f_1^n]$. Quant au champ E^n, il vérifie

$$\|E^n\|_{L^\infty} \leq \|\tilde{E}^n\|_{L^\infty} \leq \|\tilde{E}[f_1^n]\|_{L^\infty} + \left|\tilde{E}[f_1^n](1) - \tilde{E}[f_1^n](0)\right|, \tag{6.22}$$

et dans la mesure où la dérivée de E^n entre deux noeuds consécutifs x_i et x_{i+1} de Γ_x^n interpole celle de $\tilde{E}^n$ en un point $y_i \in]x_i, x_{i+1}[$, on a clairement

$$|E^n|_{W^{1,\infty}} \leq |\tilde{E}^n|_{W^{1,\infty}} \leq 2|\tilde{E}[f_1^n]|_{W^{1,\infty}}, \tag{6.23}$$

tandis que le saut $[(E^n)']_{x_i} := (E^n)'(x_i^+) - (E^n)'(x_i^-)$ est majoré par

$$\left|[(E^n)']_{x_i}\right| \leq \left|(\tilde{E}^n)'(y_i) - (\tilde{E}^n)'(y_{i-1})\right| \leq \int_{x_{i-1}}^{x_{i+1}} |(\tilde{E}^n)''| \leq \int_{x_{i-1}}^{x_{i+1}} |(\tilde{E}[f_1^n])''|. \tag{6.24}$$

6.2.4 T_n : troncature douce en vitesse

On a montré dans la section 4.2.2 que le support en vitesse

$$\Sigma_v(g) := \sup\{|v| : \exists x,\, g(x,v) > 0\}, \tag{6.25}$$

des solutions exactes était contrôlé : plus précisément, on a montré dans la proposition 4.3 qu'il augmentait au plus comme $\Sigma_v(f(t_n)) \leq \Sigma_v(f_0) + 2t_n$. C'est une propriété importante, car elle permet d'établir de nombreuses estimations pour les solutions exactes. En ce qui concerne les solutions numériques, une propriété similaire serait la bienvenue, car elle nous permettrait

- de savoir a priori quelle est la taille maximale du domaine de calcul pour une simulation donnée (une fois connus la donnée initiale f_0 et le temps final $T = N\Delta t$).
- de contrôler le manque de conservativité du schéma par ses propriétés d'approximation dans L^∞, ce qui se révèlera fort utile lors de l'analyse d'erreur.

A chaque projection, pourtant, on voit que le support des solutions numériques peut a priori augmenter d'une largeur de maille $2^{-\ell_0}$, de sorte qu'il n'y a aucune raison pour que la quantité $\sup_{n\Delta t \leq T} \Sigma_v(f^n)$ soit majorée indédendamment du pas de temps Δt .

On va surmonter cette difficulté en tronquant les solutions au delà d'une certaine distance à l'axe des x, tout en prenant soin de ne pas trop dégrader leur régularité. On pose ainsi

$$\mathsf{T}_n g(x,v) = \begin{cases} 0 & \text{si} \quad |v| > \tilde{\Sigma}_v^n + 2^{-\ell_0} \\ g(x, -\tilde{\Sigma}_v^n)(\tilde{\Sigma}_v^n + 2^{-\ell_0} + v)2^{\ell_0} & \text{si} \quad -\tilde{\Sigma}_v^n - 2^{-\ell_0} \leq v < -\tilde{\Sigma}_v^n \\ g(x,v) & \text{si} \quad -\tilde{\Sigma}_v^n \leq v \leq \tilde{\Sigma}_v^n \\ g(x, \tilde{\Sigma}_v^n)(\tilde{\Sigma}_v^n + 2^{-\ell_0} - v)2^{\ell_0} & \text{si} \quad \tilde{\Sigma}_v^n < v \leq \tilde{\Sigma}_v^n + 2^{-\ell_0}, \end{cases} \tag{6.26}$$

où

$$\tilde{\Sigma}_v^n := 2^{-\ell_0}(\lceil 2^{\ell_0}\Sigma_v^n \rceil + 1) \geq \Sigma_v^n + 2^{-\ell_0} \tag{6.27}$$

désigne le plus petit multiple entier de $2^{-\ell_0}$ supérieur à $\Sigma_v^n + 2^{-\ell_0}$, avec

$$\Sigma_v^n := \Sigma_v(f_0) + 2n\Delta t\,. \tag{6.28}$$

Grâce à cet opérateur, on peut facilement établir la proposition suivante :

Proposition 6.5 *Les solutions numériques f_i^n du schéma* (6.17) *ont toutes leur support borné en vitesse indépendamment de Δt par*

$$\sup_{i,n\leq N} \Sigma_v(f_i^n) \leq \tilde{\Sigma}_v := \Sigma_v(f_0) + 2N\Delta t\, + 7. \tag{6.29}$$

Preuve. On peut en effet écrire $\Sigma_v(\mathsf{T}_n g) \le \tilde\Sigma_v^n + 2^{-\ell_0} \le \Sigma_v^n + 3\cdot 2^{-\ell_0}$ pour une densité g quelconque, et $\Sigma_v(P_M g) \le \Sigma_v(g) + 2^{-\ell_0}$ pour tout maillage dyadique. On en déduit

$$\begin{aligned}\Sigma_v(f^{n+1}) &\le \Sigma_v(\mathcal{T}_x f_3^n) + 2^{-\ell_0} \le \Sigma_v(f_3^n) + 2^{-\ell_0} \le \Sigma_v(f_2^n) + 2\cdot 2^{-\ell_0}\\ &\le \Sigma_v(\mathsf{T}_{n+1}\mathcal{T}_v^n f_1^n) + 3\cdot 2^{-\ell_0} \le \Sigma_v^{n+1} + 6\cdot 2^{-\ell_0},\end{aligned} \tag{6.30}$$

d'où finalement $\Sigma_v(f_1^n) \le \Sigma_v(\mathcal{T}_x f^n) + 2^{-\ell_0} \le \Sigma_v(f^n) + 2^{-\ell_0} \le \Sigma_v^n + 7\cdot 2^{-\ell_0} \le \tilde\Sigma_v$. □

Remarque 6.6 *De $\Sigma_v(f(t_n)) \le \Sigma_v^n \le \tilde\Sigma_v^n$, on déduit que $f(t_n)$ s'annule en dehors de $\Omega_n := \mathbb{T}\times[-\Sigma_v^n, \Sigma_v^n]$ et de $\tilde\Omega_n := \mathbb{T}\times[-\tilde\Sigma_v^n, \tilde\Sigma_v^n]$.*

Remarque 6.7 *$\tilde\Sigma_v^n$ étant multiple de $2^{-\ell_0}$, les cellules dyadiques seront toutes soit dans $\tilde\Omega_n$, soit dans son complémentaire $(\tilde\Omega_n)^{\mathsf{c}}$ (bords compris). Ceci étant toujours vrai pour les triangles de $\mathcal{K}(M)$, les projections associées P_M vérifient*

$$\|P_M g\|_{L^\infty(\tilde\Omega_n)} \le \|g\|_{L^\infty(\tilde\Omega_n)} \quad\text{et}\quad \|P_M g\|_{L^\infty((\tilde\Omega_n)^{\mathsf{c}})} \le \|g\|_{L^\infty((\tilde\Omega_n)^{\mathsf{c}})}. \tag{6.31}$$

On aura de plus pour tout entier n, tout maillage M et toute fonction continue g :

$$(P_M g)_{|\tilde\Omega_n} = P_M(g_{|\tilde\Omega_n}),\quad (\mathcal{T}_x g)_{|\tilde\Omega_n} = \mathcal{T}_x(g_{|\tilde\Omega_n}) \quad\text{et}\quad (\mathsf{T}_n g)_{|\tilde\Omega_n} = g_{|\tilde\Omega_n}. \tag{6.32}$$

6.3 Propriétés principales du schéma

6.3.1 Estimation d'erreur

Théorème 6.1 *Si la solution initiale f_0 appartient à $W^{1,\infty}\cap BC(\Omega)$, si elle est à support compact et si elle vérifie la condition de charge nulle* (4.31)*, alors pour tout $T = N\Delta t$, il existe une constante $C = C(T, f_0)$ pour laquelle l'erreur associée au schéma* (6.17) *vérifie*

$$\|f(T) - f^N\|_{L^\infty} \le C\,(\Delta t^{\,2} + \varepsilon/\Delta t\,) \tag{6.33}$$

dès lors que

$$\varepsilon^{1/2} \le \Delta t \le 2^{-\ell_0}[8(\tilde\Sigma_v\|f_0\|_{L^\infty}+1)]^{-1}, \tag{6.34}$$

$\tilde\Sigma_v$ étant défini par (6.29)*.*

La preuve de ce théorème fait l'objet des sections 6.4 et 6.5.

Remarque 6.8 *Dans la mesure où $2^{-\ell_0}$ représente le pas de discrétisation* le plus large *des maillages dyadiques, l'hypothèse* (6.34) *n'est en aucun cas une condition de type Courant-Friedrichs-Lewy. On observera d'ailleurs qu'une telle condition serait difficilement compatible avec la présence de mailles arbitrairement fines, le pas de temps Δt étant ici non seulement constant, mais aussi* uniforme*.*

Remarque 6.9 *Une fonction à support compact vérifiant* (4.30)*, on voit que f_0 satisfait les hypothèses du théorème de Cooper et Klimas. En particulier, il existe une solution exacte f périodique en x, et on pourra utiliser les estimations de régularité données par la proposition 4.3.*

Remarque 6.10 *On peut remplacer la condition $\varepsilon \le \Delta t^{\,2}$ ci-dessus par $\varepsilon \le C\Delta t^{\,2}$ pour une constante C fixée. En équilibrant* (6.33)*, on trouve d'autre part $\varepsilon = \Delta t^{\,3}$ et*

$$\|f(N\Delta t\,) - f^N\|_{L^\infty} \le C\,\Delta t^{\,2} = C\,\varepsilon^{2/3}.$$

6.3.2 Vers un résultat de complexité

Comme on l'a déjà souligné, le théorème précédent n'est pas entièrement satisfaisant, car il ne nous dit pas quel sera le coût de calcul associé aux solutions numériques f^n pour un choix particulier des paramètres Δt et ε. On peut désigner par

$$\mathsf{N} = \mathsf{N}_{N,\Delta t\,,\varepsilon} = \sup_{n\leq N} \{\#(M^n), \#(M_1^n), \#(M_2^n), \#(M_3^n)\} \tag{6.35}$$

la taille maximale des maillages produits par notre schéma, et il est assez facile de vérifier que la complexité algorithmique d'une itération (6.17) est de l'ordre de $\mathsf{N} \log \mathsf{N}$. Pour estimer la taille de nos maillages, on peut alors utiliser

- d'une part le théorème 5.1 selon lequel la prédiction $\mathbf{T}[\mathcal{A}^n]$ des maillages préserve leur complexité à une constante multiplicative près,
- et d'autre part un raisonnement similaire à celui employé dans la preuve du théorème 3.2, selon lequel N sera de l'ordre de ε^{-1} à condition de (i) majorer la quantité $|f^n|_{BC(\mathbb{R}^2)} + \Delta t\,|f^n|_{W^{1,\infty}(\mathbb{R}^2)}$ par une constante indépendante de Δt, et (ii) garantir que les solutions numériques font décroître les indicateurs d'erreur comme
$$|f^n|_{BC(\alpha)} + \Delta t\,|\alpha||f^n|_{W^{1,\infty}(\alpha)} \leq c_s|\alpha|^s$$
pour une constante c_s indépendante de Δt.

Au terme de notre analyse d'erreur (c'est-à-dire à la fin de ce chapitre), on verra que la semi-norme lipschitzienne des solutions numériques peut être majorée de façon indépendante du pas de temps Δt. Comme les solutions numériques sont de plus toujours affines par morceaux sur une triangulation graduée, le point (ii) ne présente pas de difficulté particulière (voir en particulier la proposition 3.16). En revanche, il est moins évident d'établir une borne a priori sur la courbure totale $|f^n|_{BC(\mathbb{R}^2)}$, car les interpolations $\mathcal{P}^1$ n'ont a priori aucune raison de diminuer la courbure totale des fonctions affines par morceaux. Dans la section 3.2.3, on a montré comment construire pour une triangulation donnée $\mathcal{K}$ une courbure $|\cdot|_{\mathcal{K},\star}$ équivalente à $|\cdot|_{BC}$ qui est diminuée par les projections $\mathcal{P}^1$ associées aux raffinements uniformes de $\mathcal{K}$. Malheureusement, on n'est pas encore parvenu à étendre cette construction à des raffinements non-uniformes.

Ceci nous montre tout de même que si l'on était capable d'établir une borne sur la courbure totale des solutions f^n, on obtiendrait un résultat de complexité proche de

$$\|f(N\Delta t\,) - f^N\|_{L^\infty} \leq C\varepsilon^{2/3} \leq C\mathsf{N}^{-2/3}. \tag{6.36}$$

On peut comparer cette conjecture aux résultats de Besse [10], qui montre que lorsque la solution est dans $W^{2,\infty}(\mathbb{R}^2)$, l'erreur associée à un schéma similaire au nôtre mais n'utilisant qu'un maillage uniforme de pas h décroît comme

$$\|f(N\Delta t\,) - f_h^N\|_{L^\infty} \leq C(\Delta t^{\,2} + h^2/\Delta t\,), \tag{6.37}$$

autrement dit comme $\Delta t^{\,2} \sim h^{4/3}$ après équilibrage des termes d'erreur. La taille des maillages valant alors $\mathsf{N} = Ch^{-2}$, on en déduit le taux de convergence suivant

$$\|f(N\Delta t\,) - f_h^N\|_{L^\infty} \leq C\mathsf{N}^{-2/3}. \tag{6.38}$$

A nouveau, on sera sensible à la différence entre cette estimation et (6.36), qui serait valable pour des solutions initiales beaucoup moins régulières. On retrouverait ainsi le même type de résultats qu'avec les estimations (1.20) et (1.31), où la supériorité de l'approche adaptative se traduisait non pas par un ordre supérieur de la précision, mais par une régularité inférieure permettant d'atteindre un même ordre de précision. Ainsi, tout se passe comme si l'approximation affine par morceaux des solutions du système de Vlasov-Poisson par une méthode de transport projection basée sur le schéma de transport (6.6)-(6.7) était capable, *dans des conditions idéales de fonctionnement*, de converger dans L^∞ avec un ordre de $-2/3$. L'adaptativité nous permettrait alors d'élargir ces "conditions idéales de fonctionnement"à l'espace $W^{1,\infty} \cap BC(\mathbb{R}^2)$, bien plus grand que $W^{2,\infty}(\mathbb{R}^2)$. Le diagramme de la figure 6.1 illustre la distance entre ces différents espaces. Rappelons que, de la même façon que dans la figure 1 de l'introduction, ce diagramme indique en ordonnées le nombre de dérivées et en abscisses l'inverse $1/p$ de l'exposant de l'espace L^p dans lequel ces dérivées sont mesurées.

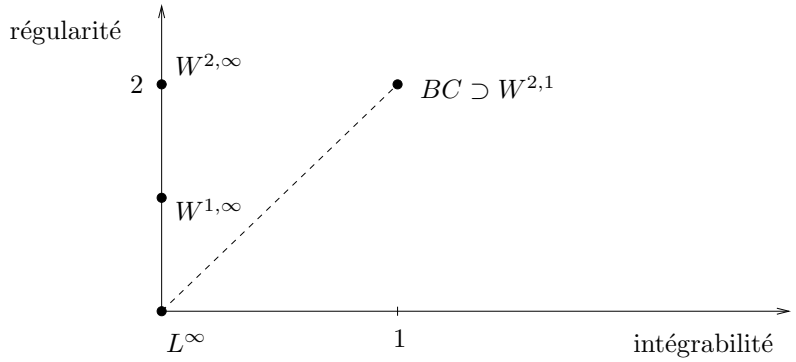


FIG. 6.1 – illustration des régularités présentes dans les estimations (6.36) et (6.38).

6.4 Propriétés des transports approchés $\mathcal{T}_x$ et $\mathcal{T}_v^n$

Pour prouver le théorème 6.1, on va suivre les mêmes étapes que dans l'analyse d'erreur de la section 5.4.1. On pourra ainsi s'assurer que les différents opérateurs de transport introduits possèdent bien des propriétés semblables à celles qu'on a énoncé dans la section 5.1. En ce qui concerne l'hypothèse de précision en temps (HT.1), on l'a établie par la proposition 6.1, et les hypothèses (HT.2) et (HT.3) seront facilement vérifiées dans la section 6.4.1 ci-dessous. En revanche, la stabilité (HT.5) des opérateurs de transport $\mathcal{T}_x$ et $\mathcal{T}_v^n$ vis-à-vis de l'indicateur d'erreur $\mathcal{E}$ fera l'objet d'une étude plus conséquente dans la section 6.4.2, et sera soumise à une borne $W^{2,\infty}$ sur le champ électrique $\tilde{E}^n$. L'hypothèse (HT.4) pourra alors être établie dans la section 6.4.3.

6.4.1 Régularité des déplacements directs et rétrogrades

En utilisant (6.23), on voit que les déplacements (6.2) et (6.16) satisfont respectivement

$$\left|\mathcal{A}_x(x,v) - \mathcal{A}_x(x',v')\right| \leq (1 + \Delta t\,/2)|(x,v) - (x',v')| \tag{6.39}$$

et

$$\left|\mathcal{A}_v^n(x,v) - \mathcal{A}_v^n(x',v')\right| \leq (1 + 2\Delta t\,|\tilde{E}^n|_{W^{1,\infty}})|(x,v) - (x',v')|. \tag{6.40}$$

On aura donc une propriété semblable à (HT.2) si l'on arrive à borner les semi-normes $|\tilde{E}^n|_{W^{1,\infty}}$ par une constante dépendant uniquement des "paramètres du problème" T et f_0. Parce que la troncature en vitesse nous permet de contrôler le support en vitesse des solutions numériques (voir la proposition 6.5), cela ne pose pas de difficulté particulière : d'après (6.5), et de façon similaire à (4.43), on a en effet

$$|\tilde{E}[f_1^n]|_{W^{1,\infty}} = \left\| \int f_1^n(\cdot,v)\,\mathrm{d}v - 1 \right\|_{L^\infty} \leq \Sigma_v(f_1^n)\|f_1^n\|_{L^\infty} + 1. \tag{6.41}$$

Comme les différents opérateurs du schéma numérique (6.18) font clairement décroître la norme L^∞, on obtient

$$\|f^{n+1}\|_{L^\infty} \leq \|f_3^n\|_{L^\infty} \leq \cdots \leq \|f^n\|_{L^\infty} \leq \cdots \leq \|f_0\|_{L^\infty}, \tag{6.42}$$

d'où l'on déduit (en utilisant (6.29) et (6.23))

$$\sup_{n\leq N} |\tilde{E}^n|_{W^{1,\infty}} \leq \tilde{\Sigma}_v\|f_0\|_{L^\infty} + 1. \tag{6.43}$$

Comme ℓ_0 est toujours positif, l'hypothèse (6.34) nous garantit alors que

$$\max\left(1 + \Delta t\,/2, 1 + 2\Delta t\,|\tilde{E}^n|_{W^{1,\infty}}\right) \leq 5/4, \quad \text{pour tout } n \leq N. \tag{6.44}$$

Les déplacements $\mathcal{A}_x$ et $\mathcal{A}_v^n$ sont donc bien lipschitziens, et comme leurs inverses $\left(\mathcal{A}_x\right)^{-1}$ et $\left(\mathcal{A}_v^n\right)^{-1}$ vérifient également (6.39) et (6.40), on déduit de (6.44) qu'ils possèdent bien une propriété de type (HT.3). Plus précisément, le théorème 5.2 s'applique, et l'on a pour tout maillage dyadique M et tout entier $n \leq N$

$$\#\big(\mathcal{I}_{M,\mathcal{A}}(\alpha)\big) \leq C \text{ pour tout } \alpha \in \boldsymbol{T}[\mathcal{A}]M, \tag{6.45}$$

$$\text{et } \ell(\beta) - \ell(\alpha) \leq C \text{ pour tout } \alpha \in \boldsymbol{T}[\mathcal{A}]M \text{ et tout } \beta \in \mathcal{I}_{M,\mathcal{A}}(\alpha) \tag{6.46}$$

avec une constante C absolue, que $\mathcal{A}$ désigne $\mathcal{A}_x$ ou $\mathcal{A}_v^n$.

6.4.2 Régularité des densités numériques transportées

Le résultat qu'on sera en mesure de montrer à la fin de cette section - et qui est l'analogue de (HT.5) - est le suivant.

Proposition 6.11 *Soit g une fonction affine par morceaux. Sous les hypothèses du théorème 6.1, la fonctionnelle $\mathcal{E}$ définie par* (6.15) *vérifie (avec des constantes dépendant uniquement de f_0 et $N\Delta t$)*

$$\mathcal{E}(\mathcal{T}_x g, \alpha) \leq C \sum\nolimits_{\beta\in\mathcal{I}_{M,\mathcal{A}_x}(\alpha)} \mathcal{E}(g,\beta) \tag{6.47}$$

pour les cellules α du maillage $\boldsymbol{T}[\mathcal{A}_x]M$ "transporté" à partir d'un maillage dyadique M arbitraire, et

$$\mathcal{E}(\mathcal{T}_v^n g, \alpha) \leq C(1 + |\tilde{E}^n|_{W^{2,\infty}}) \sum\nolimits_{\beta\in\mathcal{I}_{M,\mathcal{A}_v^n}(\alpha)} \mathcal{E}(g,\beta) \tag{6.48}$$

pour les cellules α du maillage $\boldsymbol{T}[\mathcal{A}_v^n]M_1^n$ "transporté" à partir du maillage M_1^n utilisé pour définir le déplacement $\mathcal{A}_v^n$.

Etudions alors la façon dont les opérateurs de transport $\mathcal{T}_x$ et $\mathcal{T}_v^n$ font évoluer la régularité des densités numériques, en rappelant que $\mathcal{T}_v^n$ est défini en (6.16) à partir de E^n, lui-même construit dans la section 6.2.3 comme l'interpolation du champ électrique périodisé $\tilde{E}^n$ sur la grille d'interpolation Γ_x^n. Le résultat principal est donné par le lemme suivant.

Lemme 6.12 *Si g est affine par morceaux, alors $\mathcal{T}_x g$ et $\mathcal{T}_v^n g$ le sont aussi. On a de plus pour toute cellule dyadique α*

$$|\mathcal{T}_x g|_{W^{1,\infty}(\alpha)} \leq (1 + \Delta t\,/2)\ |g|_{W^{1,\infty}(\mathcal{A}_x^{-1}(\alpha))} \tag{6.49}$$

$$|\mathcal{T}_v^n g|_{W^{1,\infty}(\alpha)} \leq (1 + \Delta t\,|\tilde{E}^n|_{W^{1,\infty}})\ |g|_{W^{1,\infty}((\mathcal{A}_v^n)^{-1}(\alpha))} \tag{6.50}$$

$$|\mathcal{T}_x g|_{\star(\alpha)} \leq (1 + \Delta t\,/2)^2\ |g|_{\star(\mathcal{A}_x^{-1}(\alpha))}. \tag{6.51}$$

Et si α appartient à un maillage dyadique $M \in \mathcal{M}(\Omega)$ dont la projection (6.21) *vérifie*

$$\Gamma_x(M) \subset \Gamma_x^n, \tag{6.52}$$

alors on a également

$$\begin{aligned}|\mathcal{T}_v^n g|_{\star(\alpha)} \leq (1 + \Delta t\,|\tilde{E}^n|_{W^{1,\infty}})^2 |g|_{\star((\mathcal{A}_v^n)^{-1}(\alpha))} \\ + 5 \cdot 2^{-2\ell(\alpha)} \Delta t\,|\tilde{E}^n|_{W^{2,\infty}} \|\partial_v g\|_{L^\infty((\mathcal{A}_v^n)^{-1}(\alpha))}.\end{aligned} \tag{6.53}$$

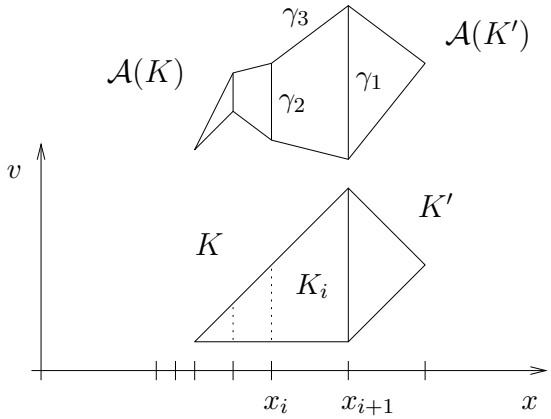


FIG. 6.2 – transport affine par morceaux.

Preuve. Pour étudier $\mathcal{T}_x$ et $\mathcal{T}_v^n$ dans un même temps, on introduit, pour une grille d'interpolation $\Gamma_x \subset \mathbb{R}$ et une fonction $\tilde{G} \in W^{2,\infty}(\mathbb{R})$, l'opérateur de transport générique

$$\mathcal{T} = \mathcal{T}(\Gamma_x, \tilde{G}) : g \to g \circ \mathcal{A}^{-1} \tag{6.54}$$

associé au difféomorphisme

$$\mathcal{A} = \mathcal{A}(\Gamma_x, \tilde{G}) : (x, v) \to (x, v + G(x)), \tag{6.55}$$

G désignant l'interpolation affine de $\tilde{G}$ sur Γ_x. Dans ces conditions, on retrouve $\mathcal{T}_v^n$ en prenant $\tilde{G}(x) = \Delta t\,\tilde{E}^n(x)$ et $\Gamma_x = \Gamma_x^n$, tandis que $\mathcal{T}_x$ s'obtient en commutant x

et v après avoir pris $G(x) = \tilde{G}(x) = x\Delta t\,/2$ (indépendamment de Γ_x). En désignant par $\mathcal{K}(g)$ la triangulation correspondant aux morceaux affines de g, et en notant x_i les nœuds de Γ_x, on peut découper chaque triangle de $\mathcal{K}(g)$ en bandes de la forme $K_i = K \cap]x_i, x_{i+1}[\times \mathbb{R}$. On vérifie alors que $\mathcal{T}g$ est continue et affine sur chaque $\mathcal{A}(K_i)$, où son gradient vaut

$$D(\mathcal{T}g)(x,v) = (\partial_x g - G'(x)\partial_v g, \partial_v g)\,(\mathcal{A}^{-1}(x,v)). \tag{6.56}$$

On peut déjà en déduire $|\mathcal{T}g|_{W^{1,\infty}(\alpha)} \leq (1 + \|G'\|_{L^\infty})\,|g|_{W^{1,\infty}(\mathcal{A}^{-1}(\alpha))}$ pour toute cellule α, ce qui nous donne (6.49) et (6.50) dans la mesure où $\|G'\|_{L^\infty} \leq \|\tilde{G}'\|_{L^\infty}$.

Pour établir les inégalités (6.51) et (6.53), on regroupe les arêtes des morceaux $\mathcal{A}(K_i)$ (portant les dérivées secondes de $\mathcal{T}g$) en trois catégories : celles qui proviennent d'une arête *verticale* d'un triangle $K \in \mathcal{K}(g)$ (comme γ_1 dans la figure 6.2), celles (comme γ_2) qui proviennent d'une arête verticale d'un K_i mais ne sont pas dans la première catégorie, et finalement celles (comme γ_3) qui proviennent d'une arête *oblique* ou *horizontale* d'un K_i. Soit alors γ une arête associée à $\mathcal{T}g$, telle que $\gamma \cap \alpha \neq \emptyset$. Si c'est une arête de troisième catégorie, on voit que le seul saut du gradient (6.56) vient d'un saut de $D(g) = (\partial_x g, \partial_v g)$, de sorte que

$$\|[D(\mathcal{T}g)]_\gamma\| \leq (1 + \|G'\|_{L^\infty})\,\|[D(g)]_{\mathcal{A}^{-1}(\gamma)}\|, \tag{6.57}$$

et un petit calcul géométrique nous montre que $|\gamma|_{\mathcal{H}^1} \leq (1 + \|G'\|_{L^\infty})\,|\mathcal{A}^{-1}(\gamma)|_{\mathcal{H}^1}$. Si γ est une arête de deuxième catégorie, le saut de gradient vient maintenant du saut de G' au point x_i de Γ_x tel que $\gamma \subset \{x_i\} \times \mathbb{R}$, et

$$\|[D(\mathcal{T}g)]_\gamma\| \leq \|\partial_v g\|_{L^\infty(\mathcal{A}^{-1}(\alpha))}\big|[G']_{x_i}\big|. \tag{6.58}$$

Enfin si γ est de première catégorie, le saut de gradient vient à la fois de $D(g)$ et G', de sorte que

$$\|[D(\mathcal{T}g)]_\gamma\| \leq (1 + \|G'\|_{L^\infty})\|[D(g)]_{\mathcal{A}^{-1}(\gamma)}\| + \|\partial_v g\|_{L^\infty(\mathcal{A}^{-1}(\alpha))}\big|[G']_{x_i}\big|, \tag{6.59}$$

où $x_i \in \Gamma_x$ vérifie à nouveau $\gamma \subset \{x_i\} \times \mathbb{R}$, et dans ces deux derniers cas, on a clairement

$$|\gamma|_{\mathcal{H}^1} = |\mathcal{A}^{-1}(\gamma)|_{\mathcal{H}^1} \leq (1 + \|G'\|_{L^\infty})\,|\mathcal{A}^{-1}(\gamma)|_{\mathcal{H}^1}. \tag{6.60}$$

En désignant alors par α_x et α_v les projections de α sur les axes x et v, on réunit les trois cas observés ci-dessus en écrivant

$$\begin{aligned}|\mathcal{T}g|_{\star(\alpha)} = \sum\nolimits_\gamma |\gamma \cap \alpha|_{\mathcal{H}^1}\|[D(\mathcal{T}g)]_\gamma\| \leq (1 + \|G'\|_{L^\infty})^2 \sum\nolimits_\lambda |\lambda \cap \mathcal{A}^{-1}(\alpha)|_{\mathcal{H}^1}\|[D(g)]_\lambda\| \\ + 2^{-\ell(\alpha)}\|\partial_v g\|_{L^\infty(\mathcal{A}^{-1}(\alpha))} \sum_{x_i \in \overline{\alpha_x}} \big|[G']_{x_i}\big|, \end{aligned} \tag{6.61}$$

où la première somme parcourt les arêtes de $\mathcal{T}g$, et la deuxième celles de g. Lorsque $\mathcal{T}$ représente $\mathcal{T}_x$, $G' = \Delta t\,/2$ est une constante, elle n'a donc pas de sauts et (6.51) se déduit immédiatement de (6.61). Dans le cas où $\tilde{G} = \Delta t\,\tilde{E}^n$ et $\mathcal{T} = \mathcal{T}_v^n$, on peut observer en raisonnant de la même façon que pour (6.24) que

$$\sum_{x_i \in \overline{\alpha_x}} \big|[(E^n)']_{x_i}\big| \leq \int_{x^-}^{x^+} |(\tilde{E}^n)''(x)|\,\mathrm{d}x, \tag{6.62}$$

où x^- et x^+ désignent les premiers nœuds de Γ_x^n situés respectivement avant et après l'intervalle fermé $\overline{\alpha_x}$. On utilise alors la condition (6.52) en voyant que la structure graduée de M se transmet à Γ_x^n, de sorte que x^- et x^+ ne peuvent pas être arbitrairement éloignés. Plus précisément, on a $|x^+ - x^-| \leq 5 \cdot 2^{-\ell(\alpha)}$, d'où l'on déduit que

$$\sum_{x_i \in \overline{\alpha_x}} \left|[(E^n)']_{x_i}\right| \leq 5 \cdot 2^{-\ell(\alpha)} |\tilde{E}^n|_{W^{2,\infty}}, \tag{6.63}$$

et l'on retrouve bien (6.53). □

Remarque 6.13 *La condition* (6.52)*, qui peut paraître contraignante, est en fait assez naturelle, car elle n'exclut que les cellules α 'trop fines'. En particulier, le lemme suivant nous montre qu'elle est toujours vérifiée par les maillages transportés avec l'algorithme $\boldsymbol{T}[\mathcal{A}_v^n]$.*

Lemme 6.14 *Pour un maillage dyadique arbitraire M, on a*

$$\Gamma_x(\boldsymbol{T}[\mathcal{A}_v^n]M) \subset \Gamma_x(M). \tag{6.64}$$

En d'autres termes, l'algorithme de transport de maillages associé au déplacement $\mathcal{A}_v^n$ n'élargit pas la projection (6.21) *des maillages dyadiques sur l'axe des x.*

Preuve. Montrons pour commencer que l'inclusion (6.64) est vérifiée par la partition *non graduée* résultant du découpage adaptatif sans raffinement :

$$\Gamma_x(\tilde{\boldsymbol{T}}[\mathcal{A}_v^n]M) \subset \Gamma_x(M). \tag{6.65}$$

Si α est une cellule de $\tilde{\boldsymbol{T}}[\mathcal{A}_v^n]M$, sa cellule parente $\tilde{\alpha} = \mathcal{P}(\alpha)$ vérifie par construction $\ell^*(\tilde{\alpha}) > \ell(\tilde{\alpha})$, de sorte qu'il existe une cellule $\beta \in M$ contenant $\left(\mathcal{A}_v^n\right)^{-1}(m_{\tilde{\alpha}})$ et telle que

$$\ell(\beta) = \ell^*(\tilde{\alpha}) \geq \ell(\tilde{\alpha}) + 1 = \ell(\alpha). \tag{6.66}$$

D'après la forme de $\mathcal{A}_v^n$, on peut observer que β_x et $\tilde{\alpha}_x$ s'intersectent nécessairement, et compte-tenu de (6.66), que le premier est strictement inclus dans le deuxième. Comme $\Gamma_x(M)$, qui contient par définition les bornes de β_x, hérite de M une structure dyadique, on en déduit facilement - et indépendamment du niveau de β - que $\Gamma_x(M)$ contient également les bornes de α_x, ce qui prouve (6.65). En utilisant le fait que $\Gamma_x(M)$ hérite en réalité de M une structure dyadique *graduée*, on peut vérifier que (6.65) implique (6.64). □

Preuve de la proposition 6.11. En utilisant successivement l'estimation (6.51) et le fait que Δt est borné par une constante, on a

$$|\mathcal{T}_x g|_{*(\alpha)} \leq C|g|_{*(\mathcal{A}_x^{-1}(\alpha))} \leq C \sum\nolimits_{\beta \in \mathcal{I}_{M,\mathcal{A}_x}(\alpha)} |g|_{*(\beta)},$$

la dernière inégalité venant du fait que les cellules de $\mathcal{I}_{M,\mathcal{A}_x}(\alpha)$ recouvrent $\mathcal{A}_x^{-1}(\alpha)$, quel que soit M. D'après l'équivalence (3.10), on en déduit

$$|\mathcal{T}_x g|_{BC(\alpha)} \leq C \sum\nolimits_{\beta \in \mathcal{I}_{M,\mathcal{A}_x}(\alpha)} |g|_{BC(\beta)}. \tag{6.67}$$

De même, on calcule d'après (6.49), et en utilisant maintenant le fait (6.46) que pour les cellules α de $\boldsymbol{T}[\mathcal{A}_x]M$, les cellules de $\mathcal{I}_{M,\mathcal{A}_x}(\alpha)$ sont de niveau "plutôt inférieur" à $\ell(\alpha)$,

$$2^{-2\ell(\alpha)}|\mathcal{T}_x g|_{W^{1,\infty}(\alpha)} \leq C2^{-2\ell(\alpha)}|g|_{W^{1,\infty}(\mathcal{A}_x^{-1}(\alpha))} \leq C\sum\nolimits_{\beta\in\mathcal{I}_{M,\mathcal{A}_x}(\alpha)} 2^{-2\ell(\beta)}|g|_{W^{1,\infty}(\beta)},$$

ce qui avec (6.67), nous donne (6.47). Si α désigne maintenant une cellule de $\boldsymbol{T}[\mathcal{A}_v^n]M$, le lemme 6.14 nous garantit qu'elle vérifie l'hypothèse (6.52), de sorte qu'on peut appliquer (6.53). En utilisant la borne (6.43), on a

$$\begin{aligned}|\mathcal{T}_v^n g|_{*(\alpha)} &\leq C|g|_{*((\mathcal{A}_v^n)^{-1}(\alpha))} + C2^{-2\ell(\alpha)}\Delta t\,|\tilde{E}^n|_{W^{2,\infty}}|g|_{W^{1,\infty}((\mathcal{A}_v^n)^{-1}(\alpha))}\\ &\leq C(1+|\tilde{E}^n|_{W^{2,\infty}})\sum\nolimits_{\beta\in\mathcal{I}_{M,\mathcal{A}_v^n}(\alpha)}\left[|g|_{*(\beta)} + 2^{-2\ell(\beta)}\Delta t\,|g|_{W^{1,\infty}(\beta)}\right]\end{aligned}$$

en utilisant à nouveau la propriété (6.46) pour les cellules de $\mathcal{I}_{M,\mathcal{A}_v^n}(\alpha)$. Et d'après (3.10), on en déduit

$$|\mathcal{T}_v^n g|_{*(\alpha)} \leq C\sum\nolimits_{\beta\in\mathcal{I}_{M,\mathcal{A}_x}(\alpha)}\mathcal{E}(g,\beta). \tag{6.68}$$

En utilisant finalement (6.50), on calcule comme précédemment

$$2^{-2\ell(\alpha)}|\mathcal{T}_v^n g|_{W^{1,\infty}(\alpha)} \leq C2^{-2\ell(\alpha)}|g|_{W^{1,\infty}((\mathcal{A}_v^n)^{-1}(\alpha))} \leq C\sum\nolimits_{\beta\in\mathcal{I}_{M,\mathcal{A}_x}(\alpha)} 2^{-2\ell(\beta)}|g|_{W^{1,\infty}(\beta)},$$

ce qui, avec (6.68), nous donne (6.48) et termine la preuve. □

Corollaire 6.15 (adéquation des maillages M^n, M_1^n et M_3^n) *On a*

$$\|(I-P_{M_1^n})\mathcal{T}_x f^n\|_{L^\infty} \leq C\varepsilon, \tag{6.69}$$

$$\|(I-P_{M_3^n})f_2^n\|_{L^\infty} \leq C\varepsilon, \tag{6.70}$$

$$\|(I-P_{M^{n+1}})\mathcal{T}_x f_3^n\|_{L^\infty} \leq C\varepsilon \tag{6.71}$$

avec des constantes indépendantes de $n \leq N$.

Preuve. Rappelons que la fonctionnelle

$$\bar{\mathcal{E}}(g,M) := \sup_{\alpha\in M}\mathcal{E}(g,\alpha) = \sup_{\alpha\in M}\left(|g|_{BC(\alpha)} + \Delta t\,2^{-2\ell(\alpha)}|g|_{W^{1,\infty}(\alpha)}\right) \tag{6.72}$$

mesure l'adéquation entre g et M, au sens où l'erreur de projection sur V_M vérifie

$$\|g - P_M g\|_{L^\infty} \leq C\bar{\mathcal{E}}(g,M). \tag{6.73}$$

Avec ce qui précède, on peut déjà estimer les erreurs de projections associées aux différentes étapes du schéma (6.17), à l'exception de (6.17c). La construction (6.17d) de M_3^n nous garantit en effet que

$$\bar{\mathcal{E}}(f_2^n, M_3^n) \leq \varepsilon, \tag{6.74}$$

et on peut vérifier qu'on a

$$\bar{\mathcal{E}}(P_M g, M) \leq C\bar{\mathcal{E}}(g,M) \tag{6.75}$$

avec une constante absolue, de sorte que l'application conjointe de (6.45) et (6.47) nous donne

$$\bar{\mathcal{E}}(\mathcal{T}_x f_3^n, M^{n+1}) \leq C\bar{\mathcal{E}}(f_3^n, M_3^n) \leq C\bar{\mathcal{E}}(f_2^n, M_3^n) \leq C\varepsilon. \tag{6.76}$$

M^0 étant d'ailleurs construit de façon à ce que $\bar{\mathcal{E}}(f_0, M^0)$ soit inférieur à ε, on peut également déduire de (6.75) que

$$\bar{\mathcal{E}}(f^n, M^n) \leq C\varepsilon \tag{6.77}$$

pour tout entier n. L'argument employé pour (6.76) nous permet alors d'écrire

$$\bar{\mathcal{E}}(\mathcal{T}_x f^n, M_1^n) \leq C\bar{\mathcal{E}}(f^n, M^n) \leq C\varepsilon, \tag{6.78}$$

et on en déduit les estimations (6.69)-(6.71). □

6.4.3 Stabilité du transport vis-à-vis des perturbations de densité

En ce qui concerne (HT.4), on s'intéressera dans notre analyse à la différence $\mathcal{T}_x\mathcal{T}_v\mathcal{T}_x f(t_n) - \mathcal{T}_x\mathcal{T}_v^n\mathcal{T}_x f(t_n) = \mathcal{T}_x(\mathcal{T}_v - \mathcal{T}_v^n)\mathcal{T}_x f(t_n)$ qu'on se propose de majorer ici en fonction de l'erreur numérique

$$e_n := \|f(t_n) - f^n\|_{L^\infty} \tag{6.79}$$

associée au schéma, et du paramètre ε. Ce passage étant relativement technique, on peut commencer par établir quelques inégalités qui nous permettront de faire quelques simplifications dans la suite.

Avant même de savoir à quelle vitesse e_n sera entraîné vers 0 par Δt et ε, on peut déduire de la stabilité L^∞ (6.42) du schéma que $e_n \leq 2\|f_0\|_{L^\infty}$ pour tout entier n. Les erreurs d'interpolation (6.69)-(6.71) étant d'autre part majorées par $C\varepsilon$, il est assez naturel de supposer que $C\varepsilon \leq e_n$ avec cette même constante (ce qui revient à remplacer e_n par $\max(C\varepsilon, e_n)$). On pourra donc supposer que

$$C\varepsilon \leq e_n \leq 2\|f_0\|_{L^\infty}. \tag{6.80}$$

Compte tenu de l'hypothèse (6.34) on pourra également utiliser

$$\varepsilon \leq \Delta t^{\,2} \leq 1 \tag{6.81}$$

pour simplifier quelques expressions. D'après la définition des opérateurs de transport, on peut majorer

$$\|(\mathcal{T}_v - \mathcal{T}_v^n)\mathcal{T}_x f(t_n)\|_{L^\infty} \leq \Delta t\, \|\tilde{E}[\mathcal{T}_x f(t_n)] - E^n\|_{L^\infty} |\mathcal{T}_x f(t_n)|_{W^{1,\infty}}. \tag{6.82}$$

En utilisant la proposition 4.3, il est facile de déduire de la définition (6.1)-(6.2) que $\mathcal{T}_x f(t_n)$ vérifie

$$|\mathcal{T}_x f(t_n)|_{W^{1,\infty}} \leq C|f(t_n)|_{W^{1,\infty}} \leq C \tag{6.83}$$

avec une constante dépendant uniquement de f_0 et $N\Delta t$. D'autre part, on peut écrire

$$\|\tilde{E}[\mathcal{T}_x f(t_n)] - E^n\|_{L^\infty} \leq \|\tilde{E}[\mathcal{T}_x f(t_n)] - \tilde{E}[f_1^n]\|_{L^\infty} + \|\tilde{E}[f_1^n] - \tilde{E}^n\|_{L^\infty} + \|\tilde{E}^n - E^n\|_{L^\infty}. \tag{6.84}$$

D'après la définition (6.4) et la borne $\|K\|_{L^\infty} = 1$, le premier terme se majore par

$$\|\tilde{E}[\mathcal{T}_x f(t_n)] - \tilde{E}[f_1^n]\|_{L^\infty} \leq \|\mathcal{T}_x f(t_n) - f_1^n\|_{L^1} \leq \tilde{\Sigma}_v \|\mathcal{T}_x f(t_n) - f_1^n\|_{L^\infty}, \tag{6.85}$$

cette dernière inégalité provenant du fait que les supports de toutes les solutions (approchées ou exactes) sont de mesure bornée par $\tilde{\Sigma}_v$. En utilisant (6.69) et (6.80), on a d'autre part

$$\|\mathcal{T}_x f(t_n) - f_1^n\|_{L^\infty} \leq \|\mathcal{T}_x(f(t_n) - f^n)\|_{L^\infty} + \|(I - P_{M_1^n})\mathcal{T}_x f^n\|_{L^\infty} \leq e_n + C\varepsilon \leq 2e_n. \tag{6.86}$$

Pour le deuxième terme de (6.84), on peut écrire d'après la définition (6.19) de $\tilde{E}^n$ que

$$\|\tilde{E}[f_1^n] - \tilde{E}^n\|_{L^\infty} \leq \left|\tilde{E}[f_1^n](1) - \tilde{E}[f_1^n](0)\right| = \left|\int_0^1 \left(\tilde{E}[f_1^n]\right)'(x)\,\mathrm{d}x\right| \leq \left|\|f_1^n\|_{L^1} - 1\right|. \tag{6.87}$$

Si le schéma numérique était conservatif, ce dernier terme serait nul en vertu de (4.31), malheureusement ce n'est pas le cas. $\mathcal{T}_x$ étant tout de même conservatif, on voit que $\|\mathcal{T}_x f(t_n)\|_{L^1} = 1$, de sorte que les inégalités (6.85) et (6.86) entraînent

$$\left|\|f_1^n\|_{L^1} - 1\right| \leq \|f_1^n - \mathcal{T}_x f(t_n)\|_{L^1} \leq 2\tilde{\Sigma}_v e_n. \tag{6.88}$$

Le troisième terme de (6.84) se majore comme une erreur d'interpolation $\mathcal{P}^1$

$$\|\tilde{E}^n - E^n\|_{L^\infty} \leq \frac{1}{8} \sup_i |x_{i+1} - x_i|^2 \|(\tilde{E}^n)''\|_{L^\infty([x_i, x_{i+1}])} \tag{6.89}$$

où les x_i désignent à nouveau les nœuds de Γ_x^n. On a alors

$$\|(\tilde{E}^n)''\|_{L^\infty([x_i,x_{i+1}])} = \left\|\int \partial_x f_1^n(\cdot, v)\,\mathrm{d}v\right\|_{L^\infty([x_i,x_{i+1}])} \leq \tilde{\Sigma}_v \sup_\alpha |f_1^n|_{W^{1,\infty}(\alpha)}, \tag{6.90}$$

le sup étant pris sur toutes les α de M_1^n dont la projection α_x intersecte $[x_i, x_{i+1}]$. D'après la construction de Γ_x^n, ceci implique $|x_{i+1} - x_i| \leq 2^{-\ell(\alpha)}$, d'où

$$\|\tilde{E}^n - E^n\|_{L^\infty} \leq C \sup_{\alpha \in M_1^n} 2^{-2\ell(\alpha)} |f_1^n|_{W^{1,\infty}(\alpha)} \leq C\Delta t^{-1} \bar{\mathcal{E}}(f_1^n, M_1^n) \leq C\varepsilon\Delta t^{-1} \tag{6.91}$$

en utilisant (6.72), (6.75) et (6.78). Si l'on met finalement bout à bout les différentes inégalités ci-dessus, on trouve

$$\|(\mathcal{T}_v - \mathcal{T}_v^n)\mathcal{T}_x f(t_n)\|_{L^\infty} \leq C(\Delta t\, e_n + \varepsilon). \tag{6.92}$$

6.5 Preuve du théorème 6.1

Pour ne pas être gêné outre mesure par l'opérateur de troncature T_n (défini en (6.26)) qui est susceptible de dégrader l'adéquation du maillage au voisinage de $|v| = \tilde{\Sigma}_v^n$, on se propose de décomposer l'analyse d'erreur en deux parties. Utilisant le domaine

$$\tilde{\Omega}_n = \mathbb{T} \times [-\tilde{\Sigma}_v^n, \tilde{\Sigma}_v^n]$$

introduit dans la remarque 6.6, l'erreur *principale* sera définie comme

$$e_n^{\text{princ}} := \|f(t_n) - f^n\|_{L^\infty(\tilde{\Omega}_n)}, \tag{6.93}$$

autrement dit sur un domaine où T_n vaut l'identité. Plus précisément, on observera attentivement à partir des égalités (6.32) que la restriction à $\tilde{\Omega}_{n+1}$ de la solution $\mathbb{S}_{\Delta t,\varepsilon} f^n$ définie par (6.18) vérifie

$$(\mathbb{S}_{\Delta t,\varepsilon} f^n)_{|\tilde{\Omega}_{n+1}} = (P_{M^{n+1}} \mathcal{T}_x P_{M_3^n} P_{M_2^n} \mathcal{T}_v^n P_{M_1^n} \mathcal{T}_x f^n)_{|\tilde{\Omega}_{n+1}}, \tag{6.94}$$

de sorte qu'on pourra "oublier" T_n dans l'analyse de e_n^{princ}. En revanche, il faudra prendre ses effets en compte pour estimer l'erreur *marginale* définie par

$$e_n^{\text{marg}} := \|f(t_n) - f^n\|_{L^\infty((\tilde{\Omega}_n)^{\mathsf{c}})}. \tag{6.95}$$

Commençons par l'étude de ce terme.

6.5.1 Estimation de l'erreur marginale

Comme $f(t_{n+1})$ s'annule en dehors de $\tilde{\Omega}_{n+1}$ (voir la remarque 6.6), on a

$$\begin{aligned} e_{n+1}^{\text{marg}} = \|f^{n+1}\|_{L^\infty((\tilde{\Omega}_{n+1})^{\mathsf{c}})} &\le \|\mathcal{T}_x f_3^n\|_{L^\infty((\tilde{\Omega}_{n+1})^{\mathsf{c}})} \le \|f_3^n\|_{L^\infty((\tilde{\Omega}_{n+1})^{\mathsf{c}})} \\ &\le \|f_2^n\|_{L^\infty((\tilde{\Omega}_{n+1})^{\mathsf{c}})} \le \|\mathsf{T}_{n+1} \mathcal{T}_v^n f_1^n\|_{L^\infty((\tilde{\Omega}_{n+1})^{\mathsf{c}})} \le \|\mathcal{T}_v^n f_1^n\|_{L^\infty((\tilde{\Omega}_{n+1})^{\mathsf{c}})}, \end{aligned}$$

les inégalités numéro 1, 2 et 4 provenant de la propriété de décroissance (6.31) des interpolations sur les ensembles $(\tilde{\Omega}_{n+1})^{\mathsf{c}}$, et les inégalités 3 et 5 venant du fait que $\mathcal{T}_x$ et T_{n+1} ne font pas croître la norme L^∞ sur $(\tilde{\Omega}_{n+1})^{\mathsf{c}}$. Pour contrôler l'amplitude de $\mathcal{T}_v^n f_1^n$ à l'extérieur de $\tilde{\Omega}_{n+1}$, on va établir l'inclusion

$$(\mathcal{A}_v^n)^{-1}\big((\tilde{\Omega}_{n+1})^{\mathsf{c}}\big) \subset \Omega_n^{\mathsf{c}}, \tag{6.96}$$

d'où l'on déduira

$$\|\mathcal{T}_v^n f_1^n\|_{L^\infty((\tilde{\Omega}_{n+1})^{\mathsf{c}})} = \|f_1^n \circ (\mathcal{A}_v^n)^{-1}\|_{L^\infty((\tilde{\Omega}_{n+1})^{\mathsf{c}})} \le \|f_1^n\|_{L^\infty((\Omega_n)^{\mathsf{c}})}.$$

En utilisant (6.69) et le fait que $f(t_n)$ s'annule en dehors de Ω_n, on pourra alors estimer

$$\begin{aligned} \|\mathcal{T}_v^n f_1^n\|_{L^\infty((\tilde{\Omega}_{n+1})^{\mathsf{c}})} \le \|f_1^n\|_{L^\infty((\Omega_n)^{\mathsf{c}})} &\le \|\mathcal{T}_x f^n\|_{L^\infty((\Omega_n)^{\mathsf{c}})} + \|(I - P_{M_1^n})\mathcal{T}_x f^n\|_{L^\infty} \\ &\le \|f^n\|_{L^\infty((\Omega_n)^{\mathsf{c}})} + C\varepsilon \le e_n + C\varepsilon. \end{aligned} \tag{6.97}$$

L'erreur marginale sera donc majorée par

$$e_{n+1}^{\text{marg}} \le e_n + C\varepsilon \tag{6.98}$$

si nous pouvons montrer l'inclusion (6.96). Pour cela, considérons un point (x, v) qui n'appartient pas à $\tilde{\Omega}_{n+1}$, autrement dit tel que $|v| \ge \tilde{\Sigma}_v^{n+1}$, et montrons que $(\mathcal{A}_v^n)^{-1}(x, v)$ est en dehors de Ω_n, autrement dit que $|v - \Delta t\, E^n(x)| \ge \Sigma_v^n$. L'inégalité $\tilde{\Sigma}_v^n \ge \Sigma_v^n + 2^{-\ell_0}$ (qui provient de la définition de $\tilde{\Sigma}_v^n$) nous permettant d'écrire

$$|v - \Delta t\, E^n(x)| \ge \tilde{\Sigma}_v^{n+1} - \Delta t\, \|E^n\|_{L^\infty} \ge \Sigma_v^n + 2^{-\ell_0} - \Delta t\, \|E^n\|_{L^\infty},$$

nous aurons montré (6.96) si nous établissons que $\Delta t\, \|E^n\|_{L^\infty}$ est inférieur à $2^{-\ell_0}$. Observons donc que le champ $\tilde{E}[f_1^n]$, défini en (6.4), vérifie d'après (6.88)

$$\|\tilde{E}[f_1^n]\|_{L^\infty} \le \|f_1^n\|_{L^1} + 1 \le 2 + 2\tilde{\Sigma}_v e_n,$$

et que les inégalités (6.22), (6.87) et (6.88) nous permettent de majorer

$$\|E^n\|_{L^\infty} \leq \|\tilde{E}[f_1^n]\|_{L^\infty} + \left|\|f_1^n\|_{L^1} - 1\right| \leq 2 + 4\tilde{\Sigma}_v e_n. \tag{6.99}$$

D'après l'hypothèse (6.34) faite sur Δt et l'estimation très large (6.80) suivant laquelle l'erreur numérique e_n est inférieure à $2\|f_0\|_{L^\infty}$, on peut alors voir que la quantité $\Delta t\,\|E^n\|_{L^\infty}$ est toujours inférieure à $2^{-\ell_0}(2 + 8\tilde{\Sigma}_v\|f_0\|_{L^\infty})(8 + 8\tilde{\Sigma}_v\|f_0\|_{L^\infty})^{-1}$ et donc à $2^{-\ell_0}$, ce qui établit l'inclusion (6.96) et finalement l'estimation (6.98).

6.5.2 Régularité lipschitzienne des solutions

Pour étudier l'erreur numérique correspondant au domaine principal de calcul, on va avoir besoin de contrôler la semi-norme $W^{1,\infty}$ des solutions numériques et la semi-norme $W^{2,\infty}$ des champs électriques associés. Jusqu'ici, on peut remarquer qu'on n'a pas cherché à préciser quelle semi-norme lipschitzienne devait être utilisée. Pour pouvoir en suivre précisément l'évolution, faisons le choix de noter

$$|g|_{W^{1,\infty}} := \|\partial_x g\|_{L^\infty} + \|\partial_v g\|_{L^\infty}.$$

On a alors le lemme suivant (qui n'a rien d'évident, et qui serait faux pour une triangulation générale).

Lemme 6.16 *Pour ce choix de semi-norme lipschitzienne, les interpolations P_M sont décroissantes. En d'autres termes, on a*

$$|P_M g|_{W^{1,\infty}} \leq |g|_{W^{1,\infty}} \tag{6.100}$$

pour toute fonction lipschitzienne g et tout maillage dyadique $M \in \mathcal{M}(\Omega)$.

Preuve. Rappelons que P_M est l'interpolation affine par morceaux correspondant à la triangulation $\mathcal{K}(M)$ construite dans la section 2.2.1. Et que les triangles K de $\mathcal{K}(M)$ sont soit obtenus par le découpage en deux d'une cellule carrée de M, soit par le recollage de deux triangles voisins K_1 et K_2 de la triangulation non conforme $\tilde{\mathcal{K}}(M)$ (voir figure 2.4). Dans le premier cas, deux des côtés de K sont parallèles aux axes x et v, d'où l'on voit que $\|\partial_x P_M g\|_{L^\infty(K)}$ et $\|\partial_v P_M g\|_{L^\infty(K)}$ sont respectivement majorés par $\|\partial_x g\|_{L^\infty(K)}$ et $\|\partial_v g\|_{L^\infty(K)}$, de sorte que

$$|P_M g|_{W^{1,\infty}(K)} \leq |g|_{W^{1,\infty}(K)} \tag{6.101}$$

est évident. Dans le deuxième cas, soit $\tilde{g} := P_{\{K_1,K_2\}} g$ l'interpolation de g sur la sous-triangulation (conforme) $\{K_1, K_2\}$ de $\tilde{\mathcal{K}}(M)$. On peut observer que l'arête commune à K_1 et K_2 est toujours parallèle à l'axe x ou à l'axe v. Notre argument étant symmétrique, supposons qu'il est parallèle à l'axe x. On a alors $(\partial_x \tilde{g})_{|K_1} = (\partial_x \tilde{g})_{|K_2}$, et l'on peut calculer que le gradient de $P_M g$, constant sur K, vérifie

$$\begin{aligned}(\partial_x P_M g)_{|K} &= \left(2(\partial_x \tilde{g})_{|K_1} - (\partial_v \tilde{g})_{|K_1} + (\partial_v \tilde{g})_{|K_2}\right)/2\\ (\partial_v P_M g)_{|K} &= \left((\partial_v \tilde{g})_{|K_1} + (\partial_v \tilde{g})_{|K_2}\right)/2.\end{aligned}$$

On a donc

$$\begin{aligned}|P_M g|_{W^{1,\infty}(K)} &= \|\partial_x P_M g\|_{L^\infty(K)} + \|\partial_v P_M g\|_{L^\infty(K)} \\ &\leq \max(|(\partial_x P_M g)_{|K} + (\partial_v P_M g)_{|K}|, |(\partial_x P_M g)_{|K} - (\partial_v P_M g)_{|K}|) \\ &\leq \|\partial_x \tilde{g}\|_{L^\infty(K)} + \|\partial_v \tilde{g}\|_{L^\infty(K)} = |\tilde{g}|_{W^{1,\infty}(K)},\end{aligned}$$

et en appliquant (6.101) à $P_{\{K_1,K_2\}}$, on en déduit que

$$|P_M g|_{W^{1,\infty}(K)} \leq |\tilde{g}|_{W^{1,\infty}(K)} = \max_{i=1,2} |\tilde{g}|_{W^{1,\infty}(K_i)} \leq \max_{i=1,2} |g|_{W^{1,\infty}(K_i)} = |g|_{W^{1,\infty}(K)},$$

ce qui prouve (6.100). □

En utilisant la définition (6.26) de l'opérateur T_n, les inégalités (6.97), (6.80) et le fait que $\|\partial_v(\mathcal{T}_v^n f_1^n)\|_{L^\infty} = \|\partial_v f_1^n\|_{L^\infty}$, on calcule alors

$$\begin{aligned}\|\partial_v(\mathsf{T}_{n+1}\mathcal{T}_v^n f_1^n)\|_{L^\infty} &\leq \max\left(\|\partial_v(\mathcal{T}_v^n f_1^n)\|_{L^\infty}, 2^{\ell_0}\|\mathcal{T}_v^n f_1^n\|_{L^\infty((\tilde{\Omega}_{n+1})^{\mathsf{c}})}\right) \\ &\leq \|\partial_v(\mathcal{T}_v^n f_1^n)\|_{L^\infty} + 2^{\ell_0}\|\mathcal{T}_v^n f_1^n\|_{L^\infty((\tilde{\Omega}_{n+1})^{\mathsf{c}})} \leq \|\partial_v f_1^n\|_{L^\infty} + 2^{\ell_0}(e_n + C\varepsilon).\end{aligned}$$

T_{n+1} diminuant d'autre part la semi-norme $\|\partial_x \cdot\|_{L^\infty}$, on a

$$\|\partial_x(\mathsf{T}_{n+1}\mathcal{T}_v^n f_1^n)\|_{L^\infty} \leq \|\partial_x(\mathcal{T}_v^n f_1^n)\|_{L^\infty} \leq \|\partial_x f_1^n\|_{L^\infty} + \Delta t\, |\tilde{E}^n|_{W^{1,\infty}}\|\partial_v f_1^n\|_{L^\infty}.$$

On déduit donc de la borne (6.43) sur le champ $\tilde{E}^n$ que

$$\begin{aligned}|\mathsf{T}_{n+1}\mathcal{T}_v^n f_1^n|_{W^{1,\infty}} &= \|\partial_x(\mathsf{T}_{n+1}\mathcal{T}_v^n f_1^n)\|_{L^\infty} + \|\partial_v(\mathsf{T}_{n+1}\mathcal{T}_v^n f_1^n)\|_{L^\infty} \\ &\leq |f_1^n|_{W^{1,\infty}}(1 + C\Delta t\,) + 2^{\ell_0}(e_n + C\varepsilon) \leq |f_1^n|_{W^{1,\infty}}(1 + C\Delta t\,) + Ce_n,\end{aligned}$$

dans la mesure où l'on peut toujours supposer que e_n est supérieur à ε. En utilisant cette estimation, la décroissance (6.100) des interpolations et le fait que $|\mathcal{T}_x g|_{W^{1,\infty}} \leq (1 + \Delta t\,/2)|g|_{W^{1,\infty}}$, on obtient finalement

$$\begin{aligned}|f^{n+1}|_{W^{1,\infty}} &\leq |\mathcal{T}_x f_3^n|_{W^{1,\infty}} \leq (1 + \Delta t\,/2)|f_3^n|_{W^{1,\infty}} \leq (1 + \Delta t\,/2)|f_2^n|_{W^{1,\infty}} \\ &\leq (1 + \Delta t\,/2)|\mathsf{T}_{n+1}\mathcal{T}_v^n f_1^n|_{W^{1,\infty}} \leq (1 + \Delta t\,/2)(1 + C\Delta t\,)|f_1^n|_{W^{1,\infty}} + Ce_n \\ &\leq (1 + C\Delta t\,)|\mathcal{T}_x f^n|_{W^{1,\infty}} + Ce_n \leq (1 + C\Delta t\,)|f^n|_{W^{1,\infty}} + Ce_n.\end{aligned} \tag{6.102}$$

6.5.3 Borne $W^{2,\infty}$ sur le champ électrique $\tilde{E}^n$

D'après leurs définitions respectives (6.4) et (6.19), les champs $\tilde{E}[f_1^n]$ et $\tilde{E}^n$ vérifient

$$|\tilde{E}^n|_{W^{2,\infty}} = |\tilde{E}[f_1^n]|_{W^{2,\infty}} = \left\|\int \partial_x f_1^n(\cdot, v)\,\mathrm{d}v\right\|_{L^\infty} \leq \tilde{\Sigma}_v |f_1^n|_{W^{1,\infty}}.$$

En simplifiant d'après (6.80), on déduira donc des inégalités (6.102) que

$$|\tilde{E}^n|_{W^{2,\infty}} \leq C(1 + |f^n|_{W^{1,\infty}}), \tag{6.103}$$

avec une constante dépendant uniquement de f_0 et $N\Delta t$.

6.5.4 Estimation de l'erreur principale

Muni de cette dernière estimation, on est enfin en mesure d'appliquer les résultats disponibles quant à l'adéquation des maillages M_2^n transportés lors de l'étape (6.17c) : en particulier, (6.45) et (6.48) nous permettent d'écrire

$$\bar{\mathcal{E}}(\mathcal{T}_v^n f_1^n, M_2^n) \leq C(1 + |\tilde{E}^n|_{W^{2,\infty}})\bar{\mathcal{E}}(f_1^n, M_1^n).$$

La stabilité (6.75) des projections et l'adéquation (6.78) de M_1^n nous garantissant que

$$\bar{\mathcal{E}}(f_1^n, M_1^n) \leq C\bar{\mathcal{E}}(\mathcal{T}_x f^n, M_1^n) \leq C\varepsilon,$$

on déduit des estimations ci-dessus que

$$\bar{\mathcal{E}}(\mathcal{T}_v^n f_1^n, M_2^n) \leq C(1 + |f^n|_{W^{1,\infty}})\varepsilon. \tag{6.104}$$

En particulier, l'erreur de projection associée à l'étape (6.17c) du schéma vérifie

$$\|(I - P_{M_2^n})\mathcal{T}_v^n f_1^n\|_{L^\infty} \leq C(1 + |f^n|_{W^{1,\infty}})\varepsilon. \tag{6.105}$$

Suivant le canevas proposé dans la section 5.4.1 pour l'analyse d'erreur, on peut à présent décomposer le terme d'erreur principal (6.93) en

$$e_{n+1}^{\text{princ}} \leq e_{n+1}^{\text{princ, t}} + e_{n+1}^{\text{princ, s}} + e_{n+1}^{\text{princ, c}}, \tag{6.106}$$

où l'on désigne respectivement par

$$e_{n+1}^{\text{princ, t}} := \|f(t_{n+1}) - \mathcal{T} f(t_n)\|_{L^\infty} \leq C(T)\Delta t^{\,3} \tag{6.107}$$

l'erreur de discrétisation en temps majorée dans la proposition 6.1 sur le domaine entier Ω, par

$$e_{n+1}^{\text{princ, s}} := \|(\mathcal{T}_x\mathcal{T}_v^n\mathcal{T}_x - \mathbb{S}_{\Delta t,\varepsilon} f^n\|_{L^\infty(\tilde{\Omega}_{n+1})},$$

une erreur de discrétisation en espace, et par

$$e_{n+1}^{\text{princ, c}} := \|\mathcal{T}_x\mathcal{T}_v\mathcal{T}_x f(t_n) - \mathcal{T}_x\mathcal{T}_v^n\mathcal{T}_x f^n\|_{L^\infty(\tilde{\Omega}_{n+1})}$$

un terme additionnel de "couplage" dû à la non-linéarité du transport. D'après la forme (6.94) prise par $\mathbb{S}_{\Delta t,\varepsilon} f^n$ sur $\tilde{\Omega}_{n+1}$, on décompose alors

$$\begin{aligned} e_{n+1}^{\text{princ, s}} \leq \|\mathcal{T}_x\mathcal{T}_v^n(I - P_{M_1^n})\mathcal{T}_x f^n\|_{L^\infty} + \|\mathcal{T}_x(I - P_{M_2^n})\mathcal{T}_v^n f_1^n\|_{L^\infty} + \|\mathcal{T}_x(I - P_{M_3^n})f_2^n\|_{L^\infty} \\ + \|(I - P_{M^{n+1}})\mathcal{T}_x f_3^n\|_{L^\infty} \leq C(1 + |f^n|_{W^{1,\infty}})\varepsilon, \end{aligned} \tag{6.108}$$

la deuxième inégalité venant des estimations (6.69)-(6.71) et (6.105). En ce qui concerne l'erreur de couplage, on peut observer que la linéarité de $\mathcal{T}_x$ et de $\mathcal{T}_v^n$ entraînent

$$\begin{aligned} e_{n+1}^{\text{princ, c}} \leq \|\mathcal{T}_x\mathcal{T}_v^n\mathcal{T}_x(f(t_n) - f^n)\|_{L^\infty} + \|\mathcal{T}_x(\mathcal{T}_v - \mathcal{T}_v^n)\mathcal{T}_x f(t_n)\|_{L^\infty}, \\ \leq e_n(1 + C\Delta t\,) + C\varepsilon \end{aligned}$$

d'après l'estimation (6.92). L'erreur principale vérifie donc

$$e_{n+1}^{\text{princ}} \leq e_n(1 + C\Delta t\,) + C[\Delta t^{\,3} + \varepsilon(1 + |f^n|_{W^{1,\infty}})]. \tag{6.109}$$

6.5.5 Fin de la preuve

En collectant les estimations (6.98) et (6.109) concernant les erreurs marginale et principale, on trouve

$$e_{n+1} \leq e_n(1 + C\Delta t\,) + C[\Delta t^{\,3} + \varepsilon(1 + |f^n|_{W^{1,\infty}})] \tag{6.110}$$

tandis que $|f^n|_{W^{1,\infty}}$ vérifie, d'après (6.102),

$$|f^{n+1}|_{W^{1,\infty}} \leq |f^n|_{W^{1,\infty}}(1 + C\Delta t\,) + Ce_n. \tag{6.111}$$

L'hypothèse (6.34) impliquant $\varepsilon \leq \Delta t^{\,2}$, le lemme de Gronwall 1.4 appliqué à $e_{n+1} + \Delta t\,|f^{n+1}|_{W^{1,\infty}}$ montre que

$$\sup_{n\leq N}(e_n + \Delta t\,|f^n|_{W^{1,\infty}}) \leq C\Delta t\,.$$

On en déduit que l'erreur numérique vérifie

$$e_{n+1} \leq e_n(1 + C\Delta t\,) + C(\Delta t^{\,3} + \varepsilon),$$

de sorte qu'une deuxième application du lemme de Gronwall nous donne l'estimation d'erreur (6.33), ce qui achève cette preuve. □

Chapitre 7

Implémentation et résultats numériques

Sans entrer dans les détails techniques de la programmation, on présente ici la façon dont on a implémenté notre schéma adaptatif semi-lagrangien. L'écriture du code de calcul ayant en réalité précédé le schéma tel qu'il est décrit au chapitre précédent, il existe entre les deux quelques différences, mais elles sont mineures. On présente alors quelques résultats numériques correspondant à la simulation d'un faisceau d'électrons à symétrie axiale, sur lesquels il apparaît d'une part que les erreurs numériques tendent vers 0 comme $\Delta t^2 \sim \varepsilon^{2/3}$, ce qui confirme l'estimation du théorème 6.1, et d'autre part que la taille maximale N des maillages générés par notre schéma est contrôlée par ε^{-1}, ce qui va dans le sens de la conjecture développée dans la section 6.3.2, au moins sur une simulation. A la fin de ce chapitre, on montre également que nos maillages adaptatifs sont très proches des maillages "optimaux" qu'on construit à partir d'une solution de référence très précise, calculée sur un maillage uniformément fin.

7.1 Le code de calcul YODA

Avant d'être écrit comme en (6.17), notre schéma adaptatif a d'abord existé sous la forme d'un code de calcul écrit en *C++* avec Michel Mehrenberger à l'occasion d'un groupe de travail proposé par Eric Sonnendrücker et co-dirigé par Albert Cohen lors de l'édition 2003 du CEMRACS (le désormais célèbre Centre d'Eté de Mathématiques et Recherche Avancées en Calcul Scientifique), à Luminy.

7.1.1 Premiers objectifs et visualisation des résultats

L'objectif premier de notre solveur (que nous avons baptisé "Yet anOther aDaptive Algorithm" dans le cadre du projet CALVI de l'INRIA où avaient déjà été développés un solveur uniforme VADOR écrit par Francis Filbet durant sa thèse de doctorat, et un code de calcul adaptatif OBIWAN [11, 38] qui utilisait des ondelettes d'interpolation de Deslauriers et Dubuc) était d'accélérer les codes uniformes existant, tout en offrant la qualité (notamment la précision) des schémas semi-lagrangiens.

Par rapport au solveur Obiwan, le choix de discrétiser la densité de plasma par des éléments finis tensoriels basés sur les maillages dyadiques du chapitre 2 était justifié par

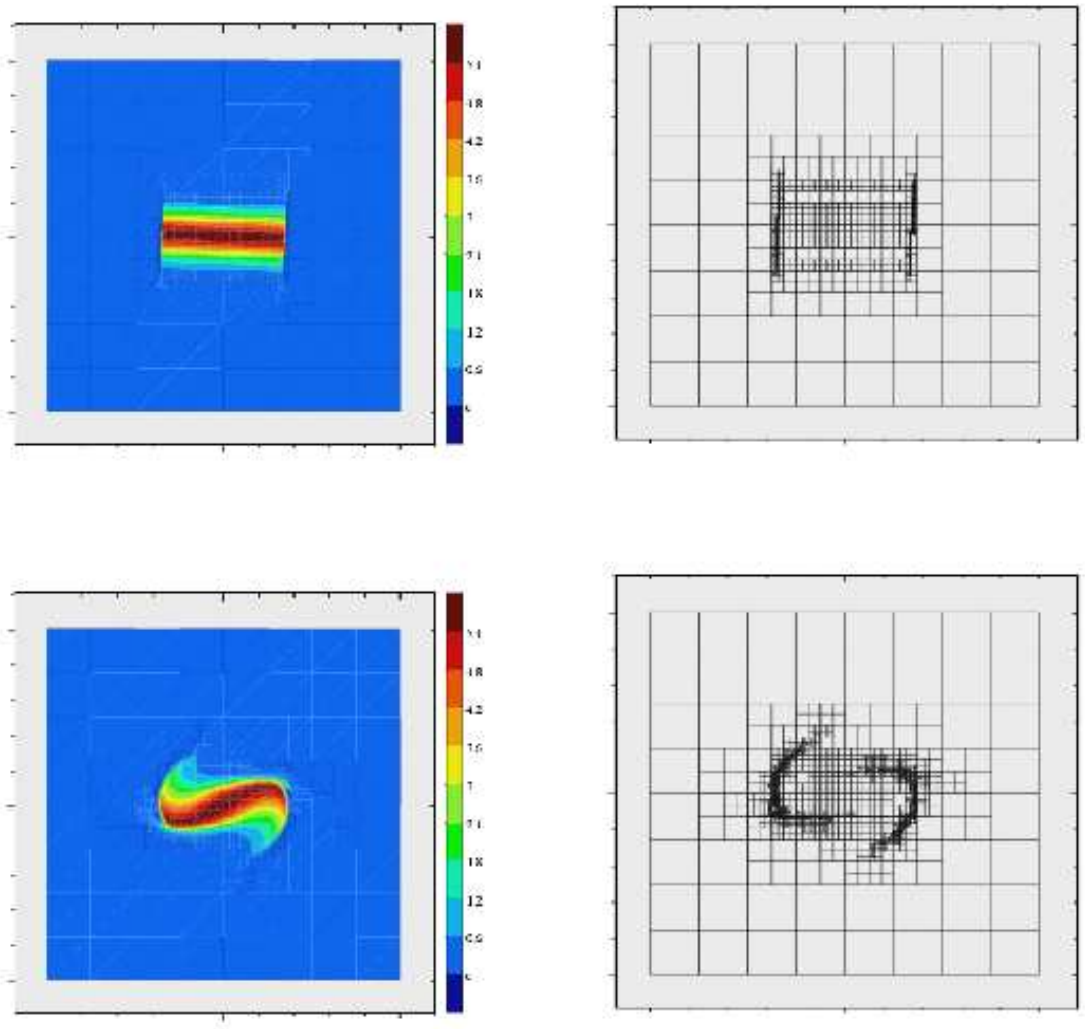


FIG. 7.1 – simulation d'un faisceau de plasma à symétrie axiale avec le code YODA. La distribution d'électrons f^n à gauche, et le maillage adaptatif M^n à droite sont représentés dans l'espace des phases avec en abscisse la distance des particules à l'axe du faisceau, et en ordonnées leur vitesse radiale.

une utilisation plus locale de la structure de données, pour une parallélisation du code qui a depuis été entreprise par Eric Violard, Michel Mehrenberger et Olivier Hoenen (voir [42] et [41]). Les figures 7.1 et 7.2 ci-après illustrent les premiers cas tests que nous avons simulés avec notre code.

7.1.2 Aspects essentiels de l'implémentation

Pour exploiter au mieux la structure arborescente de la discrétisation multi-échelles telle qu'on l'a décrite dans la section 2.1.1, chaque solution numérique f^n est d'abord matérialisée par l'*arbre de calcul* $\Lambda(M^n)$ dont les feuilles forment le maillage dyadique

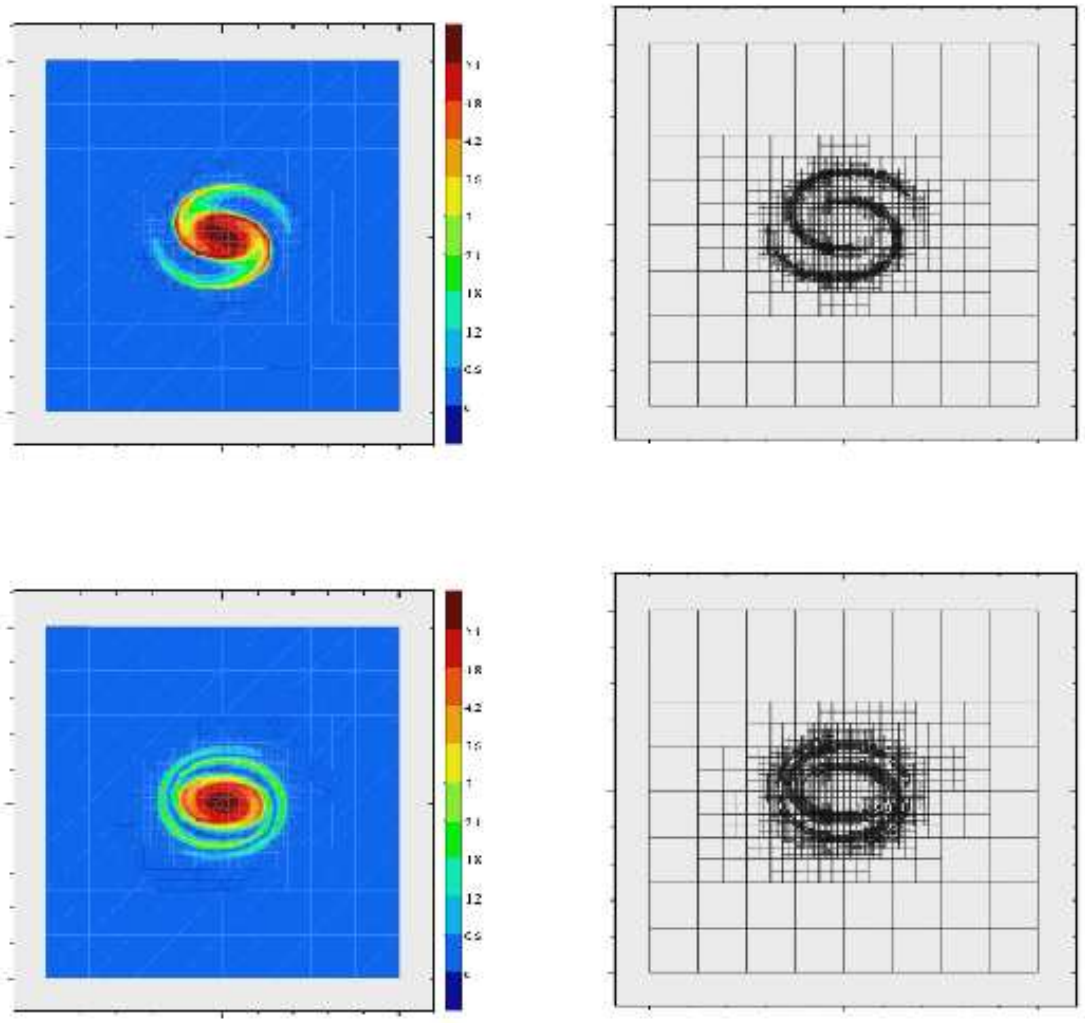


FIG. 7.2 – simulation d'un faisceau de plasma (suite de la figure 7.1).

M^n sur lequel f^n est discrétisée. Dans cet arbre, les mailles (ou les cellules) dyadiques sont des objets autonomes reliés aux autres par un jeu de pointeurs représentant les différents types de relations.

Chaque cellule $\alpha = [2^{-\ell}k_x, 2^{-\ell}(k_x+1)] \times [2^{-\ell}k_v, 2^{-\ell}(k_v+1)]$ connaît ainsi, outre son niveau ℓ et son indice $k = (k_x, k_v)$, l'adresse mémoire de sa parente $\mathcal{P}(\alpha)$, de ses filles $\mathcal{F}(\alpha)$ et de ses voisines (au même niveau) dans le maillage. La structure d'arbre gradué est alors garantie par une gestion minutieuse des diverses relations intercellulaires, notamment lorsqu'il s'agit d'ajouter ou d'enlever une cellule dans l'arbre. C'est bien sûr la plasticité de ces relations qui est à la base de l'adaptativité dynamique de nos maillages.

En parallèle avec ce graphe cellulaire dynamique, le programme gère un ensemble de nœuds partagés par les cellules de M^n au moyen d'un deuxième jeu de pointeurs,

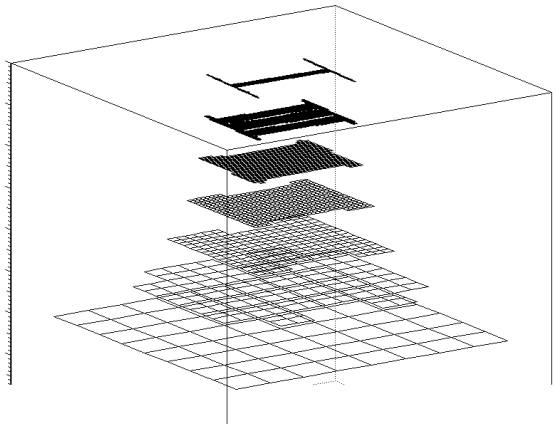


FIG. 7.3 – un arbre de cellules dyadiques $\Lambda(M^n)$ correspondant à une solution f^n.

et ce sont ces nœuds qui contiennent les valeurs des solutions numériques.

Dans notre approche, le maillage adaptatif sur lequel la solution est calculée est redessiné à chaque pas de temps. En particulier, la boucle correspondant à l'intervalle de temps $[n\Delta t\,, (n+1)\Delta t\,]$ commence par prédire un maillage $\tilde{M}^{n+1}$ en suivant le flot numérique $\mathcal{A}^n$ calculé à partir de la solution disponible (f^n, E^n). D'un point de vue algorithmique, il faut savoir que la création physique de nouvelles mailles demande qu'on alloue à chaque fois une zone de mémoire pour recevoir les informations qui composent la "carte d'identité" de cette maille. Ces allocations dynamiques de mémoire prenant un temps relativement important, il était essentiel de les réduire au minimum. On l'a fait de deux façons différentes : tout d'abord, on a créé un statut intermédiaire de cellules "recyclées" de façon à conserver des petits nombres de mailles inutilisées en attente d'être réaffectées quelque part. Mais surtout, comme le maillage prédit $\tilde{M}^{n+1}$ ne présentait généralement que peu de différences avec M^n, on a évité de multiplier des allocations redondantes en réunissant les deux arbres en un seul $\Lambda^n := \Lambda(M^n) \cup \Lambda(\tilde{M}^{n+1})$, réduisant ainsi de façon très importante le coût associé au transport des maillages dyadiques. Dans la mesure où la présence des cellules dans un arbre ou dans l'autre n'est alors plus matérialisée que par un marqueur, la plupart des créations de cellules deviennent virtuelles, le programme ne faisant plus que modifier localement le maillage.

Dans sa première version, notre code de calcul n'utilisait pas d'éléments affines par morceaux, mais des produits tensoriels entre polynômes de Lagrange, de sorte qu'on a programmé des éléments finis bilinéaires ou biquadratiques pour discrétiser les densités électroniques f^n (l'avantage de cette discrétisation étant qu'elle pouvait s'étendre aux dimensions supérieures avec la même facilité que les maillages dyadiques multiéchelles). Une deuxième différence avec notre schéma (6.17) est que les trajectoires caractéristiques n'y sont pas calculées par "time splitting", mais par une formule explicite directe. Précisons toutefois qu'à l'image du schéma "abstrait" (5.33) présenté au chapitre 5, notre code peut prendre en argument un schéma de transport $f^n \to \mathcal{A}^n = \mathcal{A}[f^n]$ arbitraire pourvu qu'il soit suffisamment régulier. La troisième différence, enfin, réside

dans la stratégie adoptée pour prédire le maillage au début de chaque pas de temps. L'algorithme implémenté regarde en effet "vers l'avant", ajoutant pour chaque cellule α du maillage de départ une cellule $\tilde{a}lpha$ de niveau identique (dans ce qui correspond à l'arbre Λ_L de l'algorithme 5.5) et contenant le point d'arrivée $\mathcal{A}^n(c_\alpha)$ de la trajectoire approchée issue du centre de la cellule α.

7.2 Faisceau d'électrons semi-gaussien

Pour tester la validité de notre code sur un cas test classique, on a simulé l'évolution temporelle de la section d'un faisceau d'électrons à symétrie axiale, initialement localisé autour de son axe et distribué de façon gaussienne en vitesses. L'allure des solutions adaptives calculées lors de ces simulations est illustrée par les figures 7.1 et 7.2. En désignant par x la distance des particules à l'axe du faisceau et par v leur vitesse radiale, la distribution initiale $f_{\rm sg}(x,v) := a\exp(-(v/b)^2)\chi_{[-c,c]\times\mathcal{R}}(x,v)$ n'est ni continue, ni à support compact. On l'a donc remplacée par

$$f_0 := a\exp(-(v/b)^2)\rho(x,v) \simeq f_{\rm sg}$$

où ρ est une approximation $W^{2,\infty}$ à support compact de l'indicatrice $\chi_{[-c,c]\times\mathcal{R}}$. Le domaine physique de calcul étant $[-0.5,0.5]^2$, on a choisi pour ces paramètres les valeurs $a = 5.794$, $b = 0.122$ et $c = 0.172$. D'autre part, on a soumis à l'équation un champ électrique extérieur affine, de façon à ce que le plasma reste bien confiné autour de l'axe.

On présente ici les résultats correspondant à des simulations adaptatives et uniformes. Les *solutions adaptives* correspondent à la stratégie adaptative telle qu'on a pu la décrire dans les sections 5.3.1 ou 6.2.2 (aux différences mineures près évoquées plus haut). On désignera ici ces solutions par $f^n = f^n_{\Delta t\,,\varepsilon}$, en précisant la valeur choisie pour le pas de temps Δt. Le paramètre de tolérance ε sera lui toujours fixé de façon à ce que l'on ait $\varepsilon = C\Delta t^{\,3}$, d'après l'équilibrage des différents termes d'erreur (avec un choix de $C = 320$). Signalons enfin que pour des raisons pratiques évidentes, on a autorisé un niveau maximal pour les cellules, fixé à $L = 10$. D'autre part, on a également calculé des *solutions uniformes* f^n_h sur des maillages uniformes de pas h en appliquant à notre schéma une tolérance $\varepsilon = 0$, et un niveau maximal $\ell_h := -\log_2(h)$. En prenant à nouveau soin d'équilibrer le rôle des paramètres Δt et h dans l'estimation d'erreur (6.37), on a imposé $\Delta t \sim h^{2/3}$ lorsque c'était possible.

La solution exacte n'étant pas connue, on a évalué la précision des solutions numériques en utilisant une solution de référence $f_L := f_{h(L)}$ calculée sur le niveau d'espace le plus fin $h(L) := 2^{-L} = 1/1024$.

7.2.1 Mesure du défaut de conservativité

Dans la mesure où notre schéma n'est pas conservatif, il était intéressant d'évaluer la perte de masse des solutions numériques au cours du temps. Rappelons que l'inégalité (6.88) prévoit que cette perte est contrôlée par l'erreur $\|f^N - f(T)\|_{L^\infty}$. En partant du principe que la masse $\|f^0_L\|_{L^1}$ de la solution initiale de référence est une bonne aproximation de la masse exacte $\|f\|_{L^1}$, on a représenté sur la figure 7.4 l'évolution du rapport $\|f^N\|_{L^1}/\|f^0_L\|_{L^1}$ en fonction du temps de simulation $T = N\Delta t$ pour différentes

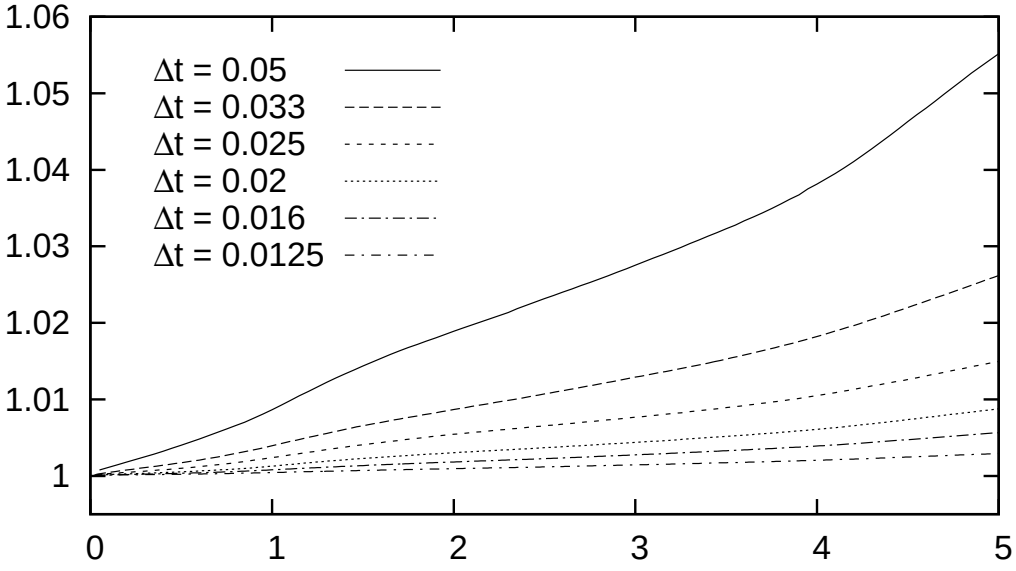


FIG. 7.4 – évolution du rapport de masse $\|f^N\|_{L^1}/\|f_L^0\|_{L^1}$ au cours du temps $T = N\Delta t$ pour diverses solutions adaptatives.

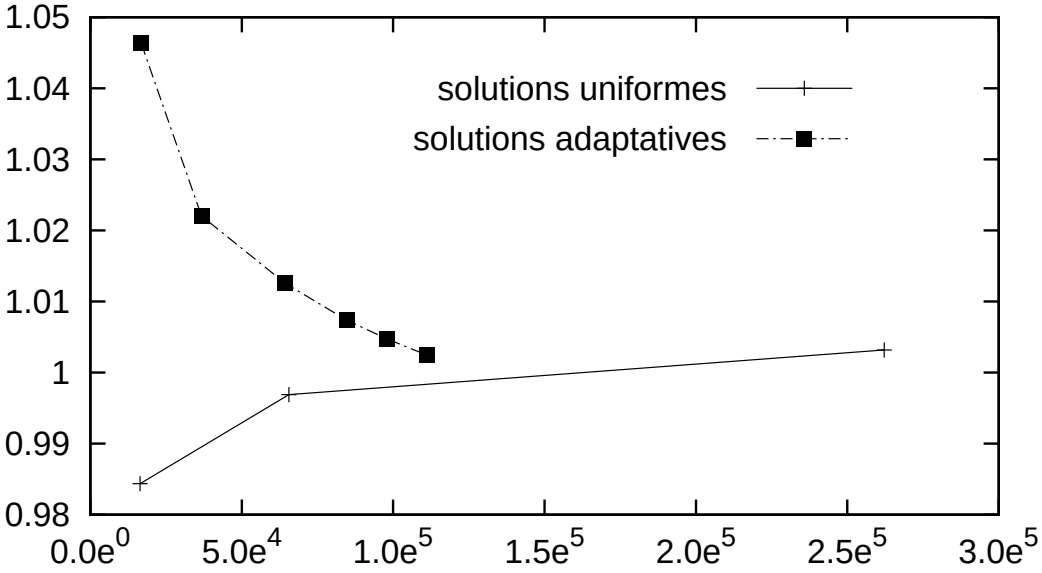


FIG. 7.5 – rapport de masse $\|f^N\|_{L^1}/\|f_L^0\|_{L^1}$ en fonction de la taille des maillages pour diverses solutions uniformes et adaptatives au temps $T = 4.5$.

valeurs de Δt, autrement dit pour différentes qualités de solutions adaptatives. Sur la figure 7.5, on a représenté ce rapport pour des solutions adaptatives et uniformes en fonction de la taille des maillages de calcul associés. Et dans les deux cas, on peut vérifier que la perte de masse tend vers zéro à mesure que la taille des maillages augmente, sans qu'une stratégie se distingue particulièrement.

Remarque 7.1 *On pourra s'étonner du fait que la masse totale des solutions adaptatives est toujours supérieure à la masse de la solution de référence f_L^0. Ceci est dû en réalité au fait que la "perte" de masse la plus importante est réalisée par les interpolations sur les mailles les plus grandes. L'allure particulière du faisceau semi-gaussien, pour lequel la distribution est convexe dans les zones les plus régulières, fait alors que les mailles les plus grandes sont précisément celles où les interpolations affines sont au-dessus de la courbe, entraînant ainsi un "gain" de masse important sur ces zones.*

7.2.2 Précision numérique

Sur les figures 7.6 et 7.7, on a représenté la distance entre f^N et cette solution de référence f_L^N en fonction de Δt. Les distances sont mesurées dans L^∞ sur la figure 7.6 et dans L^1, L^2 sur la figure 7.7 afin de vérifier que notre méthode donne également de bons résultats dans ces distances. Les pentes de ces courbes calculées par une méthode des moindres carrés sont légèrement meilleures que prévues.

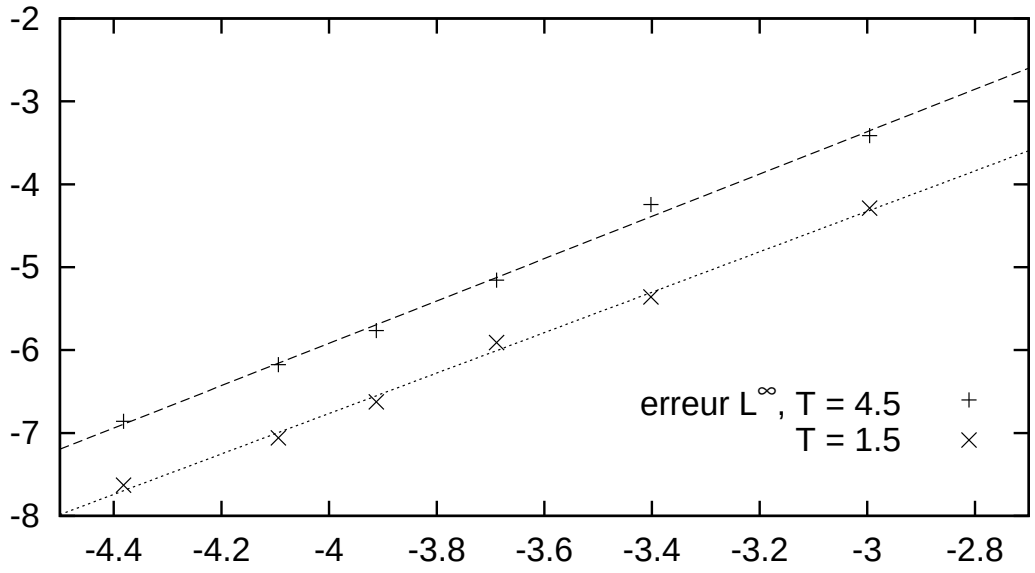


FIG. 7.6 – vitesses de convergence. Erreur numérique $\|f^N - f_L^N\|_{L^\infty}$ en fonction de Δt en échelle log-log, aux instants $T = 1.5$ et $T = 4.5$ (les pentes sont de l'ordre de 2.5).

7.2.3 Complexité optimale des maillages adaptatifs

Pour valider la discussion qu'on a menée dans la section 6.3.2, et en particulier pour éprouver notre conjecture (6.36), on a représenté sur les figures 7.8 et 7.9 les erreurs

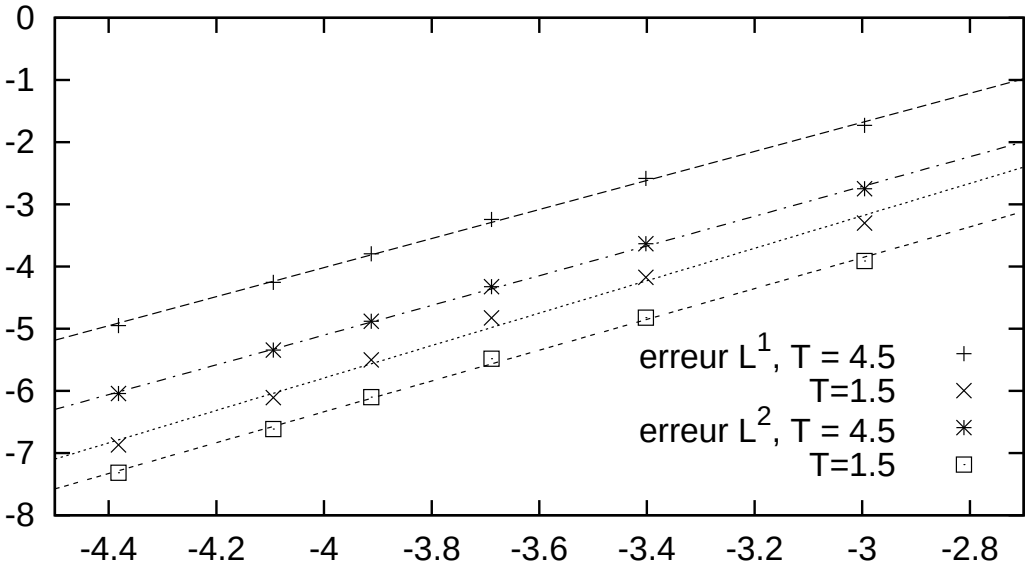


FIG. 7.7 – vitesses de convergence. Erreurs numériques mesurées dans L^1 et L^2 en fonction de Δt en échelle log-log, aux instants $T = 1.5$ et $T = 4.5$ (les pentes sont de l'ordre de 2.5).

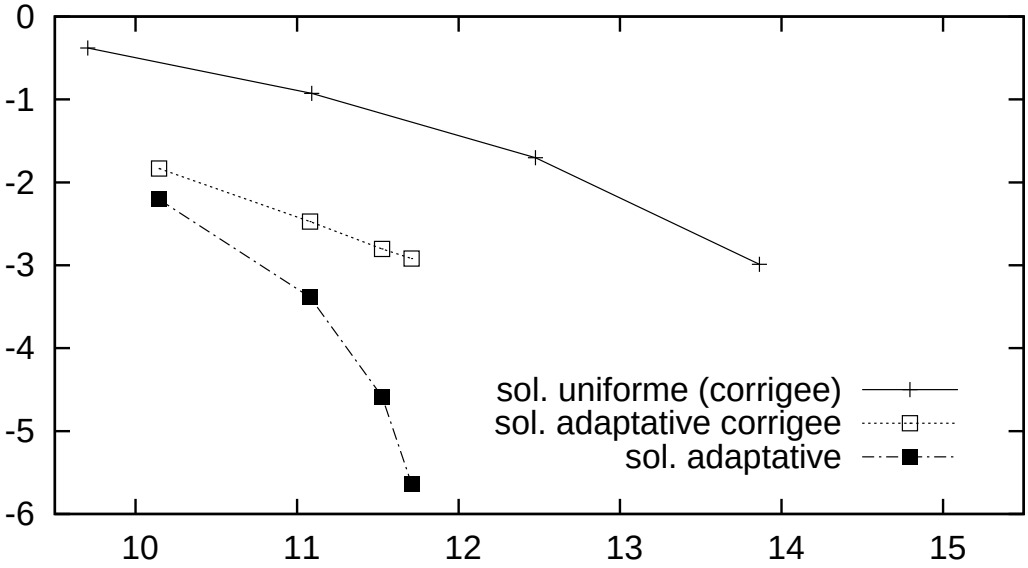


FIG. 7.8 – erreurs simples et corrigées mesurées dans L^∞ en fonction de la taille des maillages en échelle log-log pour des solutions adaptives et uniformes (à $T = 4.5$).

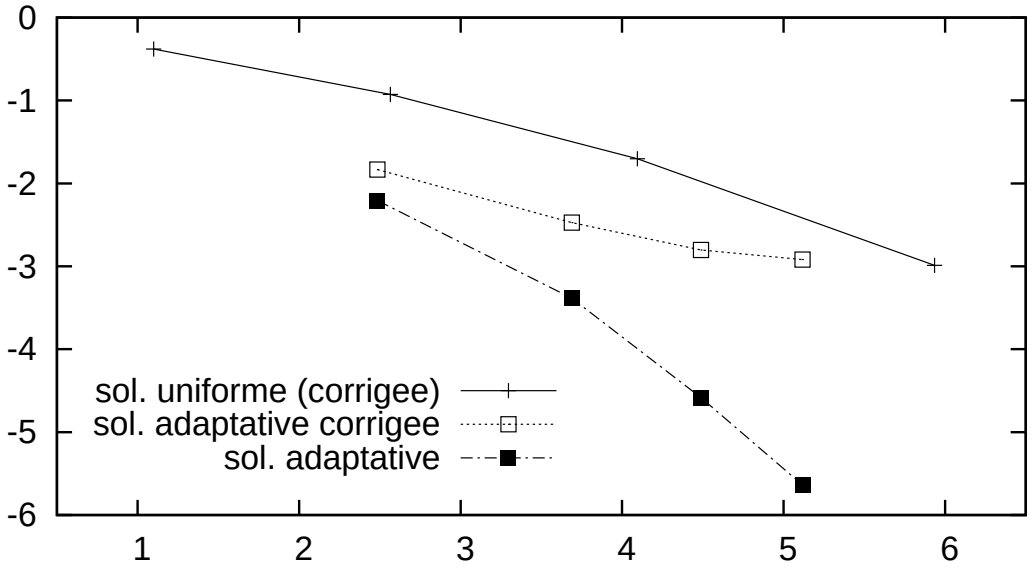


FIG. 7.9 – erreurs simples et corrigées mesurées dans L^∞ en fonction du temps cpu (en minutes) en échelle log-log pour des solutions adaptives et uniformes (à $T = 4.5$).

numériques L^∞ réalisées par les solutions uniformes f_h^N et adaptatives f^N pour différentes valeurs des paramètres h, $\Delta t \sim h^{-2/3}$ et $\varepsilon \sim \Delta t^3$. Sur la figure 7.8, ces erreurs sont tracées en fonction de la taille (6.35) des maillages, et en fonction du temps de calcul sur la figure 7.9.

Ici, toutefois, on ne peut plus se contenter d'évaluer l'erreur numérique en prenant comme référence la solution uniforme f_L^N calculée au niveau le plus fin, car on surestime de cette façon la qualité des solutions adaptatives f^N proches de f_L^N. Pour corriger ces courbes, on a ajouté aux erreurs approchées

$$\tilde{e}^N = \|f^N - f_L^N\|_{L^\infty} \tag{7.1}$$

(représentées par les carrés noirs) une estimation de l'erreur associée à la solution de référence

$$\tilde{e}_L \approx \|f_L^N - f(T)\|_{L^\infty} \tag{7.2}$$

qu'on a évaluée de la façon suivante : ayant observé que les premiers termes de la suite

$$\tilde{e}_\ell := \|f_{h(\ell)}^N - f_L^N\|_{L^\infty}, \quad \ell = \ell_0, \dots, L-1$$

avaient une décroissance quasi-géométrique, on a supposé qu'il en était de même pour la suite des erreurs "exactes" $\|f_{h(\ell)}^N - f(T)\|_{L^\infty}$, et on en a déduit $\tilde{e}_L$ par extrapolation. A coté de la courbe en carrés noirs représentant les pseudo-erreurs (optimistes) $\tilde{e}^N$ réalisées par les solutions adaptatives, on a donc représenté par des carrés blancs les erreurs "corrigées" $\tilde{e}^N + \tilde{e}_L$. Cette fois, les courbes obtenues sont clairement pessimistes, car elles empêchent les solutions adaptatives d'être plus précises que la solution uniforme de niveau L. Les performances réelles de notre schéma sont donc à chercher dans la zone comprise entre les courbes blanches et noires.

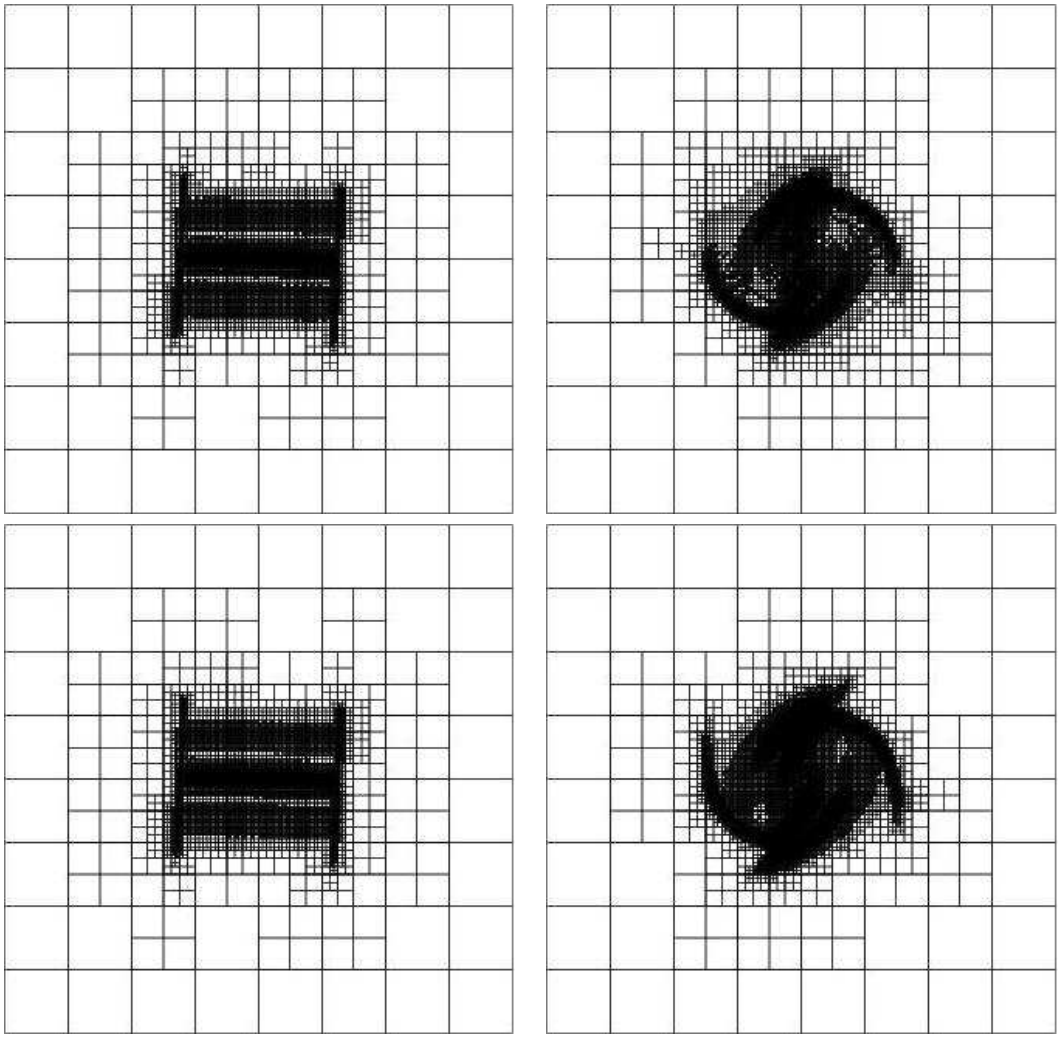


FIG. 7.10 – comparaison des maillages produits par le schéma adaptatif (en haut), et par l'algorithme de compression $\boldsymbol{A}_\varepsilon$ appliqué à la solution uniforme la plus fine (en bas), aux instants $T = 0.05$ (à gauche) et $T = 10.05$ (à droite).

On peut néanmoins mesurer la pente moyenne des courbes corrigées sur la figure 7.8, et observer qu'elle est proche de -0.7, aussi bien pour les solutions adaptatives que pour les solutions uniformes, ce qui valide l'estimation d'erreur uniforme (6.38) comme notre conjecture (6.36). En ce qui concerne le gain d'efficacité, on peut alors observer que pour une précision donnée, les maillages uniformes sont environ 100 fois plus gros que les maillages adaptatifs. Dans les estimations (6.38) et (6.36), ce rapport correspond à la différence entre les "constantes", et en particulier au fait que la courbure totale qui est présente derrière la constante de l'estimation (6.36) est bien plus petite que la semi-norme $W^{2,\infty}$ qui régit l'estimation (6.38)). Malheureusement, on ne retrouve pas un rapport aussi avantageux sur la figure 7.9 qui représente les erreurs en fonction du temps de calcul, ce qui est principalement dû au fait que le schéma adaptatif gère une structure de données complexe, et dépense un temps supérieur au traitement de chaque maille qu'un schéma uniforme. Ainsi, le rapport des temps cpu correspondant

à une erreur corrigée de $0.084 \simeq e^{-2.47}$ n'est que de 4.5. Ce constat, bien que décevant, pourra toutefois être nuancé par le fait que nous n'avons pour l'instant pas cherché à optimiser notre code informatique lui même, et par le fait que la contrainte imposée sur le niveau maximal des cellules ne nous permet pas de bien mettre en valeur la supériorité relative de l'approche adaptative.

Enfin, on a voulu comparer l'allure du maillage "transporté" par notre schéma avec celui qu'on obtiendrait à partir de la solution de référence calculée dans des conditions très proches. Pour réaliser cette mesure, on a commencé par choisir une valeur de Δt (et donc de ε) pour laquelle l'erreur adaptative (7.1) était du même ordre que l'erreur (extrapolée) (7.2), autrement dit pour laquelle les solutions f^N et f_L^N étaient de précisions comparables. On a alors représenté côte à côte sur la figure 7.10 les maillages M^1 et M^{200} produits par notre schéma (6.17) et les maillages $\boldsymbol{A}_\varepsilon(f_L^1)$ et $\boldsymbol{A}_\varepsilon(f_L^{200})$ obtenus en appliquant à la solution de référence l'algorithme (6.3) qui détermine le plus petit maillage dyadique en ε-adéquation avec f_L^N.

Le fait que les maillages M^{200} et $\boldsymbol{A}_\varepsilon(f_L^{200})$ soient de tailles comparables sur la figure 7.10 est donc un signe d'optimalité pratique de notre méthode. En particulier, ceci signifie que l'économie réalisée par l'utilisation d'un nombre réduit de mailles dans le transport n'a que très faiblement modifié la structure de la solution, et surtout que *la stratégie qu'on a mise au point pour faire évoluer ces maillages adaptatifs d'un pas de temps à l'autre a permis de suivre le maillage optimal associé à une solution de référence, et ceci sans avoir eu recours à une technique de raffinement excessif.*

Troisième partie

Analyse des lois de conservation scalaires en distance de Hausdorff

Chapitre 8

Ce qu'il convient de savoir sur les lois de conservation scalaires

On donne dans ce chapitre une présentation rapide des lois de conservation scalaires et de leurs solutions faibles entropiques, en montrant de quelle façon une solution initiale arbitrairement régulière peut devenir discontinue au bout d'un temps fini. En dimension 1, et pour des flux convexes, on rappelle la description lagrangienne proposée par Lax des solutions faibles entropiques, dans laquelle les trajectoires caractéristiques sont déterminées par une formule semi-explicite de minimisation.

8.1 Présentation des lois de conservation scalaires

On s'intéresse dans cette partie aux problèmes d'évolution s'écrivant sous la forme

$$\partial_t u(t,x) + \nabla_x [f(u(t,x))] = 0, \qquad u(0,\cdot) = u_0, \qquad t > 0, \qquad x \in \mathbb{R}^d, \tag{8.1}$$

et nos résultats concernent plus particulièrement leur version uni-dimensionnelle

$$\partial_t u(t,x) + \partial_x [f(u(t,x))] = 0, \qquad u(0,\cdot) = u_0, \qquad t > 0, \qquad x \in \mathbb{R}. \tag{8.2}$$

Dans ces équations, le *flux* f est une fonction connue et régulière de $\mathbb{R}$ dans $\mathbb{R}^d$, et l'inconnue $u(t,\cdot)\colon \mathbb{R}^d \to \mathbb{R}$ peut être vue comme une densité de masse. Ses valeurs sont *scalaires*, à la différence de ce qui se passe lorsqu'on étudie un *système* de lois de conservation. Parmi les nombreux ouvrages de référence sur ce sujet, citons (outre l'article fondateur de Kružkov [44], relativement technique), ceux de Lax [45], de Godlewski et Raviart [35, 36], de Serre [58, 57] (en français) ou plus récemment de LeFloch [46].

Une façon naturelle de voir l'équation (8.1) est de l'intégrer sur un domaine régulier ω de $\mathbb{R}^d$. On obtient alors

$$\partial_t \int_\omega u(t,x)\,\mathrm{d}x + \int_{\partial\omega} f(u(t,x)) \cdot \overrightarrow{n}\,\mathrm{d}\sigma(x) = 0, \qquad t > 0, \tag{8.3}$$

où $\overrightarrow{n}$ désigne le vecteur normal au bord $\partial\omega$ et $\mathrm{d}\sigma$ sa mesure surfacique. On peut interpréter (8.3) de la façon suivante : entre les instants t et t', la variation de la masse contenue dans le domaine ω correspond au flux du champ $f(u)$ au travers de $\partial\omega$, ce qui traduit bien un phénomène de transport. A partir du moment où f est continue, la

masse sortant du domaine ω et celle qui entre dans son complémentaire ω^{c} sont égales, de sorte que la masse "totale" est conservée.

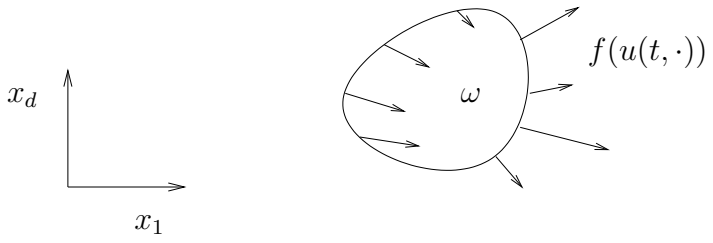


FIG. 8.1 – bilan instantané local sur un domaine ω.

On se représentera mieux la nature des solutions en voyant (8.1) comme une équation de transport "ponctuelle" dans laquelle la densité u est préservée le long des trajectoires caractéristiques $t \to X(t) = X(t;y) \in \mathbb{R}^d$ solutions de

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = f'(u(t,X)) = (f_1'(u(t,X)), \cdots, f_d'(u(t,X))), \qquad X(0) = y \in \mathbb{R}^d, \tag{8.4}$$

y désignant le point de départ cette trajectoire. Tant que ces trajectoires existent, on peut réécrire (8.1) sous la forme

$$\frac{\mathrm{d}u(t,X(t;y))}{\mathrm{d}t} = 0, \quad \text{pour tout} \quad y \in \mathbb{R}^d, \tag{8.5}$$

et remarquer que cela entraîne $\frac{\mathrm{d}X(t)}{\mathrm{d}t} = f'(u_0(y))$, ce qu'on peut aussi voir comme une conséquence du fait que la fonction de flux f ne dépend pas du temps (du moins pas autrement qu'au travers de u). Notre loi de conservation (8.1) est donc équivalente à la propriété de conservation (8.5) le long des trajectoires caractéristiques données par

$$X(t;y) = y + tf'(u_0(y)), \tag{8.6}$$

ce qui revient à écrire que pour tout couple $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$,

$$u(t,x) = u_0(y), \text{ où } y = y(t,x) \text{ est solution de } y + tf'(u_0(y)) = x, \tag{8.7}$$

$f'(u_0)\colon \mathbb{R}^d \to \mathbb{R}^d$ jouant ainsi le rôle d'un champ de vitesses constant.

Malheureusement on va bientôt s'apercevoir que l'équation (8.7), qui peut nous donner une formule explicite pour construire les solutions, n'est pas satisfaisante en dehors de quelques cas faciles comme celui où f' est constante.

Une des lois non-linéaires les plus simples s'obtient, en une dimension d'espace, avec le flux $f(u) = u^2/2$. Cela correspond à l'équation de Burgers *sans viscosité*

$$\partial_t u(t,x) + u(t,x) \cdot \partial_x u(t,x) = 0, \qquad u(0,\cdot) = u_0. \tag{8.8}$$

Pour ce flux, la formulation (8.7) devient

$$u(t,x) = u_0(y), \text{ où } y = y(t,x) \text{ est solution de } y + tu_0(y) = x. \tag{8.9}$$

8.1.1 Défauts d'existence ou d'unicité

Dans la partie précédente, on a pu voir que l'équation de Vlasov possédait des solutions continues lorsque les conditions initiales étaient elle-même continues. En particulier, on pouvait montrer que les trajectoires caractéristiques associées à de telles solutions étaient définies en tout temps.

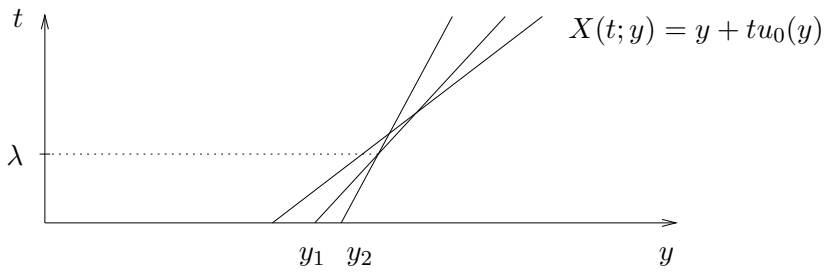


FIG. 8.2 – trajectoires caractéristiques associées à la donnée initiale de la figure 8.3 pour l'équation de Burgers en dimension 1. A partir de l'instant λ, certaines trajectoires se croisent et les courbes obtenues sont multivaluées.

La formulation (8.7), qui repose sur une définition implicite de y, ne permet de construire $u(t,\cdot)$ que lorsque l'application $y \to y + tf'(u_0(y))$ est inversible. Lorsque $f'(u_0)$ est lipschitzienne, on peut voir cette application comme une perturbation de l'identité qui sera inversible sur des petites valeurs de t, mais a priori pas pour des temps grands. La figure 8.2 nous donne une représentation "physique" de ce phénomène en représentant les trajectoires caractéristiques $X(t;y) = y + tu_0(y)$ associées à l'équation de Burgers pour une donnée initiale u_0 très régulière, tracée sur la figure 8.3, en haut. Sur la figure 8.2, les temps t pour lesquelles l'application $y \to y + tu_0(y)$ est inversible se reconnaissent au fait que les trajectoires caractéristiques ne se croisent pas entre 0 et t. A l'inverse, le fait que deux trajectoires se croisent en (t,x) indique une ambiguïté sur le point de départ y et la vitesse $u_0(y)$ permettant d'arriver en (t,x).

Plus généralement si f est convexe, cette situation se produira dès que u_0 (en dimension 1) n'est pas croissante. Il suffit en effet que deux positions initiales y_1 et y_2 vérifient

$$y_1 - y_2 = \lambda[f'(u_0(y_2)) - f'(u_0(y_1))] \text{ avec un } \lambda > 0 \tag{8.10}$$

pour que les trajectoires issues de y_1 et y_2 se croisent en $t = \lambda$ et $x = y_1 + \lambda f'(u_0(y_1)) = y_2 + \lambda f'(u_0(y_2))$. Si l'on observe la solution u, que constate-t-on ? Tant que les trajectoires ne se croisent pas, l'équation (8.7) permet de définir $u(t,\cdot)$ de façon univoque, mais à partir de $t = -1/\left[\inf_{y\in\mathbb{R}} \left(f''(u_0(y))u_0'(y)\right)\right] > 0$, la "solution" donnée par cette formule n'est plus une fonction, mais un graphe multivalué

$$G(t) = \{(y + tu_0(y), u_0(y)) : y \in \mathbb{R}\} \tag{8.11}$$

représenté sur la figure 8.3.

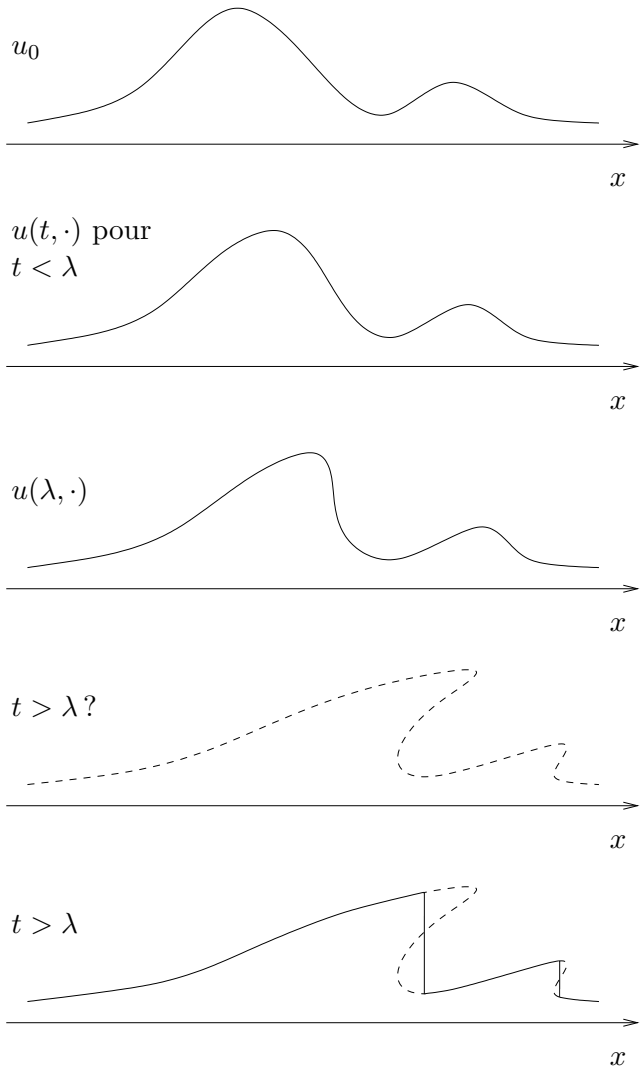


FIG. 8.3 – évolution d'une solution classique pour l'équation de Burgers jusqu'à l'apparition du premier choc. Après cet instant, la solution faible se distingue du graphe multivalué construit par l'équation (8.9).

8.1.2 Solutions faibles entropiques

A partir de ces courbes multivaluées, la façon la plus naturelle de construire des solutions qui soient des fonctions consiste à les faire "s'effondrer" sur elles-mêmes, en remplaçant les plis du graphe par des discontinuités (cette construction est d'ailleurs à la base d'un schéma numérique proposé par Brenier, voir [12]). Comme la surface définie sous le graphe (8.11) est constante, et en particulier égale à la masse initiale $\int_{\mathbb{R}} u_0(x)\,\mathrm{d}x$, la position de ces discontinuités doit être déterminée de façon à préserver cette masse, comme illustré sur la figure 8.3, en bas. Mais on reconnaîtra volontiers que cette construction, si elle permet de tracer l'allure des solutions, est loin d'être satisfaisante pour définir les solutions exactes.

Il nous faut donc donner un sens précis à l'équation (8.1) pour des solutions discontinues. Lorsque $u_0 \in L^\infty(\mathbb{R}^d)$, on dit que $u \in L^\infty(\mathbb{R}_+ \times \mathbb{R}^d)$ est une *solution faible* du problème de Cauchy (8.1) si

$$\int_0^\infty \int_{\mathbb{R}^d} (u(t,x)\partial_t\varphi(t,x) + f(u(t,x)) \cdot \nabla_x\varphi(t,x))\,\mathrm{d}x\,\mathrm{d}t + \int_{\mathbb{R}^d} u_0(x)\varphi(0,x)\,\mathrm{d}x = 0$$

pour toute fonction $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}_+ \times \mathbb{R}^d)$. Cette définition, toutefois, n'assure pas à elle seule *l'unicité* de u pour une solution initiale u_0 donnée. Pour y parvenir, on fait l'hypothèse supplémentaire que la loi de conservation (8.1) représente une sorte de limite, lorsque $\varepsilon > 0$ tends vers 0, de l'équation visqueuse

$$\partial_t u^\varepsilon(t,x) + \nabla_x \cdot [f(u^\varepsilon(t,x))] = \varepsilon\Delta_x u^\varepsilon(t,x).$$

Plus précisément, on montre que la solution u^ε de cette équation est définie de façon unique, et possède suffisamment de régularité pour que l'on puisse en extraire une sous-suite convergeant presque partout vers une solution faible (unique) de (8.1). Pour sélectionner cette solution "physique" parmi toutes les solutions possibles, on a recours à la notion d'entropie mathématique : par *fonction d'entropie*, on entend une fonction $E\colon \mathbb{R} \to \mathbb{R}$ de classe $\mathcal{C}^1$, *convexe*, à laquelle on associe un *flux d'entropie* $F\colon \mathbb{R} \to \mathbb{R}^d$ vérifiant

$$F' = E'f'. \tag{8.12}$$

Si u est une solution $\mathcal{C}^1$ de (8.1), elle vérifie

$$\begin{aligned}\partial_t[E(u)] + \nabla_x[F(u)] = E'(u)\partial_t u + F'(u) \cdot \nabla_x u &= E'(u)[\partial_t u + f'(u) \cdot \nabla_x u] \\ &= E'(u)[\partial_t u + \nabla_x(f(u))] = 0. \end{aligned} \tag{8.13}$$

Quant à la solution faible obtenue par passage à la limite des solutions u^ε, on peut montrer qu'elle vérifie au sens des distributions

$$\partial_t E(u) + \nabla_x[F(u)] \leq 0 \text{ pour toute fonction d'entropie } E. \tag{8.14}$$

L'idée présente derrière les résultats d'unicité consiste en quelque sorte à "remonter" cet argument, en montrant qu'une solution faible u qui vérifie la condition d'entropie (8.14) correspond à la solution donnée par la méthode de viscosité évanescente, et en particulier, elle est définie de façon unique. On appelle *solution faible entropique* de (8.1) une telle fonction u, et on peut citer le résultat suivant (voir en particulier [36]) :

Théorème 8.1 *Si u_0 appartient à $L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$, l'équation* (8.1) *admet une unique solution faible entropique $u \in L^\infty(\mathbb{R}_+; L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d))$. Cette solution vérifie*

1. *un principe du maximum L^∞ en temps*

$$\|u(t,\cdot)\|_{L^\infty} \leq \|u_0\|_{L^\infty}, \tag{8.15}$$

2. *une propriété de convergence L^1 vers la donnée initiale*

$$\|u(t,\cdot) - u_0\|_{L^1} \to 0 \text{ lorsque } t \to 0, \tag{8.16}$$

3. *une propriété de contraction* L^1

$$\|u(t,\cdot) - v(t,\cdot)\|_{L^1} \leq \|u_0 - v_0\|_{L^1} \tag{8.17}$$

pour toute solution entropique v *issue d'une donnée initiale* $v_0 \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$,

4. *et une propriété de monotonie*

$$u(t,\cdot) \leq v(t,\cdot) \tag{8.18}$$

pour toute solution entropique v *de* (8.1) *issue d'une donnée initiale* $v_0 \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ *vérifiant* $u_0 \leq v_0$.

Enfin si u_0 *est à variations bornées, alors* $u(t,\cdot)$ *l'est également, et on a*

$$|u(t,\cdot)|_{BV} \leq |u_0|_{BV}. \tag{8.19}$$

Remarque 8.1 *La propriété* (8.17) *de contraction* L^1 *a été démontrée en 1970 par Kružkov (voir [44]). C'est un résultat fondamental qui permet d'une part d'établir l'unicité des solutions entropiques, et d'autre part la propriété* (8.19). *On peut en effet définir la variation totale d'une fonction* v *comme la plus petite constante* C *pour laquelle* $\|v - v(\cdot - h)\|_{L^1} \leq Ch$ *est vérifiée pour tout* $h \in \mathbb{R}^d$. *Dans la mesure où la solution correspondant à une translation* $u_0(\cdot - h)$ *de la donnée initiale s'obtient par une translation identique* $u(t, \cdot - h)$ *de la solution issue de* u_0, *on déduit de* (8.17) *que*

$$\|u(t,\cdot) - u(t,\cdot - h)\|_{L^1} \leq \|u_0(\cdot) - u_0(\cdot - h)\|_{L^1} \leq h|u_0|_{BV} \tag{8.20}$$

pour tout $h \in \mathbb{R}^d$, *d'où l'inégalité* (8.19).

8.2 Une formule semi-explicite pour les lois uni-dimensionnelles à flux convexes

On se place ici en dimension $d = 1$, avec un flux f de classe $\mathcal{C}^2$ vérifiant

$$f'' > 0 \quad \text{et} \quad \lim_{\pm\infty} f' = \pm\infty. \tag{8.21}$$

Dans ces conditions, Lax démontre dans [45] que la solution faible entropique de l'équation (8.2) (dont l'existence et l'unicité sont établies par le théorème précédent), vérifie, pour tout $t > 0$ et tout $x \in \mathbb{R}$,

$$u(t,x) = u_0(y), \tag{8.22}$$

où $y = y(t,x)$ minimise globalement la fonctionnelle $z \to \mathcal{L}(z,t,x) = \mathcal{L}_{u_0,f}(z,t,x)$ définie par

$$\mathcal{L}_{u_0,f}(z,t,x) := \int_0^z u_0(s)\,\mathrm{d}s + tf^*\left(\frac{x-z}{t}\right) \tag{8.23}$$

(voir (8.28) pour le cas où u_0 est discontinue en y). La fonction f^* désigne ici la transformée de Legendre de f

$$f^*(x) := \sup_{y\in\mathbb{R}}(xy - f(y)).$$

Comme f' est strictement croissante, la borne supérieure est atteinte en $y = (f')^{-1}(x)$, d'où l'on déduit que

$$(f^*)'(x) = (f')^{-1}(x) + x[(f')^{-1}]'(x) - f'\big((f')^{-1}(x)\big)[(f')^{-1}]'(x) = (f')^{-1}(x). \tag{8.24}$$

Remarque 8.2 *Lorsque le flux vérifie* (8.21)*, la condition d'entropie* (8.14) *peut être remplacée par l'inégalité suivante, appelée* condition d'entropie d'Oleinik *:*

$$f'(u(t,x_2)) - f'(u(t,x_1)) \leq \frac{x_2 - x_1}{t} \text{ pour tout } x_1 \leq x_2 \text{ et } t > 0. \tag{8.25}$$

8.2.1 Comportement des trajectoires caractéristiques en présence de chocs

On peut vérifier que la formulation (8.22)-(8.23) nous permet de retrouver (8.7), du moins lorsque cette dernière a un sens. Le fait que y minimise $\mathcal{L}(\cdot,t,x)$ entraîne en effet

$$u_0(y^-) - (f^*)'\left(\frac{x-y}{t}\right) = \partial_z \mathcal{L}(y^-,t,x) \leq 0 \leq \partial_z \mathcal{L}(y^+,t,x) = u_0(y^+) - (f^*)'\left(\frac{x-y}{t}\right), \tag{8.26}$$

d'où l'on déduit que $u_0(y) = (f^*)'\left(\frac{x-y}{t}\right)$ lorsque u_0 est continue en y. Compte tenu de (8.24), ceci entraîne

$$x = y + tf'(u_0(y)), \tag{8.27}$$

et peut donc se lire comme "y est le point de départ d'une trajectoire de vitesse $f'(u_0(y)) = f'(u(t,x))$ passant par (t,x)". Lorsque u_0 est discontinue en y, on remplacera donc (8.22) par

$$u(t,x) = (f')^{-1}\left(\frac{x-y}{t}\right), \tag{8.28}$$

et (8.26) nous apprend que dans ce cas $u(t,x) \in [u_0(y^-), u_0(y^+)]$, autrement dit qu'il s'agit d'une discontinuité *croissante* de u_0. Sur la figure 8.4, ce phénomène se produit au point $a = (y_a, t = 0)$, à partir duquel se propage un "éventail" complet de trajectoires $\{\tau \to (\tau, y + \tau f'(\sigma)) : \sigma \in [u_0(y^-), u_0(y^+)]\}$ correspondant aux vitesses comprises entre $f'(u_0(y^-))$ et $f'(u_0(y^+))$. Pour des temps $t > 0$, en revanche, l'inégalité d'Oleinik (8.25) nous apprend que les solutions $u(t,\cdot)$ n'ont plus de discontinuités croissantes. A l'inverse, des solutions *décroissantes* peuvent se propager dans les solutions faibles. On les appelle des chocs, et une façon d'obtenir un choc est de faire aboutir en un "point" (t,x) de la figure 8.2 deux trajectoires distinctes, sur lesquelles se propagent forcément deux valeurs distinctes de la solution.

L'intérêt de la formulation "semi-explicite" (8.22)-(8.23) de Lax est qu'elle permet de décrire de façon complète les trajectoires caractéristiques en présence de telles discontinuités. Ainsi, lorsque plusieurs "trajectoires potentielles" $\tau \to (y_i + \tau f'(u_0(y_i)), \tau)$, $i = 1, 2, \ldots$ sont susceptibles de se rencontrer en un point (t,x), on interprétera les différents points de départs y_i comme des *extrema locaux* de $\mathcal{L}(\cdot,t,x)$. Sur la figure 8.4, ce phénomène se produit par exemple aux points b et c. Mais sur cet exemple, seul le point c correspond à un choc. En effet, la trajectoire issue du point y_3 est "absorbée" dans l'onde de choc tracée en pointillés, et n'atteint donc pas le point b. Pour départager ces candidats y_i, la formule de Lax leur demande d'être des *minimiseurs globaux*. Ainsi, le point $c = (x_c, t_c)$ sur la figure 8.4 correspond au cas où le minimum global de $\mathcal{L}(\cdot, t_c, x_c)$ est atteint en deux valeurs distinctes y_2 et y_3. La solution $u(t_c,\cdot)$ est donc discontinue en x_c et vérifie

$$u(t_c, x_c^-) = u_0(y_2) > u_0(y_3) = u(t_c, x_c^+). \tag{8.29}$$

De façon plus générale, les limites à gauche et à droite de $u(t,\cdot)$ en x sont respectivement données par le plus petit et le plus grand minimiseur global de $\mathcal{L}(\cdot,t,x)$.

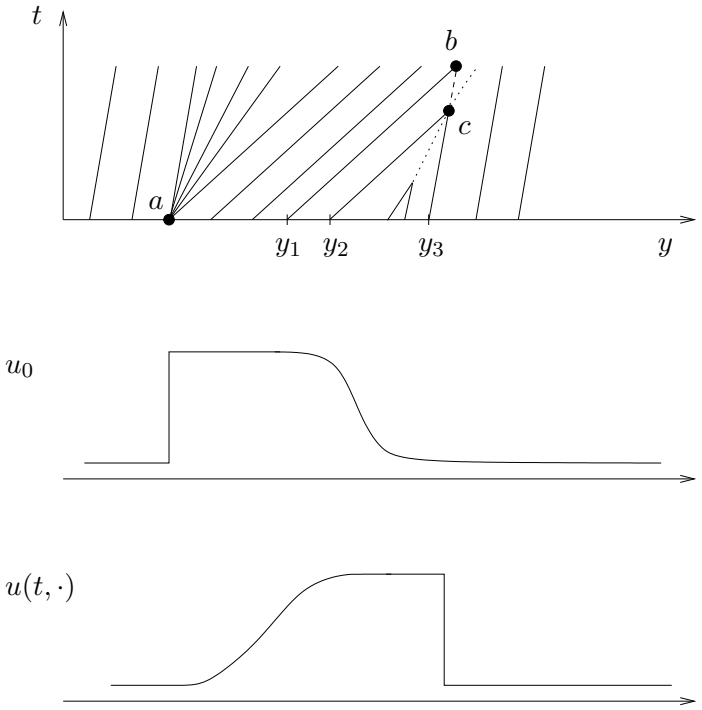


FIG. 8.4 – comportement des trajectoires caractéristiques en présence d'un choc.

8.2.2 Un petit calcul instructif

On peut conclure cette présentation par une observation intéressante, qui nous donnera un avant-goût du genre de calculs qu'on sera amené à faire dans la suite. Cette observation concerne un principe général vérifié par les solutions entropiques, selon lequel *il est toujours possible de prolonger les trajectoires caractéristiques dans le passé* (ce qui souligne au passage le caractère irréversible des solutions entropiques). On a pu ainsi observer que certaines trajectoires pouvaient *entrer* dans l'onde de choc tracée en pointillés sur la figure 8.4, mais qu'aucune n'en *sortait*. D'après ce principe, on voit facilement que pour t fixé, deux trajectoires passant par (t,x) et (t,x') avec $x \leq x'$ sont respectivement parties de $(0,y)$ et $(0,y')$ avec $y \leq y'$. Autrement dit, l'application $x \to y(t,x)$ est croissante.

En montrant que les minimiseurs de $\mathcal{L}(\cdot,t,x')$ sont toujours supérieurs à ceux de $\mathcal{L}(\cdot,t,x)$ lorsque $x > x'$ (propriété que tente d'illustrer la figure 8.5), il est possible de *retrouver* cette propriété importante à partir de la formule de Lax. Plus précisément, on peut établir que si y minimise $\mathcal{L}(\cdot,t,x)$, la quantité

$$K(z) := \mathcal{L}(z,t,x') - \mathcal{L}(y,t,x')$$

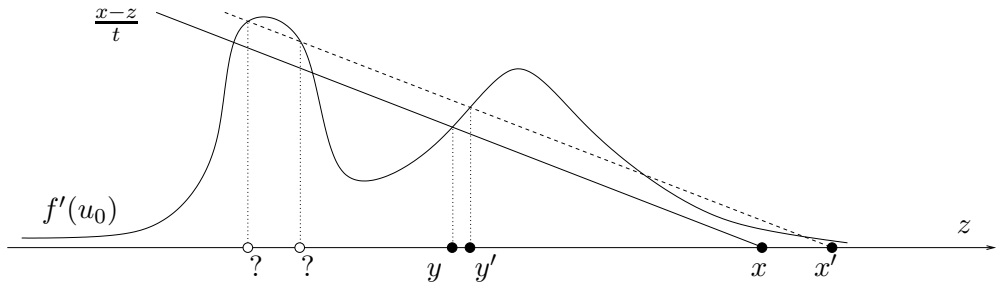


FIG. 8.5 – illustration du fait que l'application $x \to y(t,x)$ définie par la formule de Lax est croissante.

est strictement positive pour tout $z < y$. On a en effet pour tout z

$$\begin{aligned} K(z) &\geq \mathcal{L}(z,t,x') - \mathcal{L}(z,t,x) - \mathcal{L}(y,t,x') + \mathcal{L}(y,t,x) \\ &= t\big[f^*\left(\frac{x'-z}{t}\right) - f^*\left(\frac{x-z}{t}\right) - f^*\left(\frac{x'-y}{t}\right) + f^*\left(\frac{x-y}{t}\right)\big] \\ &= \int_x^{x'} \big[(f^*)'\left(\frac{u-z}{t}\right) - (f^*)'\left(\frac{u-y}{t}\right)\big]\,\mathrm{d}u. \end{aligned} \tag{8.30}$$

Lorsque $z < y$ et $x < x'$, ceci se minore par

$$K(z) \geq \int_x^{x'} \int_z^y \frac{1}{t}(f^*)''\left(\frac{u-v}{t}\right)\,\mathrm{d}v\,\mathrm{d}u \geq \frac{(x'-x)(y-z)}{t}\inf(f^*)''. \tag{8.31}$$

Comme on s'est placé dans le cas où f était strictement convexe (8.21), on peut ensuite utiliser (8.24) pour voir que $(f^*)''(w) = 1/f''\big((f')^{-1}(w)\big)$. On en déduit que

$$\inf(f^*)'' = \frac{1}{\sup f''}, \tag{8.32}$$

et quitte à supposer que f'' est majorée par une constante, on voit la borne inférieure de $(f^*)''$ est strictement positive, et finalement que $K(z)$ est bien strictement positive pour tout $z < y$.

Chapitre 9

Stabilité des solutions en distance de Hausdorff

On introduit à présent une distance définie entre deux fonctions à partir de la distance ensembliste de Hausdorff. L'intérêt principal de cette "nouvelle" distance (qui a notamment été étudiée par Sendov [56]) est qu'elle permet de considérer comme un problème bien posé l'approximation *uniforme*, c'est-à-dire sans oscillations, d'une fonction discontinue. D'une certaine façon, les résultats que nous avons obtenus (lors d'un travail effectué en collaboration avec Albert Cohen, Wolfgang Dahmen et Ronald DeVore) garantissent que dans cette distance, l'approximation des solutions de lois de conservations scalaires est elle-même un problème bien posé. Plus précisément, on montre que sous certaines hypothèses, les lois de conservation scalaires uni-dimensionnelles à flux convexes sont stables en distance de Hausdorff, au sens où les graphes des solutions s'écartent avec une vitesse au plus linéaire. La preuve de ce résultat exploite de façon importante la description semi-explicite donnée par Lax des trajectoires caractéristiques pour des flux convexes, et suppose d'autre part que les solutions initiales possèdent une régularité "semi-lipschitzienne", hypothèse nécessaire et également considérée par Tadmor et Tang dans [61], [62] et [63]. Par des raisonnements proches, on établit un résultat de stabilité des solutions vis-à-vis de perturbations lipschitziennes de la fonction de flux. Dans les cas que ces résultats ne couvrent pas, notamment lorsque le flux est non convexe ou dans le cas de problèmes multi-dimensionnels ne pouvant pas se ramener à un problème uni-dimensionnel, on construit des exemples de solutions pour lesquelles la stabilité n'est pas vérifiée.

9.1 Distance de Hausdorff entre deux fonctions

La raison principale pour laquelle on introduit cette nouvelle distance, qui ne correspond à aucune norme, est donc l'*insuffisance notoire de la distance L^∞ pour mesurer la qualité des approximations lorsque la fonction u qu'on souhaite approcher est discontinue*. De façon évidente, une méthode qui utilise des approximants continus ne pourra jamais converger vers u dans L^∞. Et même pour des approximants discontinus, la distance L^∞ est en général trop "rigide" pour que l'on puisse obtenir des propriétés intéressantes. On peut penser par exemple à la situation où l'on approche u par des approximants $u_M = \sum_{\alpha\in M} c_\alpha \chi_\alpha$ constants sur une partition dyadique arbitraire M de $\mathbb{R}^d$. Si u possède un point de discontinuité m dont aucune coordonnée m_i, $1 \leq i \leq d$,

n'est une fraction dyadique $\frac{k}{2^j}$, on peut observer que l'erreur d'approximation dans L^∞ sera toujours supérieure au demi-saut de u en ce point, *i.e.*

$$\|u - u_M\|_{L^\infty} \geq \frac{1}{2}(\overline{u} - \underline{u})(m) \tag{9.1}$$

où les fonctions $\overline{u}$ et $\underline{u}$ définies en (9.9) désignent respectivement la plus grande et la plus petite valeur d'adhérence de u en m, et ceci *quelle que soit* M !

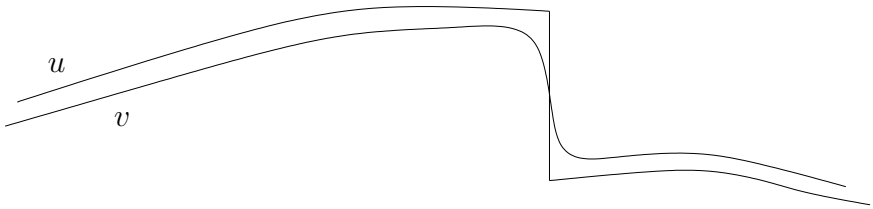


FIG. 9.1 – exemple de "mauvaise" approximation de u dans L^∞.

Pour approcher des fonctions discontinues, on utilise donc en général des normes L^p d'exposant p fini, qui sont moins sévères car elles mesurent des *erreurs moyennes*. En contrepartie, la qualité des approximations obtenues de cette façon n'est pas uniforme, et il est tout à fait possible que v soit très proche de u dans L^p tout en oscillant fortement comme cela se produit sur la figure 9.2.

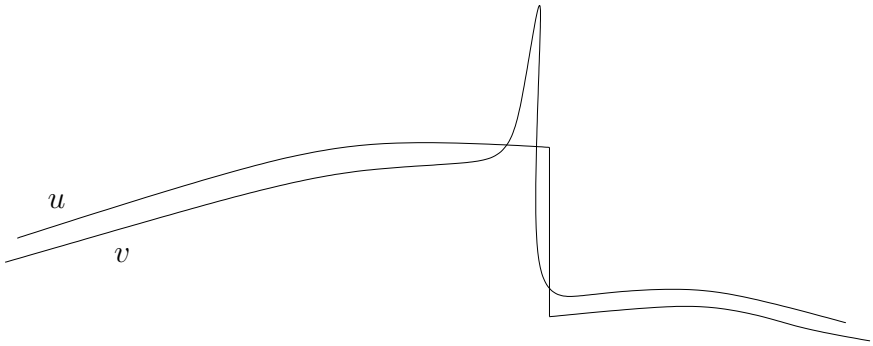


FIG. 9.2 – exemple de "bonne" approximation de u dans L^p lorsque $p < \infty$.

Dans ces conditions, la distance fonctionnelle de Hausdorff offre une alternative idéale aux distances L^p :

- d'une part, c'est une distance *uniforme*, qui est autant sensible aux oscillations que peut l'être la distance L^∞, en particulier elle pénalisera un phénomène de Gibbs tout autant que la distance L^∞.
- d'autre part, et c'est une différence essentielle avec la distance L^∞, elle est *souple*, au sens où elle permet que de bonnes approximations de u n'aient pas leurs discontinuités qui coïncident avec celles de u.

A ces deux qualités, on peut en ajouter une troisième, qui rend bien compte du caractère "visuellement satisfaisant" de la distance fonctionnelle de Hausdorff. Lorsque u est une

fonction très oscillante comme $\sin(\frac{x}{\varepsilon})$ avec $\varepsilon \ll 1$, on aurait tendance à dire que u ressemble beaucoup à $-u$, et en tout cas, qu'elle est bien plus proche de $-u$ que de la fonction nulle (voir figure 9.3). Dans la mesure où $\|u-0\| \leq \|u-(-u)\|$ pour n'importe quelle norme, on voit qu'une distance associée à une norme fonctionnelle est incapable de traduire cette propriété. En revanche, on aura

$$d_H(u,-u) \leq \pi\varepsilon \ll 1 = d_H(u,0) \tag{9.2}$$

avec la distance de Hausdorff d_H.

FIG. 9.3 – en distance de Hausdorff, u est bien plus proche de $-u$ que de la fonction nulle. Avec une distance associée à une norme, c'est toujours le contraire.

Commençons par rappeler ce qu'est la distance de Hausdorff entre deux ensembles.

9.1.1 Distance de Hausdorff entre deux ensembles (fermés)

Si A et B sont deux ensembles d'un espace métrique complet (X,δ), on désigne par

$$\delta_H(A,B) := \max\left\{\sup_{a\in A}\inf_{b\in B}\delta(a,b),\ \sup_{b\in B}\inf_{a\in A}\delta(a,b)\right\} \tag{9.3}$$

la distance de Hausdorff entre A et B. Pour se représenter graphiquement cette distance, on peut considérer l'*éloignement dissymétrique de A par rapport à B*

$$\epsilon(A,B) := \sup_{a\in A}\inf_{b\in B}\delta(a,b), \tag{9.4}$$

quantité dissymétrique au sens où A peut être "plus éloigné de B que B n'est éloigné de A", comme c'est le cas sur la figure 9.4. La distance de Hausdorff s'exprime alors comme

$$\delta_H(A,B) = \max\{\epsilon(A,B),\epsilon(B,A)\}.$$

Elle est clairement symétrique, et elle satisfait l'inégalité triangulaire car on a pour tout ensemble C

$$\epsilon(A,B) \leq \sup_{a\in A}\inf_{c\in C}\left[\delta(a,c)+\inf_{b\in B}\delta(c,b)\right] \leq \epsilon(A,C)+\epsilon(C,B)$$

d'où l'on déduit que

$$\begin{aligned}\delta_H(A,B) &= \max\{\epsilon(A,B),\epsilon(B,A)\} \leq \max\{\epsilon(A,C)+\epsilon(C,B),\epsilon(B,C)+\epsilon(C,A)\}\\ &\leq \max\{\epsilon(A,C),\epsilon(C,A)\}+\max\{\epsilon(B,C),\epsilon(C,B)\} = \delta_H(A,C)+\delta_H(C,B).\end{aligned}$$

Finalement, on peut observer que $\epsilon(A,B) = 0$ est équivalent à $A \subset \overline{B}$, de sorte que δ_H *est bien une distance sur les ensembles fermés.*

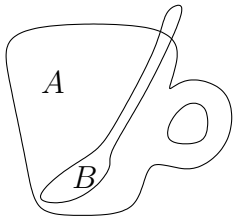


FIG. 9.4 – au sens de l'éloignement dissymétrique (9.4), l'ensemble A est ici "plus éloigné de B que B n'est éloigné de A".

9.1.2 Distance de Hausforff entre les graphes

Pour deux fonctions u et v données de $\mathbb{R}^d$ à valeurs dans $\mathbb{R}$, on définit leur distance de Hausdorff par

$$d_H(u,v) := \delta_H(G_u, G_v), \tag{9.5}$$

où G_u et G_v désignent leurs graphes respectifs "complétés", en un sens qu'on va préciser tout de suite, et où la distance ensembliste δ_H est prise sur les fermés de $\mathbb{R}^{d+1}$ muni de la distance

$$\delta((x_1,\cdots,x_{d+1}),(x'_1,\cdots,x'_{d+1})) := \max\left\{\left(\sum_{i=1}^{d}(x_i-x'_i)^2\right)^{1/2}, |x_{d+1}-x'_{d+1}|\right\}. \tag{9.6}$$

Lorsque u est une fonction continue, son graphe

$$G_u = \{(x,u(x)) : x \in \mathbb{R}^d\} \tag{9.7}$$

est un fermé de $\mathbb{R}^{d+1}$. Ce n'est plus le cas lorsque u est discontinue, et on ne saurait se satisfaire d'en prendre simplement l'adhérence $\overline{G}_u$, car celle-ci n'est en général pas connexe. En dimension 1, on peut par exemple penser au cas où u est la fonction de Heaviside $\chi_{\mathbb{R}_+}$: $\overline{G}_u$ est alors la réunion des demi-droites fermées $\mathbb{R}_- \times \{0\}$ et $\mathbb{R}_+ \times \{1\}$. Une fonction v telle que $\delta_H(\overline{G}_u, \overline{G}_v) < 1/2$ ne pouvant pas avoir un graphe connexe, elle ne saurait être continue. Autrement dit, mesurer la distance entre u et v par la distance δ_H entre $\overline{G}_u$ et $\overline{G}_v$ interdit encore d'approcher une fonction discontinue par des approximants continus. Pour éviter cet écueil, on définira le *graphe complété* de u (noté simplement G_u dans la suite), comme

$$G_u = \cup_{x\in\mathbb{R}^d}\{(x,y) : \underline{u}(x) \le y \le \overline{u}(x)\}, \tag{9.8}$$

où $\underline{u}$ et $\overline{u}$ désignent respectivement

$$\underline{u}(x) = \sup_{\varepsilon>0}\inf_{\|y-x\|_2\le\varepsilon} u(y) \quad \text{et} \quad \overline{u}(x) = \inf_{\varepsilon>0}\sup_{\|y-x\|_2\le\varepsilon} u(y). \tag{9.9}$$

Dans la suite, on se limitera à des fonctions qui sont continues en dehors d'un ensemble dénombrable de singularités isolées. En dimension $d = 1$, on utilisera la notion suivante :

Définition 9.1 *On dira qu'une fonction* $u\colon \mathbb{R} \to \mathbb{R}$ *est* admissible *si elle possède en tout point une limite à droite et une limite à gauche. Son graphe complété* (9.8) *est alors la réunion des points* $(x, u(x))$ *où* u *est continue et des segments verticaux* $\{x\} \times [\min\{u(x^-), u(x^+)\}, \max\{u(x^-), u(x^+)\}]$ *sur lesquels* u *est discontinue.*

9.1.3 Uniformité de la distance de Hausdorff

Au début de ce chapitre, on a présenté la distance d_H comme une alternative raisonnable, mais toujours uniforme, à la distance L^∞ pour approcher des fonctions discontinues. On poussera un peu plus loin la comparaison entre les deux distances en observant que l'inégalité $\|u - v\|_{L^\infty} \leq \varepsilon$ peut se traduire par l'assertion "le graphe G_u intersecte tous les *segments verticaux* de centre $m \in G_v$ et de rayon ε", tandis que $d_H(u, v) \leq \varepsilon$ revient à écrire que "le graphe G_u intersecte toutes les *boules* de centre $m \in G_v$ et de rayon ε, et inversement". On pourra donc voir d_H comme sorte de "relaxation isotrope" de la distance L^∞. En d'autres termes, la caractérisation

$$\|u - v\|_{L^\infty} \leq \varepsilon \quad \Longleftrightarrow \quad v - \varepsilon \leq u \leq v + \varepsilon \tag{9.10}$$

devient en distance de Hausdorff

$$d_H(u, v) \leq \varepsilon \quad \Longleftrightarrow \quad \mathcal{S}_\varepsilon^- v \leq u \leq \mathcal{S}_\varepsilon^+ v \ \text{ et } \ \mathcal{S}_\varepsilon^- u \leq v \leq \mathcal{S}_\varepsilon^+ u, \tag{9.11}$$

avec

$$\mathcal{S}_\varepsilon^- u(x) := \inf_{\|x-y\|_2 \leq \varepsilon} u(y) - \varepsilon \quad \text{et} \quad \mathcal{S}_\varepsilon^+ u(x) := \sup_{\|x-y\|_2 \leq \varepsilon} u(y) + \varepsilon. \tag{9.12}$$

On en déduit sans peine

$$d_H(u, v) \leq \|u - v\|_{L^\infty} \tag{9.13}$$

d'une part, et

$$\|u - v\|_{L^\infty} \leq d_H(u, v)[\|u'\|_{L^\infty} + 1] \tag{9.14}$$

d'autre part, de sorte que les distances d_H et L^∞ sont équivalentes dans les zones où l'une des fonctions est assez régulière.

Remarque 9.2 *On pourrait, parallèlement à "l'éloignement" défini par* (9.4) *entre deux ensembles, noter*

$$\epsilon(u, v) := \epsilon(G_u, G_v) \tag{9.15}$$

l'éloignement dissymétrique entre deux fonctions u *et* v*. On aurait alors*

$$\epsilon(u, v) \leq \varepsilon \quad \Longleftrightarrow \quad \mathcal{S}_\varepsilon^- v \leq u \leq \mathcal{S}_\varepsilon^+ v \tag{9.16}$$

et $d_H(u, v) = \max(\epsilon(u, v), \epsilon(v, u))$*.*

La définition suivante nous sera d'une grande utilité par la suite.

Définition 9.3 *On dit d'une fonction uni-dimensionnelle* v *qu'elle est, respectivement,* semi-lipschitzienne supérieurement *ou* inférieurement *s'il existe une constante* L *telle que*

$$v(x + h) - v(x) \leq Lh \quad \textit{ou} \quad v(x - h) - v(x) \leq Lh, \quad \textit{pour tout } h > 0, \tag{9.17}$$

autrement dit si v' *est bornée, supérieurement ou inférieurement, par* L*.*

Une fonction semi-lipschitzienne est bien sûr admissible au sens de la définition 9.1, de plus sa variation totale est localement bornée, et on a pour tout intervalle I

$$|v|_{BV(I)} \leq 2L|I|. \tag{9.18}$$

L'intérêt principal de travailler avec des fonctions semi-lipschitziennes est qu'elle n'ont pas d'oscillations arbitrairement localisées. En particulier, on peut écrire le lemme suivant.

Lemme 9.4 *Si u est semi-lipschitzienne,* i.e. *s'il existe un L positif tel que $\pm u' \leq L$, alors*

$$\epsilon(u, v) \leq (1 + 2L)\epsilon(v, u)$$

pour toute fonction v admissible.

Preuve. Notons $\varepsilon := \epsilon(v, u)$, et considérons le cas où $u' \leq L$ (le problème est symmétrique). Comme les discontinuités de u sont alors décroissantes, on peut observer que

$$u(x^+) = \underline{u}(x) \leq \overline{u}(x) = u(x^-) \tag{9.19}$$

pour tout x. On déduit alors de $u' \leq L$ que

$$u(y) - u(x^+) \leq 2L\varepsilon \quad \text{pour tout} \quad y \in]x, x + 2\varepsilon[,$$

soit

$$\mathcal{S}_\varepsilon^+ u(x + \varepsilon) = \sup_{]x, x+2\varepsilon[} u + \varepsilon \leq \underline{u}(x) + (1 + 2L)\varepsilon.$$

En utilisant (9.16), on voit alors que

$$\inf_{]x-\varepsilon, x+\varepsilon[} v \leq \overline{v}(x + \varepsilon) \leq \mathcal{S}_\varepsilon^+ u(x + \varepsilon) \leq \underline{u}(x) + (1 + 2L)\varepsilon,$$

d'où l'on déduit, en posant $\tilde{\varepsilon} := (1 + 2L)\varepsilon \geq \varepsilon$,

$$\mathcal{S}_{\tilde{\varepsilon}}^- v(x) \leq \inf_{[x-\varepsilon, x+\varepsilon]} v - (1 + 2L)\varepsilon \leq \underline{u}(x).$$

Avec un argument symmétrique, on montrerait que $\overline{u}(x) \leq \mathcal{S}_{\tilde{\varepsilon}}^+ v(x)$, ce qui conclut la preuve. □

9.2 Stabilité de l'équation de Burgers en dimension 1

On commence par traiter le cas de l'équation de Burgers uni-dimensionnelle

$$\partial_t u(t, x) + u(t, x) \cdot \partial_x u(t, x) = 0, \quad u(0, \cdot) = u_0, \quad t > 0, \quad x \in \mathbb{R}. \tag{9.20}$$

Théorème 9.1 *Si u_0 est semi-lipschitzienne (supérieurement ou inférieurement), alors on a pour toute donnée initiale v_0 admissible*

$$d_H(u(t, \cdot), v(t, \cdot)) \leq C(t) d_H(u_0, v_0) \tag{9.21}$$

avec $C(t) = \max\{1 + \tilde{L}t, \tilde{L}\}$ et $\tilde{L} := 2L + 1$, u et v désignant respectivement les solutions entropiques de l'équation (9.20) *pour les données initiales u_0 et v_0.*

Pour établir ce résultat, Cohen, Dahmen et DeVore (voir [15]) ont eu l'idée d'encadrer u_0 et v_0 par des translations de u_0, pour étudier la distance entre u et les solutions issues de ses translatées plutôt qu'entre u et v elle-même.

Avant de donner la preuve de ce théorème, on peut montrer sur un exemple simple que l'hypothèse semi-lipschitzienne (9.17) est nécessaire. Considérons pour un $\varepsilon \ll 1$ les données initiales

$$u_0(x) := \chi_{[0,\varepsilon]}(x) \quad \text{et} \quad v_0(x) := \chi_{[0,\varepsilon^2]}(x)$$

qui vérifient clairement $d_H(u_0, v_0) \leq \varepsilon$. A l'instant $t = 1$, les solutions $u = u(1, \cdot)$ et $v = v(1, \cdot)$ valent respectivement

$$u(x) = x\chi_{[0,\sqrt{2\varepsilon}]}(x) \quad \text{et} \quad v(x) = x\chi_{[0,\varepsilon\sqrt{2}]}(x), \tag{9.22}$$

de sorte que $d_H(u, v)$ est de l'ordre de $\varepsilon^{1/2} \gg \varepsilon$, ce qui exclut toute stabilité.

Preuve du Théorème 9.1. Ecrivons $\varepsilon := d_H(u_0, v_0)$, et supposons pour commencer que $(u_0)'$ est majorée par L. On peut alors vérifier que les translations

$$\tilde{\mathcal{S}}_\varepsilon^- u_0(x) := u_0(x + \varepsilon) - \tilde{L}\varepsilon \quad \text{et} \quad \tilde{\mathcal{S}}_\varepsilon^+ u_0(x) := u_0(x - \varepsilon) + \tilde{L}\varepsilon, \tag{9.23}$$

où $\tilde{L} = 2L + 1$, encadrent u_0 et v_0 :

$$\tilde{\mathcal{S}}_\varepsilon^- u_0 \leq u_0 \leq \tilde{\mathcal{S}}_\varepsilon^+ u_0 \quad \text{et} \quad \tilde{\mathcal{S}}_\varepsilon^- u_0 \leq v_0 \leq \tilde{\mathcal{S}}_\varepsilon^+ u_0. \tag{9.24}$$

L'hypothèse $(u_0)' \leq L$ entraînant en effet

$$u_0(x + \varepsilon) - 2L\varepsilon \leq u_0(y) \leq u_0(x - \varepsilon) + 2L\varepsilon \quad \text{pour tout} \quad |x - y| \leq \varepsilon, \tag{9.25}$$

on voit que les différentes fonctions décalées de u_0 suivant (9.12) et (9.23) s'ordonnent suivant

$$\tilde{\mathcal{S}}_\varepsilon^- u_0 \leq \mathcal{S}_\varepsilon^- u_0 \leq u_0 \leq \mathcal{S}_\varepsilon^+ u_0 \leq \tilde{\mathcal{S}}_\varepsilon^+ u_0, \tag{9.26}$$

ce qui établit l'encadrement de gauche dans (9.24), et celui de droite se déduit immédiatement de (9.11). L'intérêt des fonctions $\tilde{\mathcal{S}}_\varepsilon^\pm u_0$ est que leurs graphes sont des translatés de ceux de u_0, ce qui n'est pas le cas des fonctions $\mathcal{S}_\varepsilon^\pm u_0$. On peut alors observer que si u est la solution entropique de l'équation de Burgers pour la donnée initiale u_0, la fonction $\tilde{u}(t, x) := u(t, x - at) + a$ est la solution entropique correspondant à $\tilde{u}_0 = u_0 + a$, car elle vérifie à la fois l'équation (9.20) au sens faible, et la condition d'entropie d'Oleinik (8.25). L'équation étant d'autre part invariante par translation horizontale de la donnée initiale $u_0 \to u_0(\cdot - b)$, les solutions $u^\pm = u^\pm(t, \cdot)$ associées aux données initiales $\tilde{\mathcal{S}}_\varepsilon^\pm u_0$ ont la forme explicite

$$u^\pm(x) = u(x \mp (1 + \tilde{L}t)\varepsilon) \pm \tilde{L}\varepsilon. \tag{9.27}$$

En utilisant la monotonie (8.18) des solutions, on voit que les encadrements (9.24) sont préservés à l'instant t

$$u^- \leq u \leq u^+ \quad \text{et} \quad u^- \leq v \leq u^+, \tag{9.28}$$

ce qui implique que le graphe G_v est toujours compris soit entre G_{u^-} et G_{u^+}. On en déduit alors que

$$d_H(u,v) \leq \max\{d_H(u^-,u), d_H(u,u^+)\} \leq \max\{1+\tilde{L}t, \tilde{L}\}\varepsilon, \tag{9.29}$$

ce qui termine la preuve (on peut vérifier qu'une preuve similaire permet de traiter le cas où c'est $-(u_0)'$ qui est majorée par L). □

9.3 Stabilité des lois de conservation à flux convexes

Pour généraliser le Théorème 9.1, la principale difficulté est qu'avec un flux f général, on ne dispose plus d'une expression semblable à (9.27) reliant u et $u^{\pm}$.

9.3.1 Un corollaire du théorème de Lax

On a alors choisi de se placer dans le contexte de la section 8.2, où les solutions sont décrites par la formule de Lax (8.22)-(8.23), et on a établi la proposition suivante, qui en est un corollaire.

Proposition 9.5 *On considère ici un flux f de classe $\mathcal{C}^2$ qui vérifie* (8.21) *et dont la dérivée seconde est bornée sur $\mathbb{R}$*

$$f'' \leq B, \tag{9.30}$$

et on désigne par h^- et h^+ les solutions faibles entropiques de la loi de conservation scalaire (8.2) *issues respectivement de deux données initiales admissibles h_0^- et h_0^+ dont l'une, au moins, est semi-lipschitzienne supérieurement. S'il existe un $\alpha \geq 0$ pour lequel*

$$h_0^- \leq h_0^+ + \alpha, \tag{9.31}$$

alors h^- et h^+ vérifient

$$h^-(t, x^- + t\Delta) \leq h^+(t,x^+) + \alpha + tL\Delta \tag{9.32}$$

pour tout $t>0$, tout $x \in \mathbb{R}$ et tout $\Delta > \alpha B$, avec $L := \min\{\sup_{\mathbb{R}}(h_0^-)', \sup_{\mathbb{R}}(h_0^+)'\}$.

Remarque 9.6 *D'après* (8.15) *et* (8.18)*, l'enveloppe convexe de l'ensemble $I(t) = h^-(t,\mathbb{R}) \cup h^+(t,\mathbb{R}) \subset \mathbb{R}$ décroît au cours du temps. On peut donc restreindre les hypothèses faites sur le flux f à l'enveloppe convexe de $I(0) = h_0^-(\mathbb{R}) \cup h_0^+(\mathbb{R})$, les valeurs de f en dehors de cet intervalle n'ayant aucune influence sur les solutions h^- et h^+. L'hypothèse* (9.30) *est donc une simple conséquence de la continuité de f''.*

Preuve. On va utiliser le fait que les solutions h^- et h^+ sont données par (8.22)-(8.23). Pour x et t donné, si $y_+ = y_+(t,x)$ minimise la fonctionnelle $\mathcal{L}^+(\cdot,t,x) := \mathcal{L}_{h_0^+,f}(\cdot,t,x)$, on peut montrer que minimiseurs y_- de $\mathcal{L}^-(\cdot,t,x+t\Delta) := \mathcal{L}_{h_0^-,f}(\cdot,t,x+t\Delta)$ sont tous supérieurs à y_+, au sens large (ce qu'illustre la figure 9.5). Pour cela, il nous suffit que la quantité

$$K(z) := \mathcal{L}^-(z,t,x+t\Delta) - \mathcal{L}^-(y_+,t,x+t\Delta) \tag{9.33}$$

soit strictement positive lorsque $z < y_+$. En utilisant le fait que $\mathcal{L}^+(y_+,t,x) \leq \mathcal{L}^+(z,t,x)$ pour tout z, on a

$$\begin{aligned} K &\geq \mathcal{L}^-(z,t,x+t\Delta) - \mathcal{L}^-(y_+,t,x+t\Delta) - \mathcal{L}^+(z,t,x) + \mathcal{L}^+(y_+,t,x) \\ &= \int_z^{y_+} \left[h_0^+(s) - h_0^-(s)\right] \mathrm{d}s \\ &+ t\left(f^*\left(\frac{x-z+t\Delta}{t}\right) - f^*\left(\frac{x-y_+ + t\Delta}{t}\right) - f^*\left(\frac{x-z}{t}\right) + f^*\left(\frac{x-y_+}{t}\right)\right) \\ &= \int_z^{y_+} \left[h_0^+(s) - h_0^-(s) + (f^*)'\left(\frac{x-s+t\Delta}{t}\right) - (f^*)'\left(\frac{x-s}{t}\right)\right] \mathrm{d}s. \end{aligned} \tag{9.34}$$

En utilisant l'hypothèse (9.31), on en déduit que

$$K(z) \geq \int_z^{y_+} \left[-\alpha + \inf((f^*)'')\Delta\right] = (y_+ - z)\big(\inf((f^*)'')\Delta - \alpha\big)$$

dès que Δ est positif. Et lorsqu'il est strictement supérieur à αB, on voit d'après (8.32) que $K(z) > 0$ pour tout $z < y_+$.

A ce stade, on peut donc écrire que pour tout x, le plus grand minimiseur y_+ de $\mathcal{L}_{h_0^+,f}(\cdot,t,x)$ est à droite du premier minimiseur y_- de $\mathcal{L}_{h_0^-,f}(\cdot,t,x+t\Delta)$, y_+ et y_- vérifiant respectivement

$$h^+(t,x^+) = h_0^+(y_+) \text{ et } \quad h^-(t,x^- + t\Delta) = h_0^-(y_-). \tag{9.35}$$

En écrivant l'égalité (8.27) pour y_- et y_+, on obtient

$$y_- + tf'(h_0^-(y_-)) = x + t\Delta = y_+ + t\Delta + tf'(h_0^+(y_+)). \tag{9.36}$$

La convexité de f nous permet d'en déduire que y_- et y_+ vérifient soit

$$h_0^-(y_-) \leq h_0^+(y_+),$$

ce qui entraîne $h^-(t,x^- + t\Delta) \leq h^+(t,x^+)$ d'après (9.35), soit

$$y_- \leq y_+ + t\Delta.$$

Dans la mesure où $y_+ \leq y_-$, on peut alors utiliser l'hypothèse (9.31) sur h_0^- et h_0^+ pour calculer

$$h_0^-(y_-) \leq h_0^+(y_-) + \alpha \leq h_0^+(y_+) + L(y_- - y_+) + \alpha \leq h_0^+(y_+) + tL\Delta + \alpha, \tag{9.37}$$

si $(h_0^+)' \leq L$, ou bien

$$h_0^-(y_-) \leq h_0^-(y_+) + L(y_- - y_+) \leq h_0^-(y_+) + tL\Delta \leq h_0^+(y_+) + tL\Delta + \alpha \tag{9.38}$$

si c'est $(h_0^-)'$ qui est majorée par L. L'inégalité (9.32) se déduit alors immédiatement de (9.35). □

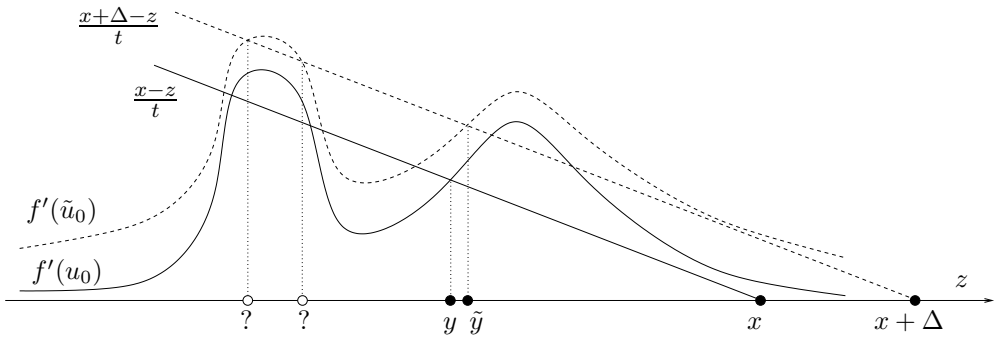


FIG. 9.5 – position des minimiseurs $\tilde{y}$ après une perturbation de la donnée initiale u_0 (illustration de la preuve de la proposition 9.5).

9.3.2 Stabilité des lois unidimensionnelles à flux convexes

On a alors établi le résultat suivant, qui généralise le théorème 9.1.

Théorème 9.2 *On considère ici un flux f de classe $\mathcal{C}^2$ qui vérifie* (8.21) *et dont la dérivée seconde est bornée sur* $\mathbb{R}$

$$f'' \leq B, \tag{9.39}$$

ce qui est toujours possible d'après la remarque 9.6. Si u_0 est une fonction semi-Lipschtizienne supérieurement

$$(u_0)' \leq L \tag{9.40}$$

et si v_0 est admissible au sens de la définition 9.1, alors les solutions faibles entropiques $u(t,\cdot)$ et $v(t,\cdot)$ de la loi de conservation scalaire (8.2) *issues des données initiales u_0 et v_0 vérifient*

$$d_H(u(t,\cdot),v(t,\cdot)) \leq C(t) d_H(u_0,v_0) \tag{9.41}$$

avec $C(t) = \tilde{L}(1+2tB\tilde{L})$ et $\tilde{L} := 2L+1$.

Preuve. Mis à part l'expression (9.27), la preuve du théorème 9.1 s'applique à nouveau. On aura donc établi l'inégalité (9.41) si l'on arrive à montrer que

$$d_H(u,u^+) \leq \varepsilon\tilde{L}(1+2tB\tilde{L}), \tag{9.42}$$

où u^+ désigne la solution issue de la translatée

$$u_0^+(x) := u_0(x-\varepsilon) + \varepsilon\tilde{L} \tag{9.43}$$

(notée $\mathcal{S}^+u_0$ en (9.23)), et ε la distance initiale $d_H(u_0,v_0)$. Pour estimer la distance entre $u(t,\cdot)$ et $u^+(t,\cdot)$, on peut rappeler dans un premier temps que la monotonie de l'équation nous assure que

$$u(t,\cdot) \leq u^+(t,\cdot). \tag{9.44}$$

Dans un deuxième temps, on peut utiliser la proposition 9.5 avec $h_0^- = u_0^+(\cdot+\varepsilon)$, $h_0^+ = u_0$ et $\alpha = \tilde{L}\varepsilon$. On a alors pour tout $t>0$ et tout x (en prenant $\Delta = 2\tilde{L}\varepsilon B$)

$$h^-(t,x^- + 2t\varepsilon\tilde{L}B) \leq h^+(t,x^+) + \varepsilon\tilde{L}(1+2tLB), \tag{9.45}$$

soit en observant que $h^+(t,\cdot) = u(t,\cdot)$ et $h^-(t,\cdot) = u^+(t,\cdot+\varepsilon)$,

$$u^+(t, x^- + \varepsilon(1 + 2t\tilde{L}B)) \leq u(t, x^+) + \varepsilon\tilde{L}(1 + 2tLB). \tag{9.46}$$

Pour conclure, on peut utiliser le fait que u ne peut avoir que des discontinuités décroissantes (garantit par exemple par l'inégalité d'Oleinik (8.25)). On en déduit qu'un point m du graphe complété de $u(t,\cdot)$ s'écrit toujours (x,u) avec $u \in [u(t,x^+), u(t,x^-)]$. Les inégalités (9.44) et (9.46) nous apprennent alors que le graphe complété de $u^+(t,\cdot)$ intersecte le rectangle de diagonale $[m, m + \varepsilon((1 + 2t\tilde{L}B), \tilde{L}(1 + 2tLB)]$ (voir figure 9.6). On en déduit qu'il est à une distance inférieure à $\varepsilon\tilde{L}(1 + 2tB\tilde{L})$ du point m, et ceci étant valable pour tous les points m de G_u, que

$$\epsilon(u, u^+) \leq \varepsilon\tilde{L}(1 + 2tB\tilde{L}) \tag{9.47}$$

où l'éloignement ϵ entre deux fonctions est défini par (9.15). Les inégalités (9.44) et (9.46) étant valable pour tout x, on peut également en déduire que le graphe complété de $u(t,\cdot)$ intersecte tous les rectangles de diagonale $[m^+ - \varepsilon((1 + 2t\tilde{L}B), \tilde{L}(1 + 2tLB)), m^+]$, lorsque m^+ décrit le graphe G_u. On en déduit que $\epsilon(u^+, u) \leq \varepsilon\tilde{L}(1 + 2tB\tilde{L})$, d'où finalement (9.42), et la preuve est terminée. □

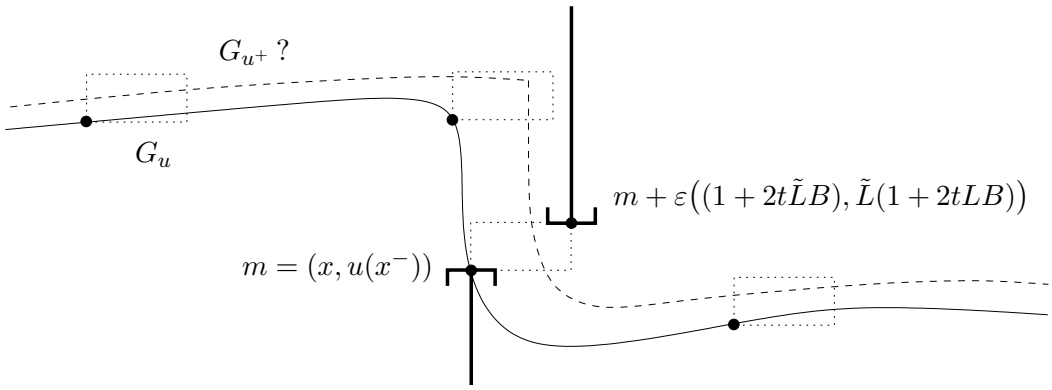


FIG. 9.6 – d'après (9.44) et (9.46), le graphe de u^+ est proche de tous les points m du graphe complété de u.

9.3.3 Stabilité vis-à-vis des perturbations de flux

Dans le même esprit, on a établi le résultat suivant concernant la stabilité des lois de conservation scalaires pour une perturbation lipschitzienne du flux.

Théorème 9.3 *On considère ici deux flux f et g de classe $\mathcal{C}^2$ vérifiant*

$$0 < A \leq f'' \leq B \quad \text{et} \quad 0 < A \leq g'' \leq B, \tag{9.48}$$

et une fonction u_0 admissible au sens de la définition 9.1. Si u et v sont respectivement solutions (faibles, entropiques) des lois de conservation (scalaires et unidimensionnelles)

$$\partial_t u(t,x) + \partial_x[f(u(t,x))] = 0, \quad u(0,\cdot) = u_0, \tag{9.49}$$

et

$$\partial_t v(t,x) + \partial_x[g(v(t,x))] = 0, \quad v(0,\cdot) = u_0 \tag{9.50}$$

issues de la même donnée initiale u_0, alors on a

$$d_H(u(t,\cdot), v(t,\cdot)) \leq C(t)\|f' - g'\|_{L^\infty} \tag{9.51}$$

avec $C(t) = \max\{tB/A, (1+tB/A)/A\}$.

Preuve. Commençons par observer que f'' et g'' étant tous deux minorés par un $A > 0$, leurs transformées de Legendre f^* et g^* vérifient

$$\|(f^*)' - (g^*)'\|_{L^\infty} \leq \varepsilon/A \tag{9.52}$$

où $\varepsilon := \|f' - g'\|_{L^\infty}$. Fixons pour cela un s et notons $u = (f^*)'(s)$ et $v = (g^*)'(s)$. Quitte à intervertir g et f (les hypothèses sont symétriques), on peut supposer que $u \leq v$, on a alors

$$s = f'(v) \geq f'(u) + A(u-v) \geq g'(u) - \varepsilon + A(u-v) = s - \varepsilon + A(u-v), \tag{9.53}$$

d'où l'on déduit $|(f^*)'(s) - (g^*)'(s)| = u - v \leq \varepsilon/A$.

On peut alors établir que u et v vérifient

$$v(t, x^- + t\Delta) \leq u(t, x^+) + (\varepsilon + \Delta)/A \tag{9.54}$$

pour tout $x \in \mathbb{R}$, tout $t > 0$ et tout $\Delta > \varepsilon B/A$. Considérons pour cela un minimiseur $y = y(t,x)$ de $\mathcal{L}(\cdot,t,x) := \mathcal{L}_{u_0,f}(\cdot,t,x)$, et montrons que les minimiseurs $\tilde{y}$ de $\tilde{\mathcal{L}}(\cdot,t,x+t\Delta) := \mathcal{L}_{u_0,g}(\cdot,t,x+t\Delta)$ vérifient tous $\tilde{y} \geq y$ dès lors que $\Delta > B\varepsilon/A$. En utilisant le fait que $\mathcal{L}(y,t,x) \leq \mathcal{L}(z,t,x)$ pour tout z, on calcule que la quantité

$$K(z) := \tilde{\mathcal{L}}(z,t,x+t\Delta) - \tilde{\mathcal{L}}(y,t,x+t\Delta) \tag{9.55}$$

vérifie pour $z < y$

$$\begin{aligned}
K &\geq \tilde{\mathcal{L}}(z,t,x+t\Delta) - \tilde{\mathcal{L}}(y,t,x+t\Delta) - \mathcal{L}(z,t,x) + \mathcal{L}(y,t,x) \\
&= t\left(g^*\left(\frac{x-z+t\Delta}{t}\right) - g^*\left(\frac{x-y+t\Delta}{t}\right) - f^*\left(\frac{x-z}{t}\right) + f^*\left(\frac{x-y}{t}\right)\right) \\
&= \int_z^y \left[(g^*)'\left(\frac{x-s+t\Delta}{t}\right) - (f^*)'\left(\frac{x-s}{t}\right)\right] \mathrm{d}s \\
&\geq \int_z^y \left[(f^*)'\left(\frac{x-s+t\Delta}{t}\right) - (f^*)'\left(\frac{x-s}{t}\right)\right] \mathrm{d}s - \varepsilon(y-z)/A \\
&\geq [\inf((f^*)'')\Delta - \varepsilon/A](y-z) \\
&> 0
\end{aligned} \tag{9.56}$$

pour tout $\Delta > B\varepsilon/A$, en utilisant à nouveau $\inf((f^*)'') = 1/\sup(f'') \geq 1/B$.

A ce stade, on peut donc écrire que pour tout x, le plus grand minimiseur y de

$\mathcal{L}_{u_0,f}(\cdot,t,x)$ est à droite du premier minimiseur $\tilde{y}$ de $\mathcal{L}_{u_0,g}(\cdot,t,x+t\Delta)$, y et $\tilde{y}$ vérifiant respectivement

$$u(t,x^+) = u_0(y) \quad \text{et} \quad v(t,x^- + t\Delta) = u_0(\tilde{y}). \tag{9.57}$$

En écrivant l'égalité (8.27) pour $\tilde{y}$ et y, on obtient

$$\tilde{y} + tg'(u_0(\tilde{y})) = x + t\Delta = y + t\Delta + tf'(u_0(y)), \tag{9.58}$$

et on déduit alors du fait que $y \leq \tilde{y}$ que

$$g'(u_0(\tilde{y})) \leq \Delta + f'(u_0(y)) \leq \Delta + g'(u_0(y)) + \varepsilon, \tag{9.59}$$

ce qui nous conduit à l'alternative suivante : soit on a $u_0(\tilde{y}) \leq u_0(y)$ et l'inégalité (9.54) est évidente, soit on a $u_0(y) \leq u_0(\tilde{y})$ et dans ce cas la forte convexité de g permet d'écrire $g'(u_0(y)) \leq g'(u_0(\tilde{y})) - (u_0(\tilde{y}) - u_0(y))A$. Ajoutée à (9.59), cette dernière inégalité nous donne $(u_0(\tilde{y}) - u_0(y))A \leq (\varepsilon + \Delta)$, ce qui d'après (9.57) correspond exactement à (9.54).

Pour en déduire que le graphe de $v(t,\cdot)$ passe à proximité de tous les points du graphe de $u(t,\cdot)$, on ne peut plus écrire que $u \leq v$ comme dans la preuve du théorème 9.2. Mais on peut observer que les hypothèses sur u et v étant symétriques, l'inégalité (9.54) entraîne également

$$v(t,x^- - t\Delta) \geq u(t,x^+) - (\varepsilon + \Delta)/A. \tag{9.60}$$

Comme les discontinuités de u sont toujours décroissantes, $u(t,x^+)$ est toujours inférieur à $u(t,x^-)$, et on a

$$v(t,x^- - t\Delta) \geq u(t,x^-) - (\varepsilon + \Delta)/A. \tag{9.61}$$

On observe alors que cette inégalité et (9.54) impliquent que le graphe de $v(t,\cdot)$ intersecte tous les rectangles de diagonale $[(x - t\Delta, u(x^-) - (\varepsilon + \Delta)/A), (x + t\Delta, u(x^-) + (\varepsilon + \Delta)/A)]$ pour $x \in \mathbb{R}$, et on en déduit que

$$\epsilon(u,v) \leq \varepsilon \max\{tB/A, (1 + tB/A)/A\}. \tag{9.62}$$

Comme à nouveau, on peut intervertir les fonctions u et v, on en déduit finalement l'inégalité (9.51). □

9.4 Résultats négatifs

Lorsque le flux f n'est pas convexe, ou bien en plusieurs dimensions lorsque l'équation (8.1) ne peut pas se réduire à une loi de conservation uni-dimensionnelle, on n'a en général pas de stabilité en distance de Hausdorff pour les solutions. Autrement dit, les hypothèses du théorème 9.2 sont optimales.

9.4.1 Cas des flux non convexes

On considère ici le cas où la dérivée seconde du flux f'' change de signe. Pour construire un contre-exemple, l'idée est de trouver une solution initiale u_0 qui donnera naissance à deux ondes de chocs distantes se déplaçant l'une vers l'autre. A l'instant de leur rencontre, le graphe de u varie brutalement (voir figure 9.7), et si l'on perturbe légèrement u_0, on déplace cet instant de façon contradictoire avec un principe de stabilité des solutions en distance de Hausdorff. Le théorème suivant décrit ce phénomène en détails.

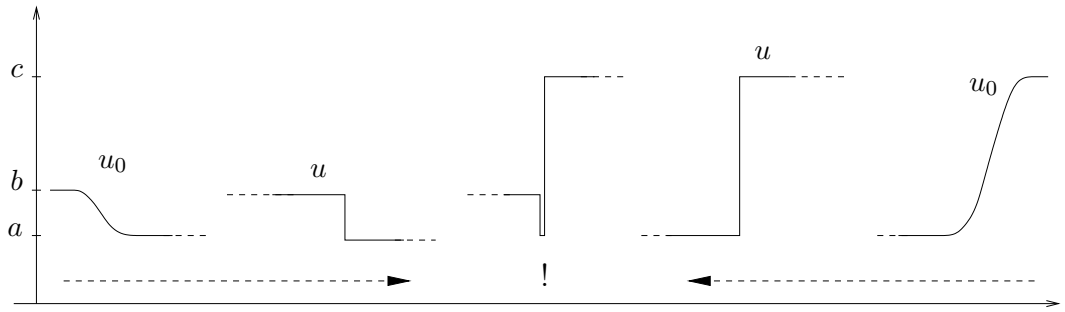


FIG. 9.7 – lorsque le flux est non convexe, il peut apparaître des chocs qui se déplacent l'un vers l'autre, faisant varier brutalement le graphe de la solution.

Théorème 9.4 *On considère ici le cas où f'' est de classe $\mathcal{C}^1$ et change de signe. Alors il existe deux réels strictement positifs K, T et une donnée initiale u_0 régulière telle que pour tout $\varepsilon > 0$, il existe une perturbation v_0 de u_0 vérifiant*

$$d_H(u_0, v_0) \leq \varepsilon \tag{9.63}$$

et pour laquelle les solutions entropiques de (8.2) *issues de u_0 et v_0 vérifient*

$$d_H(u(T,\cdot), v(T,\cdot)) \geq K. \tag{9.64}$$

Preuve. Sans perte de généralité, on peut supposer qu'il existe trois réels $a < b < c$ tels que $f''(a) > 0$, $f''(b) = 0$ et $f''(c) < 0$, et tels que f'' est positive sur $[a, b]$ et négative sur $[b, c]$. D'autre part, on peut aussi supposer que $f'(a) = 0$, car les solutions entropiques $\tilde{u}$ de (8.2) associées au flux $\tilde{f}(x) = f(x) - xf'(a)$ se déduisent de u par une translation $\tilde{u}(t, x) = u(t, x - tf'(a))$, et on a clairement $d_H(\tilde{u}, \tilde{v}) = d_H(u, v)$. D'après les hypothèses faites sur f'', on en déduit facilement que $f(b) > f(a)$, soit

$$\frac{f(b) - f(a)}{b - a} > 0. \tag{9.65}$$

Enfin, on peut observer sur la figure 9.8 que quitte à prendre a très proche de b, la valeur moyenne de f' sur $[a, c]$ peut toujours être considérée négative, de sorte que

$$\frac{f(c) - f(a)}{c - a} > 0. \tag{9.66}$$

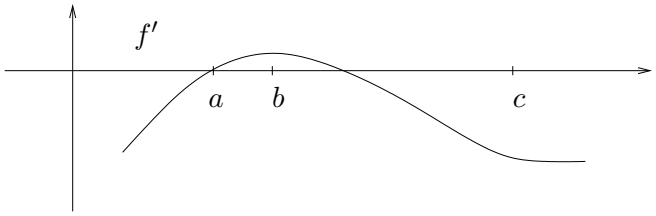


FIG. 9.8 – allure de la dérivée f' d'un flux régulier non convexe.

On considère alors une première solution initiale u_0^1 vérifiant $u_0^1(x) = b$ pour $x \leq 0$, $u_0^1(x) = a$ pour $x \geq 1$ et décroissante sur $[0,1]$. Après un temps fini T_1, cette solution devient une onde de choc pure

$$u^1(t,x) = b\chi_{x\leq y^1+tv^1}(x) + a\chi_{x\geq y^1+tv^1}(x) \tag{9.67}$$

pour $t > T_1$, où y^1 est fixé et de vitesse $v^1 = (f(b) - f(a))/(b-a) > 0$ d'après la loi de Rankine-Hugoniot et (9.65). Sur le même principe, on considère ensuite une deuxième solution initiale u_0^2 vérifiant $u_0^2(x) = a$ pour $x \leq -1$, $u_0^2(x) = c$ pour $x \geq 0$ et croissante sur $[-1,0]$. Après un temps fini T_2, cette deuxième solution devient également une onde de choc pure

$$u^2(t,x) = c\chi_{x\leq y^2+tv^2}(x) + a\chi_{x\geq y^2+tv^2}(x) \tag{9.68}$$

pour $t > T_2$, avec y^2 fixé et de vitesse $v^2 = (f(c) - f(a))/(c-a) < 0$ d'après (9.66).

On définit alors u_0 par

$$u_0(x) := u_0^1(x) + u_0^2(x-z) - a, \tag{9.69}$$

où z est tel que $z > y^1 - y^2 + (v^1 - v^2)\max\{T_1, T_2\}$, de façon à ce que les ondes de choc issues de u_0^1 et $u_0^2(\cdot - z)$ n'interagissent pas avant de s'être entièrement développées, comme c'est le cas sur la figure 9.7 : pour $t \in [\max\{T_1,T_2\}, T_3]$ où T_3 est donné par

$$z + y^2 - y^1 + T_3(v^2 - v^1) = 0, \tag{9.70}$$

on a

$$u(t,x) = u^1(t,x) + u^2(t,x-z) - a. \tag{9.71}$$

A l'instant T_3, les deux ondes de chocs se rencontrent en $y = y^1 + T_3v^1 = y^2 + z + T_3v^2$, et pour $t > T_3$, la solution devient une onde de choc pure

$$u(t,x) = b\chi_{x\leq y+(t-T_3)v}(x) + c\chi_{x\geq y+(t-T_3)v}(x) \tag{9.72}$$

de vitesse $v = (u(c) - u(b))/(c-b) < 0$.

Pour un $\varepsilon > 0$ arbitraire, on perturbe alors u_0 en "retardant" légèrement l'onde de gauche :

$$v_0(x) := u_0^1(x+\varepsilon) + u_0^2(x-z) - a, \tag{9.73}$$

d'où l'on déduit immédiatement que $d_H(v_0, u_0) \leq \varepsilon$. La discussion ci-dessus s'applique encore, et on a

$$v(t,x) = u^1(t,x+\varepsilon) + u^2(t,x-z) - a \tag{9.74}$$

pour $t \in [\max\{T_1, T_2\}, T_3^\varepsilon]$, où T_3^ε est donné par

$$z + \varepsilon + y^2 - y^1 + T_3^\varepsilon (v^2 - v^1) = 0. \tag{9.75}$$

On voit donc que pour $T_3 < t < T_3^\varepsilon$, les chocs se sont déjà confondus dans le graphe de u, mais pas encore dans le graphe de v. On en déduit alors que pour $T = (T_3 + T_3^\varepsilon)/2$,

$$d_H(u(t,\cdot), v(t,\cdot)) \geq K := b - a, \tag{9.76}$$

ce qui conclut la preuve. □

Remarque 9.7 *Ce contre-exemple n'est en réalité pas complètement satisfaisant, car la stabilité n'est violée que sur un intervalle de temps très court. Et de la même façon qu'on a relaxé la distance L^∞ en espace, on aurait envie de la* relaxer en temps. *Ce contre-exemple n'en serait alors plus un, et il n'est d'ailleurs pas évident qu'il en existe encore.*

9.4.2 Cas des dimensions supérieures

Dans le cas multi-dimensionnel, la fonction de flux s'écrit $f(y) = (f_1(y), \cdots, f_d(y))$, chaque f_i étant une fonction de $\mathbb{R}$ dans $\mathbb{R}$. On dit alors de f que c'est un flux *essentiellement uni-dimensionnel* s'il peut se mettre sous la forme

$$f(y) = ag(y) \tag{9.77}$$

où a désigne un vecteur unitaire constant de $\mathbb{R}^d$ et g une fonction de $\mathbb{R}$ dans $\mathbb{R}$, et dans le cas contraire, qu'il est *véritablement multi-dimensionnel* (voir figure 9.9).

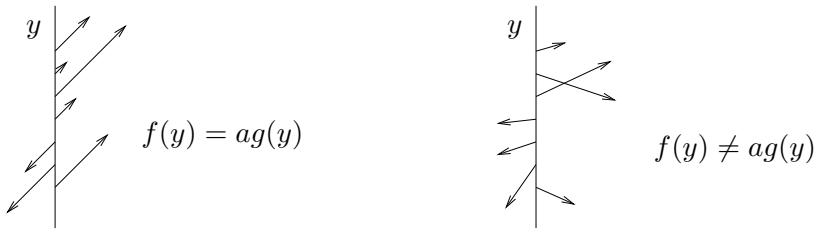


FIG. 9.9 – exemple de flux essentiellement uni-dimensionnel (à gauche), et véritablement multi-dimensionnel (à droite).

On peut d'autre part étendre la définition 9.3 à une fonction v de $\mathbb{R}^d$ en disant qu'elle est *semi-lipschitzienne dans la direction* $b \in \mathbb{R}^d$ si (b désignant un vecteur unitaire de $\mathbb{R}^d$) on a

$$v(x + sb) - v(x) \leq Ls \quad \text{pour tout} \quad s > 0, \tag{9.78}$$

avec $L > 0$ une constante fixée. Le théorème suivant montre que les seules équations stables pour la distance de Hausdorff sont celles qui sont essentiellement uni-dimensionnelles, à un terme linéaire près.

Théorème 9.5 *On considère ici que le flux f s'écrit sous la forme*

$$f = (g_1, \cdots, g_d) + ag \tag{9.79}$$

où les g_i sont des fonctions affines, a est un vecteur unitaire *fixé de $\mathbb{R}^d$, et g un flux uni-dimensionnel de classe $\mathcal{C}^2$ vérifiant $0 < g'' \leq B$. Si u_0 est une fonction continue qui est semi-lipschitzienne dans toutes les directions b telles que $(a, b) \geq 0$, alors pour toute fonction continue v_0, on a*

$$d_H(u(t,\cdot), v(t,\cdot)) \leq C(t) d_H(u_0, v_0) \tag{9.80}$$

pour tout $t > 0$, avec $C(t) = \tilde{L}(1 + 2tB\tilde{L})$ et $\tilde{L} = 1 + 2L$.

Lorsque le flux f ne peut pas s'écrire sous la forme (9.79)*, il existe deux réels strictement positifs T, K et une donnée initiale u_0 satisfaisant les hypothèses ci-dessus, telle que pour tout $\varepsilon > 0$, on puisse exhiber une deuxième donnée initiale v_0 vérifiant $d_H(u_0, v_0) \leq \varepsilon$ et pour laquelle on a*

$$d_H(u(T,\cdot), v(T,\cdot)) \geq K. \tag{9.81}$$

Preuve. Pour la première partie du théorème, on peut supposer que les g_i sont tous nuls, dans la mesure où les solutions $\tilde{u}$ et $\tilde{v}$ de (8.1) associées au flux $f - (g_1, \cdots, g_d)$ se déduisent respectivement de u et v par les translations $u(t,x) = \tilde{u}(t, x - t(g_1, \cdots, g_d))$ et $v(t,x) = \tilde{v}(t, x - t(g_1, \cdots, g_d))$, d'où l'on déduit facilement que $d_H(\tilde{u}, \tilde{v}) = d_H(u, v)$. D'autre part, le fait que la propriété (9.78) soit satisfaite dans toutes les directions b telles que $(a, b) \geq 0$ entraîne que les translations

$$u_0^+ := u_0(\cdot - \varepsilon a) + \tilde{L}\varepsilon \quad \text{et} \quad u_0^- := u_0(\cdot + \varepsilon a) - \tilde{L}\varepsilon \tag{9.82}$$

avec $\varepsilon := d_H(u_0, v_0)$ vérifient

$$u_0^- \leq u_0 \leq u_0^+ \quad \text{et} \quad u_0^- \leq v_0 \leq u_0^+, \tag{9.83}$$

car pour tout point y à distance $\|y - x\|_2 \leq \varepsilon$ de x s'écrit $y = x - \varepsilon a + \eta b$ avec $(a, b) \geq 0$ et $\eta \leq 2\varepsilon$, d'où l'on déduit que $\sup_{\|y-x\|_2 \leq \varepsilon} u_0(y) \leq u_0(x - \varepsilon a) + 2\varepsilon L$, et finalement (9.83) d'après la caractérisation (9.11)-(9.12) de la distance de Hausdorff. On désignera par u^- et u^+ les solutions issues de ces données initiales. Le flux ayant alors la forme $f = ag$, on peut vérifier que pour tout $x \in (a\mathbb{R})^\perp$, la fonction $u^{(x)}(t,s) = u(t, x + sa)$ avec $s \in \mathbb{R}$ est solution entropique de la loi de conservation uni-dimensionnelle (8.1) associée au flux g et à la donnée initiale $u_0^{(x)}$. On sait alors que $u^{(x)}$ est une fonction admissible au sens de la définition 9.1, et on désigne son graphe complété par $G_{u^{(x)}}$. D'après le théorème 9.2, on a

$$d_H(u^{(x)}(t,\cdot), u^{(\tilde{x})}(t,\cdot)) \leq C(t) d_H(u_0^{(x)}, u_0^{(\tilde{x})}) \tag{9.84}$$

et la continuité de u_0 nous permet d'écrire que

$$d_H(u_0^{(x)}, u_0^{(\tilde{x})}) \to 0 \quad \text{lorsque} \quad \tilde{x} \to x. \tag{9.85}$$

On en déduit que pour tout temps $t > 0$, le graphe complété de u peut être défini à partir des graphes $G_{u^{(x)}}$ par

$$G_u = \cup_{x\in(a\mathbb{R})^\perp}\{(x+sa, y) : (s,y) \in G_{u^{(x)}}\}, \tag{9.86}$$

autrement dit, G_u est la réunion de ses rayons uni-dimensionnels dans la direction a. Les graphes de v, u^- et u^+ peuvent être définis de la même façon. On peut alors reprendre les arguments développés dans la preuve du théorème 9.2 pour voir que

$$\begin{aligned} d_H(u,v) &\le \max\{d_H(u^-,u), d_H(u^+,u)\} \\ &\le \sup_{x\in(a\mathbb{R})^\perp} \max\{d_H\big((u^{(x)})^-, u^{(x)}\big), d_H\big((u^{(x)})^+, u^{(x)}\big)\} \le C(t)\varepsilon \end{aligned} \tag{9.87}$$

avec $C(t) = \tilde{L}(1+2tB\tilde{L})$, ce qui prouve la première partie du théorème.

Pour voir que la condition (9.79) sur le flux est nécessaire, on suppose maintenant que f est quelconque. Pour un vecteur unitaire b arbitraire de $\mathbb{R}^d$, on considère alors la fonction

$$u(t,x) := u^{(b)}(t,(x,b)) \tag{9.88}$$

construite *à partir* de la solution entropique uni-dimensionnelle associée au flux $f_b := (f,b)$ et à une donnée initiale $u_0^{(b)}$. On peut alors vérifier que u est la solution entropique associée au flux f, lorsque la donnée initiale est essentiellement uni-dimensionnelle et donnée par $u_0(x) := u_0^{(b)}((x,b))$. A chaque instant $t > 0$, la fonction $u^{(b)}$ est admissible au sens de la définition 9.1, ce qui nous permet de définir son graphe complété, de même que celui de u. D'après le théorème 9.4, on sait alors que la stabilité Hausdorff peut être mise en défaut dès lors que le signe de $f_b'' = (f'', b)$ n'est pas constant. En observant la figure 9.9, on vérifie alors sans peine qu'un flux pour lequel les projections (f'', b) sont de signe constant dans toutes les directions b est en réalité essentiellement uni-dimensionnel, et on en déduit que f doit s'écrire sous la forme (9.79) pour que la loi de conservation puisse être stable en distance de Hausdorff. □

9.4.3 Stabilité pour des temps petits

Lorsque u_0 est décroissante, la constante $C(t)$ du théorème 9.2 vaut $1+2tB$, ce qui est très intéressant lorsque l'on doit accumuler des estimations d'erreurs sur n intervalles de temps Δt, car on peut alors utiliser un lemme de Gronwall comme le lemme 1.4. Lorsque aucune des solutions initiales n'est décroissante, en revanche, on peut montrer qu'il n'existe aucune constante C pour laquelle on aurait

$$d_H(u(t,\cdot), v(t,\cdot)) \le (1+Ct) d_H(u_0, v_0) \tag{9.89}$$

pour des valeurs de t proches de 0.

Théorème 9.6 *Les solutions entropiques u et v de l'équation de Burgers* (9.20) *respectivement issues des solutions initiales*

$$u_0(x) := x\chi_{[0,1]}(x) \quad \textit{et} \quad v_0 := \sup_{|h|\le\varepsilon} u_0(x+h) + \varepsilon$$

ne vérifient

$$d_H(u(t,\cdot),v(t,\cdot)) \leq (1+Ct)d_H(u_0,v_0), \quad \textit{pour } t>0 \tag{9.90}$$

pour aucune constante C.

Preuve. D'après l'équivalence (9.11), il est clair que $d_H(u_0,v_0)=\varepsilon$. On peut d'autre part observer sur la figure 9.10 que v_0 et u_0 sont constituées d'un choc suivi par une détente. Sur u_0, la détente est au contact du choc, de sorte que l'amplitude maximale $\sup_{x\in\mathbb{R}} u(t,x)$ de u va décroître dès l'instant initial. Dans le cas de v_0, en revanche, il va s'écouler un certain temps avant que la détente ne rattrape la position du choc, et l'amplitude maximale $\sup_{x\in\mathbb{R}} v(t,x)$ va rester constante égale à $1+\varepsilon$ durant cette période. On peut donc vérifier qu'avant leur rencontre, le choc de v avance à la vitesse $\frac{1}{2}+\varepsilon$ d'après la loi de Rankine-Hugoniot, tandis que sa détente avance à la vitesse $1+\varepsilon$. Il ne se rencontrent donc pas sur des temps $t\leq 4\varepsilon$. Comme on peut d'autre part calculer que $u(t,x)=\frac{x}{1+t}\chi_{[0,\sqrt{1+t}]}(x)$, on a

$$d_H(u(t,\cdot),u(t,\cdot)) \geq \sup_{x\in\mathbb{R}} v(t,x) - \sup_{x\in\mathbb{R}} u(t,x) = 1+\varepsilon-\frac{x}{\sqrt{1+t}} \tag{9.91}$$

pour $t\leq 4\varepsilon$. S'il existait une constante C pour laquelle (9.89) était vérifiée, on aurait pour $t=4\varepsilon$

$$1-\frac{1}{1+t} \leq Ct\varepsilon \leq Ct^2, \tag{9.92}$$

d'où l'on déduirait que $\varphi(t):=\frac{1}{1+t}+Ct^2$ est supérieure à 1 à droite de 0, ce qui dans la mesure où $\varphi(0)=1$ et $\varphi'(0)=-1/2$, est notoirement faux. Et on peut également montrer que (9.89) n'est pas vérifiée pour des valeurs de t plus grandes que 4ε. □

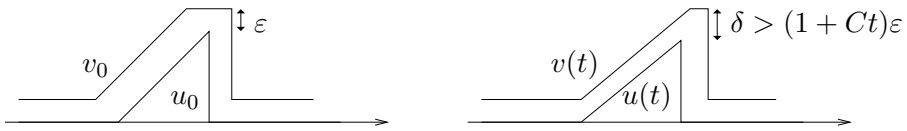


FIG. 9.10 – exemple de fonctions initiales pour lesquelles la stabilité (9.89) n'est vérifiée pour aucune constante.

Chapitre 10

Régularité "géométrique" d'ordre élevé

A partir des résultats du chapitre précédent, on établit un résultat de régularité d'ordre élevé pour les lois de conservation scalaires uni-dimensionnelles à flux convexes. Les régularités concernées par ce résultat sont celles qui caractérisent des ordres élevés de convergence pour l'approximation polynomiale par morceaux en distance de Hausdorff entre les graphes. Plus précisément, on montre que lorsque la solution initiale u_0 peut être approchée en distance de Hausdorff avec une précision de l'ordre de $N^{-\alpha}$ par une suite de fonctions polynomiales de degré fixé sur N morceaux, cette propriété est vérifiée en tout temps par les solutions faibles entropiques $u(t)$ issues de u_0. Inspirée par les travaux de DeVore et Lucier, la preuve de ce résultat repose sur la caractérisation des ordres d'approximation élevés dans L^∞ par des propriétés de régularité mesurées dans des espaces de Besov.

10.1 Présentation du résultat

Dans le contexte des lois de conservation uni-dimensionnelles à flux convexes, on interprétera la distance de Hausdorff entre deux solutions comme une distance L^∞ entre leurs graphes vus dans un repère incliné.

10.1.1 Rotation et inclinaison des graphes

On suppose donc que le flux f est de classe $\mathcal{C}^2$ et vérifie

$$0 < A \leq f''. \tag{10.1}$$

Dans ce cas, l'inégalité d'Oleinik (8.25) s'applique et nous apprend qu'à tout instant $t > 0$, la solution entropique $u = u(t, \cdot)$ de (8.2) vérifie

$$-\infty \leq u' \leq \frac{1}{At}. \tag{10.2}$$

Il est alors assez clair (voir figure 10.1) que son graphe complété G_u sera celui d'une fonction lipschitzienne dans un repère tourné d'un angle adéquat. Plus précisément, on désignera par $\mathcal{R}u$ la fonction dont le graphe est l'image de G_u par la rotation

$\Phi = \Phi_{A,t}\colon (x,z) \to (\bar{x},\bar{z})$ définie par

$$\begin{cases} \bar{x} = \mathsf{c}x - \mathsf{s}z \\ \bar{z} = \mathsf{s}x + \mathsf{c}z, \end{cases} \tag{10.3}$$

où $\theta \in]0,\pi/2[$, $\mathsf{c} := \cos\theta > 0$ et $\mathsf{s} := \sin\theta > 0$ sont tels que

$$\tau := \mathsf{s}/\mathsf{c} = \tan\theta = At/2. \tag{10.4}$$

Pour se convaincre qu'il s'agit bien d'une fonction, on peut observer que Φ est continue et qu'elle transporte G_u en une courbe $\Phi(G_u)$ qui ne se recouvre pas. En effet, deux points (x,z) et (x',z') de G_u avec $x \le x'$ sont tels que $z'-z \le (x'-x)/(2\tau)$ d'après (10.2), par conséquent leurs images vérifient

$$\bar{x}' - \bar{x} = \mathsf{c}(x'-x) - \mathsf{s}(z'-z) \ge (x'-x)\mathsf{c}/2 \ge 0. \tag{10.5}$$

D'autre part, on a

$$\bar{z}' - \bar{z} = \mathsf{c}(z'-z) + \mathsf{s}(x'-x) \le (x'-x)(\frac{\mathsf{c}}{2\tau} + \mathsf{s}) \le (\bar{x}'-\bar{x})(\tau^{-1} + 2\tau) \tag{10.6}$$

en utilisant successivement (10.2) et (10.5), tandis que

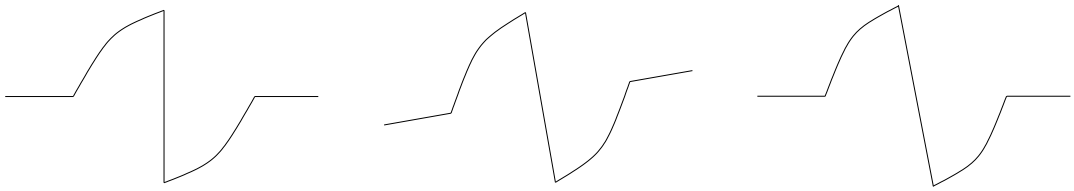
$$\bar{x}'-\bar{x} = \mathsf{c}(x'-x)-\mathsf{s}(z'-z) \ge -\mathsf{s}(z'-z) \;\text{ et }\; \bar{z}'-\bar{z} = \mathsf{s}(x'-x)+\mathsf{c}(z'-z) \ge \mathsf{c}(z'-z) \tag{10.7}$$

se déduisent de l'ordre $x \le x'$, et entraînent

$$\bar{z}' - \bar{z} \ge -\tau^{-1}(\bar{x}' - \bar{x}). \tag{10.8}$$

$\mathcal{R}u$ est donc bien lipschitzienne et vérifie

$$-\tau^{-1} \le (\mathcal{R}u)' \le \tau^{-1} + 2\tau. \tag{10.9}$$



u $\qquad$ $\mathcal{R}u$ $\qquad$ $\tilde{\mathcal{R}}u$

FIG. 10.1 – changement du système de coordonnées pour obtenir une fonction lipschitzienne à partir d'une fonction semi-lipschitzienne : rotation des graphes et retour à un support compact ("inclinaison").

De façon assez claire, $\mathcal{R}u$ n'est plus une fonction à support compact (du moins pas si u l'était), car elle vaut $\mathcal{R}u(\bar{x}) = \tau\bar{x}$ en dehors de la région correspondant au support de u. Pour préserver le caractère compact des supports, on corrige $\mathcal{R}u$ en posant

$$\tilde{\mathcal{R}}u(\bar{x}) := \mathcal{R}u(\bar{x}) - \tau\bar{x}, \tag{10.10}$$

ce qui revient encore à dire que $\tilde{\mathcal{R}}u$ est la fonction dont le graphe est l'image de G_u par l'application $\tilde{\Phi} = \tilde{\Phi}_{A,t}\colon (x,z) \to (\tilde{x},\tilde{z})$ définie par

$$\begin{cases} \tilde{x} = \bar{x} = \mathsf{c}x - \mathsf{s}z \\ \tilde{z} = \mathsf{c}^{-1}z. \end{cases} \tag{10.11}$$

Dans la suite, on parlera de fonction *inclinée* pour désigner $\tilde{\mathcal{R}}u$. Clairement, il s'agit d'une fonction lipschitzienne

$$\|(\mathcal{R}u)'\|_{L^\infty} \leq \nu \tag{10.12}$$

avec

$$\nu := \tau^{-1} + \tau = \frac{2}{At} + \frac{At}{2} \tag{10.13}$$

On pourra également observer que lorsque u est une fonction de BV, alors $\tilde{\mathcal{R}}u$ l'est aussi et vérifie

$$|\tilde{\mathcal{R}}u|_{BV(\tilde{I})} \leq \mathsf{c}^{-1}|u|_{BV(I)}, \tag{10.14}$$

ce qu'on peut voir comme une conséquence immédiate du fait que la variation totale s'écrit en dimension 1

$$|u|_{BV} = \sup \sum_{i=1}^{n} |u(x_i) - u(x_{i-1})|, \tag{10.15}$$

où la borne supérieure est prise sur tous les nuages de points $x_0 < \cdots < x_n$ dans le support de u.

10.1.2 Stabilité uniforme des solutions inclinées

On a alors la proposition suivante.

Proposition 10.1 *Dans ce nouveau repère, les lois de conservation scalaires* (8.1) *uni-dimensionnelles vérifient une propriété de stabilité L^∞ vis-à-vis des perturbations de la donnée initiale u_0 ou du flux f. Plus précisément, si u et v sont respectivement solutions entropiques de*

$$\partial_t u(t,x) + \partial_x[f(u(t,x))] = 0, \quad u(0,\cdot) = u_0, \tag{10.16}$$

et

$$\partial_t v(t,x) + \partial_x[g(v(t,x))] = 0, \quad v(0,\cdot) = v_0 \tag{10.17}$$

pour des flux de classe $\mathcal{C}^2$ vérifiant

$$0 < A \leq f'' \leq B \quad et \quad 0 < A \leq g'' \leq B \tag{10.18}$$

(la majoration par B étant toujours possible d'après la remarque 9.6), et si la solution initiale u_0 est semi-lipschitzienne supérieurement

$$(u_0)' \leq L, \tag{10.19}$$

alors on a (en identifiant $u = u(t,\cdot)$ et $v = v(t,\cdot)$)

$$\|\tilde{\mathcal{R}}u - \tilde{\mathcal{R}}v\|_{L^\infty} \leq C(t)[\|u_0 - v_0\|_{L^\infty} + \|f' - g'\|_{L^\infty}] \tag{10.20}$$

avec une constante C qui dépend de t comme $\nu(t)(1+t)$.

Preuve. D'après le théorème 9.2, les solutions u et $\tilde{u}$ issues respectivement de (10.16) et de

$$\partial_t \tilde{u}(t,x) + \partial_x[f(\tilde{u}(t,x))] = 0, \quad \tilde{u}(0,\cdot) = v_0, \tag{10.21}$$

vérifient $d_H(u,\tilde{u}) \leq C(t) d_H(u_0,v_0) \leq C(t)\|u_0 - v_0\|_{L^\infty}$. En utilisant le théorème 9.3, on trouve alors $d_H(\tilde{u},v) \leq C(t)\|f' - g'\|_{L^\infty}$, et on en déduit

$$d_H(u,v) \leq C(t)[\|u_0 - v_0\|_{L^\infty} + \|f' - g'\|_{L^\infty}]. \tag{10.22}$$

Il ne nous reste donc plus qu'à vérifier que la distance uniforme entre $\tilde{\mathcal{R}}u$ et $\tilde{\mathcal{R}}v$ est contrôlée par la distance de Hausdorff entre u et v. Pour le voir, on peut commencer par constater que la rotation des graphes (10.3) stabilise la distance de Hausdorff

$$d_H(\mathcal{R}u,\mathcal{R}v) \leq C d_H(u,v) \tag{10.23}$$

avec une constante absolue (qui serait égale à 1 si la distance d_H était construite à partir de la distance euclidienne sur $\mathbb{R}^2$ plutôt qu'avec (9.6)). En utilisant alors le fait que $\mathcal{R}u$ est lipschitzienne (10.9), on peut déduire de l'inégalité (9.14) que

$$\|\mathcal{R}u - \mathcal{R}v\|_{L^\infty} \leq (1 + 2\tau + \tau^{-1}) d_H(\mathcal{R}u,\mathcal{R}v) \leq C\nu(t) d_H(\mathcal{R}u,\mathcal{R}v). \tag{10.24}$$

Comme $\tilde{\mathcal{R}}u(x) - \tilde{\mathcal{R}}v(x) = \mathcal{R}u(x) - \tau x - (\mathcal{R}v(x) - \tau x) = \mathcal{R}u(x) - \mathcal{R}v(x)$, on trouve enfin

$$\|\tilde{\mathcal{R}}u - \tilde{\mathcal{R}}v\|_{L^\infty} = \|\mathcal{R}u - \mathcal{R}v\|_{L^\infty} \leq C\nu(t) d_H(\mathcal{R}u,\mathcal{R}v) \leq C\nu(t) d_H(u,v), \tag{10.25}$$

ce qui établit (10.20). On pourra d'ailleurs observer que l'inégalité précédente est en réalité une équivalence, dans la mesure où

$$d_H(u,v) \leq C d_H(\mathcal{R}u,\mathcal{R}v) \leq \|\mathcal{R}u - \mathcal{R}v\|_{L^\infty} = \|\tilde{\mathcal{R}}u - \tilde{\mathcal{R}}v\|_{L^\infty}. \tag{10.26}$$

□

10.1.3 Le théorème de régularité

On énonce à présent notre principal résultat. Sa preuve est basée sur une approximation des solutions exactes par des fonctions constantes par morceaux. On décrira en détails cette construction dans la section 10.2, et on y donnera une estimation inverse (dont la preuve est très technique) qui fera fonctionner la preuve de notre théorème dans la section 10.3.

Théorème 10.1 *On rappelle le flux f satisfait ici une hypothèse de forte convexité* (10.1)*. Si u_0 est une fonction semi-lipschitzienne supérieurement $u_0' \leq L$ à support compact, alors pour tout indice de régularité $\alpha > 1$ et tout instant $t > 0$, la solution entropique $u = u(t,\cdot)$ de la loi de conservation uni-dimensionnelle* (8.1) *vérifie*

$$\|\tilde{\mathcal{R}}u\|_{\tilde{B}^\alpha} \leq C(\|u_0\|_{\tilde{B}^\alpha} + 1) \tag{10.27}$$

avec une constante C indépendante de u_0, dès lors que le flux est assez régulier, i.e. *appartient à $W^{r+3,\infty}_{\mathrm{loc}}$ pour un entier $r > \alpha - 1$.*

Remarque 10.2 *L'hypothèse de régularité uniforme sur f n'est en général pas restrictive, les flux étant d'habitude très réguliers. On pourrait toutefois adapter ce résultat à des situations où f n'est pas uniformément régulière mais est bien approchée dans L^∞ par des polynômes par morceaux, par exemple lorsqu'elle est $W^{r+3,\infty}$ "par morceaux".*

10.2 Approximation polynomiale par morceaux des solutions

10.2.1 Construction des solutions initiales approchées

Dans la section 1.3.2, on a vu que les fonctions de $\tilde{B}^\alpha$ sur un intervalle I pouvaient être approchées dans $L^\infty(I)$ par des fonctions polynomiales par morceaux avec une précision de l'ordre de $N^{-\alpha}$, où N désigne la complexité des approximants. Plus précisément, on a introduit les ensembles $\Sigma_n = \Sigma_{n,r}$ des fonctions continues et polynomiales de degré au plus r sur 2^n intervalles de I, et on a montré que lorsque $r > \alpha - 1$, il existait pour toute fonction $u_0 \in \tilde{B}^\alpha$ une suite $S_n \in \Sigma_n$, $n \in \mathbb{N}$ approchant u_0 dans L^∞ de façon à ce que

$$\left(\sum_{n=-1}^{\infty} [2^{n\alpha} \|u_0 - S_n\|_{L^\infty}]^q \right)^{1/q} \leq C \|u_0\|_{\tilde{B}^\alpha} \tag{10.28}$$

avec une constante C absolue, $1/q = \alpha$ et $S_{-1} = 0$.

On va maintenant considérer une suite S_n approchant de cette façon la donnée initiale u_0 de notre théorème 10.1, et pour les besoins de notre construction, on aura besoin que les approximants aient la même propriété semi-lipschitzienne que u_0, à savoir

$$S_n' \leq L \quad \text{pour} \quad n \in \mathbb{N}. \tag{10.29}$$

Pour cela, on peut reprendre les approximants T_n polynomiaux par morceaux de u_0' introduits dans la deuxième partie de la preuve du théorème 1.2, qui vérifient

$$\sum_{n=-1}^{\infty} \|u_0' - T_n\|_{L^1}^q \leq C \|u_0'\|_{B^{\alpha-1,q}q}^q \tag{10.30}$$

avec une constante C absolue, $1/q = \alpha$ et $T_{-1} = 0$. Pour chaque T_n, on avait montré l'existence d'une partition de I en 2^{n+1} intervalles I_k sur lesquels T_n était polynôme de degré au plus $r-1$, et tels que

$$\|u_0' - T_n\|_{L^1(I_k)} \leq \frac{1}{n} \|u_0' - T_n\|_{L^1}. \tag{10.31}$$

Sur chaque I_k, on peut définir une nouvelle approximation polynomiale R_{n+1} de u_0' en projetant orthogonalement cette dernière fonction sur les polynômes de degré $r-1$, de sorte que la restriction de R_{n+1} à chaque I_k est déterminée par les r relations

$$\int_{I_k} [u_0'(x) - R_{n+1}(x)] x^s \, \mathrm{d}x = 0, \quad \text{pour} \quad s = 0, \ldots, r-1. \tag{10.32}$$

D'après le lemme 1.2, cette projection réalise une erreur d'approximation quasi-optimale dans L^1. On a donc

$$\|u_0' - R_{n+1}\|_{L^1(I_k)} \leq C \|u_0' - T_n\|_{L^1(I_k)}. \tag{10.33}$$

On peut d'autre part vérifier qu'il existe au plus $(r+1)/2$ intervalles disjoints à l'intérieur de I_k sur lesquels $R_{n+1}(x) > L$. Sur chacun d'entre eux, on remplace R_{n+1} par

L, et par $L - c(L - R_{n+1})$ sur la partie restante $\tilde{I}_k$ de I_k, où c est choisie de façon que l'intégrale de R_{n+1} sur I_k demeure inchangée. Dans la mesure où cette intégrale vaut

$$\int_{I_k} R_{n+1} = \int_{I_k} u_0' \leq L|I_k|, \tag{10.34}$$

la constante

$$c = \frac{\int_{I_k}[L - R_{n+1}]}{\int_{\tilde{I}_k}[L - R_{n+1}]} \tag{10.35}$$

appartient à $[0, 1]$, et on en déduit que $L - c(L - R_{n+1}) \leq L$ sur $\tilde{I}_k$. La fonction U_{n+a} ainsi obtenue est polynomiale de degré inférieur ou égal à r sur au plus 2^{n+a} intervalles avec $a = 1 + \log_2(r + 1)$, et vérifie $U_{n+a} \leq L$ en tout point. Enfin, on peut vérifier que les modifications effectuées sur R_{n+1} n'ont pu que diminuer l'erreur d'approximation dans L^1. En effet, on a d'une part

$$\|u_0' - U_{n+a}\|_{L^1(I_k \setminus \tilde{I}_k)} \leq \|u_0' - R_{n+1}\|_{L^1(I_k \setminus \tilde{I}_k)} - \int_{I_k \setminus \tilde{I}_k} [R_{n+1} - L], \tag{10.36}$$

et d'autre part

$$\begin{aligned}
\|u_0' - U_{n+a}\|_{L^1(\tilde{I}_k)} &= \|u_0' - L - c(R_{n+1} - L)\|_{L^1(\tilde{I}_k)} \\
&\leq \|u_0' - R_{n+1}\|_{L^1(\tilde{I}_k)} + (1 - c)\|L - R_{n+1}\|_{L^1(\tilde{I}_k)} \\
&\leq \|u_0' - R_{n+1}\|_{L^1(\tilde{I}_k)} + \left(\int_{I_k} [L - R_{n+1}] - \int_{\tilde{I}_k} [L - R_{n+1}] \right) \\
&\leq \|u_0' - R_{n+1}\|_{L^1(\tilde{I}_k)} + \int_{I_k \setminus \tilde{I}_k} [L - R_{n+1}].
\end{aligned} \tag{10.37}$$

On a donc

$$\|u_0' - U_{n+a}\|_{L^1(I_k)} \leq \|u_0' - R\|_{L^1(I_k)} \leq C\|u_0' - T_n\|_{L^1(I_k)}. \tag{10.38}$$

On définit alors S_{n+a} sur chaque intervalle $I_k = [a_k, b_k]$ par

$$S_{n+a}(x) := u_0(a_k) + \int_{a_k}^{x} U_{n+a}(s)\,\mathrm{d}s. \tag{10.39}$$

On obtient bien de cette façon une fonction de Σ_{n+a}, la continuité de S_{n+a} étant garantie par le fait que u_0 est elle-même continue et qu'on a par construction $\int_{I_k} U_{n+a} = \int_{I_k} u_0'$. D'autre part, la propriété (10.29) est clairement vérifiée. Enfin, on observe aisément que

$$\|u_0 - S_{n+a}\|_{L^\infty} \leq C \sup_k \|u_0' - T_n\|_{L^1(I_k)} \leq 2^{-n}\|u_0' - T_n\|_{L^1}, \tag{10.40}$$

et on en déduit que la suite S_n vérifie bien l'inégalité (10.28) avec une constante C absolue, $1/q = \alpha$ et $S_n = 0$ pour $-1 \leq n < a$. On peut à ce sujet remarquer que la suite S_n n'approche pas forcément u_0 de façon quasi-optimale dans L^∞, mais l'inégalité (10.28) sera suffisante pour nos besoins. Observons également que les dérivées $U_n = S_n'$ forment une suite d'approximations quasi-optimales de u_0' dans L^1. On en déduit que

$$\|S_n'\|_{L^1} \leq C\|u_0'\|_{L^1} \tag{10.41}$$

avec une constante absolue, de sorte que

$$|S_n|_{BV} \leq C|u_0|_{BV}. \tag{10.42}$$

10.2.2 Approximation du flux

On approche ensuite la fonction de flux f qu'on suppose appartenir à $W^{r+3,\infty}(\Omega)$, où Ω désigne un intervalle contenant les valeurs prises par u_0 ou un des S_n. D'après (10.42), la longueur de cet intervalle sera de l'ordre de $|u_0|_B V$. Un résultat classique d'approximation par splines nous dit alors qu'il existe pour chaque entier n une fonction g_n de classe $\mathcal{C}^{r+1}$ qui est polynomiale de degré inférieur ou égal à $r+2$ sur les intervalles $[j2^{-n}, (j+1)2^{-n}]$ avec $j \in \mathbb{Z}$, et vérifie

$$\|f^{(l)} - g_n^{(l)}\|_{L^\infty(\Omega)} \leq C 2^{-n(r+3-l)} \|f^{(r+3)}\|_{L^\infty(\Omega)} \quad \text{pour} \quad l = 0, \dots r+2. \tag{10.43}$$

On rappelle d'autre part que le flux f est fortement convexe (10.1), et vérifie même d'après la remarque 9.6

$$0 < A \leq f'' \leq B \tag{10.44}$$

avec une constante B dépendant de u_0. L'inégalité (10.43) ci-dessus prise avec $l = 2$ nous montre alors que quitte à modifier légèrement les constantes A et B, on peut supposer que les g_n vérifient également

$$0 < A \leq g'' \leq B. \tag{10.45}$$

Pour chaque entier n, on définit alors $s_n = s_n(t, \cdot)$ comme la solution entropique à l'instant t de la loi de conservation

$$\partial_t s_n(t,x) + \partial_x[g_n(s_n(t,x))] = 0, \quad s_n(0,\cdot) = S_n \tag{10.46}$$

associée aux approximations polynomiales par morceaux du flux f et de la donnée initiale u_0. Avant de décrire en détail la structure de ces solutions et de leurs inclinaisons $\tilde{\mathcal{R}} s_n$ dans le repère (10.11), on peut observer que la proposition 10.1 s'applique : on a

$$\|\tilde{\mathcal{R}} u - \tilde{\mathcal{R}} s_n\|_{L^\infty} \leq C(t)[\|u_0 - S_n\|_{L^\infty} + 2^{-nr}] \tag{10.47}$$

ainsi que

$$\|\tilde{\mathcal{R}} s_{n+1} - \tilde{\mathcal{R}} s_n\|_{L^\infty} \leq C(t)[\|S_{n+1} - S_n\|_{L^\infty} + 2^{-nr}]. \tag{10.48}$$

Autrement dit, la suite $\tilde{\mathcal{R}} s_n$ approche $\tilde{\mathcal{R}} u$ avec la même vitesse que la suite S_n approche la donnée initiale u_0, au terme additif 2^{-nr} près. Cette propriété sera fortement utilisée dans la preuve du théorème 10.1.

10.2.3 Structure des solutions approchées

En premier lieu, il convient de rappeler que les solutions inclinées sont lipschitziennes et vérifient

$$\|(\tilde{\mathcal{R}} s_n)'\|_{L^\infty} \leq \nu \tag{10.49}$$

où ν est défini en (10.13).

Le lemme suivant décrit la structure de chaque $\tilde{\mathcal{R}} s_n$ en termes de fonctions *algébriques par morceaux*. Rappelons qu'une fonction $z = z(x)$ est algébrique sur un intervalle J s'il existe un polynôme P de deux variables tel que $P(x, y(x)) = 0$ pour $x \in J$.

Lemme 10.3 *Il existe une partition du support compact de $\tilde{\mathcal{R}} s_n$ en $\mathcal{O}(2^n)$ intervalles J sur lesquels $\tilde{\mathcal{R}} s_n$ coïncide avec une fonction algébrique z d'un des deux types suivant :*

Type I : *z est solution de l'équation algébrique*

$$R(T(x)) = z(x) + \nu x \quad sur \quad J, \tag{10.50}$$

où le polynôme T est donné par

$$T(x) := z(x) + \nu x - Q(z(x)) \quad sur \quad J, \tag{10.51}$$

R et Q étant deux polynômes de degré au plus $r(r+1)$ et $r+1$ tels que

$$2 \leq Q' \leq c_1 \quad sur \quad z(J) \tag{10.52}$$

$$0 < R' \leq c_2 \quad sur \quad T(J) \tag{10.53}$$

pour deux constantes c_1 et c_2 indépendantes de n.

Type II : *z vérifie*

$$z(0) = z(x) + \nu x \quad sur \quad J, \tag{10.54}$$

i.e. *$\tilde{\mathcal{R}} s_n$ est affine sur J de pente $-\nu$.*

Preuve. Par composition, il est clair que la fonction $g_n'(S_n)$ est polynomiale par morceaux, de degré inférieur ou égal à r(r+1). Pour compter le nombre de ces morceaux, on suit la démarche de DeVore et Lucier [32] en répartissant les nœuds de $g_n'(S_n)$ en deux types particuliers. Par $\{a_i\}_{1\leq i\leq A}$, on désigne d'abord les nœuds de S_n, autrement dit les bornes des intervalles sur lesquels S_n est polynomial. Par construction, $A \leq 2^n$. On note ensuite $\{b_i\}_{0\leq i\leq B}$ les points *isolés* sur lesquels $S_n(b_i)$ est un nœud du flux approché g_n, autrement dit tels que $S_n(b_i) = j2^{-n}$ pour un $j \in \mathbb{Z}$. Pour dénombrer ces points, on désigne par $\{\tilde{b}_i\}_{0\leq i\leq \tilde{B}}$ ceux des b_j sur lesquels S_n se répète, *i.e.* $S_n(b_j) = S_n(b_{j-1})$. En notant $\{\bar{b}_i\}_{0\leq i\leq \bar{B}}$ les autres, on voit que $|S_n|_{BV[\bar{b}_i,\bar{b}_{i+1}]} \geq 2^{-n}$ pour chaque $i \leq \bar{B}$, de sorte que

$$|S_n|_{BV} = \sum_{i=0}^{\bar{B}-1} |S_n|_{BV[\bar{b}_i,\bar{b}_{i+1}]} \geq \bar{B}2^{-n}, \tag{10.55}$$

l'égalité venant du fait que S_n est une fonction continue. D'après (10.42), ceci entraîne $\bar{B} \leq C|u_0|_{BV}2^n$. D'un autre côté, si S_n coïncide avec un polynôme P_k sur un intervalle I_k, P_k' doit s'annuler au moins une fois par segment $[\tilde{b}_i, \tilde{b}_{i+1}]$ inclus dans I_k. Le degré de P_k n'excédant pas r, sa dérivée s'annule au plus $\mathcal{O}(r)$ fois sous peine d'être globalement nulle, mais dans ce cas P_k est constante et I_k ne contient aucun b_i. Il ne peut donc exister que $\mathcal{O}(r)$ points $\tilde{b}_i$ par intervalle I_k, et on en déduit que $\tilde{B}$ est de l'ordre de $\mathcal{O}(2^n)$, r étant considéré comme une constante. On a donc $B = \mathcal{O}(2^n)$, et $g_n'(S_n)$ est une fonction polynomiale sur $\mathcal{O}(2^n)$ morceaux.

Dans la section 8.2, on a vu qu'il existait une application $x \to y(x) = y(t,x)$ croissante telle que

$$s_n(x) = S_n(y(x)) \quad \text{et} \quad x = y(x) + tg_n'(S_n(y(x))). \tag{10.56}$$

Les points de discontinuité de cette fonction correspondent aux chocs σ_i de la solution s_n, et en ces points les limites à gauche et à droite $y_i^g := y(\sigma_i^-)$ et $y_i^d := y(\sigma_i^+)$ sont respectivement le plus petit et le plus grand minimiseur global de la fonctionnelle

(8.23). Géométriquement, on peut se représenter l'application y par une projection $\mathcal{P}$ de l'axe x sur le graphe complété de $g_n'(S_n)$ dans la direction $(-t, 1)$, chaque valeur $y(x)$ étant donnée par l'abscisse du point projeté $\mathcal{P}(0, x)$ (voir figure 10.2). Cette projection "enjambe" par endroits le graphe de $g_n'(S_n)$: lorsque plusieurs points de ce graphe sont situés sur une même droite $\{(x - st, s) : s \in \mathbb{R}\}$, c'est la propriété de minimisation de la fonctionnelle (8.23) qui décide lequel de ces points correspond à $y(x)$. D'autre part, le fait que la fonction y soit croissante entraîne que les points y_i^g et y_i^d s'ordonnent suivant

$$\cdots < y_i^g < y_i^d < y_{i+1}^g < \cdots \tag{10.57}$$

si les σ_i sont eux-mêmes ordonnés de façon croissante. La réciproque (éventuellement

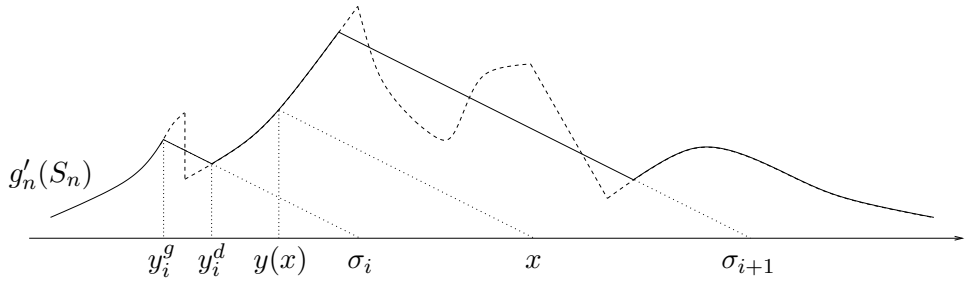


FIG. 10.2 – interprétation géométrique de la formule de Lax par l'application $y(x)$ vérifiant (10.56). Les zones en pointillés sur graphe de $g_n'(S_n)$ ont été absorbées au cours du transport par les chocs σ_i entre l'instant initial et l'instant t.

discontinue) de la fonction y coïncide donc avec

$$\mathcal{A} \colon y \to y + t g_n'(S_n(y)) \tag{10.58}$$

sur chaque intervalle $[y_i^d, y_{i+1}^g]$, et est constante égale à σ_i sur les intervalles $[y_i^g, y_i^d]$. D'après l'analyse ci-dessus, on en déduit que $\mathcal{A}$ est polynomiale de degré inférieur ou égal à $r(r+1)$ sur $\mathcal{O}(2^n)$ intervalles. Le nombre de chocs est donc lui même en $\mathcal{O}(2^n)$, et en prenant la réunion des nœuds a_i, b_i, y_i^g et y_i^d, on construit une partition $\{I_k^0\}_{1 \le k \le C2^n}$ telle que sur chaque intervalle I_k^0, $\mathcal{A}$ est un polynôme croissant qui vérifie

$$s_n(\mathcal{A}(y)) = S_n(y). \tag{10.59}$$

L'application $\mathcal{A}$ correspond donc au déplacement des abscisses lors du transport associé à l'équation (8.1). Naturellement, on désignera par $\{I_k^t := \mathcal{A}(I_k^0)\}_{1 \le k \le C2^n}$ la partition "transportée", et on peut observer que l'égalité ci-dessus entraîne

$$s_n(I_k^t) = S_n(I_k^0). \tag{10.60}$$

En remplaçant $x = \mathcal{A}(y)$ dans (10.58) et dans (10.59), on trouve alors

$$y = x - t g_n'(s_n(x)) \quad \text{pour } x \in I_k^t, \tag{10.61}$$

et en y appliquant $\mathcal{A}$,

$$x = \mathcal{A}(x - t g_n'(s_n(x))) \quad \text{sur } I_k^t. \tag{10.62}$$

Les intervalles I_k^0 étant construits de façon que g_n coïncide avec un polynôme sur chaque $S_n(I_k^0)$, on voit d'après (10.60) que g_n coïncide avec ce même polynôme sur

$s_n(I_k^t)$. On déduit alors de la dernière inégalité que s_n est bien algébrique sur I_j^t.

Il nous faut encore préciser la forme des morceaux algébriques de la solution inclinée $\tilde{\mathcal{R}}s_n$ dans le système (10.11). Le fait est que les morceaux (s_n, I_k^t) construits ci-dessus, une fois inclinés, deviennent des morceaux de type I, tandis que les chocs deviennent des morceaux de type II (voir par exemple la figure 10.1). Pour les chocs, ce que nous disons là est évident. Pour un morceau algébrique général, ça l'est moins. On considère donc un intervalle I_k^0 fixé (non réduit à un point), et on rappelle que S_n, g_n' et $\mathcal{A} := I_d + tg_n'(S_n)$ coïncident avec des polynômes sur les intervalles respectifs I_k^0, $S_n(I_k^0)$ et à nouveau I_k^0. On désigne alors respectivement par P, Q et $R := Id + Q \circ P$ les polynômes vérifiant

$$P = \mathsf{c}^{-1} S_n(\mathsf{s}\cdot) \qquad \text{sur} \quad \mathsf{s}^{-1} I_k^0 \tag{10.63}$$

$$Q = \mathsf{s}^{-1} tg_n'(\mathsf{c}\cdot) \qquad \text{sur} \quad \mathsf{c}^{-1} S_n(I_k^0) \tag{10.64}$$

$$R = \mathsf{s}^{-1} \mathcal{A}(\mathsf{s}\cdot) \qquad \text{sur} \quad \mathsf{s}^{-1} I_k^0. \tag{10.65}$$

En notant par $(\tilde{x}, \tilde{\mathcal{R}}s_n(\tilde{x}))$ le point $(x, s_n(x))$ dans le repère incliné (10.11), on trouve alors que

$$\tilde{x} = \mathsf{c}x - \mathsf{s}s_n(x) = \mathsf{c}x - \mathsf{sc}\tilde{\mathcal{R}}s_n(\tilde{x}) \tag{10.66}$$

et

$$tg_n'(s_n(x)) = tg_n'(\mathsf{c}\tilde{\mathcal{R}}s_n(\tilde{x})) = \mathsf{s}Q(\tilde{\mathcal{R}}s_n(\tilde{x})). \tag{10.67}$$

En important ces deux égalités dans (10.62), on obtient

$$\mathsf{c}^{-1}\tilde{x} + \mathsf{s}\tilde{\mathcal{R}}s_n(\tilde{x}) = x = \mathcal{A}(\mathsf{c}^{-1}\tilde{x} + \mathsf{s}\tilde{\mathcal{R}}s_n(\tilde{x}) - \mathsf{s}Q(\tilde{\mathcal{R}}s_n(\tilde{x}))), \tag{10.68}$$

ce qui s'écrit également, en observant que $\nu = \mathsf{sc}^{-1} + \mathsf{cs}^{-1} = (\mathsf{sc})^{-1}$,

$$\nu\tilde{x} + \tilde{\mathcal{R}}s_n(\tilde{x}) = R(\tilde{\mathcal{R}}s_n(\tilde{x}) + \nu\tilde{x} - Q(\tilde{\mathcal{R}}s_n(\tilde{x}))). \tag{10.69}$$

On retrouve donc bien la forme annoncée (10.50)-(10.51), aux notations près (en particulier, les abscisses x sont ici notées $\tilde{x}$). Observons que l'intervalle J correspond à $\tilde{I}_k^t = \{\tilde{x}(x) : x \in I_k^t\}$, autrement dit à la projection sur l'axe des abscisses du la tranche de graphe inclinée $\tilde{\Phi}\left(G_{s_n} \cap (I_k^t \times \mathbb{R})\right)$ associée au morceau algébrique (s_n, I_k^t). On peut d'ailleurs faire l'observation suivante qui nous servira dans la suite : d'après la forme (10.11) de l'application $\tilde{\Phi}$, on voit que $\tilde{\mathcal{R}}s_n(\tilde{I}_k^t) = \mathsf{c}^{-1}s_n(I_k^t)$. Ce dernier intervalle valant $\mathsf{c}^{-1}S_n(I_k^0)$ d'après (10.60), on peut déduire de (10.64) que

$$Q = \mathsf{s}^{-1} tg_n'(\mathsf{c}\cdot) \qquad \text{sur} \quad \tilde{\mathcal{R}}s_n(J). \tag{10.70}$$

Enfin, on peut calculer que

$$Q' = \frac{t}{\tau} g_n''(c\cdot) \quad \text{et} \quad R' = 1 + tg_n''(S_n(s\cdot))S_n'(s\cdot), \tag{10.71}$$

et les inégalités (10.52)-(10.53) découlent directement des bornes (10.29) et (10.45), avec $c_1 = 2B/A$ et $c_2 = 1 + tBL$. □

10.2.4 Une estimation inverse

D'après le lemme précédent, chaque différence $\tilde{\mathcal{R}}s_n - \tilde{\mathcal{R}}s_{n-1}$ peut s'écrire comme la somme de $\mathcal{O}(2^n)$ morceaux algébriques (z_k, J_k) qui sont des différences de morceaux de type I et de type II. En suivant la démarche de DeVore et Lucier (plus précisément, le lemme 4.2 de [32]), on peut découper à nouveau chaque intervalle J_k de façon à obtenir une partition composée de $\mathcal{O}(2^n)$ intervalles I_k sur lesquels la différence $\tilde{\mathcal{R}}s_n - \tilde{\mathcal{R}}s_{n-1}$ est algébrique et monotone, ainsi que toutes ses dérivées jusqu'à l'ordre $r+1$. On peut alors énoncer l'estimation suivante, qu'on appelle "inverse" parce qu'elle mesure la régularité des approximants, à la façon d'une inégalité de Bernstein (1.71).

Lemme 10.4 *Si (z, J) est un morceau algébrique d'une différence $\tilde{\mathcal{R}}s_n - \tilde{\mathcal{R}}s_{n+1}$ vérifiant les hypothèses de monotonie énoncées ci-dessus, alors*

$$\|z'\chi_J\|_{B_q^{\alpha-1,q}} \leq C\left[\|z\|_{L^\infty(J)} + 2^{-(r+1)n}\right] \tag{10.72}$$

avec une constante indépendante de n.

10.3 Preuve du théorème de régularité

A partir du lemme 10.4, on peut écrire une estimation inverse sur les différences $\tilde{\mathcal{R}}s_n - \tilde{\mathcal{R}}s_{n-1}$ de deux solutions approchées successives (et inclinées). Désignons en effet par $\{(z_k, J_k\}_{1\leq k\leq C2^n}$ la décomposition de $\tilde{\mathcal{R}}s_n - \tilde{\mathcal{R}}s_{n-1}$ en morceaux algébriques monotones construite dans la section 10.2.4. Comme les solutions inclinées $\tilde{\mathcal{R}}s_n$ sont continues, on a

$$(\tilde{\mathcal{R}}s_n)' - (\tilde{\mathcal{R}}s_{n-1})' = \sum_{k=1}^{C2^n} z_k'\chi_{J_k}. \tag{10.73}$$

La q-inégalité triangulaire de l'espace $B_q^{\alpha-1,q}$ nous permet alors d'écrire

$$\begin{aligned}
\|(\tilde{\mathcal{R}}s_n)' - (\tilde{\mathcal{R}}s_{n-1})'\|_{B_q^{\alpha-1,q}}^q &\leq \sum_{k=1}^{C2^n} \|z_k'\chi_{J_k}\|_{B_q^{\alpha-1,q}}^q \\
&\leq C\sum_{k=1}^{C2^n} [\|z_k\|_{L^\infty(J_k)} + 2^{-(r+1)n}]^q \\
&\leq C\left[2^n\|\tilde{\mathcal{R}}s_n - \tilde{\mathcal{R}}s_{n-1}\|_{L^\infty}^q + 2^{-n((r+1)q-1)}\right]
\end{aligned} \tag{10.74}$$

en utilisant dans la dernière inégalité l'équivalence entre les normes ℓ^1 et ℓ^q en dimension finie. D'après la propriété (10.47) des solutions approchées, on a donc

$$\|(\tilde{\mathcal{R}}s_n)' - (\tilde{\mathcal{R}}s_{n-1})'\|_{B_q^{\alpha-1,q}}^q \leq C\left[2^n\|S_n - S_{n-1}\|_{L^\infty}^q + 2^{-n((r+1)q-1)}\right]. \tag{10.75}$$

On peut d'autre part observer d'après (10.48) et (10.28) que la solution inclinée $\tilde{\mathcal{R}}u$ peut s'écrire en somme télescopique

$$\tilde{\mathcal{R}}u = \sum_{n=0}^{\infty} \tilde{\mathcal{R}}s_n - \tilde{\mathcal{R}}s_{n-1}. \tag{10.76}$$

En utilisant une nouvelle fois la q-inégalité triangulaire de $B_q^{\alpha-1,q}$, on trouve alors

$$\|(\tilde{\mathcal{R}}u)'\|^q_{B_q^{\alpha-1,q}} \leq \sum_{n=0}^{\infty} \|(\tilde{\mathcal{R}}s_n)' - (\tilde{\mathcal{R}}s_{n-1})'\|^q_{B_q^{\alpha-1,q}}$$
$$\leq C \sum_{n=0}^{\infty} \left[2^n \|S_n - S_{n-1}\|^q_{L^\infty} + 2^{-n((r+1)q-1)} \right]. \quad (10.77)$$

Grâce à l'inégalité (10.28) vérifiée par la suite d'approximations S_n, on en déduit finalement

$$\|(\tilde{\mathcal{R}}u)'\|^q_{B_q^{\alpha-1,q}} \leq C[\|u_0\|^q_{\tilde{B}^\alpha} + 1] \quad (10.78)$$

où l'on a utilisé l'hypothèse $r + 1 > \alpha = 1/q$ pour faire converger la série dyadique. Le théorème 10.1 est donc prouvé si l'on établit (10.72), ce qu'on se propose de faire à présent.

10.3.1 Une estimation intermédiaire

A partir d'ici et jusqu'à la fin du chapitre 10, on considère un entier n fixé, et on désigne par (z, J) un morceau algébrique de $\tilde{\mathcal{R}}s_n - \tilde{\mathcal{R}}s_{n-1}$ qui est monotone ainsi que toutes ses dérivées jusqu'à l'ordre $r + 1$ (on rappelle que $\tilde{\mathcal{R}}s_{-1} = 0$). Pour démontrer le lemme 10.4, on va utiliser l'estimation suivante.

Lemme 10.5 *Si (z, J) a la forme annoncée ci-dessus, alors*

$$\|z'\|_{L^\infty(J)} \leq C|J|^{-1} \left[\|z\|_{L^\infty(J)} + 2^{-(r+1)n} \right] \quad (10.79)$$

avec une constante indépendante de n.

Preuve. Ecrivons z_n et $z_{n-1} = z_n - z$ les restrictions (algébriques) de $\tilde{\mathcal{R}}s_n$ et $\tilde{\mathcal{R}}s_{n-1}$ sur l'intervalle J. D'après le lemme 10.3, plusieurs cas peuvent se présenter, suivant que z_n et z_{n-1} sont des morceaux de type I ou II. Toutefois, il n'y a rien à prouver lorsque z_n et z_{n-1} sont toutes les deux de type II. On peut donc considérer que z_n est un morceau algébrique de type I. On pose alors

$$\Theta(x) := 1 - R'(T(x))(1 - Q'(z_n(x))) \quad \text{sur } J, \quad (10.80)$$

et on se propose de montrer qu'il existe des constantes $c_i > 0$, $i = 3, \ldots, 6$ indépendantes de n et de l'intervalle J, pour lesquelles on a

$$c_3 \leq |\Theta(x)| \leq c_4 \quad \text{sur } J \quad (10.81)$$

et

$$c_4|J| \leq |T(J)| \leq c_6|J|. \quad (10.82)$$

Pour établir (10.81), on peut commencer par observer en utilisant (10.52) et (10.53) que $\|\Theta(x)\|_{L^\infty(J)} \leq 1 + c_2(1 + c_1)$. Dans l'autre direction, on trouve en dérivant respectivement (10.50) et (10.51) par rapport à x

$$R'(T)T'(x) = z_n'(x) + \nu \quad (10.83)$$

et

$$T'(x) = \nu - z_n'(x)[Q'(z_n) - 1]. \tag{10.84}$$

On a donc

$$z_n'(x)\Theta(x) = \nu[R'(T) - 1]. \tag{10.85}$$

Posons alors $J_+ := \{x \in J : |1 - R'(T)| \geq 1/2|\}$ et $J_- := J \setminus J_+$. Si $x \in J_+$, alors $|z_n'(x)\Theta(x)| \geq \nu/2$ et on déduit de (10.49) que $|\Theta(x)| \geq 1/2$. Si par contre $x \in J_-$, on a $R'(T) > 1/2$ et l'inégalité de gauche de (10.52) entraîne que

$$\begin{aligned} |\Theta(x)| = |R'(T)Q'(z_n) + 1 - R'(T)| &\geq |R'(T)Q'(z_n)| - |1 - R'(T)| \\ &\geq 1/2|Q'(z_n)| - 1/2 \geq 1/2. \end{aligned} \tag{10.86}$$

On a donc $|\Theta(x)| \geq 1/2$ sur J et (10.81) est démontré. En particulier, on déduit de (10.85) que z_n vérifie toujours

$$z_n' = \nu\Theta^{-1}[R'(T) - 1]. \tag{10.87}$$

Prouvons ensuite (10.82). D'après (10.84), il est clair que

$$\|T'\|_{L^\infty(J)} \leq \nu(2 + c_1). \tag{10.88}$$

Pour minorer T' par une constante strictement positive, supposons pour commencer que z_n est croissante sur J. Dans ce cas, les égalités (10.83) et (10.53) nous donnent

$$T'(x) = (R'(T))^{-1}(z_n' + \nu) \geq (R'(T))^{-1}\nu \geq \nu c_2^{-1}. \tag{10.89}$$

Dans le cas où z_n, qui est monotone sur J, est décroissante, ce sont (10.84) et (10.52) qui nous donnent l'inégalité voulue :

$$T'(x) = \nu - z_n'[Q'(z_n) - 1] \geq \nu - z_n' \geq \nu, \tag{10.90}$$

et dans tous les cas, on a $T'(x) \geq \nu \min\{1, c_2^{-1}\}$, ce qui ajouté à (10.88) entraîne bien l'équivalence (10.82).

On rappelle les inégalités élémentaires suivantes, valables pour un polynôme P de degré inférieur ou égal à l et deux intervalles K et K' arbitraires, tels que $K \subset K'$:

$$\|P\|_{L^\infty(K')} \leq C\left(\frac{|K'|}{|K|}\right)^l \|P\|_{L^\infty(K)} \tag{10.91}$$

$$\|P'\|_{L^\infty(K)} \leq C|K|^{-1}\|P\|_{L^\infty(K)}, \tag{10.92}$$

avec des constantes absolues.

Venons-en alors à la preuve de (10.79) : lorsque z_{n-1} est de type II, sa dérivée est constante, égale à $-\nu$. D'après (10.87) et la définition (10.80) de Θ, voit que

$$z_n' - z_{n-1}' = \nu\left[\Theta^{-1}[R'(T) - 1] + 1\right] = \nu\left[\Theta^{-1}R'(T)Q'(z_n)\right]. \tag{10.93}$$

On majore alors cette quantité par

$$\begin{aligned}\|z_n' - z_{n-1}'\|_{L^\infty(J)} &\le C\|R'(T)\|_{L^\infty(J)}\\ &\le C\|R'\|_{L^\infty(T(J))}\\ &\le C|T(J)|^{-1}\|R - z_{n-1}(0)\|_{L^\infty(T(J))}\\ &\le C|J|^{-1}\|R(T) - z_{n-1}(0)\|_{L^\infty(J)}\\ &\le C|J|^{-1}\|z_n - z_{n-1}\|_{L^\infty(J)},\end{aligned} \tag{10.94}$$

où la première inégalité vient de (10.81) et (10.52) réunies, la troisième vient de (10.92), la quatrième de (10.82) et la dernière de la forme (10.50)-(10.54) des morceaux algébriques z_n et z_{n-1}. Le lemme est donc démontré lorsque z_{n-1} est de type II.

Il nous reste à considérer le cas où z_n et z_{n-1} sont toutes deux de type I. On désigne alors par $\bar R$, $\bar T$, *etc.* les polynômes associés à z_{n-1}, et on peut observer que la relation (10.87) s'applique à nouveau : on a donc

$$z_{n-1}' = \nu\bar\Theta^{-1}[\bar R'(\bar T) - 1]. \tag{10.95}$$

On en déduit que

$$z_n' - z_{n-1}' = \nu\Theta^{-1}\bar\Theta^{-1}\left[\bar\Theta[R'(T) - 1] - \Theta[\bar R'(\bar T) - 1]\right], \tag{10.96}$$

et en utilisant l'équivalence (10.81) pour Θ et $\bar\Theta$,

$$\begin{aligned}\|z_n' - z_{n-1}'\|_{L^\infty(J)} &\le C\|\bar\Theta[R'(T) - 1] - \Theta[\bar R'(\bar T) - 1]\|_{L^\infty(J)}\\ &\le C\left[\|R'(T) - \bar R'(\bar T)\|_{L^\infty(J)} + \|\Theta - \bar\Theta\|_{L^\infty(J)}\right].\end{aligned} \tag{10.97}$$

Le lemme sera donc démontré si l'on établit que

$$\|R'(T) - \bar R'(\bar T)\|_{L^\infty(J)} \le C|J|^{-1}\left[\|z_n - z_{n-1}\|_{L^\infty(J)} + 2^{-(r+2)n}\right] \tag{10.98}$$

et

$$\|\Theta - \bar\Theta\|_{L^\infty(J)} \le C|J|^{-1}\left[\|z_n - z_{n-1}\|_{L^\infty(J)} + 2^{-(r+1)n}\right]. \tag{10.99}$$

La preuve de ces deux dernières inégalités est assez technique, et passe par les estimations suivantes :

$$\begin{aligned}&\text{(i)} \quad \|Q(z_n) - \bar Q(z_{n-1})\|_{L^\infty(J)} \le \|z_n - z_{n-1}\|_{L^\infty(J)} + 2^{-(r+2)n}\\ &\text{(ii)} \quad \|Q'(z_n) - \bar Q'(z_{n-1})\|_{L^\infty(J)} \le \|z_n - z_{n-1}\|_{L^\infty(J)} + 2^{-(r+1)n}\\ &\text{(iii)} \quad \|T - \bar T\|_{L^\infty(J)} \le \|z_n - z_{n-1}\|_{L^\infty(J)} + 2^{-(r+2)n}.\end{aligned} \tag{10.100}$$

Preuve de (10.100) (i). On se souvient peut-être que le polynôme Q coïncide avec la fonction $\mathcal{Q}_n := \mathsf{s}^{-1}tg_n'(\mathsf{c}\cdot)$ sur l'intervalle $z_n(J)$ (voir (10.70)). D'après (10.45), $\mathcal{Q}_n$ est toujours lipschitzienne et plus précisément, on a $\|\mathcal{Q}_n'\|_{L^\infty} \le \frac{\mathsf{c}t}{\mathsf{s}}B = \frac{2B}{A}$. On en déduit que

$$\|Q(z_n) - \mathcal{Q}_n(z_{n-1})\|_{L^\infty(J)} = \|\mathcal{Q}_n(z_n) - \mathcal{Q}_n(z_{n-1})\|_{L^\infty(J)} \le \frac{2B}{A}\|z_n - z_{n-1}\|_{L^\infty(J)}. \tag{10.101}$$

Pour la même raison, $\bar{Q}$ coïncide avec $\mathcal{Q}_{n-1}$ sur $z_{n-1}(J)$. D'après la vitesse de convergence (10.43) des flux approchés, on a donc

$$\|\mathcal{Q}_n(z_{n-1}) - \bar{Q}(z_{n-1})\|_{L^\infty(J)} = \|\mathcal{Q}_n(z_{n-1}) - \mathcal{Q}_{n-1}(z_{n-1})\|_{L^\infty(J)} \leq C2^{-(r+2)n}. \quad (10.102)$$

La réunion de ces deux inégalités fait l'affaire, puisque

$$\begin{aligned}\|Q(z_n) - \bar{Q}(z_{n-1})\|_{L^\infty(J)} &\leq \|Q(z_n) - \mathcal{Q}_n(z_{n-1})\|_{L^\infty(J)} + \|\mathcal{Q}_n(z_{n-1}) - \bar{Q}(z_{n-1})\|_{L^\infty(J)} \\ &\leq C\left[\|z_n - z_{n-1}\|_{L^\infty(J)} + 2^{-(r+2)n}\right]. \quad (10.103)\end{aligned}$$

Preuve de (10.100) (ii). Le même argument s'applique ici, toujours d'après (10.43).

Preuve de (10.100) (iii). En utilisant (10.100) (i), on obtient

$$\begin{aligned}\|T - \bar{T}\|_{L^\infty(J)} &\leq \|z_n - z_{n-1}\|_{L^\infty(J)} + \|Q(z_n) - \bar{Q}(z_{n-1})\|_{L^\infty(J)} \\ &\leq C\left[\|z_n - z_{n-1}\|_{L^\infty(J)} + 2^{-(r+2)n}\right]. \quad (10.104)\end{aligned}$$

Preuve de (10.98). Commençons par considérer le cas où $T(J) \cap \bar{T}(J) = \emptyset$, et supposons sans perte de généralité que

$$a := \sup T(J) < \inf \bar{T}(J). \quad (10.105)$$

On pose alors $R_e(x) := R(x)\chi_{x\leq a}(x) + (x-a)R'(a)\chi_{x>a}(x)$ qui coïncide avec R sur $T(J)$ et est affine de pente $R'(a)$ sur $\bar{T}(J)$. On en déduit que

$$\|R'(T) - R_e'(\bar{T})\|_{L^\infty(J)} \leq \|R''\|_{L^\infty(T(J))}|T(J)|, \quad (10.106)$$

et en utilisant successivement (10.92) et (10.53), que

$$\|R'(T) - R_e'(\bar{T})\|_{L^\infty(J)} \leq C\|R'\|_{L^\infty(T(J))} \leq C. \quad (10.107)$$

On utilise alors le fait que $T(J)$ et $T(J)$ sont disjoints pour voir que l'équivalence (10.82) entraîne

$$|J| \leq C\min\{|T(J)|, |\bar{T}(J)|\} \leq \|T - \bar{T}\|_{L^\infty(J)} \leq C\|T - \bar{T}\|_{L^\infty(J)}, \quad (10.108)$$

d'où l'on déduit

$$\|R'(T) - R_e'(\bar{T})\|_{L^\infty(J)} \leq C|J|^{-1}\|T - \bar{T}\|_{L^\infty(J)}. \quad (10.109)$$

D'un autre côté, $R_e - \bar{R}$ est un polynôme sur $\bar{T}(J)$. On peut donc utiliser à nouveau (10.92) et (10.82) pour voir que

$$\begin{aligned}\|R_e' - \bar{R}'\|_{L^\infty(\bar{T}(J))} &\leq |J|^{-1}\|R_e - \bar{R}\|_{L^\infty(\bar{T}(J))} \\ &\leq |J|^{-1}\left[\|R_e(\bar{T}) - R(T)\|_{L^\infty(J)} + \|R(T) - \bar{R}(\bar{T})\|_{L^\infty(J)}\right] \quad (10.110) \\ &\leq |J|^{-1}\left[\|T - \bar{T}\|_{L^\infty(J)} + \|z_n - z_{n-1}\|_{L^\infty(J)}\right],\end{aligned}$$

cette dernière inégalité étant une conséquence de (10.50) et de (10.53). En réunissant (10.109) et (10.110) et (10.100) (iii), on trouve alors

$$\begin{aligned}\|R'(T)-\bar R'(\bar T)\|_{L^\infty(J)} &\le \|R'(T)-R'_e(\bar T)\|_{L^\infty(J)}+\|R'_e-\bar R'\|_{L^\infty(\bar T(J))}\\ &\le |J|^{-1}\left[\|T-\bar T\|_{L^\infty(J)}+\|z_n-z_{n-1}\|_{L^\infty(J)}\right]\\ &\le C\left[\|z_n-z_{n-1}\|_{L^\infty(J)}+2^{-(r+2)n}\right],\end{aligned} \tag{10.111}$$

de sorte que le (10.98) est démontré, au moins dans le cas où les intervalles $T(J)$ et $\bar T(J)$ sont disjoints. Dans l'autre cas, posons $K := T(J)\cup\bar T(J)$. D'après (10.82), c'est un intervalle de longueur $\mathcal{O}(|J|)$. On peut alors appliquer successivement les inégalités (10.91), (10.53) et (10.92) pour voir que

$$\|R'\|_{L^\infty(K)} \le C\|R'\|_{L^\infty(T(J))} \le C \tag{10.112}$$

et

$$\|R''\|_{L^\infty(K)} \le C|J|^{-1}\|R'\|_{L^\infty(K)} \le C|J|^{-1}. \tag{10.113}$$

On a donc

$$\|R'(T)-R'(\bar T)\|_{L^\infty(J)} \le C|J|^{-1}\|T-\bar T\|_{L^\infty(J)} \tag{10.114}$$

d'une part, et

$$\begin{aligned}\|R'-\bar R'\|_{L^\infty(\bar T(J))} &\le C|J|^{-1}\|R-\bar R\|_{L^\infty(\bar T(J))}\\ &\le C|J|^{-1}\left[\|R(\bar T)-R(T)\|_{L^\infty(J)}+\|R(T)-\bar R(\bar T)\|_{L^\infty(J)}\right]\\ &\le C|J|^{-1}\left[\|R'\|_{L^\infty(K)}\|T-\bar T\|_{L^\infty(J)}+\|z_n-z_{n-1}\|_{L^\infty(J)}\right]\\ &\le C|J|^{-1}\left[\|T-\bar T\|_{L^\infty(J)}+\|z_n-z_{n-1}\|_{L^\infty(J)}\right]\end{aligned} \tag{10.115}$$

d'autre part, avec des arguments désormais classiques. On en déduit

$$\begin{aligned}\|R'(T)-\bar R'(\bar T)\|_{L^\infty(J)} &\le \|R'-\bar R'\|_{L^\infty(\bar T(J))}+\|R'(T)-R'(\bar T)\|_{L^\infty(J)}\\ &\le C|J|^{-1}\left[\|T-\bar T\|_{L^\infty(J)}+\|z_n-z_{n-1}\|_{L^\infty(J)}\right],\end{aligned} \tag{10.116}$$

ce qui achève de prouver (10.98), notamment grâce à (10.100) (iii).

Preuve de (10.99). L'intervalle J étant clairement de longueur bornée par une constante (indépendante de n et du morceau choisi), (10.53) entraîne en particulier que

$$\|R'\|_{L^\infty(T(J))} \le C|J|^{-1}. \tag{10.117}$$

On calcule alors d'après la définition (10.80) de Θ :

$$\begin{aligned}\|\Theta-\bar\Theta\|_{L^\infty(J)} &\le \|R'(T)(1-Q'(z_n))-\bar R'(\bar T)(1-\bar Q(z_{n-1}))\|_{L^\infty(J)}\\ &\le \|R'(T)-\bar R'(\bar T)\|_{L^\infty(J)}+\|R'(T)\|_{L^\infty(J)}\|Q'(z_n)-\bar Q'(z_{n-1})\|_{L^\infty(J)}\\ &\qquad+\|\bar Q'(z_{n-1})\|_{L^\infty(J)}\|R'(T)-\bar R'(\bar T)\|_{L^\infty(J)}\\ &\le C|J|^{-1}\left[\|z_n-z_{n-1}\|_{L^\infty(J)}+2^{-(r+1)n}\right],\end{aligned} \tag{10.118}$$

cette dernière inégalité provenant de (10.98) et de (10.100) (ii). Ceci prouve donc (10.99), et complète la preuve du lemme 10.5. □

10.3.2 Preuve de l'estimation inverse 10.4

Munis du lemme 10.5, nous pouvons à présent suivre le calcul [32] de DeVore et Lucier pour estimer la valeur de $\|z'\|_{B_q^{\alpha-1,q}}$ (par souci de simplicité, on écrira z' à la place de $z'\chi_J$). On aura besoin du lemme 4.3 de [32] :

Lemme 10.6 *Soit v une fonction de classe $\mathcal{C}^2$ sur un intervalle ouvert I, et supposons que v, v' et v'' sont de signe constant sur I. Si p et q sont tels que $0 < p \leq 1$ et $\frac{1}{p} - \frac{1}{q} > 1$, alors v' appartient à $L^p(I)$ dès que v est dans $L^q(I)$, et il existe une constante C pour laquelle*

$$\|v'\|_{L^p(I)} \leq C|I|^{\frac{1}{p}-\frac{1}{q}-1}\|v\|_{L^q(I)}. \tag{10.119}$$

D'après la définition (1.66) de la norme de Besov, il nous faut estimer la valeur de $\omega_r(z', s)_q := \sup_{0<h\leq s} \|\Delta_h^r z'\|_{L^q(\mathbb{R})}$ pour $s > 0$. Pour une valeur donnée de h, on introduit alors les ensembles suivants :

$$\Gamma := \{x \in \mathbb{R} : [x, x + rh] \subset J\}, \quad \Gamma' := \{x \in \mathbb{R} \setminus \Gamma : [x, x + rh] \cap J \neq \emptyset\}, \tag{10.120}$$

et

$$\Gamma'' := \mathbb{R} \setminus (\Gamma \cup \Gamma') = \{x \in \mathbb{R} : [x, x + rh] \cap J = \emptyset\}. \tag{10.121}$$

Si $x \in \Gamma''$, alors clairement $\Delta_h^r z'(x) = 0$ et

$$\|\Delta_h^r z'\|_{L^q(\Gamma'')} = 0. \tag{10.122}$$

Si $x \in \Gamma'$, on peut utiliser le fait que $|\Delta_h^r z'(x)| \leq 2^r(|z'(x)| + \cdots + |z'(x + rh)|)$ pour écrire

$$\int_{\Gamma'} |\Delta_h^r z'(x)|^q \, \mathrm{d}x \leq |\Gamma'|\|\Delta_h^r z'\|^q_{L^\infty(J)} \leq |\Gamma'|\|z'\|^q_{L^\infty(J)}. \tag{10.123}$$

Il n'est pas très difficile de voir que $|\Gamma'| \leq C \min\{h, |J|\}$. On applique alors le lemme 10.5, qui nous donne

$$\int_{\Gamma'} |\Delta_h^r z'(x)|^q \, \mathrm{d}x \leq C \min\{h, |J|\}|J|^{-1} \left(\|z\|_{L^\infty(J)} + 2^{-(r+1)n}\right)^q. \tag{10.124}$$

On considère alors le cas où $x \in \Gamma$. Comme cet ensemble est vide lorsque h est grand par rapport à $|J|$, ou mieux lorsque $h > |J|/r$, on peut supposer que $0 < h \leq |J|/r$. On utilise alors le fait que la différence finie vérifie $\Delta_h^r z'(x) = h^r z^{(r+1)}(\xi)$ pour un $\xi \in [x, x + rh]$. Comme $z^{(r+1)}$ est monotone sur J, et qu'on peut supposer sans perte de généralité qu'elle est décroissante, on en déduit que

$$\Delta_h^r z'(x) = h^r \min\{z^{(r+1)}(x), z^{(r+1)}(x + rh)\} = h^r z^{(r+1)}(x), \quad \text{pour tout } x \in \Gamma. \tag{10.125}$$

On fixe alors $q_0 := q = 1/\alpha$, $\varepsilon := \frac{1}{2}(\frac{\alpha}{r} - 1) > 0$, et on définit q_1, q_2,... q_r par la relation de récurrence $\frac{1}{q_j} := \frac{1}{q_{j-1}} - (1 + \varepsilon)$ pour $j = 1, \ldots, r$. Comme $\frac{1}{q_j} = \alpha - j(1 + \varepsilon)$, le dernier terme $\frac{1}{q_r}$ vaut $\alpha - r(1 + \varepsilon) > 0$, on a donc $0 < q_0 < q_1 < \cdots < q_{k-1} < 1$ et $q_r > 1$. On peut alors appliquer le lemme 10.6 r fois, pour obtenir

$$\begin{aligned} \|z^{(r+1)}\|_{L^q(J)} &\leq C|J|^\varepsilon \|z^{(r)}\|_{L^{q_1}(J)} \leq \cdots \leq C|J|^{r\varepsilon}\|z'\|_{L^{q_r}(J)} \\ &\leq C|J|^{r\varepsilon+\frac{1}{q_r}}\|z'\|_{L^\infty(J)} \leq C|J|^{\frac{1}{q}-\alpha}\|z'\|_{L^\infty(J)}. \end{aligned} \tag{10.126}$$

D'après (10.125), (10.125) et en utilisant à nouveau notre lemme "intermédiaire" 10.5, on calcule

$$\int_\Gamma |\Delta_h^r z'(x)|^q \,\mathrm{d}x \le C h^{rq} |J|^{1-q-rq} \left(\|z\|_{L^\infty(J)} + 2^{-(r+1)n} \right)^q . \tag{10.127}$$

On réunit alors (10.122), (10.124) et (10.127) :

$$\begin{aligned} \omega_r(z',s)_q^q &= \sup_{0<h\le s} \int_{\mathbb{R}} |\Delta_h^r z'(x)|^q \,\mathrm{d}x \\ &\le C \left[\min\{s,|J|\} + s^{rq} |J|^{1-rq} \chi_{[0,|J|/r}(s) \right] |J|^{-q} \left(\|z\|_{L^\infty(J)} + 2^{-(r+1)n} \right)^q , \end{aligned} \tag{10.128}$$

avec $\chi(s) = \chi_{[0,|J|/r]}(s)$. On en déduit

$$\begin{aligned} \|z'\|_{B_q^{\alpha-1,q}}^q &= \int_0^\infty s^{-(\alpha-1)q-1} \omega_r(z',s)_q^q \,\mathrm{d}s \\ &\le \left[|J|^{-q} \int_0^{|J|} s^{q-1} \,\mathrm{d}s + |J|^{1-q} \int_{|J|}^\infty s^{q-2} \,\mathrm{d}s \right. \\ &\quad \left. + |J|^{1-q-rq} \int_0^{|J|/r} s^{q+rq-2} \,\mathrm{d}s \right] \left(\|z\|_{L^\infty(J)} + 2^{-(r+1)n} \right)^q \\ &\le \left(\|z\|_{L^\infty(J)} + 2^{-(r+1)n} \right)^q , \end{aligned} \tag{10.129}$$

en utilisant le fait que $0 < q < 1$ et $rq + q - 2 = (r+1)/\alpha - 2 > -1$. Et ceci démontre le lemme 10.4.

Chapitre 11

Analyse d'un schéma numérique en distance de Hausdorff

On étudie à présent la convergence du schéma de volumes finis très classique "upwind" pour le transport linéaire, qui consiste à faire évoluer des valeurs moyennes constantes par morceaux en approchant le flux entre deux mailles par sa valeur sur la maille de gauche (les solutions étant transportées vers la droite). En utilisant un résultat élémentaire selon lequel les solutions numériques calculées sur un maillage uniforme de pas h convergent dans L^∞ vers une solution exacte $W^{2,\infty}$ avec une précision de l'ordre de h, on peut montrer que dans le cas où la solution exacte n'est que *semi-lipschitzienne*, autrement dit lorsque sa dérivée est, au choix, majorée ou minorée par une constante finie, l'ordre de convergence en distance de Hausdorff est au moins de $h^{1/3}$.

11.1 Le schéma de volumes finis "upwind" pour le transport linéaire

On considère ici le problème de transport linéaire

$$\partial_t u + a\partial_x u = 0, \quad u(0,\cdot) = u_0, \quad x \in \mathbb{R}, \quad t > 0 \tag{11.1}$$

associée à une vitesse a positive, dont la solution est bien sûr donnée par $u(t,x) = u_0(x - at)$. Le développement de méthodes de calcul itératives destinées à approcher une telle fonction u pouvant paraître saugrenu, précisons que (11.1) doit être vu comme un cas particulier (très simple) du problème plus général (8.2), et que parallèlement, le schéma upwind (11.4), (11.5) ci-dessous correspond à la forme linéaire des schémas de Godunov. L'analyse du schéma upwind présente donc un certain intérêt, d'abord parce qu'elle permet de valider la pertinence d'une conjecture donnée dans un cas particulier simple, ensuite parce qu'elle peut aider à la compréhension des phénomènes à l'œuvre dans l'approximation du problème non-linéaire.

11.1.1 La méthode de Godunov

Le principe de cette méthode (proposée par Godunov [37] en 1959) consiste à approcher la solution par une fonction constante par mailles dont les valeurs sont calculées d'un pas de temps à l'autre en résolvant des problèmes de Riemann à l'interface

entre deux mailles, et en projetant la solution obtenue par ses valeurs moyennes sur les mailles. Rappelons que le problème de Riemann consiste à chercher la solution de notre loi de conservation (8.2) pour une donnée initiale constante de part et d'autre de l'origine. Ce problème étant relativement simple (y compris dans le cas des *systèmes* de lois de conservation), il est souvent possible de calculer ses solutions de façon exacte. La méthode de Godunov s'écrit alors comme un schéma de transport-projection naturellement conservatif, dans la mesure où les projections ne changent pas la masse sur les mailles, et que le transport y est exact donc également conservatif. La solution est initialisée à

$$U^0 := Pu_0,$$

où $P = P_h$ désigne l'interpolation (constante par morceaux) des valeurs moyennes sur les intervalles $I_j := [(j-1)h, jh[$:

$$P\colon v \to \sum_{j\in\mathbb{Z}} c_j \chi_{I_j}(x) \quad \text{avec} \quad c_j := \frac{1}{h}\int_{I_j} v(x)\,\mathrm{d}x.$$

Pour aller d'un pas de temps à l'autre, on calcule alors

$$U^n := PTU^{n-1}, \tag{11.2}$$

où $T = T_{\Delta t}$ est l'opérateur *exact* d'évolution sur un pas de temps

$$T\colon v(x) \to v(x - a\Delta t\,)$$

dans le cas linéaire considéré dans cette section. On peut alors observer que sous la condition CFL

$$\lambda := \frac{a\Delta t}{h} \leq 1, \tag{11.3}$$

les valeurs prises par TU^{n-1} sur la maille I_j correspondent aux valeurs prises par U^{n-1} sur les mailles I_j et I_{j-1}. Plus précisément, les valeurs moyennes

$$U_j^n := \frac{1}{h}\int_{I_j} U^n(x)\,\mathrm{d}x \tag{11.4}$$

vérifient l'égalité suivante

$$U_j^n = (1-\lambda)U_j^{n-1} + \lambda U_{j-1}^{n-1}. \tag{11.5}$$

Dans la mesure où $U^n(x) = \sum_{j\in\mathbb{Z}} U_j^n \chi_{I_j}(x)$, le schéma upwind peut s'écrire sous la forme suivante

$$U^n = \Big[(1-\lambda)I + \lambda D_h\Big]U^{n-1} \tag{11.6}$$

où $D_h\colon v(x) \to v(x-h)$ désigne le décalage horizontal de pas h. Signalons enfin que ce schéma rentre aussi dans la catégorie des schémas de volumes finis, car les valeurs moyennes des solutions numériques évoluent suivant une loi de conservation discrète

$$U_j^n = U_j^{n-1} - F_j^{n-1} + F_{j-1}^{n-1} \tag{11.7}$$

qui reproduit la forme intégrale

$$\frac{1}{h}\int_{I_j} u(n\Delta t\,,\cdot) = \frac{1}{h}\int_{I_j} u((n-1)\Delta t\,,\cdot) - \frac{1}{\Delta t}\int_{(n-1)\Delta t}^{n\Delta t} \Big[f(u(\cdot, jh)) - f(u(\cdot,(j-1)h))\Big]$$

de l'équation (8.2). On peut donc interpréter le schéma (11.5)-(11.4) comme un schéma de volumes finis (11.7) où les flux numériques sont donnés par $F_j^{n-1} := \lambda U_j^{n-1}$.

11.1.2 Estimation d'erreur en distance L^∞

On a donc à notre disposition plusieurs expressions équivalentes du schéma upwind, mais on peut observer que toutes ne sont pas également appropriées à l'analyse d'erreur. A partir de la forme (11.2), par exemple, on décomposera l'erreur numérique comme

$$\begin{aligned}\|U^n - u(n\Delta t\,)\|_{L^\infty} &\le \|(P-I)TU^{n-1}\|_{L^\infty} + \|T\big[U^{n-1} - u((n-1)\Delta t\,)\big]\|_{L^\infty}\\ &\le \|(P-I)TU^{n-1}\|_{L^\infty} + \|U^{n-1} - u((n-1)\Delta t\,)\|_{L^\infty}\\ &\le \sum\nolimits_{n'=0}^{n-1} \|(P-I)TU^{n'}\|_{L^\infty} + \|(P-I)u_0\|_{L^\infty}.\end{aligned}$$

Les projections par valeurs moyennes étant au mieux d'ordre 1 (dans n'importe quel espace L^p), *i.e.* $\|(I-P)TU^{n-1}\|_{L^\infty} \le Ch$, on obtient de cette façon

$$\|U^n - u(n\Delta t\,)\|_{L^\infty} \le C\frac{h}{\Delta t},$$

estimation insuffisante en raison de la condition (11.3) selon laquelle $\frac{h}{\Delta t}$ est supérieur à la constante a. La forme aux différences finies (11.6), en revanche, se prête mieux à cette analyse. En désignant par S_λ l'opérateur $(1-\lambda)I + \lambda D_h$ et par

$$K_n := u((n+1)\Delta t\,) - S_\lambda u(n\Delta t\,) \tag{11.8}$$

l'*erreur de troncature* obtenue en substituant la solution exacte à la solution approchée dans la relation (11.6), on obtient

$$\begin{aligned}\|U^n - u(n\Delta t\,)\|_{L^\infty} &\le \|K_{n-1}\|_{L^\infty} + \|S_\lambda\big[U^{n-1} - u((n-1)\Delta t\,)\big]\|_{L^\infty}\\ &\le \|K_{n-1}\|_{L^\infty} + \|U^{n-1} - u((n-1)\Delta t\,)\|_{L^\infty}\\ &\le \sum\nolimits_{n'=0}^{n-1} \|K_{n'}\|_{L^\infty} + \|(P-I)u_0\|_{L^\infty}.\end{aligned}$$

où l'on a notamment utilisé le fait que l'opérateur S_λ, comme combinaison convexe d'opérateurs L^∞-contractants, était lui-même L^∞-contractant.

Lorsque la solution est régulière, on peut alors estimer les erreurs de troncature K_n en calculant

$$\begin{aligned}\lambda\big[u(n\Delta t\,,x) - u(n\Delta t\,,x-h)\big] &= -\lambda\int_0^h \partial_x u(n\Delta t\,,x-y)\,\mathrm{d}y\\ &= -\int_0^{\lambda h} \partial_x u(n\Delta t\,,x-y)\,\mathrm{d}(\lambda y),\end{aligned}$$

et d'après l'équation de transport (11.1),

$$\begin{aligned}u((n+1)\Delta t\,,x) - u(n\Delta t\,,x) &= \int_0^{\Delta t} \partial_t u(n\Delta t\,+s,x)\,\mathrm{d}s\\ &= -a\int_0^{\Delta t} \partial_x u(n\Delta t\,+s,x)\,\mathrm{d}s = -\int_0^{a\Delta t} \partial_x u(n\Delta t\,,x-as)\,\mathrm{d}(as).\end{aligned}$$

Comme $a\Delta t = \lambda h$ d'après (11.3), on en déduit en posant $z = as = \lambda y$ que

$$|K_n(x)| = \left| \int_0^{\lambda h} \left[\partial_x u(n\Delta t\,, x - z) - \partial_x u(n\Delta t\,, x - \frac{z}{\lambda}) \right] \mathrm{d}z \right| \\ \leq \int_0^{\lambda h} h \|\partial_{xx}^2 u(n\Delta t\,)\|_{L^\infty} \leq h^2 \|\partial_{xx}^2 u(n\Delta t\,)\|_{L^\infty}.$$

D'après l'inégalité (11.1.2), on en déduit

$$\|U^n - u(n\Delta t\,)\|_{L^\infty} \leq Ch\|u_0\|_{W^{2,\infty}} \tag{11.9}$$

en utilisant à nouveau le fait que $\Delta t^{-1} \leq Ch^{-1}$.

11.1.3 Régularisation d'une donnée initiale discontinue

Si u_0 est discontinue, on peut la rendre indéfiniment régulière en la "moyennant localement", autrement dit en la remplaçant par

$$u_{0,\varepsilon} := u_0 * \varphi_\varepsilon \tag{11.10}$$

où $\varphi_\varepsilon(x) := \frac{1}{\varepsilon}\varphi(\frac{x}{\varepsilon})$ et φ est une fonction positive, de classe $\mathcal{C}^\infty$ à support inclus dans l'intervalle $[-1, 1]$ et de masse $\int \varphi = 1$. On en déduit que φ_ε vérifie également ces propriétés, à l'exception de son support qui est inclus dans l'intervalle $[-\varepsilon, \varepsilon]$, comme on peut l'observer sur la figure 11.1.

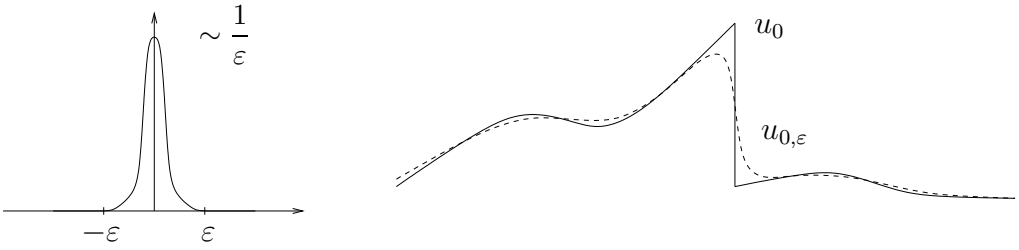


FIG. 11.1 – allure du noyau de convolution φ_ε et de la fonction régularisée $u_{0,\varepsilon}$.

On utilisera le résultat suivant.

Lemme 11.1 *Si v est une fonction semi-lipschitzienne,* i.e. *si*

$$\pm v' \leq L,$$

alors sa régularisation $v_\varepsilon := v * \varphi_\varepsilon$ vérifie

$$d_H(v, v_\varepsilon) \leq (1 + 2L)\varepsilon.$$

Preuve. En se souvenant que la distance de Hausdorff entre les fonctions v et v_ε est caractérisée par les inégalités (9.11), on déduit facilement de la positivité de φ_ε que

$$v_\varepsilon(x) = \int_{-\varepsilon}^{\varepsilon} \varphi_\varepsilon(y) v(x - y)\,\mathrm{d}y \leq \sup_{|y|\leq\varepsilon} v(x - y) \int_{-\varepsilon}^{\varepsilon} \varphi_\varepsilon(y)\,\mathrm{d}y = \sup_{|y|\leq\varepsilon} v(x - y),$$

et de même, que

$$v_\varepsilon(x) = \int_{-\varepsilon}^{\varepsilon} \varphi_\varepsilon(y) v(x-y)\,\mathrm{d}y \geq \inf_{|y|\leq\varepsilon} v(x-y) \int_{-\varepsilon}^{\varepsilon} \varphi_\varepsilon(y)\,\mathrm{d}y = \inf_{|y|\leq\varepsilon} v(x-y).$$

On en déduit que v_ε est "proche" de v, au sens où l'éloignement dissymétrique défini dans la remarque 9.2 vérifie

$$\epsilon(v_\varepsilon, v) \leq \varepsilon.$$

La proposition 9.4 nous permet alors d'en déduire que

$$\epsilon(v, v_\varepsilon) \leq (1+2L)\varepsilon,$$

ce qui prouve le lemme. □

11.1.4 Estimation d'erreur en distance de Hausdorff

On aura encore besoin du lemme suivant, qu'on appliquera en particulier à S_λ.

Lemme 11.2 *A désigne ici un opérateur linéaire monotone qui commute avec D_h, autrement dit qui vérifie*

$$u \leq v \implies Au \leq Av$$

et qui est laissé invariant par des translations d'un multiple entier de h. Pour tout couple de fonctions uni-dimensionnelles u, v admissibles au sens de la définition 9.1 et telles que v, par exemple, est semi-lipschitzienne

$$\pm v' \leq L,$$

on a

$$d_H(Au, Av) \leq (1+2L) d_H(u,v) + Lh.$$

Preuve. Notons $\varepsilon := d_H(u,v)$. Si $v' \leq L$, on peut calculer (de façon désormais classique, voir notamment la preuve du lemme 9.4) que

$$\mathcal{S}_\varepsilon^+ v(x) - \varepsilon = \sup_{]x-\varepsilon, x+\varepsilon[} v \leq \underline{v}(x-\varepsilon) + 2L\varepsilon \leq \underline{v}(x-\varepsilon') + L(\varepsilon+\varepsilon')$$

pour tout $\varepsilon' \geq \varepsilon$. En rappelant que $\lceil \varepsilon/h \rceil$ désigne le plus petit entier majorant ε/h, on a

$$\varepsilon \lceil \varepsilon/h \rceil h \leq \varepsilon + h,$$

de sorte que

$$\mathcal{S}_\varepsilon^+ v(x) \leq \underline{v}(x - \lceil \varepsilon/h \rceil h) + (1+2L)\varepsilon + Lh.$$

On déduit alors de la caractérisation (9.11) de la distance de Hausdorff que

$$u \leq \underline{v}(x - \lceil \varepsilon/h \rceil h) + (1+2L)\varepsilon + Lh.$$

Comme A est linéaire et commute avec D_h, on a

$$A\big(v(\cdot - \lceil \varepsilon/h \rceil h) + (1+2L)\varepsilon + Lh\big) = Av(\cdot - \lceil \varepsilon/h \rceil h) + (1+2L)\varepsilon + Lh,$$

et par monotonie, on trouve

$$Au \leq Av(\cdot - \lceil \varepsilon/h \rceil h) + (1+2L)\varepsilon + Lh.$$

Comme l'argument symmétrique nous donnerait

$$Av(\cdot - \lceil \varepsilon/h \rceil h) + (1+2L)\varepsilon + Lh \leq Au,$$

on en déduit que $d_H(Au, Av) \leq (1+2L)\varepsilon + Lh$. □

On est alors en mesure d'établir le théorème suivant.

Théorème 11.1 *Pour une donnée initiale u_0 semi-lipschitzienne*

$$\pm u_0' \leq L,$$

la solution numérique $U^N := S_\lambda^N u_0$ calculée par le schéma upwind (11.6) *approche la solution exacte $u = u(N\Delta t)$ de* (11.1) *avec une précision de l'ordre de $h^{1/3}$ en distance de Hausdorff,* i.e.

$$d_H(U^N, u) \leq Ch^{1/3}$$

avec une constante de l'ordre de $\|u_0\|_{L^\infty} + (1+L)^2$.

Preuve. On commence par régulariser la donnée initiale en considérant $u_{0,\varepsilon}$ défini par (11.10). L'équation (11.1) étant linéaire, la solution u_ε issue de $u_{0,\varepsilon}$ vérifie

$$u_\varepsilon = u * \varphi_\varepsilon,$$

de sorte que le lemme 11.1 s'applique et l'on a

$$d_H(u, u_\varepsilon) \leq (1+2L)\varepsilon. \tag{11.11}$$

On décompose alors

$$d_H(U^N, u) \leq d_H(S_\lambda^N u_0, S_\lambda^N u_{0,\varepsilon}) + d_H(S_\lambda^N u_{0,\varepsilon}, u_\varepsilon) + d_H(u_\varepsilon, u).$$

Le troisième terme est majoré dans l'inégalité (11.11) ci-dessus. En appliquant l'estimation (11.9), on obtient pour le deuxième (quitte à supposer $\varepsilon \leq C$)

$$d_H(S_\lambda^N u_{0,\varepsilon}, u_\varepsilon) \leq Ch\|u_{0,\varepsilon}\|_{W^{2,\infty}} \leq C(\varphi)\frac{h}{\varepsilon^2}\|u_0\|_{L^\infty}.$$

Enfin, on déduit des lemmes 11.1 et 11.2 que le premier terme vérifie

$$d_H(S_\lambda^N u_0, S_\lambda^N u_{0,\varepsilon}) \leq (1+2L)d_H(u_0, u_{0,\varepsilon}) + Lh \leq (1+2L)^2\varepsilon + Lh.$$

On en déduit que l'erreur numérique vérifie

$$d_H(U^N, u) \leq C(\varphi)\frac{h}{\varepsilon^2}\|u_0\|_{L^\infty} + C(1+L)^2\varepsilon + Lh,$$

ce qui démontre théorème en fixant $\varepsilon := h^{1/3}$. □

Perspectives

Au cours de cette thèse, on a proposé un nouveau schéma adaptatif de type semi-lagrangien, basé sur des éléments finis multi-échelles $\mathcal{P}^1$ en deux dimensions d'espace. En tirant notamment parti de la simplicité géométrique des maillages dyadiques sous-jacents, on a pu établir une estimation d'erreur prouvant la convergence en norme L^∞ des solutions numériques vers la solution exacte du système de Vlasov-Poisson sous des hypothèses de donnée initiale $W^{1,\infty} \cap W^{2,1}(\mathbb{R}^2)$.

De nombreuses extensions, toutefois, sont possibles : tout d'abord, on souhaiterait mener à son terme l'analyse de la taille des maillages créés par notre schéma, ce qui nous permettrait d'établir simultanément une vitesse de convergence pour nos approximations et l'optimalité asymptotique de notre stratégie d'adaptation de maillages. Dans cette direction, la principale difficulté consiste sans doute à évaluer l'accroissement au cours du temps de la courbure totale des solutions numériques, difficulté liée pour l'essentiel à la façon dont les interpolations font évoluer la courbure totale des fonctions affines par morceaux. D'autre part, il serait intéressant de proposer des schémas basés sur le même principe et mettant en œuvre

- des éléments finis d'ordre plus élevés, tâche aisée en pratique mais dont l'analyse reste à faire. En particulier, la démarche adoptée ici, qui consiste à prendre en compte la régularité des solutions numériques comme facteur de décision dans le processus d'adaptation des maillages, devra être adaptée à des ordres plus élevés, ce qui implique qu'on devra également étudier la stabilité des différents opérateurs pour ces nouveaux ordres de régularité.
- des dimensions supérieures. Dans cette direction, on a commencé à s'intéresser avec Albert Cohen et Eric Sonnendrücker aux "grilles d'interpolations éparses" (*sparse grids*, en anglais), qui exploitent une régularité d'ordre élevé des solutions pour atteindre des taux de convergence indépendants de la dimension par des techniques d'interpolation multi-linéaire et anisotrope.

En ce qui concerne l'utilisation de la distance de Hausdorff dans l'approximation des lois de conservation scalaires, les extensions naturelles de nos résultats sont multiples. En premier lieu, il conviendrait d'étudier plus avant les propriétés des schémas numériques classiques, et éventuellement, de proposer de nouveaux schémas dont on analyserait les performances dans cette distance. Ceci pourrait nous amener à nous poser la question, sans doute délicate, de la caractérisation des ordres de convergence dans cette distance pour des méthodes d'approximation utilisant des polynômes par morceaux. En ce qui concerne les ordres élevés, on devra considérer tout particulièrement

des techniques d'interpolation essentiellement non oscillantes (ENO, ou ENO intra-mailles, techniques intrinsèquement non-linéaires). Pour ces méthodes d'interpolation, la mise au point de bonnes estimations en distance de Hausdorff serait d'ailleurs un moyen naturel de valider leur objectif premier, qui est l'approximation uniforme de fonctions régulières par morceaux (voir [2] à ce sujet).

Notre travail s'étant d'autre part limité aux lois de conservation scalaires, on devra tôt ou tard étudier le cas des systèmes, dans lequel la description semi-explicite donnée par Lax des trajectoires caractéristiques ne s'applique plus, et où de façon plus générale, les propriétés des solutions sont beaucoup moins bien comprises.

Index

Bibliographie

[1] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[2] F. Arandiga, A. Cohen, R. Donat, and N. Dyn. Interpolation and approximation of piecewise smooth functions. *SIAM J. Numer. Anal.*, 2005.

[3] A. Arsen'ev. Global existence of a weak solution of Vlasov's system of equations. *U.S.S.R. Comp. Math. Math. Phys.*, 15(1) :131–143, 1975.

[4] I. Babuška and W. C. Rheinboldt. Error estimates for adaptive finite element computations. *SIAM J. Numer. Anal.*, 15(4) :736–754, 1978.

[5] C. Bennett and R. Sharpley. *Interpolation of operators*, volume 129 of *Pure and Applied Mathematics*. Academic Press Inc., Boston, MA, 1988.

[6] M. Berger and P. Colella. Local adaptive mesh refinement for shock hydrodynamics. *J. Comput. Phys.*, 82(1) :64–84, 1989.

[7] M. Berger and J. Oliger. Adaptive mesh refinement for hyperbolic partial differential equations. *J. Comput. Phys.*, 53(3) :484–512, 1984.

[8] S. Bertoluzza and Y Maday. Analysis of a wavelet based adaptive scheme for the Burgers equation. Prépublication du laboratoire Jacques-Louis Lions, 2001.

[9] S. Bertoluzza, Y. Maday, and J.-C. Ravel. A dynamically adaptive wavelet method for solving partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, 116(1-4) :293–299, 1994. ICOSAHOM'92 (Montpellier, 1992).

[10] N. Besse. Convergence of a semi-Lagrangian scheme for the one-dimensional Vlasov-Poisson system. *SIAM J. Numer. Anal.*, 42(1) :350–382 (electronic), 2004.

[11] N. Besse, F. Filbet, M. Gutnic, I. Paun, and E. Sonnendrücker. An adaptive numerical method for the Vlasov equation based on a multiresolution analysis. In F. Brezzi, A. Buffa, S. Escorsaro, and A. Murli, editors, *Numerical Mathematics and Advanced Applications ENUMATH 2001*, pages 437–446. Springer, 2001.

[12] Y. Brenier. Averaged multivalued solutions for scalar conservation laws. *SIAM J. Numer. Anal.*, 21(6) :1013–1037, 1984.

[13] H. Brezis. *Analyse fonctionnelle*. Collection Mathématiques Appliquées pour la Maîtrise. [Collection of Applied Mathematics for the Master's Degree]. Masson, Paris, 1983. Théorie et applications. [Theory and applications].

[14] M. Campos Pinto. A total curvature diminishing property for P_1 finite element interpolation. *Mathematisches Forschunginstitut Oberwolfach Report, 34/2004*, 2005.

[15] M. Campos Pinto, A. Cohen, W. Dahmen, and R. Devore. On the stability of nonlinear conservation laws in the Hausdorff metric. *J. Hyperbolic Differ. Equ.*, 2(1) :25–38, 2005.

[16] M. Campos Pinto, A. Cohen, and P. Petrushev. High order geometric smoothness for conservation laws. *J. Hyperbolic Differ. Equ.*, 2(1) :39–59, 2005.

[17] M. Campos Pinto and M. Mehrenberger. Adaptive numerical resolution of the Vlasov equation. *Numerical methods for hyperbolic and kinetic problems, CEMRACS 2003/IRMA Lectures in Mathematics and Theoretical Physics*, 2005.

[18] M. Campos Pinto and M. Mehrenberger. Convergence of an adaptive scheme for the one-dimensional Vlasov-Poisson system. Soumis à *SIAM Journal on Numerical Analysis*, 2005.

[19] C.Z. Cheng and G. Knorr. The integration of the Vlasov equation in configuration space. *J. Comput. Phys.*, 22 :330–351, 1976.

[20] P.G. Ciarlet. Basic error estimates for elliptic problems. In *Handbook of numerical analysis, Vol. II*, Handb. Numer. Anal., II, pages 17–351. North-Holland, Amsterdam, 1991.

[21] A. Cohen. *Numerical analysis of wavelet methods*, volume 32 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 2003.

[22] A. Cohen, I. Daubechies, and J.-C. Fauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure. and Appl. Math.*, 45 :485–560, 1992.

[23] A. Cohen, S.M. Kaber, S. Müller, and M. Postel. Fully adaptive multiresolution finite volume schemes for conservation laws. *Math. Comp.*, 72(241) :183–225 (electronic), 2003.

[24] J. Cooper and A. Klimas. Boundary value problems for the Vlasov-Maxwell equation in one dimension. *J. Math. Anal. Appl.*, 75(2) :306–329, 1980.

[25] W. Dahmen. Wavelet and multiscale methods for operator equations. *Acta Numerica*, 6 :55–228, 1997.

[26] W. Dahmen, B. Gottschlich-Müller, and S. Müller. Multiresolution schemes for conservation laws. *Numer. Math.*, 88(3) :399–443, 2001.

[27] I. Daubechies. *Ten lectures on Wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Math.* SIAM, Philadelphia, 1992.

[28] R. DeVore. A note on adaptive approximation. In *Proceedings of China-U.S. Joint Conference on Approximation Theory (Hangzhou, 1985)*, volume 3, pages 74–78, 1987.

[29] R. DeVore. Nonlinear approximation. *Acta Numerica*, 7 :51–150, 1998.

[30] R. DeVore and G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

[31] R. DeVore and B. Lucier. High order regularity for solutions of the inviscid Burgers equation. In *Nonlinear hyperbolic problems (Bordeaux, 1988)*, volume 1402 of *Lecture Notes in Math.*, pages 147–154. Springer, Berlin, 1989.

[32] R. DeVore and B. Lucier. High order regularity for conservation laws. *Indiana Univ. Math. J.*, 39(2) :413–430, 1990.

[33] R. DeVore and V. Popov. Interpolation spaces and nonlinear approximation. In *Function spaces and applications (Lund, 1986)*, volume 1302 of *Lecture Notes in Math.*, pages 191–205. Springer, Berlin, 1988.

[34] R.T. Glassey. *The Cauchy problem in kinetic theory*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

[35] E. Godlewski and P.-A. Raviart. *Hyperbolic systems of conservation laws*, volume 3/4 of *Mathématiques & Applications (Paris) [Mathematics and Applications]*. Ellipses, Paris, 1991.

[36] E. Godlewski and P.-A. Raviart. *Numerical approximation of hyperbolic systems of conservation laws*, volume 118 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996.

[37] S. K. Godunov. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Mat. Sb. (N.S.)*, 47 (89) :271–306, 1959.

[38] M. Gutnic, M. Haefele, I. Paun, and E. Sonnendrücker. Vlasov simulations on an adaptive phase-space grid. *Comput. Phys. Comm*, 164 :214–219, 2004.

[39] A. Harten. Adaptive multiresolution schemes for shock computations. *J. Comput. Phys.*, 115(2) :319–338, 1994.

[40] A. Harten. Multiresolution algorithms for the numerical solution of hyperbolic conservation laws. *Comm. Pure Appl. Math.*, 48(12) :1305–1342, 1995.

[41] O. Hoenen, M. Mehrenberger, and E. Violard. Parallelization of an adaptive Vlasov solver. *Recent Advances in Parallel Virtual Machine and Message Passing Interface : 11th European PVM/MPI Users Group Meeting Budapest, Hungary, September 19 - 22, 2004, Springer-Verlag Heidelberg*, 2004.

[42] O. Hoenen, M. Mehrenberger, E. Violard, M. Campos Pinto, and E. Sonnendrücker. A parallel adaptive Vlasov solver based on hierarchical finite element interpolation. A paraître dans *Nuclear Instruments and Methods in Physics Research, ICAP 2004*.

[43] S. V. Iordanskii. The Cauchy problem for the kinetic equation of plasma. *Amer. Math. Soc. Transl. Ser. 2*, 35 :351–363, 1964.

[44] S. N. Kružkov. First order quasilinear equations with several independent variables. *Mat. Sb.*, 81 (123) :228–255, 1970 (en russe), version anglaise dans *Mat. USSR Sb.* 10.

[45] P. D. Lax. *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1973. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 11.

[46] Ph. LeFloch. *Hyperbolic systems of conservation laws*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2002. The theory of classical and nonclassical shock waves.

[47] P.-L. Lions and B. Perthame. Propagation of moments and regularity for the 3-dimensional Vlasov-Poisson system. *Invent. Math.*, 105(2) :415–430, 1991.

[48] B. Lucier. A moving mesh numerical method for hyperbolic conservation laws. *Math. Comp.*, 46(173) :59–69, 1986.

[49] Y. Maday, V. Perrier, and J.-C. Ravel. Adaptativité dynamique sur bases d'ondelettes pour l'approximation d'équations aux dérivées partielles. *C. R. Acad. Sci. Paris Sér. I Math.*, 312(5) :405–410, 1991.

[50] S. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.*, 315 :69–87, 1989.

[51] S. Mallat. *A wavelet tour of signal processing*. Academic Press Inc., San Diego, CA, 1998.

[52] M. Mehrenberger. *Inégalités d'observabilité et résolution adaptative de l'équation de Vlasov par éléments finis hiérarchiques*. Prépublication de l'Institut de Recherche Mathématique Avancée, 2004/28. Thèse, Université Louis Pasteur (Strasbourg I), Strasbourg, 2004.

[53] Y. Meyer. *Ondelettes et opérateurs*. Hermann, Paris, 1990.

[54] P. Petrushev. Direct and converse theorems for spline and rational approximation and Besov spaces. In *Function spaces and applications (Lund, 1986)*, volume 1302 of *Lecture Notes in Math.*, pages 363–377. Springer, Berlin, 1988.

[55] P.-A. Raviart. An analysis of particle methods. In *Numerical methods in fluid dynamics (Como, 1983)*, volume 1127 of *Lecture Notes in Math.*, pages 243–324. Springer, Berlin, 1985.

[56] Bl. Sendov. *Hausdorff approximations*, volume 50 of *Mathematics and its Applications (East European Series)*. Kluwer Academic Publishers Group, Dordrecht, 1990. Translated and revised from the Russian.

[57] D. Serre. *Systèmes de lois de conservation. I.* Fondations. Diderot Editeur, Paris, 1996. Hyperbolicité, entropies, ondes de choc.

[58] D. Serre. *Systèmes de lois de conservation. II.* Fondations. Diderot Editeur, Paris, 1996. Structures géométriques, oscillation et problèmes mixtes.

[59] E. Sonnendrücker, J. Roche, P. Bertrand, and A. Ghizzo. The semi-Lagrangian method for the numerical resolution of the Vlasov equation. *J. Comput. Phys.*, 149(2) :201–220, 1999.

[60] E. M. Stein. *Singular integrals and differentiability properties of functions*. Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, N.J., 1970.

[61] E. Tadmor. Local error estimates for discontinuous solutions of nonlinear hyperbolic equations. *SIAM J. Numer. Anal.*, 28(4) :891–906, 1991.

[62] E. Tadmor and T. Tang. Pointwise error estimates for scalar conservation laws with piecewise smooth solutions. *SIAM J. Numer. Anal.*, 36(6) :1739–1758 (electronic), 1999.

[63] E. Tadmor and T. Tang. Pointwise error estimates for relaxation approximations to conservation laws. *SIAM J. Math. Anal.*, 32(4) :870–886 (electronic), 2000.

[64] A. A. Vlasov. A new formulation of the many particle problem (en russe). *Akad. Nauk SSSR. Zhurnal Eksper. Teoret. Fiz.*, 18 :840–856, 1948.