Transport et diffusion

Grégoire ALLAIRE, CMAP, Ecole Polytechnique,
Xavier BLANC, LJLL, Université Paris Diderot,
Bruno DESPRES, LJLL, Université Pierre et Marie Curie,
François GOLSE, CMLS, Ecole Polytechnique

12 novembre 2015

Table des matières

Introduction

Ce livre est issu d'un cours enseigné par les auteurs en troisième année de l'Ecole Polytechnique, ce qui correspond à un niveau de première année de Master. Le sujet en est l'étude mathématique et numérique de modèles d'équations aux dérivées partielles, dits de transport et diffusion. Ces équations modélisent l'évolution d'une densité de particules ou d'individus en interaction avec leur milieu. L'origine de ces modèles est très variée. Ils proviennent classiquement de la physique et servent à décrire des particules neutres comme les neutrons et les photons. Dans le premier cas on parle aussi de neutronique alors que le dernier cas est appelé transfert radiatif. La biologie fait aussi appel aux équations de transport pour modéliser la dynamique de populations structurées. Citons aussi pour mémoire la dynamique des gaz raréfiés, le transport d'électrons dans les semi-conducteurs ou encore la physique des plasmas qui sont des phénomènes modélisés par des équations où l'opérateur de transport est une brique de base essentielle.

Les objectifs du cours sont de familiariser le lecteur avec ces équations de transport et diffusion, leurs méthodes d'analyse mathématique et de résolution numérique, ainsi que le lien qui unit ces deux types de modèles. Afin de limiter la difficulté et de rester dans un format "compact" (typiquement neuf séances de cours et de travaux dirigés), nous nous restreignons à des modèles linéaires, déjà représentatifs dans de nombreuses applications, et nous ne disons rien des phénomènes non linéaires qui peuvent jouer un rôle essentiel dans certains cas importants (particules chargées, collisions entre particules). De même, nous ne prétendons aucunement à l'exhaustivité, ni à la plus extrême modernité en ce qui concerne les méthodes d'analyse théorique ou numérique présentées dans ce texte de niveau introductif : nous laissons le soin à des cours de niveau Master deuxième année, de présenter l'état de l'art dans ce domaine.

Le plan du cours est le suivant. Après un premier chapitre d'introduction aux principaux modèles et à leur origine en physique des réacteurs nucléaires, transfert radiatif ou dynamique des populations structurées, le Chapitre 2 est consacré à l'équation du transport libre, c'est-à-dire en l'absence de collisions des particules avec le milieu ambiant. Le Chapitre 3 traite ensuite l'équation de Boltzmann linéaire, tandis que le Chapitre 4 étudie la limite de diffusion des modèles de transport lorsque le libre parcours moyen des particules devient petit. Le Chapitre 5 porte sur quelques méthodes numériques, de type différences finies, pour la résolution des équations de transport et diffusion. Le Chapitre 6 décrit la théorie du calcul critique qui s'interprète comme la résolution d'un problème aux valeurs propres pour déterminer un état stationnaire en temps. Finalement le Chapitre 7 traite de questions d'homogénéisation et permet encore une fois de faire le lien entre transport, au niveau microscopique, et diffusion au niveau macroscopique. Des exercices avec indications terminent chaque chapitre.

Le niveau de ce cours étant introductif, il n'exige aucun prérequis particulier. Il est auto-contenu dans la mesure du possible. Il nous paraît néanmoins utile que le lecteur ait quelques notions de base, soit en analyse numérique (voir par exemple [2]), soit en analyse des équations aux dérivées partielles (voir par exemple [23]), afin qu'il ne soit pas débordé par trop de notions nouvelles. La bibliographie contient de nombreux ouvrages plus avancés sur le transport et la diffusion (notamment [12], le chapitre XXI de [20], [38], [42], [43], [51]) auxquels nous renvoyons le lecteur désireux d'en savoir plus. Des informations sur ce cours sont disponibles sur les sites web
`http://www.cmap.polytechnique.fr/~allaire/cours_map567.html`
`http://www.math.polytechnique.fr/~golse/mat567.html`
où le lecteur pourra aussi télécharger des transparents et les problèmes d'examen. Les auteurs remercient le Commissariat à l'Energie Atomique où, à divers titres et en divers lieux, ils ont travaillé et beaucoup appris sur le sujet de ce cours. Ils remercient aussi les cinq promotions d'élèves polytechniciens qui ont suivi ce cours et permis de l'améliorer au fur et à mesure par leurs remarques pertinentes.

Bien que ce livre ait été précédé de plusieurs versions d'un polycopié éponyme, il contient probablement encore d'inévitables erreurs, fautes de frappe ou imperfections : les auteur remercient à l'avance tous ceux qui voudront bien les leur signaler, par exemple par courrier électronique aux adresses
`gregoire.allaire@polytechnique.fr, blanc@ann.jussieu.fr,`
`despres@ann.jussieu.fr, francois.golse@polytechnique.fr`

G. Allaire, X. Blanc, B. Després, F. Golse
Paris, le 6 Janvier 2014

Chapitre 1

Présentation de quelques modèles

Les équations de transport et de diffusion sont des modèles mathématiques intervenant pour décrire, entre autres,

•le transfert d'énergie dans un milieu matériel sous différentes formes (thermique, rayonnement ...)

•la dynamique de particules en interaction avec la matière (neutrons dans un matériau fissile, photons dans une atmosphère planétaire ou stellaire, électrons et trous dans un semi-conducteur ...)

•l'évolution de certaines populations d'organismes vivants (dynamique des populations structurées ...)

Tous ces phénomènes proviennent de domaines extrêmement différents de la physique — ou de la biologie. Pourtant, les modèles mathématiques utilisés pour les décrire présentent, comme on va le voir, de nombreuses analogies. Le but de ce cours est

•de dégager les structures mathématiques communes à tous ces modèles pour les analyser ;

•de montrer comment les deux descriptions, par les équations de transport et par les équations de diffusion, sont intimement reliées et se complètent ;

•d'étudier et de mettre en œuvre des méthodes numériques de résolution de ces équations, et d'identifier leurs régimes de validité.

1.1 Origine des équations de transport et de diffusion

Avant de présenter quelques-uns des modèles physiques ou biologiques spécifiques (comme par exemple la neutronique, le transfert radiatif, la dynamique des populations ...) qui seront les exemples fondamentaux sur lesquels nous

baserons notre étude, nous allons commencer par expliquer succintement comment on arrive à établir les équations de diffusion et de transport, qui sont les principaux objets mathématiques étudiés dans ce cours.

1.1.1 Établissement de l'équation de diffusion

On veut étudier l'évolution d'une population de particules dans un milieu matériel, comme par exemple des neutrons dans un cœur de réacteur nucléaire. On suppose que ces particules sont absorbées par le milieu, puis réémises instantanément au même point mais dans une direction aléatoire, et que toutes les directions sont équiprobables. (Autrement dit, la direction de la particule réémise est distribuée uniformément sur la sphère unité.) On supposera de plus que le laps de temps séparant le moment où une particule est émise dans le milieu et celui où elle est absorbée est très petit devant l'échelle de temps sur laquelle on observe le système.

L'inconnue caractérisant l'état du système est la densité du nombre de particules au point x à l'instant t, notée

$$\rho \equiv \rho(t,x) \geq 0\,.$$

C'est-à-dire que, dans un volume infinitésimal dx centré au point x se trouvent environ $\rho(t,x)dx$ particules à l'instant t. Autrement dit, dans toute partie mesurable $A \subset \mathbf{R}^3$, on trouve, à l'instant t, environ

$$\int_A \rho(t,x)dx \text{ particules.}$$

(Ce calcul du nombre de particules présentes dans A à l'instant t n'est qu'approximatif puisque l'intégrale ci-dessus n'est en général pas un nombre entier.)

Soient $t_1 < t_2$ deux instants quelconques, et B une boule de $\mathbf{R}^3$; la variation du nombre de particules présentes dans la boule B entre les instants t_1 et t_2 est

$$\int_B \rho(t_2,x)dx - \int_B \rho(t_1,x)dx\,.$$

Cette différence est égale au nombre de particules entrées dans B moins le nombre de particules sorties de B entre les instants t_1 et t_2.

On cherche donc à calculer le nombre de particules ayant traversé le bord ∂B de B entre les instants t_1 et t_2. Ce calcul fait naturellement intervenir la notion de vecteur densité de courant, analogue à la notion de densité de courant électrique, et dont nous donnons la définition ci-dessous.

Cette densité de courant est le champ de vecteurs $J \equiv J(t,x) \in \mathbf{R}^3$ caractérisé par le fait que, pour tout élément de surface $dS(x)$ centré en $x \in \mathbf{R}^3$ et orienté par le vecteur unitaire n_x normal à $dS(x)$ au point x, l'on a

$$N_+ - N_- \simeq J(t,x) \cdot n_x dS(x)dt\,.$$

Dans cette formule, on a noté

$$N_\pm = \text{nombre de neutrons traversant } dS(x) \text{ dans la direction } \pm n_x \text{ dans l'intervalle de temps } [t, t+dt].$$

Donc

$$\int_B (\rho(t_2,x) - \rho(t_1,x))\, dx = -\int_{t_1}^{t_2} \int_{\partial B} J(t,x)\cdot n_x dS(x) dt\,.$$

Supposons que ρ et J sont de classe C^1 ; on transforme alors l'intégrale interne de droite par la formule de Green :

$$\int_{\partial B} J(t,x)\cdot n_x dS(x) = \int_B \mathrm{div}_x\, J(t,x) dx\,.$$

Puis on écrit que

$$\rho(t_2,x) - \rho(t_1,x) = \int_{t_1}^{t_2} \frac{\partial \rho}{\partial t}(t,x) dt\,,$$

de sorte que l'identité ci-dessus devient

$$\int_{t_1}^{t_2} \int_B \left(\frac{\partial \rho}{\partial t}(t,x) + \mathrm{div}_x\, J(t,x) \right) dt dx = 0\,.$$

Comme la fonction continue $\frac{\partial \rho}{\partial t} + \mathrm{div}_x\, J$ est d'intégrale nulle sur tout ensemble de la forme $[t_1,t_2]\times B$ où B est une boule de $\mathbf{R}^3$, on en déduit l'**équation de continuité**

$$\frac{\partial \rho}{\partial t}(t,x) + \mathrm{div}_x\, J(t,x) = 0\,.$$

Cette équation traduit la conservation locale du nombre de particules dans tout domaine à bord régulier de l'espace des positions. Autrement dit, la variation entre deux instants t_1 et t_2 du nombre de particules dans un tel domaine est égale au flux du vecteur densité de courant à travers le bord de ce domaine, intégré entre t_1 et t_2.

L'équation de continuité ne suffit pas à elle seule à déterminer l'inconnue ρ, car la densité de courant J est elle-même inconnue. L'établissement de l'équation de diffusion repose donc sur une hypothèse précisant comment le vecteur densité de courant J dépend de la densité du nombre de particules ρ.

Cette hypothèse porte, suivant les domaines d'application, le nom de **loi de Fick**, ou encore de **loi de Fourier** dans le contexte de la thermique. Elle consiste à postuler que la densité de courant J est proportionnelle au gradient spatial de ρ :

$$J(t,x) = -D\nabla_x \rho(t,x)\,, \qquad \text{avec } D>0\,.$$

Le signe de D est choisi par analogie avec l'exemple de la thermique dans un matériau. Dans ce cas, $\rho(t,x)$ est la température dans le matériau au point x à l'instant t, et J est la densité de courant de chaleur ; il est donc naturel que la chaleur "s'écoule" dans le sens opposé au gradient de température. Par exemple,

dans un mélange d'eau liquide et de glace, le flux de chaleur va de l'eau liquide (à température $> 0^0$C) vers la glace (à température $< 0^0$C) jusqu'à ce qu'un équilibre thermique soit atteint.

En substituant cette formule pour J dans l'équation de continuité, on aboutit à l'**équation de diffusion**

$$\frac{\partial \rho}{\partial t}(t,x) - D\Delta_x \rho(t,x) = 0\,. \tag{1.1}$$

Cette équation est identique à l'équation de la chaleur qu'on obtient par le même procédé de modélisation, en remplaçant la loi de Fick par celle de Fourier (voir par exemple [2] §1.2 ou [23] §9.1). Dans l'argument ci-dessus, on a supposé implicitement que $D > 0$ est une constante, ce qui correspond au cas d'un milieu homogène. Or la loi de Fick s'applique encore au cas de la diffusion de particules dans des matériaux non homogènes. Le coefficient de diffusion n'est alors plus une constante, mais une fonction de la variable de position $D \equiv D(x) > 0$. La loi de Fick s'écrit encore

$$J(t,x) = -D(x)\nabla_x \rho(t,x)\,.$$

Substituant cette expression de la densité de courant dans l'équation de continuité, on aboutit à la forme suivante de l'équation de diffusion pour un milieu inhomogène

$$\frac{\partial \rho}{\partial t}(t,x) - \mathrm{div}_x(D(x)\nabla_x \rho(t,x)) = 0\,.$$

Remarque 1.1.1 *Pour que la description de la population de particules considérées par l'équation de diffusion soit justifiée, il est absolument essentiel que l'intervalle de temps moyen entre l'apparition d'une particule dans le milieu et son absorption soit très petit par rapport à l'échelle de temps sur laquelle on observe la dynamique de cette population de particules. Comme on le verra plus loin, la modélisation par l'équation de diffusion n'est justifiée que dans une certaine limite asymptotique, incluant une hypothèse sur l'échelle de temps dans laquelle on observe le système.*

Remarque 1.1.2 *Dans le cas de neutrons dans un matériau fissile, il peut y avoir en outre dans l'équation de diffusion un terme d'amplification exponentielle de la densité de neutrons, du fait de la création de neutrons secondaires au cours des réactions de fission. Le cas de la neutronique sera étudié plus en détail dans la suite de ce cours (voir, par exemple, la section 1.2).*

1.1.2 Établissement de l'équation de transport

On considère à nouveau une population de particules interagissant avec un milieu matériel. Comme dans la section précédente, ces particules peuvent être absorbées par le milieu, puis réémises au même endroit, mais avec un vecteur vitesse différent. Pensons à l'exemple des neutrons dans un matériau fissile : la population de neutrons de vitesse donnée v est diminuée du nombre de neutrons

déviés par collision élastique avec les atomes du milieu, ou bien du fait de la capture de ces neutrons au cours d'une réaction de fission. D'autre part, cette population est augmentée des neutrons de vitesse v produits au même point du fait d'une collision élastique entre un atome du milieu et un neutron de vitesse v' avant la collision, ainsi que des neutrons secondaires de vitesse v émis par la réaction de fission.

A la différence de l'étude faite dans la section précédente, on observe cette population de particules sur des temps plus courts, de l'ordre du laps de temps moyen s'écoulant entre le moment où une particule est émise dans le milieu et celui où elle est absorbée. Le modèle de diffusion n'est plus valable à cette échelle.

Comme on recherche une compréhension plus détaillée de la dynamique de cette population de particules que dans la section précédente, on va avoir recours à une description plus complexe, qui est celle de la théorie cinétique — analogue à la théorie cinétique des gaz due à Maxwell et Boltzmann.

Dans cette nouvelle description, la fonction inconnue décrivant la population de particules est la **fonction de distribution**

$$f \equiv f(t,x,v) \geq 0\,,$$

qui est la densité du nombre de particules situées au point x et animées de la vitesse v à l'instant t. Notons que la fonction inconnue n'est plus seulement une fonction de la variable de temps t et de la position x, mais qu'elle fait intervenir en plus la variable v échantillonnant toutes les valeurs possibles du vecteur vitesse pour une particule. C'est-à-dire que, contrairement au cas de la modélisation par l'équation de diffusion, où l'espace des phases est l'ensemble de toutes les positions possibles pour une particule — par exemple l'espace euclidien $\mathbf{R}^3$ si la population de particules n'est pas confinée, ou au contraire un sous-domaine Ω de $\mathbf{R}^3$ dans le cas où ces particules sont contenues dans une enceinte — la modélisation cinétique fait intervenir comme espace des phases l'ensemble de tous les couples position-vitesse (x,v) possibles pour une particule, à savoir $\mathbf{R}^3 \times \mathbf{R}^3$ ou $\Omega \times \mathbf{R}^3$. Soulignons tout de suite le fait que la modélisation cinétique est beaucoup plus coûteuse (notamment pour ce qui est des simulations numériques) que celle par l'équation de diffusion, puisqu'elle double presque le nombre de variables de la fonction inconnue.

Comme la description cinétique et celle par l'équation de diffusion représentent la même réalité, il est important de comprendre comment elles sont reliées. La notion essentielle pour cela est celle d'**observable macroscopique** dans la modélisation cinétique.

Soit $\phi(v)$, quantité physique additive pour une seule particule de vitesse v ; par exemple

$$\begin{cases} \phi(v) = 1 & \text{(nombre de particules)}, \\ \phi(v) = mv & \text{(quantité de mouvement)}, \\ \phi(v) = \frac{1}{2}m|v|^2 & \text{(énergie cinétique)}. \end{cases}$$

La quantité globale correspondante pour l'ensemble des particules se trouvant

à l'instant t dans le domaine spatial A vaut

$$\int_A \int_{\mathbf{R}^3} \phi(v) f(t,x,v) dx dv \,.$$

La densité spatiale de cette quantité physique à l'instant t vaut donc

$$\int_{\mathbf{R}^3} \phi(v) f(t,x,v) dv \,.$$

En particulier, la densité du nombre de particules $\rho \equiv \rho(t,x)$ considérée dans la théorie de la diffusion est reliée à la fonction de distribution f par la formule

$$\rho(t,x) = \int_{\mathbf{R}^3} f(t,x,v) dv \,.$$

De même, la densité de courant s'exprime à partir de la fonction de distribution par la formule

$$J(t,x) = \int_{\mathbf{R}^3} v f(t,x,v) dv \,.$$

Remarque 1.1.3 *Supposons que le nombre total N de particules présentes dans le milieu considéré est fini et constant au cours du temps :*

$$0 < N = \iint_{\mathbf{R}^3 \times \mathbf{R}^3} f(t,x,v) dx dv < \infty \,.$$

Les observables macroscopiques ont alors une interprétation probabiliste : en effet, la mesure

$$\frac{1}{N} f(t,x,v) dx dv \text{ est une mesure de probabilité.}$$

Alors la quantité

$$\int_A \int_{\mathbf{R}^3} \phi(v) f(t,x,v) dx dv$$

s'interprète comme une espérance mathématique :

$$\int_A \int_{\mathbf{R}^3} \phi(v) f(t,x,v) dx dv = \mathbf{E}(\mathbf{1}_A(x)\phi(v)) \,.$$

De même la quantité

$$\int_{\mathbf{R}^3} \phi(v) f(t,x,v) dv$$

s'interprète comme une espérance conditionnelle connaissant la position x :

$$\int_{\mathbf{R}^3} \phi(v) f(t,x,v) dv = \mathbf{E}(\phi(v)|x) \,.$$

Pour plus de détails nous renvoyons, par exemple, à [26].

Remarque 1.1.4 *Cette notion d'observable macroscopique est évidemment à rapprocher de la notion d'observable en mécanique quantique [6]. En effet, en mécanique quantique, une quantité physique est un opérateur sur un espace de Hilbert — par exemple sur* $L^2(\mathbf{R}^3;\mathbf{C})$ *dans le cas de la mécanique quantique d'un point matériel. Si* K *est une quantité physique définie par un opérateur sur* $L^2(\mathbf{R}^3;\mathbf{C})$ *— par exemple l'énergie cinétique*

$$-\tfrac{\hbar^2}{2m}\Delta_x\,,$$

— on appelle "observable" correspondant à la quantité K *pour la fonction d'onde* ψ *la quantité*

$$(\psi|K\psi)_{L^2} = Trace(K\,|\psi\rangle\langle\psi|)\,,$$

en utilisant la notation bra-ket habituelle en mécanique quantique et en notant $(\psi|\phi)_{L^2}$ *le produit scalaire usuel dans* $L^2(\mathbf{R}^3)$*. Autrement dit, sachant que* $|\psi(x)|^2$ *est une densité de probabilité sur* $\mathbf{R}^3$*, c'est-à-dire que*

$$\int_{\mathbf{R}^3}|\psi(x)|^2dx = 1\,,$$

la notation $|\psi\rangle\langle\psi|$ *désigne l'application linéaire*

$$\phi \mapsto (\psi|\phi)_{L^2}\psi\,,$$

où l'on a noté

$$(\psi|\phi)_{L^2} := \int_{\mathbf{R}^3}\overline{\psi(x)}\phi(x)dx\,.$$

Cette application linéaire est donc la projection orthogonale dans $L^2(\mathbf{R}^3)$ *sur la droite vectorielle engendrée par* ψ*,*

Par exemple, si K *est un opérateur intégral de noyau* $k \equiv k(x,y)$*, c'est-à-dire si*

$$K\psi(x) = \int_{\mathbf{R}^3}k(x,y)\psi(y)dy\,,$$

alors l'observable associé est

$$(\psi|K\psi)_{L^2} = \iint_{\mathbf{R}^3\times\mathbf{R}^3}k(x,y)\overline{\psi(x)}\psi(y)dxdy\,.$$

Dans cette définition, la "matrice densité" $\overline{\psi(x)}\psi(y)$ *joue un rôle analogue à celui de la fonction de distribution* $f(x,v)$*, tandis que l'opérateur* K *est l'analogue de la fonction* $(x,v)\mapsto \mathbf{1}_A(x)\phi(v)$*.*

Nous allons maintenant établir l'équation régissant l'évolution de la fonction de distribution f grâce à un bilan du nombre de particules dans un domaine arbitraire de l'espace des phases, de façon tout à fait analogue à ce qui a été fait dans la section précédente dans le cas l'équation de diffusion.

Soient donc $t_1 < t_2$, et B, B' deux boules de $\mathbf{R}^3$; la variation du nombre de particules situées dans la boule B et de vecteur vitesse appartenant à la boule

B' entre les instants t_1 et t_2 vaut, en supposant la fonction de distribution f de classe C^1 :

$$\int_B \int_{B'} (f(t_2, x, v) - f(t_1, x, v))\, dxdv = \int_{t_1}^{t_2} \int_B \int_{B'} \frac{\partial f}{\partial t}(t, x, v) dtdxdv\,.$$

Cette variation peut avoir l'une des causes suivantes :
•certaines particules situées dans B et de vitesse appartenant à B' à l'instant t_1 ont été absorbées par le milieu à l'intérieur de la boule B entre les instants t_1 et t_2 ;
•entre les instants t_1 et t_2, des particules ont été créées dans la boule B avec une vitesse appartenant à la boule B' ;
•enfin, certaines particules présentes à l'instant t_1 dans le domaine $B \times B'$ de l'espace des phases en sont sorties sans être absorbées par le milieu entre les instants t_1 et t_2, ou encore, d'autres particules sont entrées dans le domaine $B \times B'$ entre les instants t_1 et t_2.

Evaluons la contribution de chacun de ces phénomènes.

Absorption. Le nombre de particules de vitesse $v \in B'$ absorbées par le milieu entre les instants t_1 et t_2 par la portion de matériau située dans la boule B vaut (environ)

$$N_A = \int_{t_1}^{t_2} \int_B \int_{B'} \sigma(x, v) f(t, x, v) dtdxdv\,,$$

où $\sigma(x, v) > 0$ est le taux d'absorption du milieu à la position x et pour des particules de vitesse v. Cette fonction σ est une caractéristique physique du matériau, qui est donnée.

Création. Le nombre de particules créées avec une vitesse $v \in B'$ entre les instants t_1 et t_2 par la portion de matériau située dans la boule B au cours d'une transition $v' \to v$, où v' est une vitesse quelconque dans $\mathbf{R}^3$, vaut environ

$$N_C = \int_{t_1}^{t_2} \int_B \int_{B'} \int_{\mathbf{R}^3} k(x, v, v') f(t, x, v') dtdxdvdv'\,,$$

où $k(x, v, v') > 0$ est le taux de transition conduisant à la création d'une particule de vitesse v à la position x à partir de l'absorption d'une particule de vitesse v' en ce même point x. A nouveau, cette fonction k est une caractéristique physique du milieu, qui est donc donnée.

Advection. Supposons pour simplifier que les particules considérées ne sont soumises à aucune force extérieure — en négligeant notamment l'effet de la gravité. Par conséquent, les équations du mouvement pour chaque particule sont

$$\begin{cases} \dot{x} = v \\ \dot{v} = 0 \end{cases}$$

— où la notation $\dot{}$ désigne la dérivée par rapport à la variable de temps. Comme chaque particule est de vitesse constante, la seule façon pour sa trajectoire dans

l'espace des phases d'entrer dans $B \times B'$ ou d'en sortir consiste à traverser la partie $\partial B \times B'$ du bord de $B \times B'$. Notons

$$N_\pm = \text{nombre de particules entrées dans/sorties de } B \times B' \text{ entre les instants } t_1 \text{ et } t_2\,;$$

alors $N_+ - N_-$ est égal au flux à travers le bord ∂B de B entre les instants t_1 et t_2 du vecteur densité de courant correspondant aux particules de vitesses appartenant à B'.

Exprimons tout d'abord la densité de courant correspondant aux particules de vitesses appartenant à B' : comme expliqué ci-dessus, on trouve

$$J_{B'}(t,x) = \int_{B'} v f(t,x,v) dv\,.$$

Par conséquent, en notant n_x le vecteur unitaire normal à ∂B au point x, dirigé vers l'extérieur de B, on trouve que

$$\begin{aligned} N_+ - N_- &= -\int_{t_1}^{t_2} \left(\int_{\partial B} J_{B'}(t,x) \cdot n_x dS(x) \right) dt \\ &= -\int_{t_1}^{t_2} \int_{\partial B} \int_{B'} f(t,x,v) v \cdot n_x dt dS(x) dv\,. \end{aligned}$$

Toujours sous l'hypothèse que f est de classe C^1, on peut échanger l'ordre des intégrations en t, x et v, puis transformer l'intégrale sur ∂B par la formule de Green, ce qui donne

$$N_+ - N_- = -\int_{t_1}^{t_2} \int_B \int_{B'} \mathrm{div}_x(v f(t,x,v)) dt dx dv\,.$$

Revenons maintenant au bilan de la variation du nombre de particules dans le domaine $B \times B'$ de l'espace des phases entre les instants t_1 et t_2 :

$$\int_{t_1}^{t_2} \int_B \int_{B'} \frac{\partial f}{\partial t}(t,x,v) dt dx dv = N_+ - N_- + N_C - N_A\,.$$

Grâce aux formules obtenues ci-dessus pour $N_\pm$, N_A et N_C, on trouve donc que

$$\int_{t_1}^{t_2} \int_B \int_{B'} \left(\frac{\partial f}{\partial t} + \mathrm{div}_x(vf) + \sigma f - Kf \right)(t,x,v) dt dx dv = 0\,,$$

en notant K l'opérateur défini comme suit :

$$Kf(t,x,v) = \int_{\mathbf{R}^3} k(x,v,v') f(t,x,v') dv'\,.$$

Supposons, outre le fait que f est de classe C^1, que les fonctions σ et k sont continues, et que $v' \mapsto k(x,v,v')$ décroît suffisamment vite pour $|v'| \to \infty$

localement uniformément en (x,v) pour que l'intégrale Kf soit bien définie. Alors l'intégrande ci-dessus

$$\frac{\partial f}{\partial t} + \mathrm{div}_x(vf) + \sigma f - Kf$$

est continue. Comme son intégrale sur tout domaine du type $[t_1,t_2]\times B\times B'$ (avec $t_1<t_2$ et B,B' boules de $\mathbf{R}^3$) est nulle, on en déduit que la fonction de distribution f vérifie l'**équation de Boltzmann linéaire**

$$\frac{\partial f}{\partial t}+\mathrm{div}_x(vf)+\sigma f-Kf=0\,,$$

c'est-à-dire que

$$\frac{\partial f}{\partial t}(t,x,v)+v\cdot\nabla_x f(t,x,v)+\sigma(x,v)f(t,x,v)-\int_{\mathbf{R}^3}k(x,v,v')f(t,x,v')dv'=0\,. \tag{1.2}$$

Un cas particulier très simple d'équation de Boltzmann linéaire est l'exemple de l'**équation de transport monocinétique conservative avec scattering isotrope**. Le terme "monocinétique" indique que toutes les particules sont de même vitesse. Sans perte de généralité, on peut se ramener au cas où cette vitesse vaut 1 : on notera donc

$$v=\omega\,,\qquad \text{avec } |\omega|=1\,.$$

Le terme "scattering isotrope" se rapporte à des processus d'absorption et de création définis par des taux d'absorption σ et de scattering k constants. Le terme "conservatif" signifie que le nombre de particules absorbées compense exactement le nombre de particules créées dans toute boule B entre deux instants quelconques $t_1<t_2$. Autrement dit, en notant $d\omega$ l'élément de surface sur la sphère unité,

$$\begin{aligned}\sigma\int_{t_1}^{t_2}\int_B\int_{|\omega|=1}f(t,x,\omega)dtdxd\omega&\\ =k\int_{t_1}^{t_2}\int_B\int_{|\omega|=1}\int_{|\omega'|=1}f(t,x,\omega')dtdxd\omega d\omega'&\end{aligned}$$

pour tous $t_1<t_2$ et toutes les boules $B\subset\mathbf{R}^3$, de sorte que

$$k=\tfrac{1}{4\pi}\sigma\,.$$

Autrement dit

$$Kf(t,x,\omega)=\sigma\langle f\rangle(t,x)\,,$$

où $\langle\cdot\rangle$ désigne la moyenne en ω :

$$\langle\psi\rangle=\tfrac{1}{4\pi}\int_{|\omega|=1}\psi(\omega)d\omega\,.$$

Par conséquent, l'équation de Boltzmann linéaire conservative dans le cas du transport monocinétique avec scattering isotrope s'écrit

$$\frac{\partial f}{\partial t}+\omega\cdot\nabla_x f+\sigma(f-\langle f\rangle)=0\,. \tag{1.3}$$

On étudiera notamment un cas particulier du transport monocinétique avec scattering isotrope, celui où la fonction de distribution f ne dépend que d'une seule variable d'espace, par exemple x_1. Cette hypothèse est connue en anglais sous le nom de **"slab symmetry"** tandis qu'en français on parle plutôt de géométrie de type **"plaque infinie"**.

Dans ce cas l'équation de Boltzmann linéaire ci-dessus devient

$$\frac{\partial f}{\partial t}+\omega_1\frac{\partial f}{\partial x_1}+\sigma(f-\langle f\rangle)=0\,.$$

Ecrivons le vecteur unitaire ω en coordonnées sphériques, l'axe des x_1 jouant le rôle d'axe polaire. Autrement dit

$$\omega=(\cos\theta,\sin\theta\cos\phi,\sin\theta\sin\phi)\,,\quad 0\le\theta\le\pi\,,\quad 0\le\phi<2\pi\,,$$

(voir Figure 1.1). Alors

$$\langle f\rangle(t,x_1)=\tfrac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} f(t,x_1,\cos\theta,\sin\theta\cos\phi,\sin\theta\sin\phi)\sin\theta d\phi d\theta\,.$$

Faisons le changement de variables

$$\mu=\cos\theta\,,\quad \text{de sorte que } \sin\theta=\sqrt{1-\mu^2}\text{ et } d\mu=-\sin\theta d\theta\,.$$

(On note que $\sin\theta\ge 0$ puisque $\theta\in[0,\pi]$.) Donc

$$\langle f\rangle(t,x_1)=\int_{-1}^{1}\left(\int_0^{2\pi} f(t,x_1,\mu,\sqrt{1-\mu^2}\cos\phi,\sqrt{1-\mu^2}\sin\phi)\frac{d\phi}{2\pi}\right)\frac{d\mu}{2}\,.$$

Posons

$$F(t,x_1,\mu)=\int_0^{2\pi} f(t,x_1,\mu,\sqrt{1-\mu^2}\cos\phi,\sqrt{1-\mu^2}\sin\phi)\frac{d\phi}{2\pi}\,; \tag{1.4}$$

alors

$$\langle f\rangle(t,x_1)=\int_{-1}^{1}F(t,x_1,\mu)\frac{d\mu}{2}\,.$$

Sous l'hypothèse que f est de classe C^1, la dérivation sous le signe somme est légitime, de sorte qu'on peut calculer la moyenne en ϕ de la fonction

$$\left(\frac{\partial f}{\partial t}+\mu\frac{\partial f}{\partial x_1}+\sigma f\right)(t,x_1,\mu,\sqrt{1-\mu^2}\cos\phi,\sqrt{1-\mu^2}\sin\phi)$$

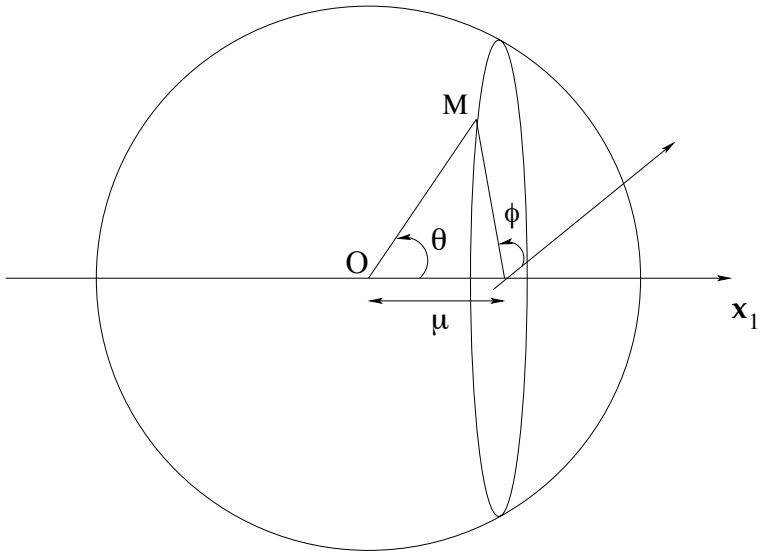

FIGURE 1.1 – Coordonnées sphériques du point M

comme suit :

$$\int_0^{2\pi} \left(\frac{\partial f}{\partial t}+\mu\frac{\partial f}{\partial x_1}+\sigma f\right)(t,x_1,\mu,\sqrt{1-\mu^2}\cos\phi,\sqrt{1-\mu^2}\sin\phi)\frac{d\phi}{2\pi} = \left(\frac{\partial F}{\partial t}+\mu\frac{\partial F}{\partial x_1}+\sigma F\right)(t,x_1,\mu)\,.$$

On déduit de ce qui précède que si la fonction $f \equiv f(t,x_1,\omega)$ de classe C^1 vérifie l'équation

$$\frac{\partial f}{\partial t}(t,x_1,\omega)+\omega_1\frac{\partial f}{\partial x_1}(t,x_1,\omega)+\sigma f(t,x_1,\omega)=\sigma\langle f\rangle(t,x_1)\,,$$

alors la fonction $F \equiv F(t,x_1,\mu)$ définie par (1.4) vérifie l'équation de Boltzmann linéaire pour la "plaque infinie"

$$\frac{\partial F}{\partial t}(t,x_1,\mu)+\mu\frac{\partial F}{\partial x_1}(t,x_1,\mu)+\sigma F(t,x_1,\mu)=\sigma\int_{-1}^{1} F(t,x_1,\mu')\frac{d\mu'}{2}\,. \qquad (1.5)$$

1.1.3 Du transport vers la diffusion

On a déjà expliqué comment la notion d'observable macroscopique permet de passer de la fonction de distribution f, qui décrit un système de particules dans le cadre de la théorie cinétique, à la densité macroscopique du nombre de particules ρ que l'on rencontre dans la théorie de la diffusion.

On verra dans la suite de ce cours comment l'équation de diffusion peut être déduite de l'équation de Boltzmann linéaire dans un certain régime asymptotique. Ce résultat jouera un rôle de premier plan dans ce cours, aussi bien sur le plan théorique que du point de vue de l'analyse numérique.

Nous allons essayer de donner un premier aperçu de ce résultat dans le cas particulier de l'équation de transport monocinétique conservative avec scattering isotrope.

Soit donc f vérifiant

$$\frac{\partial f}{\partial t}+\omega\cdot\nabla_x f+\sigma(f-\langle f\rangle)=0\,,$$

où on rappelle que

$$\langle f\rangle=\tfrac{1}{4\pi}\int_{|\omega|=1} f d\omega\,.$$

Intégrons par rapport à la variable ω chaque membre de cette égalité : toujours en supposant que f est de classe C^1, ce qui légitime la dérivation sous le signe somme, on aboutit ainsi à l'**équation de continuité** :

$$\frac{\partial}{\partial t}\int_{|\omega|=1} f(t,x,\omega)d\omega+\mathrm{div}_x\int_{|\omega|=1}\omega f(t,x,\omega)d\omega=0\,,$$

ce qui s'écrit encore

$$\frac{\partial\rho}{\partial t}+\mathrm{div}_x\, J=0\quad\text{ avec }\left\{\begin{array}{l}\rho=\langle f\rangle\,,\\ J=\langle\omega f\rangle\,.\end{array}\right.$$

Ensuite, on réécrit l'équation de Boltzmann linéaire sous la forme

$$f=\langle f\rangle-\frac{1}{\sigma}\omega\cdot\nabla_x f-\frac{1}{\sigma}\frac{\partial f}{\partial t}\,. \tag{1.6}$$

Supposons maintenant que $\sigma\gg 1$. En supposant les variations de f petites devant σ, c'est-à-dire

$$\left|\frac{\partial f}{\partial t}\right|\ll\sigma\text{ et }|\nabla_x f|\ll\sigma\,,$$

l'égalité (1.6) entraîne en particulier que f est asymptotiquement **isotrope** — c'est-à-dire indépendante de ω :

$$f(t,x,\omega)\simeq\langle f\rangle(t,x)=\rho(t,x)\,.$$

En injectant cette approximation dans (1.6), on en déduit que

$$f\simeq\rho-\frac{1}{\sigma}\omega\cdot\nabla_x\rho-\frac{1}{\sigma}\frac{\partial\rho}{\partial t}\,. \tag{1.7}$$

Evidemment, ce n'est pas parce que $f\simeq\langle f\rangle=\rho$ que $\nabla_x f\simeq\nabla_x\rho$, ni que $\partial_t f\simeq\partial_t\rho$ — sauf à se placer dans le cadre de la théorie des distributions. Mais si tel est le cas, autrement dit si l'approximation (1.7) est justifiée, on a alors

$$J=\langle\omega f\rangle\simeq\langle\omega\rho\rangle-\frac{1}{\sigma}\langle\omega\omega\cdot\nabla_x\rho\rangle-\frac{1}{\sigma}\frac{\partial\langle\omega\rho\rangle}{\partial t}\,,$$

et, comme $\langle \omega\rho\rangle = \langle\omega\rangle\rho = 0$ (puisque $\langle\omega\rangle = 0$), on obtient

$$J_k \simeq -\frac{1}{\sigma}\sum_{l=1}^{3}\langle\omega_k\omega_l\rangle\frac{\partial\rho}{\partial x_l}\,, \quad k=1,2,3\,. \tag{1.8}$$

Observons que

$$\langle\omega_k\omega_l\rangle = 0\,, \quad \text{si } k\neq l$$

et d'autre part que

$$\langle\omega_1^2\rangle = \langle\omega_2^2\rangle = \langle\omega_3^2\rangle$$

par invariance par rotation de la mesure uniforme sur la sphère unité. Donc

$$\langle\omega_1^2\rangle = \langle\omega_2^2\rangle = \langle\omega_3^2\rangle = \tfrac{1}{3}\sum_{k=1}^{3}\langle\omega_k^2\rangle = \tfrac{1}{3}\langle|\omega|^2\rangle = \tfrac{1}{3}\,.$$

Au total

$$\langle\omega_k\omega_l\rangle = \tfrac{1}{3}\delta_{kl}\,, \quad k,l=1,2,3\,.$$

En revenant à l'expression (1.8) pour J, on trouve que

$$J_k \simeq -\frac{1}{\sigma}\sum_{l=1}^{3}\langle\omega_k\omega_l\rangle\frac{\partial\rho}{\partial x_l} = -\tfrac{1}{3\sigma}\frac{\partial\rho}{\partial x_k}\,, \quad k=1,2,3\,.$$

On aboutit donc ainsi à la **loi de Fick** : sous l'hypothèse que $\sigma \gg 1$, on a

$$J \simeq -D\nabla_x\rho \quad \text{ avec } D = \tfrac{1}{3\sigma}\,.$$

Alors que la loi de Fick était *postulée* dans notre présentation de l'équation de diffusion (cf. section 1.1.1), l'argument ci-dessus permet de la *déduire* de l'équation de Boltzmann linéaire, au moins dans le cas monocinétique avec scattering isotrope. Nous retrouverons une variante de cet argument dans un cadre plus général et avec des démonstrations complètes au chapitre 4, consacré à l'approximation des solutions de l'équation de Boltzmann par les solutions de l'équation de diffusion.

En substituant cette valeur du vecteur densité de courant dans l'équation de continuité (comme on l'a fait dans la section 1.1.1 portant sur l'obtention de l'équation de diffusion), on trouve que, sous l'hypothèse que le taux d'absorption $\sigma \gg 1$, alors la fonction de distribution f, solution de l'équation de transport monocinétique conservative avec scattering isotrope, vérifie

$$f(t,x,\omega) \simeq \rho(t,x)\,,$$

et que ρ est solution de l'équation de diffusion

$$\frac{\partial\rho}{\partial t} - \frac{1}{3\sigma}\Delta_x\rho = 0\,.$$

On verra plus loin dans ce cours (voir le chapitre 4) comment rendre cette approximation de la fonction de distribution f par la solution ρ de l'équation de diffusion parfaitement rigoureuse — et même comment mesurer l'erreur correspondante.

1.2 Neutronique

1.2.1 Modélisation physique

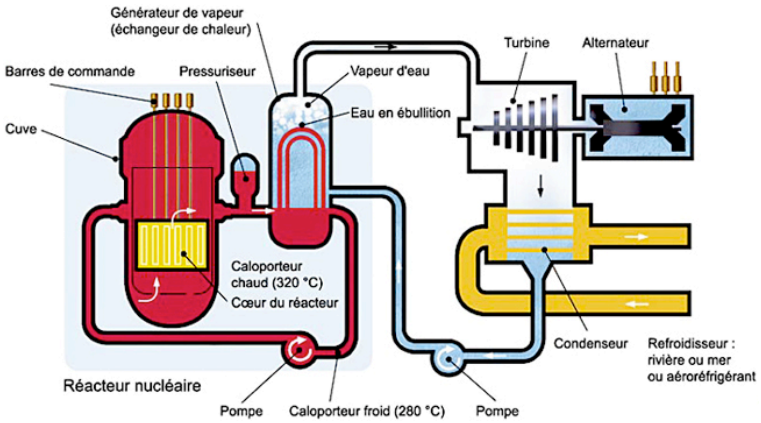

FIGURE 1.2 – Principe de fonctionnement d'un réacteur à eau pressurisée.

La neutronique est la branche de la physique, et plus particulièrement de la physique des réacteurs nucléaires, qui étudie le comportement d'une population de neutrons créés par radioactivité naturelle, ou plus généralement artificielle (voir les ouvrages de références [12], [45]). On peut distinguer au moins deux grandes classes de modèles en neutronique : les modèles de transport, qui sont précis même à de petites échelles spatiales ou temporelles, mais coûteux à résoudre, et les modèles de diffusion, qui ne sont valables qu'à des échelles macroscopiques d'espace ou de temps, mais faciles à résoudre numériquement. L'un des enjeux de ce cours est de comprendre les relations entre ces deux classes de modèles pour savoir arbitrer efficacement entre les deux.

Nous allons nous limiter dans ce bref exposé à une introduction aux équations et aux notations de la neutronique dans les cœurs de réacteurs nucléaires. Un réacteur à eau pressurisée typique (ceux de la gamme produisant 900 mégawatts électriques, voir Figure 1.2) a un cœur composé de 157 assemblages d'une hauteur d'environ 4 mètres, à section carrée de 21 centimètres de coté et répartis périodiquement sur un réseau carré inscrit dans un cercle correspondant à la section de la cuve du réacteur (voir Figure 1.3). Chaque assemblage est lui-même un réseau carré de 17 par 17 crayons de combustible (voir Figure 1.4) qui sont des tubes métalliques renfermant les pastilles de combustibles (de l'oxyde d'uranium, UO_2, enrichi, c'est-à-dire que la proportion de l'isotope $U235$ par rapport à $U238$ est augmentée de 0.7% à 3.1%). Le tout baigne dans un écoulement d'eau qui joue un double rôle de modérateur (c'est-à-dire de milieu qui ralentit les neutrons créés par fission pour qu'ils puissent à leur tour entrer en collision avec les noyaux fissiles et entretenir la réaction en chaîne) et de fluide

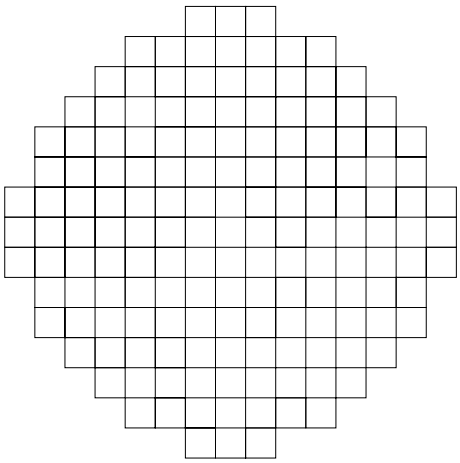

FIGURE 1.3 – Section horizontale du cœur d'un réacteur nucléaire comprenant 157 assemblages combustibles.

caloporteur (c'est-à-dire qui extrait la chaleur produite par fission dans le cœur pour la déposer dans un circuit d'eau secondaire qui produira de l'électricité en actionnant une turbine). Par ailleurs, un certain nombre de crayons de combustible sont remplacés par des tubes guides dans lesquels coulissent les barres de contrôle qui peuvent s'enfoncer verticalement plus ou moins profondément dans le cœur. Leur rôle est de contrôler la réaction de fission, voire de la stopper complètement, car elles sont constituées de matériaux absorbant les neutrons. Un cœur de réacteur est donc un milieu très hétérogène où la précision des modèles de transport est nécessaire à l'échelle du crayon ou de l'assemblage, mais où leur coût en temps de calcul peut être prohibitif à l'échelle du cœur entier. Dans ce dernier cas on leur préfère des modèles de diffusion qui sont communément utilisés pour des calculs de routine.

Les notations et les variables choisies en neutronique sont particulières à cette discipline et nous expliquons maintenant leur origine. Le point de départ est l'équation de Boltzmann linéaire (1.2) pour une population de neutrons de densité $f(t,x,v)$ où $t \in \mathbf{R}$ est le temps, $x \in \mathbf{R}^3$ la variable d'espace et $v \in \mathbf{R}^3$ la variable de vitesse. La variable de vitesse v est remplacée par sa direction ω et par l'énergie cinétique correspondante E, définies par

$$v = |v|\,\omega\ , \quad E = \frac{1}{2} m |v|^2$$

où m est la masse du neutron. L'inconnue f, qui est la densité de neutrons, est remplacée par le flux de neutrons $\phi = |v|\,f$ qui est une quantité plus facile à mesurer expérimentalement. Le gradient de ϕ, ou bien ses composantes $\omega \cdot \nabla_x \phi$, sont appelés courants. L'équation de transport pour $\phi(t,x,\omega,E)$ prend alors la

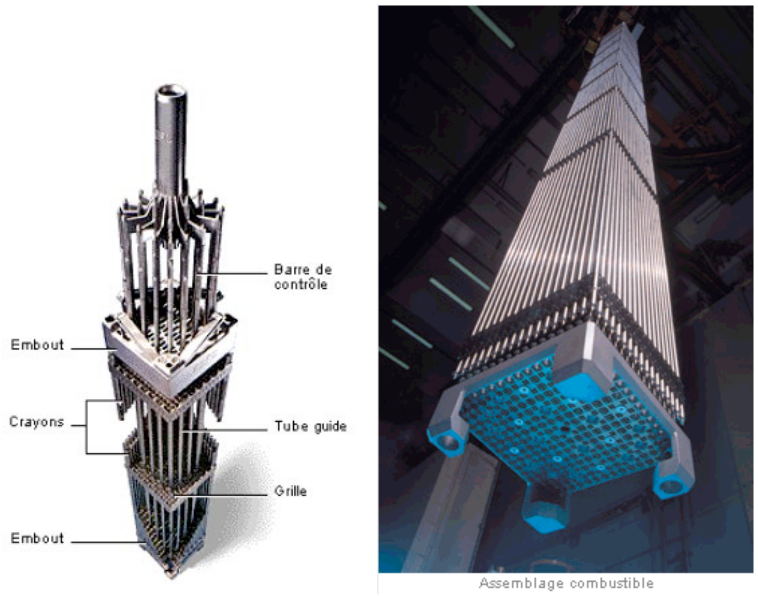

FIGURE 1.4 – Assemblage combustible composé de crayons.

forme

$$\frac{1}{|v|}\frac{\partial \phi}{\partial t} + \omega \cdot \nabla_x \phi + \sigma^t(x,\omega,E)\phi = S(x,\omega,E) + \int_{E_{min}}^{E_{max}} \int_{|\omega'|=1} \sigma^*(x,\omega,E,\omega',E')\phi(x,\omega',E')\, d\omega'\, dE', \tag{1.9}$$

où $S(x,\omega,E)$ est le terme source. Les coefficients σ^t et σ^* sont appelés sections efficaces : σ^t est la section efficace totale qui mesure le taux de disparition de neutrons au point (x,ω,E) de l'espace des phases, tandis que σ^* correspond à une source de neutrons qui apparaissent au point (x,ω,E) en provenance ou par transformation de neutrons situés auparavant au point (x,ω',E') de l'espace des phases. (Par rapport aux notations de (1.2) on a la correspondance $\sigma^t = \sigma/|v|$ et $\sigma^* = k/|v|$.) En général, le milieu est supposé isotrope, ce qui a pour conséquence que σ^t ne dépend en fait que de (x,E), mais pas de la direction ω, et que σ^* ne dépend de (ω,ω') qu'à travers l'angle que forment ces deux directions, c'est-à-dire que $\sigma^* \equiv \sigma^*(x,\omega\cdot\omega',E,E')$.

D'un point de vue physique, l'opérateur intégral dans le membre de droite de (1.9) a deux origines bien distinctes. D'une part, il modélise la collision (ou "scattering") des neutrons avec les noyaux des atomes du milieu ambiant : autrement dit, au point x, un neutron de vitesse (ω',E') rebondit sur un noyau et prend la nouvelle vitesse (ω,E). D'autre part, il représente la création par fission au point x de neutrons de vitesse (ω,E) créés grâce à la capture par un

noyau fissile d'un neutron de vitesse (ω', E'). C'est pourquoi la section efficace σ^* doit être décomposée en somme de deux termes

$$\sigma^*(x, \omega \cdot \omega', E, E') = \sigma^c(x, \omega \cdot \omega', E, E') + \nu(E)\sigma^f(x, \omega \cdot \omega', E, E')\,, \qquad (1.10)$$

où σ^c est la section efficace de collision, σ^f est la section efficace de fission et $\nu(E)$ est le nombre moyen de neutrons émis par fission (dont la valeur typique pour l'uranium est 2.42).

De même, la section efficace totale peut être décomposée en somme de trois termes

$$\begin{aligned}\sigma^t(x, E) = \sigma^a(x, E) + \int_{E_{min}}^{E_{max}} \int_{|\omega'|=1} \Big(&\sigma^c(x, \omega \cdot \omega', E, E') \\ &+ \sigma^f(x, \omega \cdot \omega', E, E')\Big) d\omega'\, dE',\end{aligned} \qquad (1.11)$$

où σ^a est la section efficace d'absorption (des neutrons capturés par des noyaux et qui "disparaissent" ainsi du bilan neutronique), les fonctions σ^c et σ^f étant les mêmes que dans (1.10). Bien sûr toutes ces sections efficaces sont positives ou nulles. Dans la suite, pour simplifier la présentation, nous n'utiliserons pas la décomposition (1.11), nous supposerons que $\nu(E) = 1$ et nous ne retiendrons la décomposition (1.10) que lorsque celle-ci sera importante d'un point de vue physique et mathématique, notamment lors de l'étude de la criticité (voir chapitre 6).

Les valeurs extrêmes de l'énergie sont très contrastées : typiquement, on aura $E_{min} = 10^{-5}$ eV et $E_{max} = 20$ MeV (eV pour électron-volt). Le support de la fonction $\sigma^f(x, \omega \cdot \omega', E, E')$ est très étroit au sens suivant : seuls les neutrons de basse énergie, dits **thermiques** (environ $E' = 1$ eV), peuvent fissionner les noyaux du combustible nucléaire, tandis que les neutrons produits par fission sont, eux, de très haute énergie (environ $E = 1$ MeV) et appelés **neutrons rapides**. Il faut donc que le combustible nucléaire soit entouré d'un milieu qui ralentisse (d'un facteur 1000 !) les neutrons rapides pour que ceux-ci deviennent thermiques et puissent engendrer de nouvelles réactions de fission : un tel milieu est dit **modérateur**. Dans les réacteurs à eau pressurisée, l'eau du circuit primaire joue le rôle de modérateur. Dans d'autres types de réacteurs (par exemple à gaz), le modérateur peut être du graphite. En première approximation, la section efficace de collision $\sigma^c(x, \omega \cdot \omega', E, E')$ définit un opérateur triangulaire au sens où, les collisions étant élastiques ou inélastiques, l'énergie après collision E ne peut être que plus petite que l'énergie avant collision E', c'est-à-dire que $E \leq E'$. Cependant, on peut assister à des remontées (faibles) en énergie, $E > E'$ (up-scattering en anglais), car une partie de l'agitation thermique de chaque noyau peut être transférée sous forme d'énergie cinétique aux neutrons entrant en collision avec lui.

Remarque 1.2.1 *L'unité de flux ϕ est le $m^{-2}s^{-1}$ (avec m pour mètre et s pour seconde), celle des sections efficaces (dites macroscopiques) σ^t et σ^* est le m^{-1}, et celle de la source S, le $m^{-3}s^{-1}$.*

Remarque 1.2.2 *En vérité tous les neutrons produits par fission ne sont pas instantanément émis après collision d'un neutron avec un noyau fissile. Une proportion infime (mais significative, de l'ordre de quelques millièmes) de ces neutrons sont émis après un certain délai correspondant à la désintégration radioactive d'isotopes intermédiaires instables. On parle dans ce cas de neutrons retardés, avec un délai de l'ordre d'une seconde à une minute. Par comparaison, les temps caractéristiques de libre parcours moyen d'un neutron (entre deux collisions) sont de l'ordre de* 10^{-4} *à* 10^{-8} *secondes. Pour tenir compte de ce phénomène, il faudrait coupler l'équation de transport à un systèmes d'équations différentielles ordinaires décrivant la rétention puis l'émission de ces neutrons retardés. Par souci de simplicité dans la présentation nous ne le ferons pas ici et nous renvoyons à [12] et [38] pour plus de détails. Néanmoins ces neutrons retardés, bien que peu nombreux, sont extrêmement importants en pratique pour pouvoir piloter un réacteur nucléaire. En effet, c'est leur présence (ou plutôt leur retard) qui permet, par exemple, d'avoir le temps d'enclancher un système d'arrêt d'un réacteur par chute des barres de contrôle dans le cœur, stoppant la réaction nucléaire. S'il n'existait pas de tels neutrons retardés, l'emballement de la réaction nucléaire (entraînant l'explosion du réacteur) en cas d'incident serait quasiment instantané, ne permettant pas d'avoir le temps matériel de réagir en insérant les barres de contrôle.*

1.2.2 Formalisme multigroupe

Bien que la variable d'énergie E soit continue dans le modèle (1.9), il est nécessaire de la discrétiser en pratique. Pour cela, on divise le spectre d'énergie $[E_{min}, E_{max}]$ en un nombre fini de sous-intervalles, appelés **groupes** (de un à quelques centaines) et numérotés par ordre décroissant d'énergie

$$E_{max} = E_0 > E_1 > \cdots > E_G = E_{min}.$$

Pour $1 \leq g \leq G$, on note $\phi_g(t, x, \omega)$ une approximation de

$$\int_{E_g}^{E_{g-1}} \phi(t, x, \omega, E)\, dE.$$

De la même manière on définit les sources

$$S_g(x, \omega) = \int_{E_g}^{E_{g-1}} S(x, \omega, E)\, dE,$$

et on note $|v_g|$ une vitesse moyenne pour le groupe g. On introduit alors des moyennes des sections efficaces sur chacun des groupes d'énergie pour obtenir le système suivant, dit **équation du transport multigroupe**

$$\begin{aligned} \frac{1}{|v_g|}\frac{\partial \phi_g}{\partial t} + \omega \cdot \nabla_x \phi_g + \sigma_g(x)\phi_g &= S_g(x, \omega) \\ &+ \sum_{g'=1}^{G} \int_{|\omega'|=1} \sigma^*_{gg'}(x, \omega \cdot \omega')\phi_{g'}(x, \omega')\, d\omega'. \end{aligned} \tag{1.12}$$

Remarquons que nous obtenons ainsi un système de G équations de transport couplées entre elles par le terme de collision-fission.

Le calcul des sections efficaces moyennes en énergie, σ_g et $\sigma^*_{gg'}$, n'est pas simple car leurs versions continues en énergie sont très oscillantes pour les hautes énergies. Cela est dû au phénomène de résonance : un neutron avec une énergie bien précise peut être capturé et créer avec le noyau qu'il percute un nouvel isotope. Le calcul des sections efficaces moyennes en présence de résonances est une procédure délicate qui repose, entre autres, sur des moyennes spatiales adéquates : on parle alors d'"autoprotection" (self-shielding en anglais). Nous renvoyons à [12] pour plus de détails.

1.2.3 Approximation par la diffusion

Revenons à la question de l'approximation du transport par la diffusion déjà présentée dans la section 1.1.3. On définit le flux scalaire comme la moyenne du flux sur toutes les directions angulaires

$$u_g(t,x) = \int_{|\omega|=1} \phi_g(t,x,\omega)\, d\omega\,,$$

ainsi que la densité de courant totale

$$j_g(t,x) = \int_{|\omega|=1} \omega\, \phi_g(t,x,\omega)\, d\omega.$$

De même la source scalaire est

$$f_g(x) = \int_{|\omega|=1} S_g(x,\omega)\, d\omega.$$

Par ailleurs, on remarque que

$$\int_{|\omega|=1} \sigma^*_{gg'}(x, \omega\cdot\omega')\, d\omega$$

est indépendant de ω' par invariance par rotation de la sphère unité, et on note $\sigma_{gg'}(x)$ la valeur de cette intégrale.

On intègre alors (1.12) par rapport à ω pour obtenir

$$\frac{1}{|v_g|}\frac{\partial u_g}{\partial t} + \operatorname{div} j_g + \sigma_g(x) u_g = f_g(x) + \sum_{g'=1}^{G} \sigma_{gg'}(x) u_{g'}. \tag{1.13}$$

Aucune approximation n'a encore été faite pour aboutir à (1.13). L'approximation de la diffusion consiste à supposer maintenant que la densité de courant totale j_g est reliée au flux scalaire u_g par la **loi de Fick** qui postule l'existence d'un coefficient de diffusion $D_g > 0$ tel que

$$j_g(t,x) = -D_g \nabla u_g(t,x). \tag{1.14}$$

En combinant (1.13) et (1.14) on obtient un système d'équations paraboliques, pour $1 \leq g \leq G$,

$$\frac{1}{|v_g|}\frac{\partial u_g}{\partial t} - \operatorname{div}(D_g \nabla u_g) + \sigma_g(x) u_g = f_g(x) + \sum_{g'=1}^{G} \sigma_{gg'}(x) u_{g'}. \tag{1.15}$$

L'approximation de (1.12) par (1.15) sera justifiée au chapitre 4 dans le cas d'un seul groupe ($G = 1$).

Un cas particulier simple de (1.15) est le modèle de diffusion à deux groupes d'énergie ($G = 2$), où u_1 est le flux de neutrons rapides et u_2 celui des neutrons thermiques. Dans ce cas la remontée en énergie lors des collisions est négligeable, et on obtient le système

$$\begin{cases} \dfrac{1}{|v_1|}\dfrac{\partial u_1}{\partial t} - \operatorname{div}(D_1 \nabla u_1) + \sigma_1(x) u_1 = f_1(x) + \sigma_{12}^f(x) u_2 \,, \\ \dfrac{1}{|v_2|}\dfrac{\partial u_2}{\partial t} - \operatorname{div}(D_2 \nabla u_2) + \sigma_2(x) u_2 = f_2(x) + \sigma_{21}^c(x) u_1 \,, \end{cases} \tag{1.16}$$

où σ_{12}^f est la section efficace de fission qui prend en compte la création de neutrons rapides à partir de neutrons thermiques, et σ_{21}^c est la section efficace de collision, ou de ralentissement, mesurant le taux de transformation des neutrons rapides en neutrons thermiques.

1.3 Le transfert radiatif

1.3.1 Les équations du transfert radiatif

Le transfert radiatif est la branche de la physique décrivant le transport d'énergie par rayonnement électromagnétique à travers un milieu matériel — par exemple une atmosphère stellaire, ou planétaire. Pour une présentation détaillée du sujet, le lecteur pourra se reporter aux ouvrages de référence que sont [15], [44] ou [40].

Dans le cadre du transfert radiatif, le rayonnement électromagnétique est modélisé de façon corpusculaire, en considérant un gaz de photons, et non pas de façon ondulatoire par les équations de Maxwell.

Dans toutes les applications dont il sera question ici, on ne considèrera que le cas de milieux d'indice [1] constant, donc non dispersifs — et même, pour simplifier, d'indice 1. Le gaz de photons considéré est donc monocinétique, la vitesse des photons étant alors c (la vitesse de la lumière dans le vide).

Chaque photon est donc caractérisé par sa position x, sa direction ω, et sa fréquence ν. Comme l'énergie d'un photon de fréquence ν vaut $h\nu$ où h est la constante de Planck, se donner la fréquence d'un photon est équivalent à se donner la norme $|v|$ de la vitesse d'une particule de masse $m > 0$, pour laquelle

1. On rappelle que l'indice d'un milieu est le rapport c/v de la vitesse de la lumière dans le vide c à la vitesse v de la lumière dans ce milieu.

l'énergie cinétique est $\frac{1}{2}m|v|^2$. Autrement dit, la donnée de la fréquence d'un photon et de sa direction est l'analogue exact de la donnée du vecteur vitesse d'une particule massique.

La modélisation cinétique de ce gaz de photons suggère donc de considérer la fonction de distribution des photons

$$f \equiv f(t,x,\omega,\nu) = \text{densité du nombre de photons situés au point } x, \text{ à l'instant } t, \text{ de direction } \omega \text{ et de fréquence } \nu.$$

En réalité, on préfère, en transfert radiatif, considérer la quantité

$$I(t,x,\omega,\nu) = ch\nu f(t,x,\omega,\nu)\,,$$

appelée **intensité radiative**.

Dans tout ce qui suit, on suppose que la matière est immobile, ou à tout le moins que son mouvement s'effectue sur des échelles de temps beaucoup plus longues que celles qui sont adaptées au transfert de rayonnement.

L'interaction entre le rayonnement et la matière se fait selon 3 mécanismes bien distincts :

a) le scattering,
b) l'absorption,
c) l'émission.

Disons quelques mots de ces trois mécanismes, sans entrer toutefois dans les détails, car il s'agit de processus physiques très complexes — voir [44], chapitre 7, pour plus de détails.

Scattering. Les photons changent brutalement de direction, et parfois de fréquence par suite de "collisions" avec les électrons du milieu. On s'intéressera essentiellement à deux cas particuliers :

•le scattering Thomson, et

•le scattering Rayleigh.

Le scattering Thomson est un mécanisme de scattering classique des ondes électromagnétiques par les électrons libres du milieu. Sous l'effet de l'onde électromagnétique incidente, un électron libre initialement immobile oscille dans la direction de la composante électrique de l'onde. On peut donc l'assimiler à un dipôle oscillant, qui rayonne donc une nouvelle onde électromagnétique, dont la direction n'est en général pas la même que celle de l'onde incidente. En revanche, la fréquence de l'onde émise est la même que celle de l'onde incidente : on dit alors que le scattering Thomson est *cohérent*. Le scattering Thomson est un mécanisme concernant des électrons initialement immobiles, et des photons incidents d'énergie négligeable devant l'énergie au repos de l'électron m_0c^2, où m_0 désigne la masse de l'électron.

Dans un volume infinitésimal $dxd\omega d\nu$ de l'espace des phases des photons centré au point (x,ω,ν), l'énergie des photons diminue de

$$\tfrac{1}{4\pi}\kappa_s^{Thomson} I(t,x,\omega,\nu)dtdxd\omega d\nu$$

dans un intervalle de temps infinitésimal de longueur dt ; d'autre part cette même énergie augmente par scattering sur les électrons libres du milieu de

$$\kappa_s^{Thomson} \left(\int_{|\omega'|=1} I(t,x,\omega',\nu) \tfrac{3}{16\pi} (1 + (\omega \cdot \omega')^2) d\omega' \right) dt dx d\omega d\nu \, .$$

Le coefficient de scattering Thomson vaut

$$\kappa_s^{Thomson} = \frac{8\pi e_0^4}{3 m_0^2 c^4} \, ,$$

où e_0 et m_0 sont respectivement la charge et la masse de l'électron. De façon remarquable, ce coefficient est constant — en particulier, il ne dépend pas de la fréquence de l'onde incidente.

Le scattering Rayleigh est un mécanisme de scattering différent — par des particules diélectriques de polarisabilité α, dont les dimensions sont petites devant la longueur de l'onde électromagnétique incidente. Ce mécanisme de scattering est décrit par les mêmes formules que le scattering Thomson, sauf que le coefficient de scattering Rayleigh vaut

$$\kappa_s^{Rayleigh} = \frac{128\pi^5 \alpha^2 \nu^4}{3c^2} \, .$$

Comme le scattering Thomson, le scattering Rayleigh est cohérent. Toutefois, au contraire du coefficient de scattering Thomson, le coefficient de scattering Rayleigh dépend de la fréquence ν de l'onde incidente — et croît comme la puissance quatrième de cette fréquence. Ce mécanisme de scattering est donc d'autant plus important que la fréquence de l'onde incidente est élevée.

Il existe plusieurs autres mécanismes de scattering jouant un rôle important pour les problèmes d'interaction rayonnement-matière, par exemple le scattering Compton par des électrons libres, dans le cas où la fréquence du photon incident n'est plus négligeable devant l'énergie au repos de l'électron, et où l'électron peut se trouver en mouvement avant l'interaction. Dans ce cas, l'électron peut communiquer une partie de son énergie cinétique au photon, ce qui en décale la fréquence. Donc, contrairement au scattering Thomson ou Rayleigh, le scattering Compton n'est pas un mécanisme de scattering cohérent : la fréquence du photon émis est en général différente de la fréquence du photon incident.

Absorption/Emission. Les mécanismes d'absorption et d'émission de photons par la matière ont été décrits par Einstein grâce au formalisme de la mécanique quantique. Considérons un atome dont les électrons ont des niveaux d'énergie notés E_k — ces niveaux d'énergie pouvant d'ailleurs être discrets ou continus. Un photon peut donc être absorbé en excitant un électron d'un état d'énergie E_m vers un état d'énergie $E_n > E_m$ pourvu que ce photon soit de fréquence $\nu_{m,n}$ telle que

$$E_n - E_m = h\nu_{m,n} \, .$$

Il peut également y avoir absorption par échange d'énergie entre un photon et un électron libre, ou encore photoionisation, suivant la fréquence — et donc

l'énergie — du photon incident. Réciproquement, un électron peut être désexcité en passant de l'état $E_n > E_m$ à l'état E_m, en émettant un photon d'énergie $h\nu_{m,n}$.

Pour avoir une description raisonnablement simple de ces mécanismes d'absorption et d'émission, on fait l'hypothèse que le milieu est en **équilibre thermodynamique local** (ETL) : en tout point x du milieu et à tout instant t, il existe une température électronique de la matière, notée $T(t,x)$.

Sous cette hypothèse, dans un élément de volume infinitésimal $dxd\omega d\nu$ de l'espace des phases, l'énergie des photons absorbés par le milieu dans un intervalle de temps de longueur dt vaut

$$\sigma_\nu(T(t,x))I(t,x,\omega,\nu)dtdxd\omega d\nu\,.$$

Le coefficient $\sigma_\nu(T) > 0$ est l'opacité — c'est-à-dire le taux d'absorption — du milieu porté à la température T aux radiations de fréquence ν.

Inversement, dans ce même élément de volume de l'espace des phases et pendant le même laps de temps, le milieu émet une énergie

$$\sigma_\nu(T(t,x))B_\nu(T(t,x))dtdxd\omega d\nu\,,$$

où $B_\nu(T)$ est la **fonction de Planck**, qui est l'intensité radiative d'un corps noir porté à la température T. Elle est donnée par la formule

$$B_\nu(T) = \frac{2h\nu^3}{c^2(e^{h\nu/kT}-1)}\,,$$

où h est la constante de Planck, k la constante de Boltzmann et c la vitesse de la lumière dans le vide.

On vérifie sans peine que l'énergie rayonnée dans toutes les directions et toutes les fréquences par un corps noir porté à la température T est donnée par la **loi de Stefan-Boltzmann** :

$$\frac{4\pi}{c}\int_0^\infty B_\nu(T)d\nu = aT^4\,,$$

avec

$$a = \frac{2\pi^5k^4}{15h^3c^3}\,.$$

La discussion ci-dessus montre toute l'importance de l'opacité $\sigma_\nu(T)$ pour décrire ces phénomènes d'absorption et d'émission en supposant qu'on est dans le régime ETL. L'opacité est une fonction donnée, qui dépend de la composition chimique de la matière, de la température électronique T et de la fréquence des photons incidents. Ceci vaut pour un matériau parfaitement homogène ; sinon, l'opacité dépend en général d'autres quantités physiques, comme par exemple la densité du milieu. En général, la dépendance en T et ν de l'opacité $\sigma_\nu(T)$ est extrêmement complexe, puisque ce coefficient contient à lui seul toute l'information relative à ces différents mécanismes d'absorption des radiations par la matière. La détermination des opacités résulte de calculs compliqués de mécanique quantique, recalés lorsque cela est possible par des données expérimentales. Plus les

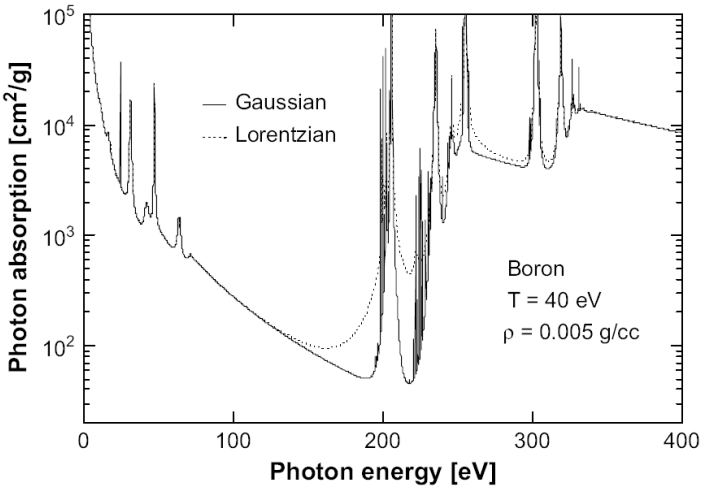

FIGURE 1.5 – Opacité du bore

atomes considérés ont de niveaux d'énergie, plus il y a de possibilités pour des transitions entre niveaux d'énergie électronique différents, et donc plus l'opacité $\sigma_\nu(T)$ est difficile à déterminer.

Le cas le plus simple — le seul où il y ait une formule élémentaire donnant l'opacité est celui des transitions entre électrons libres pour l'atome d'hydrogène :

$$\sigma_\nu(T) = \frac{4}{3} N \left(\frac{2\pi}{3m_0}\right)^{1/2} \frac{h^2 e^6}{cm_0} \left(1 + 0.1728 \left(\frac{h\nu}{I_H}\right)^{1/3} \left(1 + \frac{2kT}{h\nu}\right)\right) \frac{1 - e^{-h\nu/kT}}{kT(h\nu)^3}$$

où N est le nombre d'atomes par unité de volume et I_H le potentiel d'ionisation de l'atome d'hydrogène ($I_H = 13.6\text{eV}$). Cette formule est appelée opacité de Kramers.

Mais en général, l'opacité est une fonction dépendant des variables ν et T de manière extrêmement complexe, comme le montre la Figure 1.5. En pratique, on utilise donc des valeurs tabulées de cette fonction dans les simulations numériques.

On conclut ce rapide tour d'horizon en écrivant l'**équation du transfert radiatif** :

$$\begin{aligned} \frac{1}{c}\frac{\partial I}{\partial t}(t,x,\omega,\nu) + \omega \cdot \nabla_x I(t,x,\omega,\nu) &= \sigma_\nu(T(t,x))(B_\nu(T(t,x)) - I(t,x,\omega,\nu)) \\ &+ \kappa_s \int_{|\omega'|=1} \frac{3}{16\pi}(1 + (\omega \cdot \omega')^2)(I(t,x,\omega',\nu) - I(t,x,\omega,\nu))d\omega' \,. \end{aligned} \tag{1.17}$$

Dans cette équation, la température est donnée, ou bien au contraire couplée au rayonnement par l'équation

$$\frac{1}{c}\frac{\partial}{\partial t}\mathcal{E}(T(t,x)) = \frac{1}{4\pi}\int_0^\infty \int_{|\omega|=1} \sigma_\nu(T(t,x))(B_\nu(T(t,x)) - I(t,x,\omega,\nu))d\omega d\nu\,,$$

où $\mathcal{E}(T)$ est la densité d'énergie interne du milieu porté à la température T. Par exemple, dans le cas d'un gaz parfait de capacité calorifique constante, la densité d'énergie $\mathcal{E}(T)$ est de la forme

$$\mathcal{E}(T) = c_V T\,,$$

où c_V est la capacité calorifique à volume constant du milieu.

Nous allons conclure cette trop brève présentation du transfert radiatif en évoquant deux phénomènes physiques bien connus où il joue un rôle essentiel.

1.3.2 L'effet de serre

La Terre reçoit de l'énergie du Soleil sous forme de rayonnement lumineux, dans la gamme des fréquences correspondant à la lumière visible. La température de surface du Soleil est d'environ 5800K ; si on le considère en première approximation comme un corps noir à cette température, l'intensité radiative qu'il émet est donnée par la fonction de Planck $B_\nu(T_S)$ pour $T_S = 5800K$. La longueur d'onde maximisant son intensité radiative est de $0.5\mu m$ environ, et le flux radiatif solaire à la limite de l'atmosphère terrestre vaut $1360W/m^2$ en moyenne.

Le coefficient d'albedo A (rapport de l'énergie solaire réfléchie à l'énergie solaire incidente) de la Terre est $A \simeq 0.3$. Environ un tiers de l'énergie est donc réfléchi dans l'espace, en particulier par les nuages présents dans l'atmosphère terrestre qui contribuent beaucoup à son albedo. Les deux tiers restants sont absorbés par la surface de la Terre et, dans une moindre mesure, par l'atmosphère, qui est essentiellement transparente au rayonnement dans le domaine visible.

Comme la température terrestre est dans un état d'équilibre, l'énergie absorbée par la Terre est donc forcément rayonnée en retour dans l'espace. A nouveau, on peut considérer que la Terre est un corps noir pour simplifier le raisonnement, et que l'intensité radiative qu'elle rayonne est $B_\nu(T_r)$, où T_r est sa température d'équilibre radiatif — voir plus loin. Comme la température de la Terre T_r est très inférieure à celle du Soleil T_S, l'essentiel de cette énergie est rayonnée en retour sous forme de rayons infra-rouges.

Or l'atmosphère terrestre est plus opaque aux rayons infra-rouges qu'au rayonnement dans la gamme de fréquences correspondant aux radiations visibles. Une partie de l'énergie rayonnée par la Terre en direction de l'espace sous forme de rayons infra-rouges est donc absorbée par l'atmosphère, en particulier par les nuages, et réémise en partie vers la surface de la Terre. Ce mécanisme, connu sous le nom d'effet de serre, augmente la dose d'énergie rayonnée vers la

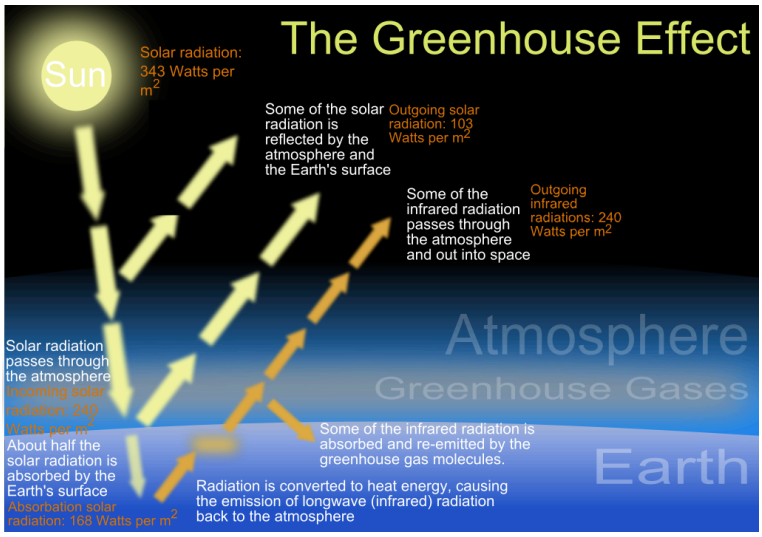

FIGURE 1.6 – Représentation schématique de l'effet de serre terrestre.

surface de la Terre, ce qui en élève la température. D'autres phénomènes physiques que le seul bilan radiatif esquissé ci-dessus interviennent aussi, comme la convection dans l'atmosphère.

On peut confirmer cette analyse par le calcul élémentaire suivant : l'énergie reçue par la surface de la Terre est

$$(1-A)\pi R^2 F_S\,,$$

où R est le rayon terrestre, et F_S le flux radiatif solaire à la limite de l'atmosphère terrestre. D'autre part, la loi de Stefan-Boltzmann stipule que l'énergie rayonnée par la surface de la Terre est

$$4\pi R^2 \cdot a T_r^4\,,$$

où a est la constante de Stefan-Boltzmann et T_r la température d'équilibre radiatif de la Terre considérée comme un corps noir. A l'équilibre et sans effet de serre, on doit donc avoir

$$(1-A)\pi R^2 F_S = 4\pi R^2 \cdot a T_r^4\,,$$

ce qui donne

$$T_r = \left(\frac{(1-A)F_S}{4a}\right)^{1/4}\,.$$

Ce calcul donne $T_r \simeq 255K = -18^0C$, ce qui est beaucoup plus froid que la température moyenne à la surface de la Terre — à peu près $288K = 15^0C$. L'effet de serre explique cette différence de température d'une trentaine de degrés.

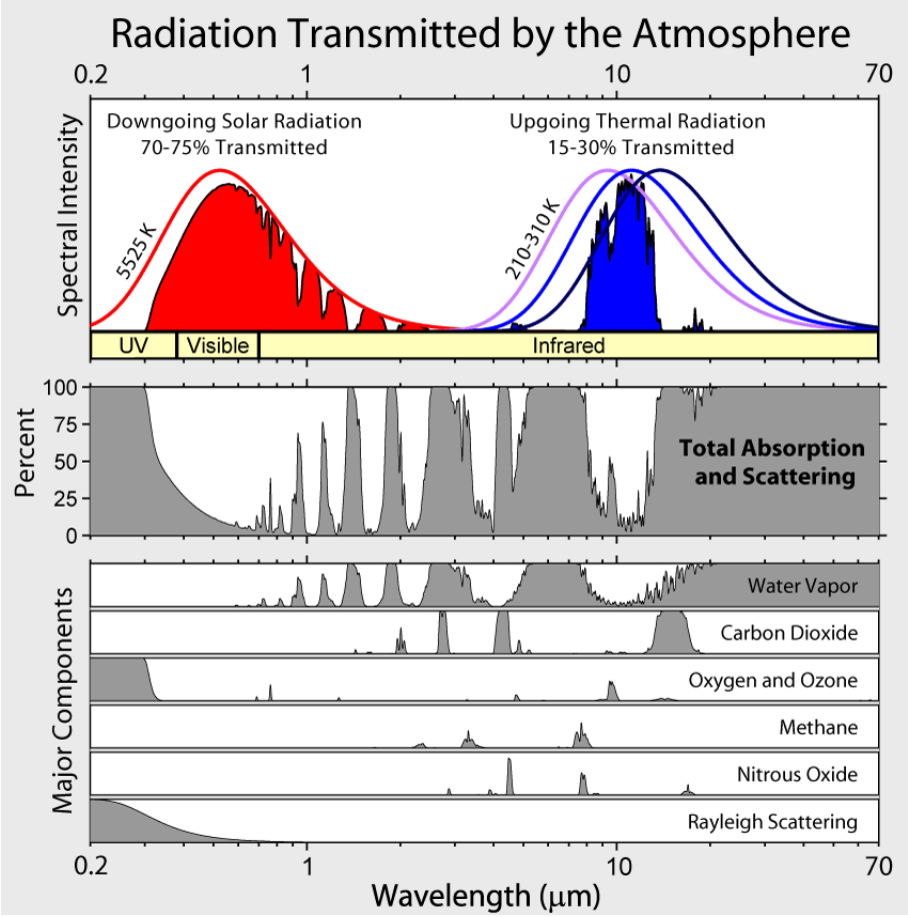

FIGURE 1.7 – Transfert radiatif dans l'atmosphère terrestre. En haut : les intensités radiatives émises par le Soleil et par la Terre en fonction de la longueur d'onde ; au milieu : spectre d'absorption totale de l'atmosphère terrestre ; en bas : spectre d'absorption des différents composants de l'atmosphère.

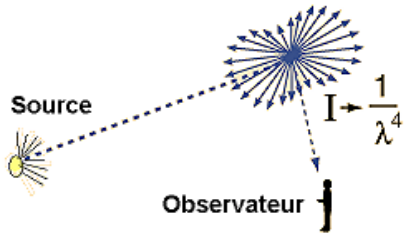

FIGURE 1.8 – Diffusion Rayleigh

L'azote et l'oxygène présents dans l'atmosphère ne jouent qu'un rôle minime dans l'effet de serre ; les deux gaz présents dans l'atmosphère et responsables pour l'essentiel de ce phénomène sont la vapeur d'eau et le dioxyde de carbone. La concentration de ce dernier dans l'atmosphère augmente en particulier avec l'activité humaine.

1.3.3 La couleur du ciel

L'effet de serre, dont nous avons sommairement décrit le principe dans la section précédente, résulte des phénomènes d'absorption et d'émission par l'atmosphère terrestre.

La couleur bleue du ciel — sans nuages — est une autre manifestation du transfert radiatif dans l'atmosphère terrestre. Une première explication en fut proposée par Tyndall et Rayleigh dans la seconde moitié du XIXème siècle.

L'idée de Tyndall était que la lumière bleue est plus fortement diffusée par les poussières et les gouttelettes d'eau présentes en suspension dans l'atmosphère que les autres couleurs de la partie visible du spectre.

Cette idée fut étudiée précisément sur le plan théorique par Rayleigh, qui arriva à la formule

$$\kappa_s^{Rayleigh} = \frac{128\pi^5\alpha^2\nu^4}{3c^2}$$

pour le taux de scattering des ondes lumineuses de fréquence ν par des sphères diélectriques de polarisabilité α.

Le point fondamental pour ce qui va suivre est que $\kappa_s^{Rayleigh}$ est proportionnel à la quatrième puissance de la fréquence incidente.

La lumière visible est constituée d'ondes électromagnétiques de longueurs d'onde (inversement proportionnelles aux fréquences) comprises entre $0.38\mu m$ correspondant au violet et $0,7\mu m$ correspondant au rouge.

Le rapport des fréquences de la lumière bleue à celle de la lumière rouge est donc environ $7/4$. La formule de Rayleigh montre donc que l'effet de scattering est environ $(7/4)^4 \simeq 10$ fois plus fort pour la lumière bleue que pour la lumière rouge.

En réalité, le raisonnement de Rayleigh ne s'applique pas seulement à des particules diélectriques, mais également à des atomes ou des molécules, dont les électrons peuvent être considérés comme des oscillateurs harmoniques. L'argument ci-dessus explique donc que la lumière bleue est plus fortement diffusée que la lumière rouge par les molécules d'azote et d'oxygène de l'atmosphère. Un observateur regardant le ciel perçoit donc davantage de lumière bleue que de lumière d'autres couleurs du spectre visible, comme expliqué par le schéma de la Figure 1.8. Toutefois, cet argument est incomplet : bien que l'effet de scattering Rayleigh soit plus fort pour la lumière violette que pour la lumière bleue, le ciel n'est pas perçu comme violet. Il faut également tenir compte, entre autres choses, du mécanisme de la vision humaine, qui joue ici un rôle important.

1.4 Biologie (dynamique des populations)

De nombreux modèles en biologie, et plus précisément en dynamique des populations, reposent sur des équations de transport. Nous nous inspirons ici du livre [42] (voir aussi [28], [39]). Les modèles de transport sont utilisés pour représenter l'évolution d'une population **structurée**, c'est-à-dire caractérisée par une variable interne, appelée **trait**, comme l'âge, la taille ou toute autre propriété attachée à chaque individu. Cette variable interne de structure jouera le rôle dévolu à la variable d'espace dans les équations de transport.

1.4.1 Population structurée par âge

On considère une population d'individus (des humains, des animaux, des cellules, etc.) caractérisés par leur âge x, évoluant au cours du temps t et de densité $n(t,x)$. On introduit deux coefficients : $d(x)$, le taux de mortalité à l'âge x, et $b(x)$, le taux de natalité à l'âge x. Par un simple bilan de population on

obtient l'équation, dite du **renouvellement**,

$$\begin{cases} \dfrac{\partial n}{\partial t}(t,x) + \dfrac{\partial n}{\partial x}(t,x) + d(x)\, n(t,x) = 0 & \text{pour } t>0,\ x>0, \\ n(t,0) = \displaystyle\int_0^{+\infty} b(y) n(t,y)\, dy & \text{pour } t>0, \\ n(0,x) = n^0(x) & \text{pour } x>0. \end{cases} \tag{1.18}$$

Ce modèle est plus simple qu'une équation de Boltzmann linéaire puisqu'il n'y a pas d'équivalent de la variable de vitesse v.

Une variante plus complexe du modèle (1.18) est le modèle de Rotenberg pour une population structurée par âge et maturation. On ajoute une variable supplémentaire de maturation $\mu \in [0,1]$ qui caractérise la vitesse de maturation ou de vieillissement. Autrement dit, x est désormais un âge biologique et x/μ un âge physique. Le taux de décès $d(x,\mu)$ peut dépendre aussi de la maturation, de même que le taux de natalité $b(x,\mu,\mu')$. La différence principale avec (1.18) provient d'un terme source additionnel qui prend en compte le changement possible de maturité par différents mécanismes représenté par un noyau $K(x,\mu,\mu')$. La densité $n(t,x,\mu)$ est alors solution de

$$\begin{cases} \left(\dfrac{\partial n}{\partial t} + \mu \dfrac{\partial n}{\partial x} + dn\right)(t,x,\mu) = \displaystyle\int_0^1 K(x,\mu,\mu') n(t,x,\mu')\, d\mu' & t,x>0,\ \mu\in[0,1], \\ n(t,0,\mu) = \displaystyle\int_0^{+\infty}\int_0^1 b(y,\mu,\mu') n(t,y,\mu')\, dy\, d\mu' & t>0,\ \mu\in[0,1], \\ n(0,x,\mu) = n^0(x,\mu) & x>0,\ \mu\in[0,1]. \end{cases}$$

Cette équation ressemble beaucoup à l'équation de Boltzmann linéaire (1.5) dans le cas de la "plaque infinie".

1.4.2 Population structurée par taille

On considère désormais une population d'individus (typiquement des cellules) caractérisés par leur taille x, évoluant au cours du temps t et de densité $n(t,x)$. On suppose que les individus croissent régulièrement en taille, puis se divisent en deux "enfants" de taille deux fois plus petite : c'est le phénomène de la **mitose** égale. Le taux de mitose est noté $b(x)$. On suppose qu'aucun individu de taille $x=0$ n'est introduit dans le système. On obtient donc l'équation

$$\begin{cases} \dfrac{\partial n}{\partial t}(t,x) + \dfrac{\partial n}{\partial x}(t,x) + b(x)\, n(t,x) = 4\, b(2x)\, n(t,2x) & \text{pour } t>0,\ x>0, \\ n(t,0) = 0 & \text{pour } t>0, \\ n(0,x) = n^0(x) & \text{pour } x>0. \end{cases} \tag{1.19}$$

Le coefficient 4 peut surprendre dans le membre de droite de (1.19) : expliquons son origine. Chaque individu de taille $2x$ produit deux enfants de taille x mais le taux de division b s'interprète en disant que l'incrément de population

$n(t,x)\,dx$ est égal à $2\,b(2x)\,n(t,2x)\,d(2x)$, d'où le coefficient 4. Autrement dit, le processus de mitose ne conserve pas le nombre mais la taille totale des individus fragmentés :

$$\int_0^{+\infty} x\,b(x)\,n(t,x)\,dx = \int_0^{+\infty} 4x\,b(2x)\,n(t,2x)\,dx. \tag{1.20}$$

C'est effectivement ce que l'on constate globalement : en intégrant par rapport à x l'équation (1.19) et en utilisant la condition aux limites en $x = 0$ (on suppose aussi que la solution $n(t,x)$ tend vers vers 0 assez vite lorsque x tend vers l'infini), on trouve que le nombre total d'individus augmente

$$\begin{aligned}\frac{d}{dt}\int_0^{+\infty} n(t,x)\,dx &= \int_0^{+\infty} 4\,b(2x)\,n(t,2x)\,dx - \int_0^{+\infty} b(x)\,n(t,x)\,dx \\ &= \int_0^{+\infty} b(x)\,n(t,x)\,dx \geq 0.\end{aligned}$$

Par ailleurs, en intégrant par rapport à x l'équation (1.19) multipliée par x et en utilisant la conservation (1.20), on en déduit que la taille moyenne globale augmente aussi

$$\frac{d}{dt}\int_0^{+\infty} x\,n(t,x)\,dx = \int_0^{+\infty} n(t,x)\,dx \geq 0.$$

Ce résultat n'est pas contradictoire avec la conservation (1.20) puisque l'opérateur de transport $\frac{\partial n}{\partial t} + \frac{\partial n}{\partial x}$ dans (1.19) indique qu'en l'absence de mitose, la population grandit uniformément avec le temps. Remarquons pour finir que ce modèle de mitose est de nouveau plus simple qu'une équation de Boltzmann linéaire puisqu'il n'y a pas d'équivalent de la variable de vitesse v.

Un modèle un peu plus général, et qui va conduire à une équation de Boltzmann linéaire, est celui de la mitose asymétrique où un individu de taille y se divise en deux enfants de taille $x \geq 0$ et $y - x \geq 0$ suivant le taux $b(x,y)$. On suppose donc que

$$b(x,y) \geq 0 \text{ et } b(x,y) = 0 \text{ si } x > y.$$

On suppose aussi que le modèle est symétrique par rapport aux deux enfants x et $y - x$, c'est-à-dire que

$$b(x,y) = b(y-x,y).$$

On note $b^*(y)$ le taux de disparition (par mitose) des individus de taille y qui est défini par

$$b^*(y) = \frac{1}{2}\int_0^{+\infty} b(x,y)\,dx = \frac{1}{2}\int_0^{y} b(x,y)\,dx,$$

et on suppose que le processus de mitose vérifie une hypothèse de conservation de la taille

$$y\,b^*(y) = \int_0^{+\infty} x\,b(x,y)\,dx = \int_0^{y} x\,b(x,y)\,dx,$$

dont on déduit l'équivalent de (1.20)

$$\int_0^{+\infty} y\, b^*(y)\, n(t,y)\, dy = \int_0^{+\infty} \int_0^{+\infty} x\, b(x,y)\, n(t,y)\, dx\, dy. \tag{1.21}$$

On obtient donc l'équation ci-dessous, de type Boltzmann linéaire,

$$\begin{cases} \dfrac{\partial n}{\partial t}(t,x) + \dfrac{\partial n}{\partial x}(t,x) + b^*(x)\, n(t,x) = \displaystyle\int_x^{+\infty} b(x,y)\, n(t,y)\, dy & t,x>0, \\ n(t,0) = 0 & t>0, \\ n(0,x) = n^0(x) & x>0. \end{cases} \tag{1.22}$$

On vérifie que les hypothèses faites sur les coefficients $b(x,y)$ et $b^*(x)$ sont bien cohérentes avec la modélisation adoptée. En premier lieu, le nombre d'individus dans la population augmente du fait de la mitose. En effet, en intégrant par rapport à x l'équation (1.22) et en utilisant la condition aux limites, on obtient

$$\begin{aligned} \frac{d}{dt}\int_0^{+\infty} n(t,x)dx &= \int_0^{+\infty}\left(\int_x^{\infty} b(x,y)n(t,y)dy\right)dx - \int_0^{+\infty} b^*(x)n(t,x)dx \\ &= \int_0^{+\infty}\left(\int_0^{y} b(x,y)dx\right)n(t,y)dy - \int_0^{+\infty} b^*(x)n(t,x)dx \\ &= \int_0^{+\infty} b^*(x)n(t,x)dx \geq 0. \end{aligned}$$

Deuxièmement, la taille moyenne de la population augmente car, en intégrant par rapport à x l'équation (1.22) multipliée par x et en utilisant la conservation (1.21), on obtient

$$\frac{d}{dt}\int_0^{+\infty} x\, n(t,x)\, dx = \int_0^{+\infty} n(t,x)\, dx \geq 0.$$

1.5 Exercices

Exercice 1.1 (Modèles structurés en âge) *Soit une population définie par sa fonction de densité $(t,a) \mapsto d(t,a)$ (t est le temps, a est l'âge).*

1. *On néglige les phénomènes de naissance et de mortalité. Montrer que la fonction d vérifie l'équation de transport*

$$\frac{\partial d}{\partial t}(t,a) + \frac{\partial d}{\partial a}(t,a) = 0, \qquad \text{pour tout } a,t>0.$$

Résoudre cette équation par la méthode des caractéristiques (voir le chapitre 2) dans le domaine $(t,a) \in \mathcal{D} = [0,+\infty[\times[0,+\infty[$ en calculant d en fonction de $d\big|_{t=0}$ et de $d\big|_{a=0}$.

2. *Le coefficient de mortalité est donné par la fonction $a \mapsto \mu(a) > 0$. Montrer que d vérifie l'équation de transport-absorption*

$$\frac{\partial d}{\partial t}(t,a) + \frac{\partial d}{\partial a}(t,a) = -\mu(a)d(t,a), \qquad \text{pour tout } (a,t) \in \mathcal{D}.$$

Résoudre cette équation comme dans le cas $\mu = 0$.

3. *La natalité est caractérisée par la fonction bornée $a \mapsto \sigma(a) \in [0, \sigma_{\max}]$ avec $\sigma_{\max} > 0$. Justifier la condition au bord $a = 0$*

$$d(t,0) = \int_0^\infty \sigma(a)d(t,a)da, \qquad \text{pour tout } t > 0.$$

4. *Pour une population donnée initiale $d(0,a) = d_0(a)$, écrire le problème complet que l'on doit étudier pour prédire l'évolution de la population. Montrer graphiquement à l'aide des caractéristiques dans $\mathcal{D}$ que le problème est bien posé.*

5. *On veut pouvoir prédire la stabilité de la population. Pour cela, dans cette question et les suivantes, on étudie les solutions du type*

$$d(t,a) = e^{\lambda t} f(a).$$

Ecrire les équations vérifiées par f.

6. *(plus difficile) Montrer que le problème se réduit à l'étude de l'équation*

$$f(0) = \left(\int_0^\infty \sigma(a) e^{-\lambda a - \int_0^a \mu(s)ds} da \right) f(0).$$

7. *(plus difficile) Montrer que si la natalité est trop faible, ce que l'on caractérise par*

$$\int_0^\infty \sigma(a)da \leq 1\,,$$

alors $d(t,a) \to 0$ pour $t \to \infty$. Montrer que c'est le cas si la natalité est non nulle uniquement pour une tranche d'âge étroite $a_- \leq a \leq a_+ = a_- + \varepsilon$.

Indications : Les questions 1, 2 et 3 se traitent en reprenant et appliquant les résultats de la section 1.4.1, la variable d'âge étant notée $a = x$. La question 4 se traite en reprenant la figure 2.2. Les questions suivantes correspondent à la théorie du calcul critique du chapitre 6. Il est conseillé de comparer plus particulièrement ces énoncés avec les résultats de la section 6.4.

Exercice 1.2 (Effet de modes) *Soit une population $d(t,a)$ très sensible aux effets de mode (t est le temps, a est l'âge). Une partie d^v porte des habits verts et l'autre $d^r = d - d^v$ des habits rouges. On suppose que l'envie de se distinguer (le dandysme) fait que les porteurs d'habits rouges changent pour des habits verts dans le cas où il y a trop de porteurs d'habits rouges (et inversement).*

1. *Justifier le système*

$$\begin{cases} \dfrac{\partial d^v}{\partial t}(t,a) + \dfrac{\partial d^v}{\partial a}(t,a) = \sigma(a)(d^r(t,a) - d^v(t,a)), \quad a \in \mathbf{R}\,,\ t > 0\,, \\ \dfrac{\partial d^r}{\partial t}(t,a) + \dfrac{\partial d^r}{\partial a}(t,a) = \sigma(a)(d^v(t,a) - d^r(t,a)), \quad a \in \mathbf{R}\,,\ t > 0\,, \end{cases}$$

où le coefficient d'échange $a \mapsto \sigma(a) \geq 0$ peut varier en fonction de l'âge.

2. *Calculer la solution en fonction de $d^v\big|_{t=0}$ et $d^r\big|_{t=0}$.*

Indication : c'est un exercice très simple qu'il faut parfaitement maîtriser. On pourra comparer cet exemple à deux groupes avec les considérations développées à la section 1.2.2. On notera que le signe de chacun des termes de couplage correspond bien au dandysme.

Exercice 1.3 (Marcheurs dans une rue) *(difficile) Soit une rue $]-\infty,+\infty[$ dans laquelle marche une population. On distingue les marcheurs à droite, de densité $d^+(t,x)$, et les marcheurs à gauche, de densité $d^-(t,x)$ (t est le temps, x est la position). Tous marchent à la même vitesse, que l'on prend égale à 1 en valeur absolue. On suppose que l'envie de discuter avec un groupe le plus fourni possible amène les marcheurs à changer de direction avec un taux $\sigma > 0$ uniforme.*

1. *En déduire le modèle*

$$\begin{cases} \dfrac{\partial d^+}{\partial t}(t,x) + \dfrac{\partial d^+}{\partial x}(t,x) = \sigma(d^-(t,x) - d^+(t,x)), \\ \dfrac{\partial d^-}{\partial t}(t,x) - \dfrac{\partial d^-}{\partial x}(t,x) = \sigma(d^+(t,x) - d^-(t,x)), \end{cases}$$

Ecrire le système pour les variables $w = d^+ + d^-$ et $z = d^+ - d^-$ (système du télégraphe).

2. *Pour $\mu > 0$ fixé, on pose $\sigma = \frac{\mu}{2\varepsilon}$ avec $\varepsilon > 0$ petit. Pour étudier ce qui se passe aux temps petits (d'ordre ε), on procède au changement de variable $t \to t/\varepsilon$. Montrer que l'on obtient le système*

$$\begin{cases} \varepsilon \dfrac{\partial w}{\partial t}(t,x) + \dfrac{\partial z}{\partial x}(t,x) = 0, \\ \varepsilon \dfrac{\partial z}{\partial t}(t,x) + \dfrac{\partial w}{\partial x}(t,x) = -\dfrac{\mu}{\varepsilon} z(t,x). \end{cases} \tag{1.23}$$

3. *Montrer la stabilité dans $L^2(\mathbf{R})$ de la solution du système (1.23), c'est-à-dire que*

$$\frac{d}{dt}\left(\int_{\mathbf{R}} \left(w(t,x)^2 + z(t,x)^2 \right) dx \right) \leq 0.$$

On supposera que la solution tend vers 0 assez vite lorsque x tend vers l'infini.

4. *On cherche à simplifier ce système. On pose*

$$w = w_0 + \varepsilon w_1 + \cdots ,$$
$$z = z_0 + \varepsilon z_1 + \cdots .$$

Montrer que w_0 satisfait l'équation de diffusion

$$\frac{\partial w_0}{\partial t} - \frac{1}{\mu}\frac{\partial^2 w_0}{\partial x^2} = 0. \tag{1.24}$$

5. *Comparer avec ce qui se passe pour une population de neutrons.*
6. *(plus difficile) On cherche à justifier que la solution w est proche de w_0. Pour cela on pose*

$$\widetilde{w} = w_0 \text{ et } \widetilde{z} = -\frac{\varepsilon}{\mu}\frac{\partial w_0}{\partial x}.$$

Montrer que

$$\begin{cases} \varepsilon\dfrac{\partial \widetilde{w}}{\partial t}(t,x) + \dfrac{\partial \widetilde{z}}{\partial x}(t,x) = 0, \\ \\ \varepsilon\dfrac{\partial \widetilde{z}}{\partial t}(t,x) + \dfrac{\partial \widetilde{w}}{\partial x}(t,x) = -\dfrac{\mu}{\varepsilon}\widetilde{z}(t,x) + R(t,x), \end{cases}$$

avec

$$R = \varepsilon\frac{\partial \widetilde{z}}{\partial t}(t,x) = -\frac{\varepsilon^2}{\mu}\frac{\partial^2 w_0}{\partial t \partial x}\,.$$

7. *On suppose que w_0 et toutes ses dérivées sont bornées. Déduire de la stabilité établie à la question 3 que la différence $(w - w_0)$ est petite en norme L^2.*

Indications : le signe des termes de couplage est opposé à celui de ceux de l'exercice précédent, car il s'agit à présent d'une phénomène d'attraction et non plus de répulsion. Les questions 4, 5 et 6 sont un exemple d'approximation d'une équation de type transport par une équation de diffusion. La théorie générale est présentée rapidement à la section 1.2.3, puis développée au chapitre 4. On consultera en particulier le théorème 4.3.1.

Chapitre 2

L'équation de transport

L'opérateur de transport est l'objet mathématique commun à tous les modèles cinétiques présentés au début de ce cours — comme l'équation de Boltzmann linéaire de la neutronique, ou les équations du transfert radiatif. C'est également le prototype des équations aux dérivées partielles (EDP) linéaires du premier ordre.

Ce chapitre est consacré à une étude systématique du problème de Cauchy et du problème aux limites pour l'équation de transport libre.

Etant donné $v \in \mathbf{R}^N$, cette équation s'écrit

$$\frac{\partial f}{\partial t}(t,x) + v \cdot \nabla_x f(t,x) + a(t,x) f(t,x) = S(t,x)\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,,$$

où $f \equiv f(t,x)$ est la fonction inconnue, tandis que $a \equiv a(t,x)$ est une fonction donnée (il s'agit d'un taux d'amplification instantanée ou bien d'amortissement, selon son signe) ainsi que $S \equiv S(t,x)$ (qui est un terme source).

Remarquons que l'équation ci-dessus ne contient aucun terme qui modélise un processus d'échange entre les vitesses des particules, comme le terme intégral

$$\int_{\mathbf{R}^N} k(t,x,v,w) f(t,x,w) d\mu(w)$$

dans l'équation de Boltzmann linéaire, de sorte que, sans perte de généralité, l'on peut raisonner à $v \in \mathbf{R}^N$ fixé. C'est pourquoi, tout au long de ce chapitre, la vitesse v joue le rôle d'un simple paramètre dans l'équation de transport ci-dessus, et ne figure donc pas comme variable dans la fonction inconnue f.

2.1 Le problème de Cauchy

Le problème de Cauchy pour l'équation de transport ci-dessus consiste à se donner une valeur initiale $f^{in} \equiv f^{in}(x)$ de la densité f, et à chercher une solution

f vérifiant à la fois l'équation de transport et la condition initiale, c'est-à-dire

$$\begin{cases} \frac{\partial f}{\partial t}(t,x) + v \cdot \nabla_x f(t,x) + a(t,x) f(t,x) = S(t,x)\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,, \\ \\ f(0,x) = f^{in}(x)\,. \end{cases}$$

2.1.1 La méthode des caractéristiques

Il existe, pour résoudre les EDP d'ordre un, une méthode systématique, la méthode des caractéristiques, que nous allons présenter d'abord dans le cas où $a = S = 0$.

Autrement dit, on cherche $f \equiv f(t,x)$, solution du problème de Cauchy pour l'équation de transport libre

$$\begin{cases} \frac{\partial f}{\partial t}(t,x) + v \cdot \nabla_x f(t,x) = 0\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,, \\ \\ f(0,x) = f^{in}(x)\,. \end{cases}$$

Soit $y \in \mathbf{R}^N$; posons, pour tout $t \in \mathbf{R}$, $\gamma(t) = y + tv$; alors γ est une application de classe C^1 de $\mathbf{R}$ dans $\mathbf{R}^N$ vérifiant

$$\frac{d\gamma}{dt}(t) = v\,.$$

Définition 2.1.1 *L'ensemble* $\{(t,\gamma(t))\,|\,t \in \mathbf{R}\}$ *est une droite de* $\mathbf{R} \times \mathbf{R}^N$*, appelée "courbe caractéristique issue de* y *à* $t = 0$ *pour l'opérateur de transport* $\frac{\partial}{\partial t} + v \cdot \nabla_x$*".*

Soit $f \in C^1(\mathbf{R}_+ \times \mathbf{R}^N)$ solution de l'équation de transport. Donc l'application $t \mapsto f(t,\gamma(t))$ est de classe C^1 sur $\mathbf{R}_+$ (comme composée des applications f et $t \mapsto (t,\gamma(t))$, toutes deux de classe C^1), et on a

$$\begin{aligned} \frac{d}{dt} f(t,\gamma(t)) &= \frac{\partial f}{\partial t}(t,\gamma(t)) + \sum_{k=1}^N \frac{\partial f}{\partial x_k}(t,\gamma(t)) \frac{d\gamma_k}{dt}(t) \\ &= \frac{\partial f}{\partial t}(t,\gamma(t)) + \sum_{k=1}^N v_k \frac{\partial f}{\partial x_k}(t,\gamma(t)) \\ &= \left(\frac{\partial f}{\partial t} + v \cdot \nabla_x f \right)(t,\gamma(t)) = 0\,. \end{aligned}$$

Ainsi, toute solution de classe C^1 de l'équation de transport reste constante le long de chaque courbe caractéristique.

Théorème 2.1.2 *Soit* $f^{in} \in C^1(\mathbf{R}^N)$*. Le problème de Cauchy d'inconnue* f

$$\begin{cases} \frac{\partial f}{\partial t}(t,x) + v \cdot \nabla_x f(t,x) = 0\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,, \\ \\ f(0,x) = f^{in}(x)\,, \end{cases}$$

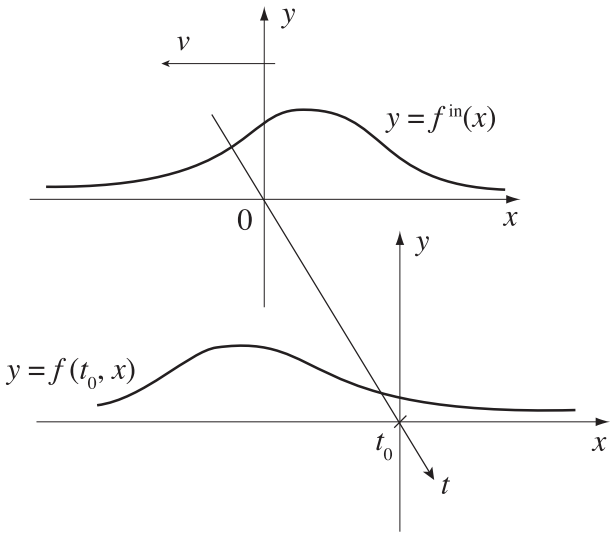

FIGURE 2.1 – Le graphe de la donnée initiale f^{in} est translaté de $t_0 v$ pour fournir le graphe de la fonction $x \mapsto f(t_0, x)$.

admet une unique solution $f \in C^1(\mathbf{R}_+ \times \mathbf{R}^N)$*, donnée par la formule*

$$f(t,x) = f^{in}(x - tv) \quad \textit{pour tout } (t,x) \in \mathbf{R}_+ \times \mathbf{R}^N .$$

Cette formule explicite pour la solution de l'équation de transport en justifie le nom : le graphe de la donnée initiale f^{in} est en effet transporté par la translation de vecteur tv.

Démonstration. Si f est une solution de classe C^1 de l'équation de transport, elle est constante le long des courbes caractéristiques, donc

$$f(t, y + tv) = f(0, y) = f^{in}(y)\,, \quad \text{pour tout } t > 0\,,\ y \in \mathbf{R}^N\,.$$

En posant $y + tv = x$, on trouve donc que

$$f(t,x) = f^{in}(x - tv)\,, \quad \text{pour tout } t > 0\,,\ x \in \mathbf{R}^N\,.$$

Réciproquement, pour $f^{in} \in C^1(\mathbf{R}^N)$, l'application $(t,x) \mapsto f^{in}(x - tv)$ est de classe C^1 sur $\mathbf{R} \times \mathbf{R}^N$ (comme composée des applications f^{in} et $(t,x) \mapsto x - tv$ qui sont toutes deux de classe C^1.) D'autre part on a

$$\nabla_x (f^{in}(x - tv)) = (\nabla f^{in})(x - tv)\,,$$

tandis que

$$\begin{aligned}\frac{\partial}{\partial t}(f^{in}(x - tv)) &= -\sum_{i=1}^{N} v_i \frac{\partial f}{\partial x_i}^{in}(x - tv) \\ &= -v \cdot (\nabla f^{in})(x - tv) = -v \cdot \nabla_x (f^{in}(x - tv))\,,\end{aligned}$$

ce qui montre que la fonction $f : (t,x) \mapsto f^{in}(x-tv)$ est bien une solution de l'équation de transport. ■

2.1.2 Problème de Cauchy avec terme source et amortissement

Appliquons maintenant la méthode des caractéristiques à la résolution de l'équation de transport avec coefficient d'amortissement ou d'amplification et second membre :

$$\begin{cases} \frac{\partial f}{\partial t}(t,x) + v \cdot \nabla_x f(t,x) + a(t,x) f(t,x) = S(t,x)\,, \quad x \in \mathbf{R}^N\,,\ t>0\,, \\ \\ f(0,x) = f^{in}(x)\,. \end{cases}$$

On supposera que a et S appartiennent à $C^1(\mathbf{R}_+ \times \mathbf{R}^N)$.

On considère à nouveau, pour tout $y \in \mathbf{R}^N$, la courbe caractéristique issue de y à $t=0$ pour l'opérateur de transport $\frac{\partial}{\partial t} + v \cdot \nabla_x$, c'est-à-dire

$$\{(t,\gamma(t))\,|\,t \geq 0\}\,, \quad \text{où } \gamma(t) = y + tv\,.$$

Si $f \in C^1(\mathbf{R}_+ \times \mathbf{R}^N)$ est solution de l'équation de transport ci-dessus, la fonction $t \mapsto f(t,\gamma(t))$ est de classe C^1 sur $\mathbf{R}_+$ et vérifie

$$\begin{aligned} \frac{d}{dt} f(t,\gamma(t)) &= \left(\frac{\partial f}{\partial t} + v \cdot \nabla_x f\right)(t,\gamma(t)) \\ &= S(t,\gamma(t)) - a(t,\gamma(t)) f(t,\gamma(t))\,, \quad t>0\,. \end{aligned}$$

On a ainsi ramené l'équation de transport, qui est une EDP, à une équation différentielle ordinaire de la forme

$$\begin{cases} u'(t) + \alpha(t) u(t) = \Sigma(t)\,, \quad t>0\,, \\ u(0) = u^{in}\,. \end{cases}$$

On sait que la solution de cette équation se calcule par la méthode "de variation de la constante" :

$$u(t) = u^{in} e^{-A(t)} + \int_0^t e^{-(A(t)-A(s))} \Sigma(s) ds\,, \quad \text{où } A(t) = \int_0^t \alpha(\tau) d\tau\,.$$

Appliquons cela à la solution f de l'équation de transport le long de la courbe caractéristique issue de y à $t=0$: on trouve que

$$\begin{aligned} f(t,y+tv) = {} & f^{in}(y) e^{-\int_0^t a(\tau,y+\tau v) d\tau} \\ & + \int_0^t e^{-\int_s^t a(\tau,y+\tau v) d\tau} S(s,y+sv) ds\,, \quad y \in \mathbf{R}^N\,,\ t \geq 0\,. \end{aligned}$$

Théorème 2.1.3 *Soient $f^{in} \in C^1(\mathbf{R}^N)$, ainsi que a et $S \in C^1(\mathbf{R}_+ \times \mathbf{R}^N)$. Le problème de Cauchy d'inconnue f*

$$\begin{cases} \frac{\partial f}{\partial t}(t,x) + v \cdot \nabla_x f(t,x) + a(t,x) f(t,x) = S(t,x)\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,, \\ \\ f(0,x) = f^{in}(x)\,, \end{cases}$$

admet une unique solution $f \in C^1(\mathbf{R}_+ \times \mathbf{R}^N)$, donnée par la formule

$$\begin{aligned} f(t,x) = {} & f^{in}(x - tv) e^{-\int_0^t a(\tau, x + (\tau - t)v) d\tau} \\ & + \int_0^t e^{-\int_s^t a(\tau, x + (\tau - t)v) d\tau} S(s, x + (s-t)v) ds\,, \end{aligned}$$

ou encore, de façon équivalente

$$\begin{aligned} f(t,x) = {} & f^{in}(x - tv) e^{-\int_0^t a(t - \tau, x - \tau v) d\tau} \\ & + \int_0^t e^{-\int_0^s a(t - \tau, x - \tau v) d\tau} S(t - s, x - sv) ds\,. \end{aligned} \tag{2.1}$$

Les deux formules du théorème ci-dessus sont un cas particulier d'une formule plus générale de la théorie des équations d'évolution, appelée formule de Duhamel — et qui fut d'ailleurs proposée initialement par Duhamel sur l'exemple de l'équation de la chaleur avec terme source. Sans entrer d'avantage dans les détails, disons que la formule de Duhamel constitue, pour la théorie des EDP linéaires d'évolution, l'analogue de la méthode de variation de la constante pour les équations différentielles ordinaires. C'est d'ailleurs en se ramenant à la méthode de variation de la constante le long des courbes caractéristiques de l'opérateur de transport $\frac{\partial}{\partial t} + v \cdot \nabla_x$ que l'on est arrivé aux deux formules du théorème.

Démonstration. En raisonnant par condition nécessaire, on a vu que, si f est une solution de classe C^1 sur $\mathbf{R}_+ \times \mathbf{R}^N$ du problème de Cauchy pour l'équation de transport, alors, pour tout $y \in \mathbf{R}^N$, l'on a

$$f(t, y + tv) = f^{in}(y) e^{-\int_0^t a(\tau, y + \tau v) d\tau} + \int_0^t e^{-\int_s^t a(\tau, y + \tau v) d\tau} S(s, y + sv) ds\,.$$

La première formule du théorème s'obtient en faisant $y = x - tv$ dans l'égalité ci-dessus.

La seconde formule du théorème s'obtient en faisant le changement de variables $\tau \mapsto t - \tau$ dans l'intégrale

$$\int_0^t a(\tau, x + (\tau - t)v) d\tau\,,$$

ainsi que le changement de variables $s \mapsto t - s$ dans l'intégrale

$$\int_0^t e^{-\int_s^t a(\tau, x + (\tau - t)v) d\tau} S(s, x + (s-t)v) ds\,.$$

Ceci démontre l'unicité dans $C^1(\mathbf{R}_+ \times \mathbf{R}^N)$ de la solution de l'équation de transport, et que cette solution, si elle existe, est donnée par les deux formules équivalentes du théorème.

Enfin, comme a et S sont de classe C^1 sur $\mathbf{R}_+ \times \mathbf{R}^N$, on démontre en dérivant sous le signe somme que la fonction f définie par l'une ou l'autre des formules du théorème est de classe C^1 sur $\mathbf{R}_+ \times \mathbf{R}^N$.

Il ne reste plus qu'à vérifier que le membre de droite de la première formule (par exemple) définit bien une solution de l'équation de transport. Revenant à la variable $y = x - tv$ dans la première formule du théorème, on obtient

$$f(t, y + tv) = f^{in}(y) e^{-\int_0^t a(\tau, y+\tau v) d\tau} + \int_0^t e^{-\int_s^t a(\tau, y+\tau v) d\tau} S(s, y + sv) ds \, .$$

On reconnaît dans le membre de droite de cette égalité la solution de l'équation différentielle ordinaire

$$\begin{cases} \frac{d}{dt} f(t, y + tv) + a(t, y + tv) f(t, y + tv) = S(t, y + tv) \, , & t > 0 \, , \\ \\ f(0, y) = f^{in}(y) \, . \end{cases}$$

Or, sachant que $f \in C^1(\mathbf{R}_+ \times \mathbf{R}^N)$, dire que la fonction $t \mapsto f(t, y + tv)$ est solution de l'équation différentielle ordinaire ci-dessus équivaut à dire que $f(t, x)$ est solution de l'équation de transport, puisque

$$\frac{d}{dt} f(t, y + tv) = \left(\frac{\partial f}{\partial t} + v \cdot \nabla_x f \right) (t, y + tv) \, .$$

Autrement dit, le membre de droite de la première formule du théorème définit bien une solution de classe C^1 sur $\mathbf{R}_+ \times \mathbf{R}^N$ du problème de Cauchy pour l'équation de transport. ■

2.2 Le problème aux limites

Dans la plupart des situations concrètes mettant en jeu des modèles cinétiques, on doit considérer des populations de particules qui ne sont pas libres de se déplacer dans tout l'espace, mais sont au contraire confinées dans une enceinte — comme les neutrons dans un réacteur nucléaire.

Il se peut également — par exemple pour des raisons liées au choix de certaines méthodes de résolution numérique — que l'on ait à considérer un sous-domaine de la région où se trouve la population de particules que l'on veut étudier.

Dans tous les cas, on est donc amené à résoudre une équation de transport dans un domaine de $\mathbf{R}^N$. Si l'on se souvient de la signification concrète de l'équation de Boltzmann linéaire, présentée au chapitre 1, qui consiste en un bilan du nombre de particules présentes à chaque instant dans tout volume infinitésimal de l'espace des phases, on conçoit aisément que la résolution d'une

équation de transport dans un domaine de $\mathbf{R}^N$ ne peut se faire sans information sur la densité de particules au bord de ce domaine.

Il est donc naturel d'étudier de manière systématique le problème aux limites pour l'équation de transport. Il faut en particulier

•préciser la nature des données au bord du domaine permettant de résoudre le problème aux limites de manière unique, et

•étudier la régularité de la solution ainsi obtenue.

2.2.1 Le problème aux limites monodimensionnel

Comme la méthode des caractéristiques réduit l'étude de l'équation de transport à celle d'une équation différentielle ordinaire le long des caractéristiques de l'opérateur de transport, il est naturel de commencer par traiter le cas de la dimension 1 d'espace. D'ailleurs, comme on va le voir, ce cas particulier contient l'essentiel des difficultés à résoudre.

Soient donc $x_L < x_R \in \mathbf{R}$. Dans tout ce paragraphe, on considèrera des équations de transport posées sur le domaine spatial $]x_L, x_R[$.

Théorème 2.2.1 *Supposons que $v > 0$. Soient donc $f_0 \in C^1([x_L, x_R])$, une densité initiale, et une donnée au bord $f_L \in C^1(\mathbf{R}_+)$. Le problème aux limites*

$$\begin{cases} \left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial x}\right) f(t,x) = 0\,, & x_L < x < x_R\,,\ t > 0\,, \\ f(t, x_L) = f_L(t)\,, \\ f(0,x) = f_0(x)\,, \end{cases}$$

admet une unique solution de classe C^1 sur $\mathbf{R}_+ \times [x_L, x_R]$, donnée par

$$f(t,x) = \begin{cases} f_0(x - tv) & \text{si } x - tv > x_L\,, \\ f_L(t - \frac{x - x_L}{v}) & \text{si } x - tv < x_L\,, \end{cases}$$

si et seulement si

$$f_L(0) = f_0(x_L) \quad \text{et} \quad f_L'(0) + v f_0'(x_L) = 0\,.$$

Cet énoncé appelle plusieurs commentaires importants pour la compréhension de la suite du cours, commentaires que nous donnons ici avant de procéder à la démonstration du théorème ci-dessus.

Tout d'abord, si $v < 0$, c'est le problème

$$\begin{cases} \left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial x}\right) f(t,x) = 0\,, & x_L < x < x_R\,,\ t > 0\,, \\ f(t, x_R) = f_R(t)\,, \\ f(0,x) = f_0(x)\,, \end{cases}$$

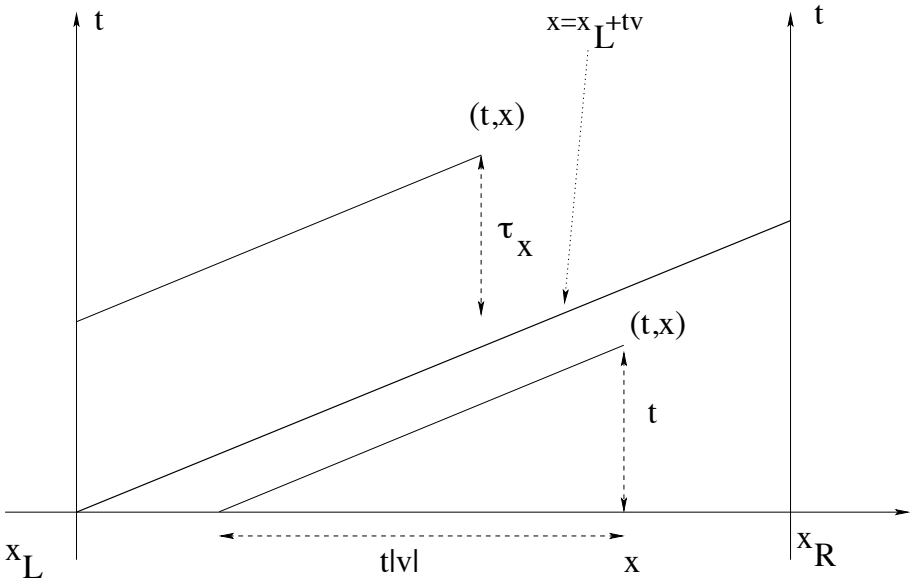

FIGURE 2.2 – Illustration graphique du Théorème 2.2.1, cas d'une vitesse $v > 0$. Au dessus de la droite d'équation $x = x_L + tv$, la solution est donnée par la formule $f(t,x) = f_L(t - \tau_x)$; en dessous de cette même droite, la solution est donnée par la formule $f(t,x) = f_0(x - tv)$. Les deux conditions de compatibilité traduisent la continuité de f et de ses dérivées partielles d'ordre 1 aux points de la droite d'équation $x = x_L + tv$ appartenant au domaine $\mathbf{R}_+ \times [x_L, x_R]$.

qui admet une unique solution de classe C^1 sur $\mathbf{R}_+ \times [x_L, x_R]$ si et seulement si $f_0 \in C^1([x_L, x_R])$ et $f_R \in C^1(\mathbf{R}_+)$ vérifient les conditions de compatibilité

$$f_0(x_R) = f_R(0) \quad \text{et} \quad f_R'(0) + v f_0'(x_R) = 0\,.$$

Supposons au contraire que $v > 0$, et considérons le même problème aux limites

$$\begin{cases} \left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial x}\right) f(t,x) = 0\,, & x_L < x < x_R\,,\ t > 0\,, \\ f(t, x_R) = f_R(t)\,, \\ f(0,x) = f_0(x)\,. \end{cases}$$

Si $f \in C^1(\mathbf{R}_+ \times [x_L, x_R])$ est solution de l'équation de transport, on déduit de la méthode des caractéristiques que, pour tout $t \geq 0$ et $x \in [x_L, x_R]$, la fonction

$$s \mapsto f(t+s, x+sv)$$

est constante pour $s \in \mathbf{R}$ tel que $(t+s, x+sv) \in \mathbf{R}_+ \times [x_L, x_R]$. Choisissons $s = \frac{x_R - x}{v}$; on trouve alors que

$$f(t,x) = f\left(t + \frac{x_R - x}{v}, x_R\right) = f_R\left(t + \frac{x_R - x}{v}\right)\,.$$

Cette formule montre que la condition initiale f_0 ne joue ici aucun rôle pour déterminer de manière unique la solution f du problème aux limites ; la condition au bord f_R y suffit à elle seule.

En faisant $t = 0$ et $x = x_R - \tau v$ dans l'égalité ci-dessus, on voit d'ailleurs que ce problème aux limites n'admet de solution de classe C^1 que si f_0 et f_R vérifient la relation de compatibilité

$$f_R(\tau) = f_0(x_R - \tau v) \text{ pour } 0 < \tau < \frac{x_R - x_L}{v}\,.$$

Autrement dit, la condition au bord f_R détermine complètement et de façon unique la donnée initiale f_0 si l'on veut que ce problème aux limites ait une solution de classe C^1.

La différence entre le problème ci-dessus et celui de l'énoncé du théorème est que, lorsque $v > 0$, la donnée au bord f_L représente la densité de particules **entrant** dans le domaine, tandis que la donnée au bord f_R représente la densité de particules **sortant** du domaine.

On peut résumer cette étude comme suit : pour l'équation de transport, le problème aux limites est bien posé dans le futur, c'est-à-dire pour $t > 0$, si l'on se donne

•la condition initiale, et

•la densité de particules **entrantes** sur le bord du domaine.

Démonstration du Théorème 2.2.1. Procédons d'abord par condition nécessaire.

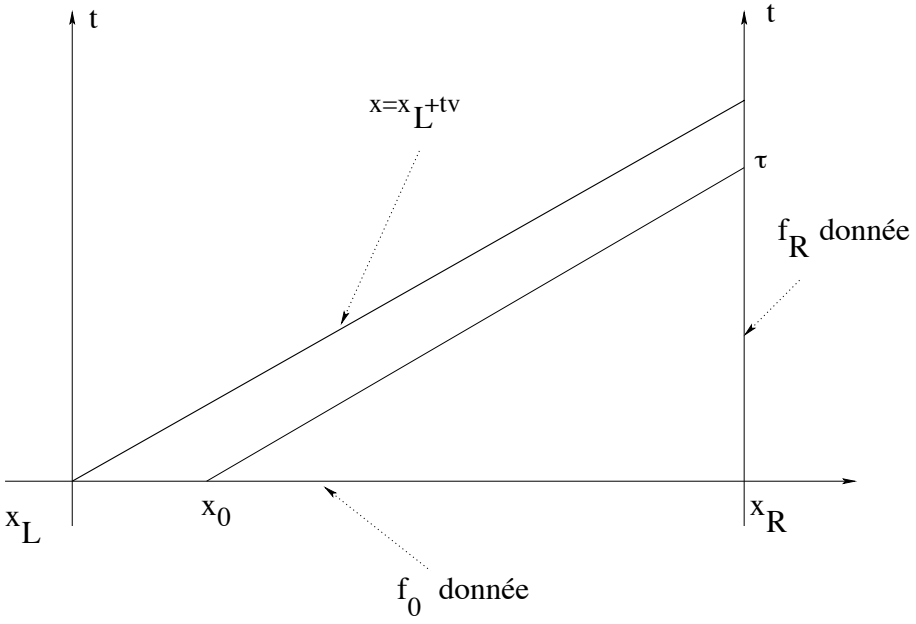

FIGURE 2.3 – Cas où $v > 0$. Comme la solution f de l'équation de transport libre est constante le long des caractéristiques, on a $f_0(x_0) = f_R(\tau)$. La donnée au bord sortant f_R restreinte à l'intervalle $[0, \frac{x_R - x_L}{v}]$ détermine donc complètement la donnée initiale.

Si f est une solution du problème aux limites de classe C^1 sur $\mathbf{R}_+ \times [x_L, x_R]$, alors f est constante sur l'intersection de chaque droite caractéristique avec $\mathbf{R}_+ \times [x_L, x_R]$. Autrement dit, pour tout $y \in \mathbf{R}$, on a

$$\frac{d}{dt} f(t, y+tv) = \left(\frac{\partial f}{\partial t} + v \frac{\partial f}{\partial x} \right)(t, y+tv) = 0\,, \quad x_L < y+tv < x_R\,,\ t > 0\,,$$

de sorte que

$$f(t, y+tv) = \text{Const.}\,, \quad x_L \leq y+tv \leq x_R\,,\ t \geq 0\,.$$

En particulier

$$f(t, y+tv) = f_0(y)\,, \quad x_L + tv \leq y+tv \leq x_R\,,\ t \geq 0\,,$$

ce qui, grâce au changement de variables $y + tv = x$, s'écrit encore

$$f(t, x) = f_0(x - tv)\,, \quad x_L + tv \leq x \leq x_R\,,\ t \geq 0\,.$$

Cette condition définit f sur le triangle

$$\left\{ (t, x) \in \mathbf{R}_+ \times [x_L, x_R] \,|\, 0 \leq t \leq \frac{x - x_L}{v} \right\}.$$

D'autre part, pour tout $\tau \geq 0$,

$$\frac{d}{ds} f(\tau+s, x_L+sv) = \left(\frac{\partial f}{\partial t} + v \frac{\partial f}{\partial x} \right)(\tau+s, x_L+sv) = 0\,, \quad x_L+sv < x_R\,,\ s > 0\,,$$

de sorte que

$$f(\tau + s, x_L + sv) = f_L(\tau)\,, \quad s \geq 0\,,\ x_L + sv \leq x_R\,.$$

Grâce au changement de variables $\tau + s = t$, $x_L + sv = x$, ceci s'écrit encore

$$f(t, x) = f_L\left(t - \frac{x - x_L}{v} \right), \quad x_L \leq x \leq x_R\,,\ \frac{x - x_L}{v} \leq t\,.$$

Donc, si $f \in C^1(\mathbf{R}_+ \times [x_L, x_R])$ est une solution du problème aux limites, f est forcément donnée par la formule du théorème. D'autre part, en posant $t = 0$ et $x = x_L$, on trouve que

$$f(0, x_L) = f_L(0) = f_0(x_L)\,,$$

ce qui est la première condition de compatibilité entre donnée initiale et donnée au bord.

La condition $f \in C^1(\mathbf{R}_+ \times [x_L, x_R])$ signifie qu'il existe une fonction $\tilde{f}$ de classe C^1 sur un voisinage ouvert de $\mathbf{R}_+ \times [x_L, x_R]$ telle que

$$\tilde{f}\big|_{\mathbf{R}_+ \times [x_L, x_R]} = f\,.$$

Par hypothèse

$$0 = \frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} = \frac{\partial \tilde{f}}{\partial t} + v\frac{\partial \tilde{f}}{\partial x} \text{ sur } \mathbf{R}_+^* \times]x_L, x_R[\,.$$

Comme $\tilde{f}$ est de classe C^1 sur un voisinage ouvert de $\mathbf{R}_+ \times [x_L, x_R]$,

$$\frac{\partial \tilde{f}}{\partial t}(t,x) \to f_L'(0) \text{ et } \frac{\partial \tilde{f}}{\partial x}(t,x) \to f_0'(x_L) \text{ lorsque } t \to 0^+ \text{ et } x \to x_L^+\,.$$

En passant à la limite dans l'égalité ci-dessus, on trouve que

$$f_L'(0) + v f_0'(x_L) = 0\,,$$

ce qui est la seconde condition de compatibilité.

Réciproquement, montrons que si f_0 et f_L vérifient les deux conditions de compatibilité, la fonction f définie par la formule du théorème est bien solution de classe C^1 du problème aux limites.

La formule du théorème montre que la fonction f est de classe C^1 sur chacun des domaines

$$\mathcal{T}^- = \left\{ (t,x) \in \mathbf{R}_+ \times [x_L, x_R] \,|\, 0 \le t < \frac{x - x_L}{v} \right\}$$

et

$$\mathcal{T}^+ = \left\{ (t,x) \in \mathbf{R}_+ \times [x_L, x_R] \,|\, t > \frac{x - x_L}{v} \right\}\,.$$

La première condition de compatibilité implique que la fonction f est continue en tout point du segment

$$\mathcal{S} = \left\{ \left(\frac{x - x_L}{v}, x \right) \,\Big|\, x_L \le x \le x_R \right\}\,,$$

de sorte que f est continue sur $\mathbf{R}_+ \times [x_L, x_R]$. Pour montrer que f est de classe C^1 sur $\mathbf{R}_+ \times [x_L, x_R]$, il suffit de montrer que les dérivées de $f\big|_{\mathcal{T}^-}$ et $f\big|_{\mathcal{T}^+}$ dans la direction orthogonale au segment $\mathcal{S}$ se recollent sur $\mathcal{S}$, puisque $f\big|_{\mathcal{T}^-}$ et $f\big|_{\mathcal{T}^+}$ sont constantes dans la direction de $\mathcal{S}$. Un vecteur orthogonal à $\mathcal{S}$ est $(-v, 1)$; d'autre part

$$\begin{aligned}
\nabla_{t,x} f\big|_{\mathcal{T}^-}(t,x) &= (-v f_0'(x - tv), f_0'(x - tv))\,, && (t,x) \in \mathcal{T}^-\,,\\
\nabla_{t,x} f\big|_{\mathcal{T}^+}(t,x) &= (f_L'(t - \tfrac{x - x_L}{v}), -\tfrac{1}{v} f_L'(t - \tfrac{x - x_L}{v}))\,, && (t,x) \in \mathcal{T}^+\,.
\end{aligned}$$

Donc

$$\begin{aligned}
(-v, 1) \cdot \nabla_{t,x} f\big|_{\mathcal{T}^-}(t,x) &= f_0'(x - tv)(1 + v^2)\,, && (t,x) \in \mathcal{T}^-\,,\\
(-v, 1) \cdot \nabla_{t,x} f\big|_{\mathcal{T}^+}(t,x) &= -f_L'(t - \tfrac{x - x_L}{v})(v + \tfrac{1}{v})\,, && (t,x) \in \mathcal{T}^+\,,
\end{aligned}$$

se recollent sur $\mathcal{S}$ si et seulement si

$$x - x_L = tv \Rightarrow f_0'(x - tv)(1 + v^2) = -f_L'(t - \tfrac{x - x_L}{v})\tfrac{1 + v^2}{v}\,.$$

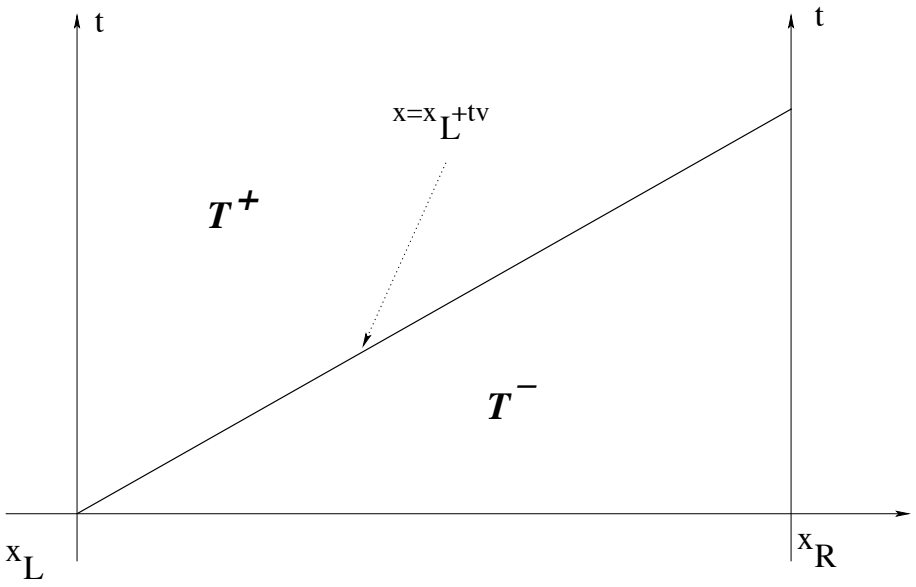

FIGURE 2.4 – Cas où $v > 0$. Les domaines $\mathcal{T}^+$ et $\mathcal{T}^-$.

Or ceci découle précisément de la condition

$$f_L'(0) + v f_0'(x_L) = 0\,.$$

Par conséquent, si les deux conditions de compatibilité portant sur f_0 et f_L sont satisfaites, la fonction f définie par la formule du théorème est bien de classe C^1 sur $\mathbf{R}_+ \times [x_L, x_R]$.

D'autre part, cette formule montre que f est constante le long des caractéristiques de l'opérateur de transport dans $\mathcal{T}^- \cup \mathcal{T}^+$. Cette fonction f vérifie donc l'équation de transport libre dans $(\mathbf{R}_+^* \times]x_L, x_R[) \setminus \mathcal{S}$, et donc dans $\mathbf{R}_+^* \times]x_L, x_R[$ puisque f est de classe C^1 sur $\mathbf{R}_+ \times [x_L, x_R]$.

Enfin, f vérifie évidement les conditions aux limites au bord de $\mathbf{R}_+ \times [x_L, x_R]$ pour $t = 0$ et pour $x = x_L$. ■

2.2.2 Le problème aux limites en dimension quelconque

Soient Ω ouvert à bord [1] de classe C^1 de $\mathbf{R}^N$ et $v \in \mathbf{R}^N \setminus \{0\}$. On notera dans tout ce qui suit n_x le vecteur unitaire normal au bord $\partial\Omega$ au point $x \in \partial\Omega$

1. On dit que Ω est un ouvert à bord de classe C^1 de $\mathbf{R}^N$ si tout point de $\overline{\Omega}$ admet un voisinage ouvert (pour la topologie induite) qui soit C^1-difféomorphe à un ouvert (toujours pour la topologie induite) d'un demi-espace fermé de $\mathbf{R}^N$ — autrement dit de $\{x \in \mathbf{R}^N \text{ t.q. } x_1 \geq 0\}$. De façon équivalente, Ω est un ouvert à bord de classe C^1 de $\mathbf{R}^N$ si (a) Ω est un ouvert de $\mathbf{R}^N$, (b) la frontière $\partial\Omega$ de Ω est une sous-variété de classe C^1 de $\mathbf{R}^N$ de dimension $N-1$ (autrement dit une courbe si $N = 2$ ou une surface si $N = 3$), et (c) l'ouvert Ω est situé localement d'un seul côté de $\partial\Omega$. Voici comment l'on traduit mathématiquement la condition (c) : pour tout $x_0 \in \partial\Omega$, il existe $\epsilon > 0$ et une fonction $\phi \in C^1(B(0,\epsilon))$ telle que $\nabla\phi(x) \neq 0$

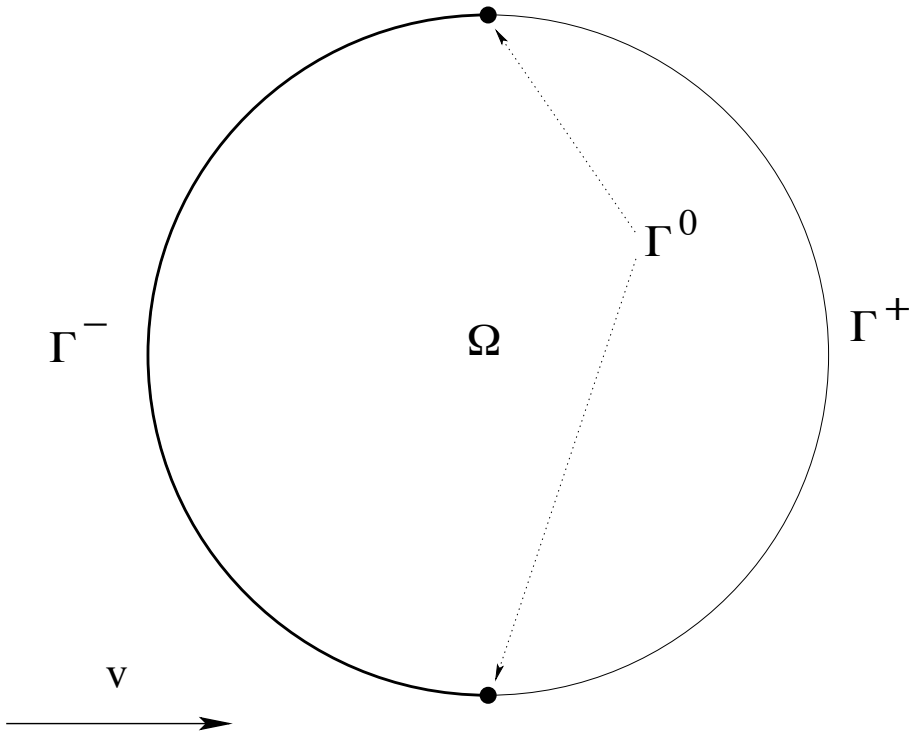

FIGURE 2.5 – Les parties entrante $\Gamma^- = \partial\Omega^-$, sortante $\Gamma^+ = \partial\Omega^+$ et caractéristique $\Gamma^0 = \partial\Omega^0$ du bord de Ω.

dirigé vers l'extérieur de Ω, ainsi que

$$\begin{aligned} \partial\Omega^- &= \{x \in \partial\Omega \,|\, v \cdot n_x < 0\} \qquad \text{(“bord entrant”)}, \\ \partial\Omega^0 &= \{x \in \partial\Omega \,|\, v \cdot n_x = 0\} \qquad \text{(“bord caractéristique”)}, \\ \partial\Omega^+ &= \{x \in \partial\Omega \,|\, v \cdot n_x > 0\} \qquad \text{(“bord sortant”)}. \end{aligned}$$

---

pour tout $x \in B(0,\epsilon)$, et

$$x \in \Omega \cap B(x_0,\epsilon) \Leftrightarrow \phi(x) > 0\,, \quad x \in \partial\Omega \cap B(x_0,\epsilon) \Leftrightarrow \phi(x) = 0\,.$$

(Quitte à diminuer $\epsilon > 0$ si besoin, il suffit de supposer que $\nabla\phi(x_0) \neq 0$, puisque $\nabla\phi$ est continu sur $B(x_0,\epsilon)$.) Pour tout $x \in \partial\Omega \cap B(x_0,\epsilon)$, les vecteurs $\pm\nabla_x\phi(x)$ sont orthogonaux à (l'espace tangent à) $\partial\Omega$ au point x. Comme ϕ est identiquement nulle sur $\partial\Omega \cap B(x_0,\epsilon)$ et strictement positive sur $\Omega \cap B(x_0,\epsilon)$, et que $\nabla\phi$ est dirigé dans le sens des valeurs croissantes de ϕ, il s'ensuit que $\nabla\phi(x)$ est dirigé vers Ω pour tout $x \in \partial\Omega \cap B(x_0,\epsilon)$. Par conséquent, le vecteur unitaire normal à $\partial\Omega$ au point x dirigé vers l'extérieur de Ω est

$$n_x = -\frac{\nabla\phi(x)}{|\nabla\phi(x)|}\,.$$

Exemple : Lorsque $N = 1$,

$$\Omega =]x_L, x_R[\quad \Rightarrow \partial\Omega = \{x_L, x_R\} \text{ avec } n_{x_L} = -1 \text{ et } n_{x_R} = +1\,,$$

de sorte que

$$v > 0 \quad \Rightarrow \quad \partial\Omega^- = \{x_L\}\,.$$

Les termes de "bord entrant" et "bord sortant" sont expliqués par le lemme ci-dessous.

Lemme 2.2.2 *Soient Ω ouvert à bord de classe C^1 de $\mathbf{R}^N$ et soit $v \in \mathbf{R}^N \setminus \{0\}$. Pour tout $x \in \partial\Omega^-$ (resp. $x \in \partial\Omega^+$), il existe $\epsilon > 0$ tel que*

$$0 < t < \epsilon \Rightarrow x + tv \in \Omega \text{ (resp. } x + tv \notin \overline{\Omega}),$$

tandis que

$$-\epsilon < t < 0 \Rightarrow x + tv \notin \overline{\Omega} \text{ (resp. } x + tv \in \Omega).$$

Démonstration. Comme $x \in \partial\Omega$, il existe $\eta > 0$ et $\phi \in C^1(B(x,\eta))$ telle que

$$y \in B(x,\eta) \cap \Omega \Leftrightarrow \phi(y) > 0\,, \quad y \in B(x,\eta) \cap \partial\Omega \Leftrightarrow \phi(y) = 0\,.$$

Pour $t \in \mathbf{R}$ tel que $|t| < \eta/|v|$, on a

$$\phi(x + tv) = \phi(x) + t\nabla\phi(x) \cdot v + o(t) = -t|\nabla\phi(x)|v \cdot n_x + o(t)\,,$$

puisque

$$n_x = -\frac{\nabla\phi(x)}{|\nabla\phi(x)|}\,.$$

Donc, si $v \cdot n_x < 0$, la quantité $\phi(x + tv)$ est du signe de t pour $|t|$ assez petit. Autrement dit, il existe $\epsilon \in]0, \eta/|v|[$ tel que

$$0 < t < \epsilon \Rightarrow \phi(x + tv) > 0 \Leftrightarrow x + tv \in \Omega$$

tandis que

$$-\epsilon < t < 0 \Rightarrow \phi(x + tv) < 0 \Leftrightarrow x + tv \notin \overline{\Omega}\,.$$

Le cas où $x \in \partial\Omega^+$ se traite de même. ■

On aura besoin de la notion de **temps de sortie** (rétrograde) de l'ouvert Ω partant de $x \in \Omega$ à la vitesse v, noté τ_x. Il est défini comme suit [2]

$$\tau_x = \inf\{t \geq 0 \,|\, x - tv \notin \overline{\Omega}\}\,.$$

Exemple : Lorsque $N = 1$, $v > 0$ et $\Omega =]x_L, x_R[$, on a

$$\tau_x = \frac{x - x_L}{v}\,.$$

L'étude du problème aux limites pour l'équation de transport en dimension d'espace $N > 1$ est absolument identique au cas monodimensionnel étudié au paragraphe précédent. Les seules complications supplémentaires proviennent de la géométrie du domaine Ω, et sont donc toutes contenues dans la fonction temps de sortie $x \mapsto \tau_x$.

2. Il faut se souvenir de la convention habituelle $\inf \varnothing = +\infty$.

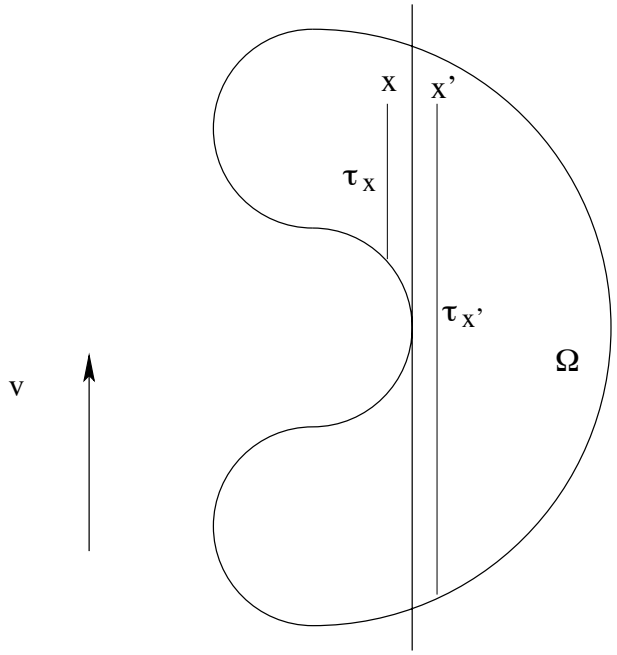

FIGURE 2.6 – Le temps de sortie est discontinu sur la trajectoire issue de la partie caractéristique du bord.

Lemme 2.2.3 *Soit Ω ouvert à bord de classe C^1 de $\mathbf{R}^N$ et $v \in \mathbf{R}^N \setminus \{0\}$. Pour tout $x \in \Omega$, l'on a $\tau_x > 0$. Si Ω est borné, la fonction $x \mapsto \tau_x$ est bornée sur Ω.*

Démonstration. Comme Ω est ouvert et $x \in \Omega$, il existe donc $\epsilon > 0$ tel que $B(x,\epsilon) \subset \Omega$. En notant $\tau_y^{B(x,\epsilon)}$ le temps de sortie de $B(x,\epsilon)$ en partant de y avec la vitesse v, on a donc

$$\tau_x \geq \tau_x^{B(x,\epsilon)} = \epsilon/|v| > 0$$

puisque la demi-droite issue de x dans la direction $-v$ sort de $B(x,\epsilon)$ avant de sortir de Ω.

Si Ω est borné, il existe $R > 0$ tel que $\Omega \subset B(0,R)$. Le même argument que ci-dessus montre que

$$\tau_x \leq \tau_x^{B(0,R)} \leq 2R/|v|$$

puisque la demi-droite issue de x dans la direction $-v$ sort de Ω avant de sortir de $B(0,R)$. Enfin, le temps de sortie d'une boule partant d'un point quelconque de cette boule est évidemment majoré par le diamètre de la boule divisé par la norme du vecteur vitesse. ■

Lorsque Ω n'est pas borné, il se peut évidemment que la fonction $x \mapsto \tau_x$ prenne la valeur $+\infty$.

Exemple : Choisissons $n \in \mathbf{R}^N$ tel que $|n| = 1$, et $v \in \mathbf{R}^N$ tel que $v \cdot n > 0$. Soit Ω le demi-espace $\Omega := \{x \in \mathbf{R}^N \,|\, x \cdot n < 0\}$. Pour tout $x \in \Omega$, l'on a

$$(x - tv) \cdot n = x \cdot n - tv \cdot n \leq x \cdot n < 0 \quad \text{pour tout } t \geq 0$$

de sorte que

$$\{t \geq 0 \,|\, x - tv \notin \overline{\Omega}\} = \varnothing \Rightarrow \tau_x = +\infty\,.$$

L'observation suivante, quoique très simple, est absolument fondamentale : lorsque Ω n'est pas convexe, même si Ω est un ouvert à bord de classe C^1, l'application

$$\Omega \ni x \mapsto \tau_x \in \mathbf{R}_+^* \cup \{+\infty\}$$

n'est pas forcément continue.

Plus précisément, lorsque Ω n'est pas convexe, il se peut que les trajectoires de particules issues du bord caractéristique $\partial\Omega^0$ passent dans Ω ; dans ce cas, le temps de sortie peut présenter des discontinuités sur de telles trajectoires (voir les Figures 2.6 et 2.7).

Après cette première remarque, étudions plus en détail la fonction $x \mapsto \tau_x$ dans le cas où Ω est convexe.

Lemme 2.2.4 *Soient Ω ouvert à bord de classe C^1 convexe de $\mathbf{R}^N$ ainsi qu'une vitesse $v \in \mathbf{R}^N \setminus \{0\}$. Alors*
(a) on a

$$(\partial\Omega^+ + \mathbf{R}_+^* v) \cap \Omega = (\partial\Omega^0 + \mathbf{R}_+^* v) \cap \Omega = \varnothing\,,$$

de sorte que

$$(\partial\Omega^- + \mathbf{R}_+^* v) \cap \Omega = (\partial\Omega + \mathbf{R}_+^* v) \cap \Omega\,;$$

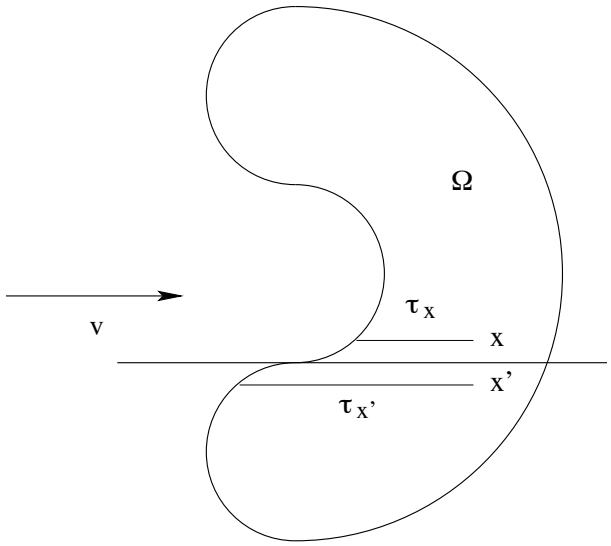

FIGURE 2.7 – Exemple de situation ou le temps de sortie n'est pas discontinu sur les trajectoires issues de la partie caractéristique du bord.

(b) pour tout $x \in \Omega$*,*

$$\tau_x < \infty \Leftrightarrow x \in (\partial\Omega^- + \mathbf{R}_+^* v) \cap \Omega \,.$$

Démonstration. Supposons qu'il existe un point $x^* \in \partial\Omega^+$ et un instant $t > 0$ tels que $x = x^* + tv \in \Omega$. D'après le Lemme 2.2.2, pour $s > 0$ assez petit, on a $x^* - sv \in \Omega$. Comme Ω est convexe, le segment $[x^* - sv, x^* + tv]$ est tout entier inclus dans Ω. En particulier, on a $x^* \in \Omega$, or ceci est impossible car $x^* \in \partial\Omega$. On vient donc de démontrer que

$$(\partial\Omega^+ + \mathbf{R}_+^* v) \cap \Omega = \varnothing \,.$$

Montrons que $(\partial\Omega^0 + \mathbf{R}_+^* v) \cap \Omega = \varnothing$. L'ouvert Ω étant convexe, il est tout entier situé du côté de l'espace tangent à $\partial\Omega$ au point x^* opposé au vecteur unitaire n_{x^*} :

$$\Omega \subset \{x \in \mathbf{R}^N \,|\, (x - x^*) \cdot n_{x^*} < 0\} \,.$$

Par conséquent, pour tout $x^* \in \partial\Omega^0$ et tout $t \in \mathbf{R}$, le point $x = x^* + tv$ n'appartient pas à Ω puisque

$$(x - x^*) \cdot n_{x^*} = tv \cdot n_{x^*} = 0 \,.$$

Ceci établit l'énoncé (a).

Soit $x \in \Omega$ tel que $\tau_x < \infty$. Le point $x^* = x - \tau_x v$ appartient donc à $\partial\Omega$, de sorte que $x = x^* + \tau_x v \in (\partial\Omega + \mathbf{R}_+^* v) \cap \Omega$ d'après le Lemme 2.2.3. D'après le (a), $x \in (\partial\Omega^- + \mathbf{R}_+^* v) \cap \Omega$.

Réciproquement, supposons que $x \in (\partial\Omega^- + \mathbf{R}_+^* v) \cap \Omega$. En particulier, il existe $x^* \in \partial\Omega^-$ et $s > 0$ tels que $x = x^* + sv$. Comme Ω est convexe, $\overline{\Omega}$ est également convexe, de sorte que le segment $[x^*, x]$ est contenu dans $\overline{\Omega}$. D'autre part, $x^* - tv \notin \overline{\Omega}$ pour tout $t > 0$. En effet, s'il existait $t^* > 0$ tel que $x^* - t^* v \in \overline{\Omega}$, on aurait $x^* - tv \in \overline{\Omega}$ pour tout $t > 0$ assez petit car $\overline{\Omega}$ est convexe. Or cela est impossible d'après le Lemme 2.2.3.

Par conséquent

$$\{t > 0 \,|\, x - tv \in \Omega\} =]0, s[$$

de sorte que $\tau_x = s < \infty$, ce qui démontre l'énoncé (b). ■

Après les différentes remarques ci-dessus portant sur le temps de sortie, nous en arrivons au point crucial pour la résolution du problème aux limites, à savoir la régularité de la fonction $x \mapsto \tau_x$.

Proposition 2.2.5 *Soient Ω ouvert à bord de classe C^1 convexe de $\mathbf{R}^N$ ainsi qu'une vitesse $v \in \mathbf{R}^N \setminus \{0\}$. Alors la fonction $x \mapsto \tau_x$ est de classe C^1 sur l'ouvert $\Omega' := \{x \in \Omega \,|\, \tau_x < +\infty\}$.*

Démonstration. La partie entrante $\partial\Omega^-$ du bord est ouverte dans $\partial\Omega$ comme image réciproque de l'intervalle ouvert $]-\infty, 0[$ par la fonction $x \mapsto v \cdot n_x$ qui est continue sur $\partial\Omega$.

L'application

$$\Phi : \partial\Omega^- \times \mathbf{R}_+^* \ni (y, t) \mapsto y + tv \in \mathbf{R}^N$$

est évidemment de classe C^1 et vérifie

$$\Phi(x_0^*, \tau_{x_0}) = x_0 \,.$$

Montrons que Φ est injective. Supposons que $\Phi(y, s) = \Phi(z, t)$ avec $t > s$. Comme y et z appartiennent à $\partial\Omega^-$, les points $y + \theta v$ et $z + \theta v$ appartiennent à Ω pour tout $\theta > 0$ assez petit. La condition $\Phi(y, s) = \Phi(z, t)$ montre que

$$\begin{aligned} z = y + \theta v + (s - t - \theta)v &= y + \theta v + \frac{s - t - \theta}{s - t}(z - y) \\ &= \frac{\theta}{s - t}(y + \theta v) + \frac{s - t - \theta}{s - t}(z + \theta v) \in \Omega \end{aligned}$$

par convexité de Ω.

De plus [3]

$$D\Phi(x_0^*, \tau_{x_0}) \cdot (u, s) = u + sv \,, \text{ pour tout } (u, s) \in T_{x_0^*}\partial\Omega \times \mathbf{R} \,.$$

Comme $v \cdot n_{x_0^*} < 0$ et $u \cdot n_{x_0^*} = 0$, la famille $\{u, v\}$ est libre, de sorte que

$$D\Phi(x_0^*, \tau_{x_0}) \cdot (u, s) = 0 \Rightarrow u = 0 \text{ et } s = 0 \,.$$

3. Lorsque Y est une sous-variété de classe au moins C^1 de $\mathbf{R}^N$, on note $T_y Y$ l'espace tangent à Y au point $y \in Y$.

Donc $D\Phi(x_0^*, \tau_{x_0})$ est un isomorphisme de $T_{x_0^*}\partial\Omega \times \mathbf{R}$ sur $\mathbf{R}^N$.

D'après le théorème d'inversion globale, $\partial\Omega^- + \mathbf{R}_+^* v = \Phi(\partial\Omega^- \times \mathbf{R}_+^*)$ est un ouvert de $\mathbf{R}^N$, et Φ est un C^1-difféomorphisme de $\partial\Omega^- \times \mathbf{R}_+^*$ sur son image.

D'après le Lemme 2.2.4, on a $\Omega' = \Omega \cap (\partial\Omega^- + \mathbf{R}_+^* v)$, ce qui montre que Ω' est ouvert dans $\mathbf{R}^N$ comme intersection de deux ouverts.

Enfin, pour tout $x \in \Omega'$, l'on a

$$\Phi^{-1}(x) = (x - \tau_x v, \tau_x)$$

de sorte que l'application $\Omega' \ni x \mapsto \tau_x \in \mathbf{R}_+^*$ est de classe C^1 comme composée de $\Phi^{-1}\big|_{\Omega'}$ et de la projection sur le second facteur dans le produit cartésien $\partial\Omega^- \times \mathbf{R}_+^*$. ■

Voici maintenant un premier résultat sur le problème aux limites pour l'équation de transport en dimension d'espace quelconque.

Théorème 2.2.6 *Supposons que Ω est un ouvert à bord de classe C^1 convexe de $\mathbf{R}^N$ avec $N \geq 2$. Soient $f_b^- \in C^1(\mathbf{R}_+ \times \partial\Omega^-)$ et $f^{in} \in C^1(\overline{\Omega})$. Le problème*

$$\begin{cases} (\frac{\partial}{\partial t} + v \cdot \nabla_x) f(t,x) = 0\,, \quad x \in \Omega\,,\ t > 0\,, \\ f\big|_{\partial\Omega^-} = f_b^-\,, \\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

admet une unique solution de classe C^1 sur $\mathbf{R}_+ \times \Omega$ et continue sur $\mathbf{R}_+ \times \overline{\Omega}$, donnée par la formule

$$f(t,x) = \begin{cases} f^{in}(x - tv) & \text{si } t \leq \tau_x\,, \\ f_b^-(t - \tau_x, x^*) & \text{si } t > \tau_x\,, \end{cases}$$

où on a noté $x^* := x - \tau_x v$, si et seulement si, pour tout $y \in \partial\Omega^-$, l'on a

$$f_b^-(0,y) = f^{in}(y)\,, \quad \frac{\partial f_b^-}{\partial t}(0,y) + v \cdot \nabla f^{in}(y) = 0\,.$$

Démonstration. La démonstration est essentiellement identique au cas monodimensionnel.

Supposons d'abord que $f \in C^1(\mathbf{R}_+ \times \Omega) \cap C(\mathbf{R}_+ \times \overline{\Omega})$ est solution du problème aux limites pour l'équation de transport. Appliquant la méthode des caractéristiques, on voit que

$$\frac{d}{ds} f(t-s, x-sv) = -\left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(t-s, x-sv) = 0\,, \quad 0 < s < \min(t, \tau_x)\,,$$

de sorte que

$$f(t-s, x-sv) = \text{Const.}$$

sur le segment $s \in [0, \min(t, \tau_x)]$. Par conséquent

$$t < \tau_x \quad \Rightarrow f(t,x) = f(0, x - tv) = f^{in}(x - tv)$$

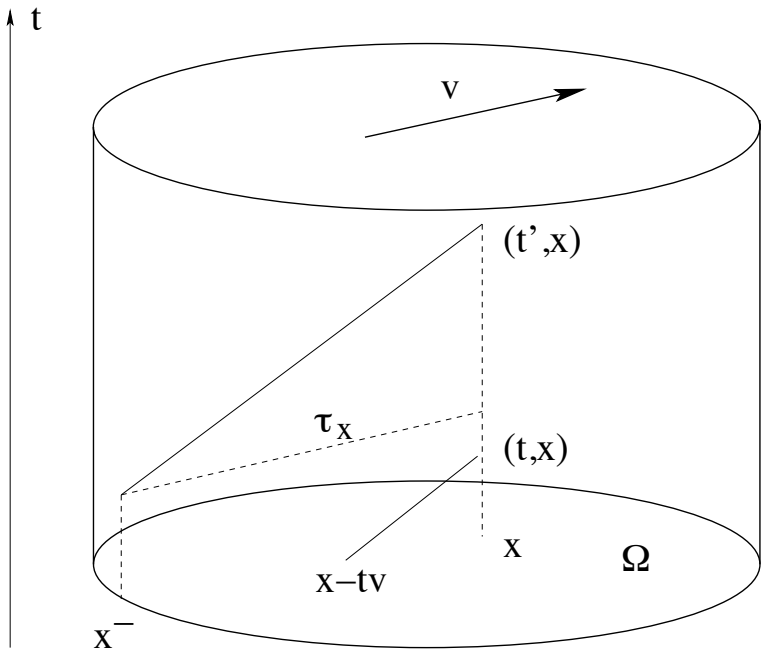

FIGURE 2.8 – La méthode des caractéristiques pour le problème aux limites. La figure représente le cas où $t < \tau_x$ et où $t' > \tau_x$.

puisque $x - tv \in \Omega$ si $0 < t < \tau_x$, tandis que

$$t > \tau_x \quad \Rightarrow f(t,x) = f(t-\tau_x, x-\tau_x v) = f_b^-(t-\tau_x, x^*)\,.$$

D'après le lemme ci-dessus, la fonction f est de classe C^1 sur $\mathcal{T}^+ \cup \mathcal{T}^-$, où

$$\mathcal{T}^+ = \{(t,x) \in \mathbf{R}_+ \times \Omega \,|\, t > \tau_x\}\,,$$
$$\mathcal{T}^- = \{(t,x) \in \mathbf{R}_+ \times \Omega \,|\, 0 \le t < \tau_x\}\,.$$

Cette fonction f est donc continue sur $\mathbf{R}_+ \times \overline{\Omega}$ si et seulement si elle est continue sur

$$\mathcal{S} = \{(t,x) \in \mathbf{R}_+ \times \Omega \,|\, t = \tau_x\}\,.$$

Or dire que f est continue sur $\mathcal{S}$ équivaut à dire que

$$\lim_{t\to\tau_x^+} f(t,x) = \lim_{t\to\tau_x^+} f_b^-(t-\tau_x, x^*) = \lim_{t\to\tau_x^-} f^{in}(x-tv) = \lim_{t\to\tau_x^-} f(t,x)$$

pour tout $x \in \Omega$, c'est-à-dire que $f_b(0,x^*) = f^{in}(x^*)$, soit

$$f_b^-(0,y) = f^{in}(y)\,, \quad \text{pour tout } y \in \partial\Omega^-\,.$$

De même, f est de classe C^1 sur $\mathbf{R}_+ \times \Omega$ si et seulement si

$$\lim_{t\to\tau_x^+} \nabla f(t,x) = \lim_{t\to\tau_x^-} \nabla f(t,x)\,, \quad x \in \Omega\,.$$

D'une part

$$\lim_{t\to\tau_x^-} \nabla f(t,x) = (-v\cdot\nabla f^{in}(x^*), \nabla f^{in}(x^*))\,.$$

D'autre part

$$\begin{aligned}\lim_{t\to\tau_x^+} \nabla f(t,x) &= \left(\frac{\partial f_b}{\partial t}^-(0,x^*), -\frac{\partial f_b}{\partial t}^-(0,x^*)\nabla\tau_x + (D_x x^*)^T \nabla_x f_b^-(0,x^*)\right)\\ &= \left(\frac{\partial f_b}{\partial t}^-(0,x^*), -\frac{\partial f_b}{\partial t}^-(0,x^*)\nabla\tau_x + (D_x x^*)^T \nabla f^{in}(x^*)\right)\end{aligned}$$

puisque $f_b^-(0,y) = f^{in}(y)$ pour tout $y \in \partial\Omega^-$ d'après la première relation de compatibilité. Ensuite

$$D_x x^* = I - v \otimes \nabla\tau_x\,,$$

de sorte que

$$\lim_{t\to\tau_x^+} \nabla f(t,x)$$
$$= \left(\frac{\partial f_b}{\partial t}^-(0,x^*), -\frac{\partial f_b}{\partial t}^-(0,x^*)\nabla\tau_x + \nabla f^{in}(x^*) - v\cdot\nabla f^{in}(x^*)\nabla\tau_x\right).$$

Donc

$$\lim_{t\to\tau_x^-} \nabla f(t,x) = \lim_{t\to\tau_x^+} \nabla f(t,x)$$

pour tout $x \in \Omega$ si et seulement si

$$-v \cdot \nabla f^{in}(x^*) = \frac{\partial f_b}{\partial t}^- (0, x^*)$$
$$\nabla f^{in}(x^*) = -\frac{\partial f_b}{\partial t}^- (0, x^*) \nabla \tau_x + \nabla f^{in}(x^*) - v \cdot \nabla f^{in}(x^*) \nabla \tau_x \,,$$

c'est-à-dire si et seulement si

$$\frac{\partial f_b^-}{\partial t}(0, y) + v \cdot \nabla f^{in}(y) = 0 \,, \quad \text{pour tout } y \in \partial\Omega^- \,,$$

ce qui est précisément la seconde condition de compatibilité.

Enfin, la formule de l'énoncé du théorème fournit bien une solution de l'équation de transport sur $\mathbf{R}_+^* \times \Omega \setminus \mathcal{S}$, puisque cette formule montre que f est constante le long des caractéristiques de l'opérateur de transport. Comme on sait que f est de classe C^1 sur $\mathbf{R}_+ \times \Omega$, la fonction f vérifie l'équation de transport sur $\mathbf{R}_+ \times \Omega$ tout entier. D'autre part, la même formule montre que la fonction f vérifie la condition initiale pour $t = 0$ et la condition aux limites sur le bord entrant $\partial\Omega^-$. ■

La formule du théorème ci-dessus montre que la solution f dépend de manière continue du temps de sortie τ_x lorsque $t > \tau_x$. Par conséquent, lorsque Ω n'est pas convexe, même lorsque la donnée initiale f^{in} et la donnée au bord f_b^- sont de classe C^1, et même si elles vérifient les conditions de compatibilité du théorème ci-dessus, la solution f peut hériter des discontinuités éventuelles du temps de sortie $x \mapsto \tau_x$.

On prendra bien garde au fait suivant : même lorsque Ω est convexe, et pour des données $f^{in} \in C^1(\overline{\Omega})$ et $f_b^- \in C^1(\mathbf{R}_+ \times \partial\Omega^-)$, la solution f du problème aux limites pour l'équation de transport définie par la formule du théorème ci-dessus n'est en général de classe C^1 que sur $\mathbf{R}_+ \times \Omega$, et non sur $\mathbf{R}_+ \times \overline{\Omega}$.

Contre-exemple : Prendre pour Ω le disque unité de $\mathbf{R}^2$ et $v = (1, 0)$, de sorte que $\tau_x < 2$ pour tout $x \in \Omega$, et $\partial\Omega^- = \{(\cos\theta, \sin\theta) \mid \frac{\pi}{2} < \theta < \frac{3\pi}{2}[\}$. Choisir

$$f^{in} \equiv 0 \quad \text{et } f_b^-(t, \cos\theta, \sin\theta) = \chi(t)\theta \,,$$

où $\chi \in C^\infty(\mathbf{R})$ vérifie

$$\chi \equiv 0 \text{ sur } [0, \tfrac{1}{2}] \text{ et } \chi \equiv 1 \text{ sur } [1, +\infty[\,.$$

Evidemment

$$f_b^-(t, \cdot)\big|_{\partial\Omega^-} = 0 \text{ pour } t \in [0, \tfrac{1}{2}] \text{ et } f^{in} = 0$$

de sorte que f_b^- et f^{in} vérifient les conditions de compatibilité du théorème ci-dessus.

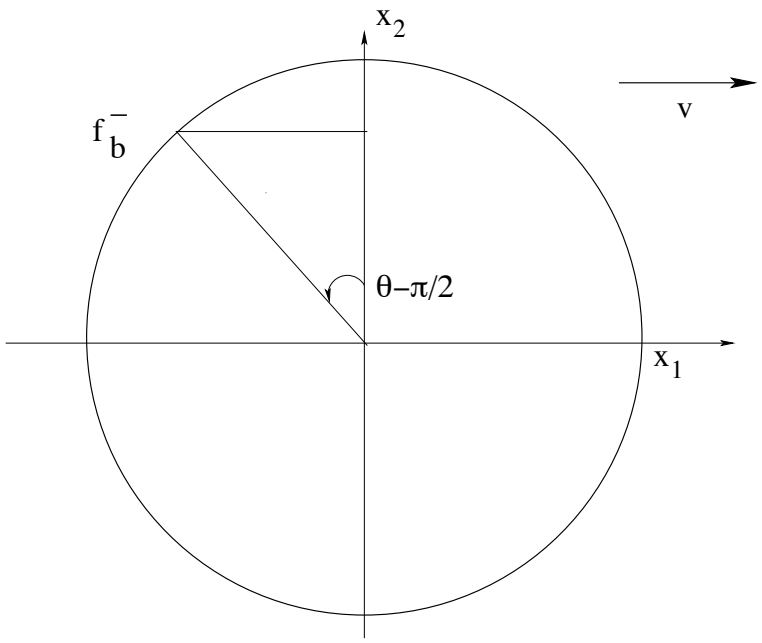

FIGURE 2.9 – Illustration graphique du contre-exemple. Supposons que $f_b^-(t, \cos\theta, \sin\theta) = \theta$ pour $t > 1$; on a alors $f(t, 0, \sin\theta) = \theta$ pour $t > 3$. La dérivée partielle $\frac{\partial f}{\partial x_2}$ n'admet pas de limite finie sur la partie caractéristique du bord.

Or un calcul simple basé sur la méthode des caractéristiques montre que la solution du problème aux limites

$$\begin{cases} \left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(t,x) = 0\,, \quad x \in \Omega\,,\ t > 0\,, \\ f\big|_{\partial\Omega^-} = f_b^-\,, \\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

vaut

$$f(t,0,x_2) = \tfrac{\pi}{2} + \arcsin x_2\,,\ \text{pour tout } x_2 \in]-1,1[\text{ et } t > 3\,.$$

Evidemment

$$\frac{\partial f}{\partial x_2}(t,0,x_2) = \frac{1}{\sqrt{1-x_2^2}}\,,$$

ce qui montre que $x_2 \mapsto f(t,0,x_2)$ n'est pas de classe C^1 sur $[-1,1]$, puisque $\frac{\partial f}{\partial x_2}(t,0,x_2) \to +\infty$ lorsque $|x_2| \to 1$. En particulier $f \notin C^1(\mathbf{R}_+ \times \overline{\Omega})$, bien que $f \in C^1(\mathbf{R}_+ \times \Omega)$.

2.2.3 Le problème aux limites avec absorption et terme source

Etendons le résultat du paragraphe précédent au cas de l'équation de transport avec terme d'amplification ou d'absorption et terme source.

Théorème 2.2.7 *Soient Ω ouvert convexe de $\mathbf{R}^N$ à bord de classe C^1, une vitesse $v \in \mathbf{R}^N \setminus \{0\}$ ainsi que des fonctions $f_b^- \in C^1(\mathbf{R}_+ \times \partial\Omega^-)$, $f^{in} \in C^1(\overline{\Omega})$ et $a, S \in C^1(\mathbf{R}_+ \times \overline{\Omega})$. Supposons que, pour tout $y \in \partial\Omega^-$,*

$$\begin{aligned} f_b^-(0,y) &= f^{in}(y)\,, \\ \frac{\partial f_b^-}{\partial t}(0,y) + v \cdot \nabla f^{in}(y) + a(0,y) f^{in}(y) &= S(0,y)\,. \end{aligned}$$

Le problème aux limites

$$\begin{cases} \left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(t,x) = -a(t,x) f(t,x) + S(t,x), \quad x \in \Omega,\, t > 0\,, \\ f\big|_{\partial\Omega^-} = f_b^-\,, \\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

admet une unique solution $f \in C^1(\mathbf{R}_+ \times \Omega) \cap C(\mathbf{R}_+ \times \overline{\Omega})$. Cette solution est

donnée par la formule

$$\begin{aligned} f(t,x) = {} & \mathbf{1}_{t\leq\tau_x} f^{in}(x-tv)\exp\left(-\int_0^t a(s,x+(s-t)v)ds\right) \\ & + \mathbf{1}_{t>\tau_x} f_b^-(t-\tau_x,x^*)\exp\left(-\int_{t-\tau_x}^t a(s,x+(s-t)v)ds\right) \\ & + \int_{(t-\tau_x)^+}^t \exp\left(-\int_s^t a(\tau,x+(\tau-t)v)ds\right) S(s,x+(s-t)v)ds\,, \end{aligned}$$

où on rappelle que $x^* = x-\tau_x v$, *et que* $z^+ = \max(z,0)$.

Remarque 2.2.8 *Dans le cas où* $\Omega = \mathbf{R}^N$, *le temps de sortie* $\tau_x = +\infty$, *de sorte que* $(t-\tau_x)^+ = 0$. *Cette formule redonne donc la solution du problème de Cauchy.*

Démonstration. La démonstration de ce résultat suit de très près celle de l'énoncé analogue dans le cas où $a = S = 0$.

Si $f \in C^1(\mathbf{R}_+\times\Omega)\cap C(\mathbf{R}_+\times\overline{\Omega})$ est solution du problème aux limites, on voit, en appliquant la méthode des caractéristiques, que

$$\begin{aligned} \frac{d}{ds}f(t-s,x-sv) = a(t-s,x-sv)f(t-s,x-sv) - S(t-s,x-sv)\,, \\ 0<s<\min(t,\tau_x)\,, \end{aligned}$$

ou encore, de manière équivalente

$$\begin{aligned} \frac{d}{ds}\left(f(t-s,x-sv)\exp\left(-\int_0^s a(t-\theta,x-\theta v)d\theta\right)\right) \\ = -S(t-s,x-sv)\exp\left(-\int_0^s a(t-\theta,x-\theta v)d\theta\right)\,, \quad 0<s<\min(t,\tau_x)\,. \end{aligned}$$

Intégrons chaque membre de cette égalité pour $0<s<\min(t,\tau_x)$: si $t<\tau_x$, on a

$$\begin{aligned} f(0,x-tv)\exp\left(-\int_0^t a(t-\theta,x-\theta v)d\theta\right) - f(t,x) \\ = -\int_0^t S(t-s,x-sv)\exp\left(-\int_0^s a(t-\theta,x-\theta v)d\theta\right)ds\,, \end{aligned}$$

ou encore

$$\begin{aligned} f(t,x) = {} & f^{in}(x-tv)\exp\left(-\int_0^t a(t-\theta,x-\theta v)d\theta\right) \\ & + \int_0^t S(t-s,x-sv)\exp\left(-\int_0^s a(t-\theta,x-\theta v)d\theta\right)ds\,. \end{aligned}$$

En faisant le changement de variables $s \mapsto t - s$ dans l'intégrale mettant en jeu le terme source, cette égalité se transforme en

$$f(t,x) = f^{in}(x - tv) \exp\left(- \int_0^t a(t-\theta, x - \theta v) d\theta \right) + \int_0^t S(s, x + (s-t)v) \exp\left(- \int_0^{t-s} a(t-\theta, x-\theta v) d\theta \right) ds \, ;$$

puis on fait le changement de variables $\theta \mapsto t - \theta$ dans le facteur d'amortissement, ce qui donne

$$f(t,x) = f^{in}(x - tv) \exp\left(- \int_0^t a(\theta, x + (\theta - t)v) d\theta \right) + \int_0^t S(s, x + (s-t)v) \exp\left(- \int_s^t a(\theta, x + (\theta - t)v) d\theta \right) ds$$

pour tout $x \in \Omega$ et tout $t \in]0, \tau_x[$.

Si au contraire $t > \tau_x$, on trouve de même que

$$f(t - \tau_x, x - \tau_x v) \exp\left(- \int_0^{\tau_x} a(t-\theta, x-\theta v) d\theta \right) - f(t,x) = - \int_0^{\tau_x} S(t-s, x - sv) \exp\left(- \int_0^s a(t-\theta, x-\theta v) d\theta \right) ds \, ,$$

ce qui donne

$$\begin{aligned} f(t,x) &= f(t - \tau_x, x - \tau_x v) \exp\left(- \int_0^{\tau_x} a(t-\theta, x-\theta v) d\theta \right) \\ &\quad + \int_0^{\tau_x} S(t-s, x-sv) \exp\left(- \int_0^s a(t-\theta, x-\theta v) d\theta \right) ds \\ &= f_b^-(t - \tau_x, x^*) \exp\left(- \int_0^{\tau_x} a(t-\theta, x-\theta v) d\theta \right) \\ &\quad + \int_{t-\tau_x}^t S(s, x + (s-t)v) \exp\left(- \int_0^{t-s} a(t-\theta, x-\theta v) d\theta \right) ds \\ &= f_b^-(t - \tau_x, x^*) \exp\left(- \int_{t-\tau_x}^t a(\theta, x + (\theta - t)v) d\theta \right) \\ &\quad + \int_{t-\tau_x}^t S(s, x + (s-t)v) \exp\left(- \int_s^t a(\theta, x + (\theta - t)v) d\theta \right) ds \, , \end{aligned}$$

pour $x \in \Omega$ et $t > \tau_x$, après les mêmes changements de variables que dans le cas où $t \in]0, \tau_x[$.

En regroupant les deux expressions ci-dessus pour $t \in]0, \tau_x[$ et pour $t > \tau_x$ on arrive à la formule de l'énoncé.

On a donc démontré que si $f \in C^1(\mathbf{R}_+ \times \Omega) \cap C(\mathbf{R}_+ \times \overline{\Omega})$ est solution du problème aux limites, elle est donnée par la formule de l'énoncé.

On laisse au lecteur le soin de vérifier

a) que les deux conditions de compatibilité de l'énoncé du théorème impliquent bien que la fonction f définie par la formule des caractéristiques est de classe C^1 sur $\mathbf{R}_+ \times \Omega$ et continue sur $\mathbf{R}_+ \times \overline{\Omega}$, et

b) que la formule de l'énoncé définit bien une solution du problème aux limites.

Ces deux vérifications suivent la démarche exposée dans la démonstration du Théorème 2.2.1. ■

Remarque 2.2.9 (Problème aux limites périodique) *L'étude de certains problèmes d'homogénéisation nécessite de résoudre des équations de transport dans la classe des fonctions périodiques en la variable de position. Voici comment formuler et résoudre ce type de problème.*

Soient $f^{in} \in C^1(\mathbf{R}^N)$ et $a, S \in C^1(\mathbf{R}_+ \times \mathbf{R})$; supposons les fonctions f^{in}, a et S périodiques de période L dans chacune des variables $x_1, \dots, x_N$ de position, c'est-à-dire que

$$f^{in}(x + Lk) = f^{in}(x)\,, \quad a(t, x + Lk) = a(t, x)\,, \quad S(t, x) = S(t, x + Lk)$$

pour tout $(t, x) \in \mathbf{R}_+ \times \mathbf{R}^N$ et tout $k \in \mathbf{Z}^N$.

Alors l'unique solution f du problème de Cauchy

$$\begin{cases} \left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(t, x) = -a(t, x) f(t, x) + S(t, x), \quad x \in \mathbf{R}^N,\, t > 0\,, \\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

est également périodique de période L dans chacune des variables $x_1, \dots, x_N$. Cela se voit évidemment sur la formule explicite donnant la solution f de ce problème de Cauchy. On peut également remarquer que, comme les fonctions f^{in}, a et S sont périodiques de période L dans chacune des variables $x_1, \dots, x_N$, pour tout $k \in \mathbf{Z}^N$, la fonction $f_k : (t, x) \mapsto f(t, x + Lk)$ est solution du problème de Cauchy ci-dessus, exactement comme la fonction f elle-même. Par unicité de la solution de ce problème de Cauchy, on en conclut que $f_k = f$ pour tout $k \in \mathbf{Z}^N$, ce qui équivaut à dire que la solution f est périodique de période L dans chacune des variables $x_1, \dots, x_N$.

Cette solution f vérifie donc

$$\begin{cases} \left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(t, x) = -a(t, x) f(t, x) + S(t, x), \quad x \in]0, L[^N,\, t > 0\,, \\ f(t, \hat{x}_j) = f(t, \hat{x}_j + L e_j)\,, \qquad\qquad x \in]0, L[^N,\, t > 0,\, 1 \le j \le N\,, \\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

où on a noté

$$\hat{x}_j = (x_1, \dots, x_{j-1}, 0, x_{j+1}, \dots, x_N) \text{ et } e_j = (\underbrace{0, \dots, 0}_{j-1}, 1, \underbrace{0, \dots, 0}_{N-j})\,.$$

Autrement dit, e_j est le j-ième vecteur de la base canonique de $\mathbf{R}^N$.

Ce nouveau problème peut être vu comme un problème aux limites, et les conditions

$$f(t,\hat{x}_j) = f(t,\hat{x}_j + Le_j)\,, \qquad\qquad x \in]0,L[^N,\, t>0,\, 1\le j\le N\,,$$

comme des conditions aux limites (ici sur chaque bord du cube $]0,L[^N$), conditions dites "de périodicité". Le cube $]0,L[^N$ est bien un ouvert convexe, mais son bord n'a pas la régularité C^1 que l'on a supposée jusqu'ici. Toutefois, cela n'est pas vraiment gênant, car le bord de $]0,L[^N$ ne joue pas un rôle important dans ce problème, comme on va le voir.

Evidemment, comme f est périodique de période L dans toutes les directions, il est équivalent de la connaître sur $\mathbf{R}^N$ et de connaître sa restriction à une période, par exemple à $[0,L[^N$. Les deux formulations ci-dessus sont donc équivalentes.

La première formulation ci-dessus du problème aux limites pour l'équation de transport avec conditions aux limites périodiques montre que la condition aux limites ne modifie pas la formule explicite donnant la solution, car celle-ci est la restriction à $x\in[0,L[^N$ de la solution du problème de Cauchy, contrairement au cas du problème aux limites étudié dans le Théorème 2.2.7.

La façon la plus commode de considérer le problème aux limites périodique ci-dessus consiste à observer qu'une fonction périodique de N variables $x_1,\dots,x_N$, de période L en chacune de ces variables, s'identifie naturellement à une fonction définie sur l'espace quotient $(\mathbf{R}/L\mathbf{Z})^N$ des N-uplets de réels modulo L. Cet espace est ce que l'on appelle un N-tore en géométrie ; on peut le voir comme le cube $[0,L]^N$ dont on aurait identifié les faces opposées — par exemple, pour $N=1$, il s'agit d'un cercle de longueur L ; pour $N=2$, cet espace ressemble à une bouée (d'où le nom de tore) dont le cercle interne serait de rayon nul.

Or un N-tore est une variété différentiable de classe C^∞ compacte et sans bord. On peut donc formuler le problème aux limites périodiques ci-dessus comme suit :

$$\begin{cases} \left(\frac{\partial}{\partial t} + v\cdot\nabla_x\right) f(t,x) = -a(t,x)f(t,x) + S(t,x), \quad x\in(\mathbf{R}/L\mathbf{Z})^N,\, t>0\,, \\ f\big|_{t=0} = f^{in}\,. \end{cases}$$

Observons que cette formulation sur le N-tore est celle d'un simple problème de Cauchy : il n'y a pas de conditions aux limites, puisque le N-tore est un variété sans bord. On peut résumer la situation en disant que le N-tore partage avec le cube $[0,L]^N$ l'avantage d'être compact, et avec l'espace entier $\mathbf{R}^N$ celui d'être sans bord. C'est pourquoi ce cadre est particulièrement commode pour l'analyse des équations aux dérivées partielles. Toutefois, comme les trois formulations ci-dessus du problème périodique sont strictement équivalentes, le lecteur peu familier avec les espaces quotients comme $(\mathbf{R}/L\mathbf{Z})^N$ est invité à utiliser indifféremment la première ou la seconde.

2.3 Solutions généralisées

On a vu dans la section précédente qu'en dehors du cas où le domaine spatial modélisant l'enceinte contenant les particules est convexe, la solution du problème aux limites n'est en général même pas continue, même si les données sont de classe C^1, et même si elles vérifient les conditions de compatibilité du théorème précédent. En pratique, on a pourtant souvent à considérer des équations de transport posées sur des domaines spatiaux non convexes.

Ceci suggère de généraliser la notion de solution de l'équation de transport, de façon à ce que certaines fonctions qui ne sont pas de classe C^1, ni même continues, puissent être considérées comme solutions de cette équation.

Une première possibilité consiste à utiliser la théorie des distributions ; toutefois, pour la commodité des lecteurs qui ne seraient pas familiers de cette théorie, nous choisirons une autre approche.

Avant de définir la notion de solution généralisée de l'équation de transport, nous allons considérer une situation simplifiée, quoiqu'exactement analogue.

Exemple. Considérons l'équation aux dérivées partielles

$$\frac{\partial f}{\partial x}(x,y) = 0\,, \quad (x,y) \in \mathbf{R}^2\,.$$

Dire que $f \in C^1(\mathbf{R}^2)$ est solution de cette équation équivaut à dire que

$$f(x,y) = C(y)\,, \quad C \in C^1(\mathbf{R})\,.$$

Cependant, il n'est pas nécessaire que la fonction f soit de classe C^1 sur $\mathbf{R}^2$, c'est-à-dire par rapport aux deux variables x et y pour que l'équation ci-dessus ait un sens. Il suffit pour cela que la fonction

$$x \mapsto f(x,y) \text{ soit de classe } C^1 \text{ sur } \mathbf{R} \text{ pour tout } y \in \mathbf{R}.$$

D'ailleurs, une telle fonction est solution de l'équation

$$\frac{\partial f}{\partial x}(x,y) = 0\,, \quad (x,y) \in \mathbf{R}^2$$

si et seulement si f est de la forme

$$f(x,y) = C(y)$$

où cette fois $\mathbf{R} \ni y \mapsto C(y) \in \mathbf{R}$ est une fonction quelconque.

Le cas de l'équation de transport est exactement identique à cet exemple — auquel il se ramène d'ailleurs par un changement de variables bien choisi. En effet, en appliquant la méthode des caractéristiques, on a, pour toute fonction $f \in C^1(\mathbf{R}_t \times \mathbf{R}_x^N)$

$$\left(\frac{\partial f}{\partial t} + v \cdot \nabla_x f\right)(t, y+tv) = \frac{d}{dt} f(t, y+tv)\,.$$

Ceci revient à faire le changement de variables $(t,x) = (t, y+tv) =: \phi(t,y)$, et à observer que

$$\left(\frac{\partial f}{\partial t} + v\cdot\nabla_x f\right)\circ\phi(t,y) = \frac{\partial}{\partial t}(f\circ\phi)(t,y)\,, \quad (t,y)\in\mathbf{R}\times\mathbf{R}^N\,.$$

Donc, dire que f annule l'opérateur de transport $\frac{\partial}{\partial t} + v\cdot\nabla_x$ équivaut à dire que $f\circ\phi$ annule la dérivée partielle $\frac{\partial}{\partial t}$, comme dans l'exemple ci-dessus.

Cette remarque suggère la définition suivante.

Définition 2.3.1 *Soient Ω ouvert à bord de classe C^1 de $\mathbf{R}^N$, ainsi que deux fonctions $a\equiv a(t,x)$ et $S\equiv S(t,x)$ continues sur $\mathbf{R}_+\times\Omega$. On dit qu'une fonction mesurable f est une solution généralisée de*

$$\left(\frac{\partial}{\partial t} + v\cdot\nabla_x\right) f(t,x) + a(t,x)f(t,x) = S(t,x)\,, \quad x\in\Omega,\, t>0\,,$$

si et seulement si la fonction

$$s\mapsto f(t+s,x+sv) \text{ est de classe } C^1 \text{ p.p. en } (t,x)\in\mathbf{R}_+\times\Omega\,,$$

et vérifie

$$\left(\frac{d}{ds} + a(t+s,x+sv)\right) f(t+s,x+sv) = S(t+s,x+sv)\,,$$
$$(t+s,x+sv)\in\mathbf{R}_+^*\times\Omega\,.$$

Evidemment, la méthode des caractéristiques montre que toute fonction f de classe C^1 sur $\mathbf{R}_+^*\times\Omega$ est une solution généralisée de l'équation de transport si et seulement si elle en est une solution au sens classique.

Il faut bien comprendre la signification de l'expression

$$\frac{\partial f}{\partial t} + v\cdot\nabla_x f$$

lorsque f est solution généralisée de l'équation de transport ci-dessus.

Comme f n'est pas de classe C^1 par rapport à toutes les variables (t,x), les dérivées partielles $\frac{\partial f}{\partial t}$ ou $\frac{\partial f}{\partial x_i}$ pour $i=1,\dots,N$ prises séparément n'existent pas en général. Mais en revenant au changement de variables avant la définition ci-dessus, on peut écrire

$$\frac{\partial f}{\partial t} + v\cdot\nabla_x f = \left(\frac{\partial}{\partial t}(f\circ\phi)\right)\circ\phi^{-1}\,.$$

où on rappelle que

$$\phi:\,(t,y)\mapsto(t,y+tv)\,.$$

Autrement dit, pour une solution généralisée de l'équation de transport, on a

$$\left(\frac{\partial}{\partial t} + v\cdot\nabla_x\right) f(t,x) + a(t,x)f(t,x) = S(t,x)\,, \quad x\in\Omega,\, t>0\,,$$

au sens où, bien que les dérivées partielles $\frac{\partial f}{\partial t}$ et $\frac{\partial f}{\partial x_i}$, $i = 1, \dots, N$ n'existent pas en général séparément, la combinaison linéaire particulière $(\frac{\partial}{\partial t} + v \cdot \nabla_x) f$ de ces dérivées partielles est bien définie p.p. comme fonction mesurable sur $\mathbf{R}_+^* \times \Omega$.

Avec cette nouvelle notion de solution, on arrive très facilement au résultat suivant d'existence et d'unicité de la solution du problème aux limites pour une équation de transport.

Théorème 2.3.2 *Soient Ω ouvert de $\mathbf{R}^N$ à bord de classe C^1 et $v \in \mathbf{R}^N \setminus \{0\}$, ainsi que deux fonctions $a \equiv a(t,x)$ et $S \equiv S(t,x)$ continues sur $\mathbf{R}_+ \times \overline{\Omega}$. Pour tout $f^{in} \in L^\infty_{loc}(\Omega)$ et tout $f_b^- \in L^\infty_{loc}(\mathbf{R}_+ \times \partial\Omega^-)$, le problème aux limites*

$$\begin{cases} \left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(t,x) = -a(t,x) f(t,x) + S(t,x), \quad x \in \Omega,\, t > 0\,, \\ f\big|_{\partial\Omega^-} = f_b^-\,, \\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

admet une unique solution généralisée f donnée par la même formule que dans le cas classique, à savoir que, p.p. en $(t,x) \in \mathbf{R}_+ \times \Omega$, l'on a

$$\begin{aligned} f(t,x) = {} & \mathbf{1}_{t \le \tau_x} f^{in}(x - tv) \exp\left(- \int_0^t a(s, x + (s-t)v) ds \right) \\ & + \mathbf{1}_{t > \tau_x} f_b^-(t - \tau_x, x^*) \exp\left(- \int_{t-\tau_x}^t a(s, x + (s-t)v) ds \right) \\ & + \int_{(t-\tau_x)^+}^t \exp\left(- \int_s^t a(\theta, x + (\theta - t)v) d\theta \right) S(s, x + (s-t)v) ds \end{aligned}$$

où on rappelle que

$$\tau_x = \inf\{t \ge 0 \,|\, x - tv \notin \overline{\Omega}\}\,, \text{ et } x^* = x - \tau_x v\,.$$

Donc, si Ω est convexe, si $f^{in} \in C^1(\overline{\Omega})$ et $f_b^- \in C^1(\mathbf{R}_+ \times \partial\Omega^-)$ et vérifient

$$f_b^-(0,y) = f^{in}(y)\,, \quad \frac{\partial f_b}{\partial t}^-(0,y) + v \cdot \nabla f^{in}(y) = S(0,y) - a(0,y) f^{in}(y)$$

pour tout $y \in \partial\Omega^-$, et si a et S sont de classe C^1 sur $\mathbf{R}_+ \times \overline{\Omega}$, alors cette solution généralisée est (p.p. égale à une fonction) de classe C^1 sur $\mathbf{R}_+ \times \Omega$.

Si Ω est convexe, si $f^{in} \in C(\overline{\Omega})$ et $f_b^- \in C(\mathbf{R}_+ \times \partial\Omega^-)$, et vérifient seulement

$$f_b^-(0,y) = f^{in}(y)\,, \quad y \in \partial\Omega^-\,,$$

alors cette solution généralisée f est (p.p. égale à une fonction) continue sur $\mathbf{R}_+ \times \overline{\Omega}$.

Précisons le sens dans lequel une solution généralisée du problème aux limites ci-dessus vérifie la condition initiale et la condition au bord. La condition initiale signifie que

$$\lim_{t \to 0^+} f(t, x + tv) = f^{in}(x) \text{ p.p. en } x \in \Omega\,,$$

tandis que la condition au bord signifie que

$$\lim_{s\to 0^+} f(t+s, y+sv) = f_b^-(t,y) \text{ p.p. en } (t,y) \in \mathbf{R}_+ \times \partial\Omega^- .$$

Remarquons que l'on n'a pas prescrit la valeur au bord de f sur la partie caractéristique $\partial\Omega^0$ du bord de Ω. Lorsque Ω n'est pas convexe, certaines trajectoires de vitesse v issues de $\partial\Omega^0$ peuvent entrer dans Ω — cf. figure 2.6. Il en résulte que la méthode des caractéristiques — et donc la formule explicite donnée dans le théorème ci-dessus ne définit pas f sur les points de Ω atteints par ces trajectoires. Mais ce n'est pas grave, comme le montre le lemme ci-dessous.

Lemme 2.3.3 (C. Bardos) *Soient Ω ouvert à bord de classe C^1 de $\mathbf{R}^N$ et une vitesse $v \in \mathbf{R}^N \setminus \{0\}$. Alors*

$$\{(x+tv) \,|\, x \in \partial\Omega^0 \text{ et } t \in \mathbf{R}\}$$

est un ensemble de mesure nulle dans $\mathbf{R}^N$.

Il est intéressant de comparer ce résultat avec le Lemme 2.2.4. Sous l'hypothèse que Ω est un ouvert à bord de classe C^1 convexe, l'énoncé (a) du Lemme 2.2.4 assure que l'ensemble

$$(\partial\Omega^0 + \mathbf{R}v) \cap \Omega$$

est vide. Sans l'hypothèse de convexité, cet ensemble n'est pas forcément vide, comme le montrent les exemples correpondant aux Figures (2.6)-(2.7). Mais le lemme de Bardos montre que ce même ensemble est de mesure nulle, et dans le cadre des solutions généralisées, cette information suffit. Cette seule observation suggère que la notion de solution généralisée introduite ci-dessus fournit le bon cadre mathématique pour l'étude des équations de transport, puisqu'il permet de traiter les problème aux limites dans les domaines les plus généraux [4].

La démonstration de ce lemme utilise le théorème de Sard — voir [33], p. 183 ; elle peut sans inconvénient être sautée par les lecteurs qui ignoreraient ce théorème.

Démonstration. Considérons à nouveau une variante de l'application Φ de la démonstration du Lemme 2.2.5 :

$$\Psi : \partial\Omega \times \mathbf{R} \ni (y,t) \mapsto y + tv \in \mathbf{R}^N .$$

4. On s'est limité ici à des ouverts à bord de classe C^1. Dans la pratique, on a affaire à des ouverts à bord de classe C^1 par morceaux. Un ouvert Ω de $\mathbf{R}^N$ est dit "à bord de classe C^1 par morceaux" si tout point de $\overline{\Omega}$ admet un voisinage ouvert pour la topologie induite qui soit C^1-difféomorphe à un ouvert d'un polyèdre de $\mathbf{R}^N$ pour la topologie induite. Contrairement au cas des ouverts de classe C^1, il n'est pas possible de définir un champ de vecteurs unitaire normal en tout point de $\partial\Omega$ lorsque Ω est un ouvert à bord de classe C^1 par morceaux. Par exemple, lorsque $\overline{\Omega}$ est un polyèdre de $\mathbf{R}^3$, le champ unitaire normal extérieur n'est défini que sur les faces du polyèdre, et pas sur les arêtes, ni sur les sommets. Mais le champ unitaire normal extérieur à un ouvert à bord de classe C^1 par morceaux est défini p.p. sur le bord, ce qui suffit à assurer la validité de la formule de Green. Tous les résultats de ce chapitre et des deux suivants peuvent être adaptés assez facilement au cas des ouverts de classe C^1 par morceaux, au prix de lourdeurs techniques supplémentaires que nous avons préféré éviter au lecteur.

Cette application est évidemment de classe C^1 et

$$D\Psi(y,t)\cdot(u,s) = u + sv \text{ pour tout } (u,s) \in T_y\partial\Omega \times \mathbf{R}\,.$$

Dire que $u \in T_y\partial\Omega$ équivaut à dire que $u \in \mathbf{R}^N$ vérifie $u\cdot n_y = 0$. Par conséquent

$$y \in \partial\Omega^0 \Rightarrow (D\Psi(y,t)\cdot(u,s))\cdot n_y = (u+sv)\cdot n_y = 0$$

pour tout $(u,s) \in T_y\partial\Omega \times \mathbf{R}$. Autrement dit

$$y \in \partial\Omega^0 \Rightarrow \text{rang}(D\Psi(y,t)) \leq N-1\,,$$

c'est-à-dire que (y,t) est un point critique de Ψ. Par conséquent, l'ensemble

$$\{(x+tv)\,|\,x \in \partial\Omega^0 \text{ et } t \in \mathbf{R}\}$$

est contenu dans l'ensemble des valeurs critiques de l'application Ψ qui est de classe C^1 entre variétés différentiables d'égales dimensions. D'après le théorème de Sard, cet ensemble est donc de mesure nulle. ■

Démonstration du Théorème 2.3.2. Soit $\mathcal{N}^0 \subset \Omega$ de mesure nulle tel que f^{in} soit définie et localement bornée sur $\Omega\backslash\mathcal{N}^0$, et soit $\mathcal{N}_b \subset \mathbf{R}_+\times\partial\Omega^-$, ensemble de mesure nulle tel que f_b^- soit définie et localement bornée sur $\mathbf{R}_+\times\partial\Omega^-\setminus\mathcal{N}_b$. Notons

$$\mathcal{N}_d := \Big(\big((\{0\}\times\mathcal{N}^0)\cup\mathcal{N}_b\cup(\mathbf{R}_+\times\partial\Omega^0)\big) + \mathbf{R}_+(1,v)\Big)\cap(\mathbf{R}_+\times\overline{\Omega})\,.$$

Observons que

$$\begin{aligned}(\{0\}\times\mathcal{N}^0) + \mathbf{R}_+(1,v) &= \tilde{\phi}(\mathcal{N}^0\times\mathbf{R}_+)\,,\\ \mathcal{N}_b + \mathbf{R}_+(1,v) &= \tilde{\psi}(\mathcal{N}_b\times\mathbf{R}_+)\,,\end{aligned}$$

où $\tilde{\phi}$ est l'application définie comme suit :

$$\tilde{\phi}:\ \mathbf{R}^N\times\mathbf{R} \ni (y,s) \mapsto (s, y+sv) \in \mathbf{R}\times\mathbf{R}^N\,,$$

tandis que

$$\tilde{\psi}:\ (\mathbf{R}\times\partial\Omega)\times\mathbf{R} \ni ((t,y),s) \mapsto (t+s, y+sv) \in \mathbf{R}\times\mathbf{R}^N\,.$$

Les applications $\tilde{\phi}$ et $\tilde{\psi}$ sont évidemment lipschitziennes. Donc les ensembles $(\{0\}\times\mathcal{N}^0)+\mathbf{R}_+(1,v)$ et $\mathcal{N}_b+\mathbf{R}_+(1,v)$ sont Lebesgue-négligeables comme images d'ensembles Lebesgue-négligeables par des applications lipschitziennes entre variétés de même dimension.

Comme

$$\mathcal{N}_d \subset A\cup B\cup C\,,$$

avec

$$\begin{aligned}A &= (\{0\}\times\mathcal{N}^0) + \mathbf{R}_+(1,v)\,,\\ B &= \mathcal{N}_b + \mathbf{R}_+(1,v)\,,\\ C &= (\mathbf{R}_+\times\partial\Omega^0) + \mathbf{R}_+(1,v)\,,\end{aligned}$$

que C est Lebesgue-négligeable d'après le lemme de Bardos, tandis que A et B sont Lebesgue-négligeables par l'argument ci-dessus, on conclut que $\mathcal{N}_d$ est de mesure nulle dans $\mathbf{R}_+ \times \overline{\Omega}$.

La formule ci-dessus définit une fonction f sur $(\mathbf{R}_+ \times \overline{\Omega}) \setminus \mathcal{N}_d$, telle que

$$\begin{aligned} f(t+s,x+sv) &= \mathbf{1}_{t+s\le\tau_{x+sv}} f^{in}(x+sv-(t+s)v) \\ &\quad \times \exp\left(-\int_0^{t+s} a(\theta,x+sv+(\theta-t-s)v)d\theta\right) \\ &\quad + \mathbf{1}_{t+s>\tau_{x+sv}} f_b^-(t+s-\tau_{x+sv},x+sv-\tau_{x+sv}v) \\ &\quad \times \exp\left(-\int_{t+s-\tau_{x+sv}}^{t+s} a(\theta,x+sv+(\theta-t-s)v)d\theta\right) \\ &\quad + \int_{(t+s-\tau_{x+sv})^+}^{t+s} \exp\left(-\int_\sigma^{t+s} a(\theta,x+sv+(\theta-t-s)v)d\theta\right) \\ &\qquad\qquad \times S(\sigma,x+sv+(\sigma-t-s)v)d\sigma\,, \end{aligned}$$

pour tout $s \in \mathbf{R}$ tel que $x+sv \in \Omega$. Comme

$$\tau_{x+sv} = \tau_x + s\,,$$

le membre de droite dans l'égalité ci-dessus devient

$$\begin{aligned} f(t+s,x+sv) &= \mathbf{1}_{t\le\tau_x} f^{in}(x-tv)\exp\left(-\int_0^{t+s} a(\theta,x+(\theta-t)v)d\theta\right) \\ &\quad + \mathbf{1}_{t>\tau_x} f_b^-(t-\tau_x,x^*)\exp\left(-\int_{t-\tau_x}^{t+s} a(\theta,x+(\theta-t)v)d\theta\right) \\ &\quad + \int_{(t-\tau_x)^+}^{t+s} \exp\left(-\int_\sigma^{t+s} a(\theta,x+(\theta-t)v)d\theta\right) S(\sigma,x+(\sigma-t)v)d\sigma\,. \end{aligned}$$

Comme tous les intégrandes ci-dessus sont des fonctions continues, la fonction

$$s \mapsto f(t+s,x+sv)$$

est donc de classe C^1 en s pour tout $(t,x) \in (\mathbf{R}_+ \times \overline{\Omega}) \setminus \mathcal{N}_d$. De plus, la formule ci-dessus entraîne que

$$\begin{aligned} &\frac{d}{ds} f(t+s,x+sv) \\ &= -\mathbf{1}_{t\le\tau_x} f^{in}(x-tv)a(t+s,x+sv)\exp\left(-\int_0^{t+s} a(\theta,x+(\theta-t)v)d\theta\right) \\ &- \mathbf{1}_{t>\tau_x} f_b^-(t-\tau_x,x^*)a(t+s,x+sv)\exp\left(-\int_{t-\tau_x}^{t+s} a(\theta,x+(\theta-t)v)d\theta\right) \\ &-a(t+s,x+sv)\int_{(t-\tau_x)^+}^{t+s} \exp\left(-\int_\sigma^{t+s} a(\theta,x+(\theta-t)v)d\theta\right) S(\sigma,x+(\sigma-t)v)d\sigma \\ &+ S(t+s,x+sv)\,, \end{aligned}$$

c'est-à-dire que

$$\frac{d}{ds}f(t+s,x+sv) = -a(t+s,x+sv)f(t+s,x+sv) + S(t+s,x+sv)$$

pour tout $(t,x) \in (\mathbf{R}_+ \times \overline{\Omega}) \setminus \mathcal{N}_d$ et pour tout s tel que $(t+s,x+sv) \in \mathbf{R}_+^* \times \Omega$.

De plus

$$\lim_{t\to 0^+} f(t,x+tv) = \lim_{t\to 0^+} f^{in}(x)\exp\left(-\int_0^t a(s,x+sv)ds\right) = f^{in}(x)$$

pour tout $x \in \Omega \setminus \mathcal{N}^0$, tandis que

$$\begin{aligned}\lim_{s\to 0^+} f(t+s,y+sv) &= \lim_{s\to 0^+} f_b^-(t,y)\exp\left(-\int_t^{t+s} a(\theta,y+(\theta-t)v)d\theta\right)\\ &= f_b^-(t,y)\end{aligned}$$

pour tout $(t,y) \in (\mathbf{R}_+ \times \partial\Omega^-) \setminus \mathcal{N}_b$, puisque $\tau_{y+sv} = s \to 0$. Cette fonction est donc bien une solution généralisée du problème aux limites.

Réciproquement, supposons que f est solution généralisée du problème aux limites. Alors il existe $\mathcal{N}_f \subset \mathbf{R}_+ \times \Omega$ de mesure nulle tel que

$$\text{la fonction } s \mapsto f(t+s,x+sv) \text{ soit de classe } C^1$$

et

$$\frac{d}{ds}f(t+s,x+sv) = -a(t+s,x+sv)f(t+s,x+sv) + S(t+s,x+sv)$$

pour tout $(t,x) \in (\mathbf{R}_+ \times \Omega) \setminus \mathcal{N}_f$ et tout $s \in]-\min(t,\tau_x),0[$, et de plus

$$\begin{aligned}&\lim_{t\to 0^+} f(t,x+tv) = f^{in}(x)\,, && x \in \Omega \setminus \mathcal{N}^0\,,\\ &\lim_{s\to 0^+} f(t+s,y+sv) = f_b^-(y)\,, && (t,x) \in (\mathbf{R}_+ \times \partial\Omega^-) \setminus \mathcal{N}_b\,.\end{aligned}$$

Alors, pour tout $(t,x) \in (\mathbf{R}_+ \times \Omega) \setminus (\mathcal{N}_f \cup \mathcal{N}_d)$, l'on a

$$\begin{aligned}&\frac{d}{ds}\left(f(t+s,x+sv)\exp\left(\int_0^s a(t+\theta,x+\theta v)d\theta\right)\right)\\ &\quad = S(t+s,x+sv)\exp\left(\int_0^s a(t+\theta,x+\theta v)d\theta\right)\end{aligned}$$

pour tout $s \in]-\min(t,\tau_x),0[$. Intégrons chaque membre de cette égalité sur

l'intervalle $]-\min(t,\tau_x),0[$. Si $t<\tau_x$, on trouve que

$$\begin{aligned}
f(t,x) &= \lim_{\epsilon\to 0^+} f(0,x-(t-\epsilon)v)\exp\left(\int_0^{-t+\epsilon} a(t+\theta,x+\theta v)d\theta\right)\\
&\quad+\lim_{\epsilon\to 0^+}\int_{-t+\epsilon}^0 S(t+s,x+sv)\exp\left(\int_0^s a(t+\theta,x+\theta v)d\theta\right)ds\\
&= f^{in}(x-tv)\exp\left(-\int_0^t a(t-\theta,x-\theta v)d\theta\right)\\
&\quad+\int_0^t S(t-s,x-sv)\exp\left(-\int_0^s a(t-\theta,x-\theta v)d\theta\right)ds\\
&= f^{in}(x-tv)\exp\left(-\int_0^t a(\theta,x+(\theta-t)v)d\theta\right)\\
&\quad+\int_0^t S(s,x+(s-t)v)\exp\left(-\int_s^t a(\theta,x+(\theta-t)v)d\theta\right)ds\,,
\end{aligned}$$

pour tout $(t,x)\in(\mathbf{R}_+\times\Omega)\setminus(\mathcal{N}_f\cup\mathcal{N}_d)$. (La seconde égalité s'obtient par le changement de variable $s\mapsto -s$, et la troisième par le changement de variables $s\mapsto t-s$.)

Au contraire, si $t>\tau_x$, on trouve que

$$\begin{aligned}
f(t,x) &= \lim_{\epsilon\to 0^+} f(t-\tau_x+\epsilon,x-(\tau_x-\epsilon)v)\exp\left(\int_0^{-\tau_x+\epsilon} a(t+\theta,x+\theta v)d\theta\right)\\
&\quad+\lim_{\epsilon\to 0^+}\int_{-\tau_x+\epsilon}^0 S(t+s,x+sv)\exp\left(\int_0^s a(t+\theta,x+\theta v)d\theta\right)ds\\
&= f_b^-(t-\tau_x,x^*)\exp\left(-\int_0^{\tau_x} a(t-\theta,x-\theta v)d\theta\right)\\
&\quad+\int_0^{\tau_x} S(t-s,x-sv)\exp\left(-\int_0^s a(t-\theta,x-\theta v)d\theta\right)ds\\
&= f_b^-(t-\tau_x,x^*)\exp\left(-\int_{t-\tau_x}^t a(\theta,x+(\theta-t)v)d\theta\right)\\
&\quad+\int_{t-\tau_x}^t S(s,x+(s-t)v)\exp\left(-\int_s^t a(\theta,x+(\theta-t)v)d\theta\right)ds
\end{aligned}$$

pour tout $(t,x)\in(\mathbf{R}_+\times\Omega)\setminus(\mathcal{N}_f\cup\mathcal{N}_d)$, grâce aux mêmes changements de variables. ■

2.4 Principe du maximum faible pour l'équation de transport libre

Notation : lorsque X est un espace topologique, on note $C_b(X)$ l'espace des fonctions continues sur X à valeurs dans $\mathbf{R}$ qui sont bornées sur X.

Dans la suite de ce cours, nous aurons souvent besoin d'estimer des solutions d'équations de transport dépendant d'un paramètre uniformément par rapport à ce paramètre. Ce sera notamment le cas pour les questions d'approximation par la diffusion dont nous avons donné une idée au chapitre précédent, ou encore d'homogénéisation, dont il sera question plus loin.

Dans ce genre de situation, bien que la famille de solutions considérée dépende d'une façon a priori inconnue d'un ou plusieurs paramètres destinés à tendre vers une ou plusieurs valeurs limites, les données au bord correspondant à ces solutions en dépendent de manière connue — ou en sont, dans certains cas, indépendantes.

Il est donc extrêmement utile pour étudier les solutions d'équations de transport de disposer de résultats permettant d'estimer la solution en fonction des données du problème aux limites.

Voici un premier énoncé de ce type.

Théorème 2.4.1 *Soient Ω ouvert de $\mathbf{R}^N$ à bord de classe C^1 et $v \in \mathbf{R}^N \setminus \{0\}$, ainsi que deux fonctions $a \equiv a(t,x)$, $S \equiv S(t,x) \in C_b(\mathbf{R}_+ \times \overline{\Omega})$. Soient une donnée initiale $f^{in} \in L^\infty(\Omega)$ et une donnée au bord $f_b^- \in L^\infty(\mathbf{R}_+ \times \partial\Omega^-)$; alors la solution généralisée f du problème*

$$\begin{cases} \left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(t,x) = -a(t,x) f(t,x) + S(t,x), \quad x \in \Omega,\, t > 0\,, \\ f\big|_{\partial\Omega^-} = f_b^-\,, \\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

vérifie, pour tout $T > 0$, et en notant $a_-(t,x) = \max(0, -a(t,x))$,

$$\|f\|_{L^\infty([0,T]\times\Omega)} \leq e^{T\|a_-\|_{L^\infty(\mathbf{R}_+\times\Omega)}} T \|S\|_{L^\infty(\mathbf{R}_+\times\Omega)}$$

$$+ e^{T\|a_-\|_{L^\infty(\mathbf{R}_+\times\Omega)}} \max\left(\|f^{in}\|_{L^\infty(\Omega)}, \|f_b^-\|_{L^\infty(\mathbf{R}_+\times\partial\Omega^-)}\right).$$

De plus, les conditions

$$f^{in} \geq 0 \text{ p.p. sur } \Omega\,,\ f_b^- \geq 0 \text{ p.p. sur } \mathbf{R}_+^* \times \partial\Omega^- \text{ et } S \geq 0 \text{ sur } \mathbf{R}_+ \times \overline{\Omega}$$

impliquent que

$$f \geq 0 \text{ p.p. sur } \mathbf{R}_+^* \times \Omega\,.$$

Démonstration. Partons de la formule exprimant la solution généralisée du problème aux limites pour l'équation de transport donnée au théorème précédent :

$$\begin{aligned} f(t,x) = {} & \mathbf{1}_{t\leq\tau_x} f^{in}(x - tv) \exp\left(-\int_0^t a(s, x + (s-t)v) ds\right) \\ & + \mathbf{1}_{t>\tau_x} f_b^-(t - \tau_x, x^*) \exp\left(-\int_{t-\tau_x}^t a(s, x + (s-t)v) ds\right) \\ & + \int_{(t-\tau_x)^+}^t \exp\left(-\int_s^t a(\theta, x + (\theta - t)v) d\theta\right) S(s, x + (s-t)v) ds \end{aligned}$$

p.p. en $(t,x) \in \mathbf{R}_+ \times \Omega$. Donc

$$\begin{aligned}
|f(t,x)| &\leq \mathbf{1}_{t\leq\tau_x}|f^{in}(x-tv)|\exp\left(\int_0^t a_-(s,x+(s-t)v)ds\right)\\
&+ \mathbf{1}_{t>\tau_x}|f_b^-(t-\tau_x,x^*)|\exp\left(\int_{t-\tau_x}^t a_-(s,x+(s-t)v)ds\right)\\
&+ \int_{(t-\tau_x)^+}^t \exp\left(\int_s^t a_-(\theta,x+(\theta-t)v)d\theta\right)|S(s,x+(s-t)v)|ds\\
&\leq \left(\mathbf{1}_{t\leq\tau_x}\|f^{in}\|_{L^\infty(\Omega)} + \mathbf{1}_{t>\tau_x}\|f_b^-\|_{L^\infty(\mathbf{R}_+\times\partial\Omega^-)}\right)e^{T\|a_-\|_{L^\infty(\mathbf{R}_+\times\Omega)}}\\
&+ T\|S\|_{L^\infty(\mathbf{R}_+\times\Omega)}e^{T\|a_-\|_{L^\infty(\mathbf{R}_+\times\Omega)}}
\end{aligned}$$

p.p. en $(t,x) \in \mathbf{R}_+ \times \Omega$, puisque toutes les intégrales intervenant au membre de droite de l'égalité ci-dessus portent sur des intervalles de longueur au plus $t \leq T$. On en déduit la majoration annoncée en remarquant que

$$\begin{aligned}
\mathbf{1}_{t\leq\tau_x}\|f^{in}\|_{L^\infty(\Omega)} + \mathbf{1}_{t>\tau_x}\|f_b^-\|_{L^\infty(\mathbf{R}_+\times\partial\Omega^-)}&\\
\leq \max\left(\|f^{in}\|_{L^\infty(\Omega)}, \|f_b^-\|_{L^\infty(\mathbf{R}_+\times\partial\Omega^-)}\right)&\ .
\end{aligned}$$

Ceci complète la démonstration. ■

Nous utiliserons très souvent, dans la suite de ce cours, les deux cas particuliers suivants de l'estimation ci-dessus.

Cas particuliers :
(1) si $a \geq 0$ sur $\mathbf{R}_+ \times \overline{\Omega}$ — autrement dit si a est un taux d'amortissement — la fonction $a_- : (t,x) \mapsto \max(0,-a(t,x))$ est identiquement nulle, et on a

$$\|f\|_{L^\infty([0,T]\times\Omega)} \leq T\|S\|_{L^\infty(\mathbf{R}_+\times\Omega)} + \max\left(\|f^{in}\|_{L^\infty(\Omega)}, \|f_b^-\|_{L^\infty(\mathbf{R}_+\times\partial\Omega^-)}\right)$$

pour tout $T \geq 0$.
(2) Si on a à la fois $a \geq 0$ et $S = 0$, alors

$$\|f\|_{L^\infty(\mathbf{R}_+\times\Omega)} \leq \max\left(\|f^{in}\|_{L^\infty(\Omega)}, \|f_b^-\|_{L^\infty(\mathbf{R}_+\times\partial\Omega^-)}\right)\ .$$

L'énoncé (2) est un cas particulier de **principe du maximum faible** : le maximum de la (valeur absolue de la) solution est "atteint" sur le bord du domaine $\mathbf{R}_+ \times \Omega$.

Rappelons l'exemple typique de ce que l'on appelle un principe du maximum : pour une fonction ϕ holomorphe sur un ouvert $\mathcal{O}$ connexe et borné de $\mathbf{C}$, et continue sur $\overline{\mathcal{O}}$, le maximum du module de ϕ est atteint sur la frontière de $\mathcal{O}$; de plus, si ce maximum est également atteint en un point de $\mathcal{O}$, alors la fonction ϕ est constante sur $\mathcal{O}$.

Le même énoncé vaut aussi pour les fonctions harmoniques sur $\mathcal{O}$ — c'est-à-dire de laplacien nul sur $\mathcal{O}$: voir [23], Théorème 7.2.6.

Dans le cas de l'équation de transport, seule la première partie de l'énoncé ci-dessus est vraie ; c'est pourquoi on appelle souvent ce type d'énoncé un principe

du maximum *faible* — par opposition au principe du maximum *fort* qui vaut pour les fonctions harmoniques ou les modules de fonctions holomorphes.

Il existe d'autres équations aux dérivées partielles que l'équation de transport vérifiant le principe du maximum faible, comme par exemple l'équation de la chaleur : voir [2], Théorème 5.2.22, ou [23], Théorème 8.3.1.

2.5 Estimation L^2 pour l'équation de transport

Dans la suite de ce cours, tout particulièrement dans l'étude des schémas numériques pour l'équation de transport, on aura besoin d'estimer la taille de la solution du problème aux limites en fonction de la taille de ses données. Le principe du maximum étudié dans la section précédente est évidemment une façon d'y parvenir. Dans cette section, nous allons présenter une autre estimation du même genre, mais en moyenne quadratique — alors que le principe du maximum fournit une estimation pour la norme de la convergence uniforme.

Proposition 2.5.1 *Soient Ω ouvert de $\mathbf{R}^N$ à bord de classe C^1 et $v \in \mathbf{R}^N \setminus \{0\}$, ainsi que deux fonctions $a \equiv a(t,x)$, $S \equiv S(t,x) \in C^1(\mathbf{R}_+ \times \overline{\Omega})$. Soient des données initiales $f^{in} \in C^1(\Omega)$ et au bord $f_b^- \in C^1(\mathbf{R}_+ \times \partial\Omega^-)$ vérifiant les conditions de compatibilité*

$$f^{in}(y) = f_b^-(0,y)\,, \quad y \in \partial\Omega^-\,,$$

et

$$\frac{\partial f_b}{\partial t}^-(0,y) + v \cdot \nabla f^{in}(y) = -a(0,y) f^{in}(y) + S(0,y) \quad y \in \partial\Omega^-\,.$$

Si $f^{in} \in L^2(\Omega)$, $S \in L^2([0,T] \times \Omega)$ et $f_b^- \in L^2([0,T] \times \partial\Omega^-; |v \cdot n_x| dt d\sigma(x))$ (en notant $d\sigma(x)$ l'élément de surface sur $\partial\Omega$ et n_x le vecteur unitaire normal à $\partial\Omega$ au point x, dirigé vers l'extérieur de Ω) et si $a(t,x) \geq 0$ pour tout $(t,x) \in \mathbf{R}_+ \times \overline{\Omega}$, alors la solution $f \in C^1(\mathbf{R}_+ \times \overline{\Omega})$ du problème

$$\begin{cases} \left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(t,x) = -a(t,x) f(t,x) + S(t,x), \quad x \in \Omega,\, t > 0\,, \\ f\big|_{\partial\Omega^-} = f_b^-\,, \\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

vérifie l'inégalité différentielle

$$\frac{d}{dt}\|f(t,\cdot)\|^2_{L^2(\Omega)} \leq \|f(t,\cdot)\|^2_{L^2(\Omega)} + \|S(t,\cdot)\|^2_{L^2(\Omega)} + \|f_b^-(t,\cdot)\|^2_{L^2(\partial\Omega^-;|v\cdot n_x| d\sigma(x))}\,.$$

De plus, pour tout $T > 0$ et tout $t \in [0,T]$,

$$\|f(t,\cdot)\|^2_{L^2(\Omega)} \leq e^t\left(\|f^{in}\|^2_{L^2(\Omega)} + \|S\|^2_{L^2([0,T]\times\Omega)} + \|f_b^-\|^2_{L^2([0,T]\times\partial\Omega^-;|v\cdot n_x| dt d\sigma(x))}\right).$$

Démonstration. La solution f étant de classe C^1 sur $\mathbf{R}_+ \times \overline{\Omega}$, par dérivation sous le signe somme, on a

$$\begin{aligned}\frac{d}{dt}\int_\Omega \tfrac{1}{2}f(t,x)^2dx &= \int_\Omega f(t,x)\frac{\partial f}{\partial t}(t,x)dx \\ &= \int_\Omega f(t,x)\left(-v\cdot\nabla_x f(t,x) - a(t,x)f(t,x) + S(t,x)\right)dx \\ &= -\int_\Omega \tfrac{1}{2}v\cdot\nabla_x f(t,x)^2dx - \int_\Omega a(t,x)f(t,x)^2dx + \int_\Omega f(t,x)S(t,x)dx\,.\end{aligned}$$

D'une part, d'après la formule de Green, en notant $d\sigma(x)$ l'élément de surface sur $\partial\Omega$ et n_x le vecteur unitaire normal à $\partial\Omega$ au point x, dirigé vers l'extérieur de Ω,

$$\begin{aligned}-\int_\Omega \tfrac{1}{2}v\cdot\nabla_x f(t,x)^2dx &= -\int_{\partial\Omega_+}\tfrac{1}{2}f(t,x)^2 v\cdot n_x d\sigma(x) \\ &\quad + \int_{\partial\Omega^-}\tfrac{1}{2}f(t,x)^2|v\cdot n_x|d\sigma(x) \\ &\le \int_{\partial\Omega^-}\tfrac{1}{2}f_b^-(t,x)^2|v\cdot n_x|d\sigma(x)\end{aligned}$$

puisque $v\cdot n_x > 0$ sur $\partial\Omega^+$.

D'autre part

$$-\int_\Omega a(t,x)f(t,x)^2dx \le 0\,,$$

de sorte que

$$\frac{d}{dt}\int_\Omega f(t,x)^2dx \le 2\int_\Omega f(t,x)S(t,x)dx + \int_{\partial\Omega^-} f_b^-(t,x)^2|v\cdot n_x|d\sigma(x)\,.$$

On en déduit que

$$\begin{aligned}\frac{d}{dt}\|f(t,\cdot)\|^2_{L^2(\Omega)} &\le \|f(t,\cdot)\|^2_{L^2(\Omega)} + \|S(t,\cdot)\|^2_{L^2(\Omega)} \\ &\quad + \|f_b^-(t,\cdot)\|^2_{L^2(\partial\Omega^-;|v\cdot n_x d\sigma(x))}\,,\end{aligned}$$

qui est l'inégalité différentielle annoncée.

Cette inégalité s'écrit encore

$$\begin{aligned}&\frac{d}{ds}\left(e^{-s}\|f(s,\cdot)\|^2_{L^2(\Omega)}\right) \\ &\quad\le e^{-s}\left(\|S(s,\cdot)\|^2_{L^2(\Omega)} + \|f_b^-(s,\cdot)\|^2_{L^2(\partial\Omega^-;|v\cdot n_x d\sigma(x))}\right) \\ &\quad\le \left(\|S(s,\cdot)\|^2_{L^2(\Omega)} + \|f_b^-(s,\cdot)\|^2_{L^2(\partial\Omega^-;|v\cdot n_x d\sigma(x))}\right)\,,\end{aligned}$$

d'où la deuxième inégalité de la proposition par intégration sur $[0,t]$ de chaque membre de l'inégalité ci-dessus. ■

2.6 Equation stationnaire du transport

Jusqu'ici, nous avons étudié uniquement l'équation de transport sous l'angle des problèmes d'évolution. Bien que les problèmes d'évolution soient en principe les problèmes fondamentaux, et que leur analyse soit nécessaire à la compréhension des régimes transitoires, il faut aussi en pratique étudier les régimes permanents, ce qui conduit à considérer des équations de transport stationnaires.

Il existe plusieurs manières d'introduire l'équation de transport stationnaire. En voici deux.

Considérons le problème aux limites

$$\begin{cases} \left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(t,x) = -a(t,x) f(t,x), \quad x \in \Omega,\, t > 0\,, \\ f\big|_{\partial\Omega^-} = f_b^-\,, \\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

où Ω est un ouvert à bord de classe C^1 de $\mathbf{R}^N$, où la fonction a est continue sur $\mathbf{R}_+ \times \overline{\Omega}$, où la donnée initiale $f^{in} \in L^\infty_{loc}(\Omega)$ et la donnée au bord $f_b^- \in L^\infty_{loc}(\partial\Omega^-)$.

Supposons qu'on cherche à en approcher la solution par un schéma d'Euler implicite en temps. Autrement dit, on se donne un pas de temps $\Delta t > 0$, et on approche la fonction $f \equiv f(t,x)$ par la suite de fonctions $(f_k(x))_{k\geq 0}$, l'idée étant que

$$f_k(x) \simeq f(k\Delta t, x) \text{ en un sens à préciser lorsque } \Delta t \to 0\,.$$

La suite de fonctions f_k est construite par récurrence en écrivant l'équation de transport en tous les points de la forme $(k\Delta t, x)$ pour $k \geq 1$, et en utilisant l'approximation rétrograde de la dérivée en temps

$$\frac{\partial f}{\partial t}(k\Delta t, x) \simeq \frac{f(k\Delta t, x) - f((k-1)\Delta t, x)}{\Delta t} \simeq \frac{f_k(x) - f_{k-1}(x)}{\Delta t}\,.$$

Le problème aux limites ci-dessus s'écrit donc, après discrétisation en temps

$$\begin{cases} \frac{f_k(x) - f_{k-1}(x)}{\Delta t} + v \cdot \nabla_x f_k(x) = -a(k\Delta t, x) f_k(x), \quad x \in \Omega,\, k \geq 1\,, \\ f_k\big|_{\partial\Omega^-} = f_b^-(k\Delta t, \cdot)\,, \\ f_0 = f^{in}\,, \end{cases}$$

ce que l'on peut encore mettre sous la forme

$$\begin{cases} \frac{1}{\Delta t} f_k(x) + v \cdot \nabla_x f_k(x) + a(k\Delta t, x) f_k(x) = \frac{1}{\Delta t} f_{k-1}(x), \quad x \in \Omega,\, k \geq 1\,, \\ f_k\big|_{\partial\Omega^-} = f_b^-(k\Delta t, \cdot)\,, \\ f_0 = f^{in}\,. \end{cases}$$

On voit donc que la fonction f_k est obtenue connaissant la fonction f_{k-1} en résolvant une équation de transport stationnaire d'inconnue $F \equiv F(x)$ de la forme

$$\begin{cases} \lambda F(x) + v \cdot \nabla_x F(x) + A(x)F(x) = Q(x), \quad x \in \Omega, \\ F\big|_{\partial\Omega^-} = F_b^- \,, \end{cases}$$

où $\lambda > 0$, où le taux d'amortissement ou d'amplification A est continu sur $\overline{\Omega}$, tandis que le terme source Q appartient à $L^\infty(\Omega)$ et la donnée au bord F_b^- à $L^\infty(\partial\Omega^-)$.

Dans le cas présent, on a $\lambda = \frac{1}{\Delta t}$, $F = f_k$, $A = a(k\Delta t, \cdot)$, $Q = \frac{1}{\Delta t} f_{k-1}$, et $F_b^- = f_b^-(k\Delta t, \cdot)$.

On voit donc qu'il est très naturel d'avoir à résoudre des équations de transport stationnaires — et nous retrouverons ce type de problème plus loin dans ce cours, lorsqu'il sera question de méthodes numériques pour l'équation de transport.

Mais il existe aussi une autre manière d'arriver à l'équation de transport stationnaire ci-dessus.

Partons du problème d'évolution

$$\begin{cases} \left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(t,x) = -a(x) f(t,x), \quad x \in \Omega,\, t > 0\,, \\ f\big|_{\partial\Omega^-} = f_b^- \,, \\ f\big|_{t=0} = f^{in}\,. \end{cases}$$

Appliquons aux deux membres de cette équation la transformation de Laplace en temps

$$\phi \equiv \phi(t) \mapsto \Phi(\lambda) := \int_0^{+\infty} e^{-\lambda t} \phi(t) dt\,.$$

Cette transformation est bien définie, par exemple pour $\phi \in L^\infty(\mathbf{R}_+)$, auquel cas $\Phi(\lambda)$ existe pour tout $\lambda > 0$, et vérifie

$$|\Phi(\lambda)| \leq \frac{1}{\lambda} \|\phi\|_{L^\infty}\,.$$

On notera

$$F(\lambda, x) = \int_0^{+\infty} e^{-\lambda t} f(t,x) dt\,.$$

Supposons que $f^{in} \in L^\infty(\Omega)$, que $f_b^- \in L^\infty(\mathbf{R}_+ \times \partial\Omega^-)$ et que $a \geq 0$. D'après le principe du maximum (Théorème 2.4.1)

$$\|f\|_{L^\infty(\mathbf{R}_+ \times \Omega)} \leq \max\left(\|f^{in}\|_{L^\infty(\Omega)}, \|f_b^-\|_{L^\infty(\mathbf{R}_+ \times \partial\Omega^-)}\right) < +\infty$$

de sorte que la fonction $F(\lambda, x)$ est bien définie pour tout $\lambda > 0$ et p.p. en $x \in \Omega$. Si f est de classe C^1 sur $\mathbf{R}_+ \times \Omega$, — ce qui est le cas lorsque Ω est convexe, les

fonctions f^{in} et a de classe C^1 sur $\overline{\Omega}$, la donnée au bord f_b^- de classe C^1 sur $\mathbf{R}_+ \times \partial\Omega^-$ et que f^{in} et f_b^- vérifient les relations de compatibilité

$$f^{in}(y) = f_b^-(0,y)\,,$$
$$\frac{\partial f_b}{\partial t}^-(0,y) + v\cdot\nabla f^{in}(y) + a(y)f^{in}(y) = 0\,,$$

pour tout $y \in \partial\Omega^-$ — on aura d'une part

$$\int_0^{+\infty} e^{-\lambda t} v\cdot\nabla_x f(t,x)dt = v\cdot\nabla_x \int_0^{+\infty} e^{-\lambda t} f(t,x)dt\,,$$

et d'autre part, en intégrant par parties,

$$\begin{aligned}\int_0^{+\infty} e^{-\lambda t}\frac{\partial f}{\partial t}(t,x)dt &= \left[e^{-\lambda t}f(t,x)\right]_{t=0}^{t=\infty} + \lambda\int_0^{+\infty} e^{-\lambda t}f(t,x)dt\\ &= \lambda F(\lambda,x) - f^{in}(x)\,.\end{aligned}$$

Bref, si $f \equiv f(t,x)$ est solution du problème d'évolution

$$\begin{cases} \left(\frac{\partial}{\partial t} + v\cdot\nabla_x\right) f(t,x) = -a(x)f(t,x), \quad x\in\Omega,\, t>0\,,\\ f\big|_{\partial\Omega^-} = f_b^-\,,\\ f\big|_{t=0} = f^{in}\,,\end{cases}$$

sa transformée de Laplace en temps, $F \equiv F(\lambda,x)$, est, pour tout $\lambda > 0$, solution du problème stationnaire

$$\begin{cases} \lambda F(\lambda,x) + v\cdot\nabla_x F(\lambda,x) + a(x)F(\lambda,x) = f^{in}(x), \quad x\in\Omega,\\ F(\lambda,\cdot)\big|_{\partial\Omega^-} = F_b^-(\lambda,\cdot)\,,\end{cases}$$

où

$$F_b^-(\lambda,y) = \int_0^{+\infty} e^{-\lambda t} f_b^-(t,y)dt\,, \quad y\in\partial\Omega^-\,.$$

Cette nouvelle manière d'arriver au problème stationnaire présente l'avantage de fournir immédiatement une formule explicite donnant sa solution. En effet, d'après le Théorème 2.3.2, on a

$$\begin{aligned}f(t,x) = {}& \mathbf{1}_{t\le\tau_x} f^{in}(x-tv)\exp\left(-\int_0^t a(x+(s-t)v)ds\right)\\ &+ \mathbf{1}_{t>\tau_x} f_b^-(t-\tau_x,x^*)\exp\left(-\int_{t-\tau_x}^t a(x+(s-t)v)ds\right)\end{aligned}$$

de sorte que, pour tout $\lambda > 0$,

$$\begin{aligned}F(\lambda,x) = {}& \int_0^{\tau_x} f^{in}(x-tv)\exp\left(-\lambda t - \int_0^t a(x-sv)ds\right)dt\\ &+ \int_{\tau_x}^{+\infty} f_b^-(t-\tau_x,x^*)\exp\left(-\lambda t - \int_0^{\tau_x} a(x-sv)ds\right)dt\,.\end{aligned}$$

Lorsque $f_b^- \equiv f_b^-(y)$ est indépendante de t, sa transformée de Laplace vaut, pour tout $\lambda > 0$,

$$F_b^-(\lambda, y) = \frac{1}{\lambda} f_b^-(y)\,, \quad y \in \partial\Omega^-\,,$$

de sorte que

$$\begin{aligned} F(\lambda, x) &= \int_0^{\tau_x} f^{in}(x - tv) \exp\left(-\lambda t - \int_0^t a(x - sv)ds\right) dt \\ &\quad + f_b^-(x^*) \exp\left(-\int_0^{\tau_x} a(x - sv)ds\right) \int_{\tau_x}^{+\infty} e^{-\lambda t} dt \\ &= \int_0^{\tau_x} f^{in}(x - tv) \exp\left(-\lambda t - \int_0^t a(x - sv)ds\right) dt \\ &\quad + \mathbf{1}_{\tau_x < +\infty} F_b^-(\lambda, x^*) \exp\left(-\lambda\tau_x - \int_0^{\tau_x} a(x - sv)ds\right)\,. \end{aligned}$$

Après ces remarques préliminaires, esquissons la théorie d'existence et unicité de la solution du problème au limites pour l'équation de transport stationnaire

$$\lambda F(x) + v \cdot \nabla F(x) + A(x)F(x) = Q(x)\,, \quad x \in \Omega\,.$$

Tout d'abord, précisons la notion de solution généralisée de l'équation de transport stationnaire.

Définition 2.6.1 *Soient Ω ouvert à bord de classe C^1 de $\mathbf{R}^N$ et $v \in \mathbf{R}^N \setminus \{0\}$. Soient $A \in C(\overline{\Omega})$, $\lambda \in \mathbf{R}$ et $Q \in C_b(\Omega)$. Une fonction mesurable F définie p.p. sur Ω est solution généralisée de l'équation de transport*

$$\lambda F(x) + v \cdot \nabla F(x) + A(x)F(x) = Q(x)\,, \quad x \in \Omega\,,$$

si et seulement si la fonction

$$s \mapsto F(x + sv)$$

est de classe C^1 p.p. en $x \in \Omega$, et vérifie

$$\lambda F(x + sv) + \frac{d}{ds} F(x + sv) + A(x + sv)F(x + sv) = Q(x + sv)\,, \quad x + sv \in \Omega\,,$$

p.p. en $x \in \Omega$.

Voici le résultat principal d'existence et unicité d'une solution généralisée pour l'équation de transport stationnaire.

Théorème 2.6.2 *Soient Ω ouvert à bord de classe C^1 de $\mathbf{R}^N$, et $v \in \mathbf{R}^N \setminus \{0\}$. Supposons que A est une fonction continue sur $\overline{\Omega}$ telle que*

$$A(x) \geq 0 \text{ pour tout } x \in \overline{\Omega}\,.$$

Soient $\lambda > 0$, $Q \in C_b(\Omega)$ et $F_b^- \in L^\infty(\partial\Omega^-)$. Alors le problème aux limites

$$\begin{cases} \lambda F(x) + v \cdot \nabla_x F(x) + A(x)F(x) = Q(x), \quad x \in \Omega, \\ F\big|_{\partial\Omega^-} = F_b^- \,, \end{cases}$$

admet une unique solution généralisée $F \in L^\infty(\Omega)$, qui est donnée par la formule

$$\begin{aligned} F(x) = &\int_0^{\tau_x} \exp\left(-\lambda s - \int_0^s A(x-\theta v)d\theta\right) Q(x-sv)ds \\ &+ \mathbf{1}_{\tau_x<+\infty} F_b^-(x^*) \exp\left(-\lambda\tau_x - \int_0^{\tau_x} A(x-\theta v)d\theta\right) \ \textit{p.p. en } x \in \Omega\,, \end{aligned} \tag{2.2}$$

où on rappelle que τ_x désigne le temps de sortie de Ω partant de $x \in \Omega$ avec la vitesse $-v$, tandis que $x^* = x - \tau_x v$.

Démonstration. Procédons par condition nécessaire.

Si F est une solution généralisée de l'équation de transport stationnaire, on a

$$\frac{d}{ds}F(x+sv) + (\lambda + A(x+sv))F(x+sv) = Q(x+sv)\,,$$

c'est-à-dire

$$\begin{aligned} &\frac{d}{ds}\left(F(x+sv)\exp\left(\lambda s + \int_0^s A(x+\theta v)d\theta\right)\right) \\ &\qquad = \exp\left(\lambda s + \int_0^s A(x+\theta v)d\theta\right) Q(x+sv) \end{aligned}$$

pour tout s tel que $x + sv \in \Omega$, p.p. en $x \in \Omega$. Intégrons chaque membre de cette égalité pour $s \in]-\tau_x, 0[$:

$$\begin{aligned} F(x) - &\lim_{s\to(-\tau_x)^+}\left(F(x+sv)\exp\left(\lambda s + \int_0^s A(x+\theta v)d\theta\right)\right) \\ &= \int_{-\tau_x}^0 \exp\left(\lambda s + \int_0^s A(x+\theta v)d\theta\right) Q(x+sv)ds \\ &= \int_0^{\tau_x} \exp\left(-\lambda s + \int_0^{-s} A(x+\theta v)d\theta\right) Q(x-sv)ds \\ &= \int_0^{\tau_x} \exp\left(-\lambda s - \int_0^s A(x-\theta v)d\theta\right) Q(x-sv)ds\,. \end{aligned}$$

Supposons dans un premier temps que $\tau_x = +\infty$. Alors, comme $F \in L^\infty(\Omega)$, $\lambda > 0$ et $A \geq 0$, pour presque tout $x \in \Omega$, il existe une suite $s_n \to -\infty$ telle que

$$|F(x+s_n v)| \exp\left(\lambda s_n + \int_0^{s_n} A(x+\theta v)d\theta\right) \leq \|F\|_{L^\infty(\Omega)} e^{\lambda s_n} \to 0$$

lorsque $s_n \to -\infty$. Par conséquent

$$F(x) = \int_0^{+\infty} \exp\left(-\lambda s - \int_0^s A(x-\theta v)d\theta\right) Q(x-sv)ds$$

p.p. en $x \in \Omega$ tel que $\tau_x = +\infty$.

Au contraire, si $\tau_x < +\infty$, alors

$$\lim_{s\to(-\tau_x)^+} \left(F(x+sv) \exp\left(\lambda s + \int_0^s A(x+\theta v) d\theta \right) \right) = F_b^-(x^*) \exp\left(-\lambda \tau_x - \int_0^{\tau_x} A(x-\theta v) d\theta \right) .$$

Donc

$$F(x) = \int_0^{\tau_x} \exp\left(-\lambda s - \int_0^s A(x-\theta v) d\theta \right) Q(x-sv) ds + \mathbf{1}_{\tau_x<+\infty} F_b^-(x^*) \exp\left(-\lambda \tau_x - \int_0^{\tau_x} A(x-\theta v) d\theta \right) .$$

On laisse au lecteur le soin de vérifier que la formule ci-dessus définit bien une solution généralisée de l'équation de transport stationnaire. ■

L'équation de transport stationnaire vérifie également une estimation L^∞ analogue au principe du maximum que nous avons déjà exposé dans le cas d'évolution.

Théorème 2.6.3 *Soient Ω ouvert à bord de classe C^1 de $\mathbf{R}^N$, un vecteur vitesse $v \in \mathbf{R}^N \setminus \{0\}$, et A, une fonction continue sur $\overline{\Omega}$ telle que*

$$A(x) \geq 0 \text{ pour tout } x \in \overline{\Omega} .$$

Soient $\lambda > 0$, $Q \in C_b(\Omega)$ et $F_b^- \in L^\infty(\partial\Omega^-)$. Alors la solution généralisée $F \in L^\infty(\Omega)$ du problème aux limites

$$\begin{cases} \lambda F(x) + v \cdot \nabla_x F(x) + A(x) F(x) = Q(x), \quad x \in \Omega, \\ F\big|_{\partial\Omega^-} = F_b^- , \end{cases}$$

vérifie l'estimation

$$\|F\|_{L^\infty(\Omega)} \leq \max\left(\frac{1}{\lambda} \|Q\|_{L^\infty(\Omega)}, \|F_b^-\|_{L^\infty(\partial\Omega^-)} \right) .$$

De plus, si $Q \geq 0$ p.p. sur Ω et $F_b^- \geq 0$ p.p. sur $\partial\Omega^-$, la solution F vérifie

$$F \geq 0 \quad \text{p.p. sur } \Omega .$$

Démonstration. Partons de la formule du théorème précédent donnant la solution généralisée F :

$$F(x) = \int_0^{\tau_x} \exp\left(-\lambda s - \int_0^s A(x-\theta v) d\theta \right) Q(x-sv) ds + \mathbf{1}_{\tau_x<+\infty} F_b^-(x^*) \exp\left(-\lambda \tau_x - \int_0^{\tau_x} A(x-\theta v) d\theta \right) ,$$

p.p. en $x \in \Omega$. Donc, p.p. en $x \in \Omega$,

$$\begin{aligned}
|F(x)| &\leq \int_0^{\tau_x} e^{-\lambda s}|Q(x-sv)|ds + |F_b^-(x^*)|e^{-\lambda\tau_x} \\
&\leq \|Q\|_{L^\infty(\Omega)} \int_0^{\tau_x} e^{-\lambda s} ds + \|F_b^-\|_{L^\infty(\partial\Omega^-)} e^{-\lambda\tau_x} \\
&= \|Q\|_{L^\infty(\Omega)} \frac{1-e^{-\lambda\tau_x}}{\lambda} + \|F_b^-\|_{L^\infty(\partial\Omega^-)} e^{-\lambda\tau_x} \\
&\leq \max\left(\frac{1}{\lambda}\|Q\|_{L^\infty(\Omega)}, \|F_b^-\|_{L^\infty(\partial\Omega^-)}\right).
\end{aligned}$$

La positivité de F lorsque $Q \geq 0$ p.p. sur Ω et $F_b^- \geq 0$ p.p. sur $\partial\Omega^-$ se lit de façon triviale sur la formule explicite donnant F. ■

2.7 Exercices

Exercice 2.1 (Cas où σ n'est pas constant) *Soit l'équation*

$$\frac{\partial f}{\partial t}(t,x) + v\frac{\partial f}{\partial x}(t,x) + \sigma(t,x)f(t,x) = S(t,x)\,, \quad x \in \mathbf{R}\,,\ t > 0\,,$$

pour une vitesse $v \in \mathbf{R}$ donnée et avec une condition initiale $f_0(x)$. Montrer que

$$f(t,x) = e^{-A(x,t,t)}f_0(x-tv) + \int_0^t e^{-A(x,t,s)}S(t-s,x-vs)ds\,,$$

où $A \equiv A(x,t,s)$ est une fonction à déterminer.

Indications : On peut utiliser le Théorème 2.1.3, et en particulier la formule (2.1) qui fournit évidemment la solution de cet exercice. Une autre approche consiste à dériver la formule ci-dessus par rapport aux variables t et x et à déterminer l'équation que doit vérifier A pour que f soit solution de l'équation de transport. Il ne reste plus qu'à montrer que la solution de cette équation est donnée par la formule du théorème :

$$A(x,t,s) = -\int_0^s \sigma(t-\tau, x-\tau v)d\tau\,.$$

Exercice 2.2 (Cas " stationnaire ") *On considère l'équation*

$$\frac{\partial f}{\partial t}(t,x) + v\frac{\partial f}{\partial x}(t,x) + \sigma f(t,x) = S(t,x)\,, \quad x \in \mathbf{R}\,,\ t > 0\,,$$

pour $x \in \mathbf{R}$, $v \in \mathbf{R}^*$ donné et σ constant. On cherche une solution de la forme

$$f(t,x) = e^{\lambda t}\widehat{f}(x).$$

1. *Montrer que dans ce cas S doit être de la forme*

$$S(t,x) = s(x)e^{\lambda t}\,.$$

2. *Quelle est l'équation satisfaite par $\widehat{f}$? Montrer, en supposant $\widehat{f}$ bornée, que*
$$\widehat{f}(x) = \frac{1}{v}\int_{-\infty}^{x} e^{-(\sigma+\lambda)\frac{x-y}{v}} s(y)dy$$
ou
$$\widehat{f}(x) = -\frac{1}{v}\int_{x}^{\infty} e^{(\sigma+\lambda)\frac{y-x}{v}} s(y)dy\,,$$
suivant le signe de $(\sigma+\lambda)/v$ (on suppose $\sigma+\lambda \neq 0$).
3. *Comment ces formules se généralisent-elles en dimension supérieure ?*

Indications : la première formule recherchée est à variables séparées : on l'obtient en insérant la forme de f directement dans l'équation. Les fonctions proposées au point 2 sont "évidemment" solutions : mais il faut vérifier l'intégrabilité, et c'est là que le signe de $(\sigma+\lambda)/v$ entre en ligne de compte. Pour la question 3, voir le Théorème 2.6.2.

Exercice 2.3 (Un modèle sans condition au bord) *On considère une population de cellules avec un modèle structuré en âge $a \geq 0$. L'inconnue est $n(t,a)$. Le vieillissement se contrôle par l'ajout journalier d'une substance chimique représenté par une fonction continue v vérifiant la condition*
$$0 \leq v(a) \leq 1\,, \quad a \geq 0\,.$$
L'équation vérifiée par n est
$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial a}(v(a)n) = 0\,.$$

1. *Interpréter les cas $v(a) \equiv 1$ et $v(a) \equiv 0$.*
2. *Soit*
$$N(t) = \int_0^{\infty} n(t,a)da$$
le nombre total de cellules au temps t. On supposera que n s'annule pour a suffisament grand. Montrer que N est croissant en temps, et constant si $n(t,0) = 0$. Cela est-il normal ? Aurait-on eu le même comportement avec un modèle de la forme
$$\frac{\partial n}{\partial t}(t,a) + v(a)\frac{\partial n}{\partial a}(t,a) = 0\,?$$
3. *Déterminer n en fonction de la condition initiale n_0, le nombre de cellules par tranche d'âge à $t = 0$.*
4. *Montrer que si $0 \leq v(a) \leq Ca$ pour a proche de 0, alors le modèle n'a pas besoin de condition au bord de type natalité donnée ($n(0,t) = \ldots$) pour être bien posé.*

5. *On suppose* $v(a) > 0$ *pour* $a > 0$. *Montrer qu'on a le même résultat sous la condition supplémentaire*

$$\int_0^1 \frac{1}{v(a)} da = \infty \,.$$

Indications : comme la fonction v est continue en 0, les conditions des questions 4 et 5 impliquent que la vitesse $v(0)$ est nulle. Dans ces conditions cela explique pourquoi il n'y a pas besoin de condition "au bord" en $a = 0$ pour que le modèle soit bien posé. La condition du 5 peut s'obtenir en calculant le temps pour atteindre 0 le long d'une caractéristique. De nouveau, on constate que, si ce temps est infini, le bord ne peut influer sur la solution.

Chapitre 3

L'équation de Boltzmann linéaire

Ce chapitre est consacré à l'analyse mathématique de l'équation de Boltzmann linéaire, qui est l'équation fondamentale du transport de particules pour tous les exemples de théories cinétiques présentés au début de ce cours.

Afin de prendre en compte l'extrême diversité de ces modèles, nous considèrerons cette équation de Boltzmann linéaire sous la forme

$$\left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(t,x,v) + a(t,x,v) f(t,x,v) = \int_{\mathbf{R}^N} k(t,x,v,w) f(t,x,w) d\mu(w)\,,$$

où l'inconnue est une densité de particules $f \equiv f(t,x,v) \geq 0$ se trouvant à l'instant t à la position x et animées de la vitesse v, et où $a \equiv a(t,x,v)$ ainsi que $k \equiv k(t,x,v,w)$ sont des fonctions continues données.

Comme l'équation de Boltzmann linéaire peut être utilisée dans des contextes extrêmement divers, il y a une grande variété de choix possibles de l'ensemble des vitesses v admissibles selon le modèle considéré — et par conséquent pour ce qui est de l'intégration par rapport à la variable v de vitesse.

La notation

$$\int_{\mathbf{R}^N} \phi(w) d\mu(w)$$

désignera donc, selon le modèle considéré [1] :

(a) l'intégrale de Lebesgue avec une fonction "poids" $E \equiv E(w) \geq 0$ mesurable sur $\mathbf{R}^N$:

$$\int_{\mathbf{R}^N} \phi(w) E(w) dw$$

pour toute fonction $\phi \in L^1(\mathbf{R}^N, Edw)$; ou bien

1. Le lecteur ayant quelque familiarité avec la théorie de la mesure pourra résumer les exemples a)-e) ci-dessous en supposant que μ est une mesure de Radon positive sur $\mathbf{R}^N$. Toutefois, aucune connaissance en théorie de la mesure n'est nécessaire pour la compréhension de ce chapitre.

(b) l'intégrale de Lebesgue

$$\int_{\mathcal{V}} \phi(w)dw$$

sur $\mathcal{V}$, boule ou couronne sphérique de $\mathbf{R}^N$, pour toute fonction $\phi \in L^1(\mathcal{V})$; ou bien encore
(c) l'intégrale de Lebesgue

$$\int_{\mathcal{V}} \phi(w)dw$$

sur $\mathcal{V}$, réunion finie de boules ou couronnes sphériques concentriques de $\mathbf{R}^N$, pour toute fonction $\phi \in L^1(V)$; ou bien encore
(d) l'intégrale de surface

$$\int_{\mathbf{S}^{N-1}} \phi(w)dw$$

où dw désigne l'élément de surface sur la sphère unité $\mathbf{S}^{N-1}$ de $\mathbf{R}^N$; ou enfin
(e) la somme

$$\sum_{v\in\mathcal{V}} \mu_w \phi(w)$$

où $\mathcal{V}$ est un ensemble fini ou dénombrable de $\mathbf{R}^N$, et $(\mu_w)_{w\in\mathcal{V}}$ une famille de réels strictement positifs ; mais aussi
(f) toute combinaison linéaire à coefficients positifs des exemples (a)-(e) ci-dessus.

Le cas (c) correspond par exemple au modèle du transport multigroupe, le cas d) à tous les modèles mettant en jeu des particules monocinétiques — comme par exemple le transfert radiatif, tandis que le cas (e) correspond à la discrétisation d'une équation cinétique en la seule variable de vitesse. Ce dernier cas n'est pas sans intérêt, même si la résolution numérique d'une équation de Boltzmann linéaire nécessite une discrétisation de toutes les variables t, x, v et pas seulement de la variable v, car il permet d'étudier par exemple l'influence de la seule discrétisation en vitesse sur la qualité de l'approximation numérique. On remarquera que les exemples (b) et (c) sont des cas particuliers de (a) (choisir pour E la fonction indicatrice $\mathbf{1}_{\mathcal{V}}$ de $\mathcal{V}$).

Dans ce chapitre, on étudiera notamment l'existence et l'unicité de la solution du problème de Cauchy pour l'équation de Boltzmann linéaire ci-dessus, ainsi que la théorie du problème aux limites pour cette même équation et sa variante stationnaire.

Contrairement au cas de l'équation de transport libre étudiée au chapitre précédent, l'équation de Boltzmann met en jeu des phénomènes d'échange par rapport au vecteur vitesse v : c'est le rôle du noyau intégral $k(t, x, v, w)$ que de modéliser ce type d'échange. Il n'est donc plus possible de considérer v comme un paramètre, comme cela a été fait dans le chapitre précédent.

3.1 Le problème de Cauchy pour l'équation de Boltzmann linéaire

Nous allons appliquer les méthodes de résolution de l'équation de transport développées dans le chapitre précédent au cas de l'équation de Boltzmann linéaire la plus générale.

On considère donc l'équation intégro-différentielle

$$\left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(t,x,v) + a(t,x,v) f(t,x,v) \\ = \int_{\mathbf{R}^N} k(t,x,v,w) f(t,x,w) d\mu(w) + Q(t,x,v)\,,$$

d'inconnue la fonction de distribution $f \equiv f(t,x,v)$. On notera dans toute la suite

$$\mathcal{K}f(t,x,v) = \int_{\mathbf{R}^N} k(t,x,v,w) f(t,x,w) d\mu(w)\,, \tag{3.1}$$

et on supposera toujours dans ce chapitre que

$$\begin{cases} a \equiv a(t,x,v) \geq 0 & \text{(taux d'absorption)}, \\ k \equiv k(t,x,v,w) \geq 0 & \text{(taux de transition } w \to v), \\ \mu \geq 0 & \text{(mesure de Radon sur } \mathbf{R}^N). \end{cases}$$

On a vu au chapitre précédent tout l'intérêt que présente, pour la résolution des équations de transport, la notion de solution généralisée — en particulier pour la théorie du problème aux limites. En effet, pour que la solution d'une équation de transport dans un domaine à bord régulier de $\mathbf{R}^N$, même convexe, soit de classe C^1, il faut imposer aux conditions initiales et sur le bord de vérifier des conditions de compatibilité assez lourdes. On a vu également que, lorsque le domaine spatial où l'équation de transport est posée n'est pas convexe, la solution du problème aux limites comporte en général des discontinuités.

A partir de maintenant, nous allons donc nous contenter d'étudier le problème de Cauchy dans le cadre des solutions généralisées.

Il existe plusieurs notions possibles de solution généralisée pour l'équation de Boltzmann linéaire. Voici celle que nous utiliserons.

Définition 3.1.1 *Soit $Q \equiv Q(t,x,v)$, fonction continue sur $]0,T[\times\Omega \times \mathbf{R}^N$. Une fonction $f \equiv f(t,x,v)$ continue sur $]0,T[\times\Omega \times \mathbf{R}^N$ est solution généralisée de l'équation de Boltzmann linéaire*

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + af = \mathcal{K}f + Q$$

si et seulement si, pour tout $(t,x,v) \in]0,T[\times\Omega \times \mathbf{R}^N$, la fonction

$$s \mapsto f(t+s, x+sv, v) \text{ est de classe } C^1 \text{ pour } x + sv \in \Omega,$$

et vérifie, pour tout $s \in \mathbf{R}$ *t.q.* $x + sv \in \Omega$,

$$\frac{d}{ds} f(t+s, x+sv, v) + a(t+s, x+sv, v) f(t+s, x+sv, v) \\ = (\mathcal{K}f + Q)(t+s, x+sv, v)\,.$$

A vrai dire, ce n'est pas la notion la plus générale de solution que l'on pourrait imaginer. Par exemple, il suffirait de demander que la fonction

$$s \mapsto f(t+s, x+sv, v)$$

soit absolument continue (c'est-à-dire primitive d'une fonction L^1), ou au moins lipschitzienne, (c'est-à-dire, de façon équivalente, primitive d'une fonction L^∞). Se restreindre au cas où cette fonction est de classe C^1 permet d'éviter le recours à des arguments quelque peu techniques sur l'intégration au sens de Lebesgue.

En contre-partie, cela exige que la solution f soit non seulement bornée mais encore continue — ce cadre excluant la possibilité de traiter le cas d'équations de Boltzmann linéaires posées dans un ouvert non convexe. Toutefois, les démonstrations que nous allons présenter ci-dessous dans le cadre de fonctions continues bornées s'étendent très facilement au cas de fonctions L^∞, à condition de savoir que toute fonction lipschitzienne sur un intervalle de $\mathbf{R}$ est dérivable p.p., et la primitive d'une fonction de L^∞ (théorème de Rademacher : cf. par exemple l'exercice 10 du chapitre 7 dans [46]).

3.1.1 Existence et unicité pour le problème de Cauchy

Commençons par un résultat d'existence et unicité de la solution généralisée du problème de Cauchy pour l'équation de Boltzmann linéaire posée dans l'espace des phases $\mathbf{R}^N_x \times \mathbf{R}^N_v$.

Théorème 3.1.2 *Soient une donnée initiale* $f^{in} \equiv f^{in}(x,v) \in C_b(\mathbf{R}^N \times \mathbf{R}^N)$ *et un terme source* $Q \equiv Q(t,x,v) \in C_b([0,T] \times \mathbf{R}^N \times \mathbf{R}^N)$. *Supposons que*

$$0 \leq a \in C_b([0,T] \times \mathbf{R}^N_x \times \mathbf{R}^N_v) \text{ et } 0 \leq k \in C_b([0,T] \times \mathbf{R}^N_x \times \mathbf{R}^N_v \times \mathbf{R}^N_w)\,,$$

et que de plus

$$\sup_{(t,x,v) \in [0,T] \times \mathbf{R}^N \times \mathbf{R}^N} \int_{\mathbf{R}^N} k(t,x,v,w) d\mu(w) < +\infty\,.$$

Alors le problème

$$\begin{cases} \frac{\partial f}{\partial t} + v \cdot \nabla_x f + af = \mathcal{K}f + Q\,, \quad t \in]0,T[\,,\ x,v \in \mathbf{R}^N \\ f\big|_{t=0} = f^{in} \end{cases}$$

admet une unique solution généralisée $f \in C_b([0,T] \times \mathbf{R}^N \times \mathbf{R}^N)$.

Contrairement au cas de l'équation de transport étudiée au chapitre précédent, on ne dispose pas, pour l'équation de Boltzmann linéaire, de formule explicite commode en donnant la solution. Il est toujours possible d'écrire la solution sous forme d'une série dont le n-ième terme général est une intégrale sur un simplexe de dimension croissant avec n, en itérant sur la formule de Duhamel (cf. la démonstration du Théorème 3.1.2 pour une présentation équivalente quoique légèrement différente). Une telle formule s'interprète très simplement comme l'analogue d'une formule des caractéristiques pour l'équation de transport dans le cadre des processus stochastiques — voir la section 3.3.

Plus précisément, la formule découlant de l'application de la méthode des caractéristiques ne donne qu'une relation implicite portant sur la solution généralisée f — autrement dit une équation intégrale vérifiée par f.

En réalité, la méthode des caractéristiques conduit aux deux formulations intégrales suivantes pour l'équation de Boltzmann linéaire :

1ère formulation intégrale : pour tout $(t,x,v) \in [0,T] \times \mathbf{R}^N \times \mathbf{R}^N$,

$$f(t,x,v) = \exp\left(-\int_0^t a(s,x+(s-t)v,v)ds\right) f^{in}(x-tv,v)$$
$$+\int_0^t (\mathcal{K}f+Q)(s,x+(s-t)v,v)\exp\left(-\int_s^t a(\theta,x+(\theta-t)v,v)d\theta\right)ds\,;$$

2ème formulation intégrale : pour tout $(t,x,v) \in [0,T] \times \mathbf{R}^N \times \mathbf{R}^N$,

$$f(t,x,v) = f^{in}(x-tv,v) + \int_0^t (\mathcal{K}f+Q-af)(s,x+(s-t)v,v)ds\,.$$

Pour obtenir la première formulation, on applique la méthode des caractéristiques à l'équation de transport

$$\left(\frac{\partial}{\partial t} + v\cdot\nabla_x\right) f + af = S$$

avec

$$S = \mathcal{K}f + Q\,.$$

Autrement dit, on écrit que

$$\frac{d}{ds}\left(f(t+s,x+sv,v)\exp\left(\int_0^s a(t+\theta,x+\theta v,v)d\theta\right)\right)$$
$$= (\mathcal{K}f+Q)(t+s,x+sv,v)\exp\left(\int_0^s a(t+\theta,x+\theta v,v)d\theta\right)\,.$$

Pour obtenir la seconde formulation, on applique cette fois la méthode des caractéristiques à l'équation de transport

$$\left(\frac{\partial}{\partial t} + v\cdot\nabla_x\right) f = S\,,$$

avec

$$S = \mathcal{K}f - af + Q\,.$$

C'est-à-dire que l'on écrit que

$$\frac{d}{ds} f(t+s, x+sv, v) = (\mathcal{K}f - af + Q)(t+s, x+sv, v)\,.$$

La démonstration du théorème 3.1.2 ci-dessus est basée sur un argument de point fixe pour la 1ère formulation intégrale. Autrement dit, on cherche une fonction $f \in C_b([0,T] \times \mathbf{R}^N \times \mathbf{R}^N)$ telle que

$$f = F[f^{in}, Q] + \mathcal{T}f\,,$$

en posant

$$\mathcal{T}g(t,x,v) := \int_0^t \mathcal{K}g(s, x+(s-t)v, v) \exp\left(-\int_s^t a(\theta, x+(\theta-t)v, v)d\theta\right) ds\,,$$

et

$$\begin{aligned} F[f^{in}, Q](t,x,v) = {} & f^{in}(x-tv, v) \exp\left(-\int_0^t a(\theta, x+(\theta-t)v, v)d\theta\right) \\ & + \int_0^t Q(s, x+(s-t)v, v) \exp\left(-\int_s^t a(\theta, x+(\theta-t)v, v)d\theta\right) ds\,. \end{aligned}$$

Démonstration. Puisqu'il s'agit de trouver un point fixe à l'équation

$$f = F[f^{in}, Q] + \mathcal{T}f\,,$$

suivant la méthode d'itérations successives de Picard, nous allons chercher la solution généralisée f sous la forme

$$f := \sum_{n \geq 0} \mathcal{T}^n F[f^{in}, Q]\,.$$

Introduisons l'espace fonctionnel

$$\mathcal{X}_T := C_b([0,T] \times \mathbf{R}^N \times \mathbf{R}^N)\,,$$

qui est un espace de Banach pour la norme de la convergence uniforme

$$\|\phi\|_{\mathcal{X}_T} := \sup_{(t,x,v) \in [0,T] \times \mathbf{R}^N \times \mathbf{R}^N} |\phi(t,x,v)|\,.$$

Les hypothèses faites sur le noyau de transition k montrent que $\mathcal{K}$ est une application linéaire continue sur $\mathcal{X}_T$, avec

$$\|\mathcal{K}\phi\|_{\mathcal{X}_T} \leq M \|\phi\|_{\mathcal{X}_T}$$

pour tout $\phi \in \mathcal{X}_T$, où l'on a noté

$$M := \sup_{(t,x,v)\in[0,T]\times\mathbf{R}^N\times\mathbf{R}^N} \int_{\mathbf{R}^N} k(t,x,v,w)d\mu(w) < +\infty\,.$$

Définissons pour tout $\phi \in C([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)$ la fonction

$$\mathcal{S}\phi(t_1,t_2,x,v) := \int_{t_1}^{t_2} \phi(\theta, x+(\theta-t_2)v, v)d\theta\,.$$

D'après le théorème de continuité des intégrales à paramètres, la fonction $\mathcal{S}\phi$ est continue sur $[0,T]^2\times(\mathbf{R}^N)^2$, et, pour tout $g \in C([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)$, la fonction

$$\mathcal{T}g(t,x,v) = \int_0^t \mathcal{K}g(s,x+(s-t)v,v)\exp(-\mathcal{S}a(s,t,x,v))ds$$

est continue sur $[0,T]\times\mathbf{R}^N\times\mathbf{R}^N$. On démontre de façon analogue que la fonction $F[f^{in},Q]$ définie plus haut est continue sur $[0,T]\times\mathbf{R}^N\times\mathbf{R}^N$.

Comme $a \geq 0$,

$$0 \leq \exp(\mathcal{S}(-a)(s,t,x,v)) \leq 1$$

pour tout $s,t\in[0,T]$ et $x,v\in\mathbf{R}^N$.

Donc, pour tout $g\in\mathcal{X}_T$

$$|\mathcal{T}g(t,x,v)| \leq \int_0^t |\mathcal{K}g(s,x+(s-t)v,v)|ds \leq MT\|g\|_{\mathcal{X}_T}\,,$$

de sorte que

$$\|\mathcal{T}g\|_{\mathcal{X}_T} \leq MT\|g\|_{\mathcal{X}_T}\,.$$

De même

$$\|F[f^{in},Q]\|_{\mathcal{X}_T} \leq \|f^{in}\|_{L^\infty(\mathbf{R}^N\times\mathbf{R}^N)} + T\|Q\|_{\mathcal{X}_T}\,.$$

Evidemment, par une récurrence immédiate,

$$\mathcal{T}^nF[f^{in},Q]\in\mathcal{X}_T \quad \text{ avec } \|\mathcal{T}^nF[f^{in},Q]\|_{\mathcal{X}_T} \leq (MT)^n\|F[f^{in},Q]\|_{\mathcal{X}_T}$$

pour tout $n\in\mathbf{N}$, estimation que l'on va améliorer.

En effet, pour tout $g\in\mathcal{X}_T$, l'on a

$$\begin{aligned}\|\mathcal{T}^ng(t,\cdot,\cdot)\|_{L^\infty(\mathbf{R}^N\times\mathbf{R}^N)} &\leq M\int_0^t \|\mathcal{T}^{n-1}g(t_1,\cdot,\cdot)\|_{L^\infty(\mathbf{R}^N\times\mathbf{R}^N)}dt_1\\ &\leq M^n\int_0^t\int_0^{t_1}\cdots\int_0^{t_{n-1}} \|g(t_n,\cdot,\cdot)\|_{L^\infty(\mathbf{R}^N\times\mathbf{R}^N)}dt_n\ldots dt_1\\ &\leq \frac{(Mt)^n}{n!}\|g\|_{L^\infty(\mathbf{R}^N\times\mathbf{R}^N)}\,.\end{aligned}$$

Par conséquent, pour tout $n \geq 1$, l'application linéaire $\mathcal{T}^n$ est continue de l'espace de Banach $\mathcal{X}_T$ dans lui-même et sa norme vérifie, pour tout $n \in \mathbf{N}$

$$\|\mathcal{T}^n\|_{\mathcal{L}(\mathcal{X}_T)} \leq \frac{(MT)^n}{n!}\,.$$

En particulier, d'après ce qui précède

$$\begin{aligned}\|\mathcal{T}^n F[f^{in},Q]\|_{\mathcal{X}_T} &\leq \frac{(MT)^n}{n!}\|F[f^{in},Q]\|_{\mathcal{X}_T}\\ &\leq \frac{(MT)^n}{n!}(\|f^{in}\|_{L^\infty(\mathbf{R}^N\times\mathbf{R}^N)} + T\|Q\|_{\mathcal{X}_T})\,,\end{aligned}$$

de sorte que la série

$$\sum_{n\geq 0}\mathcal{T}^n F[f^{in},Q]$$

converge normalement dans $\mathcal{X}_T$. L'espace $\mathcal{X}_T$ étant complet pour la norme $\|\cdot\|_{\mathcal{X}_T}$ de la convergence uniforme, on en déduit que cette série converge dans $\mathcal{X}_T$, et on pose

$$f = \sum_{n\geq 0}\mathcal{T}^n F[f^{in},Q] \in \mathcal{X}_T\,.$$

Evidemment, comme l'application $\mathcal{T}$ est linéaire et continue sur $\mathcal{X}_T$, l'on a

$$\begin{aligned}f &= F[f^{in},Q] + \sum_{n\geq 1}\mathcal{T}^n F[f^{in},Q]\\ &= F[f^{in},Q] + \mathcal{T}\left(\sum_{n\geq 0}\mathcal{T}^n F[f^{in},Q]\right) = F[f^{in},Q] + \mathcal{T}f\,,\end{aligned}$$

ce qui montre que la fonction f définie par la série ci-dessus vérifie bien la 1ère formulation intégrale de l'équation de Boltzmann linéaire sur $[0,T]\times\mathbf{R}^N\times\mathbf{R}^N$.

En particulier, posons

$$S = \mathcal{K}f + Q \in C_b([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)\,;$$

on vient de montrer que

$$\begin{aligned}f(t,x,v) &= \exp\left(-\int_0^t a(s,x+(s-t)v,v)ds\right) f^{in}(x-tv,v)\\ &+ \int_0^t S(s,x+(s-t)v,v)\exp\left(-\int_s^t a(\theta,x+(\theta-t)v,v)d\theta\right)ds\,,\end{aligned}$$

ce qui, d'après le Théorème 2.3.2 (avec $\tau_x = +\infty$), équivaut à dire que, pour tout $v\in\mathbf{R}^N$, la fonction $(t,x)\mapsto f(t,x,v)$ est solution généralisée du problème de Cauchy

$$\begin{cases}\left(\frac{\partial}{\partial t} + v\cdot\nabla_x\right) f(\cdot,\cdot,v) + a(\cdot,\cdot,v)f(\cdot,\cdot,v) = S(\cdot,\cdot,v)\,,\\ f(\cdot,\cdot,v)\big|_{t=0} = f^{in}(\cdot,v)\,.\end{cases}$$

Donc la fonction f définie par la série ci-dessus est bien solution généralisée de l'équation de Boltzmann linéaire.

Montrons que c'est la seule.

S'il existait deux solutions généralisées f_1 et f_2 du problème de Cauchy ci-dessus pour l'équation de Boltzmann linéaire, l'on aurait

$$\begin{cases} f_1 = F[f^{in}, Q] + \mathcal{T} f_1 \,, \\ f_2 = F[f^{in}, Q] + \mathcal{T} f_2 \,, \end{cases}$$

de sorte que, par linéarité de l'application $\mathcal{T}$, l'on aurait

$$(f_2 - f_1) = \mathcal{T}(f_2 - f_1) \,.$$

On conclurait alors que $f_1 = f_2$ grâce au lemme ci-dessous. ■

Lemme 3.1.3 *Avec les mêmes hypothèses et notations que dans le théorème ci-dessus, l'unique élément $g \in \mathcal{X}_T$ vérifiant*

$$g = \mathcal{T} g$$

est

$$g = 0 \,.$$

Démonstration. Si $g \in \mathcal{X}_T$ vérifie

$$g = \mathcal{T} g \,,$$

alors

$$g = \mathcal{T} g = \mathcal{T}(\mathcal{T} g) = \ldots = \mathcal{T}^n g$$

pour tout $n \in \mathbf{N}$.

Or on a établi dans la démonstration du théorème que, pour tout $n \in \mathbf{N}$,

$$\| \mathcal{T}^n \|_{\mathcal{L}(\mathcal{X}_T)} \leq \frac{(MT)^n}{n!} \,.$$

Comme

$$\frac{(MT)^n}{n!} \to 0 \text{ pour tout } T > 0 \text{ lorsque } n \to +\infty \,,$$

on conclut que

$$\| g \|_{\mathcal{X}_T} \leq \| \mathcal{T}^n \|_{\mathcal{L}(\mathcal{X}_T)} \| g \|_{\mathcal{X}_T} \leq \frac{(MT)^n}{n!} \| g \|_{\mathcal{X}_T} \to 0$$

lorsque $n \to +\infty$, d'où le résultat. ■

La démonstration du Théorème 3.1.2 fournit en particulier une expression explicite de l'unique solution généralisée du problème de Cauchy pour l'équation de Boltzmann linéaire

$$\begin{cases} \frac{\partial f}{\partial t} + v \cdot \nabla_x f + a f = \mathcal{K} f + Q \,, \quad t \in]0, T[\,, \ x, v \in \mathbf{R}^N \\ f\big|_{t=0} = f^{in} \,. \end{cases}$$

Cette solution est en effet donnée par la formule, parfois appelée "série de Duhamel",

$$f = \sum_{n\geq 0} \mathcal{T}^n F[f^{in}, Q]\,,$$

et chaque terme de la série ci-dessus est explicite en fonction de la donnée initiale f^{in} et du terme source Q. Comme nous l'avons dit plus haut, nous verrons dans la section 3.3 comment interpréter l'expression ci-dessus comme l'analogue de la formule des caractéristiques pour l'équation du transport. La différence entre la formule ci-dessus donnnant la solution de l'équation de Boltzmann linéaire et la formule des caractéristiques pour l'équation de transport libre est que, dans le cas de l'équation de Boltzmann linéaire, les caractéristiques sont des lignes brisées aléatoires.

3.1.2 Estimation L^∞ pour le problème de Cauchy

Comme pour l'équation de transport, il est important de savoir estimer la solution généralisée d'un problème de Cauchy pour l'équation de Boltzmann linéaire à partir de bornes sur sa donnée initiale et sur le terme source. En réalité, il est encore plus important de disposer de ces informations dans le cas de l'équation de Boltzmann linéaire, puisque la série de Duhamel donnant la solution de l'équation de Boltzmann linéaire en fonction de la donnée initiale et du terme source est nettement plus complexe — et donc moins facilement utilisable — que la formule des caractéristiques donnant la solution de l'équation de transport.

Proposition 3.1.4 *Soient une donnée initiale $f^{in} \equiv f^{in}(x,v) \in C_b(\mathbf{R}^N \times \mathbf{R}^N)$ et un terme source $Q \equiv Q(t,x,v) \in C_b([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)$. Supposons que*

$$0 \leq a \in C_b([0,T]\times\mathbf{R}^N_x\times\mathbf{R}^N_v) \text{ et } 0 \leq k \in C_b([0,T]\times\mathbf{R}^N_x\times\mathbf{R}^N_v\times\mathbf{R}^N_w)\,,$$

et que de plus

$$\sup_{(t,x,v)\in[0,T]\times\mathbf{R}^N\times\mathbf{R}^N} \int_{\mathbf{R}^N} k(t,x,v,w)d\mu(w) < +\infty\,.$$

Si on a

$$f^{in} \geq 0 \text{ sur } \mathbf{R}^N\times\mathbf{R}^N \text{ et } Q \geq 0 \text{ sur } [0,T]\times\mathbf{R}^N\times\mathbf{R}^N\,,$$

la solution généralisée $f \in C_b([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)$ du problème de Cauchy

$$\begin{cases} \frac{\partial f}{\partial t} + v\cdot\nabla_x f + af = \mathcal{K}f + Q\,, \quad t\in]0,T[\,,\ x,v\in\mathbf{R}^N\,,\\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

vérifie

$$f \geq 0 \text{ sur } [0,T]\times\mathbf{R}^N\times\mathbf{R}^N\,.$$

Ce résultat est évidemment très naturel du point de vue physique, puisque f est censée représenter la densité de nombre des particules dans l'espace des phases.

Démonstration. Reprenons les notations de la démonstration du Théorème 3.1.2.

La solution généralisée f a été obtenue comme somme de la série de Duhamel

$$f = \sum_{n\geq 0} \mathcal{T}^n F[f^{in}, Q]\,,$$

où on rappelle que

$$\mathcal{T}g(t,x,v) := \int_0^t \mathcal{K}g(s, x+(s-t)v, v) \exp\left(-\int_s^t a(\theta, x+(\theta-t)v, v)d\theta\right) ds\,,$$

et

$$\begin{aligned} F[f^{in}, Q](t,x,v) &= f^{in}(x-tv,v)\exp\left(-\int_0^t a(\theta, x+(\theta-t)v, v)d\theta\right) \\ &\quad + \int_0^t Q(s, x+(s-t)v, v)\exp\left(-\int_s^t a(\theta, x+(\theta-t)v, v)d\theta\right) ds\,. \end{aligned}$$

D'une part, si $f^{in} \geq 0$ sur $\mathbf{R}^N \times \mathbf{R}^N$ et $Q \geq 0$ sur $[0,T]\times\mathbf{R}^N\times\mathbf{R}^N$, la définition ci-dessus montre que

$$F[f^{in}, Q] \geq 0 \text{ sur } [0,T]\times\mathbf{R}^N\times\mathbf{R}^N\,.$$

D'autre part, comme $k \equiv k(t,x,v,w) \geq 0$ sur $[0,T]\times\mathbf{R}^N_x\times\mathbf{R}^N_v\times\mathbf{R}^N_w$,

$$g \geq 0 \text{ sur } [0,T]\times\mathbf{R}^N_x\times\mathbf{R}^N_v \Rightarrow \mathcal{K}g \geq 0 \text{ sur } [0,T]\times\mathbf{R}^N_x\times\mathbf{R}^N_v\,.$$

Donc

$$g \geq 0 \text{ sur } [0,T]\times\mathbf{R}^N_x\times\mathbf{R}^N_v \Rightarrow \mathcal{T}g \geq 0 \text{ sur } [0,T]\times\mathbf{R}^N_x\times\mathbf{R}^N_v\,,$$

de sorte que, par une récurrence immédiate,

$$\mathcal{T}^n F[f^{in}, Q] \geq 0 \text{ sur } [0,T]\times\mathbf{R}^N\times\mathbf{R}^N$$

pour tout $n \geq 0$.

La solution généralisée f est donc positive comme somme de la série à termes positifs

$$f = \sum_{n\geq 0} \mathcal{T}^n F[f^{in}, Q]\,,$$

d'après la démonstration du Théorème 3.1.2. ■

Remarque : La solution généralisée f vérifie aussi

$$\begin{aligned} f(t,x,v) &\leq \sum_{n\geq 0} \|\mathcal{T}^n\|_{\mathcal{L}(\mathcal{X}_T)} \|F[f^{in}, Q]\|_{\mathcal{X}_T} \leq \sum_{n\geq 0} \frac{(Mt)^n}{n!} \|F[f^{in}, Q]\|_{\mathcal{X}_T} \\ &\leq (\|f^{in}\|_{L^\infty(\mathbf{R}^N\times\mathbf{R}^N)} + T\|Q\|_{L^\infty([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)})e^{Mt}\,, \end{aligned}$$

avec

$$M = \sup_{(t,x,v)\in[0,T]\times\mathbf{R}^N\times\mathbf{R}^N} \int_{\mathbf{R}^N} k(x,v,w)d\mu(w)\,,$$

mais cette estimation peut être améliorée, comme on va le voir.

Proposition 3.1.5 *Soient une donnée initiale $f^{in} \equiv f^{in}(x,v) \in C_b(\mathbf{R}^N \times \mathbf{R}^N)$ et un terme source $Q \equiv Q(t,x,v) \in C_b([0,T] \times \mathbf{R}^N \times \mathbf{R}^N)$. Supposons que*

$$0 \le a \in C_b([0,T] \times \mathbf{R}^N_x \times \mathbf{R}^N_v) \text{ et } 0 \le k \in C_b([0,T] \times \mathbf{R}^N_x \times \mathbf{R}^N_v \times \mathbf{R}^N_w)\,,$$

et que de plus

$$\sup_{(t,x,v)\in[0,T]\times\mathbf{R}^N\times\mathbf{R}^N} \int_{\mathbf{R}^N} k(t,x,v,w)d\mu(w) < +\infty\,.$$

Alors la solution généralisée $f \in C_b([0,T] \times \mathbf{R}^N \times \mathbf{R}^N)$ du problème de Cauchy

$$\begin{cases} \frac{\partial f}{\partial t} + v\cdot\nabla_x f + af = \mathcal{K}f + Q\,, \quad t \in]0,T[\,,\ x,v \in \mathbf{R}^N\,, \\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

vérifie, pour tout $(t,x,v) \in [0,T] \times \mathbf{R}^N \times \mathbf{R}^N$,

$$f(t,x,v) \le (\|f^{in}\|_{L^\infty(\mathbf{R}^N\times\mathbf{R}^N)} + T\|Q\|_{L^\infty([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)})e^{Dt}\,,$$

où

$$D = \sup_{(t,x,v)\in[0,T]\times\mathbf{R}^N\times\mathbf{R}^N} \left(\int_{\mathbf{R}^N} k(t,x,v,w)d\mu(w) - a(t,x,v)\right)_+\,,$$

—avec la notation habituelle $z_+ = \max(z,0)$.

La différence entre cette majoration et celle de la remarque précédente est évidente. Dans la majoration de la remarque ci-dessus, le terme d'absorption af au membre de gauche de l'équation de Boltzmann linéaire ne joue aucun rôle, et l'estimation finale ne fait intervenir que la constante $M = \|\mathcal{K}1\|_{L^\infty}$, qui ne dépend que du terme de scattering. Au contraire, la majoration du théorème fait intervenir le taux d'absorption a (et plus précisément, l'équilibre entre absorption et scattering) à travers la constante $D = \|(\|\mathcal{K}1 - a)_+\|_{L^\infty}$ intervenant dans l'estimation finale de la solution.

L'intérêt de cette majoration par rapport à celle de la remarque précédente est qu'on peut avoir à estimer la solution d'une équation de transport dans des situations asymptotiques où

$$\mathcal{K}1 = \int_{\mathbf{R}^N} k(t,x,v,w)d\mu(w) \gg 1 \text{ et } a \gg 1\,,$$

mais où

$$(\mathcal{K}1 - a)_+ = O(1)\,.$$

Ce genre de situation est typique des cas où l'approximation de l'équation de Boltzmann linéaire par une équation de diffusion est justifiée — et d'ailleurs nous verrons que la Proposition 3.1.5 joue un rôle absolument crucial dans la démonstration de cette approximation.

Cas particuliers importants :

1) Supposons que l'absorption domine la création de particules, c'est-à-dire que

$$\mathcal{K}1(t,x,v) = \int_{\mathbf{R}^N} k(t,x,v,w)d\mu(w) \leq a(t,x,v)\,.$$

Alors, pour tout $(t,x,v) \in [0,T] \times \mathbf{R}^N \times \mathbf{R}^N$

$$f(t,x,v) \leq \|f^{in}\|_{L^\infty(\mathbf{R}^N\times\mathbf{R}^N)} + T\|Q\|_{L^\infty([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)}\,.$$

2) Principe du maximum faible : si l'absorption domine la création de particules et que de plus $Q = 0$, alors

$$f(t,x,v) \leq \|f^{in}\|_{L^\infty(\mathbf{R}^N\times\mathbf{R}^N)}\,, \quad \text{pour tout } (t,x,v) \in [0,T] \times \mathbf{R}^N \times \mathbf{R}^N\,.$$

Passons maintenant à la démonstration de la proposition ci-dessus.

Démonstration. Soit $Z \equiv Z(t) \in \mathbf{R}$ la solution de l'équation différentielle ordinaire

$$\begin{cases} \dot{Z}(t) = DZ(t) + \|Q\|_{L^\infty([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)}\,, \\ Z(0) = \|f^{in}\|_{L^\infty(\mathbf{R}^N\times\mathbf{R}^N)}\,. \end{cases}$$

Evidemment

$$Z(t) = \|f^{in}\|_{L^\infty(\mathbf{R}^N\times\mathbf{R}^N)}e^{Dt} + t\|Q\|_{L^\infty([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)}E(Dt)\,, \quad t \geq 0\,,$$

où

$$E(z) = \begin{cases} \dfrac{e^z - 1}{z} & \text{si } z \neq 0\,, \\ 1 & \text{si } z = 0\,. \end{cases}$$

Considérons la fonction

$$g(t,x,v) = Z(t) - f(t,x,v)\,, \quad (t,x,v) \in [0,T] \times \mathbf{R}^N \times \mathbf{R}^N\,.$$

Calculons

$$\begin{aligned} \frac{d}{ds}g(t+s,x+sv,v) &= \dot{Z}(t+s) - \frac{d}{ds}f(t+s,x+sv,v) \\ &= DZ(t+s) + \|Q\|_{L^\infty([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)} - \mathcal{K}f(t+s,x+sv,v) \\ &+ a(t+s,x+sv,v)f(t+s,x+sv,v) - Q(t+s,x+sv,v) \\ &= \mathcal{K}g(t+s,x+sv,v) - a(t+s,x+sv,v)g(t+s,x+sv,v) \\ &+ \|Q\|_{L^\infty([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)} - Q(t+s,x+sv,v) \\ &+ (D - \mathcal{K}1 + a)(t+s,x+sv,v)Z(t+s)\,. \end{aligned}$$

Autrement dit, la fonction g est la solution généralisée du problème de Cauchy

$$\begin{cases} \frac{\partial g}{\partial t} + v \cdot \nabla_x g + ag = \mathcal{K}g + S \,, \\ g\big|_{t=0} = \|f^{in}\|_{L^\infty(\mathbf{R}^N \times \mathbf{R}^N)} - f^{in} \,, \end{cases}$$

où

$$S(t,x,v) = \|Q\|_{L^\infty([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)} - Q(t,x,v) + (D - \mathcal{K}1 + a)(t,x,v)Z(t) \,.$$

Comme $Z \geq 0$ d'après le calcul ci-dessus, on a

$$S \geq 0 \text{ sur } [0,T] \times \mathbf{R}^N \times \mathbf{R}^N \,,$$

tandis que

$$g\big|_{t=0} = \|f^{in}\|_{L^\infty(\mathbf{R}^N \times \mathbf{R}^N)} - f^{in} \geq 0 \text{ sur } \mathbf{R}^N \times \mathbf{R}^N \,.$$

D'après la proposition précédente,

$$g \geq 0 \text{ sur } [0,T] \times \mathbf{R}^N \times \mathbf{R}^N \,,$$

d'où la majoration annoncée puisque

$$E(Dt) \leq e^{Dt} \,, \quad t \geq 0 \,.$$

En effet, pour tout $z \geq 0$, l'on a

$$E(z) = \sum_{n\geq 0} \frac{z^n}{(n+1)!} \leq \sum_{n\geq 0} \frac{z^n}{n!} = e^z \,,$$

ce qui conclut la démonstration. ■

3.2 Le problème aux limites pour l'équation de Boltzmann linéaire

On considèrera tout au long de cette section un ouvert Ω convexe borné de classe C^1 de $\mathbf{R}^N$; on note n_x le vecteur normal unitaire au point $x \in \partial\Omega$ pointant vers l'extérieur de Ω.

On notera également dans tout ce qui suit

$$\begin{aligned} \Gamma_+ &= \{(x,v) \in \partial\Omega \times \mathbf{R}^N \,|\, v \cdot n_x > 0\} \,, \\ \Gamma_0 &= \{(x,v) \in \partial\Omega \times \mathbf{R}^N \,|\, v \cdot n_x = 0\} \,, \\ \Gamma_- &= \{(x,v) \in \partial\Omega \times \mathbf{R}^N \,|\, v \cdot n_x < 0\} \,. \end{aligned}$$

On notera $\tau_{x,v}$ le temps de sortie de Ω dans la direction $-v$ en partant de $x \in \Omega$: c'est-à-dire que

$$\tau_{x,v} = \inf\{t \geq 0 \,|\, x - tv \notin \overline{\Omega}\} \,.$$

Dans le chapitre 2, on avait supposé que toutes les particules étaient transportées avec le même vecteur vitesse v, et c'est pourquoi le temps de sortie ne dépendait que de la variable de position x. Comme on l'a dit dans l'introduction du présent chapitre, l'étude de l'équation de Boltzmann linéaire impose de considérer des populations de particules de vecteurs vitesses quelconques. Le temps de sortie est donc tout naturellement, pour de tels systèmes de particules, une fonction des deux variables x et v, ce qui justifie la notation ci-dessus, différente de celle adoptée dans le chapitre 2.

Tout au long de cette section, on fait l'hypothèse suivante sur les taux d'absorption et de transition :

$$(H) \qquad \begin{cases} 0 \leq a \in C_b(\mathbf{R}_+ \times \overline{\Omega}_x \times \mathbf{R}^N_v)\,, \\ \\ 0 \leq k \in C_b(\mathbf{R}_+ \times \overline{\Omega}_x \times \mathbf{R}^N_v \times \mathbf{R}^N_w)\,, \\ \\ \displaystyle\sup_{(t,x,v)\in \mathbf{R}_+\times\overline{\Omega}\times\mathbf{R}^N} \int_{\mathbf{R}^N} k(t,x,v,w)d\mu(w) < +\infty\,. \end{cases}$$

3.2.1 Existence et unicité pour le problème aux limites

Nous allons commencer par étudier le problème aux limites où on prescrit la valeur au bord de la fonction de distribution des seules particules entrant dans le domaine spatial sur lequel est posée l'équation de Boltzmann linéaire.

Cette condition aux limites n'est pas la plus naturelle dans la plupart des applications — d'ailleurs nous donnerons plus loin d'autres exemples de conditions aux limites que l'on rencontre en pratique dans l'étude de l'équation de Boltzmann linéaire. Toutefois, on s'y ramène presque toujours en pratique. C'est pourquoi nous allons étudier cette condition aux limites en détail dans ce paragraphe.

Théorème 3.2.1 *Soient des données $f^{in} \in C_b(\overline{\Omega}\times\mathbf{R}^N)$, $f_b^- \in C_b([0,T]\times\Gamma_-)$ t.q.*

$$f_b^-(0,y,v) = f^{in}(y,v) \text{ pour tout } (y,v)\in\Gamma_-\,,$$

et $Q \in C_b([0,T]\times\overline{\Omega}\times\mathbf{R}^N)$. Il existe une unique solution généralisée f dans l'espace $C_b([0,T]\times\Omega\times\mathbf{R}^N)$ pour le problème

$$\begin{cases} \frac{\partial f}{\partial t} + v\cdot\nabla_x f + af = \mathcal{K}f + Q\,, \quad t\in]0,T[\,,\ x\in\Omega\,,\ v\in\mathbf{R}^N\,, \\ \\ f\big|_{\Gamma_-} = f_b^-\,, \\ \\ f\big|_{t=0} = f^{in}\,. \end{cases}$$

Voici comment la condition aux limites modifie la première formulation intégrale vérifiée par la solution généralisée du problème de Cauchy.

1ère formulation intégrale :

$$f(t,x,v) = \mathbf{1}_{t\leq\tau_{x,v}} f^{in}(x-tv,v)\exp\left(-\int_0^t a(s,x+(s-t)v,v)ds\right)$$
$$+\mathbf{1}_{t>\tau_{x,v}} f_b^-(t-\tau_{x,v},x_v^*,v)\exp\left(-\int_{t-\tau_{x,v}}^t a(s,x+(s-t)v,v)ds\right)$$
$$+\int_{(t-\tau_{x,v})_+}^t \exp\left(-\int_s^t a(\theta,x+(\theta-t)v,v)d\theta\right)$$
$$\times(\mathcal{K}f+Q)(s,x+(s-t)v,v)ds\,,$$

où on a noté

$$x_v^* = x-\tau_{x,v}v$$

et $z_+ = \max(z,0)$.

La démonstration du théorème ci-dessus suit de très près celle du Théorème 3.1.2 pour le problème de Cauchy. Nous nous contenterons donc de l'esquisser en insistant sur les différences entre les deux arguments.

Démonstration. On cherche une fonction f telle que

$$f = F'[f^{in},f_b^-,Q]+\mathcal{T}'f$$

en posant

$$\mathcal{T}'g(t,x,v) :=$$
$$\int_{(t-\tau_{x,v})_+}^t \mathcal{K}g(s,x+(s-t)v,v)\exp\left(-\int_s^t a(\theta,x+(\theta-t)v,v)d\theta\right)ds\,,$$

et

$$F'[f^{in},f_b^-,Q](t,x,v) :=$$
$$\mathbf{1}_{t\leq\tau_{x,v}} f^{in}(x-tv,v)\exp\left(-\int_0^t a(\theta,x+(\theta-t)v,v)d\theta\right)$$
$$+\mathbf{1}_{t>\tau_{x,v}} f_b^-(t-\tau_{x,v},x_v^*,v)\exp\left(-\int_{t-\tau_{x,v}}^t a(s,x+(s-t)v,v)ds\right)$$
$$+\int_{(t-\tau_{x,v})_+}^t Q(s,x+(s-t)v,v)\exp\left(-\int_s^t a(\theta,x+(\theta-t)v,v)d\theta\right)ds\,.$$

En procédant comme dans la Proposition 2.2.5, comme l'ouvert Ω est convexe à bord de classe C^1, on montre que l'application

$$\Psi : \Gamma_-\times\mathbf{R}_+^* \ni ((x,v),t)\mapsto(x+tv,v)\in\mathbf{R}^N\times(\mathbf{R}^N\setminus\{0\})$$

est de classe C^1 et injective, et que $D\Psi((x_0,v_0),t_0)$ est un isomorphisme de $T_{x_0,v_0}\Gamma_\times\mathbf{R}_+^*$ sur $\mathbf{R}^N\times\mathbf{R}^N$ pour tout $(x_0,v_0)\in\Gamma_-$ et tout $t_0>0$. L'application

Ψ est donc un C^1-difféomorphisme de $\Gamma_- \times \mathbf{R}_+^*$ sur son image $\Psi(\Gamma_- \times \mathbf{R}_+^*)$, qui est un ouvert de $\mathbf{R}^N \times (\mathbf{R}^N \setminus \{0\})$. Or

$$\Psi(\Gamma_- \times \mathbf{R}_+^*) \cap (\Omega \times \mathbf{R}^N) = \{(x,v) \in \Omega \times \mathbf{R}^N \text{ s.t. } \tau_{x,v} < \infty\} =: \mathcal{W}\,.$$

On en déduit que les applications

$$\mathcal{W} \ni (x,v) \mapsto \tau_{x,v} \in \mathbf{R}_+^*$$

et

$$\Omega \times \mathbf{R}^N \ni (x,v) \mapsto \tau_{x,v}|v| \in \mathbf{R}_+$$

sont continues.

De plus, comme les données f^{in} et f_b^- vérifient la relation de compatibilité sur $\{0\} \times \Gamma_-$, la fonction $F'[f^{in}, f_b^-, Q]$ est continue sur $[0,T] \times \Omega \times \mathbf{R}^N$. Et comme $a \geq 0$, on a

$$\begin{aligned}|F'[f^{in}, f_b^-, Q](t,x,v)| &\leq \mathbf{1}_{t \leq \tau_{x,v}} |f^{in}(x - tv, v)| + \mathbf{1}_{t > \tau_{x,v}} |f_b^-(t - \tau_{x,v}, x_v^*, v)| \\ &\qquad + \int_{(t-\tau_{x,v})_+}^t |Q(s, x + (s-t)v, v)| ds \\ &\leq \max\left(\|f^{in}\|_{L^\infty(\Omega \times \mathbf{R}^N)}, \|f_b^-\|_{L^\infty([0,T] \times \Gamma_-)}\right) + \int_0^t \|Q(t,\cdot,\cdot)\|_{L^\infty(\Omega \times \mathbf{R}^N)}\end{aligned}$$

puisque $(t - \tau_{x,v})_+ \geq 0$. Par conséquent

$$\begin{aligned}&\|F'[f^{in}, f_b^-, Q]\|_{L^\infty([0,T] \times \Omega \times \mathbf{R}^N)} \\ &\leq \max\left(\|f^{in}\|_{L^\infty(\Omega \times \mathbf{R}^N)}, \|f_b^-\|_{L^\infty([0,T \times \Gamma_-)}\right) + T\|Q\|_{L^\infty([0,T] \times \Omega \times \mathbf{R}^N)}\,.\end{aligned}$$

Cette fois, on va chercher la solution généralisée sous la forme

$$f := \sum_{n \geq 0} \mathcal{T}'^n F'[f^{in}, f_b^-, Q]\,.$$

Exactement comme dans le cas du problème de Cauchy dans $\mathbf{R}^N \times \mathbf{R}^N$, on montre que la série ci-dessus converge normalement dans l'espace fonctionnel

$$\mathcal{X}_T' = C_b([0,T] \times \Omega \times \mathbf{R}^N)$$

muni de la norme de la convergence uniforme

$$\|\phi\|_{\mathcal{X}_T'} := \sup_{(t,x,v) \in [0,T] \times \Omega \times \mathbf{R}^N} |\phi(t,x,v)|\,,$$

pour laquelle l'espace $\mathcal{X}_T'$ est un espace de Banach (attention, l'exposant $'$ est une notation qui ne désigne pas l'espace dual ici).

En effet, pour tout $g \in \mathcal{X}_T'$, on a

$$\begin{aligned}|\mathcal{T}'g(t,x,v)| &\leq \int_{(t-\tau_{x,v})_+}^t |\mathcal{K}g(s, x + (s-t)v, v)| ds \\ &\leq M \int_{(t-\tau_{x,v})_+}^t \|g(s,\cdot,\cdot)\|_{L^\infty(\Omega \times \mathbf{R}^N)} ds\end{aligned}$$

pour tout $(x,v) \in \Omega \times \mathbf{R}^N$ et tout $t \in [0,T]$. Par conséquent, pour tout $t \in [0,T]$,

$$\begin{aligned}\|\mathcal{T}'^n g(t,\cdot,\cdot)\|_{L^\infty(\Omega\times\mathbf{R}^N)} &\le M\int_0^t \|\mathcal{T}'^{n-1} g(s,\cdot,\cdot)|_{L^\infty(\Omega\times\mathbf{R}^N)} ds \\ &\le \frac{(MT)^n}{n!}\|g\|_{L^\infty([0,T]\times\Omega\times\mathbf{R}^N)}\end{aligned}$$

d'où

$$\|\mathcal{T}'^n\|_{\mathcal{L}(\mathcal{X}'_T)} \le \frac{(MT)^n}{n!}\,.$$

Ainsi

$$\sum_{n\ge 0}\|\mathcal{T}'^n F'[f^{in},f_b^-,Q]\|_{\mathcal{X}'_T} \le \sum_{n\ge 0\|}\frac{(MT)^n}{n!}|F'[f^{in},f_b^-,Q]\|_{\mathcal{X}'_T} < +\infty\,.$$

Autrement dit, cette série converge normalement dans l'espace $\mathcal{X}'_T$. Cette espace étant complet pour la norme $\|\cdot\|_{\mathcal{X}'_T}$, la série ci-dessus converge donc dans $\mathcal{X}'_T$, et on pose

$$f := \sum_{n\ge 0}\mathcal{T}'^n F'[f^{in},f_b^-,Q]\,,$$

ce qui définit $f \in \mathcal{X}'_T$.

Par linéarité de l'opérateur $\mathcal{T}'$, on a

$$\begin{aligned}f &= F'[f^{in},f_b^-,Q] + \sum_{n\ge 1}\mathcal{T}'^n F'[f^{in},f_b^-,Q] \\ &= F'[f^{in},f_b^-,Q] + \mathcal{T}'\left(\sum_{n\ge 0}\mathcal{T}'^n F'[f^{in},f_b^-,Q]\right) \\ &= F'[f^{in},f_b^-,Q] + \mathcal{T}' f\,,\end{aligned}$$

de sorte que la fonction continue f ainsi construite vérifie la 1ère formulation intégrale de l'équation de Boltzmann linéaire.

Posant

$$S = \mathcal{K}f + Q \in C_b([0,T]\times\Omega\times\mathbf{R}^N)\,,$$

on aboutit au fait que f vérifie

$$\begin{aligned}f(t,x,v) &= \mathbf{1}_{t\le\tau_{x,v}} f^{in}(x-tv,v)\exp\left(-\int_0^t a(s,x+(s-t)v,v)ds\right) \\ &+\mathbf{1}_{t>\tau_{x,v}} f_b^-(t-\tau_{x,v},x_v^*,v)\exp\left(-\int_{t-\tau_{x,v}}^t a(s,x+(s-t)v,v)ds\right) \\ &+\int_{(t-\tau_{x,v})_+}^t \exp\left(-\int_s^t a(\theta,x+(\theta-t)v,v)d\theta\right) S(s,x+(s-t)v,v)ds\,,\end{aligned}$$

ce qui, d'après le Théorème 2.3.2, équivaut à dire que, pour tout $v \in \mathbf{R}^N$, la fonction $(t,x) \mapsto f(t,x,v)$ est solution généralisée du problème de Cauchy

$$\begin{cases} \left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f(\cdot,\cdot,v) + a(\cdot,\cdot,v) f(\cdot,\cdot,v) = S(\cdot,\cdot,v)\,, \\ f(\cdot,\cdot,v)\big|_{[0,T]\times\partial\Omega^-} = f_b^-(\cdot,\cdot,v)\,, \\ f(\cdot,\cdot,v)\big|_{t=0} = f^{in}(\cdot,v)\,. \end{cases}$$

Donc la fonction f définie par la série ci-dessus est bien solution généralisée de l'équation de Boltzmann linéaire.

C'est la seule, car s'il en existait deux, disons f_1 et f_2, l'on aurait

$$f_1 - f_2 = \mathcal{T}'(f_1 - f_2) = \ldots = \mathcal{T}'^n(f_1 - f_2)$$

pour tout $n \in \mathbf{N}$, de sorte que

$$\|f_1 - f_2\|_{\mathcal{X}'_T} \leq \|\mathcal{T}'^n\|_{\mathcal{L}(\mathcal{X}'_T)} \|f_1 - f_2\|_{\mathcal{X}'_T} \leq \frac{(MT)^n}{n!} \|f_1 - f_2\|_{\mathcal{X}'_T} \to 0$$

lorsque $n \to +\infty$. On en déduirait donc que $\|f_1 - f_2\|_{\mathcal{X}'_T} = 0$, d'où $f_1 = f_2$. ∎

3.2.2 Estimation L^∞ pour le problème aux limites

Comme nous l'avons fait dans le cas du problème de Cauchy posé dans l'espace euclidien, nous allons maintenant étudier comment contrôler la taille de la solution en fonction de celle de la donnée initiale et de la donnée au bord du domaine spatial.

Le résultat principal dans cette direction est le

Théorème 3.2.2 *Soient des données $f^{in} \in C_b(\overline{\Omega} \times \mathbf{R}^N)$, $f_b^- \in C_b([0,T] \times \Gamma_-)$ t.q. $f_b^-\big|_{t=0} = f^{in}\big|_{\Gamma_-}$ et $Q \in C_b([0,T] \times \overline{\Omega} \times \mathbf{R}^N)$, et soit f l'unique solution généralisée du problème*

$$\begin{cases} \frac{\partial f}{\partial t} + v \cdot \nabla_x f + af = \mathcal{K}f + Q\,, \quad t \in]0,T[\,,\ x \in \Omega\,,\ v \in \mathbf{R}^N\,, \\ f\big|_{\Gamma_-} = f_b^-\,, \\ f\big|_{t=0} = f^{in}\,. \end{cases}$$

(1) Si $f^{in} \geq 0$ sur $\overline{\Omega}$, $f_b^- \geq 0$ sur $[0,T] \times \Gamma_-$, et $Q \geq 0$ sur $[0,T] \times \overline{\Omega} \times \mathbf{R}^N$,

$$\text{alors } f \geq 0 \text{ sur } [0,T] \times \Omega \times \mathbf{R}^N\,;$$

(2) de plus, pour tout $(t,x,v) \in [0,T] \times \overline{\Omega} \times \mathbf{R}^N$,

$$\begin{aligned} f(t,x,v) \leq{}& \max(\|f^{in}\|_{L^\infty(\Omega\times\mathbf{R}^N)}, \|f_b^-\|_{L^\infty([0,T]\times\Gamma_-)}) e^{Dt} \\ &+ T\|Q\|_{L^\infty([0,T]\times\Omega\times\mathbf{R}^N)} e^{Dt}\,, \end{aligned}$$

en notant

$$D = \sup_{(t,x,v)\in[0,T]\times\Omega\times\mathbf{R}^N} \left(\int_{\mathbf{R}^N} k(t,x,v,w)d\mu(w) - a(t,x,v) \right)_+ .$$

Cas particulier : si on a, pour tout $(t,x,v) \in [0,T]\times\overline{\Omega}\times\mathbf{R}^N$,

$$\mathcal{K}1(t,x,v) = \int_{\mathbf{R}^N} k(t,x,v,w)d\mu(w) \leq a(t,x,v)\,,$$

alors $D = 0$ et, pour tout $(t,x,v) \in [0,T]\times\overline{\Omega}\times\mathbf{R}^N$,

$$f(t,x,v) \leq \max(\|f^{in}\|_{L^\infty(\Omega\times\mathbf{R}^N)}, \|f_b^-\|_{L^\infty([0,T]\times\Gamma_-)}) + T\|Q\|_{L^\infty([0,T]\times\Omega\times\mathbf{R}^N)}\,.$$

Démonstration. Comme dans la démonstration de la Proposition 3.1.4, on observe d'une part que, si $f^{in} \geq 0$ sur $\overline{\Omega}$, $f_b^- \geq 0$ sur $[0,T]\times\Gamma_-$, et enfin si le terme source $Q \geq 0$ sur $[0,T]\times\overline{\Omega}\times\mathbf{R}^N$, alors, pour tout $(t,x,v) \in [0,T]\times\overline{\Omega}\times\mathbf{R}^N$

$$\begin{aligned} &F'[f^{in}, f_b^-, Q](t,x,v) \\ &= \mathbf{1}_{t\leq\tau_{x,v}} f^{in}(x-tv,v)\exp\left(-\int_0^t a(\theta, x+(\theta-t)v, v)d\theta\right) \\ &+\mathbf{1}_{t>\tau_{x,v}} f_b^-(t-\tau_{x,v}, x_v^*, v)\exp\left(-\int_{t-\tau_{x,v}}^t a(s, x+(s-t)v, v)ds\right) \\ &+\int_{(t-\tau_{x,v})_+}^t Q(s, x+(s-t)v, v)\exp\left(-\int_s^t a(\theta, x+(\theta-t)v, v)d\theta\right)ds \geq 0\,. \end{aligned}$$

D'autre part, comme $k \equiv k(t,x,v,w) \geq 0$ sur $[0,T]\times\overline{\Omega}\times\mathbf{R}^N\times\mathbf{R}^N$,

$$g \geq 0 \text{ sur } [0,T]\times\overline{\Omega}\times\mathbf{R}^N \Rightarrow \mathcal{K}g \geq 0 \text{ sur } [0,T]\times\overline{\Omega}\times\mathbf{R}^N\,,$$

ce qui entraîne que

$$\mathcal{T}'g \geq 0 \text{ sur } [0,T]\times\overline{\Omega}\times\mathbf{R}^N\,.$$

Par conséquent

$$\mathcal{T}'^n F'[f^{in}, f_b^-, Q] \geq 0 \quad \text{ sur } [0,T]\times\overline{\Omega}\times\mathbf{R}^N\,,$$

de sorte que la solution f du problème aux limites considéré, qui est donnée par la série

$$f = \sum_{n\geq 0} \mathcal{T}'^n F'[f^{in}, f_b^-, Q]\,,$$

vérifie

$$f \geq 0 \quad \text{ sur } [0,T]\times\overline{\Omega}\times\mathbf{R}^N\,,$$

comme somme d'une série à termes positifs. Ceci démontre le point (1).

Pour vérifier le point (2), on procède comme dans la preuve de la Proposition 3.1.5.

Soit $Y \equiv Y(t) \in \mathbf{R}$ la solution de l'équation différentielle ordinaire

$$\begin{cases} \dot{Y}(t) = DY(t) + \|Q\|_{L^\infty([0,T]\times\Omega\times\mathbf{R}^N)}\,, \\ Y(0) = \max\left(\|f^{in}\|_{L^\infty(\Omega\times\mathbf{R}^N)}, \|f_b^-\|_{L^\infty([0,T]\times\Gamma_-)}\right)\,. \end{cases}$$

Evidemment

$$\begin{aligned} Y(t) = \max\left(\|f^{in}\|_{L^\infty(\Omega\times\mathbf{R}^N)}, \|f_b^-\|_{L^\infty([0,T]\times\Gamma_-)}\right) e^{Dt} \\ + t\|Q\|_{L^\infty([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)} E(Dt)\,, \quad t \geq 0\,, \end{aligned}$$

où

$$E(z) = \begin{cases} \dfrac{e^z - 1}{z} & \text{si } z \neq 0\,, \\ 1 & \text{si } z = 0\,. \end{cases}$$

Considérons la fonction

$$h(t,x,v) = Y(t) - f(t,x,v)\,, \quad (t,x,v) \in [0,T]\times\mathbf{R}^N\times\mathbf{R}^N\,.$$

On vérifie comme dans la preuve de la Proposition 3.1.5 que la fonction h est la solution généralisée du problème aux limites

$$\begin{cases} \frac{\partial h}{\partial t} + v\cdot\nabla_x h + ah = \mathcal{K}h + S\,, \\ h\big|_{\Gamma_-} \geq Y(t) - f_b^-(t,\cdot,\cdot)\,, \\ h\big|_{t=0} = Y(0) - f^{in}\,, \end{cases}$$

où

$$S(t,x,v) = \|Q\|_{L^\infty([0,T]\times\mathbf{R}^N\times\mathbf{R}^N)} - Q(t,x,v) + (D - K1 + a)(t,x,v)Y(t)\,.$$

Or

$$Y(0) - f^{in} = \max\left(\|f^{in}\|_{L^\infty(\Omega\times\mathbf{R}^N)}, \|f_b^-\|_{L^\infty([0,T]\times\Gamma_-)}\right) - f^{in} \geq 0$$

sur $\Omega\times\mathbf{R}^N$, tandis que

$$\begin{aligned} Y(t) - f_b^-(t,y,v) \\ \geq \max\left(\|f^{in}\|_{L^\infty(\Omega\times\mathbf{R}^N)}, \|f_b^-\|_{L^\infty([0,T]\times\Gamma_-)}\right) - f_b^-(t,y,v) \geq 0 \end{aligned}$$

sur $[0,T]\times\Gamma_-$. D'autre part, comme $Y \geq 0$ d'après la formule explicite ci-dessus, et que $D - \mathcal{K}1 + a \geq 0$ par définition de D, on a

$$S \geq 0 \quad \text{ sur } [0,T]\times\Omega\times\mathbf{R}^N\,.$$

D'après l'énoncé (1),

$$h \geq 0 \quad \text{ sur } [0,T] \times \Omega \times \mathbf{R}^N \,,$$

d'où la majoration annoncée puisque

$$E(Dt) \leq e^{Dt} \,, \quad t \geq 0 \,.$$

■

3.2.3 Autres conditions aux limites

Jusqu'ici nous avons considéré uniquement des conditions aux limites consistant à prescrire la valeur au bord de la densité de particules entrant à tout instant dans le domaine spatial. Comme on l'a vu au chapitre précédent, cette condition est tout à fait naturelle dans le cas de l'équation de transport avec une vitesse unique donnée.

La situation est différente pour le cas de l'équation de Boltzmann, où les mécanismes d'absorption et de création (par exemple par scattering) mettent en jeu des phénomènes d'échange de vitesses.

C'est pourquoi il existe en théorie cinétique de nombreux autres exemples de conditions aux limites. Nous en présenterons quatre dans cette section.

Dans tous les cas, on considèrera l'équation de Boltzmann linéaire sous la forme

$$\left(\frac{\partial}{\partial t} + v \cdot \nabla_x \right) f(t,x,v) = \mathcal{K}f(t,x,v) - a(t,x,v) f(t,x,v) \,,$$

où $t \in [0,T]$, $x \in \Omega$ et $v \in \mathbf{R}^N$, avec Ω ouvert convexe borné à bord de classe C^1 de $\mathbf{R}^N$, en notant n_x le vecteur normal unitaire au point $x \in \partial\Omega$ pointant vers l'extérieur de Ω. Le terme de scattering $\mathcal{K}f$ est donné par un opérateur intégral de la forme

$$\mathcal{K}f(t,x,v) = \int_{\mathbf{R}^N} k(t,x,v,w) f(t,x,w) d\mu(w) \,.$$

Conditions aux limites périodiques :

L'équation de Boltzmann linéaire ci-dessus est posée sur le cube $[0,L]^N$, avec les mêmes conditions aux limites périodiques que celles de la Remarque 2.2.9 :

$$f(t,\hat{x}_j,v) = f(t,\hat{x}_j + Le_j,v) \,, \qquad\qquad x \in \partial(]0,L[^N),\ t>0,\ 1 \leq j \leq N \,,$$

où on rappelle que

$$\hat{x}_j = (x_1,\dots,x_{j-1},0,x_{j+1},\dots,x_N) \text{ et } e_j = (\underbrace{0,\dots,0}_{j-1},1,\underbrace{0,\dots,0}_{N-j}) \,.$$

En suivant le même raisonnement que dans la Remarque 2.2.9, on considère le problème de Cauchy posé dans l'espace entier

$$\begin{cases} \left(\dfrac{\partial}{\partial t} + v \cdot \nabla_x \right) F = \mathcal{K}F - aF \,, \quad (x,v) \in \mathbf{R}^N \times \mathbf{R}^N \,,\ t>0 \,, \\ \\ F\big|_{t=0} = F^{in} \,. \end{cases}$$

Dans cette formulation, a et k sont des fonctions périodiques de période L en chacune des variables spatiales $x_1, \dots, x_N$. De même, la donnée initiale F^{in} est la fonction périodique de période L en chacune des variables spatiales $x_1, \dots, x_N$ dont la restriction pour $x \in [0, L]^N$ est f^{in}.

Comme les fonctions a, k et F^{in} sont périodiques de période L en chacune des variables spatiales $x_1, \dots, x_N$, on voit que, pour tout $k \in \mathbf{Z}^N$, la fonction F_k définie par $F_k(t, x, v) = F(t, x + lk, v)$ est solution du même problème de Cauchy que F. Par unicité de la solution du problème de Cauchy (Théorème 3.1.2), on conclut que $F_k = F$ pour tout $k \in \mathbf{Z}^N$, c'est-à-dire que la solution F est périodique de période L dans chacune des variables spatiales $x_1, \dots, x_N$.

Alors la solution du problème aux limites pour l'équation de Boltzmann linéaire ci-dessus posé dans le cube $[0, L]^N$ est la restriction pour $x \in [0, L]^N$ de la solution du problème de Cauchy, c'est-à-dire que

$$f(t, x, v) = F(t, x, v) \text{ pour tout } (t, x, v) \in \mathbf{R}_+ \times [0, L]^N \times \mathbf{R}^N .$$

Réflexion spéculaire :

On suppose que chaque particule arrivant sur le bord du domaine Ω en se dirigeant vers l'extérieur de Ω y est réfléchie vers l'intérieur de Ω suivant la loi de Descartes. C'est à dire que le vecteur vitesse v' de la particule après collision au point $x \in \partial\Omega$ est le symétrique orthogonal du vecteur vitesse v de cette même particule avant collision par rapport à l'espace tangent à $\partial\Omega$ au point x. Autrement dit

$$v' = v - 2(v \cdot n_x) n_x .$$

Dans toute cette section, n_x désigne le vecteur unitaire normal à $\partial\Omega$ au point x, dirigé vers l'extérieur de Ω.

Au niveau de la fonction de distribution, cette condition de réflexion s'écrit

$$f(t, x, v) = f(t, x, v - 2(v \cdot n_x) n_x) , \quad v \in \mathbf{R}^N , \quad x \in \partial\Omega ,$$

pour tout $t \geq 0$.

Par exemple, cette loi de réflexion est vérifiée par l'intensité lumineuse éclairant une surface extrêmement lisse et parfaitement réfléchissante, typiquement un miroir (d'où l'adjectif "spéculaire").

Réflexion diffuse :

Cette loi de réflexion est adaptée par exemple au cas de rayons lumineux éclairant une surface mate — par exemple un écran de cinéma. Dans ce cas, l'intensité réfléchie est la même dans toutes les directions. Traduisons cela au moyen de la fonction de distribution :

$$f(t, x, v) = F(t, x) , \quad v \cdot n_x < 0 , \quad x \in \partial\Omega .$$

Il reste à évaluer F. Faisons l'hypothèse que le flux de particules réfléchies par le bord $\partial\Omega$ du domaine est égal, en tout point $x \in \partial\Omega$ et pour tout $t \geq 0$, au flux de particules incidentes :

$$\int_{v \cdot n_x < 0} F(t, x) |v \cdot n_x| d\mu(v) = \int_{w \cdot n_x > 0} f(t, x, w) w \cdot n_x d\mu(w) , \quad x \in \partial\Omega , \ t \geq 0 .$$

Cette identité permet d'évaluer $F(t,x)$, et d'en déduire la formulation suivante de la condition de réflexion diffuse :

$$f(t,x,v) = \frac{\displaystyle\int_{w\cdot n_x>0} f(t,x,w) w\cdot n_x d\mu(w)}{\displaystyle\int_{w\cdot n_x<0} |w\cdot n_x| d\mu(w)}\,, \qquad v\cdot n_x<0\,, \quad x\in\partial\Omega\,.$$

Dans le contexte de la photonique, cette loi de réflexion porte parfois le nom de "loi de Lambert".

On peut également panacher ces deux lois de réflexion, pour aboutir à la

Condition d'accomodation :

En tout point $x\in\partial\Omega$, et à tout instant t, la fonction de distribution des particules réémises vers l'intérieur de Ω est une pondération convexe entre la fonction de distribution obtenue par réflexion spéculaire et celle obtenue par réflexion diffuse — la pondération pouvant dépendre du point x du bord et de l'instant t. Autrement dit

$$\begin{aligned} f(t,x,v) &= (1-\alpha(t,x)) f(t,x,v-2(v\cdot n_x)n_x) \\ &\quad + \alpha(t,x)\frac{\displaystyle\int_{w\cdot n_x>0} f(t,x,w) w\cdot n_x d\mu(w)}{\displaystyle\int_{w\cdot n_x<0} |w\cdot n_x| d\mu(w)}\,, \qquad v\cdot n_x<0\,, \quad x\in\partial\Omega\,. \end{aligned}$$

Dans tous les cas, la condition aux limites se met sous la forme

$$f(t,x,v)|v\cdot n_x| = \int_{w\cdot n_x>0} r(t,x,v,w) f(t,x,w) w\cdot n_x d\mu(w)\,, \qquad v\cdot n_x<0\,, \quad x\in\partial\Omega\,,$$

où

$$r(t,x,v,w) w\cdot n_x d\mu(w) = |v\cdot n_x| \delta_0(w-v+2(v\cdot n_x)n_x)\,, \qquad v\cdot n_x<0\,, \quad x\in\partial\Omega\,,$$

dans le cas de la réflexion spéculaire, tandis que

$$r(t,x,v,w) w\cdot n_x d\mu(w) = \frac{|v\cdot n_x| w\cdot n_x d\mu(w)}{\displaystyle\int_{w\cdot n_x<0} |w\cdot n_x| d\mu(w)}\,, \quad v\cdot n_x<0\,, \quad x\in\partial\Omega\,,$$

dans le cas de la réflexion diffuse, le cas de la condition d'accomodation correspondant à

$$\begin{aligned} r(t,x,v,w) w\cdot n_x d\mu(w) &= (1-\alpha(t,x)) |v\cdot n_x| \delta_0(w-v+2(v\cdot n_x)n_x) \\ &\quad + \alpha(t,x)\frac{|v\cdot n_x| w\cdot n_x d\mu(w)}{\displaystyle\int_{w\cdot n_x<0} |w\cdot n_x| d\mu(w)}\,, \quad v\cdot n_x<0\,,\ x\in\partial\Omega\,. \end{aligned}$$

La notation δ_0 désigne la masse de Dirac au point 0 dans $\mathbf{R}^N$.

Plus généralement, on s'intéressera à des conditions de réflexion de la forme

$$f(t,x,v)|v\cdot n_x| = \int_{w\cdot n_x>0} r(t,x,v,w) f(t,x,w) w\cdot n_x d\mu(w)\,,$$
$$v\cdot n_x<0\,,\quad x\in\partial\Omega\,,$$

où

$$r(t,x,v,w)\geq 0\,,\quad (t,x,v,w)\in \mathbf{R}_+\times\partial\Omega\times\mathbf{R}^N\times\mathbf{R}^N$$

vérifie, pour tout $(t,x,w)\in\mathbf{R}_+\times\partial\Omega\times\mathbf{R}^N$,

$$\int_{v\cdot n_x<0} r(t,x,v,w)d\mu(v)=1\,,$$

et

$$\int_{w\cdot n_x>0} r(t,x,v,w) w\cdot n_x d\mu(w) = |v\cdot n_x|\,,$$

pour tout $(t,x,v)\in\mathbf{R}_+\times\partial\Omega\times\mathbf{R}^N$ tel que $v\cdot n_x<0$.

Toutes les conditions de réflexion de ce type satisfont à une inégalité élémentaire mais très importante dans la pratique, due à Darrozes et Guiraud. Nous la donnons ci-dessous dans le seul cas d'une non linéarité quadratique. (Dans le contexte de la théorie cinétique des gaz, cette inégalité porte sur le flux d'entropie au bord, et fait donc intervenir une non linéarité de la forme $z\mapsto z\ln z$.)

Lemme 3.2.3 (Darrozes-Guiraud) *Soit $f\in L^\infty([0,T]\times\partial\Omega\times\mathbf{R}^N)$ vérifiant la condition de réflexion*

$$f(t,x,v)|v\cdot n_x| = \int_{w\cdot n_x>0} r(t,x,v,w) f(t,x,w) w\cdot n_x d\mu(w)\,,$$
$$v\cdot n_x<0\,,\quad x\in\partial\Omega\,,$$

où

$$r(t,x,v,w)>0\,,\qquad \textit{p.p. en } (t,x,v,w)\in\mathbf{R}_+\times\partial\Omega\times\mathbf{R}^N\times\mathbf{R}^N$$

satisfait aux conditions

$$\int_{v\cdot n_x<0} r(t,x,v,w)d\mu(v)=1 \quad \textit{et} \int_{v\cdot n_x>0} r(t,x,v,w) w\cdot n_x d\mu(v) = |v\cdot n_x|$$

pour tout $(t,x,v)\in\mathbf{R}_+\times\partial\Omega\times\mathbf{R}^N$ tel que $v\cdot n_x<0$.

Alors

$$\int_{\mathbf{R}^N} f(t,x,v)^2 v\cdot n_x dv\geq 0 \textit{ p.p. en } (t,x)\in[0,T]\times\partial\Omega\,,$$

avec égalité si et seulement si $f(t,x,v) = F(t,x)$ μ-p.p., c'est-à-dire si la fonction $v \mapsto f(t,x,v)$ est μ-p.p. constante[2] *en v, p.p. en $(t,x) \in [0,T] \times \partial\Omega$.*

Démonstration. On a donc

$$\int_{v\cdot n_x<0} f(t,x,v)^2 |v\cdot n_x| d\mu(v)$$
$$= \int_{v\cdot n_x<0} \left(\int_{w\cdot n_x>0} r(t,x,v,w) f(t,x,w) w\cdot n_x d\mu(w) \right)^2 \frac{d\mu(v)}{|v\cdot n_x|}$$

p.p. en $(t,x) \in [0,T] \times \partial\Omega$. D'après l'inégalité de Cauchy-Schwarz

$$\left(\int_{w\cdot n_x>0} r(t,x,v,w) f(t,x,w) w\cdot n_x d\mu(w) \right)^2$$
$$\leq \left(\int_{w\cdot n_x>0} r(t,x,v,w) w\cdot n_x d\mu(w) \right)$$
$$\times \left(\int_{w\cdot n_x>0} r(t,x,v,w) f(t,x,w)^2 w\cdot n_x d\mu(w) \right)$$
$$= |v\cdot n_x| \int_{w\cdot n_x>0} r(t,x,v,w) f(t,x,w)^2 w\cdot n_x d\mu(w)\,.$$

Donc

$$\int_{v\cdot n_x<0} f(t,x,v)^2 |v\cdot n_x| d\mu(v)$$
$$\leq \int_{v\cdot n_x<0} \left(\int_{w\cdot n_x>0} r(t,x,v,w) f(t,x,w)^2 w\cdot n_x d\mu(w) \right) d\mu(v)$$
$$= \int_{w\cdot n_x>0} \left(\int_{v\cdot n_x<0} r(t,x,v,w) d\mu(v) \right) f(t,x,w)^2 w\cdot n_x d\mu(w)$$
$$= \int_{w\cdot n_x>0} f(t,x,w)^2 w\cdot n_x d\mu(w)\,,$$

d'où l'inégalité annoncée.

2. Dans les exemples (b)-(c) de l'introduction, μ est la mesure de Lebesgue restreinte à une partie de $\mathbf{R}^N$, et la notation μ-p.p. signifie "presque partout" au sens habituel. Dans l'exemple e) correspondant à une variable v appartenant à une partie finie ou dénombrable $\mathcal{V}$ de $\mathbf{R}^N$, la notation μ-p.p. en v signifie "pour tout $v \in \mathcal{V}$". Enfin, dans l'exemple d) où $d\mu(v)$ désigne l'élément de surface sur la sphère unité $\mathbf{S}^{N-1}$ de $\mathbf{R}^N$, la notation μ-p.p. signifie que la fonction $f(t,x,\cdot)$ est constante sur le complémentaire d'une partie $\mathcal{N}$ de $\mathbf{S}^{N-1}$ telle que

$$\int_{\mathcal{N}} d\mu(v) = 0\,.$$

Par exemple, lorsque $N = 3$, l'élément de surface sur la sphère unité de $\mathbf{R}^3$ s'écrit, en coordonnées sphériques $(\theta,\phi) \in]0,\pi[\times]0,2\pi[$ — comme dans la figure 1.1 — $d\mu(\theta,\phi) = \sin\theta d\phi d\theta$. Dans ce cas, la notation μ-p.p. équivaut à p.p. en (θ,ϕ).

En cas d'égalité, il y a alors égalité dans l'inégalité de Cauchy-Schwarz μ-p.p. en v, c'est-à-dire que

$$\left(\int_{w\cdot n_x>0} r(t,x,v,w)f(t,x,w)w\cdot n_x d\mu(w)\right)^2$$
$$= \left(\int_{w\cdot n_x>0} r(t,x,v,w)w\cdot n_x d\mu(w)\right)$$
$$\times \left(\int_{w\cdot n_x>0} r(t,x,v,w)f(t,x,w)^2 w\cdot n_x d\mu(w)\right).$$

Or ceci entraîne que les fonctions

$$w\mapsto \sqrt{r(t,x,v,w)w\cdot n_x} \text{ et } w\mapsto \sqrt{r(t,x,v,w)w\cdot n_x} f(t,x,w)$$

sont μ-p.p. colinéaires, ce qui équivaut à dire que la fonction $w\mapsto f(t,x,w)$ est μ-p.p. constante en la variable $w\in\mathcal{V}$, p.p. en $(t,x)\in\mathbf{R}_+\times\partial\Omega$. ■

Nous ne ferons pas en détail la théorie du problème aux limites avec ce type de condition de réflexion. Voici toutefois deux remarques importantes dans cette direction.

Supposons que le taux d'amortissement a et le noyau de transition k vérifient les hypothèses générales (H) ainsi que

$$\sup_{(t,x,w)\in\mathbf{R}_+\times\overline{\Omega}\times\mathbf{R}^N}\int_{\mathbf{R}^N} k(t,x,v,w)d\mu(v) < +\infty\,.$$

Supposons également pour simplifier que le noyau r définissant la condition de réflexion vérifie les hypothèses suivantes :

$$\begin{cases} 0\le r(t,x,v,w)\in C_b(\mathbf{R}_+\times\partial\Omega\times\mathbf{R}^N\times\mathbf{R}^N)\,,\\[2ex] \displaystyle\int_{v\cdot n_x<0} r(t,x,v,w)d\mu(v)=1\,,\quad (t,x,w)\in\mathbf{R}_+\times\partial\Omega\times\mathbf{R}^N\,,\\[2ex] \displaystyle\int_{w\cdot n_x>0} r(t,x,v,w)w\cdot n_x d\mu(w)=|v\cdot n_x|\,,\quad (t,x,v)\in\mathbf{R}_+\times\partial\Omega\times\mathbf{R}^N\,. \end{cases}$$

Evidemment, le cas de la réflexion spéculaire ou d'une condition d'accomodation ne vérifient pas la première hypothèse, car le noyau de réflexion contient alors une mesure de Dirac. Malgré tout, les raisonnements ci-dessous s'adaptent sans difficulté à ces deux cas particuliers importants.

Une première observation importante est que l'inégalité de Darrozes-Guiraud entraîne un résultat d'unicité.

Proposition 3.2.4 *Soit $f^{in}\in C^1(\overline{\Omega}\times\mathbf{R}^N)$. Sous les hypothèses ci-dessus por-*

tant sur le noyau r, le problème aux limites

$$\begin{cases} \left(\frac{\partial}{\partial t} + v\cdot\nabla_x\right) f(t,x,v) = (\mathcal{K}f - af)(t,x,v)\,, \quad (t,x,v)\in]0,T[\times\Omega\times\mathbf{R}^N\,, \\ \\ f(t,x,v)|v\cdot n_x| = \displaystyle\int_{w\cdot n_x>0} r(t,x,v,w)f(t,x,w)w\cdot n_x dw\,, \quad (x,v)\in\Gamma_-\,, \\ \\ f\big|_{t=0} = f^{in} \end{cases}$$

admet au plus une solution classique $f\in C^1(\mathbf{R}_+\times\overline{\Omega}\times\mathbf{R}^N)$.

S'il en existait deux, disons f_1 et f_2, la différence $g=f_1-f_2$ appartiendrait à $C^1(\mathbf{R}_+\times\overline{\Omega}\times\mathbf{R}^N)$ et vérifierait

$$\begin{cases} \left(\frac{\partial}{\partial t} + v\cdot\nabla_x\right) g(t,x,v) = (\mathcal{K}g - ag)(t,x,v)\,, \quad (t,x,v)\in]0,T[\times\Omega\times\mathbf{R}^N\,, \\ \\ g(t,x,v)|v\cdot n_x| = \displaystyle\int_{w\cdot n_x>0} r(t,x,v,w)g(t,x,w)w\cdot n_x dw\,, \quad (x,v)\in\Gamma_-\,, \\ \\ g\big|_{t=0} = 0\,. \end{cases}$$

Le cœur de la démonstration de la Proposition 3.2.4 repose sur l'estimation a priori suivante, obtenue par une méthode tout à fait analogue à celle décrite dans la section 2.5.

Notation : pour tout $X\subset\mathbf{R}^n$, on note $C^k_b(X)$ l'ensemble des fonctions de classe C^k sur un voisinage ouvert de X dont toutes les dérivées partielles d'ordre inférieur ou égal à k sont bornées sur X.

Lemme 3.2.5 (Estimation L^2) *Posons*

$$D = \|\tfrac{1}{2}(\mathcal{K}1 + \mathcal{K}^*1 - 2a)^+\|_{L^\infty(\mathbf{R}_+\times\Omega\times\mathbf{R}^N)}\,,$$

en notant $\mathcal{K}^*$ l'adjoint (formel) de l'opérateur $\mathcal{K}$, c'est-à-dire que

$$\mathcal{K}^*f(t,x,w) = \int_{\mathbf{R}^N} k(t,x,v,w)f(t,x,v)d\mu(v)\,.$$

Toute solution $g\in C^1_b(\mathbf{R}_+\times\overline{\Omega}\times\mathbf{R}^N)$ du problème aux limites ci-dessus vérifie l'inégalité différentielle

$$\frac{d}{dt}\iint_{\Omega\times\mathbf{R}^N} g(t,x,v)^2 dx d\mu(v) \le 2D\iint_{\Omega\times\mathbf{R}^N} g(t,x,v)^2 dx d\mu(v)\,.$$

Remarque 3.2.6 *Par rapport à la définition (3.1) de l'opérateur $\mathcal{K}$, son adjoint $\mathcal{K}^*$ est défini en inversant les rôles des vitesses v et w dans le noyau $k(t,x,v,w)$. Autrement dit, on intègre par rapport à v dans la définition de $\mathcal{K}^*$ alors qu'on intégrait par rapport à w dans celle de $\mathcal{K}$.*

Démonstration du lemme. Multiplions par g chaque membre de l'équation ci-dessus : on trouve

$$\tfrac{1}{2}\left(\frac{\partial}{\partial t}+v\cdot\nabla_x\right)g(t,x,v)^2=g(t,x,v)\mathcal{K}g(t,x,v)-a(t,x,v)g(t,x,v)^2\,.$$

De plus, en appliquant le théorème de Fubini

$$\begin{aligned}\int_{\mathbf{R}^N} g(t,x,v)\mathcal{K}g(t,x,v)d\mu(v)&\\ &=\int_{\mathbf{R}^N}\int_{\mathbf{R}^N} k(t,x,v,w)g(t,x,v)g(t,x,w)d\mu(v)d\mu(w)\\ &\le\int_{\mathbf{R}^N}\int_{\mathbf{R}^N} k(t,x,v,w)\tfrac{1}{2}(g(t,x,v)^2+g(t,x,w)^2)d\mu(v)d\mu(w)\\ &=\int_{\mathbf{R}^N}\tfrac{1}{2}(\mathcal{K}1+\mathcal{K}^*1)(t,x,u)g(t,x,u)^2d\mu(u)\,.\end{aligned}$$

Ainsi

$$\begin{aligned}\tfrac{1}{2}\int_{\mathbf{R}^N}\left(\frac{\partial}{\partial t}+v\cdot\nabla_x\right)g(t,x,v)^2d\mu(v)&\\ &\le\int_{\mathbf{R}^N}\tfrac{1}{2}(\mathcal{K}1+\mathcal{K}^*1-2a)(t,x,u)g(t,x,u)^2d\mu(u)\,.\end{aligned}$$

Intégrons ensuite chaque membre de cette inégalité par rapport à x : par dérivation sous le signe somme en la variable t, et en appliquant la formule de Green, on trouve que

$$\begin{aligned}\int_\Omega\int_{\mathbf{R}^N}\left(\frac{\partial}{\partial t}+v\cdot\nabla_x\right)g(t,x,v)^2d\mu(v)dx=\frac{d}{dt}\iint_{\Omega\times\mathbf{R}^N}\tfrac{1}{2}g(t,x,v)^2dxd\mu(v)&\\ +\iint_{\partial\Omega\times\mathbf{R}^N}\tfrac{1}{2}g(t,x,v)^2v\cdot n_x d\sigma(x)d\mu(v)\,,&\end{aligned}$$

en notant $d\sigma(x)$ l'élément de surface sur $\partial\Omega$ orientée par le champ normal extérieur n_x.

Or d'après l'inégalité de Darrozes-Guiraud, l'intégrale ci-dessus sur le bord de Ω est positive ou nulle, de sorte que

$$\begin{aligned}\frac{d}{dt}\iint_{\Omega\times\mathbf{R}^N}\tfrac{1}{2}g(t,x,v)^2dxd\mu(v)&\\ &\le\iint_{\Omega\times\mathbf{R}^N}\tfrac{1}{2}(\mathcal{K}1+\mathcal{K}^*1-2a)(t,x,u)g(t,x,u)^2d\mu(u)dx\,.\end{aligned}$$

Par hypothèse, $a\ge 0$, tandis que les fonctions positives ou nulles $\mathcal{K}1$ et $\mathcal{K}^*1$ sont bornées sur $\mathbf{R}_+\times\Omega\times\mathbf{R}^N$. En notant

$$D=\|\tfrac{1}{2}(\mathcal{K}1+\mathcal{K}^*1-2a)^+\|_{L^\infty(\mathbf{R}_+\times\Omega\times\mathbf{R}^N)}\,,$$

on trouve que l'inégalité ci-dessus implique évidemment

$$\frac{d}{dt}\iint_{\Omega\times\mathbf{R}^N} g(t,x,v)^2 dxd\mu(v) \le 2D\iint_{\Omega\times\mathbf{R}^N} g(t,x,v)^2 dxd\mu(v)\,.$$

■

Remarque 3.2.7 *On remarquera que, dans la démonstration de cette inégalité différentielle comme dans celle de la section 2.5, le point crucial est l'inégalité*

$$\iint_{\partial\Omega\times\mathbf{R}^N} g(t,x,v)^2 v\cdot n_x d\sigma(x)d\mu(v) \ge 0\,,$$

qui découle ici de l'inégalité de Darrozes-Guiraud.

Dans la section 2.5, on utilisait plutôt l'inégalité

$$\int_{\partial\Omega} \tfrac{1}{2} f(t,x)^2 v\cdot n_x d\sigma(x) \ge -\int_{\partial\Omega^-} \tfrac{1}{2} f_b^-(t,x)^2 |v\cdot n_x| d\sigma(x)$$

qui découlait, elle, de la seule positivité de $f(t,x,v)^2$.

Démonstration de la Proposition 3.2.4. D'après le lemme ci-dessus, s'il existe deux solutions classiques f_1 et f_2 du problème aux limites, leur différence $g = f_1 - f_2$ vérifie l'inégalité différentielle

$$\frac{d}{dt}\iint_{\Omega\times\mathbf{R}^N} g(t,x,v)^2 dxd\mu(v) \le 2D\iint_{\Omega\times\mathbf{R}^N} g(t,x,v)^2 dxd\mu(v)\,,$$

que l'on peut encore mettre sous la forme

$$\frac{d}{dt}\left(e^{-2Dt}\iint_{\Omega\times\mathbf{R}^N} g(t,x,v)^2 dxd\mu(v)\right) \le 0\,.$$

Intégrant chaque membre de cette inégalité sur $[0,t]$, on trouve que

$$e^{-2Dt}\iint_{\Omega\times\mathbf{R}^N} g(t,x,v)^2 dxd\mu(v) \le \iint_{\Omega\times\mathbf{R}^N} g(0,x,v)^2 dxd\mu(v) = 0\,,$$

pour tout $t \ge 0$, d'où l'on tire que $f_1 - f_2 = g = 0$. ■

La deuxième observation que l'on peut faire concerne l'existence d'une solution du problème aux limites

$$\begin{cases} \left(\frac{\partial}{\partial t} + v\cdot\nabla_x\right) f(t,x,v) = (\mathcal{K}f - af)(t,x,v)\,, \quad (t,x,v)\in]0,T[\times\Omega\times\mathbf{R}^N\,, \\ \\ f(t,x,v)|v\cdot n_x| = \displaystyle\int_{w\cdot n_x>0} r(t,x,v,w) f(t,x,w) w\cdot n_x d\mu(w)\,, \quad (x,v)\in\Gamma_-\,, \\ \\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

en supposant que $f^{in} \in C_b(\overline{\Omega}\times\mathbf{R}^N)$.

Construisons par récurrence une suite $(f_n)_{n\geq 0}$ de fonctions appartenant à $C_b([0,T]\times\Omega\times\mathbf{R}^N)$ obtenues comme suit :

$$f_0 = 0$$

et, pour tout $n \geq 1$, la fonction f_n est la solution du problème aux limites

$$\begin{cases} \left(\frac{\partial}{\partial t} + v\cdot\nabla_x\right) f_n(t,x,v) = (\mathcal{K}f_n - af_n)(t,x,v)\,, \quad (t,x,v)\in]0,T[\times\Omega\times\mathbf{R}^N\,, \\ \\ f_n(t,x,v)|v\cdot n_x| = \displaystyle\int r(t,x,v,w) f_{n-1}(t,x,w)(w\cdot n_x)_+ d\mu(w)\,, \quad (x,v)\in\Gamma_-\,, \\ \\ f_n\big|_{t=0} = f^{in}\,, \end{cases}$$

écrit sous forme intégrale, c'est-à-dire que

$$\begin{aligned} f_n(t,x,v) &= \mathbf{1}_{t\leq\tau_{x,v}} f^{in}(x-tv,v)\exp\left(-\int_0^t a(s,x+(s-t)v,v)ds\right) \\ &\quad + \mathbf{1}_{t>\tau_{x,v}}\exp\left(-\int_{t-\tau_{x,v}}^t a(s,x+(s-t)v,v)ds\right)\frac{1}{|v\cdot n_{x^-}|} \\ &\quad \times\int_{w\cdot n_{x_v^*}>0} r(t-\tau_{x,v},x_v^*,v,w) f_{n-1}(t-\tau_{x,v},x_v^*,w) w\cdot n_{x_v^*} d\mu(w) \\ &\quad + \int_{(t-\tau_{x,v})_+}^t \exp\left(-\int_s^t a(\theta,x+(\theta-t)v,v)d\theta\right)\mathcal{K}f_n(s,x+(s-t)v,v)ds\,, \end{aligned}$$

où on a noté comme d'habitude $x_v^* = x-\tau_{x,v}v$.

On vérifie par une récurrence immédiate que pour tous $t\in[0,T]$, $x\in\Omega$ et $v\in\mathbf{R}^N$, l'on a

$$\begin{aligned} 0 = f_0(t,x,v) \leq f_1(t,x,v) \leq \ldots \leq f_n(t,x,v) \\ \leq f_{n+1}(t,x,v) \leq \ldots \leq \|f^{in}\|_{L^\infty(\Omega\times\mathbf{R}^N)} e^{Dt}\,, \end{aligned}$$

où

$$D = \|(\mathcal{K}1-a)^+\|_{L^\infty([0,T]\times\Omega\times\mathbf{R}^N)}\,.$$

On en déduit par convergence monotone que f_n converge ponctuellement vers une solution de l'équation intégrale

$$\begin{aligned} f(t,x,v) &= \mathbf{1}_{t\leq\tau_{x,v}} f^{in}(x-tv,v)\exp\left(-\int_0^t a(s,x+(s-t)v,v)ds\right) \\ &\quad + \mathbf{1}_{t>\tau_{x,v}}\exp\left(-\int_{t-\tau_{x,v}}^t a(s,x+(s-t)v,v)ds\right)\frac{1}{|v\cdot n_{x_v^*}|} \\ &\quad \times\int_{w\cdot n_{x_v^*}>0} r(t-\tau_{x,v},x_v^*,v,w) f(t-\tau_{x,v},x_v^*,w) w\cdot n_{x_v^*} d\mu(w) \\ &\quad + \int_{(t-\tau_{x,v})_+}^t \exp\left(-\int_s^t a(\theta,x+(\theta-t)v,v)d\theta\right)\mathcal{K}f(s,x+(s-t)v,v)ds\,, \end{aligned}$$

qui est la formulation intégrale du problème aux limites considéré.

Mais a priori, la solution ainsi obtenue n'est pas toujours une solution classique, de sorte qu'on ne peut lui appliquer l'argument d'unicité basé sur l'inégalité de Darrozes-Guiraud tel que nous l'avons présenté ci-dessus. Par exemple, si

$$f^{in}(x,v)|v\cdot n_x| \neq \int_{\mathbf{R}^N} r(x,v,w)f^{in}(x,w)(w\cdot n_x)_+ d\mu(w)$$

pour certains $(x,v)\in\Gamma_-$, la solution f construite par le procédé ci-dessus ne sera pas continue sur $\mathbf{R}_+\times\overline{\Omega}\times\mathbf{R}^N$. L'unicité de la solution généralisée, bien que s'appuyant sur l'inégalité de Darrozes-Guiraud, nécessite une démonstration un peu plus technique, que nous ne donnerons pas ici.

3.3 Interprétation probabiliste de l'équation de Boltzmann linéaire

Contrairement au cas de l'équation de transport libre étudiée au chapitre précédent, la méthode des caractéristiques ne donne pas de formule explicite permettant de calculer la solution f de l'équation de Boltzmann linéaire en fonction de la donnée initiale et de la donnée au bord, et éventuellement du terme source. (Par "formule explicite", nous entendons une combinaison convexe des données évaluées sur les caractéristiques de l'opérateur de transport libre $\frac{\partial}{\partial t}+v\cdot\nabla_x$.) La formule obtenue à partir de la méthode des caractéristiques ne donne f que de manière implicite, puisque le terme de scattering $\mathcal{K}f$ entre dans le terme source. Autrement dit, la méthode des caractéristiques permet de transformer l'équation intégro-différentielle de Boltzmann en une équation intégrale d'inconnue f.

Evidemment, la série de Duhamel fournit une formule explicite pour la solution de l'équation de Boltzmann linéaire. A première vue, cette formule diffère notablement de la formule des caractéristiques pour l'équation de transport.

Pour aller plus loin, et avoir une formule explicite de représentation de la solution de l'équation de Boltzmann linéaire analogue à la formule des caractéristiques pour une équation de transport libre, il faut avoir recours à la théorie des processus stochastiques. Comme ces considérations sont à la base des méthodes de Monte-Carlo pour les équations de Boltzmann linéaires, qui comptent parmi les principales méthodes numériques utilisées pour résoudre numériquement ce type d'équation, nous allons en dire quelques mots. Nous n'entrerons pas dans le détail des notions du calcul des probabilités utilisées pour cela, et renvoyons au cours de C. Graham et D. Talay [26] pour plus de précisions sur ce thème.

Supposons pour fixer les idées que l'on veuille résoudre le problème de Cauchy pour l'équation de Boltzmann linéaire monocinétique dans $\mathbf{R}^3$ avec scatte-

ring isotrope

$$\begin{cases} \left(\frac{\partial}{\partial t} + v \cdot \nabla_x + \sigma\right) f(t,x,v) = \sigma \langle f \rangle (t,x) \quad (t,x,v) \in \mathbf{R}_+ \times \mathbf{R}^3 \times \mathbf{S}^2 \,, \\ f\big|_{t=0} = f^{in} \,, \end{cases}$$

où $\sigma > 0$ est une constante, et où

$$\langle f \rangle (t,x) = \tfrac{1}{4\pi} \int_{\mathbf{S}^2} f(t,x,w) ds(w) \,,$$

la notation $ds(w)$ désignant l'élément de surface sur la sphère unité $\mathbf{S}^2$ de $\mathbf{R}^3$.

Soit $(\tau_n)_{n\geq 0}$ suite de variables aléatoires à valeurs dans $\mathbf{R}_+$ distribuées selon la loi exponentielle de paramètre σ, c'est-à-dire que

$$\mathrm{Prob}(\tau_n > t) = \sigma e^{-\sigma t} \,, \quad t \geq 0 \,.$$

Notons

$$T_n = \sum_{j=0}^{n-1} \tau_j \,, \quad n \geq 1 \,.$$

Soit $(V_n)_{n\geq 1}$ suite de vecteurs aléatoires à valeurs dans $\mathbf{S}^2$ distribués uniformément, c'est-à-dire que, pour tout partie mesurable $A \subset \mathbf{S}^2$, on a

$$\mathrm{Prob}(V_n \in A) = \tfrac{1}{4\pi} \int_A dv \,.$$

Enfin, on suppose que les variables aléatoires $\tau_0, V_1, \tau_1, V_2, \ldots$ sont indépendantes.

On construit alors un processus stochastique, appelé "processus de transport", de la manière suivante.

Etant donné $(x,v) \in \mathbf{R}^3 \times \mathbf{S}^2$, on construit une trajectoire $(X_t, V_t)_{t\geq 0}$ partant de $(x,v) \in \mathbf{R}^3 \times \mathbf{S}^2$ et à valeurs dans $\mathbf{R}^3 \times \mathbf{S}^2$ comme suit :

pour $0 \leq t < T_1$: on pose

$$X_t = x - tv \,, \quad V_t = v \,;$$

pour $T_1 \leq t < T_2$: on pose

$$X_t = X_{T_1} - (t - T_1)V_1 \,, \quad V_t = V_1 \,;$$

pour $T_n \leq t < T_{n+1}$: on pose

$$X_t = X_{T_n} - (t - T_n)V_n \,, \quad V_t = V_n \,,$$

et ainsi de suite, pour tout $n \in \mathbf{N}^*$. On notera dans la suite $\mathbf{E}^{x,v}$ l'espérance sous la loi du processus de transport partant de $(x,v) \in \mathbf{R}^3 \times \mathbf{S}^2$.

Alors la fonction

$$f(t,x,v) = \mathbf{E}^{x,v}\left(f^{in}(X_t, V_t)\right)$$

est solution généralisée de l'équation de Boltzmann linéaire avec taux d'absorption σ et scattering isotrope.

Comparons cette formule avec la formule des caractéristiques

$$f(t,x,v) = f^{in}(x-tv,v)$$

pour la solution de l'équation de transport libre

$$\begin{cases} \dfrac{\partial f}{\partial t} + v\cdot\nabla_x f = 0\,, \quad x,v\in\mathbf{R}^3\,,\ t>0\,,\\ f\big|_{t=0} = f^{in}\,. \end{cases}$$

Dans les deux cas, la formule met en jeu la donnée initiale f^{in} calculée sur une courbe de l'espace des phases $\mathbf{R}^3_x\times\mathbf{R}^3_v$ issue de (x,v) à $t=0$. Alors que, dans le cas de l'équation de transport libre, la formule explicite met en jeu un unique segment de droite $[0,t]\ni s\mapsto x-sv\in\mathbf{R}^3\times\mathbf{R}^3$, dans le cas de l'équation de Boltzmann linéaire, la formule $f(t,x,v)=\mathbf{E}^{x,v}\left(f^{in}(X_t,V_t)\right)$ fait intervenir d'une part les lignes brisées $[0,t]\ni s\mapsto X_s$, qui sont paramétrées par les variables aléatoires $\tau_0,\tau_1,\dots$ et $V_1,V_2,\dots$, et d'autre part l'espérance $\mathbf{E}^{x,v}$ qui consiste à prendre une certaine moyenne sur l'ensemble de ces lignes brisées.

Donnons maintenant une idée de la démonstration de la formule explicite pour la solution de l'équation de Boltzmann linéaire. On a

$$\begin{aligned} f(t,x,v) &= \mathbf{E}^{x,v}\left(f^{in}(X_t,V_t)\mathbf{1}_{t<T_1}\right) + \mathbf{E}^{x,v}\left(f^{in}(X_t,V_t)\mathbf{1}_{t\geq T_1}\right)\\ &= f^{in}(x-tv)\mathbf{E}^{x,v}\left(\mathbf{1}_{t<T_1}\right) + \mathbf{E}^{x,v}\left(\mathbf{1}_{T_1\leq t}\mathbf{E}^{x,v}\left(f^{in}(X_t,V_t)|T_1,V_1\right)\right) \end{aligned}$$

Mais l'espérance conditionnelle sachant T_1,V_1 vaut

$$\mathbf{E}^{x,v}\left(f^{in}(X_t,V_t)|T_1\right) = f(t-T_1,X_{T_1},V_1)$$

puisque les variables aléatoires $\tau_0=T_1,V_1,\tau_1,V_2,\dots$ sont indépendantes, de sorte que, comme la loi jointe de (T_1,V_1) est la mesure de probabilité

$$\sigma e^{-\sigma t_1}\tfrac{1}{4\pi}dt_1 ds(v_1)\,,$$

on trouve que

$$\begin{aligned} f(t,x,v) &= f^{in}(x-tv)e^{-\sigma t}\\ &\quad + \iint_{\mathbf{R}_+\times\mathbf{S}^2}\mathbf{1}_{0\leq t_1\leq t}f(t-t_1,x-t_1v,v_1)\tfrac{1}{4\pi}\sigma e^{-\sigma t_1}dt_1dv_1\\ &= f^{in}(x-tv)e^{-\sigma t} + \int_0^t\left(\tfrac{1}{4\pi}\int_{\mathbf{S}^2}f(t-t_1,x-t_1v,v_1)dv_1\right)\sigma e^{-\sigma t_1}dt_1\\ &= f^{in}(x-tv)e^{-\sigma t} + \int_0^t\sigma e^{-\sigma t_1}\langle f\rangle(t-t_1,x-t_1v)dt_1\,. \end{aligned}$$

On retrouve ainsi l'équation intégrale équivalente (dans le cadre des solutions classiques) à l'équation de Boltzmann linéaire, ce qui montre, d'après le

Théorème 3.1.2, que la fonction f ainsi construite est bien la solution généralisée de l'équation de Boltzmann linéaire avec scattering isotrope.

La série de Duhamel s'interprète alors très simplement dans ce cadre. En effet

$$f(t,x,v) = \sum_{n\geq 0} \mathbf{1}_{[T_n,T_{n+1}[}(t) f(t,x,v)\,.$$

Par conséquent

$$f(t,x,v) = \sum_{n\geq 0} \mathbf{E}^{x,v}(\mathbf{1}_{[T_n,T_{n+1}[}(t) f(t,x,v))\,.$$

Or

$$\begin{aligned}&\mathbf{E}^{x,v}(\mathbf{1}_{[T_n,T_{n+1}[}(t) f(t,x,v))\\ &= \mathbf{E}^{x,v}(\mathbf{1}_{[T_n,T_{n+1}[}(t)\mathbf{E}^{x,v}(f^{in}(X_t,V_t)|\tau_0,\ldots,\tau_n,V_1,\ldots,V_n))\,.\end{aligned}$$

Pour $T_n \leq t < T_{n+1}$, la fonction $f^{in}(X_t,V_t)$ s'écrit

$$f^{in}(x-\tau_0 v-\tau_1 V_1-\tau_{n-1}V_{n-1}-(t-\tau_0-\ldots-\tau_{n-1})V_n,V_n)\,.$$

En particulier, $f^{in}(X_t,V_t)$ ne dépend que de $\tau_0,\ldots,\tau_{n-1},V_1,\ldots,V_n$, de sorte que

$$\begin{aligned}\mathbf{E}^{x,v}(f^{in}(X_t,V_t)|\tau_0,\ldots,\tau_n,V_1,\ldots,V_n) &= f^{in}(X_{T_n}-(t-T_n)V_n,V_n)\\ &= f^{in}(x-\tau_0 v-\tau_1 V_1-\tau_{n-1}V_{n-1}-(t-T_n)V_n,V_n)\,.\end{aligned}$$

Comme les variables aléatoires $\tau_0,\ldots,\tau_n,V_1,\ldots,V_n$ sont indépendantes, leur loi jointe vaut

$$(\tfrac{1}{4\pi})^n \sigma^{n+1} e^{-\sigma(t_0+\ldots+t_n)} dt_0\ldots dt_n ds(v_1)\ldots ds(v_n)\,.$$

Donc

$$\begin{aligned}&\mathbf{E}^{x,v}(\mathbf{1}_{[T_n,T_{n+1}[}(t)\mathbf{E}^{x,v}(f^{in}(X_t,V_t)|\tau_0,\ldots,\tau_n,V_1,\ldots,V_n))\\ &= (\tfrac{1}{4\pi})^n\sigma^{n+1}\int \mathbf{1}_{t_0+\ldots+t_{n-1}\leq t<t_0+\ldots+t_n} e^{-\sigma(t_0+\ldots+t_n)}\\ &f^{in}(x-t_0 v-t_1 v_1-t_{n-1}v_{n-1}-(t-t_0-\ldots-t_{n-1})v_n,v_n)\\ &\qquad dt_0\ldots dt_n ds(v_1)\ldots ds(v_n)\,.\end{aligned}$$

On calcule explicitement l'intégrale en la variable t_n :

$$\int \mathbf{1}_{t<t_0+\ldots+t_n}\sigma e^{-\sigma t_n} dt_n = e^{-\sigma(t-t_0-\ldots-t_{n-1})}$$

de sorte que l'expression au membre de droite ci-dessus se transforme en

$$\begin{aligned}&\mathbf{E}^{x,v}(\mathbf{1}_{[T_n,T_{n+1}[}(t)\mathbf{E}^{x,v}(f^{in}(X_t,V_t)|\tau_0,\ldots,\tau_n,V_1,\ldots,V_n))\\ &= (\tfrac{1}{4\pi})^n\sigma^n\int \mathbf{1}_{t_0+\ldots+t_{n-1}\leq t} e^{-\sigma t}\\ &f^{in}(x-t_0 v-t_1 v_1-t_{n-1}v_{n-1}-(t-t_0-\ldots-t_{n-1})v_n,v_n)\\ &\qquad dt_0\ldots dt_{n-1} ds(v_1)\ldots ds(v_n)\,.\end{aligned}$$

On reconnaît dans cette dernière expression le terme $\mathcal{T}^n F[f^{in},0]$ de la série de Duhamel, où on rappelle que

$$F[f^{in},0](t,x,v) := e^{-\sigma t} f^{in}(x-tv,v)\,,$$

et que

$$\mathcal{T}g(t,x,v) := \tfrac{1}{4\pi}\int_0^t e^{-\sigma(t-\tau)} \int_{\mathbf{S}^2} \sigma g(\tau, x+(\tau-t)v, w) ds(w) d\tau\,.$$

Autrement dit, le n-ième terme de la série de Duhamel correspond à la contribution du n-ième segment de la ligne brisée $\mathbf{R}_+ \ni t \mapsto X_t \in \mathbf{R}^3$ dans le calcul de l'espérance donnant la solution de l'équation de Boltzmann linéaire.

3.4 Le problème stationnaire pour l'équation de Boltzmann linéaire

Nous allons terminer ce chapitre avec quelques remarques sur le problème aux limites pour l'équation de Boltzmann stationnaire.

Nous gardons dans ce qui suit les hypothèses de la section 3.2, sauf que les fonctions $a \equiv a(x,v)$ (taux d'amortissement) et $k \equiv k(x,v,w)$ (noyau de transition $w \to v$) sont indépendantes de la variable de temps t.

On considère donc le problème aux limites

$$\begin{cases} (\lambda + v\cdot\nabla_x + a(x,v))F(x,v) = \mathcal{K}F(x,v) + Q(x,v)\,, \quad (x,v)\in\Omega\times\mathbf{R}^N\,, \\ F\big|_{\Gamma_-} = F_b^- \end{cases}$$

avec la notation

$$\mathcal{K}F(x,v) = \int_{\mathbf{R}^N} k(x,v,w)F(x,w)d\mu(w)\,.$$

Théorème 3.4.1 *Soit*

$$D = \|(\mathcal{K}1 - a)^+\|_{L^\infty(\Omega\times\mathbf{R}^N)}\,.$$

Pour tout $\lambda > D$, le problème aux limites pour l'équation de Boltzmann linéaire stationnaire ci-dessus admet une unique solution, qui vérifie

$$\|F\|_{L^\infty(\Omega\times\mathbf{R}^N)} \le \frac{1}{\lambda - D}\max\left(\|Q\|_{L^\infty(\Omega\times\mathbf{R}^N)}, \lambda\|F_b^-\|_{L^\infty(\Gamma_-)}\right)\,.$$

Comme on l'a expliqué au chapitre 2, le problème ci-dessus est obtenu à partir du problème aux limites d'évolution

$$\begin{cases} \left(\frac{\partial}{\partial t} + v\cdot\nabla_x + a(x,v)\right) f(t,x,v) = \mathcal{K}f(t,x,v)\,, \quad (t,x,v)\in\mathbf{R}_+\times\Omega\times\mathbf{R}^N\,, \\ f(t,\cdot,\cdot)\big|_{\Gamma_-} = \lambda F_b^-\,, \\ f\big|_{t=0} = Q\,, \end{cases}$$

par la transformation de Laplace

$$F(x,v) = \int_0^{+\infty} e^{-\lambda t} f(t,x,v) dt \,.$$

La solution du problème d'évolution vérifie l'estimation L^∞ suivante :

$$\|f(t,\cdot,\cdot)\|_{L^\infty(\Omega\times\mathbf{R}^N)} \leq \max\left(\|Q\|_{L^\infty(\Omega\times\mathbf{R}^N)}, \lambda\|F_b^-\|_{L^\infty(\Gamma_-)}\right) e^{Dt} \,,$$

— cf. Proposition 3.1.5 — de sorte que la transformée de Laplace de f est définie pour tout $\lambda > D$. Le théorème ci-dessus se démontre alors sans difficulté, et sa preuve est laissée au lecteur à titre d'exercice.

3.5 Exercices

Exercice 3.1 (Existence et unicité ; cadre simplifié) *Considérons l'équation du transport en dimension deux pour un gaz de photons de fonction de répartition $f(x,t,\theta)$*

$$\left(\frac{1}{c}\frac{\partial f}{\partial t} + \omega\frac{\partial f}{\partial x} + \sigma f\right)(t,x,\theta) = \frac{\sigma_s}{2\pi}\int_0^{2\pi} f(t,x,\theta')d\theta' \,,$$
$$x \in \mathbf{R}^2 \,,\ \theta \in [0,2\pi[\,,\ t > 0 \,,$$

avec $\omega = (\cos\theta, \sin\theta)$, la direction d'un photon, et des coefficients constants $\sigma, \sigma_s > 0$. On suppose pour simplifier que la donnée initiale est périodique

$$f_0(x_1+1,x_2,\theta) = f_0(x_1,x_2+1,\theta) = f_0(x_1,x_2,\theta) \,, \qquad \text{pour tous } x_1, x_2 \text{ et } \theta \,.$$

On se ramène ainsi à étudier le problème sur $\mathbf{T} = [0,1]\times[0,1]\times[0,2\pi]$ avec des conditions de périodicité au bord.

1. *Soit C un partie mesurable quelconque de $\mathbf{R}^N$. On rappelle l'inégalité de Hölder : pour tous $p, q \geq 1$ tels que $\frac{1}{p}+\frac{1}{q} = 1$ et pour tout couple (u,v) de fonctions mesurables sur C à valeurs réelles ou complexes, l'on a*

$$\int_C |u(x)v(x)|dx \leq \left(\int_C |u(x)|^p dx\right)^{\frac{1}{p}} \left(\int_C |v(x)|^q dx\right)^{\frac{1}{q}} \,.$$

Lorsque $p, q \geq 1$ vérifient $\frac{1}{p}+\frac{1}{q} = 1$, on dit que p et q sont des exposants conjugués. Lorsque $p = \infty$ et $q = 1$, on a encore $\frac{1}{p}+\frac{1}{q} = 1$ avec la convention $1/\infty = 0$, et l'inégalité de Hölder s'écrit[3]

$$\int_C |u(x)v(x)|dx \leq \operatorname*{supess}_{x\in C} |u(x)| \int_C |v(x)|dx \,.$$

3. On rappelle que la borne supérieure essentielle d'une fonction f mesurable sur C à valeurs réelles, notée $\operatorname{supess}_{x\in C} f(x)$, est le plus petit réel M tel que $f(x) \leq M$ pour presque tout $x \in C$.

Montrer que pour $p \geq 1$

$$\frac{1}{2\pi}\int_0^{2\pi} |f(t,x,\theta)|d\theta \leq \left(\frac{1}{2\pi}\int_0^{2\pi} |f(x,t,\theta)|^p d\theta\right)^{\frac{1}{p}} .$$

2. *On considère la suite* f^k *définie par*

$$\left(\frac{1}{c}\frac{\partial f^{k+1}}{\partial t} + \omega\cdot\nabla f^{k+1} + \sigma f^{k+1}\right)(t,x,\theta) = \frac{\sigma_s}{2\pi}\int_0^{2\pi} f^k(x,t,\theta')d\theta'$$

avec la condition initiale $f\big|_{t=0} = f_0$. *On pose* $\mathcal{D} = [0,T]\times\mathbf{T}$ *pour* $T>0$. *Montrer que*

$$\|f^{k+1}\|_{L^p(\mathcal{D})} \leq E\|f_0\|_{L^p(\mathbf{T})} + D\|f^k\|_{L^p(\mathcal{D})}$$

avec $E \geq 0$ *et* $0 < D < 1$ *dans le cas* $\sigma > \sigma_s$.

3. *Sous cette hypothèse, montrer que la suite* $(f^k)_{k\geq 1}$ *converge dans* $L^p(\mathcal{D})$ *vers une limite notée* f, *qui satisfait une équation que l'on donnera.*
4. *En quoi les résultats de la section 3.2 sont-ils plus puissants ?*
5. *Montrer que si* $f_0 \geq 0$ *alors* $f \geq 0$.
6. *Que se passe-t-il si* $\mathbf{T}$ *est remplacé par un domaine borné avec des conditions au bord de Dirichlet, de réflexion diffusive ou spéculaire ?*

Indications : cet exercice construit une solution par une technique de suite contractante, l'inégalité du 2 impliquant que

$$\|f^{k+2} - f^{k+1}\|_{L^p(\mathcal{D})} \leq D\|f^{k+1} - f^k\|_{L^p(\mathcal{D})}$$

pour des données initiales quelconques. Les restrictions sont : (a) f^{k+1} est construit a priori à l'aide de la méthode des caractéristiques ; (b) la convergence a lieu dans $\mathcal{D}$, donc sur un domaine qui est borné en temps.

Exercice 3.2 (Le contre-exemple de Jeffrey Rauch) *On pourrait penser que l'intuition physique permet de se passer d'une analyse rigoureuse de convergence. Il n'en est rien, comme le démontre ce contre-exemple.*

Pour $v \neq 0$ *et* $\sigma > 0$, *soit l'équation*

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} + \sigma f = 0\,, \quad x\in\mathbf{R}\,,\ t>0\,.$$

La donnée initiale $f_0(x) = a(x)\cos\left(\frac{x}{\varepsilon}\right)$ *correspond à un paquet d'ondes de petite longueur d'onde* $\varepsilon > 0$. *La fonction* a *est de classe* C^∞ *à support compact sur* $\mathbf{R}$.

1. *Ecrire la solution exacte du problème de Cauchy ci-dessus.*

2. *Vérifier que*

$$\frac{\partial f}{\partial t} = O(1/\varepsilon)\,, \quad \frac{\partial f}{\partial x} = O(1/\varepsilon)\,, \quad f = O(1)\,.$$

En "déduire" que f peut s'approcher par g, solution de l'équation

$$\frac{\partial g}{\partial t} + v\frac{\partial g}{\partial x} = 0\,, \quad x \in \mathbf{R}\,,\ t > 0\,,$$

avec la condition initiale $g\big|_{t=0} = f_0$.

3. *Calculer g. La démarche est-elle correcte ?*
4. *Pour comprendre l'erreur, il faut mettre en place une stratégie d'étude de l'écart $e = g - f$. Pour cela on commence par déterminer le terme source b tel que*

$$\frac{\partial e}{\partial t} + v\frac{\partial e}{\partial x} + \sigma e = b\,.$$

Montrer ensuite que

$$\frac{d}{dt}\|e(t)\|^2_{L^2(\mathbf{R})} \leq 2\|e(t)\|_{L^2(\mathbf{R})}\|b(t)\|_{L^2(\mathbf{R})}\,.$$

En déduire que e est petit dès que b l'est. En déduire l'explication du "paradoxe" ci-dessus.

Exercice 3.3 (Un problème d'absorption très simple) *Soit une colonne ($x \geq 0$) d'un gaz à température uniforme T. Le rayonnement $I(t, x; \nu, \theta)$ ($0 < \nu < \infty$ et $0 \leq \theta < 2\pi$) entre en $x = 0$. D'où l'équation*

$$\frac{1}{c}\frac{\partial I}{\partial t} + \cos\theta\frac{\partial I}{\partial x} = \sigma\left(B_\nu(T) - I\right)$$

où $\sigma > 0$ et $B_\nu(T)$ est la fonction de Planck (voir la section 1.3.1). On s'intéresse aux solutions stationnaires et on étudie la pénétration du rayonnement dans la colonne.

1. *Ecrire le problème stationnaire qu'il faut résoudre. Montrer que si I est bornée,*

$$I(x; \nu, \theta) = B_\nu(T) \text{ pour } \frac{\pi}{2} < \theta < \frac{3\pi}{2}\,.$$

2. *Montrer que le rayonnement devient rapidement Planckien en fonction du produit σx pour $x > 0$.*

Indications : montrer que $u(x) = I(x; \nu, \theta) - B_\nu(T)$ vérifie $\cos\theta u'(x) + u(x) = 0$. Ecrire la solution exacte et montrer que si $\cos\theta < 0$, alors $u \equiv 0$ en considérant la condition physique à l'infini. Dans le cas $\cos\theta > 0$ montrer que $|u(x)| \to 0$ pour x croissant.

Exercice 3.4 (Principe du maximum et conditions au bord) *On considère l'équation du transport*

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} + \sigma f = Q(f)\,,$$

avec $\sigma > 0$*, une donnée initiale* f_0 *et*

$$Q(f) = \sigma \frac{\int_{|v|=1} f dv}{\int_{|v|=1} dv} = \sigma \langle f \rangle\,,$$

pour des particules monocinétiques où $|v| = 1$ *(pour simplifier). On suppose que* x *décrit un domaine borné* $\Omega \subset \mathbf{R}^N$*.*

1. *On considère les conditions au bord usuelles de Dirichlet ou de réflexion. Les écrire.*
2. *Soit* $f \mapsto \varphi(f)$ *une fonction positive deux fois dérivable avec* $\varphi''(f) \geq 0$*,* $\varphi(f) \geq 0$ *et* $\varphi(0) = 0$*. Montrer que*

$$\frac{d}{dt}\int_\Omega \int_{|v|=1} \varphi(f) dv dx \leq 0\,.$$

3. *Montrer que, pour* $t \geq 0$*,*

$$\int_\Omega \int_{|v|=1} \varphi(f(t)) dv dx \leq \int_\Omega \int_{|v|=1} \varphi(f_0) dv dx$$

   lorsque $\varphi(f) = \max(0, -f)$*. En déduire que si* $f_0 \geq 0$ *alors* $f(t) \geq 0$ *pour tout* $t \geq 0$*.*

4. *Que dire pour la borne supérieure de* f *?*

Remarque : cette méthode est issue des techniques d'entropie pour les lois de conservation non linéaires. Les fonctions convexes φ sont appelées "entropies" dans la théorie des lois de conservation.

Chapitre 4

Limite de diffusion

Dans ce chapitre, nous allons expliquer en détail les relations existant entre les modélisations du transport de particules par l'équation de Boltzmann linéaire et par l'équation de diffusion.

4.1 Rappels sur l'équation de diffusion

Soit Ω ouvert borné de $\mathbf{R}^N$ à bord de classe C^∞. Comme on l'a fait jusqu'ici, nous noterons dans toute la suite de ce chapitre n_x le vecteur unitaire normal au bord $\partial\Omega$ au point $x \in \partial\Omega$ dirigé vers l'extérieur de Ω.

Nous allons commencer par quelques rappels nécessaires des résultats fondamentaux portant sur l'équation de la chaleur. Ces résultats se trouvent par exemple dans [23] (chapitre 8) ou dans [2] (chapitre 8, §8.4).

4.1.1 Equation de la chaleur avec condition de Dirichlet

Soit $\kappa \in \mathbf{R}^*$; on considère le problème aux limites suivant, d'inconnue $u \equiv u(t,x) \in \mathbf{R}$, où $t \geq 0$ et $x \in \Omega$:

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) - \frac{1}{2}\kappa^2 \Delta_x u(t,x) = 0\,, \quad x \in \Omega\,,\ t > 0\,, \\ u\big|_{\partial\Omega} = u_b\,, \\ u\big|_{t=0} = u^{in}\,. \end{cases}$$

Les données de ce problème sont
a) la donnée initiale $u^{in} \equiv u^{in}(x)$, et
b) la donnée au bord $u_b \equiv u_b(t,x)$.

La condition au bord

$$u(t,y) = u_b(t,y)\,, \quad y \in \partial\Omega\,,\ t \geq 0\,,$$

consistant à prescrire la valeur de la solution sur $\partial\Omega$ pour tout temps, porte le nom de "condition de Dirichlet".

Voici comment on peut résumer la théorie d'existence et d'unicité de la solution de l'équation de diffusion ci-dessus.

Théorème 4.1.1 *Supposons que $u^{in} \in C^\infty(\overline{\Omega})$ et que $u_b \in C^\infty([0,T] \times \partial\Omega)$ vérifient les relations de compatibilité suivantes pour tout $k \geq 0$:*

$$\frac{\partial^k u_b}{\partial t^k}(0,y) = \left(\tfrac{1}{2}\kappa^2 \Delta_x\right)^k u^{in}(y)\,, \quad y \in \partial\Omega\,.$$

Il existe une unique solution $u \in C^\infty([0,T] \times \overline{\Omega})$ au problème de Dirichlet

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) - \frac{1}{2}\kappa^2 \Delta_x u(t,x) = 0\,, \quad x \in \Omega\,,\ t > 0\,, \\ u\big|_{\partial\Omega} = u_b\,, \\ u\big|_{t=0} = u^{in}\,. \end{cases}$$

Pour avoir une idée de la preuve, on pourra se reporter aux §8.2 et 8.4 de [2].

Afin d'estimer la taille de la solution — et de ses dérivées successives — nous utiliserons de façon systématique le principe du maximum faible, que nous rappelons ci-dessous.

Théorème 4.1.2 (Principe du maximum) *Sous les hypothèses du théorème précédent, la solution u du problème de Dirichlet pour l'équation de la chaleur*

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) - \frac{1}{2}\kappa^2 \Delta_x u(t,x) = 0\,, \quad x \in \Omega\,,\ t > 0\,, \\ u\big|_{\partial\Omega} = u_b\,, \\ u\big|_{t=0} = u^{in}\,, \end{cases}$$

vérifie l'estimation L^∞ suivante :

$$\|u\|_{L^\infty([0,T]\times\Omega)} \leq \max\left(\|u^{in}\|_{L^\infty(\Omega)}, \|u_b\|_{L^\infty([0,T]\times\partial\Omega)}\right)\,,$$

ainsi que la propriété de positivité :

$$\left.\begin{array}{l} u^{in} \geq 0 \text{ sur } \overline{\Omega} \\ u_b \geq 0 \text{ sur } [0,T] \times \partial\Omega \end{array}\right\} \Rightarrow u \geq 0 \text{ sur } [0,T] \times \overline{\Omega}\,.$$

Pour une démonstration de ce résultat, voir la Proposition 8.4.2 de [2].

4.1.2 Le problème de Cauchy pour l'équation de la chaleur dans $\mathbf{R}^N$

Les énoncés sur l'équation de la chaleur rappelés dans la section précédente sont obtenus à partir d'une formulation variationnelle du problème aux limites dans un espace de Sobolev — ou bien par une méthode de projection en dimension finie de type Galerkin.

Lorsque $\Omega = \mathbf{R}^N$, on dispose d'une méthode directe de résolution de l'équation de la chaleur en utilisant une formule explicite basée sur la notion de solution élémentaire de l'opérateur de la chaleur — voir par exemple [23], chapitre 8.

On considère donc le problème de Cauchy d'inconnue $u \equiv u(t,x)$, soit

$$\left\{ \begin{array}{l} \frac{\partial u}{\partial t}(t,x) - \frac{1}{2}\kappa^2 \Delta_x u(t,x) = 0\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,, \\ \\ u\big|_{t=0} = u^{in}\,, \end{array} \right.$$

avec $u \equiv u^{in}(x)$ donnée.

Rappelons la formule explicite donnant la solution élémentaire dans le futur pour l'opérateur de la chaleur, solution élémentaire qui est la densité gaussienne centrée, de matrice de covariance tI :

$$E(t,x) := \frac{1}{(2\pi t)^{N/2}} \exp\left(-\frac{|x|^2}{2t}\right)\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,.$$

Théorème 4.1.3 *Pour tout $u^{in} \in L^2(\mathbf{R}^N)$, le problème de Cauchy ci-dessus admet une unique solution $u \in C^\infty(\mathbf{R}_+^* \times \mathbf{R}^N)$, solution qui est donnée par la formule*

$$u(t,x) = \int_{\mathbf{R}^N} E(\kappa^2 t, x-y) u^{in}(y) dy\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,.$$

Si de plus $u^{in} \in C^\infty(\mathbf{R}^N)$, alors la solution $u \in C^\infty(\mathbf{R}_+ \times \mathbf{R}^N)$.

Voir par exemple le §8.2 de [23] pour une démonstration de ce résultat.

Comme dans le cas du problème de Dirichlet posé dans un domaine borné de $\mathbf{R}^N$, on estimera la taille de la solution ci-dessus au moyen du principe du maximum, lorsque u^{in} appartient à $L^\infty(\mathbf{R}^N)$.

Théorème 4.1.4 (Principe du maximum) *La solution $u \in C^\infty(\mathbf{R}_+^* \times \mathbf{R}^N)$ du problème de Cauchy*

$$\left\{ \begin{array}{l} \frac{\partial u}{\partial t}(t,x) - \frac{1}{2}\kappa^2 \Delta_x u(t,x) = 0\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,, \\ \\ u\big|_{t=0} = u^{in}\,, \end{array} \right.$$

pour l'équation de la chaleur vérifie l'encadrement suivant :

$$\inf_{y \in \mathbf{R}^N} u^{in}(y) \leq u(t,x) \leq \sup_{y \in \mathbf{R}^N} u^{in}(y)\,,$$

pour tout $(t,x) \in \mathbf{R}_+ \times \mathbf{R}^N$.

Voir [23], Théorème 8.3.1, pour une démonstration de ce résultat — qui n'utilise que le fait que, pour tout $t > 0$, la fonction $x \mapsto E(t,x)$ est une densité de probabilité sur $\mathbf{R}^N$.

4.2 Approximation par la diffusion : calculs formels

Dans tout ce qui suit, on supposera que l'ensemble des vitesses admissibles pour les particules considérées ici est $v \in \mathcal{V} = \overline{B(0,R)}$, avec $R > 0$ — ou un de ses sous-ensembles mesurables invariants par le groupe orthogonal $O_N(\mathbf{R})$.

On étudiera dans tout ce paragraphe l'équation de Boltzmann linéaire suivante

$$\left(\frac{\partial}{\partial t} + v \cdot \nabla_x\right) f + af = a(1+\gamma)\mathcal{K}f$$

d'inconnue $f \equiv f(t,x,v)$, avec $t \geq 0$, $x \in \overline{\Omega}$ et $v \in \mathcal{V}$, où $a > 0$ et γ sont deux constantes réelles, et où le terme de création de particules est de la forme

$$\mathcal{K}f(t,x,v) = \int_{\mathcal{V}} k(v,w) f(t,x,w) d\mu(w)\,.$$

Dans toute la suite de chapitre, nous ferons les hypothèses suivantes sur les différentes données (a, γ, μ et k) intervenant dans l'équation de Boltzmann ci-dessus.

Hypothèses :
(H1) on suppose que, pour tout $\phi \in C(\mathcal{V})$

$$\phi(v) \geq 0 \text{ pour tout } v \in \mathcal{V} \Rightarrow \int_{\mathcal{V}} \phi(v) d\mu(v) \geq 0\,,$$

et que

$$0 < \int_{\mathcal{V}} d\mu(v) < +\infty \qquad \text{et} \qquad \int_{\mathcal{V}} v d\mu(v) = 0\,;$$

(H2) on suppose que $k \in C(\mathcal{V} \times \mathcal{V})$ vérifie

$$\begin{cases} 0 < k(v,w) = k(w,v) \text{ pour tout } v, w \in \mathcal{V}\,, \\ \\ \mathcal{K}1(v) = \displaystyle\int_{\mathcal{V}} k(v,w) d\mu(w) = 1 \text{ pour tout } v \in \mathcal{V}\,. \end{cases}$$

En pratique, la notation $d\mu(v)$ désignera
(a) soit une mesure ayant une densité par rapport à la mesure de Lebesgue sur $\mathcal{V}$, autrement dit

$$\int_{\mathcal{V}} \phi(v) d\mu(v) = \int_{\mathcal{V}} \phi(v) E(v) dv$$

où $E \in L^1(\mathcal{V})$ vérifie $m \geq 0$ p.p. sur $\mathcal{V}$;
b) soit l'élément de surface sur une sphère centrée en 0 de rayon $< R$, donc contenue dans $\mathcal{V}$.

4.2.1 Loi d'échelle de la diffusion

Comme on l'a déjà suggéré dans notre présentation heuristique de l'approximation par la diffusion pour l'équation de Boltzmann linéaire monocinétique avec scattering isotrope au chapitre 1, les descriptions d'une population de particules par l'équation de diffusion et par l'équation de Boltzmann linéaire coïncident seulement dans un certain régime asymptotique que nous allons examiner plus en détail.

Plus précisément, l'approximation par la diffusion de l'équation de Boltzmann linéaire concerne un régime asymptotique où

(1) $a \gg 1$ (correspondant à un taux élevé d'échange par absorption/création dans le milieu hôte),

(2) $\gamma \ll 1$ (correspondant à un régime de quasi-équilibre entre absorption et création de particules),

(3) $|\frac{\partial f}{\partial t}| \ll 1$ (correspondant à une dynamique lentement variable en temps autour de ce quasi-équilibre).

Nous commenterons plus loin ces différentes hypothèses d'échelle, au cours de notre dérivation systématique de l'équation de diffusion à partir de l'équation de Boltzmann linéaire.

Les trois hypothèses asymptotiques ci-dessus sont réalisées au moyen d'un seul petit paramètre $0 < \epsilon \ll 1$, comme suit :

(1') on suppose que a est de la forme

$$a = \frac{\hat{a}}{\epsilon}\,,$$

avec $\hat{a} > 0$ de l'ordre de l'unité ;

(2') on suppose que γ est de la forme

$$\gamma = \epsilon^2 \hat{\gamma}\,,$$

avec $\hat{\gamma}$ de l'ordre de l'unité ;

(3') on suppose enfin que

$$f(t,x,v) = f_\epsilon(\epsilon t, x, v)\,,$$

avec

$$f_\epsilon \equiv f_\epsilon(\hat{t},x,v)\,, \quad \text{et} \quad \frac{\partial f_\epsilon}{\partial \hat{t}} = O(1)\,.$$

Sous ces hypothèses, l'équation de Boltzmann linéaire devient donc

$$\epsilon \frac{\partial f_\epsilon}{\partial \hat{t}} + v \cdot \nabla_x f_\epsilon + \frac{\hat{a}}{\epsilon}\Big(f_\epsilon - (1+\epsilon^2\hat{\gamma})\mathcal{K} f_\epsilon\Big) = 0\,.$$

Dans ce qui suit on va étudier le comportement asymptotique des solutions de cette équation lorsque le petit paramètre $\epsilon \to 0^+$.

A partir de maintenant, on va d'ailleurs oublier les paramètres originels a et γ et ne considérer que l'équation ci-dessus, écrite après mise à l'échelle au moyen du petit paramètre ϵ. C'est pourquoi on oubliera les notations $\hat{a}$, $\hat{\gamma}$ et $\hat{t}$ et on notera ces quantités a, γ et t respectivement.

Autrement dit, on étudiera dans la suite de ce chapitre le comportement asymptotique de familles de solutions f_ϵ de l'équation de Boltzmann linéaire

$$\epsilon \frac{\partial f_\epsilon}{\partial t} + v \cdot \nabla_x f_\epsilon + \frac{a}{\epsilon}\Big(f_\epsilon - (1+\epsilon^2\gamma)\mathcal{K} f_\epsilon \Big) = 0\,. \tag{4.1}$$

4.2.2 Solution série formelle de Hilbert

On cherche pour commencer une solution f_ϵ de l'équation

$$\epsilon \frac{\partial f_\epsilon}{\partial t} + v \cdot \nabla_x f_\epsilon + \frac{a}{\epsilon}\Big(f_\epsilon - (1+\epsilon^2\gamma)\mathcal{K} f_\epsilon \Big) = 0$$

qui soit une série **formelle** en puissances de ϵ à coefficients réguliers (de classe C^∞ en t, x et continus en v) :

$$f_\epsilon(t,x,v) = \sum_{n\geq 0} \epsilon^n f_n(t,x,v) \in A[[\epsilon]]\,.$$

La notation $f_\epsilon \in A[[\epsilon]]$ signifie que $f_n \in A$ pour tout $n \geq 0$, où

$$A := \left\{ \phi \equiv \phi(t,x,v) \in C_b(\mathbf{R}_+ \times \Omega \times \mathcal{V}) \,|\, \frac{\partial^\alpha}{\partial t^\alpha}\frac{\partial^\beta}{\partial x^\beta}\phi \in C_b(\mathbf{R}_+ \times \Omega \times \mathcal{V}) \right\}.$$

L'idée de chercher une solution de l'équation de Boltzmann linéaire qui soit une série formelle en le paramètre ϵ est due à Hilbert [27]. L'idée originale de Hilbert était d'appliquer la théorie des équations intégrales à la construction de solutions de l'équation de Boltzmann de la théorie cinétique des gaz au voisinage d'états déquilibre (qui sont des fonctions de distribution maxwelliennes, c'est-à-dire proportionnelles à une gaussienne en v de matrice de covariance de la forme θI).

On prendra bien garde au fait suivant : cette série ne converge, en général, pour aucune valeur de $\epsilon > 0$. D'ailleurs ce n'est pas grave : seuls les premiers termes de cette série formelle nous serviront pour justifier l'approximation de l'équation de Boltzmann linéaire par l'équation de diffusion.

Quoiqu'il en soit, on commence par écrire que cette série formelle est solution de l'équation de Boltzmann linéaire ci-dessus paramétrée par ϵ. C'est-à-dire qu'on remplace dans l'équation de Boltzmann la solution f_ϵ par la série formelle, et que l'on identifie les coefficients des puissances successives de ϵ.

On trouve donc, ordre par ordre en ϵ, la hiérarchie infinie d'équations suivante :

Ordre ϵ^{-1} :

$$a(f_0 - \mathcal{K} f_0) = 0\,;$$

Ordre ϵ^0 :

$$v \cdot \nabla_x f_0 + a(f_1 - \mathcal{K} f_1) = 0\,;$$

Ordre ϵ^1 :

$$\frac{\partial f_0}{\partial t} + v \cdot \nabla_x f_1 + a(f_2 - \mathcal{K} f_2) - a\gamma \mathcal{K} f_0 = 0\,.$$

De façon générale, on trouve, pour tout $n \geq 1$, la relation

Ordre ϵ^n :

$$\frac{\partial f_{n-1}}{\partial t} + v \cdot \nabla_x f_n + a(f_{n+1} - \mathcal{K} f_{n+1}) - a\gamma \mathcal{K} f_{n-1} = 0\,.$$

Avant d'aller plus loin et d'étudier en détail chacune de ces équations, remarquons qu'elles sont toutes de la même forme, comme nous allons le montrer ci-dessous.

Pour tout $n \geq 1$, supposons que les termes $f_0, \ldots, f_n$ sont déjà connus. L'équation correspondant aux termes d'ordre ϵ^n dans l'équation de Boltzmann linéaire est de la forme

$$a(I - \mathcal{K}) f_{n+1} = S[f_{n-1}, f_n]$$

où $S[f_{n-1}, f_n]$ est une expression fonctionnelle connue des termes f_{n-1} et f_n, supposés connus également.

L'égalité ci-dessus est donc une équation d'inconnue f_{n+1} permettant de déterminer f_{n+1} une fois que f_{n-1} et f_n ont été calculées. (En réalité, les choses sont, comme on va le voir, un peu plus compliquées que cela.) En tout cas, cette remarque montre qu'il sera naturel d'étudier en détail l'équation intégrale d'inconnue ϕ

$$(I - \mathcal{K})\phi = \Sigma\,,$$

où Σ est une fonction de v donnée. Nous y reviendrons plus loin.

L'équation à l'ordre ϵ^{-1} :

Commençons par étudier la variante homogène de cette équation intégrale

$$a(f_0 - \mathcal{K} f_0) = 0\,.$$

Lemme 4.2.1 *Pour tout $\phi \in L^2(\mathcal{V}, d\mu)$, on a*

$$\int_{\mathcal{V}} \phi(v)(I - \mathcal{K})\phi(v) d\mu(v) \geq 0\,,$$

avec égalité si et seulement si $\phi =$ Const. μ-p.p. sur $\mathcal{V}$.

En particulier, si $\phi = \mathcal{K}\phi$, alors $\phi =$ Const. μ-p.p. sur $\mathcal{V}$, de sorte que

$$\mathrm{Ker}(I - \mathcal{K}) = \{ \textit{ fonctions constantes } \mu\textit{-p.p. sur } \mathcal{V} \}\,.$$

Démonstration. D'une part, en appliquant le théorème de Fubini,

$$\begin{aligned}\int_{\mathcal{V}} \phi(v)\mathcal{K}\phi(v)d\mu(v) &= \int_{\mathcal{V}} \phi(v)\left(\int_{\mathcal{V}} k(v,w)\phi(w)d\mu(w)\right)d\mu(v)\\ &= \iint_{\mathcal{V}\times\mathcal{V}} k(v,w)\phi(v)\phi(w)d\mu(v)d\mu(w)\,.\end{aligned}$$

Calculons ensuite, toujours en appliquant le théorème de Fubini,

$$\begin{aligned}&\iint_{\mathcal{V}\times\mathcal{V}} k(v,w)\tfrac{1}{2}\left(\phi(v)^2+\phi(w)^2\right)d\mu(v)d\mu(w)\\ &\qquad= \tfrac{1}{2}\int_{\mathcal{V}}\left(\int_{\mathcal{V}} k(v,w)d\mu(w)\right)\phi(v)^2 d\mu(v)\\ &\qquad+ \tfrac{1}{2}\int_{\mathcal{V}}\left(\int_{\mathcal{V}} k(v,w)d\mu(v)\right)\phi(w)^2 d\mu(w)\\ &\qquad= \tfrac{1}{2}\int_{\mathcal{V}}\phi(v)^2 d\mu(v)+\tfrac{1}{2}\int_{\mathcal{V}}\phi(w)^2 d\mu(w)\\ &\qquad= \int_{\mathcal{V}}\phi(u)^2 d\mu(u)\,,\end{aligned}$$

puisque, d'après l'hypothèse (H2) sur k

$$\begin{cases}\displaystyle\int_{\mathcal{V}} k(v,w)d\mu(w)=1 \text{ pour tout } v\in\mathcal{V},\\[2ex] \displaystyle\int_{\mathcal{V}} k(v,w)d\mu(v)=1 \text{ pour tout } w\in\mathcal{V}.\end{cases}$$

Au total

$$\begin{aligned}&\int_{\mathcal{V}}\phi(v)(I-\mathcal{K})\phi(v)d\mu(v)\\ &\qquad= \tfrac{1}{2}\iint_{\mathcal{V}\times\mathcal{V}} k(v,w)\left(\phi(v)^2+\phi(w)^2-2\phi(v)\phi(w)\right)d\mu(v)d\mu(w)\\ &\qquad= \tfrac{1}{2}\iint_{\mathcal{V}\times\mathcal{V}} k(v,w)\left(\phi(v)-\phi(w)\right)^2 d\mu(v)d\mu(w)\,.\end{aligned}$$

Comme $k>0$ sur $\mathcal{V}\times\mathcal{V}$, l'identité ci-dessus montre que

$$\int_{\mathcal{V}}\phi(v)(I-\mathcal{K})\phi(v)d\mu(v)\geq 0$$

pour tout $\phi\in L^2(\mathcal{V},d\mu)$, avec égalité si et seulement si [1]

$$k(v,w)\left(\phi(v)-\phi(w)\right)^2=0 \quad \mu\otimes\mu\text{-p.p. en } (v,w)\in\mathcal{V}\times\mathcal{V}\,.$$

1. Dans le cas où

$$\int_{\mathcal{V}}\phi(v)d\mu(v) \text{ est de la forme } \int_{\mathcal{V}}\phi(v)E(v)dv$$

Comme $k > 0$ sur $\mathcal{V} \times \mathcal{V}$, cette condition se ramène à la condition

$$\phi(v) - \phi(w) = 0 \quad \mu \otimes \mu\text{-p.p. en } (v, w) \in \mathcal{V} \times \mathcal{V}\,.$$

Intégrons cette identité en w :

$$\phi(v) \int_{\mathcal{V}} d\mu(w) - \int_{\mathcal{V}} \phi(w) d\mu(w) = 0 \quad \mu\text{-p.p. en } v \in \mathcal{V}\,,$$

c'est-à-dire que

$$\phi(v) = \langle \phi \rangle \ \mu\text{-p.p. en } v \in \mathcal{V}.$$

On vient donc de démontrer que

$$\mathrm{Ker}(I - \mathcal{K}) \subset \{\text{fonctions constantes } \mu\text{-p.p. sur } \mathcal{V}\}\,.$$

D'autre part,

$$\mathcal{K}1 = \int_{\mathcal{V}} k(v, w) d\mu(w) = 1$$

d'après l'hypothèse (H2) sur k, de sorte que

$$\{\text{fonctions constantes } \mu\text{-p.p. sur } \mathcal{V}\} \subset \mathrm{Ker}(I - \mathcal{K})\,.$$

Ceci conclut la démonstration. ■

Revenons à l'équation

$$f_0 - \mathcal{K} f_0 = 0\,.$$

Comme $f_0 \equiv f_0(t, x, v)$ est une fonction continue bornée des variables t, x, v, alors en particulier $f_0(t, x, \cdot) \in L^2(\mathcal{V}, d\mu)$ pour tous t, x puisque μ est une mesure positive bornée d'après l'hypothèse (H1). Donc

$$f_0 \equiv f_0(t, x) \text{ est indépendante de } v\,.$$

L'équation à l'ordre ϵ^0 :

Passons maintenant à la 2ème équation :

$$(I - \mathcal{K}) f_1(t, x, v) = -\frac{1}{a} v \cdot \nabla_x f_0(t, x)\,.$$

Alors que, pour l'équation à l'ordre ϵ^{-1}, il suffisait d'étudier l'équation intégrale homogène — c'est à dire sans second membre

$$(I - \mathcal{K})\phi = 0\,,$$

avec $E \in L^1(\mathcal{V})$ t.q. $E(v) \geq 0$ p.p. sur $\mathcal{V}$, la notation "$\mu \otimes \mu$-p.p." signifie "p.p. sur $\mathcal{V}_+ \times \mathcal{V}_+$", où $\mathcal{V}_+ = \{v \in \mathcal{V} \,|\, E(v) > 0\}$". Dans le cas où $N = 3$, où $\mathbf{S}^2 \subset \mathcal{V}$ et où $d\mu(v)$ désigne l'élément de surface sur la sphère unité, on paramètre la sphère unité $\mathbf{S}^2$ par les coordonnées sphériques (θ, ϕ) comme sur la figure 1.1. Le point courant de $\mathbf{S}^2 \times \mathbf{S}^2$ est alors paramétré par $(\theta, \phi, \theta', \phi') \in [0, \pi] \times [0, 2\pi] \times [0, \pi] \times [0, 2\pi]$, et donc "$\mu \otimes \mu$-p.p." signifie "p.p. en $(\theta, \phi, \theta', \phi') \in [0, \pi] \times [0, 2\pi] \times [0, \pi] \times [0, 2\pi]$".

il nous faut maintenant étudier la même équation avec second membre.

Les hypothèses faites sur le noyau intégral k entraînent que

$$\iint_{\mathcal{V}\times\mathcal{V}} k(v,w)^2 d\mu(v)d\mu(w) < +\infty$$

c'est-à-dire que $\mathcal{K}$ est un opérateur de Hilbert-Schmidt sur $L^2(\mathcal{V}, d\mu)$. De plus, l'opérateur $\mathcal{K}$ est auto-adjoint, car $k \equiv k(v,w)$ est symétrique en v et w, d'après l'hypothèse (H2).

Or un opérateur de Hilbert-Schmidt vérifie l'énoncé suivant, connu sous le nom d'**alternative de Fredholm**.

Théorème 4.2.2 (Alternative de Fredholm) *Soit T opérateur intégral sur $L^2(\mathcal{V}, d\mu)$ de la forme*

$$T\phi(v) = \int_{\mathcal{V}} t(v,w)\phi(w)d\mu(w)\,,$$

où $t \equiv t(v,w)$ est une fonction mesurable sur $\mathcal{V}\times\mathcal{V}$ telle que

$$\iint_{\mathcal{V}\times\mathcal{V}} t(v,w)^2 d\mu(v)d\mu(w) < +\infty\,.$$

Notons T^* l'opérateur adjoint de T, qui est défini par la formule

$$T^*\phi(v) = \int_{\mathcal{V}} t(w,v)\phi(w)d\mu(w)\,.$$

Alors

$$\begin{aligned}\mathrm{Im}(I-T) &= \mathrm{Ker}(I-T^*)^\perp \\ &= \left\{\phi \in L^2(\mathcal{V}, d\mu) \,\middle|\, T^*\psi = \psi \Rightarrow \int_{\mathcal{V}} \phi(v)\psi(v)d\mu(v) = 0\right\}\,.\end{aligned}$$

Nous admettrons ce résultat ; le lecteur curieux d'en connaître une démonstration pourra consulter les Théorèmes VI.6 et VI.12 dans [11].

Dans le cas présent, comme $k \equiv k(v,w)$ est symétrique en v et w, on a $\mathcal{K}^* = \mathcal{K}$. Donc, d'après l'alternative de Fredholm,

$$\mathrm{Im}(I-\mathcal{K}) = \mathrm{Ker}(I-\mathcal{K}^*)^\perp = \mathrm{Ker}(I-\mathcal{K})^\perp\,.$$

D'autre part d'après le Lemme 4.2.1, on sait que

$$\mathrm{Ker}(I-\mathcal{K}) = \{\phi \in L^2(\mathcal{V}, d\mu) \,|\, \phi = \text{Const. p.p.}\} \simeq \mathbf{R}\,.$$

Donc

$$\begin{aligned}\mathrm{Im}(I-\mathcal{K}) &= \mathbf{R}^\perp \\ &= \left\{\phi \in L^2(\mathcal{V}, d\mu) \,\middle|\, \int_{\mathcal{V}} \phi(v)d\mu(v) = 0\right\}\end{aligned}$$

est le sous-espace de $L^2(\mathcal{V}, d\mu)$ formé des fonctions de moyenne nulle.

Donc pour résoudre l'équation intégrale d'inconnue $\phi \in L^2(\mathcal{V}, d\mu)$

$$\phi - \mathcal{K}\phi = \psi\,,$$

où $\psi \in L^2(\mathcal{V}, d\mu)$ est une fonction donnée, on applique l'alternative de Fredholm, qui nous dit alors que, de deux choses l'une

(a) ou bien

$$\int_{\mathcal{V}} \psi(v) d\mu(v) \neq 0\,,$$

auquel cas l'équation

$$\phi - \mathcal{K}\phi = \psi \text{ n'a pas de solution dans } L^2(\mathcal{V}, d\mu)\,;$$

(b) ou bien

$$\int_{\mathcal{V}} \psi(v) d\mu(v) = 0\,,$$

auquel cas l'équation

$$\phi - \mathcal{K}\phi = \psi \text{ admet au moins une solution dans } L^2(\mathcal{V}, d\mu)\,.$$

Dans ce dernier cas, étant donnée une solution $\phi \in L^2(\mathcal{V}, d\mu)$, la fonction

$$\phi_0 = \phi - \langle\phi\rangle$$

vérifie également

$$\phi_0 - \mathcal{K}\phi_0 = \psi$$

puisque $\mathcal{K}\langle\phi\rangle = \langle\phi\rangle$ — rappelons que $\mathrm{Ker}(I - \mathcal{K}) = \mathbf{R}$.

Autrement dit, dans le cas (b), il existe une unique solution ϕ_0 de l'équation

$$\phi_0 - \mathcal{K}\phi_0 = \psi \quad \text{t.q.} \quad \int_{\mathcal{V}} \phi_0(v) d\mu(v) = 0\,;$$

l'ensemble de toutes les solutions de l'équation intégrale

$$(I - \mathcal{K})\phi = \psi$$

est alors

$$\{\phi = \phi_0 + C \cdot 1 \,|\, C \in \mathbf{R}\} = \phi_0 + \mathbf{R} = \phi_0 + \mathrm{Ker}(I - \mathcal{K})\,.$$

Remarque 4.2.3 *On note parfois*

$$\phi_0 = (I - \mathcal{K})^{-1}\psi\,,$$

bien que l'opérateur $I - \mathcal{K}$ ne soit pas inversible. En effet, l'opérateur $(I - \mathcal{K})^{-1}$ n'est défini que de $\mathrm{Im}(I - \mathcal{K}) = \mathrm{Ker}(I - \mathcal{K})^{\perp}$ dans lui-même, et non sur $L^2(\mathcal{V}, d\mu)$. On appelle l'opérateur

$$(I - \mathcal{K})^{-1} : \mathrm{Im}(I - \mathcal{K}) = \mathrm{Ker}(I - \mathcal{K})^{\perp} \to \mathrm{Im}(I - \mathcal{K}) = \mathrm{Ker}(I - \mathcal{K})^{\perp}$$

le pseudo-inverse de $I - \mathcal{K}$.

Revenons maintenant à l'équation à l'ordre $O(\epsilon^0)$, que nous écrivons sous la forme

$$(I-\mathcal{K})f_1(t,x,v) = -\frac{1}{a}v\cdot\nabla_x f_0(t,x) = -\frac{1}{a}\sum_{j=1}^N v_j\frac{\partial f_0}{\partial x_j}(t,x)\,.$$

Considérons pour commencer l'équation auxiliaire suivante :

Equation auxiliaire : pour tout $j=1,\dots,N$, résoudre le problème d'inconnue $b_j(v)$ suivant :

$$\left\{\begin{array}{l} (I-\mathcal{K})b_j(v) = v_j\,,\\ \\ \displaystyle\int_{\mathcal{V}} b_j(v)d\mu(v) = 0\,. \end{array}\right.$$

Comme $\mathcal{V}$ est une partie bornée de $\mathbf{R}^N$ et que μ est une mesure bornée d'après l'hypothèse (H1), l'application $\mathcal{V}\ni v\mapsto v\in\mathbf{R}^N$ appartient bien à $L^2(\mathcal{V},d\mu)$.

Appliquons l'alternative de Fredholm : on a, d'après l'hypothèse (H1),

$$\int_{\mathcal{V}} v_j d\mu(v) = \left(\int_{\mathcal{V}} v d\mu(v)\right)_j = 0$$

pour tout $j=1,\dots,N$. On est donc dans le cas b), où il existe au moins une solution ; en particulier il existe une unique solution $b_j(v)\in\mathrm{Ker}(I-\mathcal{K})^\perp=\mathbf{R}^\perp$. Autrement dit

$$v_j\in\mathrm{Im}(I-\mathcal{K}) = \mathrm{Ker}(I-\mathcal{K})^\perp \text{ et } b_j(v) = (I-\mathcal{K})^{-1}v_j\,,$$

où $(I-\mathcal{K})^{-1}$ est le pseudo-inverse de $I-\mathcal{K}$.

Les solutions de l'équation d'inconnue f_1 sont donc toutes les fonctions de la forme

$$f_1(t,x,v) = -\frac{1}{a}\sum_{j=1}^N b_j(v)\frac{\partial f_0}{\partial x_j}(t,x) + C_1(t,x)$$

où la fonction $C_1(t,x)$, constante en $v\in\mathcal{V}$, est une solution arbitraire de l'équation homogène

$$(I-\mathcal{K})C_1 = 0\,.$$

Dans toute la suite, on notera $b(v)$ le vecteur

$$b(v) = (b_1(v),\dots,b_N(v))\,,$$

de sorte que la fonction f_1 s'écrit

$$f_1(t,x,v) = -\frac{1}{a}b(v)\cdot\nabla_x f_0(t,x) + C_1(t,x)\,.$$

L'équation à l'ordre ϵ^1 :

Passons maintenant à l'étude de l'équation à l'ordre $O(\epsilon^1)$, que l'on réécrit sous la forme :

$$\begin{aligned} a(f_2 - \mathcal{K}f_2) &= -\frac{\partial f_0}{\partial t} - v \cdot \nabla_x f_1 + a\gamma \mathcal{K} f_0 \\ &= -\left(\frac{\partial f_0}{\partial t} + v \cdot \nabla_x f_1 - a\gamma f_0 \right) \end{aligned}$$

puisque $\mathcal{K}f_0 = f_0$.

De nouveau, il s'agit d'une équation intégrale de la forme

$$(I - \mathcal{K})\phi = \Sigma$$

à laquelle on applique l'alternative de Fredholm.

Pour que f_2 existe — c'est-à-dire, pour que $v \mapsto f_2(t,x,v)$ existe pour tout $(t,x) \in \mathbf{R}_+^* \times \Omega$ — le terme f_0 doit vérifier la condition de compatibilité

$$\frac{\partial f_0}{\partial t}(t,x,\cdot) + v \cdot \nabla_x f_1(t,x,\cdot) - a\gamma f_0(t,x,\cdot) \in \mathrm{Ker}(I - \mathcal{K})^\perp \,.$$

Autrement dit, on doit avoir

$$\left\langle \frac{\partial f_0}{\partial t} + v \cdot \nabla_x f_1 - a\gamma f_0 \right\rangle = 0 \,,$$

c'est-à-dire que

$$\frac{\partial f_0}{\partial t} + \langle v \cdot \nabla_x f_1 \rangle - a\gamma f_0 = 0 \,,$$

puisque f_0 est indépendante de v. Substituons dans cette égalité la formule générale donnant f_1 : ainsi

$$\langle v \cdot \nabla_x f_1 \rangle = -\frac{1}{a} \sum_{i,j=1}^{N} \langle v_i b_j(v) \rangle \frac{\partial^2 f_0}{\partial x_i \partial x_j} \,.$$

Le terme C_1 disparaît car

$$\langle v \cdot \nabla_x C_1(t,x) \rangle = \langle v \rangle \cdot \nabla_x C_1(t,x) = 0 \,,$$

puisque

$$\int_{\mathcal{V}} v d\mu(v) = 0$$

d'après l'hypothèse (H1).

Conclusion : Le terme f_2 existe si et seulement si

$$\frac{\partial f_0}{\partial t} - \frac{1}{a} \sum_{i,j=1}^{N} \langle v_i b_j(v) \rangle \frac{\partial^2 f_0}{\partial x_i \partial x_j} = 0 \,,$$

qui est une variante a priori anisotrope de l'équation de la chaleur.

Autrement dit, l'équation de diffusion correspond à la condition de compatibilité assurant l'existence du terme f_2 dans la solution formelle de Hilbert de l'équation de Boltzmann linéaire.

L'équation à l'ordre $O(\epsilon^1)$ devient alors, en tenant compte de la condition de compatibilité correspondant à l'équation de diffusion

$$(f_2 - \mathcal{K}f_2) = \frac{1}{a^2}\sum_{i,j=1}^{N}(v_i b_j(v) - \langle v_i b_j(v)\rangle)\frac{\partial^2 f_0}{\partial x_i \partial x_j} - \frac{1}{a}v\cdot\nabla_x C_1\,.$$

Procédons comme dans l'étude de l'équation à l'ordre $O(\epsilon^0)$.

Soit donc $d_{ij}(v)$ la solution de l'équation intégrale

$$(I-\mathcal{K})d_{ij}(v) = v_i b_j(v) - \langle v_i b_j\rangle\,, \quad \langle d_{ij}\rangle = 0$$

pour tout $i,j = 1,\dots,N$.

Vérifions d'abord que le membre de droite de cette équation appartient bien à $L^2(\mathcal{V},d\mu)$. Or $b(v) = (I-\mathcal{K})^{-1}v \in L^2(\mathcal{V},d\mu)$ par définition, et d'autre part la fonction $\mathcal{V}\ni v\mapsto v\in\mathbf{R}$ est bornée puisque $\mathcal{V}$ est une partie bornée de $\mathbf{R}^N$. Par conséquent, pour tout $i,j=1,\dots,N$, la fonction $\mathcal{V}\ni v\mapsto v_i b_j(v)\in\mathbf{R}$ appartient à $L^2(\mathcal{V},d\mu)$. Comme d'après l'hypothèse (H1) la mesure μ est bornée, toute fonction constante appartient à $L^2(\mathcal{V},d\mu)$, de sorte que la fonction

$$\mathcal{V}\ni v\mapsto v_i b_j(v) - \langle v_i b_j\rangle \in \mathbf{R}$$

appartient bien à $L^2(\mathcal{V},d\mu)$.

Comme le membre de droite de cette équation intégrale est une fonction de moyenne nulle par construction, l'alternative de Fredholm garantit qu'il existe au moins une solution de cette équation dans $L^2(\mathcal{V},d\mu)$. On choisit donc

$$d_{ij} = (I-\mathcal{K})^{-1}(v_i b_j(v) - \langle v_i b_j\rangle) \in \mathbf{R}^\perp = \mathrm{Ker}(I-\mathcal{K})^\perp\,,$$

où $(I-\mathcal{K})^{-1}$ désigne le pseudo-inverse de $(I-\mathcal{K})$.

Alors

$$f_2(t,x,v) = \frac{1}{a^2}\sum_{i,j=1}^{N} d_{ij}(v)\frac{\partial^2 f_0}{\partial x_i\partial x_j}(t,x) - \frac{1}{a}\sum_{k=1}^{N} b_k(v)\frac{\partial C_1}{\partial x_k}(t,x) + C_2(t,x)$$

où $(t,x)\mapsto C_2(t,x)$ est une fonction constante en $v\in\mathcal{V}$ arbitraire, c'est-à-dire une solution arbitraire de l'équation homogène

$$(I-\mathcal{K})C_2 = 0\,.$$

4.2.3 Le coefficient de diffusion

Comme on vient de le voir, la condition garantissant l'existence du terme d'ordre ϵ^2 dans la solution formelle de Hilbert pour l'équation de Boltzmann

linéaire paramétrée par ϵ fait intervenir a priori, non pas un *coefficient* de diffusion, mais une *matrice* de diffusion dont l'élément figurant à la i-ème ligne et la j-ième colonne vaut

$$M_{ij} = \frac{1}{a}\langle v_i b_j \rangle \,, \quad i,j = 1,\dots,N \,.$$

On va voir qu'une hypothèse géométrique très simple portant sur la mesure μ et le noyau intégral k permet de garantir que cette matrice de diffusion est proportionnelle à l'identité — autrement dit qu'elle se réduit bien à un unique coefficient de diffusion, et que ce coefficient de diffusion est strictement positif.

Hypothèse : Supposons k invariant par les transformations orthogonales

(H3) $\qquad k(Qv,Qw) = k(v,w)$ pour tous $v,w \in \mathcal{V}$ et tout $Q \in O_N(\mathbf{R})$

ainsi que la mesure μ :

(H4) $$\int_{\mathcal{V}} g(Qv)d\mu(v) = \int_{\mathcal{V}} g(v)d\mu(v)$$

pour tout $g \in C(\mathcal{V})$ et tout $Q \in O_N(\mathbf{R})$. (On rappelle que le domaine $\mathcal{V}$ lui-même est supposé invariant par le groupe orthogonal.)

L'hypothèse (H3) est vérifiée lorsque le noyau intégral $k(v,w)$ est de la forme

$$k(v,w) = K(v \cdot w) \,.$$

Ce type de noyau intégral se rencontre dans de nombreux exemples physiques, comme dans le cas des processus de scattering Thomson et Rayleigh en transfert radiatif (cf. section 1.3.1).

L'hypothèse (H4) est trivialement vérifiée lorsque $d\mu(v)$ désigne l'élément de surface sur la sphère unité $\mathbf{S}^{N-1}$. Lorsque $d\mu(v)$ est une mesure de densité $E \equiv E(v)$ par rapport à la mesure de Lebesgue, l'hypothèse (H4) est vérifiée si et seulement si μ est radiale, c'est-à-dire de la forme $E(v) = M(|v|)$.

Lemme 4.2.4 *Sous les hypothèses (H1)-(H4), le champ de vecteurs*

$$b(v) = (b_1(v),\dots,b_N(v))$$

défini ci-dessus par

$$b_j(v) = (I - \mathcal{K})^{-1} v_j \,, \quad j = 1,\dots,N \,,$$

est de la forme

$$b_j(v) = \beta(|v|)v_j \quad \text{p.p. en } v \in \mathcal{V} \,,$$

où $\beta : \mathbf{R}_+ \to \mathbf{R}$ *est une fonction mesurable telle que*

$$\int_{\mathcal{V}} \beta(|v|)^2 |v|^2 d\mu(v) < +\infty \,.$$

Démonstration. Le vecteur $b(v)$ est caractérisé par le fait que

$$b(v) - \mathcal{K}b(v) = v \quad \text{et} \int_{\mathcal{V}} b(v)d\mu(v) = 0\,.$$

Evidemment, pour tout $Q \in O_N(\mathbf{R})$, on a

$$Qb(v) - \mathcal{K}(Qb(v)) = Q(b(v) - \mathcal{K}b(v)) = Qv$$

et

$$\int_{\mathcal{V}} Qb(v)d\mu(v) = Q\int_{\mathcal{V}} b(v)d\mu(v) = 0\,.$$

Mais, on a aussi

$$\int_{\mathcal{V}} Qvd\mu(v) = Q\int_{\mathcal{V}} vd\mu(v) = Q0 = 0\,,$$

de sorte que

$$(Qv)_j \in \mathrm{Im}(I-\mathcal{K}) = \mathrm{Ker}(I-\mathcal{K})^\perp = \mathbf{R}^\perp\,, \quad j = 1,\dots,N\,.$$

Il existe donc un unique élément $\phi_j \in L^2(\mathcal{V}, d\mu)$ tel que

$$(I-\mathcal{K})\phi_j(v) = (Qv)_j \quad \text{et} \int_{\mathcal{V}} \phi_j(v)d\mu(v) = 0\,, \quad j = 1,\dots,N\,.$$

D'après ce qui précède

$$\phi_j(v) = (Qb(v))_j\,, \quad j = 1,\dots,N\,.$$

Or, par invariance de k, puis de μ sous l'action de $O_N(\mathbf{R})$, on a aussi

$$\begin{aligned}\mathcal{K}(b_j\circ Q)(v) &= \int_{\mathcal{V}} k(v,w)b_j(Qw)d\mu(w)\\ &= \int_{\mathcal{V}} k(Qv,Qw)b_j(Qw)d\mu(w)\\ &= \int_{\mathcal{V}} k(Qv,w)b_j(w)d\mu(w) = (\mathcal{K}b_j)(Qv)\end{aligned}$$

de sorte que

$$b_j(Qv) - \mathcal{K}(b_j\circ Q)(v) = (b_j - \mathcal{K}b_j)(Qv) = (Qv)_j\,, \quad j = 1,\dots,N\,.$$

D'autre part, toujours par invariance de μ sous l'action de $O_N(\mathbf{R})$,

$$\int_{\mathcal{V}} b_j(Qv)d\mu(v) = \int_{\mathcal{V}} b_j(v)d\mu(v) = 0\,.$$

Autrement dit, $b_j \circ Q \in L^2(\mathcal{V}, d\mu)$ vérifie les deux égalités caractérisant la fonction $\phi_j = (I-\mathcal{K})^{-1}((Qv)_j)$. On en déduit donc que

$$b(Qv) = (\phi_1(v),\dots,\phi_N(v)) = Qb(v)\,, \quad \text{p.p. en } v \in \mathcal{V}\,.$$

Soit donc $v \in \mathcal{V} \setminus \{0\}$ pour lequel la relation ci-dessus a lieu. Soit

$$O_N(\mathbf{R})_v = \{Q \in O_N(\mathbf{R}) \,|\, Qv = v\}\,,$$

le stabilisateur de v. On sait que $O_N(\mathbf{R})_v \simeq O_{N-1}(\mathbf{R})$ est le groupe des transformations orthogonales de l'hyperplan $(\mathbf{R}v)^\perp$ orthogonal à v dans $\mathbf{R}^N$. En particulier, $O_N(\mathbf{R})_v$ agit transitivement sur les sphères de l'hyperplan $(\mathbf{R}v)^\perp$ centrées en l'origine, c'est-à-dire que

$$\text{pour tous } y, z \in (\mathbf{R}v)^\perp \text{ t.q. } |y| = |z|\,, \quad \text{il existe } Q \in O_N(\mathbf{R})_v \text{ t.q. } Qy = z\,.$$

Ecrivons la relation d'invariance vérifiée par b pour tout $Q \in O_N(\mathbf{R})_v$:

$$b(v) = b(Qv) = Qb(v)\,, \quad \text{pour tout } Q \in O_N(\mathbf{R})_v\,,$$

et notons P la projection orthogonale de $\mathbf{R}^N$ sur $\mathbf{R}v$. Comme $PQ = QP = PQP$ pour tout $Q \in O_N(\mathbf{R})_v$, on a donc

$$(I-P)b(v) = (I-P)Qb(v) = Q(I-P)b(v)\,, \quad \text{pour tout } Q \in O_N(\mathbf{R})_v\,.$$

Ceci entraîne que $(I-P)b(v) = 0$; sinon, étant donné $w \perp v$ quelconque tel que $|w| = |(I-P)b(v)|$ et $w \neq (I-P)b(v)$, il existerait $Q \in O_N(\mathbf{R})_v$ tel que

$$Q(I-P)b(v) = w \neq (I-P)b(v)\,.$$

Par conséquent,

$$b(v) = Pb(v)\,.$$

On déduit donc de ce qui précède que $b(v)$ est colinéaire à v p.p. en $v \in \mathcal{V}$, c'est-à-dire que

$$b(v) = B(v)v \quad \text{p.p. en } v \in \mathcal{V}$$

où $B : \mathcal{V} \to \mathbf{R}$ est une fonction mesurable définie p.p. sur $\mathcal{V}$. Ecrivant de nouveau que

$$b(Qv) = Qb(v) \quad \text{p.p. en } v \in \mathcal{V}\,, \text{ pour tout } Q \in O_N(\mathbf{R})\,,$$

on trouve que

$$B(Qv) = B(v) \quad \text{p.p. en } v \in \mathcal{V}\,, \text{ pour tout } Q \in O_N(\mathbf{R})\,,$$

c'est-à-dire que B est une fonction radiale.

Autrement dit, il existe $\beta : [0, R] \to \mathbf{R}$ telle que

$$B(v) = \beta(|v|) \text{ p.p. en } v \in \mathcal{V}\,.$$

Enfin, comme $b_j(v) \in L^2(\mathcal{V}, d\mu)$ pour tout $j = 1, \ldots, N$, on a

$$\int_{\mathcal{V}} \beta(|v|)^2 |v|^2 d\mu(v) = \sum_{j=1}^N \int_{\mathcal{V}} |b_j(v)|^2 d\mu(v) < +\infty\,,$$

ce qui conclut la démonstration. ■

Corollaire 4.2.5 *Sous les hypothèses (H1)-(H4), la matrice de diffusion est alors proportionnelle à l'identité :*

$$\langle v_i b_j(v)\rangle = \tfrac{1}{N}\langle |w|^2\beta(|w|)\rangle\delta_{ij}\,, \quad i,j=1,\dots,N\,,$$

et de plus

$$\langle |w|^2\beta(|w|)\rangle > 0\,.$$

Démonstration. D'après le lemme ci-dessus

$$b_j(v) = \beta(|v|)v_j \quad \text{p.p. en } v\in\mathcal{V}\,, \quad j=1,\dots,N\,.$$

Donc

$$\langle v_i b_j\rangle = \langle \beta(|v|)v_iv_j\rangle\,, \quad i,j=1,\dots,N\,.$$

D'une part

$$i\neq j \Rightarrow \langle\beta(|v|)v_iv_j\rangle = 0\,.$$

En effet, soit $Q_i\in O_N(\mathbf{R})$ définie par

$$Q_iv = (v_1,\dots,v_{i-1},-v_i,v_{i+1}\dots v_N)\,.$$

Alors, par invariance de μ sous l'action de $O_N(\mathbf{R})$, on a

$$\int_{\mathcal{V}}\beta(|v|)v_iv_jd\mu(v) = \int_{\mathcal{V}}\beta(|Q_iv|)(Q_iv)_i(Q_iv)_jd\mu(v) = \int_{\mathcal{V}}\beta(|v|)(-v_i)v_jd\mu(v)$$

d'où

$$\int_{\mathcal{V}}\beta(|v|)v_iv_jd\mu(v) = 0\,.$$

D'autre part,

$$i\neq j \Rightarrow \langle\beta(|v|)v_i^2\rangle = \langle\beta(|v|)v_j^2\rangle\,.$$

En effet, soit $Q_{ij}\in O_N(\mathbf{R})$ définie par

$$Q_{ij}(v_1,\dots,v_i,\dots,v_j,\dots,v_N) = (v_1,\dots,v_j,\dots,v_i,\dots,v_N)\,.$$

Alors, toujours par invariance de μ sous l'action de $O_N(\mathbf{R})$, on a

$$\int_{\mathcal{V}}\beta(|v|)v_i^2d\mu(v) = \int_{\mathcal{V}}\beta(|Q_{ij}v|)(Q_{ij}v)_i^2d\mu(v) = \int_{\mathcal{V}}\beta(|v|)v_j^2d\mu(v)\,.$$

On déduit de ce qui précède que

$$\langle\beta(|v|)v_iv_j\rangle = c\delta_{ij}\,, \quad i,j=1,\dots,N\,.$$

Pour calculer la constante c, on observe que

$$Nc = \sum_{i=1}^N\langle\beta(|v|)v_i^2\rangle = \left\langle\beta(|v|)\sum_{i=1}^N v_i^2\right\rangle = \langle\beta(|v|)|v|^2\rangle\,,$$

de sorte que

$$c = \tfrac{1}{N}\langle \beta(|v|)|v|^2 \rangle \, .$$

Ainsi

$$\langle v_i b_j(v) \rangle = \tfrac{1}{N}\langle \beta(|v|)|v|^2 \rangle \delta_{ij} \, , \quad i,j = 1, \ldots, N \, .$$

Il reste à vérifier que $\langle \beta(|v|)|v|^2 \rangle > 0$. Pour cela, on écrit que

$$v_1 = (I - \mathcal{K})(\beta(|v|)v_1) \, ,$$

de sorte que

$$\int_{\mathcal{V}} \beta(|v|)v_1^2 d\mu(v) = \int_{\mathcal{V}} \beta(|v|)v_1(I - \mathcal{K})(\beta(|v|)v_1) d\mu(v) \geq 0$$

d'après le Lemme 4.2.1. De plus, l'égalité est exclue, car sinon on aurait

$$\beta(|v|)v_1 = \text{Const.} \quad \text{p.p. en } v \in \mathcal{V} \, ,$$

c'est-à-dire que

$$\beta(|v|)v_1 = \langle \beta(|v|)v_1 \rangle \quad \text{p.p. en } v \in \mathcal{V} \, .$$

Or

$$\int_{\mathcal{V}} \beta(|v|)v_1 d\mu(v) = 0 \, ,$$

puisque

$$\beta(|v|)v_1 = (I - \mathcal{K})^{-1}(v_1) \in \mathrm{Ker}(I - \mathcal{K})^{\perp} = \mathbf{R}^{\perp} \, .$$

Donc l'égalité forcerait $\beta(|v|)v_1 = 0$ μ-p.p. en $v \in \mathcal{V}$, qui entraînerait à son tour que

$$v_1 = (I - \mathcal{K})(\beta(|v|)v_1) = 0 \text{ p.p. en } v \in \mathcal{V} \, ,$$

ce qui est évidemment une contradiction.

Par conséquent

$$0 < \langle \beta(|v|)v_1^2 \rangle = \tfrac{1}{N}\langle \beta(|v|)|v|^2 \rangle \, ,$$

ce qui conclut la démonstration. ■

D'après le corollaire ci-dessus, on voit que, sous les hypothèses additionnelles (H3)-(H4), la condition de compatibilité assurant l'existence de f_2 obtenue dans la section précédente se met sous la forme de l'équation de diffusion

$$\frac{\partial f_0}{\partial t} - \tfrac{1}{2}\kappa^2 \Delta_x f_0 - a\gamma f_0 = 0 \, ,$$

avec un coefficient de diffusion donné par la formule

$$\tfrac{1}{2}\kappa^2 = \frac{\langle |w|^2 \beta(|w|) \rangle}{Na} > 0 \, .$$

On sait que l'équation de la chaleur engendre un semi-groupe régularisant

$$e^{\frac{1}{2}t\Delta_x} : L^2(\mathbf{R}^N) \mapsto C^{\infty}(\mathbf{R}^N) \, , \qquad t > 0 \, ,$$

— voir par exemple [23], Théorème 8.3.2 dans le cas de l'espace entier, et [2], §8.4.4 dans le cas du problème aux limites.

Cette circonstance montre que le semi-groupe en question ne peut en aucun cas s'étendre pour $t < 0$. Donc le problème de Cauchy dans le futur pour l'équation de diffusion à coefficient de diffusion c n'est bien posé que si $c \geq 0$. En effet, résoudre le problème

$$\begin{cases} \frac{\partial u}{\partial t} - c\Delta_x u = 0\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,, \\ u\big|_{t=0} = u^{in} \end{cases}$$

revient à définir

$$u(t,\cdot) = e^{ct\Delta_x} u^{in}$$

de sorte que résoudre ce problème pour $c < 0$ reviendrait à savoir étendre le semi-groupe $e^{\frac{1}{2}t\Delta_x}$ pour tout $t \in \mathbf{R}$.

C'est pourquoi, dans le contexte de l'approximation de l'équation de Boltzmann linéaire par la diffusion, il est absolument capital de vérifier que le coefficient de diffusion obtenu à la limite est bien positif. (On retrouvera d'ailleurs cette même question dans l'étude de certains problèmes d'homogénéisation.)

4.3 Démonstration de l'approximation par la diffusion

4.3.1 Conditions aux limites indépendantes de $v \in \mathcal{V}$

Les calculs effectués dans la section précédente ont fait apparaître l'équation de diffusion de façon naturelle comme condition de compatibilité assurant l'existence du terme d'ordre 2 dans la solution formelle de Hilbert.

Pourtant ces calculs ne sauraient à eux seuls justifier l'utilisation de l'équation de diffusion comme modèle approché pour décrire la dynamique de particules sous les hypothèses d'échelle énoncées dans la section précédente, car la solution formelle de Hilbert n'a a priori aucun sens physique — nous reviendrons plus loin sur ce point.

Cependant, les calculs menés sur cette solution formelle de Hilbert ne l'ont pas été en vain, car ils vont nous permettre de justifier très simplement que la solution de l'équation de diffusion approche effectivement la solution de l'équation de Boltzmann linéaire paramétrée par ϵ, par un argument de stabilité élémentaire basé sur le principe du maximum pour l'équation de Boltzmann linéaire.

On rappelle que Ω est un ouvert convexe borné à bord de classe C^∞ de $\mathbf{R}^N$, dont on note $\partial\Omega$ le bord. Pour tout $x \in \partial\Omega$, on note n_x le vecteur unitaire normal à $\partial\Omega$ au point $x \in \partial\Omega$, dirigé vers l'extérieur de Ω. Enfin on pose

$$\Gamma_- = \{(x,v) \in \partial\Omega \times \mathcal{V} \,|\, v \cdot n_x < 0\}\,.$$

On considère maintenant, pour tout $\epsilon > 0$, le problème aux limites pour l'équation de Boltzmann linéaire

$$\begin{cases} \epsilon \frac{\partial f_\epsilon}{\partial t} + v \cdot \nabla_x f_\epsilon + \frac{a}{\epsilon}\Big(f_\epsilon - (1+\epsilon^2\gamma)\mathcal{K} f_\epsilon \Big) = 0\,, \quad (x,v) \in \Omega \times \mathcal{V}\,,\ t > 0\,, \\ f_\epsilon\big|_{\Gamma_-} = f_b\,, \\ f_\epsilon\big|_{t=0} = f^{in}\,, \end{cases}$$

où la donnée initiale et la donnée au bord vérifient

$$f^{in} \equiv f^{in}(x) \in C^\infty(\overline{\Omega})\,, \quad f_b \equiv f_b(t,x) \in C^\infty([0,T] \times \partial\Omega)\,.$$

Rappelons enfin les hypothèses (H1)-(H4) faites sur la mesure μ et le noyau intégral $k \equiv k(v,w)$:

(H1) on suppose que μ est une mesure de Radon positive sur $\mathcal{V}$ t.q.

$$\int_{\mathcal{V}} d\mu(v) < +\infty \qquad \text{et} \qquad \int_{\mathcal{V}} v d\mu(v) = 0\,;$$

(H2) on suppose que $k \in C(\mathcal{V} \times \mathcal{V})$ vérifie

$$\begin{cases} 0 < k(v,w) = k(w,v) \text{ pour tout } v, w \in \mathcal{V}\,, \\ \mathcal{K}1(v) = \displaystyle\int_{\mathbf{R}^N} k(v,w) d\mu(w) = 1 \text{ pour tout } v \in \mathcal{V}\,; \end{cases}$$

(H3) on suppose en outre que le noyau intégral k est invariant par les transformations orthogonales :

$$k(Qv,Qw) = k(v,w)\,, \qquad \text{pour tous } v, w \in \mathcal{V} \text{ et tout } Q \in O_N(\mathbf{R})\,;$$

(H4) on suppose enfin que la mesure μ est également invariante par les transformations orthogonales :

$$\int_{\mathcal{V}} g(Qv) d\mu(v) = \int_{\mathcal{V}} g(v) d\mu(v)\,,$$

pour tout $g \in C(\mathcal{V})$ et tout $Q \in O_N(\mathbf{R})$. (Rappelons que le domaine $\mathcal{V}$ lui-même invariant par le groupe orthogonal $O_N(\mathbf{R})$.)

Théorème 4.3.1 *Sous les hypothèses (H1)-(H4), on suppose en outre que a et γ sont deux réels strictement positifs.*

Soient $f^{in} \equiv f^{in}(x) \in C^\infty(\overline{\Omega})$ et $f_b \equiv f_b(t,x) \in C^\infty([0,T] \times \partial\Omega)$ vérifiant les relations de compatibilité

$$\frac{\partial^k f_b}{\partial t^k}(0,y) = \big(\tfrac{1}{2}\kappa^2 \Delta_x + a\gamma \big)^k f^{in}(y)\,, \quad y \in \partial\Omega\,, \quad k \in \mathbf{N}\,,$$

avec

$$\tfrac{1}{2}\kappa^2 = \frac{1}{N}\langle \beta(|v|)|v|^2\rangle\,,$$

où

$$\beta(|v|)v_j = (I-\mathcal{K})^{-1}(v_j)\,,\quad j=1,\dots,N\,.$$

Soit $u \equiv u(t,x)$ la solution du problème aux limites

$$\begin{cases} \frac{\partial u}{\partial t} - \frac{1}{2}\kappa^2\Delta_x u - a\gamma u = 0\,, \quad x\in\Omega\,,\ t>0\,,\\ u\big|_{\Gamma_-} = f_b\,,\\ u\big|_{t=0} = f^{in}\,. \end{cases}$$

Alors il existe $C_T > 0$ t.q. la famille $(f_\epsilon)_{\epsilon>0}$ de solutions généralisées du problème aux limites pour l'équation de Boltzmann linéaire

$$\begin{cases} \epsilon\frac{\partial f_\epsilon}{\partial t} + v\cdot\nabla_x f_\epsilon + \frac{a}{\epsilon}\Big(f_\epsilon - (1+\epsilon^2\gamma)\mathcal{K}f_\epsilon\Big) = 0\,, \quad (x,v)\in\Omega\times\mathcal{V}\,,\ t>0\,,\\ f_\epsilon\big|_{\Gamma_-} = f_b\,,\\ f_\epsilon\big|_{t=0} = f^{in}\,, \end{cases}$$

vérifie l'estimation

$$\sup_{(t,x,v)\in[0,T]\times\overline{\Omega}\times\mathcal{V}} |f_\epsilon(t,x,v) - u(t,x)| \le C_T\epsilon\,.$$

Démonstration.

La preuve se décompose en plusieurs étapes.

Etape 1. A partir de la solution u du problème aux limites pour l'équation de diffusion, formons la solution formelle de Hilbert tronquée à l'ordre ϵ^2.

Autrement dit, on définit

$$\begin{aligned} f_0(t,x,v) &:= u(t,x)\\ f_1(t,x,v) &:= -\frac{1}{a}\beta(|v|)v\cdot\nabla_x u(t,x)\\ f_2(t,x,v) &:= \frac{1}{a^2}\sum_{i,j=1}^N d_{ij}(v)\frac{\partial^2 u}{\Delta x_i \Delta x_j}(t,x) \end{aligned}$$

où

$$d_{ij}(v) = (I-\mathcal{K})^{-1}(v_i b_j - \langle v_i b_j\rangle)\,,\qquad i,j=1,\dots,N\,.$$

On pose alors

$$F_\epsilon(t,x,v) = f_0(t,x,v) + \epsilon f_1(t,x,v) + \epsilon^2 f_2(t,x,v)\,.$$

Le cœur de la démonstration consiste à estimer la différence entre la vraie solution f_ϵ et la solution formelle de Hilbert tronquée à l'ordre 2 en norme uniforme.

Remarque : Le lecteur attentif pourra s'interroger sur l'absence des fonctions $C_1(t,x)$ et $C_2(t,x)$ qui intervenaient dans la construction de la solution série formelle de Hilbert. On se souvient que ces fonctions constantes en v n'intervenaient que comme éléments arbitraires de $\mathrm{Ker}(I-\mathcal{K})$ apparaissant dans la solution générale de l'équation intégrale $(I-\mathcal{K})\phi=\psi$. Lorsqu'on interrompt la construction de la série formelle de Hilbert à l'ordre 2 en ϵ, ces fonctions C_1 et C_2 demeurent arbitraires. Comme leur contribution à la série formelle de Hilbert est en $O(\epsilon)$ et que l'on s'intéresse ici à une approximation à l'ordre dominant de la solution de l'équation de Boltzmann linéaire, on peut choisir $C_1=C_2=0$ sans le moindre inconvénient, et c'est ce que nous avons fait.

Etape 2. La fonction F_ϵ est de classe C^∞ par rapport aux variables t et x, et un calcul élémentaire basé sur la construction de la solution formelle de Hilbert à la section précédente montre que

$$\frac{\partial F_\epsilon}{\partial t}+\frac{1}{\epsilon}v\cdot\nabla_x F_\epsilon+\frac{a}{\epsilon^2}\Big(F_\epsilon-(1+\epsilon^2\gamma)\mathcal{K}F_\epsilon\Big)=-S_\epsilon$$

où

$$S_\epsilon=-\epsilon\frac{\partial f_1}{\partial t}-\epsilon^2\frac{\partial f_2}{\partial t}-\epsilon v\cdot\nabla_x f_2+a\epsilon\gamma\mathcal{K}f_1+a\epsilon^2\gamma\mathcal{K}f_2\,.$$

Par linéarité de l'équation de Boltzmann, la fonction

$$R_\epsilon(t,x,v)=f_\epsilon(t,x,v)-F_\epsilon(t,x,v)\,,$$

est solution généralisée du problème aux limites suivant :

$$\begin{cases}\frac{\partial R_\epsilon}{\partial t}+\dfrac{1}{\epsilon}v\cdot\nabla_x R_\epsilon+\dfrac{a}{\epsilon^2}\Big(R_\epsilon-(1+\epsilon^2\gamma)\mathcal{K}R_\epsilon\Big)=S_\epsilon\,,\\[2ex] R_\epsilon\big|_{\Gamma_-}=R_\epsilon^b\,,\\[2ex] R_\epsilon\big|_{t=0}=R_\epsilon^{in}\,.\end{cases}$$

Les données initiale et au bord pour ce problème aux limites sont évidemment les fonctions

$$\begin{cases}R_\epsilon^{in}=-\epsilon f_1\big|_{t=0}-\epsilon^2 f_2\big|_{t=0}\,,\\[2ex] R_\epsilon^b\ =-\epsilon f_1\big|_{\Gamma_-}-\epsilon^2 f_2\big|_{\Gamma_-}\,.\end{cases}$$

Etape 3. Par définition,

$$\beta(|v|)v_j=(I-\mathcal{K})^{-1}(v_j)\in L^2(\mathcal{V},d\mu)\,.$$

Comme $k\in C(\mathcal{V}\times\mathcal{V})$ et que $\mathcal{V}$ est compact car fermé borné dans $\mathbf{R}^N$,

$$M:=\sup_{v,w\in\mathcal{V}}k(v,w)<+\infty\,.$$

Alors

$$\begin{aligned}|\mathcal{K}(\beta(|v|)v_j)| &= \left|\int_{\mathcal{V}} k(v,w)\beta(|w|)w_j d\mu(w)\right| \\ &\leq \left(\int_{\mathcal{V}} k(v,w)^2 d\mu(w)\right)^{1/2}\left(\int_{\mathcal{V}} \beta(|w|)^2 w_j^2 d\mu(w)\right)^{1/2} \\ &\leq \left(\int_{\mathcal{V}} d\mu(v)\right)^{1/2} M\|\beta(|v|)v_j\|_{L^2(\mathcal{V},d\mu)}\,,\end{aligned}$$

de sorte qu'en utilisant l'équation

$$\beta(|v|)v_j - \mathcal{K}\beta(|v|)v_j = v_j \quad j = 1,\dots,N,$$

μ-p.p. en $v \in \mathcal{V}$, on trouve que

$$\|\beta(|v|)v_j\|_{L^\infty(\mathcal{V},d\mu)} \leq \left(\int_{\mathcal{V}} d\mu(v)\right)^{1/2} M\|\beta(|v|)v_j\|_{L^2(\mathcal{V},d\mu)} + R$$

pour tout $j = 1,\dots,N$. On établit de même que

$$\|d_{ij}\|_{L^\infty(\mathcal{V},d\mu)} < +\infty \quad \text{pour tout } i,j = 1,\dots,N\,.$$

Etape 4. D'après le Théorème 4.1.1 rappelé au début de ce chapitre, la fonction $u \in C^\infty([0,T]\times\overline{\Omega})$. Comme Ω est borné dans $\mathbf{R}^N$, $[0,T]\times\overline{\Omega}$ est compact dans $\mathbf{R}^{N+1}$ de sorte que, pour tout $m \in \mathbf{N}$ et tout multi-indice $\alpha \in \mathbf{N}^N$,

$$\sup_{(t,x)\in[0,T]\times\overline{\Omega}}\left|\left(\frac{\partial}{\partial t}\right)^m \nabla_x^\alpha u(t,x)\right| \leq C_{m,\alpha,T}\,.$$

Les formules définissant f_1, f_2, R_ϵ^{in} et R_b montrent que, pour tout ϵ vérifiant $0 < \epsilon < 1$

$$\begin{aligned}\|R_\epsilon^{in}\|_{L^\infty(\Omega\times\mathcal{V})} &\leq \frac{\epsilon}{a}\sum_{|\alpha|=1} C_{0,\alpha,T}\max_{1\leq j\leq N}\|\beta(|v|)v_j\|_{L^\infty(\mathcal{V},d\mu)} \\ &\quad + \frac{\epsilon^2}{a^2}\sum_{|\alpha|=2} C_{0,\alpha,T}\max_{1\leq i,j\leq N}\|d_{ij}\|_{L^\infty(\mathcal{V},d\mu)} \leq C_T^{b,in}\epsilon\,,\end{aligned}$$

et de même

$$\begin{aligned}\|R_\epsilon^{b}\|_{L^\infty([0,T]\times\Gamma_-)} &\leq \frac{\epsilon}{a}\sum_{|\alpha|=1} C_{0,\alpha,T}\max_{1\leq j\leq N}\|\beta(|v|)v_j\|_{L^\infty(\mathcal{V},d\mu)} \\ &\quad + \frac{\epsilon^2}{a^2}\sum_{|\alpha|=2} C_{0,\alpha,T}\max_{1\leq i,j\leq N}\|d_{ij}\|_{L^\infty(\mathcal{V},d\mu)} \leq C_T^{b,in}\epsilon\,,\end{aligned}$$

en posant

$$\begin{aligned}C_T^{b,in} :=& \frac{1}{a}\sum_{|\alpha|=1} C_{0,\alpha,T}\max_{1\leq j\leq N}\|\beta(|v|)v_j\|_{L^\infty(\mathcal{V},d\mu)} \\ &+ \frac{1}{a^2}\sum_{|\alpha|=2} C_{0,\alpha,T}\max_{1\leq i,j\leq N}\|d_{ij}\|_{L^\infty(\mathcal{V},d\mu)}\,.\end{aligned}$$

D'autre part,

$$\begin{aligned}\|S_\epsilon\|_{L^\infty([0,T]\times\Omega\times\mathcal{V})} \le{}& \frac{\epsilon}{a}\sum_{|\alpha|=1} C_{1,\alpha,T}\max_{1\le j\le N}\|\beta(|v|)v_j\|_{L^\infty(\mathcal{V},d\mu)}\\ &+\frac{\epsilon^2}{a^2}\sum_{|\alpha|=2} C_{1,\alpha,T}\max_{1\le i,j\le N}\|d_{ij}\|_{L^\infty(\mathcal{V},d\mu)}\\ &+\frac{\epsilon}{a^2}\sum_{|\alpha|=3} C_{0,\alpha,T}R\max_{1\le i,j\le N}\|d_{ij}\|_{L^\infty(\mathcal{V},d\mu)}\\ &+\epsilon\gamma\|\mathcal{K}1\|_{L^\infty(\mathcal{V},d\mu)}\sum_{|\alpha|=1} C_{0,\alpha,T}\max_{1\le j\le N}\|\beta(|v|)v_j\|_{L^\infty(\mathcal{V},d\mu)}\\ &+\frac{\epsilon^2}{a}\gamma\|\mathcal{K}1\|_{L^\infty(\mathcal{V},d\mu)}\sum_{|\alpha|=2} C_{0,\alpha,T}\max_{1\le i,j\le N}\|d_{ij}\|_{L^\infty(\mathcal{V},d\mu)}\\ &\le C_T^s\epsilon\,.\end{aligned}$$

Etape 5. Appliquons le principe du maximum (Proposition 3.1.5 du chapitre 3) au problème aux limites vérifié par la fonction R_ϵ dans l'étape 2.

Dans le cas présent, on a

$$\frac{a}{\epsilon^2}\left((1+\epsilon^2\gamma)\mathcal{K}1-1\right)=\frac{a}{\epsilon^2}\left((1+\epsilon^2\gamma)1-1\right)=a\gamma\,.$$

Avec les notations de la Proposition 3.1.5 du chapitre 3, on trouve donc que

$$D=\left\|\frac{a}{\epsilon^2}\left((1+\epsilon^2\gamma)\mathcal{K}1-1\right)_+\right\|_{L^\infty}=a\gamma_+\,.$$

Donc

$$\|R_\epsilon\|_{L^\infty([0,T]\times\Omega\times\mathcal{V})}\le\epsilon C_T^{b,in}e^{a\gamma_+T}+\epsilon TC_T^se^{a\gamma_+T}\,.$$

On en déduit immédiatement que

$$\begin{aligned}\|f_\epsilon-u\|_{L^\infty([0,T]\times\Omega\times\mathcal{V})}&\le\|R_\epsilon\|_{L^\infty([0,T]\times\Omega\times\mathcal{V})}\\ &\quad+\epsilon\|f_1\|_{L^\infty([0,T]\times\Omega\times\mathcal{V})}\\ &\quad+\epsilon^2\|f_2\|_{L^\infty([0,T]\times\Omega\times\mathcal{V})}\\ &\le\epsilon C_T^{b,in}(1+e^{a\gamma_+T})+\epsilon C_T^se^{a\gamma_+T}\end{aligned}$$

puisque les définitions de $C_T^{b,in}$, de $C_{m,\alpha,T}$, et de f_1 et f_2 montrent que

$$\epsilon\|f_1\|_{L^\infty([0,T]\times\Omega\times\mathcal{V})}+\epsilon^2\|f_2\|_{L^\infty([0,T]\times\Omega\times\mathcal{V})}\le C_T^{b,in}\epsilon\,.$$

Ceci démontre l'inégalité du Théorème 4.3.1 avec

$$C_T=C_T^{b,in}(1+e^{a\gamma_+T})+C_T^se^{a\gamma_+T}\,.$$

■

4.3.2 Conditions aux limites dépendant de $v \in \mathcal{V}$

La démonstration ci-dessus peut s'étendre dans différentes directions. Nous allons en discuter tout particulièrement une, qui porte sur la possibilité de généraliser l'énoncé du Théorème 4.3.1 au cas de données au bord plus générales.

En effet, dans le Théorème 4.3.1 ci-dessus, on a supposé que les données initiale et au bord sont indépendantes de la variable v :

$$f^{in} \equiv f^{in}(x)\,, \quad \text{et} \quad f_b \equiv f_b(t,x)\,.$$

En général, il faudrait pouvoir traiter le cas de données au bord de la forme

$$f^{in} \equiv f^{in}(x,v)\,, \quad \text{et} \quad f_b \equiv f_b(t,x,v)\,.$$

Dans ce cas, la solution formelle de Hilbert tronquée à n'importe quel ordre ne peut pas en général être utilisée comme dans la preuve du Théorème 4.3.1. En effet, en gardant les notations de la démonstration du Théorème 4.3.1, on a

$$f_\epsilon(0,x,v) - F_\epsilon(0,x,v) = f^{in}(x,v) - f_0(0,x) + O(\epsilon) = O(1)$$

et de même

$$f_\epsilon(t,x,v) - F_\epsilon(t,x,v) = f(t,x,v) - f_b(t,x) + O(\epsilon) = O(1)\,, \quad (x,v) \in \Gamma_-\,.$$

On ne peut donc déduire du principe du maximum que

$$f_\epsilon - F_\epsilon = O(\epsilon)$$

en norme uniforme sur $[0,T] \times \overline{\Omega} \times \mathcal{V}$, puisque cette estimation fait défaut sur le bord.

Expliquons comment y remédier dans le cas très simple de l'équation de Boltzmann linéaire monocinétique avec scattering isotrope, en dimension 1 d'espace et avec la symétrie du de la plaque infinie. Pour l'écriture de ce modèle particulier, nous renvoyons le lecteur au chapitre 1, en particulier à la section 1.1.2.

On considère donc le problème stationnaire d'inconnue $f_\epsilon \equiv f_\epsilon(x,\mu)$

$$\begin{cases} \epsilon f_\epsilon + \mu \frac{\partial f_\epsilon}{\partial x} + \dfrac{a}{\epsilon}(f_\epsilon - \langle f_\epsilon \rangle) = \epsilon q(x)\,, \quad x \in]x_L, x_R[\,,\ |\mu| \le 1\,, \\ \\ f_\epsilon(x_L,\mu) = f_L(\mu)\,, \qquad 0 < \mu < 1\,, \\ \\ f_\epsilon(x_R,-\mu) = f_R(\mu)\,, \quad 0 < \mu < 1\,, \end{cases}$$

où $a > 0$ et

$$\langle f_\epsilon \rangle(x) = \tfrac{1}{2} \int_{-1}^{1} f_\epsilon(x,\mu) d\mu\,, \qquad x_L < x < x_R\,.$$

La fonction donnée $q \equiv q(x)$ appartient à $C^\infty([x_L, x_R])$.

On montre que, lorsque $\epsilon \to 0^+$, la famille f_ϵ de solutions de l'équation de Boltzmann linéaire ci-dessus vérifie $f_\epsilon \to f_0 \equiv f_0(x)$, où f_0 est solution du problème aux limites de diffusion

$$\begin{cases} f_0 - \frac{1}{3a}\frac{d^2 f_0}{dx^2} = q\,, \quad x_L < x < x_R\,, \\ f_0(x_L) = F_L\,, \\ f_0(x_R) = F_R\,, \end{cases}$$

où

$$F_L = \frac{\sqrt{3}}{2}\int_0^1 \mu H(\mu) f_L(\mu) d\mu\,,$$
$$F_R = \frac{\sqrt{3}}{2}\int_0^1 \mu H(\mu) f_R(\mu) d\mu\,.$$

La fonction H intervenant dans les expressions de f_L et f_R ci-dessus est la *fonction de Chandrasekhar* — qui n'est pas connue explicitement, mais se calcule très facilement sur le plan numérique : voir [15]. On trouve qu'une valeur approchée de la fonction H est

$$H(\mu) \simeq \frac{2+3\mu}{\sqrt{3}}\,, \quad 0 < \mu < 1\,.$$

Pour arriver au résultat ci-dessus, on modifie la solution formelle de Hilbert en y rajoutant des termes de couche limite, localisés près des extrémités de l'intervalle $[x_L, x_R]$. Plus précisément, on cherche une solution formelle du type

$$F_\epsilon(x,\mu) = \sum_{n\geq 0}\epsilon^n f_n(x,\mu) + \sum_{n\geq 0}\epsilon^n \phi_n^L\left(\frac{x-x_L}{\epsilon},\mu\right) + \sum_{n\geq 0}\epsilon^n \phi_n^R\left(\frac{x_R-x}{\epsilon},-\mu\right)\,.$$

Le terme

$$\sum_{n\geq 0}\epsilon^n f_n(x,\mu)$$

est habituellement appelé "solution formelle intérieure", tandis que les termes

$$\sum_{n\geq 0}\epsilon^n \phi_n^L\left(\frac{x-x_L}{\epsilon},\mu\right) \quad \text{et} \quad \sum_{n\geq 0}\epsilon^n \phi_n^R\left(\frac{x_R-x}{\epsilon},-\mu\right)$$

sont les termes de couche limite localisés près de x_L et de x_R respectivement.

A l'ordre dominant, ces termes de couches limites satisfont les équations

$$\begin{cases} \mu\frac{\partial \phi_0^L}{\partial z}(z,\mu) + a(\phi_0^L(z,\mu) - \langle\phi_0^L\rangle(z)) = 0\,, & z > 0\,,\ |\mu| \leq 1\,, \\ \phi_0^L(0,\mu) = f_L(\mu) - f_0(x_L)\,, & 0 < \mu \leq 1\,, \end{cases}$$

pour ce qui est de ϕ_0^L, la fonction ϕ_0^R étant solution d'un problème analogue.

Le problème aux limites ci-dessus est posé sur la demi-droite $z > 0$; c'est un exemple de *problème de demi-espace* pour l'équation de Boltzmann linéaire, parfois appelée *problème de Milne*.

Le fait crucial concernant ce problème de Milne est que, sous l'hypothèse

$$f_0(x_L) = \tfrac{\sqrt{3}}{2} \int_0^1 \mu H(\mu) f_L(\mu) d\mu \,,$$

la fonction ϕ_0^L converge vers 0 à l'infini en z, à vitesse exponentielle :

$$|\phi_0^L(z,\mu)| = O(e^{-cz}) \text{ pour tout } z > 0 \text{ et tout } \mu \in]-1,1[$$

quelque soit $c < a$: voir [5].

Ainsi, en choisissant comme conditions aux limites pour l'équation de diffusion stationnaire

$$\begin{cases} f_0(x_L) = F_L = \tfrac{\sqrt{3}}{2} \displaystyle\int_0^1 \mu H(\mu) f_L(\mu) d\mu \,, \\ \\ f_0(x_R) = F_R = \tfrac{\sqrt{3}}{2} \displaystyle\int_0^1 \mu H(\mu) f_L(\mu) d\mu \,, \end{cases}$$

on trouve que les termes de couche limite dans la solution formelle tronquée à l'ordre 2 vérifient

$$\sum_{n=0}^{2} \epsilon^n \phi_n^L \left(\frac{x - x_L}{\epsilon}, \mu \right) = O(e^{-c(x-x_L)/\epsilon}) \,,$$
$$\sum_{n=0}^{2} \epsilon^n \phi_n^R \left(\frac{x_R - x}{\epsilon}, -\mu \right) = O(e^{-c(x_R-x)/\epsilon}) \,,$$

pour tout $x \in]x_L, x_R[$, c'est-à-dire qu'ils restent confinés au bord de l'intervalle $]x_L, x_R[$. Autrement dit, le rôle du terme de couche limite

$$\sum_{n=0}^{2} \epsilon^n \phi_n^L \left(\frac{x - x_L}{\epsilon}, \mu \right)$$

consiste à raccorder la condition aux limites pour l'équation de Boltzmann linéaire

$$f_\epsilon(x_L, \mu) = f_L(\mu) \,, \quad 0 < \mu < 1$$

— condition aux limites qui est incompatible avec l'approximation

$$f_\epsilon(x, \mu) \simeq f_0(x)$$

— avec la condition au bord $x = x_L$ pour l'équation de diffusion, qui est

$$f_0(x_L) = F_L = \tfrac{\sqrt{3}}{2} \int_0^1 \mu H(\mu) f_L(\mu) d\mu \,.$$

Nous n'en dirons pas plus sur ce sujet, et renvoyons le lecteur intéressé aux articles [5], [24] et à l'Exercice 4.4, où ces problèmes de demi-espace sont étudiés en détail.

On peut également chercher à obtenir une approximation plus précise que celle du Théorème 4.3.1, par exemple une approximation à $O(\epsilon^2)$ près de la solution f_ϵ.

Dans ce cas, même si les données au bord sont indépendantes de μ, c'est à dire si on considère le problème

$$\begin{cases} \epsilon f_\epsilon + \mu \frac{\partial f_\epsilon}{\partial x} + \frac{a}{\epsilon}(f_\epsilon - \langle f_\epsilon \rangle) = \epsilon q(x)\,, \quad x \in]x_L, x_R[\,,\ |\mu| \leq 1\,, \\ f_\epsilon(x_L, \mu) = f_L\,, \qquad 0 < \mu < 1\,, \\ f_\epsilon(x_R, \mu) = f_R\,, \quad 0 < \mu < 1\,, \end{cases}$$

où f_L et f_R sont des constantes, l'approximation par la diffusion exige d'avoir recours à des termes de couches limites.

En effet, pour avoir une approximation à l'ordre $O(\epsilon^2)$, on doit ajuster la solution formelle intérieure de Hilbert à la condition aux limites de telle sorte que

$$F_\epsilon(x_L, \mu) - f_L = O(\epsilon^2)\,, \quad F_\epsilon(x_R, \mu) - f_R = O(\epsilon^2)\,.$$

Si F_ϵ est une solution intérieure série formelle en puissance de ϵ de l'équation de Boltzmann linéaire stationnaire monocinétique avec scattering isotrope et symétrie du slab, c'est-à-dire si

$$F_\epsilon(x, \mu) = \sum_{n \geq 0} \epsilon^n f_n(x, \mu)\,,$$

on trouve comme dans la section précédente que

$$\begin{aligned} f_0 &\equiv f_0(x)\,, \\ f_1 &\equiv C_1(x) - \frac{1}{a}\mu \frac{df_0}{dx}(x)\,, \end{aligned}$$

de sorte que

$$F_\epsilon(x_L, \mu) - f_L = f_0(x_L) + \epsilon C_1(x_L) - f_L - \frac{\epsilon}{a}\mu \frac{df_0}{dx}(x_L) + O(\epsilon^2)\,.$$

Il est donc impossible de satisfaire à la condition

$$F_\epsilon(x_L, \mu) - f_L = O(\epsilon^2)\,,$$

sauf dans le cas très particulier où $\frac{df_0}{dx}(x_L) = 0$. Le même problème se pose évidemment en $x = x_R$.

Pour le résoudre, on ajoute donc à la solution formelle intérieure ci-dessus des termes de couche limite d'ordre $O(\epsilon)$, c'est-à-dire qu'on considère une solution formelle du type

$$F_\epsilon(x,\mu) = \sum_{n\geq 0} \epsilon^n f_n(x,\mu) + \sum_{n\geq 1} \epsilon^n \phi_n^L\left(\frac{x-x_L}{\epsilon},\mu\right) + \sum_{n\geq 1} \epsilon^n \phi_n^L\left(\frac{x_R-x}{\epsilon},-\mu\right).$$

A l'ordre dominant, le terme de couche limite en $x = x_L$ vérifie le problème de demi-espace

$$\begin{cases} \mu \frac{\partial \phi_1^L}{\partial z}(z,\mu) + a(\phi_1^L(z,\mu) - \langle \phi_1^L \rangle(z)) = 0\,, \quad z>0\,,\ |\mu|\leq 1\,, \\ \\ \phi_1^L(0,\mu) = C_1(x_L) - \frac{1}{a}\mu \frac{df_0}{dx}(x_L)\,. \end{cases}$$

On montre alors que, lorsque $\epsilon \to 0^+$, la solution de l'équation de Boltzmann linéaire f_ϵ vérifie

$$f_\epsilon = g_\epsilon + O(\epsilon^2)$$

où la fonction g_ϵ est la solution du problème au limites

$$\begin{cases} g_\epsilon - \frac{1}{3a}\frac{d^2}{dx^2} g_\epsilon = q\,, \quad x_L < x < x_R\,, \\ \\ g_\epsilon(x_L) - \epsilon \frac{\Lambda}{a}\frac{dg_\epsilon}{dx}(x_L) = f_L\,, \\ \\ g_\epsilon(x_R) + \epsilon \frac{\Lambda}{a}\frac{dg_\epsilon}{dx}(x_R) = f_R\,, \end{cases}$$

avec

$$\Lambda = \frac{\sqrt{3}}{2}\int_0^1 \mu^2 H(\mu) d\mu\,,$$

où H est la fonction de Chandrasekhar. Un calcul numérique montre que l'on a environ $\Lambda \simeq 0.7104$.

On remarquera que la condition aux limites pour le problème de diffusion a changé, entre le cas de l'approximation à l'ordre $O(\epsilon)$ et celui de l'approximation à l'ordre $O(\epsilon^2)$. Dans le cas de l'approximation à l'ordre $O(\epsilon)$, l'équation de diffusion apparaît avec des conditions de Dirichlet, tandis que dans le cas de l'approximation à l'ordre $O(\epsilon^2)$, la même équation apparaît avec des conditions aux limites mélangeant les valeurs de la fonction et de sa dérivée normale au bord du domaine spatial, qui sont appelées *conditions de Robin* ou parfois (en neutronique) conditions d'albedo (voir par exemple le chapitre 11 dans [43]).

La quantité $\epsilon\frac{\Lambda}{a}$ est homogène à une longueur et porte le nom de *longueur d'extrapolation*. En effet, la formule de Taylor pour g_ϵ en x_L permet de voir la condition

$$g_\epsilon(x_L) - \epsilon \frac{\Lambda}{a}\frac{dg_\epsilon}{dx}(x_L) = f_L$$

comme analogue à la condition de Dirichlet

$$g_\epsilon\left(x_L - \epsilon\frac{\Lambda}{a}\right) = f_L + O(\epsilon^2)\,.$$

Autrement dit, la condition de Robin est équivalente à une condition de Dirichlet sur un intervalle un peu plus grand que $[x_L, x_R]$, auquel on a rajouté un segment de longueur $\epsilon\frac{\Lambda}{a}$ à chaque extrémité. Bien que la fonction g_ϵ ne soit définie a priori que sur l'intervalle $[x_L, x_R]$, prescrire la condition de Robin est asymptotiquement équivalent à extrapoler g_ϵ sur un intervalle un peu plus grand, sur les bords duquel la condition de Robin est équivalente — à $O(\epsilon^2)$ près — à une condition de Dirichlet.

4.4 Diffusion à flux limité

De nouveau, nous allons restreindre notre attention à l'équation de Boltzmann linéaire monocinétique avec scattering isotrope et symétrie de la plaque infinie, présentée au chapitre 1.

Considérons le problème de Cauchy d'inconnue $f_\epsilon(t,x,\mu)$

$$\begin{cases} \epsilon\frac{\partial f_\epsilon}{\partial t} + \mu\partial_x f_\epsilon + \dfrac{a}{\epsilon}(f_\epsilon - \langle f_\epsilon\rangle) = 0\,, \quad x\in\mathbf{R}\,,\ |\mu|\le 1\,, \\ \\ f_\epsilon\big|_{t=0} = f^{in}\,, \end{cases}$$

où $a>0$ et $f^{in}\equiv f^{in}(x)$ est — pour fixer les idées — de classe C^∞ à support compact sur $\mathbf{R}$. La notation $\langle f_\epsilon\rangle$ désigne la moyenne angulaire de f_ϵ, c'est-à-dire que

$$\langle\phi\rangle := \tfrac{1}{2}\int_{-1}^{1}\phi(\mu)d\mu\,,$$

pour tout $\phi\in L^1([-1,1])$.

L'approximation par la diffusion consiste à approcher f_ϵ par la solution u de l'équation

$$\begin{cases} \frac{\partial u}{\partial t} - \frac{1}{3a}\frac{\partial^2 u}{\partial x^2} = 0\,, \quad x\in\mathbf{R}\,,\ t>0\,, \\ \\ u\big|_{t=0} = f^{in}\,. \end{cases}$$

Rappelons (cf. chapitre 1, section 1.1.3) que cette équation de diffusion correspond à écrire que f_ϵ satisfait l'équation de continuité

$$\frac{\partial}{\partial t}\langle f_\epsilon\rangle + \frac{\partial}{\partial x}\left(\frac{1}{\epsilon}\langle\mu f_\epsilon\rangle\right) = 0\,,$$

puis à écrire la loi de Fick, sous la forme

$$\frac{1}{\epsilon}\langle\mu f_\epsilon\rangle \simeq -\frac{1}{3a}\frac{\partial}{\partial x}\langle f_\epsilon\rangle\,.$$

Bien que l'approximation par la diffusion soit une théorie asymptotique qui n'est valable *stricto sensu* qu'après passage à la limite pour $\epsilon \to 0^+$, on souhaite pouvoir l'utiliser en pratique comme approximation pour $\epsilon > 0$ éventuellement très petit, mais pas forcément nul. Si d'autre part la donnée initiale présente des zones de fort gradient spatial, il peut arriver que l'on ait

$$\frac{1}{3a}\left|\frac{\partial}{\partial x}\langle f_\epsilon\rangle(t,x)\right| \gg \frac{1}{\epsilon}\langle f_\epsilon\rangle(t,x)\,,$$

ce qui est en contradiction avec la loi de Fick

$$\frac{1}{\epsilon}\langle \mu f_\epsilon\rangle \simeq -\frac{1}{3a}\frac{\partial}{\partial x}\langle f_\epsilon\rangle\,.$$

En effet, si $f_\epsilon \geq 0$, on doit avoir

$$\frac{1}{\epsilon}|\langle \mu f_\epsilon\rangle(t,x)| \leq \frac{1}{\epsilon}\langle f_\epsilon\rangle\,,$$

puisque $|\mu| \leq 1$.

Evidemment, ceci est lié à la perte éventuelle de positivité dans la solution formelle de Hilbert pour l'équation de Boltzmann linéaire. En effet, rappelons que cette solution formelle est donnée par

$$F_\epsilon = \sum_{n\geq 0} \epsilon^n f_n(t,x,\mu)\,,$$

avec

$$f_0 \equiv f_0(t,x)\,,$$
$$f_1 \equiv C_1(t,x) - \frac{1}{a}\mu\frac{\partial f_0}{\partial x}(t,x)\,.$$

Il est donc parfaitement possible que, si $\epsilon > 0$ n'est pas pris assez petit, et en présence de valeurs élevées de la dérivée spatiale $\partial_x f_0(t,x)$, la somme des deux premiers termes de la solution formelle soit négative, c'est-à-dire que l'on ait

$$(f_0 + \epsilon C_1)(t,x) - \frac{\epsilon}{a}\mu\frac{\partial f_0}{\partial x}(t,x) < 0,$$

pour certaines valeurs de (t,x,μ).

Il existe plusieurs façons de remédier à ce genre de difficulté, qui est particulièrement gênante lorsqu'on veut utiliser la modélisation par la diffusion dans des régimes où le taux de scattering est grand sans être très grand — c'est-à-dire, de manière équivalente, où $\epsilon > 0$ est petit, mais pas infinitésimalement petit. L'idée consiste à modifier l'équation de diffusion de façon à limiter la taille du flux calculé par la loi de Fick. Bien souvent, ces limiteurs de flux sont introduits de façon *ad hoc*, et sont adaptés au contexte applicatif particulier dans lequel on veut les utiliser.

Il existe pourtant une méthode systématique pour obtenir l'équation de diffusion à flux limités, qui a été proposée en 1979 par C.D. Levermore, et dont nous allons donner une brève présentation.

Une idée permettant d'éviter toute perte de positivité dans la solution formelle de Hilbert consiste à la chercher sous la forme

$$F_\epsilon = \exp\left(\sum_{n\geq 0} \epsilon^n A_n(t,x,\mu)\right) .$$

(Là encore, on ne cherchera pas à montrer que cette série converge : l'objet ci-dessus est seulement une expression formelle en puissances de ϵ.)

Substituons cette expression dans l'équation de Boltzmann linéaire monocinétique avec scattering isotrope et symétrie du slab : on aboutit aux équations suivantes.

Ordre ϵ^{-1} :

$$a\left(e^{A_0(t,x,\mu)} - \left\langle e^{A_0(t,x,\cdot)}\right\rangle\right) = 0\,,$$

Ordre ϵ^0 :

$$\mu\frac{\partial}{\partial x}e^{A_0(t,x,\mu)} + a\left(e^{A_0(t,x,\mu)}A_1(t,x,\mu) - \left\langle e^{A_0(t,x,\cdot)}A_1(t,x,\cdot)\right\rangle\right) = 0\,.$$

Nous n'expliciterons pas les conditions suivantes — dont ne nous servirons d'ailleurs pas.

L'équation à l'ordre $O(\epsilon^{-1})$ entraîne que

$$e^{A_0(t,x,\mu)} \text{ est indépendant de } \mu\,,$$

c'est-à-dire que

$$A_0 \equiv A_0(t,x)\,.$$

L'équation à l'ordre $O(\epsilon^0)$ devient alors

$$\mu e^{A_0(t,x)}\frac{\partial A_0}{\partial x}(t,x) + a e^{A_0(t,x)}(A_1(t,x,\cdot) - \langle A_1(t,x,\cdot)\rangle) = 0\,,$$

d'où l'on tire que

$$A_1(t,x,\mu) = C_1(t,x) - \frac{1}{a}\mu\frac{\partial A_0}{\partial x}(t,x)\,, \quad C_1 = \langle A_1\rangle\,.$$

Quitte à regrouper les termes $A_0 + \epsilon\langle A_1\rangle$ en une seule fonction

$$B_\epsilon(t,x) := A_0 + \epsilon\langle A_1\rangle$$

la solution formelle ci-dessus devient

$$F_\epsilon(t,x,\mu) = \exp\left(B_\epsilon(t,x) - \frac{\epsilon}{a}\mu\frac{\partial B_\epsilon}{\partial x}(t,x) + O(\epsilon^2)\right) .$$

Ecrivons que cette solution formelle vérifie l'équation de continuité

$$\frac{\partial}{\partial t}\langle F_\epsilon\rangle + \frac{\partial}{\partial x}\left(\frac{1}{\epsilon}\langle \mu F_\epsilon\rangle\right) = 0\,,$$

obtenue en moyennant chaque membre de l'équation de Boltzmann linéaire par rapport à μ.

Observons [2] que

$$\tfrac{1}{2}\int_{-1}^{1} e^{z\mu} d\mu = \frac{\sinh z}{z}\,,$$

de sorte que

$$\tfrac{1}{2}\int_{-1}^{1} \mu e^{z\mu} d\mu = \frac{d}{dz}\left(\frac{\sinh z}{z}\right) = \frac{z\cosh z - \sinh z}{z^2}\,.$$

Donc, en posant

$$\rho_\epsilon(t,x) := \langle F_\epsilon\rangle \simeq e^{B_\epsilon(t,x)}\frac{\sinh((\epsilon/a)\partial B_\epsilon/\partial x)}{(\epsilon/a)\partial B_\epsilon/\partial x} \simeq e^{B_\epsilon(t,x)}\,,$$

on voit que

$$\frac{1}{\epsilon}\langle \mu F_\epsilon\rangle = \frac{1}{\epsilon}\frac{\langle \mu F_\epsilon\rangle}{\langle F_\epsilon\rangle}\langle F_\epsilon\rangle \simeq -\frac{1}{\epsilon}D\left(\frac{\epsilon}{a}\frac{\partial B_\epsilon}{\partial x}\right)\rho_\epsilon \simeq -\frac{1}{\epsilon}D\left(\frac{\epsilon}{a}\frac{\partial \rho_\epsilon/\partial x}{\rho_\epsilon}\right)\rho_\epsilon$$

où D est la fonction holomorphe sur la bande $|\Im z| < \pi$ définie pour $z \neq 0$ par

$$D(z) = \coth z - \frac{1}{z}\,.$$

Le calcul formel ci-dessus suggère donc de remplacer l'équation de diffusion habituelle

$$\partial_t u - \frac{1}{3a}\frac{\partial^2 u}{\partial x^2} = 0\,, \quad x \in \mathbf{R}\,,\ t > 0\,,$$

par l'équation de diffusion non linéaire

$$\partial_t \rho_\epsilon - \partial_x\left(\frac{1}{\epsilon}D\left(\frac{\epsilon}{a}\frac{\partial \rho_\epsilon/\partial x}{\rho\epsilon}\right)\rho_\epsilon\right) = 0\,.$$

Comme $D(z) \sim \frac{1}{3}z$ lorsque $z \to 0$, on trouve que, si

$$\rho_\epsilon \to u \quad \text{et} \quad \frac{\partial \rho_\epsilon}{\partial x} \to \frac{\partial u}{\partial x}$$

ponctuellement en (t,x) lorsque $\epsilon \to 0^+$, alors

$$\frac{1}{\epsilon}\langle \mu F_\epsilon\rangle \simeq -\frac{1}{\epsilon}D\left(\frac{\epsilon}{a}\frac{\partial \rho_\epsilon/\partial x}{\rho\epsilon}\right)\rho_\epsilon \to -\frac{1}{3a}\frac{\partial u}{\partial x}\,,$$

de sorte que l'équation de diffusion non linéaire ci-dessus se ramène à l'équation de diffusion usuelle dans la limite $\epsilon \to 0^+$.

2. La fonction $z \mapsto \frac{\sinh z}{z}$ se prolonge en une fonction holomorphe sur $\mathbf{C}$, prenant la valeur 1 en $z = 0$.

D'autre part, l'équation de diffusion non linéaire vérifie évidemment que

$$\left|\frac{\langle \mu F_\epsilon\rangle}{\langle F_\epsilon\rangle}\right| \simeq \left|D\left(\frac{\epsilon}{a}\frac{\partial \rho_\epsilon/\partial x}{\rho\epsilon}\right)\right| \leq 1\,,$$

c'est-à-dire que la condition de flux limité est automatiquement vérifiée par l'équation non linéaire — on a tout fait pour cela, notamment en imposant à la solution formelle F_ϵ d'être inconditionnellement positive. En effet,

$$D'(z) = \frac{1}{z^2} - \frac{1}{\sinh^2 z} \geq 0 \text{ pour tout } z \in \mathbf{R}^*\,,$$

de sorte que D est une fonction croissante impaire sur $\mathbf{R}$ vérifiant

$$D(0) = 0\,, \quad \text{et } \lim_{z\to+\infty} D(z) = 1\,.$$

La justification rigoureuse de cette approximation par la diffusion limitée reste à établir.

4.5 Interprétation probabiliste de l'équation de diffusion

Considérons de nouveau le problème de Cauchy pour l'équation de la chaleur posée dans $\mathbf{R}^N$

$$\begin{cases} \dfrac{\partial u}{\partial t} - \frac{1}{2}\Delta_x u = 0\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,, \\ \\ u\big|_{t=0} = u^{in}\,. \end{cases}$$

Pour tout $u^{in} \in L^2(\mathbf{R}^N)$, on sait que le problème de Cauchy ci-dessus admet une unique solution $u \in C^\infty(\mathbf{R}_+^* \times \mathbf{R}^N)$ donnée par la formule

$$u(t,x) = \int_{\mathbf{R}^N} E(t, x-y)u^{in}(y)dy\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,,$$

où

$$E(t,x) = \frac{1}{(2\pi t)^{N/2}} \exp\left(-\frac{|x|^2}{2t}\right)\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,.$$

Cette formule montre que l'on ne peut espérer résoudre l'équation de la chaleur par une méthode des caractéristiques. Autrement dit, même si la donnée initiale vérifie $u^{in} \in C^\infty(\mathbf{R}^N)$, il n'est pas possible d'exprimer la solution u de l'équation de la chaleur ci-dessus par une formule du type

$$u(t,x) = u^{in}(\gamma(t;x))\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,,$$

où $\mathbf{R}_+ \ni x \mapsto \gamma(t;x) \in \mathbf{R}^N$ est, pour tout $x \in \mathbf{R}^N$, une courbe paramétrée tracée dans $\mathbf{R}^N$ telle que $\gamma(0;x) = x$, vérifiant par exemple $|\gamma(t;x)| \to +\infty$ lorsque $|x| \to +\infty$ et $\gamma \in C^2(\mathbf{R}_+ \times \mathbf{R}^N; \mathbf{R}^N)$.

Exercice 4.1 *Démontrer cette impossibilité. (Indication : supposer que la donnée initiale $u^{in} \geq 0$ est une fonction continue à support compact dans $\mathbf{R}^N$, et démontrer que la solution u du problème de Cauchy pour l'équation de la chaleur vérifie* $\mathrm{supp}(u(t,\cdot)) = \mathbf{R}^N$ *pour tout $t > 0$).*

Soit $S(t)$ l'application linéaire $u^{in} \mapsto S(t)u^{in} = u(t,\cdot)$ pour tout $t > 0$. Si on note

$$\hat{u}(t,\xi) := \int_{\mathbf{R}^N} e^{-i\xi\cdot x} u(t,x)dx$$

la transformation de Fourier partielle en x, on voit que

$$\hat{u}(t,\xi) = \exp(-\tfrac{1}{2}t|\xi|^2)\hat{u}^{in}(\xi)\,,$$

puisque

$$u(t,\cdot) = E(t,\cdot) \star u^{in}\,, \qquad \text{et } \hat{E}(t,\xi) = \exp(-\tfrac{1}{2}t|\xi|^2)\,.$$

En particulier, pour tout $t_1, t_2 \geq 0$, on a

$$S(t_1 + t_2) = S(t_1) \circ S(t_2)\,.$$

Soit $t > 0$ et $n \in \mathbf{N}^*$; on définit le pas de temps $\Delta t = t/n$. Alors

$$u(t,\cdot) = S(\Delta t)^n u^{in}\,.$$

Exprimons cette égalité en variables physiques, au lieu des variables de Fourier. On trouve que

$$u(t,x) = \underbrace{E(\Delta t,\cdot) \star \ldots \star E(\Delta t,\cdot)}_{n \text{ termes}} \star u^{in}(x)\,,$$

ce qui s'exprime explicitement comme suit :

$$\begin{aligned} u(t,x) &= \int_{\mathbf{R}^N} \ldots \int_{\mathbf{R}^N} E(\Delta t, x - x_1)\ldots E(\Delta t; x_{n-1} - x_n))u^{in}(x^n)dx_1\ldots dx_n \\ &= \int_{\mathbf{R}^N} \ldots \int_{\mathbf{R}^N} \exp\left(-\tfrac{1}{2}\Delta t \sum_{k=1}^{n} \frac{|x_k - x_{k-1}|^2}{\Delta t^2}\right) u^{in}(x^n) \frac{dx_1\ldots dx_n}{(2\pi\Delta t)^{nN/2}}\,, \end{aligned}$$

en posant $x_0 = x$.

A partir de maintenant, on se contente d'un raisonnement purement formel. L'idée consiste à voir les points x_k pour $k = 0,\ldots,n$ comme les points situés sur une courbe paramétrée $t \mapsto X(t)$ correspondant aux valeurs $t = k\Delta t$ du paramètre. Lorsque $n \to +\infty$, on a $\Delta t = t/n \to 0$ de sorte que

$$\frac{|x_k - x_{k-1}|^2}{\Delta t^2} \simeq |\dot{X}((k-1)\Delta t)|^2\,.$$

L'argument de l'exponentielle sous l'intégrale évoque donc la somme de Riemann

$$\tfrac{1}{2}\Delta t \sum_{k=1}^{n} \frac{|x_k - x_{k-1}|^2}{\Delta t^2} \simeq \tfrac{1}{2}\Delta t \sum_{k=1}^{n} |\dot{X}((k-1)\Delta t)|^2 \simeq \tfrac{1}{2}\int_0^t |\dot{X}(s)|^2 ds\,.$$

Quant à l'intégrale n-uple

$$\underbrace{\int_{\mathbf{R}^N}\cdots\int_{\mathbf{R}^N}}_{n \text{ fois}} \cdots \frac{dx_1\dots dx_n}{(2\pi\Delta t)^{nN/2}}$$

dans la limite $n \to +\infty$, on peut y penser comme une intégration par rapport à la "mesure de Lebesgue produit"

$$\prod_{0\le s\le t} dX(s)\,,$$

où X décrit une certaine classe de courbes tracées dans $\mathbf{R}^N$ telles que $X(0) = x$. On peut penser à cette expression comme à l'analogue, dans l'espace $(\mathbf{R}^N)^{[0,t]}$ de toutes les applications de $[0,t]$ dans $\mathbf{R}^N$, de la mesure de Lebesgue $dy_1\dots dy_n$ dans $\mathbf{R}^n$.

Ces remarques nous conduisent à introduire la notation

$$\mathcal{D}X = \lim_{n\to+\infty} \frac{dx_1\dots dx_n}{(2\pi\Delta t)^{nN/2}}\,,$$

de sorte que

$$u(t,x) = \int \exp\left(-\tfrac{1}{2}\int_0^t |\dot X(s)|^2 ds\right) u^{in}(X(t))\mathcal{D}X\,.$$

Ce raisonnement est grossièrement faux du point de vue mathématique. Pour commencer, on ne peut pas donner un sens à la "mesure de Lebesgue produit" ci-dessus, qui fait intervenir le produit d'une infinité non dénombrable de termes. De plus ce raisonnement ne tient pas compte du facteur de normalisation $\frac{1}{(2\pi\Delta t)^{nN/2}}$, dont la présence ne s'explique qu'en raison du terme exponentiel. Ceci suggère donc de ne pas séparer les facteurs

$$\exp\left(-\tfrac{1}{2}\int_0^t |\dot X(s)|^2 ds\right) \quad \text{et} \quad \mathcal{D}X$$

comme on l'a fait ci-dessus. Au contraire, c'est le produit

$$\exp\left(-\tfrac{1}{2}\int_0^t |\dot X(s)|^2 ds\right)\mathcal{D}X$$

qui a un sens et qui s'interprète comme une mesure de probabilité sur l'espace $C(\mathbf{R}_+;\mathbf{R}^N)_0$ des applications continues de $\mathbf{R}_+$ dans $\mathbf{R}^N$ prenant la valeur 0 en $t = 0$.

Plus précisément, on démontre qu'il existe une unique mesure de probabilité — appelée "mesure de Wiener" — sur l'espace $C(\mathbf{R}_+;\mathbf{R}^N)_0$ telle que, pour tout $m \ge 1$, toute suite finie de parties mesurables $\vec A = (A_1,\dots,A_m)$ de $\mathbf{R}^N$ et tout $\vec t = (t_1,\dots,t_m) \in (\mathbf{R}_+^*)^m$, la probabilité de l'"ensemble cylindrique"

$$C(\vec t,\vec A) := \{\gamma \in C(\mathbf{R}_+;\mathbf{R}^N)_0 \,|\, \gamma(t_k) \in A_k\,,\ 1 \le k \le m\}$$

soit

$$\mathrm{Prob}(C(\vec{t},\vec{A})) = \int_{A_1}\int_{A_2}\dots\int_{A_{m-1}}\int_{A_m} E(t_1,x_1)E(t_2-t_1,x_2-x_1)\dots \\ E(t_m-t_{m-1},x_m-x_{m-1})dx_1dx_2\dots dx_{m-1}dx_m\,,$$

où on rappelle que

$$E(t,x) = \frac{1}{(2\pi t)^{N/2}}\exp\left(-\frac{|x|^2}{2t}\right)\,,\quad x\in\mathbf{R}^N\,,\ t>0\,.$$

Une fois que l'on dispose de la mesure de Wiener, la solution u du problème de Cauchy pour l'équation de la chaleur ci-dessus s'exprime par la formule

$$u(t,x) = \mathbf{E}(u^{in}(x+\gamma(t)))$$

où γ décrit l'espace $C(\mathbf{R}_+;\mathbf{R}^N)_0$ et où l'espérance $\mathbf{E}$ est prise sous la mesure de Wiener.

Cette formule est l'analogue exact de celle donnant la solution du problème de Cauchy pour l'équation de Boltzmann linéaire à partir du processus de transport, obtenue au chapitre précédent (voir section 3.3.)

En effet, l'argument de la donnée initiale dans l'expression ci-dessus, soit $x+\gamma(t)$, peut-être vu comme un processus stochastique X_t partant de x pour $t=0$, en posant $X_t(\gamma) = x+\gamma(t)$, où l'aléa est le chemin $\gamma\in C(\mathbf{R}_+;\mathbf{R}^N)_0$ et la mesure de probabilité la mesure de Wiener. Ainsi, la formule ci-dessus exprimant la solution u du problème de Cauchy pour l'équation de la chaleur se met-elle sous la forme

$$u(t,x) = \mathbf{E}^x(u^{in}(X_t))\,,$$

où X_t est le processus stochastique ainsi obtenu, et où $\mathbf{E}^x$ désigne l'espérance prise sur les trajectoires de ce processus partant de x à $t=0$.

En réalité, le processus stochastique naturel dans ce contexte est le *mouvement brownien* standard habituellement noté B_t (ou W_t, en l'honneur du mathématicien Norbert Wiener) défini en posant $B_t(\gamma) = \gamma(t)$, où l'aléa est le chemin $\gamma\in C(\mathbf{R}_+;\mathbf{R}^N)_0$ et la mesure de probabilité la mesure de Wiener. La solution du problème de Cauchy pour l'équation de la chaleur s'écrit donc

$$u(t,x) = \mathbf{E}(u^{in}(x+B_t))$$

Dans la formule ci-dessus, l'espérance mathématique est prise par rapport à la mesure de Wiener définie sur $C(\mathbf{R}_+;\mathbf{R}^N)_0$.

Le mouvement brownien intervient dans des contextes très variés. Le nom même de mouvement brownien évoque les observations de R. Brown (1827) sur le mouvement de particules des grains de pollen de *Clarckia Pulchella* immergées dans de l'eau. Les premières études mathématiques sur le processus stochastique appelé mouvement brownien remontent au début du XXème siècle, avec les travaux de L. Bachelier sur les cours de la bourse, qui sont à l'origine

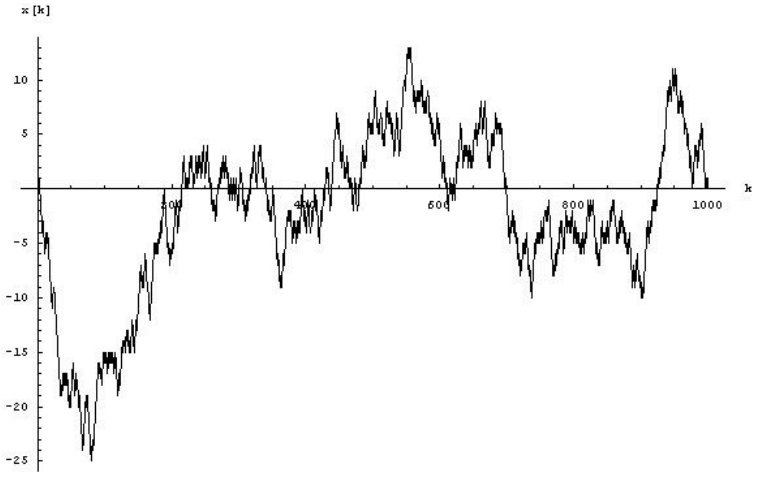

FIGURE 4.1 – Une trajectoire du mouvement brownien : graphe de $t \mapsto B_t$ dans le cas de dimension $N = 1$.

des applications des mathématiques à la finance, et avec ceux d'A. Einstein en mécanique statistique. Mais le mouvement brownien joue également un rôle important en théorie conforme des champs, en topologie (théorie des cartes)...

Le mouvement brownien peut être caractérisé par les propriétés suivantes (en supposant pour simplifier que $N = 1$) :

(a) $B_0 = 0$ p.s. ;
(b) les trajectoires de $(B_t)_{t\geq 0}$ sont toutes continues ;
(c) pour tout $t > s \geq 0$, la variable aléatoire $B_t - B_s$ est indépendante de la tribu engendrée par les B_r, pour tout $0 \leq r \leq s$;
(d) pour tout $t > s \geq 0$, la variable aléatoire $B_t - B_s$ suit la loi gaussienne de moyenne 0 et de variance $t - s$.

En particulier, la propriété (d) implique que, pour tout $t > s \geq 0$

$$\mathbf{E}(B_t - B_s) = 0\,, \quad \mathbf{E}(|B_t - B_s|^2) = t - s\,.$$

Bien que les trajectoires du processus brownien soient toutes continues, elles sont presque sûrement non dérivables ; voir la figure 4.5 pour avoir une idée de l'allure des trajectoires du mouvement brownien.

Le lecteur intéressé à approfondir ces questions pourra consulter le chapitre 14 du cours de J.-F. Le Gall [35].

Retournons à la formule explicite donnant la solution du problème de Cauchy pour l'équation de la chaleur à partir du mouvement brownien, soit

$$u(t, x) = \mathbf{E}(u^{in}(x + B_t))\,.$$

Cette formule est l'analogue probabiliste de la formule déduite de la méthode

des caractéristiques pour résoudre l'équation de transport

$$\begin{cases} \dfrac{\partial f}{\partial t} + v \cdot \nabla_x f = 0\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,, \\ f\big|_{t=0} = f^{in}\,, \end{cases}$$

à savoir

$$f(t,x) = f^{in}(x - tv)\,.$$

En effet, cette formule peut s'interpréter comme

$$f(t,x) = \mathbf{E}(f^{in}(x + \gamma(t)))$$

où $\mathbf{E}$ désigne l'espérance par rapport à la mesure de probabilité sur $C(\mathbf{R}_+;\mathbf{R}^N)_0$ définie par la masse de Dirac concentrée sur la droite $t \mapsto \gamma(t) = -tv$.

Comme la solution de l'équation de Boltzmann linéaire

$$\begin{cases} \epsilon \frac{\partial f_\epsilon}{\partial t} + v \cdot \nabla_x f_\epsilon + \dfrac{a}{\epsilon}\Big(f_\epsilon - (1+\epsilon^2\gamma)\mathcal{K} f_\epsilon\Big) = 0\,, \quad (x,v) \in \mathbf{R}^N \times \mathcal{V}\,,\ t > 0\,, \\ \\ f_\epsilon\big|_{t=0} = f^{in}\,, \end{cases}$$

est donnée par la formule

$$f_\epsilon(t,x,v) = \mathbf{E}^{x,v}(f^{in}(X_t^\epsilon(x,v), V_t^\epsilon(x,v)))\,,$$

où $(X_t^\epsilon, V_t^\epsilon)_{t\geq 0}$ est le processus de transport associé, la convergence de f_ϵ vers la solution u de l'équation de la chaleur

$$\begin{cases} \dfrac{\partial u}{\partial t} - \frac{1}{2}\Delta_x u = 0\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,, \\ \\ u\big|_{t=0} = u^{in}\,, \end{cases}$$

donnée par la formule

$$u(t,x) = \mathbf{E}(u^{in}(x + B_{\kappa^2 t}))\,,$$

s'interprète comme la convergence en un certain sens du "processus des positions" du processus de transport, à savoir $(X_t^\epsilon)_{t\geq 0}$, vers le processus de diffusion $x + B_{\kappa^2 t}$.

Nous renvoyons au cours de C. Graham et D. Talay [26] pour approfondir les applications de ce point de vue.

4.6 Exercices

Exercice 4.2 *Nous considérons l'équation de Boltzmann linéaire monocinétique en dimension 2 d'espace. La vitesse s'écrit $v = (\cos\theta, \sin\theta)$ et la position est notée $x \in \mathbf{R}^2$. La fonction de distribution $f = f(t,x,\theta)$ est donc solution de*

$$\frac{\partial f}{\partial t} + \frac{1}{\varepsilon} v \cdot \nabla_x f = \frac{\sigma}{\varepsilon^2}(K(f) - f)\,,$$

avec

$$K(f) := \frac{\int_0^{2\pi} k(\theta-\mu) f(x,t,\mu) d\mu}{\int_0^{2\pi} k(\mu) d\mu}\,.$$

Le noyau de collision k est une fonction continue 2π-périodique strictement positive et bornée.

1. *On rappelle le développement en série de Fourier d'une fonction g*

$$g(\theta) = \frac{1}{2\pi}\sum_{n\in\mathbf{Z}} g_n e^{in\theta}\,, \quad g_n = \int_0^{2\pi} g(\theta) e^{-in\theta} d\theta\,.$$

   Montrer que

$$K(f) = \frac{1}{2\pi k_0}\sum_{n\in\mathbf{Z}} k_n f_n e^{in\theta}$$

   et que $|k_n| < k_0$ pour tout $n \neq 0$.

2. *Pour étudier la limite de diffusion, on pose*

$$f = f^0 + \varepsilon f^1 + \varepsilon^2 f^2 + \dots$$

   et on écrit la hiérarchie d'équations suivant les puissances de ε. Appliquer la transformée de Fourier à ces équations.

3. *En déduire que f^0 est une fonction constante en la variable angulaire θ. En utilisant les deux équations suivantes (permettant de calculer f^1 et f^2) retrouver la limite de diffusion et montrer que cette équation dépend de σ ainsi que de k_0, k_1 et k_{-1}.*

4. *Mettre en place une stratégie pour démontrer que la solution $f(t,x,\theta)$ est proche de la solution de l'équation de diffusion dans L^2.*

Exercice 4.3 (Temps de sortie du soleil pour des photons) *Soit un soleil monodimensionnel dans lequel sont présents (et créés) des photons. On considère l'équation en dimension 1 d'espace*

$$\frac{1}{c}\left(\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x}\right) = \sigma(\langle f\rangle - f)\,, \quad |v| = c\,,\ |x| \le R\,,$$

où on a posé

$$\langle f\rangle(t,x) = \frac{1}{2}\Big(f(x,c,t) + f(x,-c,t)\Big)\,.$$

1. *Expliquer à partir de considérations élémentaires pourquoi le terme $\langle f\rangle - f$ peut s'interpréter comme une redistribution aléatoire de particules (les photons) sur la matière lourde composant le soleil.*

2. *Quelle est la dimension de la constante σ ?*

   Pour déterminer une valeur raisonnable de cette constante, on revient sur le problème en dimension 3 d'espace. Montrer que, pour un soleil de rayon R et de masse M, alors

$$\sigma \approx \left(\frac{M}{\frac{4\pi}{3}R^3 m}\right)^{\frac{1}{3}} \propto \frac{\left(\frac{M}{m}\right)^{\frac{1}{3}}}{R}$$

convient, en notant m la masse du nucléon.

3. *Déterminer σ pour les valeurs physiques*

$$M = 2 \cdot 10^{30} kg\,, \quad R = 10^6 km\,, \quad m = 1.6 \cdot 10^{-27} kg\,.$$

4. *Quelle est la limite de diffusion ? Interpréter la solution d'un point de vue probabiliste.*
5. *En déduire que la plupart des photons issus du coeur du soleil mettent (au moins) plusieurs milliers d'années pour en sortir. Comparer avec le temps d'arrivée sur terre (la distance terre-soleil est $D = 1.5 \cdot 10^8 km$).*

Indication : il est utile de faire un parallèle avec la théorie probabiliste qui est très puissante pour arriver à une interprétation physique correcte. Essentiellement les photons à l'intérieur du soleil ont des trajectoires de type brownien avec des "rebonds" multiples : c'est pour cela qu'ils mettent un temps très grand pour sortir du soleil. Pour la deuxième partie de la question 2, l'idée est d'interpréter le libre parcours moyen comme un multiple de la distance moyenne entre les nucléons. S'ils sont répartis à peu près régulièrement dans le volume occupé par le soleil, on a donc une densité de masse qui vérifie $\rho \propto m\sigma^3$. Or la densité totale s'écrit aussi $\rho = \left(\frac{4\pi}{3} R^3\right)^{-1} M$. D'où le résultat.

Exercice 4.4 (Le problème de Milne) *Ce problème concerne l'absorption d'un groupe de neutrons incidents monocinétiques (la vitesse étant normalisée à 1) sur un demi espace avec un angle de vol θ par rapport à la normale. On note $\mu = \cos\theta$ et on considère l'équation*

$$\mu \frac{\partial u}{\partial x} + \sigma u = \frac{\sigma_s}{2} \int_{-1}^{1} u(x, \mu') d\mu' \,, \quad x \in]-\infty, 0]\,, \ \mu \in [-1, +1]\,,$$

munie de la condition aux limites

$$u(0, \mu) = g(-\mu) \quad \text{pour tout } \mu \in [-1, 0[\,.$$

On pose $u^+(x, \mu) = u(x, \mu)$ et $u^-(x, \mu) = u(x, -\mu)$ pour $\mu \in]0, 1]$. Les fonctions u^+ et u^- sont définies sur $]-\infty, 0] \times]0, 1]$.

L'objectif de cet exercices est de caractériser la distribution sortante de neutrons $u^+(0, \mu)$ en fonction des neutrons entrants dont la distribution est $u^-(0, \mu) = g(\mu)$, qui est donnée. Ce problème est connu sous le nom de "problème de l'albedo".

1. *Montrer que le problème stationnaire peut s'écrire, à un changement de variables près,*

$$\begin{cases} \mu \dfrac{\partial u^+}{\partial x} + u^+ - \dfrac{c}{2} \displaystyle\int_0^1 (u^+ + u^-) d\mu' = 0\,, \\ -\mu \dfrac{\partial u^-}{\partial x} + u^- - \dfrac{c}{2} \displaystyle\int_0^1 (u^+ + u^-) d\mu' = 0\,, \end{cases}$$

avec $c = \frac{\sigma_s}{\sigma}$.

2. *Expliquer formellement pourquoi on peut se ramener à écrire*

$$\mu u^+(0,\mu) = \int_0^1 R(\mu,\mu')\mu' u^-(0,\mu')d\mu' \,,$$

où R est une fonction à déterminer. Expliquer pourquoi on a aussi, pour tout $x < 0$,

$$\mu u^+(x,\mu) = \int_0^1 R(\mu,\mu')\mu' u^-(x,\mu')d\mu' \,.$$

3. *Vérifier qu'on peut éliminer u^+ et les dérivées spatiales dans les équations de la question 1 grâce à la question 2. Montrer par un calcul (assez lourd) qu'on trouve une équation intégrale non linéaire*

$$\int_0^1 \left(\frac{1}{\lambda}+\frac{1}{\mu}\right) R(\mu,\lambda)u^-(x,\lambda)\lambda d\lambda = \frac{c}{2}\left(1+\int_0^1 R(\mu,\lambda)d\lambda\right) \times \int_0^1 \left(u^-(x,\mu')+\frac{1}{\mu'}\int_0^1 R(\mu',\lambda)\lambda u^-(x,\lambda)d\lambda\right) d\mu' \,.$$

En déduire l'identité

$$\left(\frac{1}{\mu}+\frac{1}{\mu'}\right) R(\mu,\mu') = \frac{c}{2}\left(1+\int_0^1 R(\mu,\lambda)d\lambda\right)\left(\frac{1}{\mu'}+\int_0^1 \frac{R(\lambda,\mu')}{\lambda}d\lambda\right) \,.$$

4. *On pose $S(\mu,\mu') = \mu' R(\mu,\mu')$. Montrer que la fonction S est symétrique :*

$$S(\mu,\mu') = S(\mu',\mu) \,.$$

Proposer une interprétation physique de cette formule.

5. *On définit alors la fonction de Chandrasekhar*[3] *par*

$$H(\mu) = 1 + \int_0^1 \frac{S(\lambda,\mu)}{\lambda}d\lambda.$$

Montrer que la connaissance de la fonction H est suffisante pour résoudre le problème de l'albedo. Montrer que H vérifie l'équation intégrale non linéaire

$$H(\mu) = 1 + \frac{c}{2}\mu H(\mu)\int_0^1 \frac{H(\mu')}{\mu+\mu'}d\mu'.$$

6. *On suppose finalement que c est petit. Montrer que $H = 1$ en première approximation. En déduire qu'une solution approchée du problème de l'albedo est donnée par la formule*

$$u^+(0,\mu) \approx \frac{c}{2}\int_0^1 \frac{\mu' g(\mu')}{\mu+\mu'}d\mu' \,.$$

3. Subrahmanyan Chandrasekhar (1910-1995), lauréat du prix Nobel de physique en 1983 "pour ses travaux théoriques sur les processus physiques jouant un rôle important dans la structure et l'évolution des étoiles".

Indications : cet exercice qui peut ne sembler que calculatoire est en fait profond, car il s'agit de déterminer la fonction H de Chandrasekhar. Le rôle important joué par cette fonction a déjà été évoqué dans la section 4.3.2. La question 4 est particulièrement délicate. Pour la traiter, on pourra par exemple poser

$$T(\mu, \lambda) = S(\mu, \lambda) - S(\lambda, \mu)\,,$$

et démontrer que

$$\frac{T(\mu, \lambda)}{\lambda} = \frac{c}{2} \frac{\mu}{\mu + \lambda} \int_0^1 \frac{T(\mu, \mu') - T(\lambda, \mu')}{\mu'} d\mu'\,,$$

puis intégrer chaque membre de cette égalité par rapport à λ.

Chapitre 5

Méthodes numériques

5.1 Rappels sur la méthode des différences finies

Hormis quelques cas très particuliers, il est impossible de calculer explicitement les solutions des équations aux dérivées partielles. Il est donc nécessaire d'avoir recours au calcul numérique sur ordinateur pour estimer qualitativement et quantitativement ces solutions. Le principe de toutes les méthodes de résolution numérique des équations aux dérivées partielles est d'obtenir des valeurs numériques discrètes (c'est-à-dire en nombre fini) qui **"approchent"** (en un sens convenable à préciser) la solution exacte. Nous présentons ici une telle méthode, dite des différences finies, qui **discrétise** le problème en représentant des fonctions par un nombre fini de valeurs.

5.1.1 Principes de la méthode pour l'équation de diffusion

Nous rappelons les notations et résultats essentiels de la méthode des différences finies sur un premier exemple, à savoir l'équation de diffusion (pour plus de détails nous renvoyons à [2]). Pour simplifier la présentation, nous nous limitons à la dimension un d'espace. Nous considérons l'équation de diffusion dans le domaine borné $(0,1)$

$$\begin{cases} \dfrac{\partial u}{\partial t} - \nu \dfrac{\partial^2 u}{\partial x^2} = 0 \text{ pour } (x,t) \in (0,1) \times \mathbf{R}_*^+ \\ u(t,0) = u(t,1) = 0 \text{ pour } t \in \mathbf{R}_*^+ \\ u(0,x) = u_0(x) \text{ pour } x \in (0,1), \end{cases} \tag{5.1}$$

où $\nu > 0$ est un coefficient de diffusion constant.

Pour discrétiser le continuum spatio-temporel, on introduit un **pas d'espace** $\Delta x = 1/(N+1) > 0$ (N entier positif) et un **pas de temps** $\Delta t > 0$ qui seront les plus petites échelles représentées par la méthode numérique. On définit un maillage (voir la Figure 5.1) ou des coordonnées discrètes de l'espace et du temps

$$(t_n, x_j) = (n\Delta t, j\Delta x) \text{ pour } n \geq 0, j \in \{0, 1, ..., N+1\}.$$

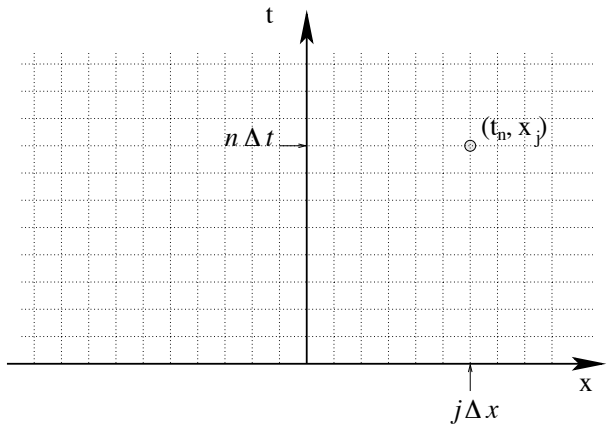

FIGURE 5.1 – Maillage en différences finies.

On note $u(t,x)$ la solution exacte (inconnue) de (5.1) et u_j^n une solution discrète approchée au point (t_n, x_j), c'est-à-dire que $u_j^n \approx u(t_n, x_j)$ (dans un sens à préciser). Le principe de la méthode des différences finies est de remplacer les dérivées par des différences finies en utilisant des formules de Taylor dans lesquelles on néglige les restes. Par exemple, on approche la dérivée seconde en espace (le Laplacien en dimension un) par

$$-\frac{\partial^2 u}{\partial x^2}(t_n, x_j) \approx \frac{-u_{j-1}^n + 2u_j^n - u_{j+1}^n}{(\Delta x)^2} \tag{5.2}$$

où l'on reconnaît la formule de Taylor

$$\frac{-u(t,x-\Delta x) + 2u(t,x) - u(t,x+\Delta x)}{(\Delta x)^2} = -\frac{\partial^2 u}{\partial x^2}(t,x) - \frac{(\Delta x)^2}{12}\frac{\partial^4 u}{\partial x^4}(t,x) + \mathcal{O}\Big((\Delta x)^4\Big).$$

Si Δx est "petit", la formule (5.2) est une "bonne" approximation (elle est naturelle mais pas unique). La formule (5.2) est dite **centrée** car elle est symétrique en j.

Pour discrétiser l'équation de diffusion (5.1) il faut aussi discrétiser la dérivée en temps. On a le choix entre deux formules principalement.

1. Un premier choix est le **schéma d'Euler implicite** ou rétrograde

$$\frac{u_j^n - u_j^{n-1}}{\Delta t} + \nu \frac{-u_{j-1}^n + 2u_j^n - u_{j+1}^n}{(\Delta x)^2} = 0 \tag{5.3}$$

qui est basé sur la formule

$$\frac{\partial u}{\partial t}(t_n, x_j) \approx \frac{u_j^n - u_j^{n-1}}{\Delta t}.$$

Il faut résoudre un système d'équations linéaires pour calculer les valeurs $(u_j^n)_{1\leq j\leq N}$ en fonctions des valeurs précédentes $(u_j^{n-1})_{1\leq j\leq N}$.

2. Un deuxième choix est le symétrique du précédent : il s'agit du **schéma d'Euler explicite** ou progressif

$$\frac{u_j^{n+1}-u_j^n}{\Delta t}+\nu\frac{-u_{j-1}^n+2u_j^n-u_{j+1}^n}{(\Delta x)^2}=0 \tag{5.4}$$

qui utilise la formule

$$\frac{\partial u}{\partial t}(t_n,x_j)\approx\frac{u_j^{n+1}-u_j^n}{\Delta t}.$$

Toutes ces formules de schémas s'entendent pour les indices j tels que $1\leq j\leq N$. On les complète par les conditions aux limites

$$u_0^n=u_{N+1}^n=0 \text{ pour } n\geq 1.$$

Il y a bien sûr une donnée initiale pour démarrer les itérations en n : les valeurs initiales $(u_j^0)_{0\leq j\leq N+1}$ sont définies, par exemple, par $u_j^0=u_0(j\Delta x)$ où u_0 est la donnée initiale de l'équation de diffusion (5.1).

Remarque 5.1.1 *Bien sûr, si on considère l'équation stationnaire de diffusion dans le domaine borné* $(0,1)$

$$\begin{cases} -\nu\dfrac{\partial^2 u}{\partial x^2}=f \text{ pour } x\in(0,1) \\ u(0)=u(1)=0, \end{cases}$$

pour un second membre donné $f(x)$, un schéma adapté à ce cas est simplement

$$\nu\frac{-u_{j-1}+2u_j-u_{j+1}}{(\Delta x)^2}=f_j \text{ pour } 1\leq j\leq N,$$

avec f_j une approximation de $f(x_j)$ et les conditions aux limites $u_0=u_{N+1}=0$.

5.1.2 Consistance, stabilité et convergence

Pour formaliser l'approximation d'une équation aux dérivées partielles par des différences finies, on introduit la notion de **consistance** et de **précision** d'un schéma. Bien que pour l'instant nous ne considérons que l'équation de diffusion (5.1), nous allons donner une définition de la consistance valable pour n'importe quelle équation aux dérivées partielles que nous notons $F(u)=0$. Remarquons que $F(u)$ est une notation pour une fonction de u et de ses dérivées partielles en tout point (t,x). De manière générale un schéma aux différences finies est caractérisé, pour tous les indices possibles n,j, par la formule

$$F_{\Delta t,\Delta x}\left(\{u_{j+k}^{n+m}\}_{m^-\leq m\leq m^+,\ k^-\leq k\leq k^+}\right)=0 \tag{5.5}$$

où les entiers m^-,m^+,k^-,k^+ définissent la largeur du stencil du schéma.

Définition 5.1.2 *Le schéma aux différence finies (5.5) est dit consistant avec l'équation aux dérivées partielles $F(u) = 0$, si* **l'erreur de troncature** *du schéma, définie pour toute fonction régulière $u(t,x)$ par*

$$F_{\Delta t,\Delta x}\left(\{u(t+m\Delta t, x+k\Delta x)\}_{m^-\le m\le m^+,\ k^-\le k\le k^+}\right),$$

tend vers zéro lorsque Δt et Δx tendent vers zéro indépendemment, **si et seulement si** *$u(t,x)$ est une solution de cette équation.*

De plus, on dit que le schéma est précis à l'ordre p en espace et à l'ordre q en temps si l'erreur de troncature (définie ci-dessus) tend vers zéro comme $\mathcal{O}\Big((\Delta x)^p + (\Delta t)^q\Big)$ lorsque Δt et Δx tendent vers zéro.

Remarque 5.1.3 *Il faut prendre garde dans la formule (5.5) à une petite ambiguïté quant à la définition du schéma. En effet, on peut toujours multiplier n'importe quelle formule par une puissance suffisamment élevée de Δt et Δx de manière à ce que l'erreur de troncature tende vers zéro. Cela rendrait consistant n'importe quel schéma ! Pour éviter cet inconvénient, on demande à ce que l'erreur de troncature ne tende pas vers 0 si $u(t,x)$ n'est pas solution de l'équation.*

Concrètement on calcule l'erreur de troncature d'un schéma en remplaçant u_{j+k}^{n+m} dans la formule (5.5) par $u(t+m\Delta t, x+k\Delta x)$ et en procédant à un développement de Taylor. Comme application de la Définition 5.1.2, nous allons montrer le lemme suivant.

Lemme 5.1.4 *Le schéma explicite (5.4) est consistant, précis à l'ordre 1 en temps et 2 en espace.*

Démonstration. Soit $v(t,x)$ une fonction de classe $\mathcal{C}^6$. Par développement de Taylor autour du point (t,x), on calcule l'erreur de troncature du schéma (5.4)

$$\begin{aligned}\frac{v(t+\Delta t,x)-v(t,x)}{\Delta t} &+\nu\frac{-v(t,x-\Delta x)+2v(t,x)-v(t,x+\Delta x)}{(\Delta x)^2}\\ &=\left(\frac{\partial v}{\partial t}-\nu\frac{\partial^2 v}{\partial x^2}+\frac{\Delta t}{2}\frac{\partial^2 v}{\partial t^2}-\frac{\nu(\Delta x)^2}{12}\frac{\partial^4 v}{\partial x^4}\right)(t,x)\\ &+\mathcal{O}\Big((\Delta t)^2+(\Delta x)^4\Big).\end{aligned}$$

Si v est une solution de l'équation de la chaleur (5.1), on obtient ainsi aisément la consistance ainsi que la précision à l'ordre 1 en temps et 2 en espace. ■

Tous les schémas consistants avec l'équation à résoudre ne sont pas des "bons" schémas. Certains peuvent présenter de violentes oscillations numériques lors de leur application : on dit qu'ils sont instables. Pour éviter cet inconvénient on se restreint aux schémas stables que l'on définit comme suit. Nous introduisons deux normes classiques pour la solution numérique $u^n = (u_j^n)_{1\le j\le N}$:

$$\|u^n\|_2 = \left(\sum_{j=1}^{N}\Delta x|u_j^n|^2\right)^{1/2} \quad\text{et}\quad \|u^n\|_\infty = \max_{1\le j\le N}|u_j^n|.$$

Définition 5.1.5 *Un schéma aux différences finies est dit* **stable** *pour la norme* $\| \ \|_p$, $p = 2, \infty$, *s'il existe une constante* $K > 0$ *indépendante de* Δt *et* Δx *(lorsque ces pas tendent vers zéro) telle que*

$$\|u^n\|_p \leq K \|u^0\|_p \text{ pour tout } n \geq 0,$$

quelle que soit *la donnée initiale* u^0.

Si cette inégalité n'a lieu que pour des pas Δt *et* Δx *astreints à certaines conditions, on dit que le schéma est* **conditionnellement stable**.

La stabilité en norme L^∞ est très liée avec le principe du maximum discret.

Définition 5.1.6 *On dit qu'un schéma aux différences finies vérifie le* **principe du maximum discret** *si pour tout* $n \geq 0$ *et tout* $1 \leq j \leq N$ *on a*

$$\min\left(0, \min_{0 \leq j \leq N+1} u_j^0\right) \leq u_j^n \leq \max\left(0, \max_{0 \leq j \leq N+1} u_j^0\right)$$

quelle que soit la donnée initiale u^0.

Dans la Définition 5.1.6 les inégalités tiennent compte non seulement du minimum et du maximum de u^0 mais aussi de zéro qui est la valeur imposée au bord par les conditions aux limites de Dirichlet. Cela est nécessaire si la donnée initiale u^0 ne vérifie pas les conditions aux limites de Dirichlet (ce qui n'est pas exigé), et inutile dans le cas contraire. Nous suggérons au lecteur de faire l'exercice (facile) suivant pour vérifier que le principe du maximum discret conduit bien à la stabilité en norme L^∞.

Exercice 5.1 *Montrer que le schéma explicite (5.4) est stable en norme* L^∞ *si et seulement si la condition CFL (Courant, Friedrichs, Lewy),* $2\nu\Delta t \leq (\Delta x)^2$, *est satisfaite. Montrer que le schéma implicite (5.3) est inconditionnellement stable en norme* L^∞, *c'est-à-dire quels que soient les pas de temps* Δt *et d'espace* Δx.

De nombreux schémas ne vérifient pas le principe du maximum discret mais sont néanmoins de "bons" schémas. Pour ceux-là, il faut vérifier la stabilité dans une autre norme que la norme L^∞. La norme L^2 se prête très bien à l'étude de la stabilité pour deux raisons disctinctes. D'une part, lorsque les conditions aux limites sont de type périodiques, on peut utiliser l'outil très puissant des **séries de Fourier** que nous allons rappeler brièvement (voir [2] pour plus de détails). D'autre part, quel que soit le type des conditions aux limites, on peut utiliser la notion **d'inégalité d'énergie** (ou estimation a priori) que nous décrirons un peu plus loin. Cette dernière méthode sera utilisée de manière systématique par la suite.

Plutôt que de décrire de manière exhaustive la méthode des séries de Fourier nous allons simplement en livrer l'essence sous la forme d'une "recette" connue sous le nom de **condition nécessaire de stabilité de Von Neumann**. Tout

d'abord, on suppose que les conditions aux limites pour l'équation de diffusion (5.1) sont des **conditions aux limites de périodicité**, qui s'écrivent $u(t, x+1) = u(t, x)$ pour tout $x \in [0, 1]$ et tout $t \geq 0$. Pour les schémas numériques, elles conduisent aux égalités $u_0^n = u_{N+1}^n$ pour tout $n \geq 0$, et plus généralement $u_j^n = u_{N+1+j}^n$ pour tout n, j. L'idée de Von Neumann est de tester si des solutions discrètes particulières d'un schéma sont stables ou non. Ces solutions particulières sont choisies sous la forme d'un mode de Fourier, pour $k \in \mathbf{Z}$,

$$u_j^n = A(k)^n \exp(2i\pi k x_j) \quad \text{avec} \quad x_j = j\Delta x. \tag{5.6}$$

On injecte la formule (5.6) dans le schéma et on en déduit la valeur du facteur d'amplification $A(k) \in \mathbf{C}$. Différentes valeurs du nombre d'onde k correspondent à différentes données initiales u^0. Une fois calculée $A(k)$, la solution particulière (5.6) est stable si et seulement si l'inégalité suivante est vérifiée

$$|A(k)| \leq 1 \text{ pour tout mode } k \in \mathbf{Z}. \tag{5.7}$$

Puisqu'on n'a pas testé **toutes** les solutions discrètes, l'inégalité (5.7) est seulement une condition **nécessaire** de stabilité, dite de Von Neumann.

Remarque 5.1.7 *En fait, dans de nombreux cas, comme celui de l'équation de la diffusion, on peut montrer que la condition de stabilité de Von Neumann est non seulement nécessaire mais aussi suffisante (voir [2]). La base de cette observation est le fait que toute fonction de $L^2(0,1)$ peut se décomposer en une série de Fourier (voir [23]) et que, de même, toute solution discrète d'un schéma numérique est une superposition de modes du type (5.6) pour k parcourant $\mathbf{Z}$.*

Exercice 5.2 *Montrer que le schéma explicite (5.4) vérifie la condition nécessaire de stabilité en norme L^2 de Von Neumann si et seulement si la condition CFL $2\nu\Delta t \leq (\Delta x)^2$ est satisfaite.*
Indication : montrer que $A(k) = 1 - 4\frac{\Delta t}{\Delta x^2}\sin^2 \pi k \Delta x$.

Exercice 5.3 *Montrer que le schéma implicite (5.3) vérifie toujours la condition nécessaire de stabilité en norme L^2 de Von Neumann.*
Indication : montrer que $A(k) = \left(1 + 4\frac{\Delta t}{\Delta x^2}\sin^2 \pi k \Delta x\right)^{-1}$.

Remarque 5.1.8 *La Définition 5.1.5 de stabilité d'un schéma est simple mais parfois restrictive car elle ne s'applique pas aux équations aux dérivées partielles dont la solution croît naturellement en temps. C'est le cas, par exemple, pour la solution de l'équation*

$$\frac{\partial u}{\partial t} - \nu \frac{\partial^2 u}{\partial x^2} = cu \text{ pour } (t, x) \in \mathbf{R}^+ \times \mathbf{R},$$

qui, par le changement d'inconnue $v(t, x) = e^{-ct}u(t, x)$, se ramène à l'équation de la chaleur (donc pour $c > 0$ suffisamment grand, la solution u croît exponentiellement en temps). Aucun schéma, consistant avec cette équation, ne pourrait donc être stable au sens de la Définition 5.1.5 qui implique que la solution reste

bornée quand t tend vers l'infini. C'est pourquoi il existe une autre définition de la stabilité, moins restrictive mais plus complexe. Dans cette définition le schéma est dit stable pour la norme $\| \ \|_p$ si pour tout temps $T > 0$ il existe une constante $K(T) > 0$ indépendante de Δt et Δx telle que

$$\|u^n\|_p \leq K(T)\|u^0\|_p \text{ pour tout } 0 \leq n \leq T/\Delta t,$$

quelle que soit la donnée initiale u^0. Cette nouvelle définition permet à la solution de croître avec le temps puisque la constante $K(T)$ dépend du temps final T. Avec une telle définition de la stabilité, la condition nécessaire de Von Neumann devient l'inégalité

$$|A(k)| \leq 1 + C\Delta t \text{ pour tout mode } k \in \mathbf{Z}.$$

Par souci de simplicité nous préférons nous en tenir à la Définition 5.1.5 de la stabilité.

Venons-en maintenant à la méthode d'inégalité d'énergie et commençons par démontrer un résultat sur l'équation de diffusion (5.1) avec ses conditions aux limites de Dirichlet (tout autre type "raisonnable" de conditions aux limites conviendrait aussi).

Lemme 5.1.9 *Soit $u(t,x)$ une solution régulière de (5.1). Alors elle vérifie l'inégalité, dite d'énergie, pour tout $t > 0$,*

$$\int_0^1 |u(t,x)|^2 dx \leq \int_0^1 |u_0(x)|^2 dx.$$

Remarque 5.1.10 *Le mot "énergie" est à prendre ici au sens mathématique d'une quantité intégrale ayant des propriétés de stabilité (un peu comme une fonction de Lyapunov pour les systèmes dynamiques). Elle peut parfois correspondre à l'énergie physique (notamment dans les applications mécaniques) mais pas toujours. Par exemple, dans le cadre du modèle de diffusion la quantité $\int_0^1 |u(t,x)|^2 dx$ n'est pas interprétable comme une énergie physique quelconque.*

Démonstration. On multiplie l'équation (5.1) par u et on intègre par parties en espace pour obtenir

$$\int_0^1 u\frac{\partial u}{\partial t} dx + \nu \int_0^1 \left(\frac{\partial u}{\partial x}\right)^2 dx - \nu \left(u\frac{\partial u}{\partial x}(t,1) - u\frac{\partial u}{\partial x}(t,0)\right) = 0.$$

Les termes de bord s'annulent à cause des conditions aux limites et, en intégrant en temps, on obtient

$$\frac{1}{2}\int_0^1 |u(t,x)|^2 dx - \frac{1}{2}\int_0^1 |u(0,x)|^2 dx + \nu \int_0^t \int_0^1 \left(\frac{\partial u}{\partial x}(s,x)\right)^2 dx\, ds = 0$$

d'où l'on déduit le résultat en minorant par zéro la dernière intégrale. Remarquons que l'hypothèse d'avoir une solution régulière de (5.1) est automatiquement satisfaite dès que la donnée initiale u_0 est suffisamment régulière. ■

Nous allons voir que le même type d'inégalité d'énergie que le Lemme 5.1.9 peut s'obtenir pour le schéma implicite (5.3) en suivant le même raisonnement et en remplaçant les intégrations par parties par des réarrangements de sommes.

Lemme 5.1.11 *Soit $u^n = (u_j^n)_{1\leq j\leq N}$ la solution discrète du schéma implicite (5.3). Alors elle vérifie l'inégalité*

$$\|u^n\|_2 \leq \|u^0\|_2.$$

Autrement dit, le schéma implicite (5.3) est inconditionnellement stable.

Démonstration. On multiplie la formule du schéma implicite (5.3) par u_j^n et on somme en j (équivalent de l'intégration en espace) pour obtenir

$$\Delta x \sum_{j=1}^{N} u_j^n (u_j^n - u_j^{n-1}) + \Delta t \sum_{j=1}^{N} u_j^n \Big((u_j^n - u_{j+1}^n) - (u_{j-1}^n - u_j^n) \Big) = 0.$$

On réarrange la dernière somme (équivalent d'une intégration par parties), ce qui conduit à

$$\Delta x \sum_{j=1}^{N} u_j^n (u_j^n - u_j^{n-1}) + \Delta t \sum_{j=1}^{N} u_j^n (u_j^n - u_{j+1}^n) - \Delta t \sum_{j=0}^{N-1} u_{j+1}^n (u_j^n - u_{j+1}^n) = 0,$$

et, en utilisant la condition aux limites de Dirichlet,

$$\Delta x \sum_{j=1}^{N} u_j^n (u_j^n - u_j^{n-1}) + \Delta t \sum_{j=0}^{N} (u_j^n - u_{j+1}^n)^2 = 0.$$

On en déduit

$$\Delta x \sum_{j=1}^{N} (u_j^n)^2 \leq \Delta x \sum_{j=1}^{N} u_j^n u_j^{n-1}$$

et par Cauchy-Schwarz $\|u^n\|_2 \leq \|u^{n-1}\|_2$, d'où le résultat. ■

Nous avons maintenant tous les outils pour démontrer la convergence des schémas de différences finies. Le Théorème de Lax, ci-dessous, affirme que, pour un schéma linéaire, **consistance et stabilité impliquent convergence**.

Rappelons que l'on dit qu'un schéma aux différences finies est **linéaire** si la formule $F_{\Delta t,\Delta x}(\{u_{j+k}^{n+m}\}) = 0$ qui le définit est linéaire par rapport à ses arguments u_{j+k}^{n+m}, et qu'il est **à deux niveaux** s'il ne fait intervenir que deux indices de temps.

Théorème 5.1.12 (Lax) *Soit $u(t,x)$ la solution régulière de l'équation de diffusion (5.1). Soit u_j^n la solution numérique discrète obtenue par un schéma de différences finies avec la donnée initiale $u_j^0 = u_0(x_j)$. On suppose que le schéma*

est linéaire, à deux niveaux, consistant et stable pour une norme $\|\ \|$. Alors le schéma est convergent au sens où

$$\forall\, T>0, \quad \lim_{\Delta t,\Delta x\to 0}\left(\sup_{t_n\le T}\|e^n\|\right)=0,$$

avec e^n le vecteur "erreur" défini par ses composantes $e_j^n=u_j^n-u(t_n,x_j)$.

De plus, si le schéma est précis à l'ordre p en espace et à l'ordre q en temps, alors pour tout temps $T>0$ il existe une constante $C_T>0$ telle que

$$\sup_{t_n\le T}\|e^n\|\le C_T\Big((\Delta x)^p+(\Delta t)^q\Big). \tag{5.8}$$

Démonstration. On démontre seulement l'inégalité (5.8) : la preuve de la simple convergence vers zéro de l'erreur est similaire. Un schéma linéaire à deux niveaux peut s'écrire sous la forme condensée

$$u^{n+1}=Au^n, \tag{5.9}$$

où A est la matrice d'itération (carrée de taille N). On note $\tilde u^n=(\tilde u_j^n)_{1\le j\le N}$ avec $\tilde u_j^n=u(t_n,x_j)$ où u est la solution de (5.1). Comme le schéma est consistant, il existe un vecteur ϵ^n tel que

$$\tilde u^{n+1}=A\tilde u^n+\Delta t\epsilon^n \text{ avec } \lim_{\Delta t,\Delta x\to 0}\left(\sup_{t_n\le T}\|\epsilon^n\|\right)=0. \tag{5.10}$$

Si le schéma est précis à l'ordre p en espace et à l'ordre q en temps, alors $\|\epsilon^n\|\le C((\Delta x)^p+(\Delta t)^q)$. En posant $e_j^n=u_j^n-u(t_n,x_j)$ on obtient par soustraction de (5.10) à (5.9)

$$e^{n+1}=Ae^n-\Delta t\epsilon^n$$

d'où par récurrence

$$e^n=A^ne^0-\Delta t\sum_{k=1}^{n}A^{n-k}\epsilon^{k-1}. \tag{5.11}$$

Or, la stabilité du schéma veut dire que $\|u^n\|=\|A^nu^0\|\le K\|u^0\|$ pour toute donnée initiale, c'est-à-dire que $\|A^n\|\le K$ où la constante K ne dépend pas de n. D'autre part, $e^0=0$, donc (5.11) donne

$$\|e^n\|\le\Delta t\sum_{k=1}^{n}\|A^{n-k}\|\|\epsilon^{k-1}\|\le\Delta tnKC\Big((\Delta x)^p+(\Delta t)^q\Big),$$

ce qui donne l'inégalité (5.8) avec la constante $C_T=TKC$. ∎

Remarque 5.1.13 *Le Théorème de Lax 5.1.12 est en fait valable pour toute équation aux dérivées partielles linéaire. Remarquer que la vitesse de convergence dans (5.8) est exactement la précision du schéma.*

5.1.3 Équation de transport

Nous considérons désormais l'équation de transport (ou équation d'advection) en une dimension d'espace dans le domaine borné $(0,1)$ avec une vitesse constante $V > 0$ et une condition aux limites "d'entrée" de type Dirichlet

$$\begin{cases} \dfrac{\partial u}{\partial t} + V\dfrac{\partial u}{\partial x} = 0 \text{ pour } (x,t) \in (0,1) \times \mathbf{R}_*^+ \\ u(t, x=0) = g(t) \text{ pour } t \in \mathbf{R}_*^+ \\ u(t=0,x) = u_0(x) \text{ pour } x \in (0,1), \end{cases} \tag{5.12}$$

où u_0 est la donnée initiale et g la donnée au bord entrant. Si on étend $u_0(x)$ par 0 en dehors de l'intervalle $(0,1)$ et $g(t)$ par 0 pour $t < 0$, la solution exacte de (5.12) est (le vérifier !)

$$u(t,x) = u_0(x - Vt) + g(t - \frac{x}{V}).$$

Bien évidemment, si l'on choisit une vitesse de signe opposée $V < 0$, alors la condition aux limites d'entrée doit être imposée en $x = 1$, et non plus en $x = 0$, autrement dit "l'entrée" est toujours en amont de la vitesse. On garde les mêmes notations pour la discrétisation : $\Delta x = 1/(N+1) > 0$ (N entier positif), $\Delta t > 0$, $(t_n, x_j) = (n\Delta t, j\Delta x)$ pour $n \geq 0, j \in \{0, 1, ..., N+1\}$, et u_j^n la valeur d'une solution discrète approchée au point (t_n, x_j). Comme d'habitude on choisit la donnée initiale discrète sous la forme

$$u_j^0 = u_0(x_j).$$

La condition aux limites se traduit par l'égalité $u_0^n = g(t_n)$ pour tout $n \geq 0$. Il n'y a pas de conditions aux limites en "sortie" mais certains schémas nécessitent de se donner une valeur u_{N+1}^n (qu'on ne peut pas calculer autrement). Dans ce cas on utilise une condition aux limites "numérique", dite transparente ou de sortie, du type Neumann, c'est-à-dire qu'on impose $u_{N+1}^n = u_N^n$. Par conséquent, l'inconnue discrète à chaque pas de temps est un vecteur $u^n = (u_j^n)_{1 \leq j \leq N} \in \mathbf{R}^N$.

Un des schémas les plus simples (mais pas très précis) pour l'équation de transport (5.12) est le **schéma de Lax-Friedrichs**

$$\frac{2u_j^{n+1} - u_{j+1}^n - u_{j-1}^n}{2\Delta t} + V\frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} = 0.$$

Lemme 5.1.14 *Le schéma de Lax-Friedrichs est stable en norme L^∞ sous la condition CFL*

$$|V|\Delta t \leq \Delta x.$$

Si le rapport $\Delta t/\Delta x$ est gardé constant lorsque Δt et Δx tendent vers zéro, il est consistant avec l'équation de transport (5.12) et précis à l'ordre 1 en espace et en temps. Par conséquent, il est conditionnellement convergent.

Démonstration. On réécrit le schéma de Lax-Friedrichs sous la forme

$$u_j^{n+1} = \frac{1}{2}\left(1 - \frac{V\Delta t}{\Delta x}\right) u_{j+1}^n + \frac{1}{2}\left(1 + \frac{V\Delta t}{\Delta x}\right) u_{j-1}^n$$

qui est une combinaison convexe des valeurs au temps t_n si la condition CFL $|V|\Delta t \leq \Delta x$ est satisfaite. Le schéma vérifie le principe du maximum discret et est donc conditionnellement stable en norme L^∞. Pour étudier la consistance, on calcule l'erreur de troncature en effectuant un développement de Taylor autour de (t_n, x_j) pour une fonction régulière u :

$$\begin{aligned}\frac{2u(t_{n+1},x_j) - u(t_n,x_{j+1}) - u(t_n,x_{j-1})}{2\Delta t} + V\frac{u(t_n,x_{j+1}) - u(t_n,x_{j-1})}{2\Delta x} \;=\\ \left(\frac{\partial u}{\partial t} + V\frac{\partial u}{\partial x}\right)(t_n,x_j) - \frac{(\Delta x)^2}{2\Delta t}\left(1 - \frac{(V\Delta t)^2}{(\Delta x)^2}\right)\frac{\partial^2 u}{\partial x^2}(t_n,x_j)\\ + \mathcal{O}\Big((\Delta x)^2 + \frac{(\Delta x)^4}{\Delta t}\Big).\end{aligned}$$

Comme l'erreur de troncature contient un terme en $\mathcal{O}\Big((\Delta x)^2/\Delta t\Big)$, le schéma n'est pas consistant si Δt tend vers zéro plus vite que $(\Delta x)^2$. Par contre, il est consistant et précis d'ordre 1 si le rapport $\Delta t/\Delta x$ est constant. Pour obtenir la convergence on reprend la démonstration du Théorème de Lax 5.1.12. L'erreur e^n est toujours majorée par l'erreur de troncature, et donc ici

$$\|e^n\| \leq \Delta t n K C\Big(\frac{(\Delta x)^2}{\Delta t} + \Delta t\Big).$$

Si on garde fixe le rapport $\Delta x/\Delta t$ l'erreur est donc majoré par une constante fois Δt qui tend bien vers zéro, d'où la convergence. ■

Le schéma de Lax-Friedrichs est en fait une amélioration d'un schéma beaucoup plus simple et "naturel" mais qui est inutilisable en pratique car instable ! Il s'agit du **schéma explicite centré**

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + V\frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} = 0$$

qui est désastreux en théorie comme en pratique...

Exercice 5.4 *Montrer que le schéma explicite centré est consistant avec l'équation de transport (5.12), précis à l'ordre 1 en temps et 2 en espace, mais inconditionnellement instable en norme L^2.*
Indication : on supposera que les conditions aux limites sont périodiques et on utilisera la condition de stabilité de Von Neumann.

Une idée fondamentale pour obtenir de "bons" schémas pour l'équation de transport (5.12) est le **décentrement amont**. Nous donnons la forme générale

du **schéma décentré amont**

$$\begin{aligned} &\frac{u_j^{n+1}-u_j^n}{\Delta t}+V\frac{u_j^n-u_{j-1}^n}{\Delta x}=0 \quad \text{si} \quad V>0, \\ &\frac{u_j^{n+1}-u_j^n}{\Delta t}+V\frac{u_{j+1}^n-u_j^n}{\Delta x}=0 \quad \text{si} \quad V<0. \end{aligned} \tag{5.13}$$

L'idée principale du décentrement amont est d'aller chercher "l'information" en remontant le courant. Grâce à cette astuce (ou plutôt cette intuition physique) on obtient un schéma stable qui vérifie le principe du maximun discret. Malheureusement le schéma décentré amont n'est précis qu'à l'ordre 1.

Exercice 5.5 *Montrer que le schéma explicite décentré amont (5.13) est consistant avec l'équation de transport (5.12), précis à l'ordre 1 en espace et en temps, stable en norme L^∞ et convergent si la condition CFL $|V|\Delta t \leq \Delta x$ est satisfaite.*
Indication : en supposant $V>0$, récrire le schéma sous la forme explicite $u_j^{n+1} + \alpha u_j^n + \beta u_{j-1}^n$ avec $\alpha, \beta \geq 0$ et $\alpha+\beta=1$, puis conclure.

Exercice 5.6 *Montrer que, sous condition CFL, le schéma explicite décentré amont (5.13) vérifie la condition nécessaire de stabilité en norme L^2 de Von Neumann.*
Indication : montrer que $A(k) = 1 - 2\frac{V\Delta t}{\Delta x}\left(1-\frac{V\Delta t}{\Delta x}\right)\sin^2 k\pi\Delta x$.

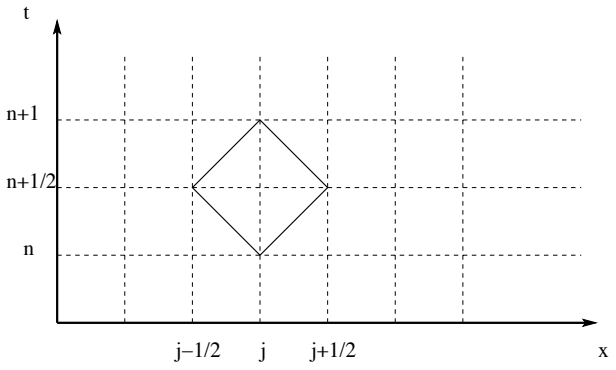

FIGURE 5.2 – Schéma diamant ou en croix.

Finalement nous présentons le **schéma diamant** (ou en croix), dû à Carlson [14], qui est un schéma centré ayant l'avantage d'être précis à l'ordre 2 et inconditionnellement stable, et qui se généralise facilement pour l'équation de Boltzmann linéaire (voir la Section 5.2). Pour cette dernière équation le schéma diamant est "le" schéma de référence, extrêmement populaire dans les applications industrielles. Ce schéma est un peu plus compliqué que les précédents car il utilise les inconnues intermédiaires $u_{j+1/2}^{n+1/2}$ qui sont des approximations de la

solution au temps $t_{n+1/2} = (n+1/2)\Delta t$ et au point $x_{j+1/2} = (j+1/2)\Delta x$. Il s'écrit

$$\begin{cases} \dfrac{u_j^{n+1} - u_j^n}{\Delta t} + V \dfrac{u_{j+1/2}^{n+1/2} - u_{j-1/2}^{n+1/2}}{\Delta x} = 0, \\ u_j^{n+1} + u_j^n = u_{j+1/2}^{n+1/2} + u_{j-1/2}^{n+1/2}. \end{cases} \tag{5.14}$$

La première ligne de (5.14) discrétise de manière centrée l'équation de transport (5.12) tandis que la deuxième ligne est purement algébrique. Cette deuxième relation de (5.14) est dite "diamant" à cause de la figure obtenue en reliant les points $(u_j^{n+1}, u_j^n, u_{j+1/2}^{n+1/2}, u_{j-1/2}^{n+1/2})$ sur un maillage espace-temps (voir la Figure 5.2). Les relations (5.14) sont valables pour les indices $1 \leq j \leq N$ et on les complète par la conditions aux limites d'entrée

$$u_0^n = g(t_n) \quad \text{pour } n \geq 0. \tag{5.15}$$

Pour calculer la solution de (5.14), on élimine l'inconnue $u_j^{n+1} = u_{j+1/2}^{n+1/2} + u_{j-1/2}^{n+1/2} - u_j^n$ pour se ramener au schéma, a priori implicite,

$$u_{j+1/2}^{n+1/2} \left(1 + \frac{V\Delta t}{\Delta x}\right) + u_{j-1/2}^{n+1/2} \left(1 - \frac{V\Delta t}{\Delta x}\right) = 2u_j^n, \tag{5.16}$$

qui permet de calculer les valeurs $(u_{j+1/2}^{n+1/2})_j$ en fonction des valeurs précédentes $(u_j^n)_j$. On calcule ensuite facilement les valeurs $(u_j^{n+1})_j$ avec la relation diamant, deuxième ligne de (5.14). La relation (5.16) est valable pour les indices $0 \leq j \leq N$ et on la complète par la condition aux limites d'entrée (obtenue en injectant (5.15) dans la relation diamant)

$$u_{-1/2}^{n+1/2} = -u_{1/2}^{n+1/2} + g(t_n) + g(t_{n+1}),$$

qui conduit à

$$u_{1/2}^{n+1/2} = \frac{g(t_n)\left(1 + \frac{V\Delta t}{\Delta x}\right) - g(t_{n+1})\left(1 - \frac{V\Delta t}{\Delta x}\right)}{2\frac{V\Delta t}{\Delta x}}. \tag{5.17}$$

Bien que le schéma (5.16) semble être implicite, il n'en est rien car il est possible de calculer de proche en proche les valeurs $u_{j+1/2}^{n+1/2}$ en allant dans le sens des j croissants (car la vitesse est positive $V > 0$; c'est l'équivalent discret de la méthode des caractéristiques). Autrement dit, la résolution du système linéaire associé à (5.16) est immédiate car la matrice correspondante A est triangulaire inférieure (et inversible puisque $1 + c > 0$)

$$A = \begin{pmatrix} 1+c & 0 & & & 0 \\ 1-c & 1+c & & 0 & \\ & \ddots & \ddots & \ddots & \\ & & 1-c & 1+c & 0 \\ 0 & & & 1-c & 1+c \end{pmatrix} \quad \text{avec } c = \frac{V\Delta t}{\Delta x}.$$

Autrement dit, le schéma diamant (5.14) est quasiment explicite alors qu'il va hériter des propriétés usuelles des schémas implicites (stabilité inconditionnelle). Pour montrer la stabilité de ce schéma nous allons utiliser la méthode d'inégalité d'énergie et pour commencer nous établissons cette inégalité pour l'équation de transport (5.12) (comme dans la Sous-section 2.5).

Lemme 5.1.15 *Toute solution régulière de (5.12) vérifie*

$$\int_0^1 |u(T,x)|^2 dx \leq \int_0^1 |u_0(x)|^2 dx + \int_0^T V|g(t)|^2\, dt.$$

Démonstration. On multiplie (5.12) par u et on intègre en espace

$$\frac{1}{2}\frac{d}{dt}\left(\int_0^1 |u(t,x)|^2 dx\right) + \frac{V}{2}\left(|u(t,1)|^2 - |u(t,0)|^2\right) = 0.$$

On remplace $u(t,0)$ par la condition aux limites $g(t)$, on minore par zéro le terme de bord en $x=1$ et on intègre en temps pour obtenir

$$\int_0^1 |u(T,x)|^2 dx - \int_0^1 |u(0,x)|^2 dx - V\int_0^T |g(t)|^2\, dt \leq 0$$

qui est l'inégalité d'énergie recherchée. ■

Lemme 5.1.16 *Le schéma diamant (5.14) est inconditionnellement stable en norme L^2.*

Démonstration. On va reproduire l'argument de la preuve du Lemme 5.1.15 mais adapté au cas discret, c'est-à-dire en remplaçant les intégrations par des réarrangements de sommes discrètes. Pour cela on multiplie la première équation de (5.14) par $(u_j^{n+1} + u_j^n)$ et on utilise la deuxième équation de (5.14) pour le deuxième terme, ce qui donne

$$\frac{(u_j^{n+1})^2 - (u_j^n)^2}{\Delta t} + V\frac{(u_{j+1/2}^{n+1/2})^2 - (u_{j-1/2}^{n+1/2})^2}{\Delta x} = 0.$$

En sommant sur j les termes en $(j+1/2)$ s'éliminent deux à deux (sauf les extrêmes pour lesquels on utilise la condition aux limites d'entrée) et on obtient

$$\sum_{j=1}^N \Delta x (u_j^{n+1})^2 - \sum_{j=1}^N \Delta x (u_j^n)^2 - V\Delta t (u_{1/2}^{n+1/2})^2 \leq 0.$$

On somme alors en n pour en déduire

$$\sum_{j=1}^N \Delta x (u_j^{n+1})^2 \leq \sum_{j=1}^N \Delta x (u_j^0)^2 + V\sum_{k=0}^n \Delta t (u_{1/2}^{k+1/2})^2. \tag{5.18}$$

La condition aux limites d'entrée (5.17) nous dit que

$$\begin{aligned} u_{1/2}^{n+1/2} &= \frac{g(t_n)+g(t_{n+1})}{2} + \frac{g(t_n)-g(t_{n+1})}{2\frac{V\Delta t}{\Delta x}} \\ &= g(t_{n+1/2}) - \frac{\Delta x}{2V} g'(t_{n+1/2}) + \mathcal{O}\Big((\Delta t)^2 + (\Delta t)^2 \Delta x\Big), \end{aligned} \tag{5.19}$$

ce qui implique que la dernière somme dans (5.18) est une approximation consistante de $\int_0^T |g(t)|^2\, dt$ et donc que (5.18) est bien une inégalité de stabilité en norme L^2 valable pour tout choix des pas de temps et d'espace. ■

Exercice 5.7 *Montrer que le schéma diamant (5.14) (avec conditions aux limites de périodicité) vérifie la condition nécessaire de stabilité en norme L^2 de Von Neumann.*
Indication : incorporer une donnée de type Fourier $u_j^n = A(k)^n exp(2\pi k x_j)$ directement dans (5.18), et en déduire une majoration sur $|A(k)|$.

Lemme 5.1.17 *Le schéma diamant (5.14) est consistant avec l'équation de transport (5.12), précis à l'ordre 2 en espace et en temps.*

Démonstration. On calcule les erreurs de troncature des deux relations du schéma

$$E_1 = \frac{u(t_{n+1},x_j) - u(t_n,x_j)}{\Delta t} + V \frac{u(t_{n+1/2},x_{j+1/2}) - u(t_{n+1/2},x_{j-1/2})}{\Delta x}$$

et

$$E_2 = u(t_{n+1},x_j) + u(t_n,x_j) - u(t_{n+1/2},x_{j+1/2}) - u(t_{n+1/2},x_{j-1/2})$$

en faisant un développement de Taylor autour du point $(t_{n+1/2}, x_j)$. On obtient

$$E_1 = \left(\frac{\partial u}{\partial t} + V\frac{\partial u}{\partial x}\right)(t_{n+1/2},x_j) + \mathcal{O}\left((\Delta t)^2 + (\Delta x)^2\right)$$

et

$$E_2 = \mathcal{O}\left((\Delta t)^2 + (\Delta x)^2\right),$$

ce qui prouve que le schéma est consistant et d'ordre 2. ■

Puisque le schéma diamant (5.14) est stable et consistant, il est automatiquement convergent par application du Théorème de Lax 5.1.12. Néanmoins nous allons redémontrer ce résultat (en suivant exactement la même méthode !) car le lecteur attentif aura remarqué que la définition du schéma diamant n'est pas standard puisqu'elle comprend deux relations dont une (la deuxième) est purement algébrique. Stricto sensu, la démonstration que nous avons donnée du Théorème de Lax ne s'applique pas au schéma diamant qui est un schéma à trois niveaux $(n,\, n+1/2,\, n+1)$. Par ailleurs, la définition de la stabilité du schéma dans le Lemme 5.1.16 tient compte, non seulement de la donnée initiale, mais aussi de la condition aux limites. De même, dans la démonstration du Lemme 5.1.17 sur la précision du schéma on peut se demander s'il ne fallait pas diviser la deuxième relation de (5.14) par Δt ou Δx. Pour clarifier ces points nous allons donc démontrer le résultat suivant.

Lemme 5.1.18 *Le schéma diamant (5.14) est convergent en norme L^2.*

Démonstration. Comme dans la démonstration du Théorème de Lax 5.1.12 nous introduisons la notation $\tilde{u}_j^n = u(t_n, x_j)$ où u est la solution exacte de (5.12) et on définit l'erreur $e_j^n = u_j^n - u(t_n, x_j)$. Comme le schéma est précis à l'ordre 2, on a

$$\left\{ \begin{array}{l} \dfrac{\tilde{u}_j^{n+1} - \tilde{u}_j^n}{\Delta t} + V \dfrac{\tilde{u}_{j+1/2}^{n+1/2} - \tilde{u}_{j-1/2}^{n+1/2}}{\Delta x} = \mathcal{O}\left((\Delta t)^2 + (\Delta x)^2\right), \\ \tilde{u}_j^{n+1} + \tilde{u}_j^n = \tilde{u}_{j+1/2}^{n+1/2} + \tilde{u}_{j-1/2}^{n+1/2} + \mathcal{O}\left((\Delta t)^2 + (\Delta x)^2\right). \end{array} \right. \tag{5.20}$$

Par soustraction avec (5.14) on en déduit

$$\left\{ \begin{array}{l} \dfrac{e_j^{n+1} - e_j^n}{\Delta t} + V \dfrac{e_{j+1/2}^{n+1/2} - e_{j-1/2}^{n+1/2}}{\Delta x} = \epsilon_j^n, \\ e_j^{n+1} + e_j^n = e_{j+1/2}^{n+1/2} + e_{j-1/2}^{n+1/2} + \eta_j^n, \end{array} \right. \tag{5.21}$$

avec

$$|\epsilon_j^n| \leq \delta \quad \text{et} \quad |\eta_j^n| \leq \delta \quad \text{où} \quad \delta = \mathcal{O}\left((\Delta t)^2 + (\Delta x)^2\right). \tag{5.22}$$

Nous suivons alors la démonstration du Lemme 5.1.16 sur la stabilité, c'est-à-dire que nous multiplions la première relation de (5.21) par $e_{j+1/2}^{n+1/2} + e_{j-1/2}^{n+1/2}$, que nous sommons en j et que nous utilisons dans la deuxième relation pour obtenir

$$\begin{array}{l} \Delta x \displaystyle\sum_{j=1}^{N} \left((e_j^{n+1})^2 - (e_j^n)^2\right) + V \Delta t \left((e_{N+1/2}^{n+1/2})^2 - (e_{1/2}^{n+1/2})^2\right) = \\ \Delta x \displaystyle\sum_{j=1}^{N} \left(\Delta t \epsilon_j^n (e_j^{n+1} + e_j^n) + \eta_j^n (e_j^{n+1} - e_j^n) - \Delta t \epsilon_j^n \eta_j^n\right). \end{array} \tag{5.23}$$

Puis on somme en n, en se rappelant que l'erreur initiale est nulle, $e^0 = 0$, en introduisant la norme L^2 discrète et en minorant par zéro le terme $(e_{N+1/2}^{n+1/2})^2$ pour obtenir

$$\begin{array}{rl} \|e^{n_0}\|_2^2 \leq & V \Delta t \displaystyle\sum_{n=0}^{n_0-1} (e_{1/2}^{n+1/2})^2 \\ & + \Delta x \displaystyle\sum_{j=1}^{N} \sum_{n=0}^{n_0-1} \left(\Delta t \epsilon_j^n (e_j^{n+1} + e_j^n) + \eta_j^n (e_j^{n+1} - e_j^n) - \Delta t \epsilon_j^n \eta_j^n\right). \end{array} \tag{5.24}$$

On réarrange la somme suivante

$$\sum_{n=0}^{n_0-1} \eta_j^n (e_j^{n+1} - e_j^n) = \sum_{n=0}^{n_0} e_j^n (\eta_j^{n-1} - \eta_j^n),$$

puis on majore tous les termes du membre de droite de (5.24). Remarquons que η_j^n est une erreur de troncature et que, si on avait poursuivi le développement de Taylor dans la démonstration du Lemme 5.1.17, on aurait aisément vu que

$$\eta_j^{n-1} - \eta_j^n = \Delta t\, \mathcal{O}\left((\Delta t)^2 + (\Delta x)^2\right).$$

En utilisant la condition aux limites d'entrée (5.19), il est facile de voir par un développement de Taylor que l'on a

$$|e_{1/2}^{n+1/2}| \leq \delta \quad \text{où} \quad \delta = \mathcal{O}\left((\Delta t)^2 + (\Delta x)^2\right).$$

Par conséquent, on en déduit

$$\|e^{n_0}\|_2^2 \leq V n_0 \Delta t\, \delta^2 + \Delta t \Delta x \sum_{j=1}^{N} \sum_{n=0}^{n_0-1} \left(\delta(|e_j^{n+1}| + |e_j^n|) + \delta|e_j^n| + \delta^2\right). \quad (5.25)$$

Pour un temps final $T = n_T \Delta t$, on choisit n_0 qui donne l'erreur maximale, $\|e^{n_0}\|_2 = \max_{1\leq n\leq n_T} \|e^n\|_2$, ce qui permet de majorer encore (5.25)

$$\|e^{n_0}\|_2^2 \leq (V+1) T\, \delta^2 + 3T\delta \Delta x \sum_{j=1}^{N} |e_j^{n_0}|. \quad (5.26)$$

Comme $\Delta x \sum_{j=1}^{N} \leq 1$, par Cauchy-Schwarz on a

$$\Delta x \sum_{j=1}^{N} |e_j^{n_0}| \leq \|e^{n_0}\|_2,$$

et (5.26) conduit à

$$\left(\delta^{-1}\|e^{n_0}\|_2\right)^2 \leq C\left(\delta^{-1}\|e^{n_0}\|_2 + 1\right),$$

d'où l'on déduit l'existence d'une constante C, indépendante de Δt et Δx, telle que $\|e^{n_0}\|_2 = \max_{1\leq n\leq n_T} \|e^n\|_2 \leq C\delta$. Le schéma diamant est donc convergent et sa vitesse de convergence est bien d'ordre 2. ■

Remarque 5.1.19 *Le seul inconvénient du schéma diamant (5.14) est qu'il n'est pas positif en général, c'est-à-dire qu'il peut produire des valeurs négatives de la solution même si la donnée initiale (et éventuellement la condition aux limites) est positive. C'est bien sûr contraire au principe du maximum pour l'équation de transport (5.12).*

Remarque 5.1.20 (A propos des conditions aux limites) *Afin de rester consistant avec le début de cette section, nous avons présenté le schéma diamant en discrétisant le domaine spatial* $(0,1)$ *à l'aide des points* $x_j = j\Delta x$ *qui, en particulier, redonnent les deux extrémités de l'intervalle,* $x_0 = 0$ *et* $x_{N+1} = 1$*. Les points supplémentaires* $x_{j+1/2} = (j+1/2)\Delta x$ *sont définis ultérieurement comme*

les points milieux des mailles (x_j, x_{j+1}). Cette façon de faire n'est pas très commode pour définir les conditions aux limites du schéma numérique. En effet, puisque les bords de l'intervalle sont des points entiers, x_0 et x_{N+1}, la condition aux limites d'entrée (5.15) porte sur la valeur discrète en x_0, mais il faut la transférer à l'inconnue sur le point milieu $x_{1/2}$ (voir (5.17)) puisque le calcul effectif du schéma diamant porte sur les valeurs $u_{j+1/2}^{n+1/2}$ plutôt que sur les valeurs u_j^n. C'est un peu compliqué mais encore faisable sur l'exemple que nous venons d'étudier mais cela devient vite inextricable pour l'équation de Boltzmann. C'est pourquoi, dans ce cas, l'usage est de changer les rôles des points entiers x_j et des points milieux $x_{j+1/2}$. Autrement dit, on discrétise le domaine spatial $(0,1)$ par des points $x_{j+1/2} = j\Delta x$, pour $0 \leq j \leq N+1$, avec $\Delta x = 1/(N+1)$, et on considère les points x_j comme les milieux des segments $(x_{j-1/2}, x_{j+1/2})$. On garde alors les mêmes formules (5.14) pour le schéma diamant, mais la condition aux limites d'entrée est beaucoup plus simple : $u_{1/2}^n = g(t_n)$. Cette approche est parfois appelée méthode des volumes finis. C'est ce que nous allons faire dans la suite mais avec une notation différente du domaine spatial afin d'éviter d'éventuelles confusions...

5.2 Différences finies pour l'équation de Boltzmann linéaire

5.2.1 Le cas stationnaire sans collisions

Nous commençons par un cas particulièrement simple qui revient peu ou prou à considérer l'équation précédente de transport dans le cas stationnaire. On étudie donc l'équation de Boltzmann linéaire stationnaire, sans collisions, en géométrie de plaque infinie ("slab" en anglais) dans le cas mono-groupe isotrope, voir (1.5),

$$\left\{ \begin{array}{l} \mu \dfrac{\partial u}{\partial x}(x,\mu) + \sigma(x) u(x,\mu) = f(x,\mu) \\ \qquad\qquad\qquad\qquad \text{pour } (x,\mu) \in (-\ell,+\ell) \times (-1,+1) \\ \\ u(-\ell,\mu) = 0 \text{ pour } \mu > 0,\ u(+\ell,\mu) = 0 \text{ pour } \mu < 0, \end{array} \right. \tag{5.27}$$

où $\sigma(x) \geq 0$ est la section efficace d'absorption, $f(x,\mu)$ est un terme source, $2\ell > 0$ est la largeur du domaine géométrique et les conditions aux limites sont de type Dirichlet ou vide (pas de particules rentrantes). Il est, bien sûr, très facile de calculer la solution exacte du problème aux limites (5.27) par la méthode des caractéristiques. Néanmoins, nous allons décrire une méthode de différences finies en espace et vitesse, appelée aussi **méthode S_N ou des ordonnées discrètes** qui sera la brique de base des schémas numériques pour des modèles de transport plus complets. Les points du maillage sont équirépartis en x,

$$x_{j+1/2} = -\ell + j\Delta x \text{ pour } j \in \{0,1,...,N\} \text{ et } \Delta x = \frac{2\ell}{N}, \tag{5.28}$$

mais pas forcément en vitesse : (μ_k) est une famille finie de vitesses discrétisant l'intervalle $[-1,+1]$. Pour l'instant, la seule propriété que l'on exige de cette famille est qu'elle évite la vitesse nulle, pour une raison qui sera claire dans quelques lignes. Autrement dit, on suppose que $\mu_k \neq 0$ pour tout indice k.

Le schéma le plus répandu est le **schéma diamant** qui est ici la version stationnaire du schéma introduit à la section précédente et qui utilise les points milieux définis par

$$x_j = -\ell + (j-1/2)\Delta x \text{ pour } j \in \{1,2,...,N\}.$$

Pour tout indice de vitesse k on notera u_j^k une approximation de $u(x_j,\mu_k)$, pour $j \in \{1,...,N\}$, et $u_{j+1/2}^k$ une approximation de $u(x_{j+1/2},\mu_k)$, pour $j \in \{0,...,N\}$. Pour $j \in \{1,...,N\}$ et tous les indices k, le schéma diamant est donné par

$$\mu_k \frac{u_{j+1/2}^k - u_{j-1/2}^k}{\Delta x} + \sigma_j\, u_j^k = f_j^k \tag{5.29}$$

et la formule "diamant"

$$u_j^k = \frac{u_{j+1/2}^k + u_{j-1/2}^k}{2}. \tag{5.30}$$

Comme d'habitude σ_j et f_j^k sont des approximations de $\sigma(x_j)$ et $f(x_j,\mu_k)$ respectivement.

La résolution de (5.29)-(5.30) est totalement explicite et très simple par l'équivalent discret de la méthode des caractéristiques. En effet, pour $\mu_k > 0$ on résout (5.29)-(5.30) selon les valeurs de j croissantes en partant de la condition aux limites

$$u_{1/2}^k = 0 \text{ pour } \mu_k > 0,$$

et en écrivant

$$u_{j+1/2}^k = \frac{(2\mu_k - \sigma_j\Delta x)u_{j-1/2}^k + 2\Delta x f_j^k}{2\mu_k + \sigma_j\Delta x} \quad \text{pour } j \geq 1, \tag{5.31}$$

tandis que pour $\mu_k < 0$ on résout (5.29)-(5.30) selon les valeurs de j décroissantes en partant de la condition aux limites

$$u_{N+1/2}^k = 0 \text{ pour } \mu_k < 0,$$

et en écrivant

$$u_{j-1/2}^k = \frac{(-2\mu_k - \sigma_j\Delta x)u_{j+1/2}^k + 2\Delta x f_j^k}{-2\mu_k + \sigma_j\Delta x} \quad \text{pour } j \leq N. \tag{5.32}$$

Il est clair maintenant pourquoi l'on évite la valeur nulle de la vitesse μ qui ne permet pas de résoudre (5.29)-(5.30) par cet algorithme de marche en espace (le sens de la condition aux limites n'est pas non plus clair pour $\mu = 0$).

Pour simplifier l'analyse nous allons supposer désormais que l'absorption $\sigma(x)$ est constante, c'est-à- dire que $\sigma_j \equiv \sigma \geq 0$ pour tout indice j. On déduit de (5.31)-(5.32) les formules exactes pour la solution discrète : pour $\mu_k > 0$,

$$u^k_{j+1/2} = \frac{2\Delta x}{2\mu_k + \sigma\Delta x} \sum_{i=1}^{j} (A_k)^{j-i} f^k_i \quad \text{avec } A_k = \frac{2\mu_k - \sigma\Delta x}{2\mu_k + \sigma\Delta x},$$

tandis que, pour $\mu_k < 0$,

$$u^k_{j-1/2} = \frac{2\Delta x}{-2\mu_k + \sigma\Delta x} \sum_{i=j}^{N} (A_k)^{i-j} f^k_i \quad \text{avec } A_k = \frac{-2\mu_k - \sigma\Delta x}{-2\mu_k + \sigma\Delta x},$$

Lemme 5.2.1 *Le schéma diamant (5.29)-(5.30) est consistant et précis d'ordre 2 en espace. De plus il est stable au sens L^∞, c'est-à-dire que, pour tout $j \in \{0, 1, ..., N\}$, on a*

$$|u^k_{j+1/2}| \leq \frac{2\ell}{|\mu_k|} \max_{x\in(-\ell,+\ell)} |f(x, \mu_k)|.$$

Démonstration. Par développement de Taylor en x autour du point x_j il est évident que le schéma diamant est précis à l'ordre 2. Par ailleurs, on vérifie que $|A_k| \leq 1$ pour tout k, puisque $\sigma \geq 0$. On en déduit donc

$$|u^k_{j+1/2}| \leq \frac{\Delta x}{|\mu_k|} N \max_{x\in(-\ell,+\ell)} |f(x, \mu_k)|$$

d'où l'on déduit le résultat de stabilité. ■

Exercice 5.8 *(difficile) En s'inspirant de la démonstration du Lemme 5.1.16 montrer que le schéma diamant (5.29)-(5.30) est stable au sens L^2.*

Lemme 5.2.2 *Le schéma diamant (5.29)-(5.30) vérifie le principe du maximum discret sous la condition*

$$\Delta x \leq \frac{2 \min_k |\mu_k|}{\sigma}.$$

Démonstration. Il faut vérifier que, si $f^k_j \geq 0$ pour tout j, alors la solution discrète vérifie aussi $u^k_{j+1/2} \geq 0$ pour tout j. C'est vrai si $A_k \geq 0$ ce qui correspond à la condition annoncée lorsque k varie. ■

Dans la pratique, la plus petite des vitesses est faible ou bien l'absorption est forte, ce qui entraine que la condition, de type CFL, du Lemme 5.2.2 est violée sauf pour des maillages très fins. Le schéma diamant n'est donc pas positif dans de nombreux cas, c'est-à-dire qu'il peut donner des valeurs négatives de la solution même si les sources sont positives, ce qui peut être génant du point de vue physique. Pour remédier à cet inconvénient on peut utiliser un autre

schéma comme le **schéma décentré amont** (appelé aussi "step scheme" en neutronique)

$$\begin{cases} \mu_k \dfrac{u_j^k - u_{j-1}^k}{\Delta x} + \sigma_j\, u_j^k = f_j^k & \text{pour } \mu_k > 0 \\ \mu_k \dfrac{u_{j+1}^k - u_j^k}{\Delta x} + \sigma_j\, u_j^k = f_j^k & \text{pour } \mu_k < 0 \end{cases} \tag{5.33}$$

qui est évidemment similaire au schéma décentré amont (5.13) pour l'équation d'advection. (Dans ce cas, il vaut mieux revenir à l'ancienne définition des points x_j qui permet d'écrire simplement la condition aux limites d'entrée, $u_0^k = 0$ pour $\mu_k > 0$, et $u_{N+1}^k = 0$ pour $\mu_k < 0$). Encore une fois on résout (5.33) pour les j croissants lorsque $\mu_k > 0$ et pour les j décroissants lorsque $\mu_k < 0$.

Lemme 5.2.3 *Le schéma décentré amont (5.33) vérifie le principe du maximum discret (sans condition sur les pas de discrétisation). Par ailleurs, il est consistant et précis à l'ordre 1 seulement.*

Démonstration. On réécrit le schéma sous la forme

$$u_j^k = \frac{\mu_k u_{j-1}^k + \Delta x f_j^k}{\mu_k + \sigma_j} \text{ pour } \mu_k > 0$$

et

$$u_j^k = \frac{-\mu_k u_{j+1}^k + \Delta x f_j^k}{-\mu_k + \sigma_j} \text{ pour } \mu_k < 0.$$

On vérifie sans peine, par récurrence, que $f_j^k \geq 0$, pour tout j, implique $u_j^k \geq 0$ pour tout j. Par développement de Taylor en x il est évident que le schéma décentré amont est précis à l'ordre 1 mais pas plus. ■

De la même manière que le schéma explicite centré pour l'équation d'advection était inutilisable, quoique pourtant très naturel, le schéma centré est instable pour l'équation de transport comme le montre l'exercice suivant.

Exercice 5.9 *Montrer que le schéma suivant, dit centré,*

$$\mu_k \frac{u_{j+1}^k - u_{j-1}^k}{2\Delta x} + \sigma_j\, u_j^k = f_j^k$$

est instable, c'est-à-dire que ses solutions exactes exhibent une partie à croissance exponentielle en j.
Indication : on pourra prendre $f_j^k \equiv 0$ et $\sigma_j = \sigma$.

5.2.2 Formules d'intégration numérique

Cette section est consacrée au choix de la famille de vitesses discrètes (μ_k) qui doivent appartenir à l'intervalle $[-1; +1]$ et être non nulles. Lorsqu'on doit discrétiser des opérateurs de collision en transport il est nécessaire d'évaluer des

intégrales par rapport à la vitesse μ. Pour cela on introduit des poids $\omega_k \in \mathbf{R}$ et, pour toute fonction f, on approche l'intégrale exacte par une somme de Riemann

$$\int_{-1}^{+1} f(\mu)d\mu \approx \sum_k \omega_k f(\mu_k).$$

Par conséquent, la motivation principale du choix des vitesses discrètes (μ_k) est la précision de cette formule d'intégration numérique ou quadrature .

Pour des raisons, physiques et mathématiques, de symétrie, on se restreint à des distributions symétriques par rapport à l'origine, c'est-à-dire que les vitesses sont impaires et les poids pairs

$$\mu_{-k} = -\mu_k \ , \quad \omega_{-k} = \omega_k \quad \text{pour tout } k.$$

On note $2K$ le nombre total de vitesses que l'on ordonne ainsi

$$-1 \leq \mu_{-K} < \mu_{-K+1} < \cdots < \mu_{-1} < 0 < \mu_1 < \cdots < \mu_{K-1} < \mu_K \leq +1.$$

Pour des raisons de positivité il est aussi souhaitable d'avoir des poids positifs $\omega_k \geq 0$. Avec toutes ces conditions il existe encore de nombreuses formules de quadrature possibles

$$\int_{-1}^{+1} f(\mu)d\mu \approx \sum_{k=-K, k\neq 0}^{K} \omega_k f(\mu_k). \tag{5.34}$$

Par exemple, une idée simple, mais un peu naïve, est de prendre des vitesses équidistribuées, c'est-à-dire

$$\mu_k = (1/2+k)/K \text{ pour } -K \leq k \leq -1, \quad \mu_k = (-1/2+k)/K \text{ pour } 1 \leq k \leq K,$$

et d'appliquer la formule des trapèzes sur chaque sous-intervalle ce qui donne $\omega_k = 1/K$ pour tout k.

Une méthode fréquemment utilisée car plus précise est la **formule d'intégration de Gauss** à $2K$ points qui repose sur l'utilisation des polynômes de Legendre, définis par

$$P_0(\mu) = 1 \ , \quad P_n(\mu) = \frac{1}{2^n n!}\frac{d^n}{d\mu^n}\Big((\mu^2-1)^n\Big) \quad \text{pour } n \geq 1. \tag{5.35}$$

Un calcul explicite facile conduit aux expressions suivantes pour les premiers polynômes

$$P_1(\mu) = \mu, \ P_2(\mu) = \frac{1}{2}(3\mu^2-1), \ P_3(\mu) = \frac{1}{2}(5\mu^3-3\mu).$$

Plus généralement nous laissons au lecteur le soin de vérifier les propriétés suivantes en exercice.

Exercice 5.10 *Les polynômes de Legendre, définis par (5.35), ont les propriétés suivantes :*

1. *le degré de P_n est exactement n, P_{2n} est pair et P_{2n+1} est impair,*
2. *ils sont orthogonaux au sens où*

$$\frac{1}{2}\int_{-1}^{+1} P_n(\mu)P_m(\mu)\,d\mu = \frac{\delta_{nm}}{2n+1} \tag{5.36}$$

   avec δ_{nm} le symbole de Kronecker (égal à 1 si $n = m$ et à 0 sinon),
3. *ils vérifient la relation de récurrence*

$$\mu P_n(\mu) = \frac{1}{2n+1}\Big((n+1)P_{n+1}(\mu) + n\,P_{n-1}(\mu)\Big), \tag{5.37}$$

4. *ils vérifient l'équation différentielle*

$$(1-\mu^2)\frac{d^2}{d\mu^2}P_n(\mu) - 2\mu\frac{d}{d\mu}P_n(\mu) + n(n+1)P_n(\mu) = 0,$$

5. *les racines de $P_n(\mu)$ sont toutes réelles, distinctes, comprises dans l'intervalle ouvert $(-1;+1)$ et symétriques par rapport à l'origine, c'est-à-dire que si μ est racine alors $-\mu$ aussi.*

Pour plus de détails, on pourra consulter tout ouvrage sur les fonctions spéciales, tel le célèbre Handbook of Mathematical functions *par Abramowitz et Stegun [1].*

Comme P_{2K} a $2K$ racines distinctes non nulles, on les choisit pour être les $2K$ vitesses discrètes (μ_k) et on prend

$$\omega_k = \int_{-1}^{+1} \Big(l_k(\mu)\Big)^2 d\mu \quad \text{avec} \quad l_k(\mu) = \prod_{j=-K, j\neq 0, j\neq k}^{K} \frac{\mu-\mu_j}{\mu_k-\mu_j}. \tag{5.38}$$

Le polynôme $l_k(\mu)$ est appelé polynôme d'interpolation de Lagrange. Il vaut exactement 1 pour $\mu = \mu_k$ et 0 pour toutes les autres valeurs $\mu = \mu_j$, $j \neq k$.

Lemme 5.2.4 *La formule de quadrature de Gauss, basée sur les vitesses discrètes (μ_k) égales aux racines du polynôme de Legendre P_{2K} et sur les poids (ω_k) définis par (5.38), est exacte pour tous les polynômes d'ordre inférieur où égal à $4K-1$. On dit qu'elle est d'ordre $4K-1$.*

Démonstration. Pour simplifier les notations, on désigne par $\mathcal{K}$ l'ensemble des indices entiers $-K \leq k \leq +K$ avec $k \neq 0$ (le cardinal de $\mathcal{K}$ est $2K$). Il est bien connu que les polynômes d'interpolation de Lagrange, $l_k(\mu)$ pour $k \in \mathcal{K}$, forment une base de l'ensemble $\mathbb{P}_{2K-1}$ des polynômes de degré inférieur ou égal à $2K-1$. De plus, tout polynôme $q \in \mathbb{P}_{2K-1}$ s'écrit

$$q(\mu) = \sum_{k\in\mathcal{K}} q(\mu_k) l_k(\mu).$$

La formule de quadrature $\sum_{k\in\mathcal{K}} \tilde{\omega}_k q(\mu_k)$ avec $\tilde{\omega}_k = \int_{-1}^{+1} l_k(\mu)\, d\mu$ est donc exacte pour tout polynôme $q \in \mathbb{P}_{2K-1}$

$$\int_{-1}^{+1} q(\mu)d\mu = \sum_{k\in\mathcal{K}} \tilde{\omega}_k q(\mu_k) \quad \forall q \in \mathbb{P}_{2K-1}.$$

Vérifions qu'elle l'est aussi pour les polynômes $q \in \mathbb{P}_{4K-1}$. Par division euclidienne, il existe deux polynômes d et r, dont les degrés sont l'un et l'autre inférieurs ou égaux à $2K-1$, tels que

$$q(\mu) = d(\mu) \prod_{k\in\mathcal{K}} (\mu - \mu_k) + r(\mu).$$

Or $\prod_{k\in\mathcal{K}}(\mu - \mu_k) = P_{2K}(\mu)$ qui est orthogonal à l'espace $\mathbb{P}_{2K-1}$ (voir (5.36) dans l'Exercice 5.10). Donc

$$\int_{-1}^{+1} q(\mu)d\mu = \int_{-1}^{+1} r(\mu)d\mu = \sum_{k\in\mathcal{K}} \tilde{\omega}_k r(\mu_k) = \sum_{k\in\mathcal{K}} \tilde{\omega}_k q(\mu_k),$$

c'est-à-dire que la formule de quadrature est exacte dans $\mathbb{P}_{4K-1}$. Finalement, en prenant $q(\mu) = (l_k(\mu))^2 \in \mathbb{P}_{4K-1}$, on vérifie que

$$\tilde{\omega}_k = \int_{-1}^{+1} l_k(\mu)\, d\mu = \int_{-1}^{+1} \Big(l_k(\mu)\Big)^2 d\mu = \omega_k.$$

■

Remarque 5.2.5 *Lorsque l'on choisit une discrétisation en vitesses μ_k, le dilemne est de choisir entre une grande précision obtenue grâce à une discrétisation fine (grande valeur de K) et un coût de calcul modéré pour une discrétisation plus grossière (petite valeur de K). Lorsque les phénomènes de collision sont importants (autrement dit, quand on est proche d'une limite de diffusion), un faible nombre de vitesses discrètes est suffisant en pratique. Néanmoins, lorsque les milieux sont "transparents" (c'est-à-dire qu'il y a peu de collisions), une trop grossière discrétisation en vitesse conduit à des artefacts numériques connus sous le nom* **d'effets de raie**. *Certaines directions sont privilégiées, ce qui peut conduire à des solutions non monotones en espace.*

Remarque 5.2.6 *Pour mieux éviter encore la singularité créée par la vitesse nulle dans la résolution de l'équation du transport, un autre choix populaire est la* **formule de double quadrature de Gauss**. *L'idée est prendre sur chacun des intervalles $[-1, 0]$ et $[0, +1]$ une formule de quadrature de Gauss à K points et d'ordre $2K+1$.*

Remarque 5.2.7 *Lorsque l'on définit la formule de quadrature (5.34) avec les vitesses discrètes μ_k, qui sont les racines du polynôme de Legendre P_{2K}, et avec les poids ω_k donnés par (5.38), la méthode des ordonnées discrètes, ou* **méthode**

S_N, est équivalente à la **méthode** *P_N, ou méthode des polynômes de Legendre. L'idée de cette dernière est, à partir de l'équation de Boltzmann (5.27), de ne pas discrétiser en vitesse, mais plutôt d'approcher l'inconnue $u(x,\mu)$ dans la base des polynômes de Legendre*

$$u(x,\mu) \approx \sum_{k=0}^{2K} (2k+1)\phi_k(x)P_k(\mu).$$

On utilise alors la propriété d'orthogonalité des polynômes P_k, ainsi que la relation de récurrence (5.37) pour montrer que l'équation de Boltzmann peut être approchée par un système d'équations différentielles ordinaires

$$\frac{k}{2k+1}\frac{d\phi_{k-1}}{dx} + \frac{k+1}{2k+1}\frac{d\phi_{k+1}}{dx} + \sigma(x)\phi_k = f_k(x) \quad 0 \le k \le 2K,$$

avec $f_k(x) = 1/2\int_{-1}^{+1} f(x,\mu)P_k(\mu)d\mu$ et, par convention, $\phi_{2K+1} \equiv 0$. Il n'y a plus alors qu'à discrétiser par différences finis ces équations différentielles ordinaires. L'intérêt de la méthode P_N est que les noyaux de collision (que nous étudierons dans la section suivante) s'écrivent aussi très simplement dans la base des polynômes de Legendre. Nous laissons au lecteur le soin de vérifier que les deux méthodes S_N et P_N sont équivalentes dans ce cas uni-dimensionnel (voir [38] si nécessaire).

5.2.3 Le cas stationnaire avec collisions

On considère maintenant l'équation de Boltzmann linéaire stationnaire dans les mêmes conditions que la section précédente mais en tenant compte désormais des collisions

$$\begin{cases} \mu\dfrac{\partial u}{\partial x}(x,\mu) + \sigma(x)u(x,\mu) = \dfrac{\sigma^*(x)}{2}\displaystyle\int_{-1}^{+1} u(x,\mu')\,d\mu' + f(x,\mu) \\ \qquad\qquad\qquad\qquad \text{pour } (x,\mu) \in (-\ell,+\ell)\times(-1,+1) \\ u(-\ell,\mu) = 0 \text{ pour } \mu > 0,\ u(+\ell,\mu) = 0 \text{ pour } \mu < 0. \end{cases} \tag{5.39}$$

Pour que le problème aux limites (5.39) soit bien posé (voir la Section 3.4) nous faisons l'hypothèse que le milieu est sous-critique, c'est-à-dire qu'il existe une constante $\sigma_0 > 0$ telle que

$$0 < \sigma_0 \le \sigma(x) - \sigma^*(x) \text{ pour } x \in (-\ell,+\ell). \tag{5.40}$$

Nous décrivons la **méthode S_N ou des ordonnées discrètes** dans ce contexte. Le maillage en espace est toujours défini par les points $x_{j+1/2}$ définis dans (5.28) et on choisit une des discrétisations symétriques en vitesse décrites dans la section 5.2.2, (μ_k) où l'indice k varie dans $\{-K,...,-1\}\cup\{1,...,K\}$ mais ne prend pas la valeur 0 afin qu'aucune vitesse μ_k ne soit nulle. Les poids ω_k sont aussi symétriques et positifs et la formule de quadrature est (5.34).

Dans ce cadre le **schéma diamant** est donné, pour $1 \leq j \leq N$, par

$$\begin{cases} \mu_k \dfrac{u^k_{j+1/2} - u^k_{j-1/2}}{\Delta x} + \sigma_j u^k_j = \sigma^*_j \bar{u}_j + f^k_j \\ u^k_j = \dfrac{u^k_{j+1/2} + u^k_{j-1/2}}{2} \end{cases} \tag{5.41}$$

où la deuxième ligne est la relation diamant (5.30) et $\bar{u}_j$ est la moyenne angulaire définie par

$$\bar{u}_j = \frac{1}{2} \sum_{k=-K, k\neq 0}^{K} \omega_k u^k_j. \tag{5.42}$$

Comme d'habitude σ_j, σ^*_j et f^k_j sont des approximations de $\sigma(x_j)$, $\sigma(x_j)$ et $f(x_j, \mu_k)$ respectivement. Les conditions aux limites de flux entrant nul sont

$$u^k_{1/2} = 0 \text{ pour } \mu_k > 0\,, \quad \text{et} \quad u^k_{N+1/2} = 0 \text{ pour } \mu_k < 0\,.$$

Comme pour le schéma sans collisions (5.29)-(5.30) de la section précédente, la précision de (5.41)-(5.42) est d'ordre 2 en espace. Nous expliquerons un peu plus loin (voir le Lemme 5.2.10) comment on calcule en pratique la solution discrète du schéma diamant. Auparavant nous étudions sa stabilité en nous inspirant d'une inégalité d'énergie dans le cas continu (voir le Lemme 3.2.5).

Lemme 5.2.8 *La solution $u(x, \mu)$ de l'équation de transport (5.39) vérifie*

$$\int_{-\ell}^{+\ell} \int_{-1}^{+1} |u(x,\mu)|^2 dx\, d\mu \leq \frac{1}{\sigma_0^2} \int_{-\ell}^{+\ell} \int_{-1}^{+1} |f(x,\mu)|^2 dx\, d\mu.$$

Démonstration. On multiplie l'équation (5.39) par u et on intègre par parties. Le terme de transport devient

$$\int_{-\ell}^{+\ell} \int_{-1}^{+1} \mu \frac{\partial u}{\partial x} u\, dx\, d\mu = \frac{1}{2} \int_{-1}^{+1} \mu |u(+\ell,\mu)|^2\, d\mu - \frac{1}{2} \int_{-1}^{+1} \mu |u(-\ell,\mu)|^2\, d\mu \geq 0$$

à cause des conditions aux limites imposées. Par conséquent

$$\int_{-\ell}^{+\ell} \int_{-1}^{+1} \sigma |u|^2 dx\, d\mu \leq \frac{1}{2} \int_{-\ell}^{+\ell} \sigma^* \left(\int_{-1}^{+1} u\, d\mu \right)^2 dx + \int_{-\ell}^{+\ell} \int_{-1}^{+1} f\, u\, dx\, d\mu.$$

Or, par Cauchy-Schwarz,

$$\left(\int_{-1}^{+1} u\, d\mu \right)^2 \leq 2 \int_{-1}^{+1} |u|^2\, d\mu,$$

d'où l'on déduit

$$\sigma_0 \int_{-\ell}^{+\ell} \int_{-1}^{+1} |u|^2 dx\, d\mu \leq \int_{-\ell}^{+\ell} \int_{-1}^{+1} (\sigma - \sigma^*) |u|^2 dx\, d\mu \leq \int_{-\ell}^{+\ell} \int_{-1}^{+1} f\, u\, dx\, d\mu.$$

Une nouvelle application de l'inégalité de Cauchy-Schwarz permet de conclure. ■

On adapte l'argument de la preuve du Lemme 5.2.8 pour obtenir le résultat suivant de stabilité.

Lemme 5.2.9 *Le schéma diamant (5.41-5.42) est inconditionnellement stable L^2 au sens où sa solution discrète u_j^k vérifie*

$$\|(u_j^k)\| \leq \frac{1}{\sigma_0}\|(f_j^k)\|.$$

avec la norme discrète définie par

$$\|(u_j^k)\|^2 = \sum_{j=1}^{N} \Delta x \sum_{k=-K,k\neq 0}^{K} \omega_k |u_j^k|^2.$$

Démonstration. On multiplie (5.41) par $\Delta x \omega_k (u_{j+1/2}^k + u_{j-1/2}^k) = 2\Delta x \omega_k u_j^k$ et on somme sur j et k. Le terme de transport devient

$$\begin{aligned} \sum_{j=1}^{N} \sum_{k=-K,k\neq 0}^{K} \omega_k \mu_k \left((u_{j+1/2}^k)^2 - (u_{j-1/2}^k)^2 \right) &= \\ \sum_{k=-K,k\neq 0}^{K} \omega_k \mu_k (u_{N+1/2}^k)^2 - \sum_{k=-K,k\neq 0}^{K} \omega_k \mu_k (u_{1/2}^k)^2 &\geq 0 \end{aligned}$$

à cause des conditions aux limites imposées. Par conséquent, le reste de (5.41) donne

$$2\sum_{j=1}^{N} \sum_{k=-K,k\neq 0}^{K} \Delta x \sigma_j \omega_k (u_j^k)^2 \leq 4\sum_{j=1}^{N} \Delta x \sigma_j^* (\bar{u}_j)^2 + 2\sum_{j=1}^{N} \sum_{k=-K,k\neq 0}^{K} \Delta x \omega_k u_j^k f_j^k.$$

Or, par Cauchy-Schwarz,

$$\begin{aligned} (\bar{u}_j)^2 = \left(\frac{1}{2} \sum_{k=-K,k\neq 0}^{K} \omega_k u_j^k \right)^2 &\leq \left(\frac{1}{2} \sum_{k=-K,k\neq 0}^{K} \omega_k \right) \left(\frac{1}{2} \sum_{k=-K,k\neq 0}^{K} \omega_k (u_j^k)^2 \right) \\ &\leq \frac{1}{2} \sum_{k=-K,k\neq 0}^{K} \omega_k (u_j^k)^2, \end{aligned}$$

d'où l'on déduit, grâce à l'hypothèse de sous-criticité (5.40),

$$\sigma_0 \sum_{j=1}^{N} \sum_{k=-K,k\neq 0}^{K} \Delta x \omega_k (u_j^k)^2 \leq \sum_{j=1}^{N} \sum_{k=-K,k\neq 0}^{K} \Delta x \omega_k u_j^k f_j^k.$$

Une nouvelle application de l'inégalité de Cauchy-Schwarz permet de conclure. ■

Nous revenons maintenant à la définition du schéma diamant et expliquons comment on peut calculer sa solution discrète. Il est possible de résoudre simultanément toutes les équations du schéma (5.41)-(5.42) en résolvant un grand système linéaire pour le vecteur inconnu ayant $2KN$ composantes $(u^k_{j+1/2})$. Mais cela requiert beaucoup de place mémoire et de temps de calcul, pas tant en une dimension d'espace mais pour les dimensions 2 ou 3 d'espace. En général on préfère utiliser une méthode itérative très simple, connue sous le nom **d'itération sur les sources**. Son principe est de supposer connu le membre de droite de (5.41) (y compris la moyenne angulaire), de résoudre l'équation de transport sans collision par un schéma de la Section 5.2.1, de mettre à jour le membre de droite de (5.41), puis d'itérer ce procédé jusqu'à convergence. Cet algorithme itératif est l'exact analogue, en discret, de l'argument de point fixe utilisé pour démontrer l'existence d'une solution de l'équation de Boltzmann au Théorème 3.1.2.

Plus précisément, on note $n \geq 0$ le numéro d'itération. On initialise l'algorithme (dit d'itération sur les sources) en posant, pour $n = 0$,

$$\bar{u}^0_j = 0,$$

puis à l'itération $n \geq 1$ on résout

$$\mu_k \frac{u^{k,n}_{j+1/2} - u^{k,n}_{j-1/2}}{\Delta x} + \sigma_j \frac{u^{k,n}_{j+1/2} + u^{k,n}_{j-1/2}}{2} = \sigma^*_j \bar{u}^{n-1}_j + f^k_j \tag{5.43}$$

et on met à jour la moyenne angulaire

$$\bar{u}^n_j = \frac{1}{2} \sum_{k=-K, k\neq 0}^{K} \omega_k \frac{u^{k,n}_{j+1/2} + u^{k,n}_{j-1/2}}{2}. \tag{5.44}$$

La résolution de (5.43) est identique à celle de (5.29)-(5.30) dans la section précédente et est donc très facile par simple remontée des caractéristiques. L'intérêt de cet algorithme est qu'il ne nécessite aucun stockage de matrice ni résolution de système linéaire.

Lemme 5.2.10 *L'algorithme d'itération sur les sources (5.43)-(5.44) converge, lorsque n tend vers l'infini, vers la solution discrète du schéma (5.41)-(5.42).*

Démonstration. Afin d'étudier sa convergence lorsque n tend vers $+\infty$ nous réécrivons (5.43)-(5.44) sous une forme matricielle plus compacte. On note U^n le vecteur de composantes $(u^{k,n}_{j+1/2})$, F le vecteur de composantes (f^k_j), T la matrice de l'opérateur de transport discrétisé dans le membre de gauche de (5.43) et enfin K la matrice de l'opérateur de collision discrétisé défini par (5.44) que multiplie le coefficient σ^*. Avec ces notations U^n est la solution de

$$TU^n = KU^{n-1} + F. \tag{5.45}$$

La suite U^n converge, c'est-à-dire que la méthode itérative converge (pour tout second membre F), si et seulement si le rayon spectral de $T^{-1}K$ est strictement plus petit que 1

$$\rho(T^{-1}K) < 1.$$

(Rappelons que le rayon spectral d'une matrice est défini comme le maximum des modules de ses valeurs propres.) Comme $\rho(T^{-1}K) \leq \|T^{-1}K\| \leq \|T^{-1}\|\,\|K\|$ (cf. Proposition 13.1.7 dans [2]), il suffit de montrer que $\|T^{-1}\|\,\|K\| < 1$. Pour cela on s'inspire de la démonstration du Lemme 5.2.9. Pour un second membre G on appelle U la solution de

$$TU = G. \tag{5.46}$$

On multiplie (5.46) par U et on somme sur toutes les composantes, ce qui est équivalent à multiplier le terme de transport de (5.41) par $\Delta x \omega_k (u^k_{j+1/2} + u^k_{j-1/2})$ et à sommer sur j et k, calcul que nous avons déjà fait dans la démonstration du Lemme 5.2.9. En notant $\|U\|$ la norme (déjà introduite au Lemme 5.2.9)

$$\|U\| = \left(\sum_{j=1}^{N} \Delta x \sum_{k=-K, k\neq 0}^{K} \omega_k |u^k_j|^2 \right)^{1/2},$$

où on a utilisé la relation diamant $2u^k_j = u^k_{j+1/2} + u^k_{j-1/2}$, on obtient que

$$\sigma \|U\|^2 \leq TU \cdot U = G \cdot U \leq \|G\| \|U\|,$$

c'est-à-dire que

$$\|T^{-1}G\| = \|U\| \leq \frac{1}{\sigma} \|G\|.$$

Par ailleurs, on vérifie que

$$\|KU\|^2 = (\sigma^*)^2 \sum_{j=1}^{N} \Delta x \sum_{k=-K, k\neq 0}^{K} \omega_k |\bar{u}_j|^2 = 2(\sigma^*)^2 \sum_{j=1}^{N} \Delta x |\bar{u}_j|^2$$

et, par Cauchy-Schwarz pour $\bar{u}_j = \frac{1}{2} \sum_{k=-K, k\neq 0}^{K} \omega_k u^k_j$,

$$\|KU\|^2 \leq (\sigma^*)^2 \sum_{j=1}^{N} \Delta x \sum_{k=-K, k\neq 0}^{K} \omega_k |u^k_j|^2 = (\sigma^*)^2 \|U\|^2,$$

d'où l'on déduit que $\|T^{-1}\|\,\|K\| \leq \sigma^*/\sigma < 1$ à cause de l'hypothèse de sous-criticité (5.40). ■

5.2.4 Accélération par la diffusion

Dans cette section nous allons expliquer comment l'algorithme d'itération sur les sources (5.43)-(5.44) peut être accéléré afin qu'il converge plus rapidement (c'est-à-dire, avec un plus petit nombre d'itérations), et ceci grâce, encore une fois, à l'approximation du transport par la diffusion. Afin d'expliquer le principe de cette méthode d'accélération (ou de diffusion synthétique), commençons par réécrire l'équation de Boltzmann (5.39) sous la forme abstraite suivante

$$Tu = Ku + f,$$

où T est l'opérateur de transport, muni de ses conditions aux limites,

$$\left\{ \begin{array}{l} Tu = \mu \dfrac{\partial u}{\partial x} + \sigma\, u \text{ pour } (x,\mu) \in (-\ell,+\ell) \times (-1,+1) \\ u(-\ell,\mu) = 0 \text{ pour } \mu > 0,\ u(+\ell,\mu) = 0 \text{ pour } \mu < 0, \end{array} \right.$$

et K est l'opérateur de collision

$$Ku(x,\mu) = \frac{\sigma^*(x)}{2} \int_{-1}^{+1} u(x,\mu')\, d\mu'.$$

On note $\overline{\cdot}$ l'opérateur de moyennisation angulaire défini par

$$\overline{u}(x) = \frac{1}{2} \int_{-1}^{+1} u(x,\mu')\, d\mu'.$$

Remarquons que $Ku = K\overline{u}$. Introduisons un opérateur de diffusion D qui soit une approximation convenable du transport, au sens de la Section 1.1.3 : il ne s'applique qu'à des fonctions $\overline{u}(x)$ ne dépendant pas de la vitesse μ,

$$D\overline{u} = -\operatorname{div}(\mathcal{D}\nabla\overline{u}) + \sigma_{\mathcal{D}}\overline{u},$$

et il est muni de conditions aux limites convenables.

Une version continue de l'algorithme d'itération sur les sources (5.43)-(5.44) s'écrit

$$\left\{ \begin{array}{l} Tv^n = Ku^{n-1} + f \equiv K\overline{u}^{n-1} + f, \\ u^n = v^n, \end{array} \right. \tag{5.47}$$

où nous avons fait exprès d'introduire une inconnue supplémentaire, inutile ici, v^n. Remarquons aussi que seul $\overline{u}^{n-1}$ est nécessaire dans le membre de droite. L'idée est de modifier la relation donnant u^n en fonction de v^n. Pour cela on remarque qu'en moyennant l'équation abstraite $Tu = Ku + f$ et en additionnant/soustrayant l'opérateur D on obtient

$$D\overline{u} = \overline{f} - \overline{(T-K-D)u}.$$

On propose alors le nouveau schéma itératif

$$\left\{ \begin{array}{l} Tv^n = K\overline{u}^{n-1} + f, \\ D\overline{u}^n = \overline{f} - \overline{(T-K-D)v^n}, \end{array} \right.$$

que l'on peut réécrire plus simplement en remarquant que la seconde équation est équivalente à

$$D\overline{(u^n - v^n)} = \overline{f} - \overline{(T-K)v^n} = \overline{K(v^n - u^{n-1})} = K(\overline{v}^n - \overline{u}^{n-1}),$$

où l'on a utilisé une moyenne de la première équation pour éliminer la source f. Ainsi donc, l'algorithme d'itération sur les sources accéléré par diffusion est

$$\left\{ \begin{array}{l} Tv^n = K\overline{u}^{n-1} + f, \\ \overline{u}^n = \overline{v}^n + D^{-1}K(\overline{v}^n - \overline{u}^{n-1}). \end{array} \right. \tag{5.48}$$

Formellement, si l'opérateur de diffusion D "vaut l'infini", on retombe sur l'algorithme précédent (5.47). Sinon, comme D^{-1} et K sont des opérateurs positifs, la deuxième relation de (5.48) s'interprète comme une extrapolation entre $\overline{v}^n$ et $\overline{u}^{n-1}$ pour obtenir $\overline{u}^n$. La résolution de (5.48) est un peu plus chère, à chaque itération, que celle de (5.47) puisqu'il faut résoudre d'abord une équation de transport pour obtenir v^n puis une équation de diffusion pour obtenir $\overline{u}^n$.

D'un point de vue algébrique et discret, l'algorithme précédent (5.47) s'interprétait selon (5.45) comme

$$TU^n = KU^{n-1} + F,$$

où U^n est le vecteur des composantes de la discrétisation de u^n. Si on élimine v^n dans (5.48), sachant que

$$v^n = (I + D^{-1}K)^{-1}(\overline{u}^n + D^{-1}Ku^{n-1}),$$

ce nouvel algorithme est modifié (on dit préconditionné) comme suit

$$T(I + D^{-1}K)^{-1}\Big(U^n + D^{-1}KU^{n-1}\Big) = KU^{n-1} + F. \tag{5.49}$$

La matrice d'itération de (5.49) est

$$T^{-1}K - D^{-1}K\Big(I - T^{-1}K\Big)$$

dont on espère que le rayon spectral est plus petit que celui de $T^{-1}K$, ce qui est vrai si on sait déjà que $\rho(T^{-1}K) < 1$ et que $D^{-1}K$ est suffisamment petit. Bien sûr, si la suite U^n, définie par (5.49), converge, alors elle converge vers la même limite que celle définie par (5.45).

5.2.5 Equation instationnaire ou cinétique

On considère maintenant l'équation complète de Boltzmann linéaire dépendant du temps (ou modèle cinétique) pour l'inconnue $u(t, x, \mu)$

$$\begin{cases} \dfrac{\partial u}{\partial t} + \mu\dfrac{\partial u}{\partial x} + \sigma(x)u = \dfrac{\sigma^*(x)}{2}\displaystyle\int_{-1}^{+1} u(x,\mu')\,d\mu' + f(x,\mu) \\ \qquad\qquad\qquad\qquad \text{pour } (t,x,\mu) \in \mathbf{R}^+ \times (-\ell,+\ell) \times (-1,+1), \\ u(t=0,x,\mu) = u^0(x,\mu) \qquad\quad \text{pour } (x,\mu) \in (-\ell,+\ell)\times(-1,+1), \\ u(-\ell,\mu) = 0 \text{ pour } \mu > 0,\ u(+\ell,\mu) = 0 \text{ pour } \mu < 0. \end{cases} \tag{5.50}$$

Pour que le problème aux limites (5.50) admette une solution qui ne croît pas exponentiellement en temps, nous supposons encore que le milieu est sous-critique, mais avec une hypothèse un peu plus faible que (5.40), à savoir

$$0 \le \sigma(x) - \sigma^*(x) \text{ pour } x \in (-\ell,+\ell). \tag{5.51}$$

Reprenons la **méthode S_N ou des ordonnées discrètes** dans ce contexte. On note $u_j^{n,k}$ une approximation de la solution $u(t_n, x_j, \mu_k)$. Le **schéma diamant**, obtenu par combinaison des schémas (5.14) et (5.41)-(5.42), s'écrit, pour $1 \leq j \leq N$,

$$\begin{aligned} \frac{u_j^{n+1,k} - u_j^{n,k}}{\Delta t} + \mu_k \frac{u_{j+1/2}^{n+1/2,k} - u_{j-1/2}^{n+1/2,k}}{\Delta x} + \sigma_j u_j^{n+1/2,k} \\ = \sigma_j^* \bar{u}_j^{n+1/2} + f_j^{n+1/2,k} \end{aligned} \tag{5.52}$$

avec les relations diamant

$$\begin{aligned} u_j^{n+1,k} + u_j^{n,k} &= u_{j+1/2}^{n+1/2,k} + u_{j-1/2}^{n+1/2,k} \\ 2u_j^{n+1/2,k} &= u_{j+1/2}^{n+1/2,k} + u_{j-1/2}^{n+1/2,k} \end{aligned} \tag{5.53}$$

et la moyenne angulaire

$$\bar{u}_j^{n+1/2} = \frac{1}{2} \sum_{k=-K, k\neq 0}^{K} \omega_k \, u_j^{n+1/2,k}. \tag{5.54}$$

Comme d'habitude σ_j, σ_j^* et $f_j^{n+1/2,k}$ sont des approximations de $\sigma(x_j)$, $\sigma^*(x_j)$ et $f(t_{n+1/2}, x_j, \mu_k)$ respectivement. La première relation diamant (5.53) permet d'éliminer l'inconnue $u_j^{n+1,k}$ tandis que la seconde relation diamant permet d'éliminer $u_j^{n+1/2,k}$ et d'obtenir un schéma implicite pour les valeurs $u_{j+1/2}^{n+1/2,k}$ en fonction des valeurs $u_j^{n,k}$

$$\begin{aligned} \mu_k \frac{u_{j+1/2}^{n+1/2,k} - u_{j-1/2}^{n+1/2,k}}{\Delta x} + \left(\sigma_j + \frac{2}{\Delta t}\right) \frac{u_{j+1/2}^{n+1/2,k} + u_{j-1/2}^{n+1/2,k}}{2} \\ = \frac{\sigma_j^*}{2} (\bar{u}_{j+1/2}^{n+1/2} + \bar{u}_{j-1/2}^{n+1/2}) + f_j^{n+1/2,k} + \frac{2}{\Delta t} u_j^{n,k}. \end{aligned} \tag{5.55}$$

On retrouve ensuite les valeurs $u_j^{n+1,k}$ grâce à la première relation diamant dans (5.53).

Le schéma (5.55) est complètement similaire au schéma stationnaire (5.41)-(5.42) avec simplement un terme source modifié et une absorption σ_j augmentée de $2/\Delta t$. En particulier, on utilise les mêmes conditions aux limites de flux nul en entrée, et on peut encore résoudre (5.55) par une méthode d'itération sur les sources (cf. Lemme 5.2.10).

Lemme 5.2.11 *Le schéma diamant (5.52)-(5.53)-(5.54) est inconditionnellement stable L^2 au sens où, pour tout temps final $T > 0$, il existe une constante $C(T) > 0$ telle que la solution discrète $u_j^{n,k}$ vérifie pour tout $n \leq T/\Delta t$*

$$\|(u_j^{n,k})\|^2 \leq C(T) \left(\|(u_j^{0,k})\|^2 + \sum_{m=0}^{n} \Delta t \|(f_j^{n+1/2,k})\|^2 \right). \tag{5.56}$$

avec la norme discrète définie par

$$\|(u_j^{n,k})\|^2 = \sum_{j=1}^{N} \Delta x \sum_{k=-K,k\neq 0}^{K} \omega_k |u_j^{n,k}|^2.$$

Démonstration. On multiplie le schéma (5.52) par $\Delta t \omega_k (u_j^{n+1,k} + u_j^{n,k})$ et on utilise les deux relations diamant (5.53) pour obtenir, après sommation,

$$\sum_{j=1}^{N} \sum_{k=-K,k\neq 0}^{K} \omega_k \left(|u_j^{n+1,k}|^2 - |u_j^{n,k}|^2 \right) + 2\Delta t \sum_{j=1}^{N} \sum_{k=-K,k\neq 0}^{K} \omega_k \sigma_j |u_j^{n+1/2,k}|^2$$
$$\leq 4\Delta t \sum_{j=1}^{N} \sigma_j^* |\bar{u}_j^{n+1/2}|^2 + \Delta t \sum_{j=1}^{N} \sum_{k=-K,k\neq 0}^{K} \omega_k f_j^{n+1/2,k} (u_j^{n+1,k} + u_j^{n,k})$$

car le terme de transport donne une contribution positive comme dans la démonstration du Lemme 5.2.9. Or, par Cauchy-Schwarz,

$$|\bar{u}_j^{n+1/2}|^2 \leq \frac{1}{2} \sum_{k=-K,k\neq 0}^{K} \omega_k |u_j^{n+1/2,k}|^2$$

et, comme $\sigma_j \geq \sigma_j^*$, les termes d'absorption peuvent s'éliminer dans l'inégalité. D'autre part,

$$\sum_{j=1}^{N} \sum_{k=-K,k\neq 0}^{K} \omega_k f_j^{n+1/2,k} (u_j^{n+1,k} + u_j^{n,k}) \leq \|f_j^{n+1/2,k}\| \left(\|u_j^{n+1,k}\| + \|u_j^{n,k}\| \right)$$
$$\leq \|f_j^{n+1/2,k}\|^2 + \frac{1}{2} \left(\|u_j^{n+1,k}\|^2 + \|u_j^{n,k}\|^2 \right)$$

car, pour trois nombre positifs a, b, c, on a $a(b+c) \leq a^2 + (b^2+c^2)/2$. En combinant ces inégalités on déduit

$$(1 - \Delta t/2)\|u_j^{n+1,k}\|^2 \leq (1 + \Delta t/2)\|u_j^{n,k}\|^2 + \Delta t \|f_j^{n+1/2,k}\|^2.$$

Comme il existe une constante $C > 0$ telle que, pour tout $T > 0$ et $\Delta t > 0$ petit, on a

$$\left(\frac{1 + \Delta t/2}{1 - \Delta t/2} \right)^{T/\Delta t} \leq e^{C\,T},$$

on en déduit l'inégalité (5.56). ■

Exercice 5.11 *(difficile) En s'inspirant du Lemme 5.2.8, démontrer l'estimation d'énergie dans $L^\infty((0,T); L^2((-\ell,+\ell) \times (-1,+1)))$ pour l'équation (5.50) qui soit équivalente à l'inégalité de stabilité discrète (5.56).*

5.2.6 Généralisation à la dimension d'espace $N = 2$.

Nous nous sommes limités jusqu'ici à la discrétisation d'équations en une seule dimension d'espace. Nous expliquons maintenant comment généraliser ce qui précède en dimension plus grande. Sans perte de généralité nous nous concentrons sur le cas de la dimension d'espace $N = 2$; le cas $N = 3$ est complètement similaire dans le principe même si les temps de calcul et l'encombrement en mémoire sont beaucoup plus importants et limitent singulièrement la taille des problèmes que l'on peut résoudre.

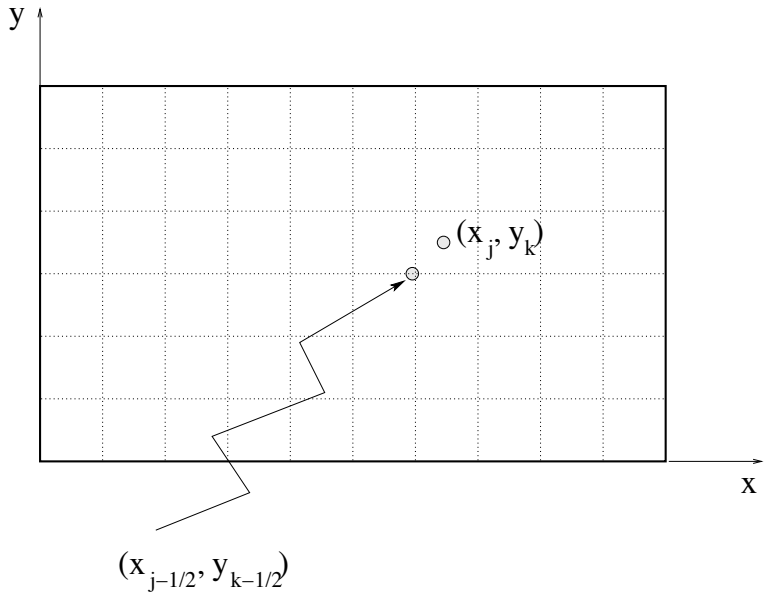

FIGURE 5.3 – Maillage 2d de type volumes finis où les points (x_j, y_k) sont au centre des mailles.

Pour une présentation de la méthode des différences finies en dimension $N = 2$ pour l'équation de diffusion, nous renvoyons à [2]. Nous discutons donc ici uniquement de l'équation de Boltzmann. En fait, l'algorithme d'itérations sur les sources, vu à la Section 5.2.3, est le même quelle que soit la dimension, et permet de se ramener à la résolution d'une simple équation de transport. Pour simplifier nous supposons que cette équation est stationnaire et mono-groupe, c'est-à-dire que l'ensemble des vitesses possibles est caractérisé par le vecteur $\omega = (\omega_x, \omega_y)$ dans la sphère unité $|\omega| = 1$, et que le domaine de calcul est un rectangle $R = (0, \ell_x) \times (0, \ell_y)$ dans le plan. Nous considérons donc l'équation suivante de transport sans collision

$$\left\{ \begin{array}{l} \omega_x \dfrac{\partial u}{\partial x} + \omega_y \dfrac{\partial u}{\partial y} + \sigma(x,y)u = f(x,y,\omega) \text{ dans } R \times \{|\omega| = 1\} \\ u(x,y,\omega) = 0 \text{ pour } (x,y,\omega) \in \Gamma^-, \end{array} \right. \tag{5.57}$$

où Γ^- est la frontière rentrante dans l'espace des phases

$$\Gamma^- = \{(x,y) \in \partial R \text{ tel que } n(x,y)\cdot\omega < 0\},$$

$\sigma(x,y) \geq 0$ est la section efficace d'absorption, et $f(x,y,\omega)$ est un terme source. Nous décrivons la méthode des ordonnées discrètes qui est une méthode de différences finies mais s'interprète aussi comme une méthode de volumes finis. On utilise un maillage cartésien défini par les pas d'espace $\Delta x = \frac{\ell_x}{N_x}$ et $\Delta y = \frac{\ell_y}{N_y}$ et par les points

$$x_{j+1/2} = j\Delta x \text{ pour } j \in \{0,1,...,N_x\},$$
$$y_{k+1/2} = k\Delta y \text{ pour } k \in \{0,1,...,N_y\}.$$

Dans l'approche volumes finis les points (x_j, y_k) sont vus comme les centres des cellules rectangulaires de sommets les points milieux $(x_{j+1/2}, y_{k+1/2})$ (voir la Figure 5.3). Autrement dit, $x_j = (j-1/2)\Delta x$, pour $j \in \{1,...,N_x\}$, et $y_k = (k-1/2)\Delta y$, pour $k \in \{1,...,N_y\}$.

L'espace des vitesses $|\omega| = 1$ est discrétisé par une famille finie (ω_p). Nous ne disons rien sur la façon dont on discrétise la sphère unité et nous renvoyons le lecteur, par exemple, à [38]. La seule hypothèse que nous faisons est que, comme en dimension un, aucune des composantes des vitesses discrètes $\omega_{p,x}, \omega_{p,y}$ ne s'annule.

L'idée du schéma diamant est d'intégrer l'équation de transport (5.57) sur une cellule rectangulaire autour du point (x_j, y_k) en supposant les coefficients constants dans cette maille. On obtient ainsi

$$\omega_{p,x}\frac{u^p_{j+1/2,k} - u^p_{j-1/2,k}}{\Delta x} + \omega_{p,y}\frac{u^p_{j,k+1/2} - u^p_{j,k-1/2}}{\Delta y} + \sigma_{j,k}u^p_{j,k} = f^p_{j,k} \qquad (5.58)$$

où $u^p_{j,k}$ est une approximation de $u(x_j, y_k, \omega_p)$ et des définitions similaires pour les autres termes. On complète (5.58) par deux relations diamant supplémentaires (autant que de dimensions d'espace)

$$u^p_{j,k} = \frac{u^p_{j+1/2,k} + u^p_{j-1/2,k}}{2}, \qquad (5.59)$$

$$u^p_{j,k} = \frac{u^p_{j,k+1/2} + u^p_{j,k-1/2}}{2}. \qquad (5.60)$$

Avant de voir comment résoudre le système d'équations discrètes (5.58)-(5.59)-(5.60), comptons le nombre d'équations et celui d'inconnues, pour une vitesse ω_p donnée. Nous avons 3 équations par cellule ou par point (x_j, y_k), c'est-à-dire $3N_xN_y$ équations. Nous comptons aussi N_xN_y inconnues $u^p_{j,k}$ ainsi que $(N_x+1)N_y$ inconnues $u^p_{j+1/2,k}$ et $(N_y+1)N_x$ inconnues $u^p_{j,k+1/2}$, soit au total $3N_xN_y + N_x + N_y$ inconnues. Grâce aux conditions aux limites sur le bord rentrant Γ^- (et avec l'hypothèse usuelle qu'aucune des composantes de la vitesse ω_p n'est nulle) nous pouvons rajouter N_x+N_y relations supplémentaires donnant

les valeurs de $u^p_{j+1/2,k}$ et $u^p_{j,k+1/2}$ pour les indices j et k correspondant à Γ^-. On a donc autant d'inconnues que d'équations dans ce système linéaire.

Comme en dimension un, on peut résoudre explicitement ce système par l'analogue discret de la méthode des caractéristiques. L'idée consiste à **balayer** le domaine de calcul R en partant des conditions aux limites et en suivant le sens des caractéristiques. On doit donc distinguer 4 cas selon le signe des composantes de la vitesse ω_p :

(1) si $\omega_{p,x} > 0$ et $\omega_{p,y} > 0$, on balaye de gauche à droite et de bas en haut,

(2) si $\omega_{p,x} > 0$ et $\omega_{p,y} < 0$, on balaye de gauche à droite et de haut en bas,

(3) si $\omega_{p,x} < 0$ et $\omega_{p,y} > 0$, on balaye de droite à gauche et de bas en haut,

(4) si $\omega_{p,x} < 0$ et $\omega_{p,y} < 0$, on balaye de droite à gauche et de haut en bas.

On réécrit aussi les relations diamant (5.59)-(5.60) suivant le signe des composantes de ω_p afin d'éliminer l'inconnue dans le membre de gauche des équation qui suivent :

$$u^p_{j+1/2,k} = 2u^p_{j,k} - u^p_{j-1/2,k} \quad \text{si } \omega_{p,x} > 0, \tag{5.61}$$

$$u^p_{j-1/2,k} = 2u^p_{j,k} - u^p_{j+1/2,k} \quad \text{si } \omega_{p,x} < 0, \tag{5.62}$$

$$u^p_{j,k+1/2} = 2u^p_{j,k} - u^p_{j,k-1/2} \quad \text{si } \omega_{p,y} > 0, \tag{5.63}$$

$$u^p_{j,k-1/2} = 2u^p_{j,k} - u^p_{j,k+1/2} \quad \text{si } \omega_{p,y} < 0. \tag{5.64}$$

En reportant ces relations dans (5.58) on obtient un système d'équation pour les inconnues $u^p_{j,k}$ que l'on résout par le procédé de balayage évoqué ci-dessus.

Explicitons cette résolution dans le cas particulier (1) ci-dessus, c'est-à-dire $\omega_{p,x} > 0$ et $\omega_{p,y} > 0$ (les autres cas s'obtiennent de manière similaire : nous laissons au lecteur le soin de le vérifier en exercice). On calcule les inconnues discrètes dans l'ordre indiqué sur la Figure 5.4. Les valeurs $u^p_{j,k}$ sont calculés par la formule suivante obtenue en reportant (5.61) et (5.63) dans (5.58)

$$u^p_{j,k} = \frac{f^p_{j,k} + \frac{2\omega_{p,x}}{\Delta x} u^p_{j-1/2,k} + \frac{2\omega_{p,y}}{\Delta y} u^p_{j,k-1/2}}{\sigma_{j,k} + \frac{2\omega_{p,x}}{\Delta x} + \frac{2\omega_{p,y}}{\Delta y}}. \tag{5.65}$$

5.3 Autres méthodes numériques

5.3.1 Méthodes intégrales

Les méthodes intégrales de résolution numérique de l'équation de Boltzmann linéaire sont basées sur la formule de représentation intégrale de la solution (2.1), appelée formule de Duhamel. Rappelons la brièvement. On considère l'équation (pour l'instant sans collisions)

$$\begin{cases} \partial_t f(t,x) + v \cdot \nabla_x f(t,x) + \sigma(x) f(t,x) = S(t,x)\,, \quad x \in \mathbf{R}^N\,,\ t > 0\,, \\ \\ f(0,x) = f^{in}(x)\,, \end{cases} \tag{5.66}$$

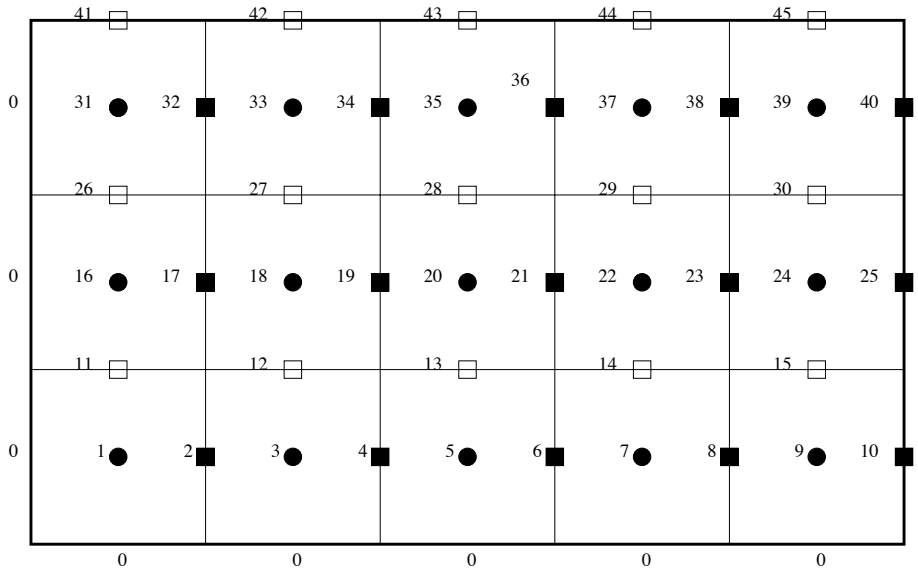

FIGURE 5.4 – Ordre de calcul des inconnues lorsque $\omega_{p,x} > 0$ et $\omega_{p,y} > 0$ avec $N_x = 5$ et $N_y = 3$. Les valeurs $u^p_{j,k}$ aux centres des mailles (marquées par un rond noir) se calculent avec (5.65), les valeurs $u^p_{j,k+1/2}$ sur les arêtes horizontales (marquées par un carré blanc) se calculent avec (5.63), les valeurs $u^p_{j+1/2,k}$ sur les arêtes verticales (marquées par un carré noir) se calculent avec (5.61). Les valeurs 0 sont les conditions aux limites imposées.

qui admet une unique solution $f \in C^1(\mathbf{R}_+ \times \mathbf{R}^N)$, donnée par la formule de Duhamel

$$f(t,x) = f^{in}(x-tv)e^{-\theta(x,x-tv)} + \int_0^t e^{-\theta(x,x-(t-s)v)} S(t-s,x-sv)\,ds\,, \quad (5.67)$$

avec le **trajet optique** θ défini par

$$\theta(x,x-tv) = \int_0^t \sigma(x-sv)\,ds\,.$$

(On n'indique pas la dépendance de la solution f par rapport à v qui est un simple paramètre dans le cas présent.) Par analogie avec la propagation de la lumière le long de ses rayons, le trajet optique $\theta(x,x-tv)$ est une mesure de l'absorption totale entre les points x et $x-tv$. Le trajet optique est toujours positif et, plus il est grand, plus grande est l'atténuation d'une particule partie de $x-tv$ pour aller en x, c'est-à-dire plus grande est la probabilité pour que cette particule ait été absorbée en chemin.

Le formule (5.67) permet de calculer une solution approchée de l'équation (5.66) **sans avoir à utiliser de discrétisation** de l'opérateur de transport. Evidemment, il y a un prix à payer qui est l'évaluation du trajet optique θ et le calcul de l'intégrale du terme source le long de la trajectoire. C'est précisément le principe des **méthodes intégrales** de résolution numérique. Bien sûr, l'équation (5.66) est trop simple, puisqu'elle ne comporte pas de collisions, pour être un exemple représentatif d'application des méthodes intégrales. Mais l'essentiel est là : il faut savoir calculer des trajets optiques et des intégrales de convolution. Cela se fait de manière numérique et nous renvoyons à [38] pour plus de détails.

Remarque 5.3.1 *A cause de cette interprétation en terme de probabilité d'absorption du trajet optique, certaines méthodes intégrales de résolution de l'équation de Boltzmann sont appelées* **méthodes des probabilités de collision**.

Considérons maintenant un cas plus compliqué en présence de collisions. Dans ce cas le second membre $S(t,x)$ de (5.66) n'est plus une donnée mais vaut

$$S(t,x) = \sigma^*(x) \int_{|v'|=1} f(t,x,v')\,dv' + Q(t,x)\,, \quad (5.68)$$

où on a supposé que Q est le terme source et l'espace des vitesses est la sphère unité, pour simplifier. Désormais (5.67) devient (en indiquant à nouveau la dépendance en v)

$$\begin{aligned} f(t,x,v) \;&= \int_0^t e^{-\theta(x,x-(t-s)v)} \sigma^*(x-sv) \int_{|v'|=1} f(t-s,x-sv,v')\,dv'\,ds \\ &\quad + f^{in}(x-tv,v)e^{-\theta(x,x-tv)} + \int_0^t e^{-\theta(x,x-(t-s)v)} Q(t-s,x-sv)\,ds\,. \end{aligned}$$

On intègre cette équation par rapport à v pour faire apparaître comme seule inconnue la moyenne du flux sur toutes les directions angulaires, appelée aussi flux scalaire,

$$\bar{f}(t,x) = \int_{|v|=1} f(t,x,v)\, dv\,,$$

et on obtient

$$\begin{aligned}\bar{f}(t,x) &= \int_0^t \int_{|v|=1} e^{-\theta(x,x-(t-s)v)} \sigma^*(x-sv) \bar{f}(t-s,x-sv)\, dv\, ds \\ &+ \int_{|v|=1} f^{in}(x-tv,v) e^{-\theta(x,x-tv)}\, dv \\ &+ \int_0^t \int_{|v|=1} e^{-\theta(x,x-(t-s)v)} Q(t-s,x-sv)\, ds\, dv\,.\end{aligned} \tag{5.69}$$

qui n'est rien d'autre qu'une formulation intégrale comme on en a vu au Chapitre 3. Remarquons que la double intégrale en (s,v) devient une intégrale en espace sur la boule de centre x et de rayon t. De manière abstraite, (5.69) est équivalent à une équation linéaire

$$\bar{f} = \mathcal{T}\bar{f} + F[f^{in}, Q]\,, \tag{5.70}$$

où $\mathcal{T}$ est un opérateur intégral et $F[f^{in}, Q]$ est un second membre fixé. L'idée des méthodes intégrales est de discrétiser (5.70) et de se ramener ainsi à la résolution d'un système linéaire pour calculer une approximation de la solution exacte $\bar{f}$. Remarquons encore une fois que l'avantage principal de ces méthodes est d'éviter de discrétiser les dérivées partielles de l'opérateur de transport.

Néanmoins, la contrepartie ou l'inconvénient est que la matrice du système linéaire issu de (5.70) est **dense**, c'est-à-dire que tous ses éléments sont non nuls a priori. C'est un contraste fort avec les méthodes de différences finies ou d'éléments finis qui conduisent à des matrices creuses (i.e., avec beaucoup d'éléments nuls), et c'est, bien sûr, très pénalisant pour le stockage en mémoire et les temps de calcul. Dans le cas présent, la matrice est pleine ou dense car l'opérateur $\mathcal{T}$ correspond à une intégrale de volume autour du point x avec un noyau particulier.

En conclusion les méthodes intégrales sont très précises mais coûteuses en temps de calcul. On ne peut les utiliser que pour des problèmes de dimension spatiale réduite. Les méthodes intégrales s'appliquent à de nombreux modèles physiques : nous renvoyons par exemple à [41].

5.3.2 Méthode du flux pair

Nous expliquons brièvement le principe de la méthode du flux pair qui permet d'utiliser tout l'arsenal des formulations variationnelles et des méthodes d'éléments finis comme étudiées dans [2]. On suppose que tous les coefficients, ainsi que le terme source de l'équation de Boltzmann linéaire sont isotropes,

c'est-à-dire indépendants de la variable de vitesse. Par ailleurs, on se place dans un cas stationnaire. Autrement dit, on considère l'équation

$$v \cdot \nabla_x f(x,v) + \sigma(x) f(x,v) - \sigma^*(x) \int_{|v'|=1} f(x,v')dv' = S(x)\,, \tag{5.71}$$

où, pour simplifier, on a pris la sphère unité comme espace des vitesses. On introduit deux nouvelles inconnues : le flux pair défini par

$$f^+(x,v) = \frac{1}{2}\Big(f(x,v) + f(x,-v)\Big)$$

et le flux impair

$$f^-(x,v) = \frac{1}{2}\Big(f(x,v) - f(x,-v)\Big).$$

Bien sûr, on retrouve que

$$f(x,v) = f^+(x,v) + f^-(x,v) \quad \text{ et } f(x,-v) = f^+(x,v) - f^-(x,v).$$

On écrit l'équation (5.71) pour la vitesse $-v$

$$-v \cdot \nabla_x f(x,-v) + \sigma(x) f(x,-v) - \sigma^*(x) \int_{|v'|=1} f(x,v')dv' = S(x)\,. \tag{5.72}$$

Par addition de (5.71) et (5.72) on obtient

$$v \cdot \nabla_x f^-(x,v) + \sigma(x) f^+(x,v) - \sigma^*(x) \int_{|v'|=1} f^+(x,v')dv' = S(x)\,, \tag{5.73}$$

car $\int_{|v'|=1} f dv' = \int_{|v'|=1} f^+ dv'$, tandis que par soustraction

$$v \cdot \nabla_x f^+(x,v) + \sigma(x) f^-(x,v) = 0\,. \tag{5.74}$$

L'équation (5.74) permet de calculer f^- en fonction du courant de f^+ sous l'hypothèse que le coefficient d'absorption $\sigma(x) > 0$ est strictement positif. En reportant dans (5.73) on obtient une équation du deuxième ordre pour f^+

$$-v \cdot \nabla_x \Big(\frac{1}{\sigma(x)} v \cdot \nabla_x f^+(x,v)\Big) + \sigma(x) f^+(x,v) - \sigma^*(x) \int_{|v'|=1} f^+ dv' = S(x)\,. \tag{5.75}$$

Si on note D le domaine spatial, les conditions aux limites de flux nul pour (5.71)

$$f(x,v) = 0 \text{ pour } (x,v) \in \Gamma^- = \{(x,v) \in \partial D \times \{|v'| = 1\} \text{ tel que } v \cdot n < 0\}$$

deviennent

$$0 = f^+(x,v) + f^-(x,v) = f^+(x,v) - \frac{1}{\sigma(x)} v \cdot \nabla_x f^+(x,v) \text{ pour } v \cdot n(x) < 0.$$

Mais la condition aux limites de flux nul dit aussi que $f(x,-v)=0$ pour $v\cdot n(x)>0$ et donc que

$$0=f^+(x,v)-f^-(x,v)=f^+(x,v)+\frac{1}{\sigma(x)}v\cdot\nabla_x f^+(x,v) \text{ pour } v\cdot n(x)>0.$$

On vérifie aisément qu'au total la condition aux limites est équivalente à

$$v\cdot\nabla_x f^+(x,v)+\mathrm{sign}(v\cdot n(x))\,\sigma(x)f^+(x,v)=0 \text{ pour } x\in\partial D. \tag{5.76}$$

On introduit alors la forme bilinéaire symétrique

$$\begin{aligned} a(f,g) &= \int_D\int_{|v|=1}\Big(\frac{1}{\sigma(x)}v\cdot\nabla_x f\; v\cdot\nabla_x g+\sigma(x)fg\Big)dx\,dv \\ &\quad -\int_D\sigma^*(x)\left(\int_{|v|=1}f dv\right)\left(\int_{|v|=1}g dv\right)dx \\ &\quad +\int_{\partial D}\int_{|v|=1}fg|n\cdot v|\,dx\,dv \end{aligned}$$

et la forme linéaire

$$L(g)=\int_D\int_{|v|=1}g(x,v)\,S(x)\,dx\,dv.$$

Lemme 5.3.2 *Soit une fonction régulière $f^+(x,v)$. Alors $f^+(x,v)$ est une solution de l'équation (5.75) avec la condition aux limites (5.76) si et seulement si $f^+(x,v)$ est une solution de la formulation variationnelle,*

$$\text{trouver } f^+\in W \text{ tel que } a(f^+,g)=L(g)\quad\forall g\in W, \tag{5.77}$$

avec l'espace $W=\{g(x,v)\in L^2(D\times\{|v|=1\})$ et $v\cdot\nabla_x g\in L^2(D\times\{|v|=1\})\}$.

Démonstration. On multiplie l'équation (5.75) par une fonction test $g(x,v)$ et on intègre par parties pour obtenir

$$\begin{aligned} &\int_D\int_{|v|=1}\Big(\frac{1}{\sigma(x)}v\cdot\nabla_x f^+\; v\cdot\nabla_x g+\sigma(x)fg\Big)dx\,dv \\ &\qquad -\int_D\sigma^*(x)\left(\int_{|v|=1}f^+dv\right)\left(\int_{|v|=1}g dv\right)dx \\ &-\int_{\partial D}\int_{|v|=1}v\cdot\nabla_x f^+\,g\,n\cdot v\,dx\,dv=\int_D\int_{|v|=1}g(x,v)\,S(x)\,dx\,dv. \end{aligned}$$

Comme d'habitude on a supposé que la mesure dv est normalisée de manière que $\int_{|v|=1}dv=1$. En remplaçant $v\cdot\nabla_x f^+$ par la valeur de la condition aux limites dans l'intégrale de bord, on trouve bien la formulation variationnelle (5.77). Réciproquement, le même calcul en sens inverse permet de passer de la formulation variationnelle à l'équation (5.75) avec la condition aux limites (5.76). ■

A partir de la formulation variationnelle (5.77) il est classique de construire des méthodes d'éléments finis (voir par exemple [2]).

5.3.3 Eléments finis

On peut, comme on l'a déjà dit, utiliser des méthodes d'éléments finis pour la formulation variationnelle en flux pair de l'équation de Boltzmann. Il est aussi possible d'utiliser directement des éléments finis pour l'équation de Boltzmann sous sa forme usuelle. Nous renvoyons pour cela aux travaux de P. Lesaint et P.-A. Raviart [36] et aux ouvrages [17], [30].

5.3.4 Méthode de Monte-Carlo

Les algorithmes probabilistes de type Monte-Carlo sont aussi très populaires pour la résolution de l'équation de Boltzmann. Ils sont basés sur l'interprétation probabiliste de l'équation de Boltzmann, voir la Section 3.3. Nous renvoyons à [26] et [31] pour plus de détails.

5.4 Exercices

Exercice 5.12 (Schémas pour la diffusion) *On passe ici en revue quelques schémas numériques historiques pour la résolution numérique de l'équation de la chaleur*

$$\frac{\partial u}{\partial t}(t,x) - \frac{\partial^2 u}{\partial x^2}(t,x) = 0.$$

On impose des conditions de périodicité au bord de $[0,1]$ *en espace.*

1. *Etudier le schéma semi-discret*

$$u_i'(t) - \frac{u_{i+1}(t) - 2u_i(t) + u_{i-1}(t)}{\Delta x^2} = 0.$$

Montrer qu'il est d'ordre 2 en espace, et stable en norme L^2*. Ce schéma est-il convergent ? Et pourquoi ?*

2. *Etudier le schéma de Richardson*

$$\frac{u_i^{n+1} - u_i^{n-1}}{2\Delta t} - \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\Delta x^2} = 0.$$

Montrer qu'il est d'ordre 2 en espace et 2 en temps. Monrer qu'il est inconditionellement instable.

3. *Etudier le schéma de Dufort et Frankel*

$$\frac{u_i^{n+1} - u_i^{n-1}}{2\Delta t} - \frac{u_{i+1}^n - u_i^{n-1} - u_i^{n+1} + u_{i-1}^n}{\Delta x^2} = 0.$$

Monrer qu'il est inconditionellement stable. Quel est le problème ?

4. *Etudier le schéma de Crank-Nicolson*

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} - \frac{1}{2}\frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\Delta x^2} - \frac{1}{2}\frac{u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}}{\Delta x^2} = 0.$$

Quel est son avantage sur le schéma de Gear qui suit ?

5. *Etudier le schéma de Gear*

$$\frac{3u_i^{n+1} - 4u_i^n + u_i^{n-1}}{2\Delta t} - \frac{u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}}{\Delta x^2} = 0.$$

Indications : pour la stabilité utiliser le critère de Von Neumann.

Pour la question 3, le schéma présente en fait deux problèmes. Le premier est qu'il est à deux pas en temps, et qu'il est donc nécessaire d'avoir une donnée initiale à la fois pour u^0 et u^1. Le deuxième est que, dans l'étude de la consistance de ce schéma, on trouve que l'erreur de troncature contient un terme de l'ordre de $(\Delta t/\Delta x)^2$ qui n'est petit que si $\Delta t \ll \Delta x$ alors même que le schéma est inconditionnellement stable.

En ce qui concerne la question 4, le problème du schéma de Gear est le même que celui du schéma de Dufort-Frankel : il est à deux pas en temps. Un des intérêts du schéma de Crank-Nicolson est d'ailleurs qu'il est d'ordre 2 tout en restant à un seul pas de temps.

Exercice 5.13 (Schéma saute-mouton en transport) *Etudions l'équation du transport libre*

$$\frac{\partial u}{\partial t}(t,x) + a\frac{\partial u}{\partial x}(t,x) = 0.$$

Le schéma "saute-mouton" (ou leap-frog) se définit par

$$\frac{u_i^{n+1} - u_i^{n-1}}{2\Delta t} + a\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} = 0.$$

1. *Montrer que ce schéma vérifie la condition nécessaire de stabilité de Von Neumann dans L^2, sous condition CFL.*
2. *Démontrer que ce schéma est consistant à l'ordre 2.*
3. *Dans le cas $a > 0$, on considère le problème pour $x \in]0,1[$. La condition aux limites est de Dirichlet en $x = 0$, $u(t,0) = 0$. Le pas d'espace est $\Delta x = 1/(N+1)$.*
   Montrer qu'un schéma raisonnable s'écrit

$$\frac{u_i^{n+1} - u_i^{n-1}}{2\Delta t} + a\frac{u_{i+1}^n}{2\Delta x} = 0 \quad i = 1,$$

$$\frac{u_i^{n+1} - u_i^{n-1}}{2\Delta t} + a\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} = 0, \quad 2 \le i \le N,$$

$$\frac{u_i^{n+1} - u_i^{n-1}}{2\Delta t} + a\frac{u_i^n - u_{i-1}^n}{\Delta x} = 0, \quad i = N+1.$$

4. *Montrer que ce schéma est instable au sens de Von Neumann, à cause de la condition au bord à droite.*

Indications : pour la dernière question on écrira le schéma sous forme matricielle et on remarquera que la matrice d'itération possède au moins une valeur propre de partie réelle strictement positive. On construira alors une solution particulière instable à l'aide du vecteur propre correspondant.

Exercice 5.14 (Schéma décentré amont en transport) *Soit l'équation*

$$\frac{\partial u}{\partial t}(t,x) + a\frac{\partial u}{\partial x}(t,x) = 0, \quad a > 0,$$

sur le segment $x \in]0,1[$ avec condition de Dirichlet à gauche, $u(t,0) = 0$.

1. *Ecrire le schéma décentré amont pour cette équation.*
2. *Montrer qu'il est stable dans L^q pour tout $q \geq 1$ sous condition CFL.*

Indication : utiliser l'inégalité de Hölder à partir de l'indication de l'exercice 5.5.

Exercice 5.15 (Discrétisation en angle) *On note $P_n(\mu)$ les polynômes de Legendre, définis par (5.35) et étudiés dans l'exercice 5.10.*

1. *Soit l'équation*

$$\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial x} = \sigma\left(\frac{1}{2}\int_{-1}^{1} f(\mu')d\mu' - f\right).$$

On approche f en tronquant la série

$$f(t,x,\mu) \approx \sum_{n=0}^{N} f_n(t,x)P_n(\mu).$$

Justifier le système

$$\frac{\partial f_n}{\partial t} + \frac{n+1}{2n+1}\frac{\partial f_{n+1}}{\partial x} + \frac{n}{2n+1}\frac{\partial f_{n-1}}{\partial x} = \sigma\left(\delta_{n0} - 1\right) f_n$$

pour $0 \leq n \leq N$. Par convention $f_{-1} = f_{N+1} = 0$.

2. *Montrer l'inégalité*

$$\frac{d}{dt}\left(\sum_{n=1}^{N} \frac{2n+1}{2}\int_{\mathbb{R}} f_n^2(t,x)dx\right) \leq 0.$$

3. *En déduire que*

$$\sum_{n=0}^{N} f_n(t,x)P_n(\mu) \underset{N\to\infty}{\longrightarrow} f$$

dans $L^2(\mathbb{R})$. On fera toutes les hypothèses de régularité nécessaires.

4. *Proposer une méthode numérique pour la résolution de*

$$\frac{\partial X}{\partial t} + A\frac{\partial X}{\partial x} = 0.$$

où $A = A^t$ est une matrice symétrique de taille N. Appliquer à la discrétisation du système de la question 1.

Indication : pour la dernière question : commencer par diagonaliser A, qui s'écrit donc $A = Q^t DQ$, puis étudier l'équation vérifiée par $Y = QX$.

Exercice 5.16 (Limite de diffusion du schéma diamant) *Considérons le schéma diamant en dimension 1 pour l'équation de Boltzmann linéaire stationnaire, avec des vitesses*

$$\mu_k \in [-1,1], \quad 1 \leq k \leq K.$$

1. *Ecrire ce schéma pour un scattering*

$$\overline{u} = \frac{\sigma}{2}\int_{-1}^{1} u(\mu)d\mu$$

et une absorption σu.

2. *On suppose que la source est de l'ordre de ε et que $\sigma = \overline{\sigma}\varepsilon^{-1}$. Montrer que la limite de diffusion est correcte (ou autrement dit que le coefficient de diffusion du schéma vaut $\frac{1}{3\overline{\sigma}}$) dès que les relations*

$$\sum_k \omega_k \mu_k = 0 \ \textit{et} \ \sum_k \omega_k \mu_k^2 = \frac{2}{3}$$

sont vérifiées.

Indications : on se bornera à utiliser un développement de Hilbert pour le schéma diamant et à vérifier que, formellement, la solution converge bien vers celle d'un schéma de diffusion. Ceci revient à adapter au cas discret les méthodes du cas continu développées dans la section 4.2.2. Voir également l'exercice 4.2.

Exercice 5.17 (Accélération par la diffusion) *On analyse le schéma explicite en 1D*

$$\mu_k \frac{u^{k,n}_{j+\frac{1}{2}} - u^{k,n}_{j-\frac{1}{2}}}{\Delta x} + \sigma \frac{u^{k,n}_{j+\frac{1}{2}} + u^{k,n}_{j-\frac{1}{2}}}{2} = \sigma^\star \overline{u}^{n-1}_j + f_j \tag{5.78}$$

avec la relation diamant et avec seulement deux vitesses

$$\mu_k = \pm 1.$$

1. *Ecrire la relation diamant. Que vaut $\overline{u}^{n-1}_j$?*

2. *Analyser le schéma en modes de Fourier. On cherchera la solution sous la forme*

$$\begin{pmatrix} u^{k=1,n}_{j+\frac{1}{2}} \\ u^{k=2,n}_{j+\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} \alpha_n \\ \beta_n \end{pmatrix} e^{i\theta(j+\frac{1}{2})\Delta x} \quad \forall n \geq 0.$$

Déterminer la relation de récurrence sur les (α_n, β_n).

3. *Trouver les valeurs propres de la matrice d'itération et retrouver le fait que l'algorithme converge dès que $\frac{\sigma^\star}{\sigma} < 1$, mais que cette convergence est très lente si $\frac{\sigma^\star}{\sigma}$ est très proche de 1.*

4. *Pour accélérer la convergence, on utilise la méthode d'accélération par la diffusion, ou diffusion synthétique. Pour cela, écrire le schéma sous la forme*

$$Tu^n = K\overline{u}^{n-1} + f,$$

en explicitant les matrices T et K. La"limite de diffusion" du schéma s'écrit (en omettant l'itération sur les sources)

$$-\frac{1}{\sigma\Delta x^2}\left(u_{j+1} - 2u_j - u_{j-1}\right) + \sigma u_j = \sigma^\star u_j + f_j.$$

L'écrire également sous la forme :

$$Du = f.$$

5. *On considère maintenant le schéma*

$$\left\{\begin{array}{l} Tv^n = K\overline{u}^{n-1} + f, \\ D\overline{u}^n = D\overline{v}^n + K\left(\overline{v}^n - \overline{u}^{n-1}\right). \end{array}\right.$$

Faire le même travail en Fourier, et vérifier que cet algorithme est beaucoup plus rapide que l'algorithme initial (5.78).

Exercice 5.18 (Méthode du flux pair) *(exercice très complet qui reprend la section correspondante) Soit $\varphi(x,v)$ la solution de l'équation du transport*

$$v\cdot\nabla\varphi + \sigma\varphi = \sigma_s\phi + S$$

avec $\sigma \geq \sigma_s > 0$. La source est $S(x)$. Le domaine est $(x,v) \in \mathcal{C}\times\mathbf{S}^{N-1}$ où $\mathcal{C}\subset\mathbf{R}^N$ est un ouvert borné et $\mathbf{S}^{N-1}$ est la sphère unité dans l'espace des directions $|v|=1$. Par définition

$$\phi(x) = \frac{1}{|\mathbf{S}^{N-1}|}\int_{\mathbf{S}^{N-1}}\varphi(x,v)\,dv, \quad \left|\mathbf{S}^{N-1}\right| = \int_{\mathbf{S}^{N-1}} dv.$$

Le bord se décompose en deux parties $\partial\mathcal{C} = \overline{\Gamma_n\cup\Gamma_r}$. Soit $\mathbf{n}$ la normale extérieure. Les conditions au bord sont du type neutre (ou flux entrant nul) sur Γ_n

$$\varphi(x,v) = 0 \quad x\in\Gamma_n, \quad \mathbf{n}\cdot v < 0,$$

et réflexion sur Γ_r

$$\varphi(x,v) = \varphi(x,v') \quad x\in\Gamma_r, \quad v' = v - 2\left(v\cdot\mathbf{n}\right)\mathbf{n}.$$

1. *Faire un dessin et interpréter les conditions au bord.*
2. *On pose*

$$\varphi^{\pm}(x,v) = \frac{1}{2}\left(\varphi(x,v)\pm\varphi(x,-v)\right).$$

Montrer que φ^+ est pair en v et que φ^- est impair en v. Montrer que

$$\int_{\mathbf{S}^{N-1}} v\varphi^+ dv = 0.$$

3. *Montrer que* $\phi^- = 0$ *et*

$$\phi = \phi^+ \equiv \frac{1}{|\mathbf{S}^{N-1}|}\int_{\mathbf{S}^{N-1}} \varphi^+ dv.$$

4. *Montrer que le couple* (φ^+, φ^-) *est solution du système du premier ordre*

$$\left\{ \begin{array}{l} v\cdot\nabla\varphi^- + \sigma\varphi^+ = \sigma_s\phi + S, \\ v\cdot\nabla\varphi^+ + \sigma\varphi^- = 0. \end{array} \right.$$

5. *En déduire que* φ^+ *est solution de l'équation du second ordre*

$$-v\cdot\nabla\left(\frac{1}{\sigma}v\cdot\nabla\varphi^+\right) + \sigma\varphi^+ = \sigma_s\phi^+ + S.$$

6. *Montrer que les conditions aux bords peuvent s'écrire sur le bord neutre*

$$v\cdot\nabla\varphi^+ \pm \sigma\varphi^+ = 0, \quad x\in\Gamma_n, \quad \pm\mathbf{n}\cdot v > 0,$$

et sur le bord réflexif

$$\varphi^+(x,v) = \varphi^+(x,v') \quad x\in\Gamma_r, \quad v' = v - 2\,(v\cdot\mathbf{n})\,\mathbf{n}.$$

Exercice 5.19 (Forme variationnelle du flux pair) *(difficile) On reprend l'exercice 5.18, et on pose*

$$J(\varphi^+) = \frac{1}{2}\int_{\mathcal{C}}\int_{\mathbf{S}^{N-1}}\left(\frac{1}{\sigma}\left(v\cdot\nabla\varphi^+\right)^2 + (\sigma-\sigma_s)\left(\varphi^+\right)^2\right) + \frac{1}{2}\int_{\Gamma_v}\int_{\mathbf{S}^{N-1}}\left(\varphi^+\right)^2|\mathbf{n}\cdot v| - \int_{\mathcal{C}}\int_{\mathbf{S}^{N-1}} S\varphi^+. \quad (5.79)$$

1. *Montrer formellement que* $J(\varphi^+) \le J(\widetilde{\varphi^+})$ *pour toute fonction test* $\widetilde{\varphi^+}$ *suffisamment régulière.*
2. *Définir la forme bilinéaire* $a(\varphi^+, \widetilde{\varphi^+})$ *et la forme linéaire* $b(\widetilde{\varphi^+})$ *associées, et montrer que la solution* φ^+ *est solution du problème variationnel* $a(\varphi^+, \widetilde{\varphi^+}) = b(\widetilde{\varphi^+})$ *pour tout* $\widetilde{\varphi^+}$ *dans un espace de fonctions suffisamment régulières.*
3. *Faire le lien avec la théorie variationnelle des équations elliptiques (voir [2] par exemple). Pour cela on considérera le cas monodimensionel* $N = 1$ *et on posera* $u(x) = \varphi^+(x,1)$*. Ecrire le problème pour* u*. Montrer que l'espace* $H^1(\mathcal{C})$ *est le bon cadre fonctionel. En déduire l'existence et l'unicité de* $u \in H^1(\mathcal{C})$*.*

Exercice 5.20 (Méthode du flux pair : analyse numérique en 1D) *On reprend les exercices 5.18 et 5.19, et on discrétise la formulation variationnelle* (5.79) *en éléments finis.*

1. *Rappeler la définition des fonctions chapeaux ($\mathbb{P}_1$) en dimension $N = 1$. Ecrire la formulation variationnelle discrète.*
2. *Montrer l'existence et l'unicité de la solution variationnelle discrète.*

Exercice 5.21 (Méthode du flux pair : approximation de diffusion) *(exercice de modélisation dont l'hypothèse de départ constitue un raccourci souvent utilisé dans l'art de l'ingénieur) Dans le cadre de l'exercice 5.18 nous allons retrouver l'approximation de diffusion en admettant que le flux φ dépend linéairement de la variable angulaire v*

$$\varphi(x, v) = \phi(x) + v \cdot J(x).$$

1. *Montrer que $\varphi^+ = \phi$. Ecrire la formulation variationnelle correspondante dont l'inconnue est la fonction ϕ. On rappelle que*

$$\frac{1}{|\mathbf{S}^{N-1}|} \int_{\mathbf{S}^{N-1}} v \otimes v \, dv = \frac{1}{N} I_N.$$

2. *Montrer que cette formulation variationnelle est bien posée dans $H^1(\mathcal{C})$.*
3. *On considère une base d'éléments finis $\mathbb{P}^1$ en espace, notée $(\chi_j)_{1 \leq j \leq P}$. Ecrire l'approximation de Galerkin, dans l'espace engendré par ces éléments finis, de la formulation variationnelle de la question précédente. Démontrer que ce problème admet une unique solution.*

Chapitre 6

Théorie du calcul critique

6.1 Comportement asymptotique en temps

6.1.1 Position du problème

Nous étudions le comportement en temps grand d'une équation de Boltzmann que, pour simplifier, nous choisissons sous la forme

$$\begin{cases} \dfrac{\partial \phi}{\partial t} + v \cdot \nabla \phi + \sigma(x)\phi = \displaystyle\int_{|v'|=1} \sigma^*(x, v \cdot v')\phi(x, v')\, dv' \\ \phi(t = 0, x, v) = \phi^0(x, v) \end{cases} \tag{6.1}$$

en l'absence de terme source. La donnée initiale est une fonction positive $\phi^0 \geq 0$ puisqu'elle représente une densité de particules. On se pose la question de savoir quelle est la limite, lorsque le temps t tend vers $+\infty$, de la solution $\phi(t, x, v)$. Plus précisément, nous voulons pouvoir décider dans lequel des trois cas suivants nous pouvons nous trouver :

1. la solution $\phi(t, x, v)$ converge vers 0, ce qui correspond à l'extinction de la population de particules,
2. la solution $\phi(t, x, v)$ converge vers $+\infty$, ce qui correspond à "l'explosion" de la population de particules,
3. la solution $\phi(t, x, v)$ converge vers une limite finie non nulle $\phi^\infty(x, v)$, ce qui correspond à l'existence d'un état stationnaire de la population de particules.

Le dernier cas est dit **critique**, le premier **sous-critique** tandis que le second est **sur-critique**. Nous restons volontairement vague sur la notion de convergence utilisée ci-dessus (pour quelle norme ?) et nous remarquons qu'a priori d'autres cas pourraient être possibles : par exemple, la solution n'admet aucune limite ou bien elle converge vers un cycle limite périodique en temps. Il se trouve que, sous des hypothèses adéquates, on peut se limiter au trois cas ci-dessus qui sont les seuls possibles.

Remarque 6.1.1 *Le même genre de questions se pose aussi pour une ou des équations de diffusion. La notion ci-dessus de criticité se retrouve dans de nombreux domaines d'applications : en neutronique [12], [45], comme en dynamique des populations [39].*

L'obtention de résultats mathématiquement rigoureux sur le comportement asymptotique en temps grand de (6.1) est d'un niveau technique qui dépasse celui de ce cours, sauf dans le cas particulier d'une dimension d'espace (voir la Section 6.4). C'est pourquoi nous allons introduire un problème analogue en dimension finie, pour lequel nous allons donner une théorie assez complète. Par la suite, nous admettrons des résultats équivalents pour les équations de transport ou de diffusion.

6.1.2 Analogie en dimension finie

Pour comprendre comment étudier le comportement asymptotique de l'équation de Boltzmann (6.1) nous allons faire une étude préalable d'un modèle, beaucoup plus simple, d'équations différentielles ordinaires. A chaque instant t l'inconnue, au lieu d'être une fonction de (x, v), sera un vecteur dans $\mathbf{R}^n$: il s'agit donc d'une approximation en dimension finie. On peut toujours se ramener à ce cas par discrétisation de l'espace des phases, i.e., du domaine des points (x, v). Par analogie, le comportement asymptotique en temps grand de ce modèle simplifié nous permettra de comprendre ce qui va se passer en dimension infinie, c'est-à-dire pour l'équation de Boltzmann (6.1).

On considère donc une fonction du temps t à valeurs vectorielles, $u(t) \in C^1(\mathbf{R}^+ : \mathbf{R}^n)$, solution de l'équation différentielle linéaire

$$\left\{ \begin{array}{l} \dfrac{du}{dt} + Au = 0, \\ u(t=0) = u^0, \end{array} \right. \tag{6.2}$$

où A est une matrice réelle de taille $n \times n$.

Supposons dans un premier temps que la matrice A soit diagonalisable sur $\mathbf{R}$, et notons λ_k ses valeurs propres, r_k ses vecteurs propres et l_k les vecteurs propres adjoints, normalisés, correspondant à sa transposée ou adjointe A^*, $1 \leq k \leq n$. Autrement dit,

$$Ar_k = \lambda_k r_k, \quad A^* l_k = \lambda_k l_k, \quad r_k \cdot l_j = \delta_{jk}$$

où δ_{jk} est le symbole de Kronecker. On choisit de plus d'ordonner les valeurs propres (répétées avec leur multiplicité) par ordre croissant

$$\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n.$$

La solution exacte de (6.2) est

$$u(t) = \sum_{k=1}^{n} e^{-\lambda_k t} (u^0 \cdot l_k)\, r_k \,.$$

Le comportement asymptotique de cette solution est clair et suit le signe des valeurs propres. Nous ne ferons pas l'injure au lecteur de démontrer le résultat suivant.

Lemme 6.1.2 *Supposons que A est diagonalisable sur $\mathbf{R}$ et que $u^0 \cdot l_1 \neq 0$. La solution de l'équation différentielle ordinaire (6.2) admet une limite finie non nulle quand t tend vers l'infini si et seulement si la plus petite valeur propre vérifie $\lambda_1 = 0$.*

Malheureusement en pratique il est difficile de savoir si une matrice est diagonalisable sur $\mathbf{R}$, sauf si elle est normale (i.e. $AA^* = A^*A$) ou plus simplement symétrique réelle. Or, les équations de Boltzmann, mais aussi les systèmes d'équations de diffusion, conduisent, lorsqu'on les discrétise, à des matrices réelles non symétriques, ni mêmes normales. On ne peut donc pas se contenter du Lemme 6.1.2. C'est pourquoi nous allons introduire un autre cas particulier de matrices qui peuvent paraître très spécifiques au premier abord mais qui sont en fait assez fréquentes dans la pratique comme nous le montrera la théorie de Perron-Frobenius que nous développerons dans la section suivante.

Lemme 6.1.3 *Soit λ_k, $1 \leq k \leq n$, les valeurs propres (éventuellement complexes) d'une matrice réelle A. On suppose qu'il existe une valeur propre, disons λ_1, qui soit réelle, simple et qui vérifie*

$$\lambda_1 < \mathcal{R}(\lambda_k) \quad \text{pour tout } k \neq 1, \tag{6.3}$$

où $\mathcal{R}$ désigne la partie réelle. De plus, on fait l'hypothèse que $u^0 \cdot l_1 \neq 0$ avec l_1 le vecteur propre adjoint associé à λ_1. Alors la solution de l'équation différentielle ordinaire (6.2) admet une limite finie non nulle quand t tend vers l'infini si et seulement si on a $\lambda_1 = 0$.

Remarque 6.1.4 *Rappelons qu'une valeur propre d'une matrice est dite simple si sa multiplicité comme racine du polynôme caractéristique est un.*

Démonstration. La matrice A est semblable à sa forme de Jordan qui est une matrice diagonale par blocs avec des blocs du type

$$D_m = \begin{pmatrix} \lambda_m & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_m \end{pmatrix},$$

où les valeurs propres $(\lambda_m)_{1 \leq m \leq M}$ forment un sous-ensemble maximal de toutes les valeurs propres admettant des vecteurs propres indépendants. Dans la base de la forme de Jordan l'équation (6.2) est équivalente aux équations découplées

$$\begin{cases} \dfrac{dv_m}{dt} + D_m v_m = 0, \\ v_m(t=0) = v_m^0, \end{cases}$$

On peut calculer de manière explicite la solution de ce dernier système : chaque composante de la solution est le produit de l'exponentielle $e^{-\lambda_m t}$ et d'un polynôme en t de degré inférieur ou égal à $(\dim D_m - 1)$. A cause de l'hypothèse (6.3) et de la simplicité de λ_1 (qui signifie que $\dim D_1 = 1$) tous les autres termes sont négligeables devant $v_1(t) = v_1^0 e^{-\lambda_1 t}$ pour t tendant vers l'infini. D'où le résultat. ■

Remarque 6.1.5 *Dans la section suivante nous verrons que l'hypothèse $u^0 \cdot l_1 \neq 0$ n'est pas restrictive pour une large classe de matrices lorsque toutes les composantes de u_0 sont positives et $u_0 \neq 0$. En effet, pour ces matrices les composantes de l_1 seront toutes strictement positives et on aura donc $u^0 \cdot l_1 > 0$.*

Pour revenir aux modéles de diffusion ou de transport, comme (6.1), la détermination de la première valeur propre et fonction propre est appelée **calcul critique**. Le terme critique correspond à l'existence d'un état stationnaire non nul quand t tend vers l'infini. Si la limite en temps grand est nulle on parle de problème sous-critique, tandis que si la limite est infinie il s'agit d'un problème sur-critique.

6.2 M-matrices et théorie de Perron-Frobenius

Cette section présente la théorie de Perron-Frobenius en algèbre matricielle qui joue un rôle important dans de nombreux domaines mathématiques (par exemple, dans l'étude des chaînes de Markov [7] ou autres processus de branchement [39]) et qui, dans le cadre de ce cours, permet de donner un sens à la notion de calcul critique. Toutes les matrices considérées ci-dessous sont réelles et de taille $n \times n$.

6.2.1 M-matrices

Un ouvrage de référence sur les M-matrices est [10] dont nous nous inspirons largement (voir aussi [34], [50]).

Définition 6.2.1 *On dit que A est une M-matrice si elle s'écrit*

$$A = \begin{pmatrix} a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & -a_{n-1n} \\ -a_{n1} & \cdots & -a_{nn-1} & a_{nn} \end{pmatrix} \tag{6.4}$$

avec des coefficients $a_{ij} \geq 0$ positifs ou nuls tels que

$$a_{ii} - \sum_{j=1, j\neq i}^{n} a_{ij} \geq 0\,, \quad \text{pour tout } 1 \leq i \leq n. \tag{6.5}$$

Si l'inégalité (6.5) est stricte pour tout i, on dit que A est une M-matrice stricte.

Définition 6.2.2 *On dit qu'une matrice B est positive si tous ses coefficients sont positifs ou nuls, $b_{ij} \geq 0$. On dit qu'elle est stritement positive s'ils sont strictement positifs, $b_{ij} > 0$.*

Il faut faire attention que la terminologie de la Définition 6.2.2 est différente de la notion de positivité d'une matrice au sens des formes quadratiques.

Exercice 6.1 *Montrer que, si A est une M-matrice, alors il existe une matrice positive B et un réel $c \geq \rho(B)$ tels que $A = c\,\mathrm{Id} - B$ (avec $\rho(B)$ le rayon spectral de B). Indication : choisir $c \geq \max_{1\leq i\leq n} a_{ii} \geq 0$.*

Dans la littérature on appelle parfois M-matrice une matrice de la forme (6.4) telle que $A = c\,\mathrm{Id} - B$ avec $c \geq \rho(B)$ et B une matrice positive. Dans ce cas, la Définition 6.2.1 est un cas particulier de M-matrice.

Exercice 6.2 *On considère l'équation de diffusion suivante*

$$\begin{cases} -\nu \dfrac{\partial^2 u}{\partial x^2} + \sigma(x)u = f(x) \text{ pour } x \in (0,1) \\ u(0) = u(1) = 0, \end{cases}$$

avec l'hypothèse que $\sigma(x) \geq \sigma_0 > 0$ pour tout $x \in (0,1)$. Montrer que la matrice issue de la discrétisation par différences finies de cette équation (voir la Section 5.1.1) est une M-matrice stricte.

Exercice 6.3 *Pour une vitesse donnée $\mu > 0$, on considère l'équation de transport suivante*

$$\begin{cases} \mu \dfrac{\partial u}{\partial x} + \sigma(x)u = f(x) \text{ pour } x \in (0,1) \\ u(0) = 0, \end{cases}$$

avec l'hypothèse que $\sigma(x) \geq \sigma_0 > 0$ pour tout $x \in (0,1)$. Montrer que la matrice issue de la discrétisation par différences finies de cette équation (pour le schéma décentré amont comme pour le schéma diamant) est une M-matrice stricte.

Lemme 6.2.3 *Toute M-matrice stricte est inversible.*

Démonstration. Soit $x \in \mathbf{R}^n$ un vecteur tel que $Ax = 0$. On note i un indice tel que $|x_i| = \max_{1\leq j\leq n} |x_j|$. Si on suppose que $|x_i| > 0$, alors

$$a_{ii}|x_i| \leq \sum_{j\neq i} a_{ij}|x_j| \leq \sum_{j\neq i} a_{ij}|x_i| < a_{ii}|x_i|,$$

ce qui est une contradiction. Donc $x = 0$ et A est inversible. ■

Définition 6.2.4 *On dit qu'une matrice A est irréductible s'il n'existe pas de matrice de permutation P telle que PAP^* se mette sous forme trianglaire par blocs*

$$PAP^* = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}. \tag{6.6}$$

Lemme 6.2.5 *A toute matrice A on associe le graphe de nœuds $1, 2, ..., n$ et d'arêtes orientées reliant i à j si $a_{ij} \neq 0$. Alors A est irréductible si et seulement si pour tout couple $i \neq j$ il existe une chaîne d'arêtes qui permet d'aller de i à j,*

$$a_{ik_1} \neq 0 \to a_{k_1k_2} \neq 0 \to ... \to a_{k_mj} \neq 0.$$

Démonstration. Si A n'est pas irréductible, alors il existe une matrice de permutation P telle que PAP^* ait la forme (6.6). Autrement dit, quitte à renuméroter les nœuds du graphe (c'est l'effet de la multiplication à gauche par P et à droite par son adjoint), les premiers nœuds $1 \leq i \leq \dim A_{11}$ ne sont pas reliés par une chaîne d'arêtes aux derniers nœuds $\dim A_{11} + 1 \leq j \leq n$. Réciproquement, supposons qu'il existe un couple d'indices $i_0 \neq j_0$ qui ne soient reliés par aucune chaîne d'arêtes. On définit les ensembles

$$\begin{aligned} &I = \{i \in \{1, ..., n\} \text{ tel que } i_0 \text{ est relié à } i\}, \\ &J = \{j \in \{1, ..., n\} \text{ tel que } j \text{ est relié à } j_0\}, \\ &K = \{1, ..., n\} \setminus (I \cup J). \end{aligned}$$

L'intersection $I \cap J$ est vide sinon i_0 serait relié à j_0. De même, $I \cap K = \emptyset$. Par conséquent, si on applique la permutation qui numérote en premier les éléments de I puis ceux de $J \cup K$, on obtient une matrice trianglaire par blocs de la forme (6.6). Donc A n'est pas irréductible. ■

Exercice 6.4 *Vérifiez que la matrice de l'Exercice 6.2 est irréductible, y compris avec l'hypothèse plus faible que $\sigma(x) \geq 0$ pour tout $x \in (0, 1)$. Montrer, au contraire, que la matrice de l'Exercice 6.3 n'est pas irréductible.*

Exercice 6.5 *Montrer que la transposée ou adjointe d'une matrice irréductible est aussi irréductible.*

Lemme 6.2.6 *Soit A une M-matrice irréductible telle qu'il existe un indice i_0 pour lequel*

$$a_{i_0i_0} - \sum_{j=1, j\neq i_0}^{n} a_{i_0j} > 0.$$

Alors A est inversible.

Démonstration. Soit $x \in \mathbf{R}^n$ un vecteur tel que $Ax = 0$. On note i un indice tel que $|x_i| = \max_{1\leq j\leq n} |x_j|$. On a

$$a_{ii}|x_i| \leq \sum_{j\neq i} a_{ij}|x_j| \leq \sum_{j\neq i} a_{ij}|x_i| \leq a_{ii}|x_i|,$$

ce qui implique que chacune de ces inégalités est en fait une égalité. En particulier, on a $|x_j| = |x_i|$ pour tous les indices j tels que $a_{ij} \neq 0$. Soit la chaîne d'indices qui relie i à i_0 : le long de cette chaîne on a $a_{jk} \neq 0$ et $|x_j| = |x_k| = |x_i|$. Par conséquent, $|x_{i_0}| = \max_{1\leq j\leq n} |x_j|$ et on peut conclure comme dans la démonstration du Lemme 6.2.3 que $x = 0$ et A est inversible. ■

Lemme 6.2.7 *Soit A une M-matrice inversible irréductible. Alors son inverse A^{-1} est strictement positive au sens où tous ses coefficients sont strictement positifs.*

Démonstration. Soit $x \in \mathbf{R}^n$ le k-ème vecteur colonne de A^{-1} qui vérifie donc $Ax = e_k$ avec e_k le k-ème vecteur de la base canonique. On note i un indice tel que $x_i = \min_{1 \leq j \leq n} x_j$. Supposons que $x_i \leq 0$. On a

$$0 \leq a_{ii}x_i - \sum_{j \neq i} a_{ij}x_j \leq \left(a_{ii} - \sum_{j \neq i} a_{ij} \right) x_i \leq 0, \tag{6.7}$$

ce qui implique que chacune de ces inégalités est en fait une égalité. En particulier, on a $x_j = x_i$ pour tous les indices j tels que $a_{ij} \neq 0$. Grâce à l'irréductibilité de A on en déduit que $x_k = \min_{1 \leq j \leq n} x_j$. On reprend alors l'inégalité (6.7) pour $i = k$ mais, dans ce cas, le terme le plus à gauche vaut 1 comme la k-ème composante de e_k, ce qui est une contradiction. Par conséquent $x_i > 0$. ■

Remarque 6.2.8 *L'hypothèse d'irréductibilité de la matrice A est essentielle dans le Lemme 6.2.7 (penser au cas $A =$ Id).*

Exercice 6.6 *Montrer, à l'aide de l'Exercice 6.1, que la partie réelle de chaque valeur propre d'une M-matrice stricte est strictement positive. Montrer le même résultat avec une inégalité large pour une M-matrice.*

6.2.2 Théorème de Perron-Frobenius

Nous suivons la présentation de [34].

Théorème 6.2.9 (Perron-Frobenius) *Soit K une matrice strictement positive au sens où tous ses coefficients sont strictement positifs. Alors K a une valeur propre dominante λ_{max} qui vérifie les propriétés suivantes.*

1. *$\lambda_{max} > 0$ et un vecteur propre associé x (tel que $Kx = \lambda_{max}x$) a toutes ses composantes strictement positives.*
2. *λ_{max} est simple (sa multiplicité comme racine du polynôme caractéristique est un).*
3. *Toute autre valeur propre λ de K vérifie $|\lambda| < \lambda_{max}$.*
4. *La matrice K n'a pas d'autre vecteur propre dont toutes les composantes sont positives.*

Pour démontrer le Théorème 6.2.9 nous avons besoin du lemme suivant. Rappelons que, si x est un vecteur de $\mathbf{R}^n$, on note $x \geq 0$ pour dire que toutes ses composantes sont positives, $x_i \geq 0$ pour $1 \leq i \leq n$. De même $x \geq y$ si $x - y \geq 0$.

Lemme 6.2.10 *Pour une matrice K on désigne par $p(K)$ l'ensemble des réels positifs ou nuls $\lambda \geq 0$ tels qu'il existe un vecteur non nul $x \geq 0$ vérifiant*

$$Kx \geq \lambda x\,. \tag{6.8}$$

Si K est une matrice strictement positive, l'ensemble $p(K)$ est fermé, borné, non vide et contient un nombre strictement positif.

Démonstration. Soit $x > 0$ un vecteur à composantes strictement positives. Alors $Kx > 0$ est aussi à composantes strictement positives et, pour $\lambda > 0$ suffisamment petit, l'inégalité (6.8) a lieu, ce qui prouve que $p(K)$ contient au moins un nombre strictement positif. Soit $\mathbf{1}$ le vecteur de composantes toutes égales à 1. En prenant le produit scalaire de (6.8) avec $\mathbf{1}$, sachant que $x \geq 0$, on obtient

$$\lambda \leq \frac{x \cdot K^*\mathbf{1}}{\sum_{i=1}^n x_i} \leq \max_{1\leq i\leq n} (K^*\mathbf{1})_i,$$

ce qui prouve que $p(K)$ est borné. Si on prend une suite $\lambda_n \in p(K)$ qui converge vers une limite $\lambda \in \mathbf{R}^+$, on peut lui associer une suite de vecteurs x_n qui vérifient (6.8). Quitte à les normaliser par la condition $\mathbf{1}\cdot x_n = 1$, on peut en extraire une sous-suite qui converge vers $x \geq 0$, non nul car vérifiant la même condition de normalisation. On passe facilement à la limite dans (6.8), ce qui implique que $\lambda \in p(K)$ qui est donc fermé. ■

Démonstration du Théorème 6.2.9. Grâce au Lemme 6.2.10, $p(K)$, étant fermé borné non vide, admet un plus grand élément $\lambda_{max} = \max p(K) > 0$ dont nous allons montrer qu'il est une valeur propre de K. Puisque $\lambda_{max} \in p(K)$ il existe un vecteur non nul $y \geq 0$ tel que $Ky \geq \lambda_{max} y$. Montrons que cette inégalité est en fait une égalité et donc que y est un vecteur propre. Supposons qu'il existe un indice k tel que

$$\sum_{j=1}^n K_{kj} y_j > \lambda_{max} y_k \quad \text{et} \quad \sum_{j=1}^n K_{ij} y_j \geq \lambda_{max} y_i \quad \text{pour } i \neq k\,.$$

Pour $\epsilon > 0$ on définit le vecteur $x = y + \epsilon e_k$ où e_k est le k-ème vecteur de la base canonique. Comme K est strictement positive, on a $Kx > Ky$ tandis que seule la k-ème composante de x diffère de y. Par conséquent, pour $\epsilon > 0$ suffisamment petit, on en déduit une inégalité stricte

$$Kx > \lambda_{max} x\,.$$

On peut donc augmenter légèrement λ_{max} dans cette inégalité, ce qui contredit le caractère maximal de $\lambda_{max} = \max p(K)$. Ainsi, $Ky = \lambda_{max} y$ et λ_{max} est bien une valeur propre de K. De plus, y a toutes ses composantes strictement positives car K est strictement positive, $y \geq 0$ et $y = Ky/\lambda_{max}$. Cela termine la preuve du point 1.

Démontrons maintenant le point 2. Commençons par montrer que la valeur propre λ_{max} est géométriquement simple, c'est-à-dire qu'il n'y a pas d'autres

vecteurs propres que des multiples de y. Supposons qu'il en exite un autre z. Comme z n'est pas proportionnel à y, on peut trouver une constante c suffisamment petite telle que $y + cz$ (qui est toujours un vecteur propre de K pour la valeur propre λ_{max}) est positif, $y + cz \geq 0$, non nul mais a au moins une composante nulle. Ceci contredit notre argumentation précédente qui montrait que tout vecteur propre positif associé à λ_{max} est en fait strictement positif. Montrons ensuite que λ_{max} est algébriquement simple, c'est-à-dire qu'elle est racine simple du polynôme caractéristique de K. Si cela n'était pas le cas, au vu de la forme de Jordan de la matrice K, il existerait un vecteur non nul z tel que

$$Kz = \lambda_{max} z + c\, y\,,$$

où, quitte à changer z en $-z$, on peut supposer $c > 0$, et quitte à changer z en $z + d\, y$ avec $d > 0$, on peut aussi supposer $z \geq 0$. On en déduit l'inégalité

$$Kz > \lambda_{max} z$$

et on peut donc augmenter un peu λ_{max} en préservant cette inégalité, ce qui contredit encore le caractère maximal de λ_{max}.

Vérifions le point 3. Soit une valeur propre $\lambda \in \mathbf{C}$ et un vecteur propre non nul $z \in \mathbf{C}^n$ vérifiant $Kz = \lambda z$. Comme K est positive, on a

$$|\lambda|\,|z_i| \leq \sum_{j=1}^{n} K_{ij} |z_j| \quad \text{pour tout } 1 \leq i \leq n\,, \tag{6.9}$$

ce qui n'est rien d'autre que (6.8) pour le vecteur z^+ de composantes $|z_i|$. Ainsi $|\lambda|$ appartient à l'ensemble $p(K)$ et $|\lambda| \leq \lambda_{max}$. Cette inégalité est stricte car sinon z^+ serait proportionnel à y et l'inégalité (6.9) serait en fait une égalité. Or, on ne peut avoir

$$\left| \sum_{j=1}^{n} K_{ij} z_j \right| = \sum_{j=1}^{n} K_{ij} |z_j|$$

que s'il existe un unique nombre $\theta \in \mathbf{R}$ tel que

$$z_j = e^{i\theta} |z_j| \quad \text{pour tout } 1 \leq j \leq n\,.$$

Autrement dit, z serait proportionnel à y. Donc, pour tout valeur propre $\lambda \neq \lambda_{max}$ on a bien $|\lambda| < \lambda_{max}$.

Terminons par le point 4. Supposons que K ait un autre vecteur propre $z \neq 0$, réel à composantes positives, pour une autre valeur propre λ (forcément réelle) différente de λ_{max}. On sait que K et sa transposée (ou adjointe) K^* ont les mêmes valeurs propres. En particulier, puisque K^* est aussi une matrice strictement positive, elle admet λ_{max} (le même que pour K) comme valeur propre dominante avec un vecteur propre à composantes strictement positives y. Or les vecteurs propres z et y sont orthogonaux car

$$\lambda z \cdot y = Kz \cdot y = z \cdot K^* y = \lambda_{max} z \cdot y \quad \text{et } \lambda \neq \lambda_{max}.$$

Mais comme toutes les composantes de y sont trictement positives, on ne peut pas avoir $z \cdot y = 0$. Ainsi, il n'y a pas d'autre vecteur propre de K à composantes positives. ■

Nous pouvons maintenant combiner les résultats sur les M-matrices et le théorème de Perron-Frobenius pour obtenir le théorème principal de cette section.

Théorème 6.2.11 *Soit A une M-matrice irréductible. Alors A a une plus petite valeur propre λ_{min} qui vérifie les propriétés suivantes.*

1. *λ_{min} est réelle et simple.*
2. *Son vecteur propre associé x (tel que $Ax = \lambda_{min}x$) a toutes ses composantes strictement positives.*
3. *La matrice A n'a pas d'autre vecteur propre dont toutes les composantes sont positives.*
4. *Toute autre valeur propre $\lambda \in \mathbf{C}$ de A vérifie $\lambda_{min} < \mathcal{R}(\lambda)$.*

Pour démontrer ce théorème nous utilisons le résultat intermédiaire suivant.

Lemme 6.2.12 *Soit B une matrice irréductible positive, dont les coefficients diagonaux sont strictement positifs, autrement dit*

$$b_{ii} > 0 \ \forall\, i\,, \ \text{ et } b_{ij} \geq 0 \ \forall\, i, j\,.$$

Alors il existe un entier m, avec $1 \leq m \leq n-1$, tel que B^m est strictement positive, c'est-à-dire que tous ses coefficients sont strictement positifs.

Démonstration. Par une récurrence facile on vérifie que le coefficient $b_{ij}^{(m)}$ de B^m en i-ème ligne et j-ème colonne est donnée par une somme sur $m-1$ indices

$$b_{ij}^{(m)} = \sum_{k_1=1}^{n} \cdots \sum_{k_{m-1}=1}^{n} b_{ik_1} b_{k_1 k_2} ... b_{k_{m-1} j}.$$

Remarquons que tous les termes de cette somme sont positifs ou nuls. Si $b_{ij} > 0$ alors en prenant $k_1 = k_2 = ... = k_{m-1} = i$, on est sûr que $b_{ij}^{(m)} > 0$ car $b_{ii} > 0$ par hypothèse. Si $b_{ij} = 0$, l'hypothèse d'irréductibilité de B implique qu'il existe une chaîne de coefficients non nuls $b_{ik_1}, b_{k_1 k_2}, ..., b_{k_{m-1} j}$ qui relie les indices i et j. La longueur m de cette chaîne est inférieure à $n-1$ qui est le cas où la chaîne visite tous les indices autres que i. Par conséquent, au moins un terme de la somme ci-dessus est non nul et on a bien $b_{ij}^{(m)} > 0$. ■

Démonstration du Théorème 6.2.11. Pour $\alpha > \max_i a_{ii}$, la matrice $(\alpha\,\mathrm{Id} - A)$ est une matrice positive irréductible avec des coefficients diagonaux strictement positifs. En vertu du Lemme 6.2.12 il existe m tel que $(\alpha\,\mathrm{Id} - A)^m$ est strictement positive. On peut alors lui appliquer le Théorème 6.2.9 de Perron-Frobenius qui affirme que $(\alpha\,\mathrm{Id} - A)^m$ admet une valeur propre dominante. Les

valeurs propres (répétées avec leur multiplicité) de A, notées λ, et celles de $(\alpha\,\mathrm{Id} - A)^m$, notées μ, vérifient

$$\mu = (\alpha - \lambda)^m \text{ et } \lambda = \alpha - \mu^{1/m}$$

avec les mêmes vecteurs propres, où la racine $\mu^{1/m}$ est l'unique racine positive si $\mu \geq 0$ et est une des racines possibles si μ est plus généralement un nombre complexe. De la relation $\mu_{max} > |\mu|$ on déduit

$$\mu_{max}^{1/m} > |\mu^{1/m}| \geq \mathcal{R}(\mu^{1/m}),$$

c'est-à-dire

$$\lambda_{min} = \alpha - \mu_{max}^{1/m} < \alpha - \mathcal{R}(\mu^{1/m}) \leq \mathcal{R}(\lambda)$$

ce qui prouve les propriétés annoncées à partir de celles données par le Théorème 6.2.9. ■

Remarque 6.2.13 *En fait, dans les hypothèses du Théorème 6.2.11 il n'est pas nécessaire que A soit une M-matrice mais simplement qu'il existe un réel $\beta \geq 0$ tel que $(\beta\,\mathrm{Id} + A)$ soit une M-matrice. En effet, la seule propriètè de M-matrice qui soit utilisée est que les coefficients extra-diagonaux sont positifs et rajouter $\beta\,\mathrm{Id}$ ne fait que décaler de β le spectre de A.*

Exercice 6.7 *Montrer que toute M-matrice est limite d'une suite de M-matrices irréductibles. En déduire que toute M-matrice admet une valeur propre réelle λ_{min} admettant un vecteur propre positif et telle que, pour toute autre valeur propre $\lambda \in \mathbf{C}$, on a $\lambda_{min} \leq \mathcal{R}(\lambda)$. Donner un contre-exemple où λ_{min} n'est pas simple.*

On peut se demander comment le Théorème 6.2.9 de Perron-Frobenius se généralise en dimension infinie. Dans ce cas il n'y a évidemment plus de notion de M-matrice et l'hypothèse essentielle sera la propriètè de positivité de l'opérateur. Sans vouloir rentrer dans les détails (qui dépassent le niveau de ce cours) il est instructif d'énoncer le théorème de Krein-Rutman qui généralise celui de Perron-Frobenius (voir, par exemple, le chapitre VI de [11]).

Théorème 6.2.14 (Krein-Rutman) *Soit E un espace de Banach et soit C un cône convexe de sommet 0, c'est-à-dire que $(\lambda x + \mu y) \in C$ pour tout $x, y \in C$ et $\lambda \geq 0, \mu \geq 0$. On suppose que C est fermé, d'intérieur $\mathrm{Int}C$ non vide et que $C \cap (-C) = \{0\}$. Soit K un opérateur compact de E dans E tel que $K(C \setminus \{0\}) \subset \mathrm{Int}C$. Alors il existe $u \in \mathrm{Int}C$ et $\lambda > 0$ tels que $Ku = \lambda u$. De plus λ est l'unique valeur propre associée à un vecteur propre dans C, est simple et*

$$\lambda = \max\{|\mu|\,,\ \textit{avec } \mu \textit{ valeur propre de } K\}.$$

Dans le Théorème 6.2.14 il faut penser que E est un espace de fonctions et C est le cône des fonctions positives ou nulles. Dans ce cas, l'hypothèse $K(C \setminus \{0\}) \subset \mathrm{Int}C$ est une propriété de positivité (stricte) de l'opérateur K.

6.2.3 Application

On revient à l'équation différentielle ordinaire (6.2)

$$\begin{cases} \dfrac{du}{dt} + Au = 0, \\ u(t=0) = u^0. \end{cases} \tag{6.10}$$

Proposition 6.2.15 *On suppose d'une part que le vecteur $u^0 \geq 0$ est positif et d'autre part que A est une M-matrice irréductible. Alors la solution $u(t)$ de l'équation différentielle ordinaire (6.10) admet une limite non nulle lorsque t tend vers $+\infty$ si et seulement si la plus petite valeur propre de A est nulle, $\lambda_{min} = 0$.*

Dans tous les cas le comportement asymptotique de $u(t)$ est

$$u(t) \approx (u^0 \cdot l_1) e^{-\lambda_1 t} r_1 \quad \text{quand } t \to +\infty\,,$$

où $r_1 > 0$ et $l_1 > 0$ sont les vecteurs propres de A et A^*, respectivement, associés à la plus petite valeur propre λ_1 commune de A et A^* et normalisés par

$$Ar_1 = \lambda_1 r_1\,,\ A^* l_1 = \lambda_1 l_1 \quad \text{et } r_1 \cdot l_1 = 1\,.$$

Remarque 6.2.16 *Rappelons que l'hypothèse sur la matrice A est vérifiée pour les matrices issues de la discrétisation d'équations de diffusion ou de Boltzmann (voir l'Exercice 6.2 et le Lemme 6.5.8). L'hypothèse $u^0 \geq 0$ est naturelle aussi si les composantes de la solution $u(t)$ représentent une densité de particules.*

Remarque 6.2.17 *Le vecteur propre l_1 est dit adjoint car il est le vecteur propre de la matrice adjointe A^* (qui est aussi une M-matrice irréductible). La Proposition 6.2.15 est la première occurence de la notion d'adjoint que nous reverrons un peu plus loin. La formule asymptotique pour la solution $u(t)$ à l'infini montre que son profil est toujours proportionnel au premier vecteur propre r_1 (quel que soit u^0) et que la constante de proportionnalité se calcule à l'aide du premier vecteur propre adjoint l_1.*

Démonstration. Il s'agit d'une simple combinaison du Théorème 6.2.11 et du Lemme 6.1.3. ■

6.3 Problème aux valeurs propres pour l'équation de diffusion

Dans cette section, nous allons passer en revue quelques résultats élémentaires sur le problème aux valeurs propres pour le laplacien avec condition de Dirichlet.

De façon générale, on se pose le problème suivant, sur un ouvert connexe borné Ω de $\mathbf{R}^N$, que l'on supposera à bord régulier (de classe C^∞) :

$$\begin{cases} -\Delta\phi = \lambda\phi\,, \quad x \in \Omega\,, \\ \phi\big|_{\partial\Omega} = 0\,. \end{cases} \tag{6.11}$$

Il s'agit d'un problème aux valeurs propres : l'inconnue est le couple (λ, ϕ), et on cherche tous les $\lambda \in \mathbf{R}$ pour lesquels le problème ci-dessus admette une solution $\phi \equiv \phi(x)$ à valeurs dans $\mathbf{R}$ et non identiquement nulle. Lorsque tel est le cas, on dira que λ est valeur propre du laplacien avec condition de Dirichlet sur l'ouvert Ω, et que ϕ en est une fonction propre pour la valeur propre λ.

Commençons par quelques remarques générales.

Disons quelques mots de l'espace fonctionnel où chercher la fonction ϕ. Pour que les différents termes intervenant dans le problème ci-dessus aient un sens, un choix naturel serait de demander que $\phi \in C^2(\Omega) \cap C(\overline{\Omega})$. D'autre part, avec le formalisme des distributions ou la formulation variationnelle des problèmes elliptiques, on pourrait aussi chercher ϕ dans l'espace de Sobolev $H^2(\Omega) \cap H^1_0(\Omega)$: cf. [2], chapitre 7. La propriété de régularisation elliptique du laplacien (cf. [11], Théorème IX.25) entraîne alors par une récurrence immédiate que $\phi \in H^m(\Omega)$ pour tout $m \in \mathbf{N}$. Puis on déduit des injections de Sobolev (cf. [11], Corollaire IX.15, note 2) que $\phi \in C^m(\overline{\Omega})$ pour tout $m \in \mathbf{N}$. Par conséquent, on ne restreint pas la généralité de l'étude en cherchant a priori les fonctions propres ϕ dans $C^\infty(\overline{\Omega})$.

Remarque 6.3.1 *On pourrait plus généralement chercher des valeurs propres $\lambda \in \mathbf{C}$ associées à des fonctions propres $\phi \equiv \phi(x)$ à valeurs dans $\mathbf{C}$, solutions de (6.11). Le calcul élémentaire suivant montre que cela n'est pas nécessaire, autrement dit que toutes les valeurs propres de (6.11) sont réelles et que les fonctions propres peuvent être choisies réelles.*

En effet, pour tout $\phi \in C^2(\overline{\Omega})$ à valeurs complexes, on a, d'après la formule de Green,

$$\begin{aligned}\int_\Omega \overline{\phi(x)}(-\Delta)\phi(x)dx &= -\int_\Omega \mathrm{div}(\overline{\phi(x)}\nabla\phi(x))dx + \int_\Omega |\nabla\phi(x)|^2 dx \\ &= -\int_\Omega \overline{\phi(x)}\nabla\phi(x)\cdot n_x dS(x) + \int_\Omega |\nabla\phi(x)|^2 dx\end{aligned}$$

en notant $dS(x)$ l'élément de surface sur $\partial\Omega$, n_x le vecteur unitaire normal à $\partial\Omega$ au point x et $\overline{\phi}$ la fonction complexe conjuguée de ϕ. Si ϕ est fonction propre du laplacien avec condition de Dirichlet sur Ω, on a donc

$$\int_\Omega |\nabla\phi(x)|^2 dx = \lambda \int_\Omega |\phi(x)|^2 dx\,. \tag{6.12}$$

Si λ est valeur propre, et que ϕ est une fonction propre associée à λ, on sait d'une part que ϕ est (au moins) continue sur $\overline{\Omega}$, et non identiquement nulle, de sorte que

$$\int_\Omega |\phi(x)|^2 dx > 0$$

Donc toute valeur propre λ du laplacien avec condition de Dirichlet sur Ω appartient nécessairement à $\mathbf{R}_+$.

Examinons plus en détail le cas $\lambda = 0$: l'égalité (6.12) entraînerait que, si $\lambda = 0$ est valeur propre et que ϕ est une fonction propre associée

$$\int_\Omega |\nabla\phi(x)|^2 dx = 0\,.$$

Par conséquent, $\nabla\phi = 0$, de sorte que ϕ est constante sur Ω puisque ce dernier est connexe. La condition de Dirichlet entraîne alors que $\phi = 0$, ce qui contredit le fait que ϕ soit une fonction propre.

En réalité, en utilisant la théorie des opérateurs compacts (cf. [11], chapitre VI), on peut aller plus loin, en établissant le résultat suivant, qui résume l'essentiel de ce qui est connu sur les valeurs propres du laplacien avec condition de Dirichlet.

Théorème 6.3.2 (Spectre du laplacien/conditions de Dirichlet) *Soit Ω ouvert connexe borné de $\mathbf{R}^N$, à bord de classe C^∞. Il existe une suite croissante de réels strictements positifs*

$$0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots \leq \lambda_n \leq \ldots$$

et une suite $(\phi_n)_{n\geq 1}$ de fonctions de classe C^∞ sur $\overline{\Omega}$ à valeurs réelles vérifiant les propriétés suivantes :

a) les λ_n sont les valeurs propres du laplacien avec condition de Dirichlet sur Ω ;
b) pour tout $n \geq 1$, la fonction ϕ_n est fonction propre du laplacien avec condition de Dirichlet sur Ω pour la valeur propre λ_n ;
c) la suite $(\phi_n)_{n\geq 1}$ est une base hilbertienne de $L^2(\Omega)$;
d) lorsque $n \to +\infty$, on a $\lambda_n \sim Cn^{2/N}$, où C est une constante positive.

La démonstration de ce théorème dépasse le cadre de notre étude et nous l'admettrons donc (voir [2], chapitre 7). Voir également le Théorème IX.31 dans [11], ainsi que le point 13 des compléments du chapitre IX de [11].

Nous allons conclure cette section en étudiant de manière détaillée le cas de la dimension $N = 1$, où l'on peut tout calculer de manière explicite. Ce cas particulier suffira d'ailleurs pour la suite de notre étude.

On considère donc le problème aux valeurs propres

$$\begin{cases} -\dfrac{d^2\phi}{dx^2}(x) = \lambda\phi(x)\,, \quad |x| < L \\ \phi(\pm L) = 0 \end{cases}$$

D'après la Remarque 6.3.1, on sait qu'il faut chercher les valeurs propres λ dans $\mathbf{R}_+^*$. Si l'on ne tient pas compte des conditions aux limites, une base de solutions de cette équation différentielle linéaire d'ordre 2 est $\{\sin(\sqrt{\lambda}x), \cos(\sqrt{\lambda}x)\}$, de sorte que la solution générale peut se mettre sous la forme

$$\phi(x) = C\sin(\sqrt{\lambda}(x - x_0))$$

où C et x_0 sont deux réels arbitraires.

Pour satisfaire la condition $\phi(-L) = 0$, il est naturel de poser $x_0 = -L$, de sorte que ϕ est de la forme $\phi(x) = \sin(\sqrt{\lambda}(x+L))$ (en faisant $C = 1$ puisque les fonctions propres sont définies à une constante multiplicative près). Pour vérifier la condition $\phi(+L) = 0$, la fonction propre devra donc satisfaire

$$\sqrt{\lambda}2L = k\pi\,, \quad k \in \mathbf{N}^*\,.$$

On obtient ainsi la suite des fonctions propres du laplacien sur $]-L, L[$ avec condition de Dirichlet :

$$\phi_k(x) = \sin\left(\frac{k\pi}{2L}(x+L)\right)\,, \quad k \in \mathbf{N}^*\,,$$

ainsi que la suite des valeurs propres associées

$$\lambda_k = \frac{k^2\pi^2}{4L^2}\,, \quad k \in \mathbf{N}^*\,.$$

On observera en particulier que la plus petite valeur propre est

$$\lambda_1 = \frac{\pi^2}{4L^2}$$

et que la fonction propre associée

$$\phi_1(x) = \sin\left(\frac{\pi}{2L}(x+L)\right) = \cos\left(\frac{\pi x}{2L}\right)$$

est strictement positive sur $]-L, L[$.

On peut résumer ces quelques remarques en disant que le problème aux valeurs propres pour le laplacien avec condition de Dirichlet sur un ouvert connexe borné de $\mathbf{R}^N$ est analogue à la recherche des valeurs propres pour une matrice symétrique réelle — ou hermitienne — mais en dimension infinie.

Cependant, il ne faudrait pas croire que les problèmes spectraux pour les opérateurs différentiels soient toujours aussi simples.

Par exemple, si on ne suppose plus l'ouvert Ω borné, le problème aux valeurs propres pour le laplacien est complètement différent. Considérons par exemple le cas où $\Omega = \mathbf{R}$ et où le problème spectral est

$$-\frac{d^2\phi}{dx^2}(x) = \lambda\phi(x)\,.$$

Une première différence avec le cas où Ω est un intervalle borné est qu'il n'existe pas de fonction propre $\phi \in L^2(\mathbf{R})$ pour ce problème — en effet, les solutions sont des combinaisons linéaires de $\cos(\sqrt{\lambda}x)$ et $\sin(\sqrt{\lambda}x)$ si $\lambda > 0$, ou bien des combinaisons linéaires de $\exp(\sqrt{-\lambda}x)$ et $\exp(-\sqrt{-\lambda}x)$ si $\lambda < 0$.

Une deuxième différence est que, si l'on accepte pour fonctions propres des fonctions qui ne sont pas dans $L^2(\mathbf{R})$, il n'y a aucune restriction sur les valeurs propres : tout réel est valeur propre. (En réalité, comme les fonctions propres associées ne sont pas de carré sommable, il ne s'agit pas vraiment de valeurs propres ou de fonctions propres au sens habituel, mais en un sens généralisé dont la définition dépasse le cadre de notre étude.)

6.4 Problème spectral pour l'équation de Boltzmann linéaire monocinétique avec scattering isotrope

Si nous voulions étudier le problème aux valeurs propres pour l'équation de Boltzmann linéaire, nous nous heurterions à une difficulté théorique importante, qui est liée au caractère non auto-adjoint de l'opérateur intervenant dans cette équation.

Déjà dans le cas de l'algèbre linéaire en dimension finie, on sait bien que l'étude spectrale d'une matrice auto-adjointe ou normale — c'est-à-dire commutant avec son adjointe — est plus simple que pour une matrice quelconque, dans la mesure où une matrice normale est toujours diagonalisable et admet une base de vecteurs propres orthonormée pour le produit scalaire (ou hermitien) canonique de $\mathbf{R}^N$ (ou de $\mathbf{C}^N$).

Outre les difficultés relatives au caractère non auto-adjoint de l'opérateur de Boltzmann linéaire, l'étude du problème aux valeurs propres pour l'équation de Boltzmann linéaire est compliquée par le fait qu'il s'agit d'un problème en dimension infinie, de sorte que son spectre ne se réduit pas forcément aux seules valeurs propres — voir ci-dessous.

Toutefois, dans certains cas particuliers, on peut ramener l'étude spectrale de l'équation de Boltzmann linéaire à celle d'un problème auto-adjoint faisant intervenir un opérateur de Hilbert-Schmidt, pour lequel le spectre se réduit aux valeurs propres.

Nous allons étudier dans cette section l'exemple de l'équation de Boltzmann linéaire pour des particules monocinétiques avec scattering isotrope, dans le cas de la symétrie de la plaque infinie.

On considère donc l'équation de Boltzmann linéaire

$$\left(\frac{\partial}{\partial t} + \mu\frac{\partial}{\partial x}\right) f(t,x,\mu) = \sigma(1+\gamma)\langle f\rangle(t,x) - \sigma f(t,x,\mu)$$

avec la notation

$$\langle\phi\rangle = \tfrac{1}{2}\int_{-1}^{1} \phi(\mu)d\mu\,.$$

On supposera dans toute cette étude que l'équation est posée pour $x \in [-L,L]$ et $\mu \in [-1,1]$, que $\sigma > 0$ et $\gamma > -1$, tandis que la solution f vérifie les conditions aux limites

$$f(t,-L,\mu) = f(t,L,-\mu) = 0\,, \quad 0 < \mu < 1\,.$$

Ces conditions aux limites sont dites "absorbantes" : les particules situées au bord de l'intervalle $[-L,L]$ peuvent seulement sortir de l'intervalle mais en aucun cas y entrer. Autrement dit, l'extérieur de l'intervalle absorbe les particules, ce qui explique la terminologie adoptée pour cette condition aux limites.

En mettant l'équation de Boltzmann linéaire ci-dessus sous la forme

$$\frac{\partial f}{\partial t} = Af\,, \qquad \text{avec } Af = -\mu\frac{\partial f}{\partial x} + \sigma(1+\gamma)\langle f\rangle - \sigma f\,,$$

on voit que, si $\phi \equiv \phi(x,\mu)$ est fonction propre de A, c'est-à-dire que

$$\begin{cases} A\phi = \lambda\phi\,, & \phi \neq 0\,, \\ \phi(-L,\mu) = \phi(L,-\mu) = 0\,, & 0 < \mu \leq 1\,, \end{cases}$$

alors la fonction $f \equiv f(t,x,\mu)$ définie par

$$f(t,x,\mu) = e^{\lambda t}\phi(x,\mu)$$

est une solution de l'équation de Boltzmann linéaire vérifiant la condition absorbante pour tout $t \geq 0$.

On va donc étudier dans la suite le problème spectral

$$\begin{cases} -\mu\frac{\partial\phi}{\partial x}(x,\mu) + \sigma(1+\gamma)\langle\phi\rangle(x) - \sigma\phi(x,\mu) = \lambda\phi(x,\mu)\,, & \phi \neq 0\,, \\ \phi(-L,\mu) = \phi(L,-\mu) = 0\,, & 0 < \mu \leq 1\,. \end{cases}$$

6.4.1 Le résultat principal

Dans la suite de cette section, nous allons établir le résultat suivant, qui donne une description complète des valeurs propres de l'opérateur de transport monocinétique avec scattering isotrope, en faisant l'hypothèse de la symétrie de type plaque infinie.

Théorème 6.4.1 *Soient $\sigma > 0$ et $\gamma > -1$. Il existe un unique réel $\lambda_L(\sigma,\gamma)$ dans l'intervalle $]-\sigma,+\infty[$ qui est la plus grande valeur propre de l'opérateur A défini sur $]-L,L[\times[-1,1]$ par*

$$A\phi = -\mu\frac{\partial\phi}{\partial x} + \sigma(1+\gamma)\langle\phi\rangle - \sigma\phi\,, \tag{6.13}$$

avec conditions aux limites absorbantes. De plus, il existe une fonction propre de A p.p. positive ou nulle pour la valeur propre $\lambda_L(\sigma,\gamma)$.

La démonstration de ce résultat n'est pas triviale, en premier lieu parce que, comme nous l'avons dit plus haut, l'opérateur de Boltzmann linéaire n'est pas auto-adjoint.

On peut montrer que les autres valeurs propres de A sont toutes réelles, de multiplicités finies, et appartiennent à l'intervalle $]-\sigma,\lambda_L(\sigma,\gamma)]$. Toutefois la démonstration de ce point est assez technique et nous ne la donnerons pas.

La preuve se décompose en plusieurs étapes.

Etape 1. On réduit l'équation aux valeurs propres pour l'opérateur de Boltzmann linéaire à une équation intégrale de la forme

$$\langle\phi\rangle = B_\lambda\langle\phi\rangle\,.$$

Ici, B_λ est un opérateur intégral de la forme

$$B_\lambda \psi(x) = \int_{-L}^{L} b_\lambda(x,y)\psi(y)dy\,,$$

où $b_\lambda(x,y) = b_\lambda(y,x) \in \mathbf{R}$ vérifie

$$\iint_{[-L,L]^2} b_\lambda(x,y)^2 dxdy < +\infty\,.$$

Le paramètre λ est valeur propre de l'équation de Boltzmann linéaire si et seulement si 0 est valeur propre de $I - B_\lambda$. L'étape 1 fait l'objet de la section 6.4.2 ci-dessous.

Etape 2. On étudie ensuite les valeurs propres de l'opérateur intégral B_λ vu comme application linéaire sur $L^2([-L,L])$: on montrera notamment que ces valeurs propres s'organisent en une suite

$$\rho_0(\lambda) \geq \rho_1(\lambda) \geq \ldots \geq \rho_n(\lambda) \geq \ldots \geq 0\,.$$

On conclut avec une caractérisation variationnelle de la plus grande valeur propre $\rho_0(\lambda)$. L'étape 2 fait l'objet de la section 6.4.3 ci-dessous.

Etape 3. On utilise la caractérisation variationnelle de $\rho_0(\lambda)$ pour étudier sa dépendance par rapport à λ. On montre en particulier que ρ_0 est une fonction continue et strictement décroissante de λ tendant vers 0 pour $\lambda \to +\infty$, et vers $+\infty$ pour $\lambda \to -\sigma$. Par conséquent, il existe un unique réel $\lambda \in]-\sigma, +\infty[$ tel que $\rho_0(\lambda) = 1$, et ce réel est la plus grande valeur propre cherchée pour l'équation de Boltzmann. Cette dernière étape fait l'objet de la section 6.4.4 ci-dessous.

6.4.2 Réduction à un problème spectral auto-adjoint

On résout le problème aux limites ci-dessus en exprimant ϕ en fonction de $\langle\phi\rangle$, grâce à la formule (2.2), comme suit :

$$\phi(x,\mu) = \int_{-L}^{x} \tfrac{\sigma(1+\gamma)}{\mu} e^{-(\sigma+\lambda)(x-y)/\mu} \langle\phi\rangle(y)dy\,, \qquad |x| \leq L\,, \quad 0 < \mu \leq 1\,,$$

$$\phi(x,\mu) = \int_{x}^{L} \tfrac{\sigma(1+\gamma)}{|\mu|} e^{-(\sigma+\lambda)(y-x)/|\mu|} \langle\phi\rangle(y)dy\,, \qquad |x| \leq L\,, \quad -1 \leq \mu < 0\,.$$

Puis, en moyennant en μ les deux membres de l'égalité ci-dessus, on trouve que

$$\begin{aligned}\langle\phi\rangle(x) = &\int_{-L}^{x} \left(\tfrac{1}{2} \int_0^1 \tfrac{\sigma(1+\gamma)}{\mu} e^{-(\sigma+\lambda)(x-y)/\mu} d\mu \right) \langle\phi\rangle(y)dy \\ &+ \int_{x}^{L} \left(\tfrac{1}{2} \int_{-1}^0 \tfrac{\sigma(1+\gamma)}{|\mu|} e^{-(\sigma+\lambda)(y-x)/|\mu|} d\mu \right) \langle\phi\rangle(y)dy\,.\end{aligned}$$

Dans les deux intégrales ci-dessus, l'intégrale interne s'exprime au moyen de la même fonction

$$E(z) = \tfrac{1}{2}\int_0^1 e^{-z/\mu}\tfrac{d\mu}{\mu} = \tfrac{1}{2}\int_1^\infty e^{-zu}\tfrac{du}{u} = \tfrac{1}{2}\int_z^\infty e^{-v}\tfrac{dv}{v}\,, \quad z > 0\,, \qquad (6.14)$$

(qui est, au facteur $\frac{1}{2}$ près, la fonction exponentielle intégrale). Autrement dit, la relation ci-dessus sur la moyenne $\langle\phi\rangle$ s'écrit

$$\langle\phi\rangle(x) = \sigma(1+\gamma)\int_{-L}^{L} E((\sigma+\lambda)|x-y|)\langle\phi\rangle(y)dy\,, \quad |x| \le L\,. \tag{6.15}$$

On voit déjà que ce calcul a transformé le problème original pour l'opérateur A qui n'est pas auto-adjoint, en le problème (6.15) pour $\langle\phi\rangle$ qui, lui, est bien auto-adjoint — par analogie avec le système linéaire

$$\psi_i = \sum_j E_{ij}\psi_j\,, \quad \text{où } E_{ij} = E_{ji}\,, \qquad i,j = 1,2,\dots$$

On peut en effet penser à discrétiser l'équation intégale (6.15) en posant

$$\psi_i = \langle\phi\rangle(x_i)\,, \quad E_{ij} = (\sigma+\gamma)\int_{-\Delta x/2}^{\Delta x/2} E((\sigma+\lambda)|x_i+z-x_j|)dz\,,$$

pour tous $i,j = 1\dots,n$, où $\Delta x = \frac{2L}{n}$ est un pas d'espace uniforme sur l'intervalle $[-L,L]$.

En faisant le changement de variables $z \mapsto -z$, on trouve que

$$\begin{aligned} E_{ij} &= (\sigma+\gamma)\int_{-\Delta x/2}^{\Delta x/2} E((\sigma+\lambda)|x_i+z-x_j|)dz \\ &= (\sigma+\gamma)\int_{-\Delta x/2}^{\Delta x/2} E((\sigma+\lambda)|x_i-x_j-z|)dz \\ &= (\sigma+\gamma)\int_{-\Delta x/2}^{\Delta x/2} E((\sigma+\lambda)|x_j+z-x_i|)dz = E_{ji} \end{aligned}$$

pour tous $i,j = 1,\dots,N$, de sorte que la matrice $(E_{ij})_{1\le i,j\le n}$ est symétrique réelle.

Avant d'aller plus loin, il nous faut étudier de plus près la fonction E.

Lemme 6.4.2 *La fonction E, définie par (6.14), vérifie les propriétés suivantes :*
(a) $E \in C^1(\mathbf{R}_+^*)$ est une fonction strictement décroissante ;
(b) lorsque $z \to +\infty$

$$E(z) \sim \tfrac{1}{2}\frac{e^{-z}}{z}\,;$$

(c) lorsque $z \to 0^+$

$$E(z) \sim -\tfrac{1}{2}\ln z\,.$$

En particulier, $E \in L^p(\mathbf{R}_+)$ pour tout $p \in [1,+\infty[$.

Démonstration. Le point (a) est évident puisque E est la primitive de la fonction $v \mapsto -\frac{e^{-v}}{v}$, continue et strictement négative sur $\mathbf{R}_+^*$.

Pour ce qui est du point (b), on remarque que

$$\left(\frac{e^{-v}}{v}\right)' = -\frac{e^{-v}}{v} - \frac{e^{-v}}{v^2} \sim -\frac{e^{-v}}{v}$$

pour $v \to +\infty$. On conclut grâce au fait que, si f et g sont deux fonctions continues sur $\mathbf{R}_+^*$ telles que $0 \leq f(t) \sim g(t)$ pour $t \to +\infty$ et si

$$\int_1^{+\infty} f(t)dt < +\infty\,,$$

alors

$$\int_x^{+\infty} f(t)dt \sim \int_x^{+\infty} g(t)dt$$

lorsque $x \to +\infty$.

Enfin, on démontre le point (c) en écrivant

$$\begin{aligned} E(z) &= \tfrac{1}{2}\int_z^1 \frac{dv}{v} + \tfrac{1}{2}\int_z^1 (e^{-v}-1)\frac{dv}{v} + \tfrac{1}{2}\int_1^{+\infty} e^{-v}\frac{dv}{v} \\ &= -\tfrac{1}{2}\ln z + \tfrac{1}{2}\int_z^1 (e^{-v}-1)\frac{dv}{v} + \tfrac{1}{2}\int_1^{+\infty} e^{-v}\frac{dv}{v} \end{aligned}$$

et en remarquant que la fonction

$$z \mapsto \int_z^1 (e^{-v}-1)\frac{dv}{v} \text{ est de classe } C^1 \text{ sur } \mathbf{R}\,,$$

puisque l'intégrande $v \mapsto \frac{e^{-v}-1}{v}$ se prolonge par continuité en $v = 0$. ∎

Pour tout $\lambda > 0$, on introduit l'opérateur K_λ défini par la formule

$$K_\lambda\psi(x) = \int_{-L}^{L} E((\sigma+\lambda)|x-y|)\psi(y)dy\,.$$

Lemme 6.4.3 *Pour tout $\lambda > -\sigma$, l'opérateur K_λ est un opérateur de Hilbert-Schmidt sur $L^2([-L,L])$. De plus, pour tout $\psi \in L^2([-L,L])$, la fonction $K_\lambda\psi$ est (p.p. égale à) une fonction continue sur $[-L,L]$.*

Démonstration. La fonction $(x,y) \mapsto E((\sigma+\lambda)|x-y|)$ est continue sur

$$[-L,L]\times[-L,L] \setminus \{(x,x)\,|\,|x| \leq L\}$$

et vérifie

$$E((\sigma+\lambda)|x-y|) \sim -\tfrac{1}{2}\ln|x-y| \text{ pour } x-y \to 0\,.$$

Donc

$$\iint_{[-L,L]^2} E((\sigma+\lambda)|x-y|)^2 dxdy < +\infty\,,$$

ce qui montre que l'opérateur K_λ est de Hilbert-Schmidt sur $L^2([-L,L])$.

D'autre part la fonction $F : z \mapsto E((\sigma+\lambda)|z|)$ définit un élément de $L^2(\mathbf{R})$ d'après le point (c) du lemme précédent, de sorte que la fonction

$$K_\lambda \psi = F \star (\psi \mathbf{1}_{[-L,L]})\big|_{[-L,L]}$$

est continue sur $[-L,L]$ comme produit de convolution de fonctions appartenant à deux espaces de Lebesgue en dualité [1]. ■

Précisons la relation entre le problème aux valeurs propres pour l'équation de Boltzmann linéaire de départ portant sur la fonction $\phi \equiv \phi(x,\mu)$ et celui portant sur la fonction $\langle\phi\rangle$ moyennée en μ.

Proposition 6.4.4 *Soit $\lambda > -\sigma$. Une condition nécessaire et suffisante pour qu'il existe une fonction $\phi \in L^2([-L,L]\times[-1,1])$ telle que $x \mapsto \phi(x,\mu)$ soit p.p. en $\mu \in [-1,1]$ égale à une fonction continue sur $[-L,L]$, solution généralisée du problème aux valeurs propres*

$$\begin{cases} A\phi = \lambda\phi\,, & \phi \neq 0\,, \\ \\ \phi(-L,\mu) = \phi(L,-\mu) = 0\,, & 0 < \mu \leq 1\,, \end{cases}$$

avec A défini par (6.13), est que 1 soit valeur propre de l'opérateur $\sigma(1+\gamma)K_\lambda$ de Hilbert-Schmidt sur $L^2([-L,L])$.

Démonstration. Les calculs ci-dessus montrent que si $\phi \in L^2([-L,L]\times[-1,1])$ est une fonction propre de A pour la valeur propre λ au sens généralisé — c'est-à-dire que la fonction $x \mapsto \phi(x,\mu)$ est continue p.p. en $\mu \in [-1,1]$ et est solution généralisée du problème aux valeurs propres pour A — alors $\langle\phi\rangle \in L^2([-L,L])$ est soit nulle p.p., soit fonction propre de l'opérateur $\sigma(1+\gamma)K_\lambda$ pour la valeur propre 1.

Si $\langle\phi\rangle = 0$ p.p. sur $[-L,L]$, alors ϕ vérifie

$$\begin{gathered} (\lambda+\sigma)\phi(x,\mu) + \mu\frac{\partial\phi}{\partial x}(x,\mu) = 0\,, \quad |x| < L\,,\ |\mu| \leq 1\,, \\ \phi(-L,\mu) = \phi(L,-\mu) = 0\,, \quad 0 < \mu \leq 1\,. \end{gathered}$$

On vérifie alors en appliquant la méthode des caractéristiques que $\phi = 0$ p.p. sur $[-L,L]\times[-1,1]$, ce qui est en contradiction avec le fait que ϕ soit une fonction propre de l'opérateur A.

Réciproquement, si $u \in L^2([-L,L])$ est fonction propre de $\sigma(1+\gamma)K_\lambda$ pour la valeur propre 1, alors $u = \sigma(1+\gamma)K_\lambda u$ est (p.p. égale à) une fonction continue sur $L^2([-L,L])$ et la fonction $\phi \equiv \phi(x,\mu)$ définie pour tout $\mu \in [-1,1]\setminus\{0\}$ par

1. On rappelle en effet que, pour tout $p \in [1,\infty]$, et pour tout $f \in L^p(\mathbf{R})$ et tout $g \in L^{p'}(\mathbf{R})$ avec $\frac{1}{p} + \frac{1}{p'} = 1$ (convenant que $1/\infty = 0$), le produit de convolution $f \star g$ est (p.p. égal à) une fonction bornée uniformément continue sur $\mathbf{R}$.

les formules

$$\phi(x,\mu) = \int_{-L}^{x} \tfrac{\sigma(1+\gamma)}{\mu} e^{-(\sigma+\lambda)(x-y)/\mu} u(y)dy\,, \qquad |x| \le L\,, \quad 0 < \mu \le 1\,,$$

$$\phi(x,\mu) = \int_{x}^{L} \tfrac{\sigma(1+\gamma)}{|\mu|} e^{-(\sigma+\lambda)(y-x)/|\mu|} u(y)dy\,, \qquad |x| \le L\,, \quad -1 \le \mu < 0\,,$$

est évidemment continue sur $[-L,L] \times ([-1,1] \setminus \{0\})$. D'autre part

$$|\phi(x,\mu)| \le \frac{\sigma(1+\gamma)}{\sigma+\lambda} \sup_{|x|\le L} |u(y)|\,, \quad |x| \le L\,,\ 0 < |\mu| \le 1\,,$$

de sorte que $\phi \in L^2([-L,L]\times[-1,1])$, puisque $\sigma > -\lambda$.

De plus, en moyennant l'expression définissant ϕ en fonction de u par rapport à la variable μ, on trouve comme dans les calculs ci-dessus que

$$\langle \phi \rangle = \sigma(1+\gamma) K_\lambda u = u\,.$$

D'autre part, les formules ci-dessus définissant ϕ en fonction de u signifient précisément que ϕ est solution généralisée du problème au valeurs propres pour l'opérateur A et la valeur propre λ. De plus, ϕ est bien fonction propre : si on avait $\phi(x,\mu) = 0$ p.p. en (x,μ), on aurait alors $\langle\phi\rangle = 0$ p.p. sur $[-L,L]$, or ceci est impossible puisqu'on a supposé $u = \langle\phi\rangle$ est fonction propre de $\sigma(1+\gamma)K_\lambda$ (pour la valeur propre 1). ■

6.4.3 Le problème spectral pour K_λ

A l'intérieur de la classe des opérateurs bornés sur l'espace de Hilbert L^2, les opérateurs de Hilbert-Schmidt jouissent de propriétés tout à fait semblables à celles vérifiées par les endomorphismes définis sur des espaces vectoriels de dimension finie. Par exemple, nous avons déjà vu que ces opérateurs vérifient l'alternative de Fredholm, qui est une condition de compatibilité analogue aux conditions de résolubilité d'un système linéaire dans un espace vectoriel de dimension finie. Il y a plus : l'analyse spectrale des opérateurs de Hilbert-Schmidt se réduit à la recherche de leurs valeurs propres — ce qui n'est évidemment pas le cas pour un opérateur borné quelconque, comme le montre l'exemple ci-dessous.

Exemple : Sur l'espace de Hilbert $H = L^2([0,1])$, considérons l'opérateur T : $\phi \mapsto T\phi$ défini par

$$T\phi(x) = f(x)\phi(x)$$

où $f \in C([0,1])$. L'opérateur T est un opérateur borné — c'est-à-dire une application linéaire continue de H dans lui-même, car

$$\|T\phi\|_{L^2([0,1])} \le \|f\|_{L^\infty([0,1])} \|\phi\|_{L^2([0,1])}\,.$$

Pour tout $x_0 \in [0,1]$, on a

$$(T - f(x_0)I)\phi(x) = (f(x) - f(x_0))\phi(x)\,,$$

et, comme

$$\sup_{\substack{0\le x\le 1\\ x\neq x_0}} \left| \frac{1}{f(x)-f(x_0)} \right| = +\infty \,,$$

l'opérateur $T - f(x_0)I$ n'admet pas d'inverse dans l'algèbre des opérateurs bornés sur H. On montre ainsi sans aucune difficulté que le spectre de T — c'est-à-dire le complémentaire dans $\mathbf{C}$ de l'ensemble des nombres complexes λ tels que $\lambda I - T$ soit inversible et d'inverse borné sur H — est donné par

$$\text{spectre}(T) = \{ f(x_0) \,|\, x_0 \in [0,1] \} \,.$$

Par exemple, si $f(x) = 2x$, on voit ainsi que $\lambda = 1 = f(\frac{1}{2}) \in \text{spectre}(T)$. Mais 1 n'est pas valeur propre de T : en effet, $\text{Ker}(T - I) = \{0\}$, puisque

$$(T-I)\phi(x) = 2(x-\tfrac{1}{2})\phi(x) = 0 \text{ p.p. en } x \in [0,1]$$

ce qui entraîne que

$$\phi(x) = 0 \text{ pour tout } x \in [0,1] \setminus (\mathcal{N} \cup \{\tfrac{1}{2}\}) \,,$$

où $\mathcal{N}$ est un ensemble de mesure nulle dans $[0,1]$. Donc $\mathcal{N} \cup \{\frac{1}{2}\}$ est aussi un ensemble de mesure nulle, de sorte que $\phi(x) = 0$ p.p. en $x \in [0,1]$. Autrement dit, bien que 1 appartienne au spectre de T, c'est-à-dire que $T - I$ n'admette pas d'inverse qui soit une application linéaire continue sur H, l'application linéaire $T - I$ est injective, de sorte qu'il n'existe pas de fonction ϕ non nulle dans H telle que l'on ait $T\phi = \phi$. Ceci est dû au fait que H est de dimension infinie, car tout endomorphisme d'un espace vectoriel de dimension finie est continu et inversible si et seulement si il est injectif.

Revenons au cas des opérateurs de Hilbert-Schmidt.

Théorème 6.4.5 *Soit J intervalle de $\mathbf{R}$ et $H = L^2(J)$. Soit $k \in L^2(J \times J)$, et soit K l'opérateur intégral de noyau k défini sur H, c'est-à-dire que*

$$K\phi(x) = \int_J k(x,y)\phi(y)dy \,, \quad \phi \in H \,.$$

Supposons que

$$k(x,y) = \overline{k(y,x)} \text{ p.p. en } (x,y) \in J \times J \,,$$

et que

$$\int_J \overline{\phi(x)} K\phi(x) dx \ge 0 \,, \quad \phi \in H \,.$$

Alors il existe une base hilbertienne $(e_n)_{n\ge 0}$ de H, et une suite de réels

$$\lambda_0 \ge \lambda_1 \ge \ldots \ge \lambda_n \ge \ldots \ge 0$$

telle que

$$Ke_n = \lambda_n e_n \,, \quad n \ge 0 \,.$$

De plus la suite λ_n *tend vers* 0 *et, si* $\lambda_n \neq 0$, *alors* $\mathrm{Ker}(K-\lambda_n I)$ *est de dimension finie. Enfin*

$$\lambda_0 = \max_{\substack{\phi\in L^2([-L,L])\\ \phi\neq 0}} \frac{\displaystyle\int_J \overline{\phi(x)} K\phi(x)dx}{\displaystyle\int_J |\phi(x)|^2 dx}\,.$$

Voir Théorèmes VI.11 et VI.12 dans [11].

Exercice 6.8 *Démontrer la formule donnant* λ_0. *(Indication : on pourra représenter tout élément* ϕ *de* H *sous la forme*

$$\phi = \sum_{n\geq 0} \phi_n e_n$$

et calculer $K\phi$ *à partir de cette représentation.)*

Admettons ce résultat — qui est tout à fait analogue au théorème spectral pour une matrice auto-adjointe ou symétrique réelle, et voyons ce qu'il implique dans le cas qui nous intéresse.

On sait déjà que l'opérateur K_λ est de Hilbert-Schmidt sur $L^2([-L,L])$; vérifions qu'il est positif au sens des formes bilinéaires sur H.

Comme le noyau intégral de K_λ est à valeurs réelles, il suffit de le faire opérer sur des fonctions à valeurs réelles. Si ϕ est une fonction propre à valeurs complexes de K_λ pour la valeur propre $\rho \in \mathbf{R}$, alors la partie réelle de ϕ ou sa partie imaginaire sont également fonctions propres de K_λ pour la valeur propre ρ. On pourra donc supposer que, dans le Théorème 6.4.5, toutes les fonctions considérées, y compris les fonctions propres e_n, sont à valeurs réelles.

Lemme 6.4.6 *Pour tout* $\psi \in H$ *à valeurs réelles, on a*

$$\int_{-L}^{L} \psi(x) K_\lambda \psi(x) dx \geq 0\,.$$

Démonstration. Il suffit de se souvenir que

$$K_\lambda \psi(x) = \langle\phi\rangle(x)$$

pour presque tout $x \in [-L,L]$, où $\phi \equiv \phi(x,\mu)$ est la solution généralisée du problème aux limites suivant pour l'équation de transport :

$$\begin{cases} (\lambda+\sigma)\phi(x,\mu) + \mu\frac{\partial\phi}{\partial x}(x,\mu) = \psi(x)\,, & |x|\leq L\,,\ 0<\mu\leq 1\,,\\ \\ \phi(-L,\mu) = \phi(L,-\mu) = 0\,, & 0<\mu\leq 1\,. \end{cases}$$

Par conséquent

$$\begin{aligned}\int_{-L}^{L}\psi(x)K_\lambda\psi(x)dx &= \int_{-L}^{L}\psi(x)\langle\phi\rangle(x)dx = \tfrac{1}{2}\int_{-L}^{L}\int_{-1}^{1}\psi(x)\phi(x,\mu)d\mu dx\\ &= \tfrac{1}{2}\int_{-L}^{L}\int_{-1}^{1}\left((\lambda+\sigma)\phi(x,\mu)+\mu\frac{\partial\phi}{\partial x}(x,\mu)\right)\phi(x,\mu)d\mu dx\\ &= \tfrac{1}{2}\int_{-L}^{L}\int_{-1}^{1}\left((\lambda+\sigma)\phi(x,\mu)^2+\tfrac{1}{2}\mu\frac{\partial}{\partial x}\left(\phi(x,\mu)^2\right)\right)d\mu dx\\ &\geq \tfrac{\lambda+\sigma}{2}\int_{-L}^{L}\int_{-1}^{1}\phi(x,\mu)^2 d\mu dx \geq 0\,.\end{aligned}$$

En effet, compte-tenu des conditions aux limites vérifiées par ϕ, on a

$$\int_{-L}^{L}\int_{-1}^{1}\mu\frac{\partial}{\partial x}\left(\phi(x,\mu)^2\right)d\mu dx = \int_0^1 \mu\left(\phi(L,\mu)^2+\phi(-L,-\mu)^2\right)d\mu \geq 0\,.$$

■

Appliquons le Théorème 6.4.5 à l'opérateur K_λ. Soit donc

$$\rho_0(\lambda)\geq\rho_1(\lambda)\geq\ldots\geq\rho_n(\lambda)\geq\ldots\geq 0\,,$$

la suite des valeurs propres de l'opérateur K_λ rangées par ordre décroissant et comptées avec leurs ordres de multiplicité. Partons de la caractérisation variationnelle de $\rho_0(\lambda)$ pour écrire que

$$\begin{aligned}\rho_0(\lambda) &= \sigma(1+\gamma)\max_{\substack{\psi\in L^2([-L,L])\\ \psi\neq 0}}\frac{\displaystyle\int_{-L}^{L}\psi(x)K_\lambda\psi(x)dx}{\displaystyle\int_{-L}^{L}|\psi(x)|^2dx}\\ &= \sigma(1+\gamma)\max_{\substack{\psi\in L^2([-L,L])\\ \psi\neq 0}}\frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L}E((\sigma+\lambda)|x-y|)\psi(x)\psi(y)dx}{\displaystyle\int_{-L}^{L}|\psi(x)|^2dx}\,.\end{aligned}$$

Evidemment, le maximum ci-dessus est atteint pour $\psi_\lambda = e_0$ (fonction propre de K_λ pour la valeur propre $\rho_0(\lambda)$). Comme la fonction E est à valeurs positives ou nulles,

$$\begin{aligned}E\geq 0 \quad\Rightarrow\quad &\frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L}E((\sigma+\lambda)|x-y|)\psi(x)\psi(y)dx}{\displaystyle\int_{-L}^{L}|\psi(x)|^2dx}\\ &\leq\frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L}E((\sigma+\lambda)|x-y|)|\psi(x)||\psi(y)|dx}{\displaystyle\int_{-L}^{L}|\psi(x)|^2dx}\,,\end{aligned}$$

de sorte que

$$\rho_0(\lambda) \leq \sigma(1+\gamma) \max_{\substack{\psi \in L^2([-L,L]) \\ \psi \neq 0}} \frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\psi(x)||\psi(y)|dx}{\displaystyle\int_{-L}^{L}|\psi(x)|^2 dx}\,.$$

D'autre part

$$\begin{aligned}
&\sigma(1+\gamma) \max_{\substack{\psi \in L^2([-L,L]) \\ \psi \neq 0}} \frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\psi(x)||\psi(y)|dx}{\displaystyle\int_{-L}^{L}|\psi(x)|^2 dx} \\
&= \sigma(1+\gamma) \max_{\substack{0\leq\phi \in L^2([-L,L]) \\ \phi \neq 0}} \frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)\phi(x)\phi(y)dx}{\displaystyle\int_{-L}^{L}\phi(x)^2 dx} \\
&\leq \sigma(1+\gamma) \max_{\substack{\psi \in L^2([-L,L]) \\ \psi \neq 0}} \frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)\psi(x)\psi(y)dx}{\displaystyle\int_{-L}^{L}|\psi(x)|^2 dx} = \rho_0(\lambda)\,.
\end{aligned}$$

Par conséquent

$$\rho_0(\lambda) = \sigma(1+\gamma) \max_{\substack{\psi \in L^2([-L,L]) \\ \psi \neq 0}} \frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\psi(x)||\psi(y)|dx}{\displaystyle\int_{-L}^{L}|\psi(x)|^2 dx}\,.$$

Lemme 6.4.7 *L'opérateur K_λ admet une fonction propre presque partout positive pour la valeur propre $\rho_0(\lambda)$.*

Démonstration. Soit e_0, fonction propre de K_λ pour la valeur propre $\rho_0(\lambda)$, telle que $\|e_0\|_{L^2([-L,L])} = 1$. Le raisonnement ci-dessus montre que

$$\begin{aligned}
\rho_0(\lambda) &= \sigma(1+\gamma)\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)e_0(x)e_0(y)dxdy \\
&= \sigma(1+\gamma)\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|e_0(x)||e_0(y)|dxdy\,.
\end{aligned}$$

Comme $E(z) > 0$ pour tout $z > 0$, la condition

$$\begin{aligned}&\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)e_0(x)e_0(y)dxdy\\ &= \int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|e_0(x)||e_0(y)|dxdy\\ &= \left|\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)e_0(x)e_0(y)dxdy\right|\end{aligned}$$

implique qu'il existe une constante $C \in \mathbf{R}$ telle que

$$e_0(x)e_0(y) = C|e_0(x)||e_0(y)|$$

pour presque tout $(x, y) \in [-L, L]^2$ (voir par exemple Théorème 1.39 (c) de [46]).

Evidemment $C \neq 0$; sinon on aurait $e_0(x)e_0(y) = 0$ pour presque tout $(x, y) \in [-L, L]^2$, ce qui est impossible puisque

$$\iint_{[-L,L]^2} |e_0(x)e_0(y)|^2 dxdy = \int_{-L}^{L} |e_0(x)|^2 dx \int_{-L}^{L} |e_0(y)|^2 dy = 1\,.$$

En multipliant chaque membre de cette égalité par $|e_0(y)|$ et en intégrant par rapport à y, on trouve que

$$e_0(x)\int_{-L}^{L} e_0(y)|e_0(y)|dy = C|e_0(x)|$$

pour presque tout $x \in [-L, L]$, puisque $\|e_0\|_{L^2([-L,L])} = 1$. Notons

$$C' = \int_{-L}^{L} e_0(y)|e_0(y)|dy\,,$$

en remarquant que la fonction $y \mapsto e_0(y)|e_0(y)|$ est intégrable sur $[-L, L]$ par hypothèse. On a évidemment $C' \neq 0$, sinon $|e_0(x)| = 0$ pour presque tout $x \in [-L, L]$ puisque $C \neq 0$, ce qui est impossible car $\|e_0\|_{L^2([-L,L])} = 1$. Il s'ensuit que

$$e_0(x) = \frac{C}{C'}|e_0(x)|\,, \qquad \text{pour presque tout } x \in [-L, L]\,.$$

En particulier

$$K_\lambda|e_0| = \rho_0(\lambda)|e_0|\,,$$

d'où le résultat annoncé. ■

6.4.4 Dépendance en λ de la valeur propre ρ_0

Le résultat principal de cette section est la proposition ci-dessous.

Proposition 6.4.8 *La fonction $\lambda \mapsto \rho_0(\lambda)$ vérifie les propriétés suivantes :*
(a) $\rho_0(\lambda) \to 0$ lorsque $\lambda \to +\infty$;
(b) $\rho_0(\lambda) \to +\infty$ lorsque $\lambda \to -\sigma$;
(c) ρ_0 est strictement décroissante ;
(d) ρ_0 est continue sur $]-\sigma, +\infty[$.

Démonstration. Pour établir le a), on rappelle que, d'après la preuve du Lemme 6.4.6

$$K_\lambda \psi(x) = \langle \phi \rangle \,,$$

où ϕ est la solution généralisée du problème aux limites

$$\begin{cases} (\lambda+\sigma)\phi(x,\mu) + \mu \frac{\partial \phi}{\partial x}(x,\mu) = \psi(x)\,, & |x| \le L\,,\ 0 < \mu \le 1\,, \\ \\ \phi(-L,\mu) = \phi(L,-\mu) = 0\,, & 0 < \mu \le 1\,. \end{cases}$$

Donc

$$\begin{aligned} \int_{-L}^{L} \psi(x) K_\lambda \psi(x) dx &= \tfrac{1}{2} \int_{-L}^{L} \int_{-1}^{1} \psi(x)\phi(x,\mu) d\mu dx \\ &\ge \tfrac{1}{2}(\lambda+\sigma) \int_{-L}^{L} \int_{-1}^{1} \phi(x,\mu)^2 d\mu dx \,. \end{aligned}$$

En appliquant l'inégalité de Cauchy-Schwarz, on trouve que

$$\begin{aligned} \tfrac{1}{2}(\lambda+\sigma) \int_{-L}^{L} \int_{-1}^{1} \phi(x,\mu)^2 d\mu dx &\le \tfrac{1}{2} \int_{-L}^{L} \int_{-1}^{1} \psi(x)\phi(x,\mu) d\mu dx \\ &\le \left(\tfrac{1}{2} \int_{-L}^{L} \int_{-1}^{1} \phi(x,\mu)^2 d\mu dx \right)^{1/2} \left(\tfrac{1}{2} \int_{-L}^{L} \int_{-1}^{1} \psi(x)^2 d\mu dx \right)^{1/2} \end{aligned}$$

ce qui montre que

$$\begin{aligned} \left(\int_{-L}^{L} \langle \phi \rangle(x)^2 dx \right)^{1/2} &\le \left(\tfrac{1}{2} \int_{-L}^{L} \int_{-1}^{1} \phi(x,\mu)^2 d\mu dx \right)^{1/2} \\ &\le \frac{1}{\lambda+\sigma} \left(\int_{-L}^{L} \psi(x)^2 dx \right)^{1/2} . \end{aligned}$$

Autrement dit

$$\| K_\lambda \psi \|_{L^2([-L,L])} \le \frac{1}{\lambda+\sigma} \| \psi \|_{L^2([-L,L])} \,.$$

En particulier

$$\int_{-L}^{L} \psi(x) K_\lambda \psi(x) dx \le \| \psi \|_{L^2([-L,L])} \| K_\lambda \psi \|_{L^2([-L,L])} \le \frac{1}{\lambda+\sigma} \| \psi \|^2_{L^2([-L,L])} \,,$$

d'où

$$\rho_0(\lambda) = \sigma(1+\gamma) \max_{\substack{\psi \in L^2([-L,L]) \\ \psi \neq 0}} \frac{\displaystyle\int_{-L}^{L} \psi(x) K_\lambda \psi(x) dx}{\displaystyle\int_{-L}^{L} |\psi(x)|^2 dx} \leq \frac{\sigma(1+\gamma)}{\sigma+\lambda}\,,$$

ce qui montre que $\rho_0(\lambda) \to 0$ lorsque $\lambda \to +\infty$.

Démontrons le point (b). Pour cela, on utilise la minoration de $\rho_0(\lambda)$ obtenue en remplaçant la fonction test ψ du problème de maximisation par une fonction constante :

$$\begin{aligned}\rho_0(\lambda) = \sigma(1+\gamma) \max_{\substack{\psi \in L^2([-L,L]) \\ \psi \neq 0}} & \frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)\psi(x)\psi(y)dxdy}{\displaystyle\int_{-L}^{L} |\psi(x)|^2 dx} \\ & \geq \sigma(1+\gamma) \frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)dxdy}{\displaystyle\int_{-L}^{L} dx}\,.\end{aligned}$$

Or, comme E est une fonction décroissante sur $\mathbf{R}_+$ et que $E(z) \to +\infty$ lorsque $z \to 0^+$, on conclut par convergence monotone que

$$\lim_{\lambda \to -\sigma} \int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)dxdy = +\infty\,,$$

ce qui entraîne que

$$\lim_{\lambda \to -\sigma} \rho_0(\lambda) = +\infty\,.$$

Le point (c) découle de la formule

$$\rho_0(\lambda) = \sigma(1+\gamma) \max_{\substack{\psi \in L^2([-L,L]) \\ \psi \neq 0}} \frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\psi(x)||\psi(y)|dxdy}{\displaystyle\int_{-L}^{L} |\psi(x)|^2 dx}$$

établie plus haut. Soient donc $\lambda > \lambda' \in]-\sigma, +\infty[$, et soit $\psi_\lambda \in L^2([-L,L])$ réalisant le maximum ci-dessus. (Rappelons que ce maximum est atteint pour

ψ_λ, fonction propre de l'opérateur K_λ pour la valeur propre $\rho_0(\lambda)$). Alors

$$\rho_0(\lambda) = \sigma(1+\gamma)\frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\psi_\lambda(x)||\psi_\lambda(y)|dxdy}{\displaystyle\int_{-L}^{L}|\psi_\lambda(x)|^2dx}$$

$$< \sigma(1+\gamma)\frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda')|x-y|)|\psi_\lambda(x)||\psi_\lambda(y)|dxdy}{\displaystyle\int_{-L}^{L}|\psi_\lambda(x)|^2dx} \leq \rho_0(\lambda')\,.$$

En effet, la deuxième inégalité ci-dessus découle de la formule

$$\rho_0(\lambda') = \sigma(1+\gamma)\max_{\substack{\psi\in L^2([-L,L])\\ \psi\neq 0}} \frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda')|x-y|)|\psi(x)||\psi(y)|dxdy}{\displaystyle\int_{-L}^{L}|\psi(x)|^2dx}\,.$$

D'autre part, l'inégalité stricte, qui est le point crucial de l'argument, s'obtient en remarquant que E est strictement décroissante d'après le Lemme 6.4.2 (a). Par conséquent, pour tous $x \neq y \in [-L, L]$, on a

$$E((\sigma+\lambda')|x-y|) > E((\sigma+\lambda)|x-y|)\,,$$

de sorte que

$$\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda')|x-y|)|\psi_\lambda(x)||\psi_\lambda(y)|dxdy$$
$$> \int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\psi_\lambda(x)||\psi_\lambda(y)|dxdy\,.$$

En effet, si les deux intégrales ci-dessus étaient égales, on aurait alors

$$(E((\sigma+\lambda')|x-y|)-E((\lambda+\sigma)|x-y|))|\psi_\lambda(x)||\psi_\lambda(y)| = 0 \text{ p.p. en } (x,y)\in[-L,L]^2\,,$$

c'est-à-dire que

$$|\psi_\lambda(x)||\psi_\lambda(y)| = 0 \text{ p.p. en } (x,y)\in[-L,L]^2\setminus\{(z,z) \text{ t.q. } |z|\leq L\}\,.$$

On en déduirait alors que

$$0 = \iint_{[-L,L]^2}|\psi_\lambda(x)||\psi_\lambda(y)|dxdy = \left(\int_{-L}^{L}|\psi_\lambda(z)|dz\right)^2$$

de sorte que $\psi_\lambda = 0$ p.p. sur $[-L, L]$. Or ceci est impossible puisque par hypothèse ψ_λ réalise

$$\max_{\substack{\psi\in L^2([-L,L])\\ \psi\neq 0}} \frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\psi(x)||\psi(y)|dxdy}{\displaystyle\int_{-L}^{L}|\psi(x)|^2dx}$$

ce qui entraîne en particulier que ψ_λ est non p.p. nul. Ceci démontre l'inégalité stricte, et donc que le point (c) est bien vérifié.

Passons à la démonstration du point (d). Soit donc $\lambda \in]-\sigma, +\infty[$ et une suite $\lambda_n \to \lambda$ pour $n \to +\infty$. Soit ψ_λ réalisant

$$\max_{\substack{\psi \in L^2([-L,L]) \\ \psi \neq 0}} \frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\psi(x)||\psi(y)|dxdy}{\displaystyle\int_{-L}^{L}|\psi(x)|^2dx}\,.$$

Evidemment

$$\rho_0(\lambda_n) \geq \sigma(1+\gamma)\frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda_n)|x-y|)|\psi_\lambda(x)||\psi_\lambda(y)|dxdy}{\displaystyle\int_{-L}^{L}|\psi_\lambda(x)|^2dx}$$

$$\to \sigma(1+\gamma)\frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\psi_\lambda(x)||\psi_\lambda(y)|dxdy}{\displaystyle\int_{-L}^{L}|\psi_\lambda(x)|^2dx} = \rho_0(\lambda)\,,$$

de sorte que

$$\underline{\lim}_{n\to+\infty}\, \rho_0(\lambda_n) \geq \rho_0(\lambda)\,.$$

Soit d'autre part, pour tout $n \geq 1$ une fonction $\psi_n \in L^2([-L,L])$ réalisant

$$\max_{\substack{\psi \in L^2([-L,L]) \\ \psi \neq 0}} \frac{\displaystyle\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda_n)|x-y|)|\psi(x)||\psi(y)|dxdy}{\displaystyle\int_{-L}^{L}|\psi(x)|^2dx}\,,$$

et soit $\eta > 0$.

D'une part, d'après le théorème des accroissements finis

$$\begin{aligned}|E((\sigma+\lambda_n)|x-y|) - E((\sigma+\lambda)|x-y|)| &\leq \frac{e^{-\sigma\eta/2}}{\sigma\eta/2}2L|\lambda-\lambda_n| \\ &= \frac{4Le^{-\sigma\eta/2}}{\sigma\eta}|\lambda-\lambda_n|\end{aligned}$$

pour tout n assez grand tel que $\lambda_n > -\sigma/2$, et tous $x, y \in [-L, L]$ tels que $|x-y| \geq \eta$.

D'autre part, comme $\lambda_n \to \lambda$ lorsque $n \to +\infty$, il existe un réel $M > 0$ tel que $\lambda_n \leq M$ pour tout $n \geq 1$. Choisissons η tel que $0 < \eta < \frac{1}{\sigma+M}$. D'après le Lemme 6.4.2 (c), $E(z) \sim -\frac{1}{2}\ln z$ pour $z \to 0^+$, de sorte qu'il existe $C > 0$ tel que

$$|E((\sigma+\lambda_n)|x-y|) - E((\sigma+\lambda)|x-y|)| \leq C|\ln(\tfrac{\sigma}{2}|x-y|)|$$

pour tout n assez grand tel que $\lambda_n > -\sigma/2$, et tous $x, y \in [-L, L]$ tels que $0 < |x-y| < \eta < 1/(\sigma + M)$.

Donc, en notant $\chi_n = \psi_n / \|\psi_n\|_{L^2([-L,L])}$, on a

$$\begin{aligned}
&\left| \int_{-L}^{L} \int_{-L}^{L} E((\sigma+\lambda_n)|x-y|)|\chi_n(x)||\chi_n(y)|dxdy \right. \\
&\qquad \left. - \int_{-L}^{L} \int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\chi_n(x)||\chi_n(y)|dxdy \right| \\
&\leq \iint_{[-L,L]^2} |E((\sigma+\lambda_n)|x-y|) - E((\sigma+\lambda)|x-y|)||\chi_n(x)||\chi_n(y)|dxdy \\
&= \iint_{\substack{|x|,|y|\leq L \\ |x-y|\geq\eta}} |E((\sigma+\lambda_n)|x-y|) - E((\sigma+\lambda)|x-y|)||\chi_n(x)||\chi_n(y)|dxdy \\
&+ \iint_{\substack{|x|,|y|\leq L \\ 0<|x-y|<\eta}} |E((\sigma+\lambda_n)|x-y|) - E((\sigma+\lambda)|x-y|)||\chi_n(x)||\chi_n(y)|dxdy \\
&\leq \frac{4Le^{-\sigma\eta/2}}{\sigma\eta} |\lambda - \lambda_n| \iint_{\substack{|x|,|y|\leq L \\ |x-y|\geq\eta}} |\chi_n(x)||\chi_n(y)|dxdy \\
&+ C \iint_{\substack{|x|,|y|\leq L \\ 0<|x-y|<\eta}} |\ln(\tfrac{\sigma}{2}|x-y|)| \tfrac{1}{2} \left(\chi_n(x)^2 + \chi_n(y)^2\right) dxdy \,.
\end{aligned}$$

D'une part

$$\iint_{\substack{|x|,|y|\leq L \\ |x-y|\geq\eta}} |\chi_n(x)||\chi_n(y)|dxdy \leq \iint_{|x|,|y|\leq L} |\chi_n(x)||\chi_n(y)|dxdy \leq 2L$$

en appliquant l'inégalité de Cauchy-Schwarz, tandis que

$$\begin{aligned}
&\iint_{\substack{|x|,|y|\leq L \\ 0<|x-y|<\eta}} |\ln(\tfrac{\sigma}{2}|x-y|)| \tfrac{1}{2} \left(\chi_n(x)^2 + \chi_n(y)^2\right) dxdy \\
&= \iint_{\substack{|x|,|y|\leq L \\ 0<|x-y|<\eta}} |\ln(\tfrac{\sigma}{2}|x-y|)| \chi_n(y)^2 dxdy \\
&\leq \iint_{\mathbf{R}^2} \mathbf{1}_{0<|x-y|<\eta} |\ln(\tfrac{\sigma}{2}|x-y|)| \chi_n(y)^2 dxdy \\
&= \int_{|z|<\eta} |\ln(\tfrac{\sigma}{2}|z|)| dz \int_{\mathbf{R}} \chi_n(y)^2 dy \\
&\leq \tfrac{2}{\sigma} \int_0^{\sigma\eta/2} -\ln z dz = \eta + \eta |\ln(\tfrac{\sigma}{2}\eta)| \,,
\end{aligned}$$

(en prolongeant χ_n par 0 dans $\mathbf{R} \setminus [-L, L]$).

Au total, on a montré que

$$\left|\rho_0(\lambda_n) - \int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\chi_n(x)||\chi_n(y)|dxdy\right| \leq \frac{8L^2 e^{-\sigma\eta/2}}{\sigma\eta}|\lambda-\lambda_n| + C\eta + C\eta|\ln(\tfrac{\sigma}{2}\eta)|\,,$$

de sorte qu'en optimisant en $\eta > 0$, on trouve que

$$\rho_0(\lambda_n) - \int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\chi_n(x)||\chi_n(y)|dxdy \to 0$$

lorsque $n \to +\infty$. Donc

$$\begin{aligned}\overline{\lim_{n\to+\infty}}\rho_0(\lambda_n) &\leq \overline{\lim_{n\to+\infty}}\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\chi_n(x)||\chi_n(y)|dxdy \\ &\leq \max_{\|\chi\|_{L^2}=1}\int_{-L}^{L}\int_{-L}^{L} E((\sigma+\lambda)|x-y|)|\chi(x)||\chi(y)|dxdy = \rho_0(\lambda)\,.\end{aligned}$$

Au total, on a donc montré que

$$\overline{\lim_{n\to+\infty}}\,\rho_0(\lambda_n) \leq \rho_0(\lambda) \leq \underline{\lim_{n\to+\infty}}\,\rho_0(\lambda_n)\,,$$

d'où

$$\rho_0(\lambda_n) \to \rho_0(\lambda) \text{ lorsque } n \to +\infty\,.$$

Comme ceci vaut pour tout λ et toute suite $\lambda_n \to \lambda$ lorsque $n \to +\infty$, on conclut que ρ_0 est continue sur $]-\sigma, +\infty[$. ■

6.4.5 Valeur propre principale de l'opérateur de Boltzmann linéaire

Grâce aux préparations ci-dessus, nous pouvons maintenant conclure notre étude de la criticité pour l'équation de Boltzmann linéaire monocinétique avec scattering isotrope posée dans l'intervalle $[-L, L]$.

En effet, d'après la proposition ci-dessus, la plus grande valeur propre $\rho_0(\lambda)$ de l'opérateur de Hilbert-Schmidt K_λ définit une fonction

$$\begin{aligned}]-\sigma, +\infty[\ni \lambda \mapsto \rho_0(\lambda) \in]0, +\infty[&\text{ continue, strictement décroissante} \\ &\text{et telle que } \lim_{\lambda\to-\sigma}\rho_0(\lambda) = +\infty\,, \quad \lim_{\lambda\to+\infty}\rho_0(\lambda) = 0\,.\end{aligned}$$

D'après le théorème des valeurs intermédiaires, il existe une unique valeur

$$\lambda_L(\sigma,\gamma) \in]-\sigma, +\infty[$$

telle que

$$\rho_0(\lambda_L(\sigma,\gamma)) = \frac{1}{\sigma(1+\gamma)}\,.$$

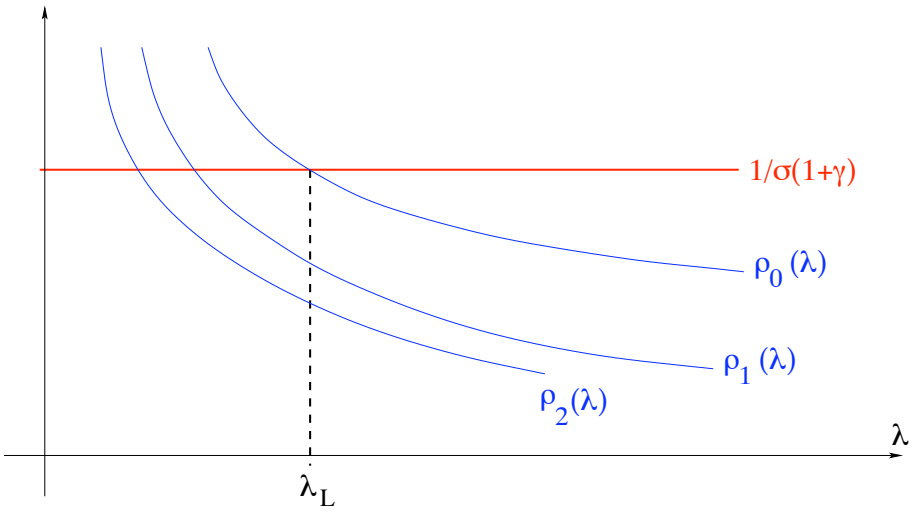

FIGURE 6.1 – Valeurs propres $\rho_0(\lambda) \geq \rho_1(\lambda) \geq \rho_2(\lambda) \geq \dots$ de l'opérateur K_λ et plus grande valeur propre $\lambda_L \equiv \lambda_L(\sigma,\gamma)$ de l'opérateur de Boltzmann linéaire A avec conditions aux limites absorbantes (cas du transfert radiatif avec scattering isotrope, dans la géométrie de la plaque infinie).

Autrement dit, 1 est valeur propre de l'opérateur $\sigma(1+\gamma)K_{\lambda_L(\sigma,\gamma)}$, ce qui veut dire, d'après la Proposition 6.4.4 que $\lambda_L(\sigma,\gamma)$ est valeur propre de l'opérateur de Boltzmann linéaire A, défini par (6.13), avec conditions aux limites absorbantes sur $[-L,L]\times[-1,1]$.

Vérifions que $\lambda_L(\sigma,\gamma)$ est la plus grande valeur propre de A avec conditions aux limites absorbantes. Soit donc $\lambda > \lambda_L(\sigma,\gamma)$; si λ était valeur propre de A avec conditions aux limites absorbantes, 1 serait valeur propre de l'opérateur $\sigma(1+\gamma)K_\lambda$, d'après la Proposition 6.4.4. Or, comme ρ_0 est strictement décroissante,

$$\rho_0(\lambda) < \rho_0(\lambda_L(\sigma,\gamma)) = \frac{1}{\sigma(1+\gamma)}\,,$$

ce qui contredit le fait que 1 soit valeur propre de $\sigma(1+\gamma)K_\lambda$.

Enfin, d'après le Lemme 6.4.7, l'opérateur $K_{\lambda_L(\sigma,\gamma)}$ admet une fonction propre e_0 presque partout positive ou nulle sur $[-L,L]$ pour la valeur propre $\rho_0(\lambda_L(\sigma,\gamma)) = \frac{1}{\sigma(1+\gamma)}$. Soit ϕ_0 définie par

$$\phi_0(x,\mu) = \int_{-L}^{x} \tfrac{\sigma(1+\gamma)}{\mu} e^{-(\sigma+\lambda)(x-y)/\mu} e_0(y)dy\,, \qquad |x|\leq L\,, \quad 0<\mu\leq 1\,,$$

$$\phi_0(x,\mu) = \int_{x}^{L} \tfrac{\sigma(1+\gamma)}{|\mu|} e^{-(\sigma+\lambda)(y-x)/|\mu|} e_0(y)dy\,, \qquad |x|\leq L\,, \quad -1\leq\mu<0\,.$$

Ces formules montrent que $\phi_0 \geq 0$ presque partout sur $[-L,L]\times[-1,1]$; et c'est une fonction propre de A pour la valeur propre $\lambda_L(\sigma,\gamma)$ d'après la démonstration de la Proposition 6.4.4. Ceci conclut la démonstration du Théorème 6.4.1.

6.4.6 Taille critique pour l'équation de Boltzmann linéaire

Terminons maintenant notre étude du problème aux valeurs propres pour l'équation de Boltzmann linéaire en dégageant la notion de taille critique.

Si $\phi \equiv \phi(x,\mu)$ est solution du problème aux valeurs propres sur le domaine $[-L,L]\times[-1,1]$ correspondant à la plus grande valeur propre $\lambda_L(\sigma,\gamma)$ de l'opérateur de Boltzmann linéaire A, défini par (6.13), alors la fonction

$$\Phi : (x,\mu) \mapsto \phi(Lx,\mu)$$

est solution du problème aux valeurs propres sur $[-1,1]\times[-1,1]$

$$\begin{cases} -\mu\frac{\partial\Phi}{\partial x} + \sigma L(1+\gamma)\langle\Phi\rangle - \sigma L\Phi = L\lambda_L(L\sigma,\gamma)\Phi\,, & \Phi \neq 0\,, \\ \\ \Phi(-1,\mu) = \Phi(1,-\mu) = 0\,, & 0 < \mu \leq 1\,, \end{cases}$$

de sorte que

$$\lambda_L(\sigma,\gamma) = \frac{\lambda_1(\sigma L,\gamma)}{L}\,.$$

Définition 6.4.9 *Soient $\sigma > 0$ et $\gamma > -1$ donnés. On dira que $2L$ est la taille critique pour l'opérateur de Boltzmann linéaire A, défini par (6.13), avec conditions aux limites absorbantes, posé sur $[-L,L]\times[-1,1]$, si la plus grande valeur propre de cet opérateur est nulle, c'est-à-dire si*

$$\lambda_L(\sigma,\gamma) = 0\,.$$

L'argument ci-dessus montre que, de façon équivalente, $2L$ est la taille critique pour l'opérateur de Boltzmann linéaire avec conditions aux limites absorbantes si et seulement si

$$\lambda_1(L\sigma,\gamma) = 0\,.$$

(Dans les applications à la neutronique, on parle souvent de "masse critique", ce qui est la même chose, quitte à multiplier la taille critique par la densité massique du matériau fissile considéré.)

Nous allons conclure cette section par quelques remarques qualitatives sur cette notion de masse critique, sans chercher à obtenir de résultat rigoureux, ni de démonstration.

Considérons le problème aux valeurs propres pour l'opérateur de Boltzmann linéaire A, défini par (6.13), posé sur le domaine $[-L,L]\times[-1,1]$ avec conditions aux limites absorbantes.

Nous allons étudier de plus près le cas où le paramètre γ est strictement positif mais $\ll 1$ — de sorte que le milieu est très légèrement amplificateur. D'autre part, nous allons supposer que le taux de perte de particules au bord du domaine spatial, du fait de la condition aux limites absorbantes, est compensé par l'épaisseur $2L$ du domaine, que l'on suppose $\gg 1$.

En même temps que le changement de variables $x = Lz$ ci-dessus, de sorte que $|z| \leq 1$, on réalise l'hypothèse $0 < \gamma \ll 1$ en posant

$$\gamma = \hat{\gamma}/L^2\,, \qquad \hat{\gamma} > 0\,,$$

et en supposant que $L \gg 1$.

Le problème spectral pour l'opérateur A à la plus grande valeur propre $\lambda_L(\sigma, \hat{\gamma}/L^2)$ mis à l'échelle sur le domaine $[-1,1] \times [-1,1]$ est donc

$$\begin{cases} -\mu \frac{\partial \Phi}{\partial x} + \sigma L \left(1 + \frac{\hat{\gamma}}{L^2}\right) \langle \Phi \rangle - \sigma L \Phi = L \lambda_L(\sigma, \hat{\gamma}/L^2) \Phi \,, & \Phi \neq 0 \,, \\ \\ \Phi(-1, \mu) = \Phi(1, -\mu) = 0 \,, & 0 < \mu \leq 1 \,. \end{cases}$$

Lorsque $L \gg 1$, le problème ci-dessus se présente naturellement sous une forme suggérant d'utiliser l'approximation par la diffusion. Sans rentrer dans les détails, l'idée est que

$$\lambda_L(\sigma, \hat{\gamma}/L^2) = O(1/L^2) \,,$$

ce qui suggère d'utiliser l'approximation par la diffusion de l'équation de Boltzmann linéaire ci-dessus et de chercher Φ sous la forme

$$\Phi(x, \mu) \simeq \langle \Phi \rangle(x) - \frac{1}{\sigma L} \mu \frac{d}{dx} \langle \Phi \rangle(x) + \dots$$

En moyennant le problème ci-dessus par rapport à μ, on trouve d'abord que

$$-\frac{d}{dx} \langle \mu \Phi \rangle + \frac{\sigma \hat{\gamma}}{L} \langle \Phi \rangle = L \lambda_L(L\sigma, \hat{\gamma}/L^2) \langle \Phi \rangle \,,$$

puis, en remplaçant Φ par son approximation ci-dessus,

$$\begin{cases} \frac{1}{3\sigma L} \frac{d^2}{dx^2} \langle \Phi \rangle + \frac{\sigma \hat{\gamma}}{L} \langle \Phi \rangle \simeq L \lambda_L(\sigma, \hat{\gamma}/L^2) \langle \Phi \rangle \,, \\ \\ \langle \Phi \rangle(-1) = \langle \Phi \rangle(+1) = 0 \,, \end{cases}$$

ce qui est précisément l'approximation par la diffusion de l'équation de Boltzmann linéaire ci-dessus vérifiée par Φ.

Ceci suggère alors de rechercher λ_{diff}, la plus grande valeur propre pour le problème aux limites de diffusion

$$\begin{cases} \frac{1}{3\sigma} \frac{d^2}{dx^2} q(x) + \sigma \hat{\gamma} q(x) = \lambda_{diff} q(x) \,, \\ \\ q(-1) = q(+1) = 0 \,, \end{cases}$$

que l'on trouve sans difficulté :

$$\lambda_{diff} = \sigma \hat{\gamma} - \frac{\pi^2}{12\sigma} \,.$$

Cette valeur propre correspond à la fonction propre

$$q(x) = \cos(\tfrac{\pi}{2} x) \,,$$

— cf. section 6.3. Puis

$$L^2 \lambda_L(\sigma, \hat{\gamma}/L^2) \simeq \lambda_{diff} = \sigma \hat{\gamma} - \frac{\pi^2}{12\sigma} \,,$$

c'est-à-dire que

$$\lambda_L\left(\sigma, \frac{\hat{\gamma}}{L^2}\right) \simeq \frac{\lambda_{diff}}{L^2} = \sigma\frac{\hat{\gamma}}{L^2} - \frac{\pi^2}{12\sigma L^2}\,.$$

Autrement dit,

$$\lambda_L(\sigma,\gamma) \simeq \sigma\gamma - \frac{\pi^2}{12\sigma L^2} + \dots\,,$$

ce qui suggère la valeur approchée suivante de la taille critique pour la bande $[-L, L]$, dans la limite où $L \gg 1$ et $\gamma \ll 1$

$$2L_c \simeq \frac{\pi}{\sqrt{3\gamma}\sigma}\,.$$

Cette estimation de la taille critique peut être améliorée de façon significative en utilisant l'approximation par la diffusion à l'ordre un, et la notion de longueur d'extrapolation obtenue au moyen de la fonction de Chandrasekhar (cf. chapitre 4). Sans entrer dans les détails, disons seulement qu'on trouverait la valeur approchée suivante de la taille critique

$$2L_c \simeq 2\frac{1}{\sqrt{3\gamma}\sigma}\arctan\left(\frac{\sigma}{\sqrt{3\gamma}\mathcal{L}}\right)\,,$$

où $\mathcal{L} \simeq 0.7104$ est le coefficient d'extrapolation. On remarquera que les deux valeurs approchées ci-dessus diffèrent de $O(\sqrt{\gamma})$ pour $\gamma \ll 1$.

6.5 Problèmes aux valeurs propres et criticité

6.5.1 Equations de diffusion

Les résultats de cette section sont valides pour une large classe de modèles de diffusion mais, pour simplifier l'exposé, nous considérons seulement le modèle (1.16) de diffusion à deux groupes d'énergie en neutronique

$$\begin{cases} \dfrac{\partial u_1}{\partial t} - \operatorname{div}\left(D_1\nabla u_1\right) + \sigma_1 u_1 = \sigma_{12}^f u_2 & \text{dans } \Omega\times\mathbf{R}^+, \\ \dfrac{\partial u_2}{\partial t} - \operatorname{div}\left(D_2\nabla u_2\right) + \sigma_2 u_2 = \sigma_{21}^c u_1 & \text{dans } \Omega\times\mathbf{R}^+, \\ u_1 = u_2 = 0 & \text{sur } \partial\Omega\times\mathbf{R}^+, \\ u_1(t=0,x) = u_1^0(x)\,, u_2(t=0,x) = u_2^0(x) & \text{dans } \Omega, \end{cases} \tag{6.16}$$

où Ω est un ouvert borné régulier et les coefficients sont des fonctions de $L^\infty(\Omega)$ vérifiant

$$D_1(x)\,, D_2(x)\,, \sigma_{12}^f(x)\,, \sigma_{21}^c(x) \geq C > 0 \quad \text{et } \sigma_1 \geq \sigma_{12}^f\,, \sigma_2 \geq \sigma_{21}^c\,. \tag{6.17}$$

L'hypothèse de positivité de σ_1 et σ_2 n'est pas restrictive car on peut toujours remplacer $u_1(t), u_2(t)$ par $v_1(t)e^{Ct}, v_2(t)e^{Ct}$ dans (6.16) et se ramener à un tel cas

pour $C > 0$ suffisamment grand. Le problème aux valeurs propres correspondant à ce problème d'évolution est : trouver $\lambda \in \mathbf{C}$ et $(\psi_1, \psi_2) \neq 0$ tels que

$$\left\{ \begin{array}{ll} -\lambda\psi_1 - \mathrm{div}\,(D_1\nabla\psi_1) + \sigma_1\psi_1 = \sigma_{12}^f\psi_2 & \text{dans } \Omega, \\ -\lambda\psi_2 - \mathrm{div}\,(D_2\nabla\psi_2) + \sigma_2\psi_2 = \sigma_{21}^c\psi_1 & \text{dans } \Omega, \\ \psi_1 = \psi_2 = 0 & \text{sur } \partial\Omega. \end{array} \right. \tag{6.18}$$

Une conséquence du Théorème 6.2.14 de Krein-Rutman (qui généralise celui de Perron-Frobenius) est le résultat suivant que nous énonçons volontairement sans préciser les espaces fonctionnels (voir [43] pour plus de détails).

Proposition 6.5.1 *Si Ω est un ouvert borné régulier, il existe une plus petite valeur propre réelle et simple λ solution de (6.18) telle que sa fonction propre associée (ψ_1, ψ_2) est la seule à être strictement positive dans Ω et pour toute autre valeur propre $\mu \in \mathbf{C}$ on a $\lambda < |\mu|$.*

Remarque 6.5.2 *Puisque seule la fonction propre associée à la plus petite valeur propre est positive, c'est la seule qui puisse s'interpréter comme une densité de particules. Autrement dit, les autres fonctions propres, si elles existent, n'ont pas d'interprétation physique claire.*

Pour se convaincre du résultat de la Proposition 6.5.1 on peut démontrer son analogue en dimension finie pour la matrice issue d'une discrétisation spatiale de (6.18).

Lemme 6.5.3 *Sous les hypothèses (6.17) la matrice de discrétisation de (6.18) par différences finies en dimension $N = 1$ d'espace est une M-matrice irréductible (à laquelle on peut donc appliquer le Théorème 6.2.11).*

Démonstration. Pour simplifier on suppose tous les coefficients constants dans (6.18). On range les inconnues discrètes en les regroupant par groupe d'énergie, c'est-à-dire qu'on discrétise d'abord ψ_1 puis ψ_2. Cette matrice de discrétisation (pour $\lambda = 0$) a une forme par blocs

$$A = \begin{pmatrix} A_{11} & -\sigma_{12}^f\,\mathrm{Id} \\ -\sigma_{21}^c\,\mathrm{Id} & A_{22} \end{pmatrix},$$

avec les blocs diagonaux, pour $k = 1, 2$,

$$A_{kk} = \begin{pmatrix} \sigma_k + 2c_k & -c_k & & & 0 \\ -c_k & \sigma_k + 2c_k & -c_k & & \\ & \ddots & \ddots & \ddots & \\ & & -c_k & \sigma_k + 2c_k & -c_k \\ 0 & & & -c_k & \sigma_k + 2c_k \end{pmatrix} \text{ avec } c_k = \frac{D_k}{(\Delta x)^2}.$$

Il s'agit bien d'une M-matrice car tous les coefficients extra-diagonaux sont négatifs ou nuls tandis que les coefficients diagonaux sont positifs et la dominance

diagonale est bien satisfaite. Comme σ_{12}^f et σ_{21}^c sont non nuls on peut construire une chaîne d'arêtes, $a_{ij} \neq 0$, reliant n'importe quels nœuds i et j du graphe de connectivité de A. Elle est donc bien irréductible. ■

Pour voir les conséquences de la Proposition 6.5.1 sur le comportement en temps grand de (6.16) nous avons besoin d'introduire le problème adjoint de (6.18), de la même manière que nous avons dû introduire la matrice adjointe dans la Proposition 6.2.15 pour un modèle analogue en dimension finie. La définition formelle de l'adjoint A^* d'un opérateur A dans un espace de Hilbert V, muni du produit scalaire $\langle u, v\rangle$, est

$$\langle Au, v\rangle = \langle u, A^*v\rangle \quad \text{pour tout } u, v \in V.$$

Ici, on choisit $V = L^2(\Omega)^2$ et on définit, pour des fonctions $u = (u_1, u_2)$ suffisamment régulières, l'opérateur A par

$$Au = \begin{pmatrix} -\operatorname{div}(D_1 \nabla u_1) + \sigma_1 u_1 - \sigma_{12}^f u_2 \\ -\operatorname{div}(D_2 \nabla u_2) + \sigma_2 u_2 - \sigma_{21}^c u_1 \end{pmatrix}.$$

Un calcul simple, pour une fonction régulière $v = (v_1, v_2)$, montre que

$$\langle Au, v\rangle = \int_\Omega \Big(D_1 \nabla u_1 \cdot \nabla v_1 + D_2 \nabla u_2 \cdot \nabla v_2 + \sigma_1 u_1 v_1 + \sigma_2 u_2 v_2 - \sigma_{12}^f u_2 v_1 - \sigma_{21}^c u_1 v_2 \Big) dx$$

et donc que

$$A^*v = \begin{pmatrix} -\operatorname{div}(D_1 \nabla v_1) + \sigma_1 v_1 - \sigma_{21}^c v_2 \\ -\operatorname{div}(D_2 \nabla v_2) + \sigma_2 v_2 - \sigma_{12}^f v_1 \end{pmatrix}.$$

On définit donc le **problème adjoint** de (6.18)

$$\begin{cases} -\lambda \psi_1^* - \operatorname{div}(D_1 \nabla \psi_1^*) + \sigma_1 \psi_1^* = \sigma_{21}^c \psi_2^* & \text{dans } \Omega, \\ -\lambda \psi_2^* - \operatorname{div}(D_2 \nabla \psi_2^*) + \sigma_2 \psi_2^* = \sigma_{12}^f \psi_1^* & \text{dans } \Omega, \\ \psi_1^* = \psi_2^* = 0 & \text{sur } \partial\Omega. \end{cases}$$

Remarquons que ce problème adjoint ressemble à (6.18) à ceci près que les rôles et places de σ_{21}^c et σ_{12}^f ont été échangés. On peut, bien sûr, appliquer la Proposition 6.5.1 à ce problème adjoint qui admet, comme en dimension finie, la même plus petite valeur propre λ que le problème "direct" (6.18), avec un vecteur propre positif (ψ_1^*, ψ_2^*). On peut alors compléter la Proposition 6.5.1 par le résultat suivant.

Proposition 6.5.4 *On suppose que la donnée initiale de (6.16) est positive, non nulle. Une condition nécessaire pour que le problème d'évolution (6.16) admette une limite non nulle quand $t \to +\infty$ est que la plus petite valeur propre de (6.18) soit $\lambda = 0$.*

Si cette limite existe, alors elle est du type

$$\lim_{t \to +\infty} (u_1(t), u_2(t)) = c\,(\psi_1, \psi_2) \quad \textit{avec } c = \int_\Omega \left(u_1^0 \psi_1^* + u_2^0 \psi_2^* \right) dx\,.$$

On voit que, comme en dimension finie, le profil spatial asymptotique est toujours le premier et seul vecteur propre positif (ψ_1, ψ_2). Par ailleurs, le coefficient de proportionnalité c est le produit scalaire de la donnée initiale et du premier vecteur propre adjoint (ψ_1^*, ψ_2^*), positif lui aussi. De ce fait, ce vecteur propre adjoint est aussi appelé **fonction d'importance** par les neutroniciens. Nous reviendrons par la suite sur le rôle de cette fonction d'importance.

En neutronique les physiciens et ingénieurs préfèrent une autre formulation du problème aux valeurs propres (6.18), appelée **problème critique** que nous allons maintenant décrire. Dans (6.18) la valeur propre λ est introduite naturellement comme un taux de décroissance exponentiel en temps. Dans le problème critique on introduit une autre valeur propre $1/k$ qui est un facteur de proportionnalité du taux de fission

$$\begin{cases} -\operatorname{div}(D_1 \nabla \psi_1) + \sigma_1 \psi_1 = \dfrac{1}{k} \sigma_{12}^f \psi_2 & \text{dans } \Omega, \\ -\operatorname{div}(D_2 \nabla \psi_2) + \sigma_2 \psi_2 = \sigma_{21}^c \psi_1 & \text{dans } \Omega, \\ \psi_1 = \psi_2 = 0 & \text{sur } \partial\Omega. \end{cases} \tag{6.19}$$

On peut encore appliquer le Théorème 6.2.14 de Krein-Rutman (qui généralise celui de Perron-Frobenius) pour obtenir :

Proposition 6.5.5 *Si Ω est un ouvert borné régulier, il existe une plus petite valeur propre réelle et simple $1/k_{\text{eff}}$ solution du problème critique (6.19) telle que son vecteur propre associé (ψ_1, ψ_2) est le seul à être strictement positif dans Ω et pour toute autre valeur propre $1/k \in \mathbf{C}$ on a $1/k_{\text{eff}} < 1/|k|$.*

La valeur k_{eff} est appelée **facteur multiplicatif effectif**.

Remarquons que $k_{\text{eff}} = 1$ dans (6.19) si et seulement si la plus petite valeur propre de (6.18) est $\lambda = 0$ et que, dans ce cas, les vecteurs propres associés sont les mêmes. Si on a $k_{\text{eff}} > 1$, alors on aurait $\lambda = 0$ pour le problème (6.18) où le taux de fission σ_{12}^f serait divisé par k_{eff}. Autrement dit, il ne peut exister une limite finie du problème d'évolution (6.16) que si on diminue les fissions. Inversement, si $k_{\text{eff}} < 1$, il ne peut exister une limite finie non nulle du problème d'évolution (6.16) que si on augmente les fissions.

Ces constatations nous amènent à l'interprétation suivante du facteur multiplicatif effectif :

1. si $k_{\text{eff}} = 1$, le milieu est dit **critique** (les réactions de fission sont équilibrées par la diffusion et l'absorption),
2. si $k_{\text{eff}} > 1$, le milieu est dit **sur-critique** (les réactions de fission dominent la diffusion et l'absorption),
3. si $k_{\text{eff}} < 1$, le milieu est dit **sous-critique** (les réactions de fission sont trop faibles en comparaison de la diffusion et l'absorption).

Un des intérêts de la formulation (6.19) du problème critique par rapport au problème aux valeurs propres (6.18) est que la connaissance du facteur multiplicatif effectif k_{eff} permet de savoir de combien il faudrait modifier la section

efficace de fission pour que le milieu soit critique. Par exemple, dans un réacteur nucléaire on peut jouer sur les barres de contrôle, les "poisons" consommables (bore dilué dans l'eau, carbure de bore ou cadmium solide) qui absorbent préférentiellement les neutrons, pour régler la réactivité et avoir en permanence $k_{\text{eff}} = 1$. Nous reviendrons plus loin sur cet aspect de sensibilité du k_{eff} aux sections efficaces.

Remarque 6.5.6 *Bien sûr, les mêmes résultats s'appliquent au cas d'une seule équation de diffusion. Comme on l'a vu dans la Section 6.3, la théorie est beaucoup plus simple dans ce cas particulier car l'opérateur correspondant est* **auto-adjoint** *et on peut donc trouver une base hilbertienne de fonctions propres, ce qui n'est pas forcément le cas pour le modèle à deux groupes (6.16).*

6.5.2 Equation de transport

Les résultats de la Section 6.5.1 se transposent aux modèles de transport. Nous écrivons directement le problème critique pour l'équation de Boltzmann (6.1). Il s'agit de trouver $1/k \in \mathbf{C}$ et une fonction non nulle $\phi(x,v)$ tels que

$$\left\{\begin{array}{l} v\cdot\nabla\phi + \sigma(x)\phi - \displaystyle\int_{|v'|=1} \sigma^c(x, v\cdot v')\phi(x,v')\,dv' = \\ \qquad\qquad \dfrac{1}{k}\displaystyle\int_{|v'|=1} \sigma^f(x, v\cdot v')\phi(x,v')\,dv' \text{ dans } \Omega\times\{|v|=1\}, \\ \phi(x,v) = 0 \quad \text{ sur } \Gamma^-, \end{array}\right. \tag{6.20}$$

avec le bord entrant $\Gamma^- = \{x \in \partial\Omega \,|\, v\cdot n_x < 0\}$. Suivant la modélisation (1.10) nous avons séparé dans (6.20) les collisions pures, représentées par la section efficace σ^c, des créations de particules par fission, représentées par la section efficace σ^f. C'est seulement à cette dernière que s'applique le facteur de multiplication effectif.

Comme d'habitude nous supposons que l'absorption totale est positive ou nulle, c'est-à-dire que

$$\sigma(x) - \int_{|v'|=1} \sigma^c(x, v\cdot v')\,dv' \geq 0 \quad \forall\, x, v.$$

Il est essentiel de supposer que la section efficace de fission est minorée par une constante strictement positive

$$\sigma^c(x, v\cdot v') \geq 0, \quad \sigma^f(x, v\cdot v') \geq C > 0 \quad \forall\, x, v\cdot v', \tag{6.21}$$

car l'opération de moyennisation angulaire est à la base d'une propriété de compacité essentielle dans l'application du Théorème 6.2.14 de Krein-Rutman (voir le Chapitre XXI de [20] pour plus de détails).

Proposition 6.5.7 *Si Ω est un ouvert borné régulier, il existe une plus petite valeur propre réelle et simple $1/k_{\text{eff}}$, appelée facteur multiplicatif effectif, solution*

du problème critique (6.20) telle que son vecteur propre associé ϕ est le seul à être strictement positif dans $\Omega \times \{|v| = 1\}$ et pour toute autre valeur propre $1/k \in \mathbf{C}$ on a $1/k_{\text{eff}} < 1/|k|$.

Pour se convaincre du résultat de la Proposition 6.5.7 on peut démontrer son analogue en dimension finie pour la matrice issue d'une discrétisation spatiale de (6.20).

Lemme 6.5.8 *Soit A la matrice de discrétisation de (6.20) par différences finies en dimension $N = 1$ d'espace. Sous l'hypothèse (6.21) il existe un réel $\alpha \geq 0$ tel que $(A + \alpha\,\mathrm{Id})$ est une M-matrice irréductible (à laquelle on peut donc appliquer le Théorème 6.2.11).*

Démonstration. Pour simplifier on suppose tous les coefficients constants dans (6.20) et on choisit le schéma décentré amont (un argument similaire fonctionne pour le schéma diamant). On note $(\mu_l)_{1\leq l\leq K}$ les vitesses discrètes et on range les inconnues discrètes ϕ_j^l en les regroupant par vitesse commune. Pour simplifier on suppose aussi des poids uniformes dans la règle de quadrature pour calculer les moyennes angulaires. Cette matrice de discrétisation (pour $1/k = 1$) a une forme par blocs

$$A = \begin{pmatrix} A_{11} & -s\,\mathrm{Id} & \dots & \dots & -s\,\mathrm{Id} \\ -s\,\mathrm{Id} & A_{22} & -s\,\mathrm{Id} & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & -s\,\mathrm{Id} & A_{K-1\,K-1} & -s\,\mathrm{Id} \\ -s\,\mathrm{Id} & \dots & \dots & -s\,\mathrm{Id} & A_{KK} \end{pmatrix} \text{ avec } s = (\sigma^c + \sigma^f)/K,$$

avec les blocs diagonaux, indicés par le numéro l de la vitesse, du type, lorsque $\mu_l > 0$,

$$A_{ll} = \begin{pmatrix} \sigma + c_l & 0 & & & 0 \\ -c_l & \sigma + c_l & 0 & & \\ & \ddots & \ddots & \ddots & \\ & & -c_l & \sigma + c_l & 0 \\ 0 & & & -c_l & \sigma + c_l \end{pmatrix} \text{ avec } c_l = \frac{\mu_l}{\Delta x},$$

et, lorsque $\mu_l < 0$, une expression transposée. Il s'agit bien d'une M-matrice car tous les coefficients extra-diagonaux sont négatifs ou nuls tandis que les coefficients diagonaux sont positifs et la dominance diagonale est bien satisfaite, quitte à rajouter $\alpha\,\mathrm{Id}$. Comme $(\sigma^c + \sigma^f) > 0$ on peut construire une chaîne d'arêtes, $a_{ij} \neq 0$, reliant n'importe quels nœuds i et j du graphe de connectivité de A. Elle est donc bien irréductible. ■

6.6 Calcul critique

6.6.1 Problèmes à sources légèrement sous-critiques

Le but de cette section est d'expliquer en quoi la criticité est utile aussi pour résoudre des problèmes stationnaires à sources données. Nous avons déjà vu à la Section 3.4 une condition suffisante de solvabilité de l'équation de Boltzmann linéaire stationnaire qui consiste à dire que la disparition de particules par absorption domine l'apparition de particules par collision. Nous allons préciser cette condition d'existence de solution en fonction de la valeur du facteur multiplicatif effectif k_{eff}. Comme les résultats s'appliquent aussi bien au transport qu'à la diffusion, nous allons faire une présentation unifiée en termes d'opérateurs. Pour ne pas alourdir l'exposé et rentrer dans des considérations techniques (notamment sur la définition des espaces fonctionnels dans lesquels sont définis ces opérateurs) nous allons supposer qu'il s'agit de simples matrices, prolongeant ainsi l'analogie développée dans la Section 6.2.

On réécrit les problèmes critiques (6.19), en diffusion, ou (6.20), en transport, sous la forme

$$A\psi = \frac{1}{k} F\psi \,, \tag{6.22}$$

où F est l'opérateur de fission et A est l'opérateur de diffusion ou bien de Boltzmann linéaire, qui tient compte des conditions aux limites. Nous faisons l'hypothèse que A est inversible, A^{-1} et F positifs et que $K = A^{-1}F$ est strictement positif. En particulier, le Théorème 6.2.9 de Perron-Frobenius affirme que K admet une plus grande valeur propre réelle et simple k_{eff}.

Soit une source ou second membre b. On cherche à résoudre le problème stationnaire à source

$$Au = Fu + b \,. \tag{6.23}$$

Proposition 6.6.1 *Soit $1/k_{\text{eff}}$ la plus petite valeur propre de (6.22). Il existe une unique solution de (6.23) si et seulement si $k = 1$ n'est pas valeur propre de (6.22) (ce qui est le cas si le milieu est sous-critique).*

Si le milieu est sous-critique, $k_{\text{eff}} < 1$, alors (6.23) vérifie le principe du maximum, c'est-à-dire que $b \geq 0$ implique que $u \geq 0$.

Si le milieu est sur-critique, $k_{\text{eff}} > 1$, alors la solution de (6.23) n'a pas de sens physique, au sens où, si $b \geq 0$, la solution ne vérifie pas $u \geq 0$.

Remarque 6.6.2 *La violation du principe du maximum dans le cas sur-critique est évidemment non physique, mais pas surprenante. En effet, si on avait considéré le problème d'évolution pour (6.23) la solution croîtrait exponentiellement en temps et ne pourrait converger vers une limite stationnaire.*

Démonstration. On note $K = A^{-1}F$. L'équation (6.23) est équivalente à

$$(\,\mathrm{Id} - K)u = A^{-1}b$$

et l'alternative de Fredholm (voir le Théorème 4.2.2 pour ce résultat dans un cadre différent) dit qu'il en existe une solution unique si et seulement si 1 n'est pas valeur propre de K.

Lorsque $k_{\text{eff}} < 1$, on sait grâce au Théorème 6.2.9 de Perron-Frobenius que le rayon spectral de K est strictement plus petit que 1. Par conséquent, $(\,\mathrm{Id} - K)^{-1}$ est égal à la série convergente

$$(\,\mathrm{Id} - K)^{-1} = \sum_{p \geq 0} K^p.$$

Comme K est positif, on en déduit que $b \geq 0$ implique $u \geq 0$.

Supposons maintenant que $k_{\text{eff}} > 1$. Ecrivons alors le problème critique adjoint

$$A^* \psi^* = \frac{1}{k_{\text{eff}}} F^* \psi^* \,,$$

où $\psi^* > 0$ est le premier vecteur propre positif. On multiplie l'équation (6.23) par ψ^* pour obtenir

$$\langle u, A^* \psi^* \rangle = \langle u, F^* \psi^* \rangle + \langle b, \psi^* \rangle$$

qui devient

$$\left(\frac{1}{k_{\text{eff}}} - 1 \right) \langle u, F^* \psi^* \rangle = \langle b, \psi^* \rangle \,.$$

Si $b \geq 0$ et $b \neq 0$, comme $\psi^* > 0$, le membre de droite est strictement positif et, puisque $1/k_{\text{eff}} < 1$, on doit avoir $\langle u, F^* \psi^* \rangle = \langle Fu, \psi^* \rangle < 0$ ce qui oblige Fu, donc u, à avoir des composantes négatives. ■

Remarque 6.6.3 *Lorsque $k_{\text{eff}} = 1$ on peut résoudre (6.23) grâce à l'alternative de Fredholm. Il existe alors une solution si le second membre vérifie $\langle b, \psi^* \rangle = 0$ et cette solution est unique à l'addition d'un multiple de ψ près. Mais si, pour des raisons physiques, la source est positive $b \geq 0$ et non nulle $b \neq 0$, alors, comme $\psi^* > 0$, on ne peut satisfaire cette condition de compatibilité et (6.23) n'a pas de solution.*

La criticité ne sert pas qu'à décider si on peut résoudre ou non un problème stationnaire. Elle peut aussi donner une formule asymptotique de la solution lorsque le problème est **légèrement sous-critique**. Dans la proposition suivante on va supposer que le milieu est critique mais qu'on peut agir sur les sections efficaces (par exemple dans un réacteur nucléaire en insérant les barres de contrôle ou en injectant des poisons consommables) de manière à ce que l'opérateur de fission soit légèrement diminué et que le problème devienne sous-critique.

Proposition 6.6.4 *On suppose que le problème (6.22) est critique, c'est-à-dire que la première valeur propre est $k_{\text{eff}} = 1$. Soit $\epsilon > 0$ petit devant 1. On considère le problème à sources, petites de l'ordre de ϵ,*

$$Au_\epsilon = (1 - \epsilon) F u_\epsilon + \epsilon b \,. \tag{6.24}$$

Alors la solution vérifie

$$u_\epsilon = \langle A^{-1}b, \psi^* \rangle \, \psi + \mathcal{O}(\epsilon) \,, \tag{6.25}$$

où ψ et ψ^* sont les premiers vecteurs propres direct et adjoint de (6.22) définis, lorsque $k_{\text{eff}} = 1$, par

$$A\psi = F\psi \,, \quad A^*\psi^* = F^*\psi^* \,,$$

et normalisés par $\langle \psi, \psi^* \rangle = 1$.

Remarque 6.6.5 *Dans la formule (6.25) on a remplacé l'inversion de $(A - (1-\epsilon)F)$, qui est un opérateur presque singulier lorsque ϵ tend vers 0, par celle de A qui, en général, est un opérateur facile à inverser indépendamment de la valeur de ϵ.*

Démonstration. L'équation (6.24) est équivalente à

$$(\,\mathrm{Id} - (1-\epsilon)K)u_\epsilon = \epsilon z \quad \text{ avec } K = A^{-1}F \text{ et } z = A^{-1}b \,.$$

Par le Théorème 6.2.9 de Perron-Frobenius on sait que le sous-espace propre associé à $k_{\text{eff}} = 1$ est de dimension 1 engendré par ψ, noté $\mathrm{Vect}(\psi)$. On remarque que tout vecteur $v \in \mathbf{R}^n$ peut s'écrire

$$v = \langle v, \psi^* \rangle \psi + \tilde{v} \quad \text{ avec } \tilde{v} = v - \langle v, \psi^* \rangle \psi \,,$$

qui vérifie $\langle \tilde{v}, \psi^* \rangle = 0$. Autrement dit, $\mathrm{Vect}(\psi^*)^\perp$ est un sous-espace supplémentaire de $\mathrm{Vect}(\psi)$ dans $\mathbf{R}^n$, qui est stable par K. Comme toutes les valeurs propres k de la matrice K, autres que k_{eff}, vérifient $|k| < k_{\text{eff}} = 1$, on en déduit que $(\,\mathrm{Id} - (1-\epsilon)K)$ est inversible sur $\mathrm{Vect}(\psi^*)^\perp$ et que la norme de l'inverse est bornée indépendamment de ϵ. On écrit alors

$$z = \langle z, \psi^* \rangle \psi + \tilde{z} \;\text{ et }\; u_\epsilon = \langle u_\epsilon, \psi^* \rangle \psi + \tilde{u}_\epsilon \quad \text{ avec } \tilde{z} \,, \tilde{u}_\epsilon \in \mathrm{Vect}(\psi^*)^\perp,$$

et un simple calcul montre que

$$\langle u_\epsilon, \psi^* \rangle = \langle z, \psi^* \rangle \;\text{ et }\; |\tilde{u}_\epsilon| \leq \left\| (\,\mathrm{Id} - (1-\epsilon)K)^{-1}\big|_{\mathrm{Vect}(\psi^*)^\perp} \right\| \, |\epsilon \tilde{z}| \leq C\epsilon |\tilde{z}|$$

ce qui implique (6.25). ■

6.6.2 Analyse de sensibilité

Dans cette section nous allons étudier la variation de la criticité d'un milieu en fonction des variations des coefficients ou sections efficaces de ce milieu. Ces dernières varient, par exemple dans un réacteur nucléaire, à cause de l'insertion plus ou moins grande des barres de contrôle ou bien de la déplétion (l'usure) du combustible nucléaire avec le temps. Ici encore nous adoptons un formalisme d'opérateurs que, pour simplifier, nous supposons être de simples matrices.

Rappelons que le problème critique et son adjoint sont

$$A\psi = \frac{1}{k_{\text{eff}}} F\psi\,, \quad A^*\psi^* = \frac{1}{k_{\text{eff}}} F^*\psi^*\,, \tag{6.26}$$

où l'on suppose qu'on a normalisé ψ par $\|\psi\| = 1$, tandis que ψ^* est normalisé par $\langle \psi, \psi^* \rangle = 1$. On rappelle le résultat classique suivant (voir par exemple le chapitre 9 de [34]).

Lemme 6.6.6 *Si une valeur propre d'une matrice est simple alors elle et son vecteur propre, convenablement normalisé, sont (localement) continuement dérivables comme fonctions de cette matrice.*

Remarque 6.6.7 *L'hypothèse de simplicité de la valeur propre est essentielle dans le Lemme 6.6.6. D'une part, elle permet d'éviter les problèmes de croisement de valeurs propres et donc d'ambiguité dans la classification des valeurs propres. Par exemple, si on considère la matrice* 2×2

$$A = \begin{pmatrix} +a & 0 \\ 0 & -a \end{pmatrix}$$

qui admet $\pm a$ *comme valeur propre, on peut, suivant l'usage appeler* λ_1 *la plus petite valeur propre et* λ_2 *la plus grande, mais alors ces fonctions ne sont pas dérivables en* $a = 0$ *car* $\lambda_1(a) = -|a|$ *et* $\lambda_2(a) = |a|$*. D'autre part, et de manière plus fondamentale, on ne peut pas toujours trouver des vecteurs propres dérivables, ni même continus, pour une valeur propre double, comme le montre l'exemple* $A = PDP^{-1}$ *avec*

$$D = \begin{pmatrix} e^{-1/a^2} & 0 \\ 0 & -e^{-1/a^2} \end{pmatrix} \quad et \quad P = \begin{pmatrix} \cos(1/a) & -\sin(1/a) \\ \sin(1/a) & \cos(1/a) \end{pmatrix}$$

qui est une matrice de classe C^∞ *par rapport à* a*, qui admet une valeur propre double en* $a = 0$ *et aucun choix de vecteurs propres continus en* $a = 0$*.*

Proposition 6.6.8 *La variation du facteur multiplicatif effectif sous l'effet de variations* δA *et* δF *des opérateurs de transport/diffusion et fission est*

$$\delta k = k_{\text{eff}}^2 \frac{\langle \left(-\delta A + \frac{1}{k_{\text{eff}}} \delta F \right) \psi, \psi^* \rangle}{\langle F\psi, \psi^* \rangle}\,. \tag{6.27}$$

Remarque 6.6.9 *Encore une fois la formule (6.27) justifie le nom de fonction d'importance donnée à la fonction propre adjointe* ψ^**. Puisque* ψ *et* ψ^**, comme* F*, sont positifs, on en déduit, fort logiquement, qu'une augmentation des fissions,* $\delta F \geq 0$*, contribue à une augmentation de la criticité, tandis qu'une augmentation de l'absorption,* $\delta A \geq 0$*, engendre une diminution de la criticité.*

Démonstration. On dérive la première égalité de (6.26) pour obtenir

$$\left(A - \frac{1}{k_{\text{eff}}} F \right) \delta\psi = \left(-\delta A + \frac{1}{k_{\text{eff}}} \delta F - \frac{\delta k}{k_{\text{eff}}^2} F \right) \psi. \tag{6.28}$$

On ne peut résoudre (6.28) que si le second membre est orthogonal à ψ^*. On multiplie alors par ψ^*

$$\begin{aligned}\langle\left(A-\frac{1}{k_{\text{eff}}}F\right)\delta\psi,\psi^*\rangle &= \langle\delta\psi,\left(A^*-\frac{1}{k_{\text{eff}}}F^*\right)\psi^*\rangle = 0\\ &= \langle\left(-\delta A+\frac{1}{k_{\text{eff}}}\delta F-\frac{\delta k}{k_{\text{eff}}^2}F\right)\psi,\psi^*\rangle\end{aligned}$$

d'où l'on déduit la formule (6.27). Remarquons au passage que la solution $\delta\psi$ de (6.28) est définie, a priori, à l'addition d'un multiple de ψ près. Or, comme on a la normalisation $\|\psi\|=1$, par différentiation on en déduit $\langle\psi,\delta\psi\rangle=0$, ce qui fixe l'indétermination pour $\delta\psi$. ■

Le facteur multiplicatif effectif n'est pas la seule quantité dont on souhaite calculer la sensibilité aux variations des sections efficaces. Par exemple, des quantités importantes en physique des réacteurs nucléaires sont les taux de réactions, $\langle\sigma,\psi\rangle$, où σ est une section efficace de fission ou d'absorption pour un groupe d'énergie précis. Plus généralement, nous allons donner la variation d'une fonction régulière $j(\psi)$ à valeurs réelles.

Proposition 6.6.10 *On définit le flux adjoint ou fonction d'importance associée à j par*

$$\left(A^*-\frac{1}{k_{\text{eff}}}F^*\right)p = j'(\psi)-\langle j'(\psi),\psi\rangle\psi\,, \tag{6.29}$$

où j' est la dérivée de j. La variation de $j(\psi)$ sous l'effet de variations δA et δF des opérateurs de transport/diffusion et fission est

$$\delta j = \left\langle\left(-\delta A+\frac{1}{k_{\text{eff}}}\delta F\right)\psi,\left(p-\frac{\langle F\psi,p\rangle}{\langle F\psi,\psi^*\rangle}\psi^*\right)\right\rangle\,. \tag{6.30}$$

Démonstration. Commençons par montrer que le le problème adjoint (6.29) est bien posé. Par l'alternative de Fredholm (voir le Théorème 4.2.2 pour ce résultat dans un cadre différent) il existe une solution p, unique à l'addition d'un multiple de ψ^* près, si le second membre est orthogonal à ψ ce qui est le cas par construction (rappelons la normalisation $\|\psi\|^2=\langle\psi,\psi\rangle=1$).

En dérivant $j(\psi)$ on obtient

$$\delta j = \langle j'(\psi),\delta\psi\rangle\,,$$

où $\delta\psi$ est solution de (6.28). La résolution de l'équation (6.28) est très coûteuse car pour chaque composante des variations δA et δF il faut la résoudre. Remarquons d'ailleurs que nous n'avions pas besoin de $\delta\psi$ dans le résultat final de la Proposition 6.6.8. C'est donc pour éliminer $\delta\psi$ que nous introduisons l'adjoint p solution de (6.29). En effet, en multipliant (6.28) par p on obtient

$$\langle\left(A-\frac{1}{k_{\text{eff}}}F\right)\delta\psi,p\rangle = \langle\left(-\delta A+\frac{1}{k_{\text{eff}}}\delta F-\frac{\delta k}{k_{\text{eff}}^2}F\right)\psi,p\rangle,$$

tandis qu'en multipliant (6.29) par $\delta\psi$ on a

$$\langle\left(A^* - \frac{1}{k_{\text{eff}}}F^*\right)p, \delta\psi\rangle = \langle j'(\psi), \delta\psi\rangle - \langle j'(\psi), \psi\rangle\langle\psi, \delta\psi\rangle = \langle j'(\psi), \delta\psi\rangle,$$

car $\langle\psi, \delta\psi\rangle = 0$ puisque $\|\psi\| = 1$. On remarque que les membres de gauche des deux dernières égalités sont identiques, d'où l'on déduit que

$$\langle j'(\psi), \delta\psi\rangle = \langle\left(-\delta A + \frac{1}{k_{\text{eff}}}\delta F - \frac{\delta k}{k_{\text{eff}}^2}F\right)\psi, p\rangle.$$

A cause de la formule (6.27) pour δk on obtient exactement la formule (6.30). Remarquons que cette formule est invariante par addition à p d'un multiple de ψ^*, ce qui est consistant avec la classe d'unicité de p. ■

6.6.3 Calcul numérique de la criticité

Pour résoudre un problème de criticité, c'est-à-dire trouver la plus petite valeur propre du problème spectral (6.19) (en diffusion) ou (6.20) (en transport), la méthode numérique la plus populaire, et la plus simple, est la méthode de la puissance. Cette méthode permet de calculer la plus grande ou la plus petite valeur propre d'une matrice, ainsi qu'un vecteur propre associé. Pour une matrice quelconque, ou au contraire pour une matrice symétrique, la méthode de la puissance n'est pas la plus efficace qui soit. Mais elle est est bien adaptée aux M-matrices irréductibles (pour lesquelles le Théorème 6.2.9 de Perron-Frobenius s'applique) et justement les matrices de discrétisation de (6.19) et (6.20) en sont.

C'est cette méthode de la puissance pour calculer la plus grande valeur propre d'une matrice K, réelle de taille $n \times n$, que nous décrivons maintenant. Nous faisons l'hypothèse que K est strictement positive et donc qu'on peut lui appliquer le Théorème de Perron-Frobenius. En notant $(\lambda_1, \cdots, \lambda_n)$ les valeurs propres de K, il existe une valeur propre dominante simple

$$\lambda_n > |\lambda_i| \quad \text{pour tout } 1 \leq i \leq n-1,$$

et son vecteur propre associé e_n peut être choisi strictement positif.

La méthode de la puissance pour calculer la plus grande valeur propre λ_n est définie par l'algorithme ci-dessous.

1. Initialisation : $x_0 \in \mathbf{R}^n$ tel que $x_0 > 0$.
2. Itérations : pour $k \geq 1$
   1. $y_k = Kx_{k-1}$
   2. $x_k = y_k / \max(y_k)$ où $\max(y)$ désigne la plus grande composante en module du vecteur y,
   3. test de convergence : si $\|x_k - x_{k-1}\| \leq \varepsilon$, on arrête.

Dans le test de convergence ε est un petit nombre réel, typiquement égal à 10^{-6}. Si $\delta_k = x_k - x_{k-1}$ est petit, alors x_k est un vecteur propre approché de K de valeur propre approchée $\max(y_k)$ car $Kx_k - \max(y_k)x_k = K\delta_k$.

Proposition 6.6.11 *On suppose que la matrice K est strictement positive. Alors la méthode de la puissance converge, c'est-à-dire que*

$$\lim_{k\to+\infty} \max(y_k) = \lambda_n, \quad \lim_{k\to+\infty} x_k = e_n/\max(e_n).$$

De plus, la vitesse de convergence est proportionnelle au rapport $|\lambda_{n-1}|/\lambda_n$.

Démonstration. Supposons pour simplifier que K soit diagonalisable avec des vecteurs propres $(e_1,\cdots,e_n)$ correspondant aux valeurs propres $(\lambda_1,\cdots,\lambda_n)$. Soit $x_0 = \sum_{i=1}^n \beta_i e_i$ le vecteur initial. Comme $x_0 > 0$ et que le vecteur propre adjoint e_n^* de K^* est aussi strictement positif en vertu du Théorème de Perron-Frobenius, on a $\beta_n = x_0 \cdot e_n^* > 0$. Une récurrence facile montre que

$$x_k = \frac{K^k x_0}{\max(K^k x_0)} = \frac{\sum_{i=1}^n \beta_i (\lambda_i)^k e_i}{\max\left(\sum_{i=1}^n \beta_i (\lambda_i)^k e_i\right)} = \frac{e_n + \sum_{i=1}^{n-1} \frac{\beta_i}{\beta_n}\left(\frac{\lambda_i}{\lambda_n}\right)^k e_i}{\max\left(e_n + \sum_{i=1}^{n-1} \frac{\beta_i}{\beta_n}\left(\frac{\lambda_i}{\lambda_n}\right)^k e_i\right)}.$$

Comme $|\lambda_i| < \lambda_n$ on en déduit que x_k converge vers $e_n/\max(e_n)$. De même, puisque $y_k = Kx_k$ converge vers $\lambda_n e_n/\max(e_n)$, on en déduit que $\max(y_k)$ converge vers λ_n.

Si K n'est pas diagonalisable, alors il faut utiliser la base $(e_1,\cdots,e_n)$ de sa forme de Jordan. Pour fixer les idées et simplifier les notations, supposons que tous les vecteurs e_i sont en fait des vecteurs propres sauf e_{n-2} qui appartient au sous-espace spectral de la valeur propre $\lambda_{n-1} = \lambda_{n-2}$ sans être vecteur propre. Autrement dit, on a

$$Ke_{n-2} = \lambda_{n-1}e_{n-2} + e_{n-1} \quad \text{et } Ke_i = \lambda_i e_i \text{ pour } i \neq (n-2)\,.$$

Dans ce cas, la formule ci-dessus pour x_k doit être modifiée comme suit

$$x_k = \frac{\beta_n e_n + (\beta_{n-1} + k\beta_{n-2}/\lambda_{n-1})\left(\frac{\lambda_{n-1}}{\lambda_n}\right)^k e_{n-1} + \sum_{i=1}^{n-2} \beta_i \left(\frac{\lambda_i}{\lambda_n}\right)^k e_i}{\max\left(\beta_n e_n + (\beta_{n-1} + k\beta_{n-2}/\lambda_{n-1})\left(\frac{\lambda_{n-1}}{\lambda_n}\right)^k e_{n-1} + \sum_{i=1}^{n-2} \beta_i \left(\frac{\lambda_i}{\lambda_n}\right)^k e_i\right)}.$$

On a toujours les mêmes convergences de x_k et $\max(y_k)$ puisque $k(\lambda_{n-1}/\lambda_n)^k$ tend toujours vers zéro lorsque k tend vers $+\infty$, même si la convergence est un peu plus lente. ■

En pratique, les matrices issues de la discrétisation d'un problème critique sont les inverses de matrices strictement positives, et on cherche plutôt leur **plus petite** valeur propre. Décrivons l'algorithme de la puissance "inverse" dans ce cas. On suppose que A est une M-matrice irréductible. On peut donc lui appliquer le Théorème 6.2.11 qui affirme que A admet une plus petite valeur propre réelle simple λ_1 telle que

$$\lambda_1 < |\lambda_i| \quad \text{pour tout } 2 \leq i \leq n,$$

et son vecteur propre associé peut être choisi strictement positif. L'algorithme s'écrit alors.

1. Initialisation : $x_0 \in \mathbf{R}^n$ tel que $x_0 > 0$.
2. Itérations : pour $k \geq 1$
   1. résoudre $Ay_k = x_{k-1}$
   2. $x_k = y_k / \max(y_k)$
   3. test de convergence : si $\|x_k - x_{k-1}\| \leq \varepsilon$, on arrête.

Si $\delta_k = x_k - x_{k-1}$ est petit, alors x_{k-1} est un vecteur propre approché de valeur propre approchée $1/\max(y_k)$ car $Ax_{k-1} - \frac{x_{k-1}}{\max(y_k)} = -A\delta_k$.

Proposition 6.6.12 *On suppose que A est une M-matrice irréductible. Alors la méthode de la puissance inverse converge, c'est-à-dire que*

$$\lim_{k\to+\infty} \frac{1}{\max(y_k)} = |\lambda_1|, \quad \lim_{k\to+\infty} x_k = e_1/\max(e_1).$$

La vitesse de convergence est proportionnelle au rapport $\lambda_1/|\lambda_2|$.

La démonstration est similaire à celle de la Proposition 6.6.11 et nous la laissons au lecteur en guise d'exercice.

6.7 Exercices

Exercice 6.9 (Une autre preuve du théorème de Perron) *(difficile)*
Le théorème de Perron (1907) s'énonce : "Une matrice à coefficients strictement positifs admet un vecteur propre à coefficients strictement positifs, et la valeur propre associée est le rayon spectral." Ce qui suit est une partie de la preuve de Wielandt.

Soit A une matrice $n \times n$ à coefficients dans $\mathbf{R}_+^*$. Soit

$$f : (\mathbf{R}_+^*)^n \to \mathbf{R}_+^*,\ f(x) = \max_{1\leq i\leq n} \frac{(Ax)_i}{x_i} \ ,$$

1. *Montrer que f admet un point de minimum u dans $(\mathbf{R}_+^*)^n$.*
2. *Montrer que u est vecteur propre de A.*
3. *On rappelle (cf. Proposition 13.1.7 dans [2]) que pour toute matrice M*

$$\rho(M) \leq ||M||_\infty = \max_{1\leq i\leq n} \sum_{1\leq j\leq n} |m_{ij}|.$$

   Montrer que le minimum de f est le rayon spectral $\rho(A)$ de A.
4. *En quoi le théorème 6.2.9 est-il plus puissant ?*

Indications : pour la question 2, on peut procéder par l'absurde, et supposer qu'il existe $i < j$ tels que $\frac{(Au)_i}{u_i} < \frac{(Au)_j}{u_j}$. On montrera alors qu'on peut diminuer $f(u)$ en modifiant légèrement u. Pour la question 3, on considérera la matrice diagonale D de coefficients diagonaux $u_1, \ldots, u_n$, et on calculera $\|D^{-1}AD\|_\infty$.

Exercice 6.10 (Application au calcul critique en neutronique) *Soit un modèle de réacteur à deux groupes de neutrons : les rapides (groupe 1, $v \approx 10^7$ cm s^{-1}) et les lents (groupe 2, $v \approx 10^5$ cm s^{-1}). On sait que les lents fissionnent préférentiellement. Le modèle "à l'équilibre" s'écrit*

$$\begin{cases} -\nabla \cdot (d_1 \nabla \phi_1) + \sigma_{a1}\phi_1 = \nu \left(\sigma_{f1}\phi_1 + \sigma_{f2}\phi_2\right), \\ -\nabla \cdot (d_2 \nabla \phi_2) + \sigma_{a2}\phi_2 = \sigma_r \phi_1. \end{cases}$$

On a les encadrements

$$\sigma_{a1} \geq \sigma_{f1} + \sigma_r, \quad \sigma_{a2} \geq \sigma_{f2}, \quad \sigma_{f2} > \sigma_{f1}.$$

Le nombre de neutrons créés lors d'une fission est $\nu > 1$.

1. *Il est courant de considérer une succession d'états "d'équilibre" sous la forme d'une suite*

$$\begin{cases} -\nabla \cdot \left(d_1 \nabla \phi_1^{(p+1)}\right) + \sigma_{a1}\phi_1^{(p+1)} = \nu \left(\sigma_{f1}\phi_1^{(p)} + \sigma_{f2}\phi_2^{(p)}\right), \\ -\nabla \cdot \left(d_2 \nabla \phi_2^{(p+1)}\right) + \sigma_{a2}\phi_2^{(p+1)} = \sigma_r \phi_1^{(p+1)}. \end{cases}$$

   On parle des neutrons de la génération p, $p+1$,... A partir d'un certain rang, on admet que

$$\phi_1^{(p+1)} \approx \lambda \phi_1^{(p)}, \quad \lambda > 0.$$

   Justifier le problème aux valeurs propres

$$\begin{cases} -\nabla \cdot (d_1 \nabla \phi_1) + \sigma_{a1}\phi_1 = \frac{1}{\lambda}\nu \left(\sigma_{f1}\phi_1 + \sigma_{f2}\phi_2\right), \\ -\nabla \cdot (d_2 \nabla \phi_2) + \sigma_{a2}\phi_2 = \sigma_r \phi_1. \end{cases}$$

   Pour $\lambda = 1$ le système est critique ($\lambda > 1$ sur-critique, $\lambda < 1$ sous-critique).

2. *On munit le domaine d'étude de conditions au bord de Dirichlet ou de Neumann. Soient G_1 et G_2 les inverses formels des opérateurs à gauche des signes "=", c'est-à-dire*

$$G_1^{-1}\phi = -\nabla \cdot (d_1 \nabla \phi) + \sigma_{a1}\phi, \quad G_2^{-1}\phi = -\nabla \cdot (d_2 \nabla \phi) + \sigma_{a2}\phi.$$

   Montrer que

$$G_1 \left(\nu\sigma_{f1} I + \nu\sigma_{f2} G_2 \sigma_r\right) \phi_1 = \lambda \phi_1.$$

   On pose $s_f = \nu \left(\sigma_{f1}\phi_1 + \sigma_{f2}\phi_2\right)$ le nombre de neutrons qui fissionent par unité. Montrer que

$$\left(\nu\sigma_{f1} I + \nu\sigma_{f2} G_2 \sigma_r\right) G_1 s_f = \lambda s_f.$$

3. *On utilise une discrétisation de type différences finies sur grille régulière pour les différents opérateurs en présence.*
   Montrer que les hypothèses du théorème de Perron sont vérifiées. En déduire l'existence d'une solution discrète au problème de valeur propre-vecteur propre précédent.

Indication : pour la question 1 on montrera que la même relation existe pour les itérées de ϕ_2.

Exercice 6.11 (Application à l'analyse spectrale de graphes) *On va se baser dans cet exercice sur un exemple est tiré de l'article* Spectral analysis of Internet topologies *[22].*

1. *La méthode du* page-ranking *peut se concevoir dans la version simplifiée suivante. Chaque page est repérée par un indice j entre 1 et n. La matrice d'adjacence $A = (a_{ij})$ des n pages considérées est telle que*

$$a_{ij} = 1 \text{ ssi la page } i \text{ pointe vers la page } j,$$

$a_{ij} = 0$ dans le cas contraire. Par convention, $a_{ii} = 0$. On pose

$$d_{\text{out}}(i) = \sum_j a_{ij}.$$

Soit $0 \leq \alpha < 1$ une "probabilité" de transition. La matrice de toutes les transitions possibles est P définie par

$$p_{ij} = \frac{\alpha}{d_{\text{out}}(i)} + \frac{1-\alpha}{n} \text{ si } a_{ij} \neq 0,$$

$$p_{ij} = \frac{1-\alpha}{n} \text{ si } a_{ij} = 0.$$

Cette matrice représente une marche au hasard dans la graphe de A, avec une probabilité α d'aller vers une autre page choisie au hasard dans la liste des pages "adjacentes". A priori α est proche de 1.

Montrer que les hypothèses du théorème de Perron sont satisfaites par P.

2. *Montrer que la valeur propre maximale de P est $\lambda = 1$ (cette matrice est dite stochastique). Le vecteur propre à gauche donne le* page-rank.

Remarque : cet algorithme, qui accorde à une page une importance d'autant plus grande que son coefficient dans le page rank (vecteur propre à gauche pour la valeur propre 1) est grand, est la base de fonctionnement du moteur de recherche Google.

Exercice 6.12 (Application à l'analyse spectrale de graphes) *Voici encore nouveau un exemple est aussi tiré de l'article* Spectral analysis of Internet topologies *[22].*

1. *Soit un graphe construit à partir d'un maillage. Ce maillage est de type éléments finis par exemple. On veut le partionner de façon équilibrée (équilibrage de tâches pour les calculs parallèles). Le maillage peut aussi représenter la toile Internet, et on désire analyser le trafic sur la toile. On note*

$$a_{ij} = 1 \text{ si une arête existe entre les nœuds } i \text{ et } j,$$

et $a_{ij} = 0$ sinon. Par convention, on pose $a_{ii} = 0$. Noter que la matrice $A = (a_{ij})$ est symétrique $A = A^t$.
Soit D la matrice diagonale dont le $i^{ème}$ coefficient est égal à la somme des coefficients de A sur la ligne i :

$$d_i = \sum_j a_{ij} = \sum_j a_{ji}.$$

soit

$$M^\varepsilon = D - A + \varepsilon I = (M^\varepsilon)^t, \quad \varepsilon > 0.$$

Cette matrice est appelée Laplacien du graphe pour $\varepsilon = 0$, qui est le cas vraiment intéressant.

2. *On prend $\varepsilon > 0$. Montrer que la plus petite valeur propre de M^ε peut se déterminer par l'application du théorème de Perron à $(M^\varepsilon)^{-1}$.*
3. *On prend $\varepsilon = 0$. Montrer que la plus petite valeur propre de M^0 est nulle, $\lambda_1 = 0$. La suivante $\lambda_2 \geq 0$ est appelée* connectivité algébrique.
   Montrer que $\lambda_2 > 0$ dès qu'un chemin permet de passer de tout indice i à tout indice j.
   Le vecteur propre correspondant x^2 est appelé vecteur de Fiedler. Montrer que

$$\sum_j x_j^2 = 0.$$

   Montrer également qu'aucune composante de x^2 ne peut être nulle. La méthode standard consiste alors à séparer les indices en deux groupes

$$I^+ = \{j,\ x_j^2 > 0\},\ I^- = \{j,\ x_j^2 < 0\}.$$

4. *Soit la matrice du Laplacien en dimension 1 d'espace. Montrer que I^+ et I^- sont connexes.*
   L'application successive de cette méthode prend le nom de : méthode de bissection spectrale récursive.

Indication : pour la question 3, pour montrer qu'aucune composante de x^2 ne peut être nulle, on peut raisonner par l'absurde, et prouver que, si $x_{i_0}^2 = 0$ pour un indice i_0, alors le sommet i_0 du graphe est isolé.

Exercice 6.13 *Soit le modèle structuré en âge pour une population de cellules*

$$\begin{cases} \dfrac{\partial n}{\partial t}(t,x) + \dfrac{\partial n}{\partial x}(t,x) = 0, & t > 0,\ x > 0, \\ n(t, x = 0) = \int_0^\infty B(y) n(t,y) dy, \\ n(t = 0, x) = n_0(x). \end{cases}$$

Le coefficient de natalité B vérifie

$$0 \leq B(y) \leq C, \quad 1 < \int_0^\infty B(y) dy < \infty.$$

1. *On cherche des solutions particulières du type* $n(t,x) = e^{\lambda t}N(x)$*, ignorant la condition initiale. Ecrire le problème spectral (direct) dont est solution le couple* (λ, N)*. Montrer qu'il possède une unique solution* $\lambda > 0$*,* $N > 0$*.*
2. *Ecrire le problème spectral adjoint (de solution* ϕ*). Montrer qu'il possède une unique solution* $\phi > 0$*. Calculer* ϕ *pour* $B(x) = 1.2 \times \mathbf{1}_{\{x<1\}}$*, et pour* $B(x) = 3 \times \mathbf{1}_{\{0.5<x<1\}}$*.*
3. *Comment s'appelle* ϕ *dans un calcul de neutronique ?*
4. *Ecrire formellement ce que vaut* $n(t,\cdot)$ *pour* t *grand. La preuve complète se trouve dans* Transport Equations in Biology *[42], page 64.*

Exercice 6.14 *Nous reprenons l'exercice 6.13 avec des hypothèses différentes. Soit le modèle structuré en taille dans lequel les cellules de taille* x *se décomposent en deux cellules de taille* $\frac{x}{2}$

$$\begin{cases} \dfrac{\partial n}{\partial t}(t,x) + \dfrac{\partial n}{\partial x}(t,x) + B(x)n(t,x) = 4B(2x)n(t,2x), \quad t > 0,\ x > 0, \\ n(t, x = 0) = 0, \\ n(t = 0, x) = n_0(x). \end{cases}$$

Il n'y a pas de création de cellules de taille $x = 0$*.*

1. *Vérifier que le nombre total de cellules*

$$\int_0^\infty n(t,x)dx$$

et la taille totale

$$\int_0^\infty xn(t,x)dx$$

sont croissants.

2. *Ecrire les équations pour les problèmes critiques direct et adjoint, dont les solutions sont notées* (λ, N, ϕ)*.*
Ecrire formellement la solution asymptotique en temps du problème de départ. Expliquer pourquoi la valeur propre λ *est aussi appelée paramètre de Malthus.*
3. *On étudie le cas* $B(x) = b$ *constant dans lequel les calculs sont explicites. Montrer que les solutions sont* $\lambda = b$*,* $\phi(x) = 1$ *et*

$$N(x) = \overline{N} \sum_{n=0}^{\infty} (-1)^n \alpha_n e^{-2^{n+1}bx}$$

avec $\alpha_0 = 1$ *et* $\alpha_n = \frac{2}{2^n - 1}\alpha_{n-1}$*.*
4. *Montrer que* $N(x) > 0$ *pour* $2bx \geq 1$*.*
5. *A partir de l'égalité*

$$\frac{\partial |N(x)|}{\partial x} + 2b|N(x)| = 4bN(2x)\mathrm{sgn}(N(x)),$$

montrer que

$$\int_0^\infty |N(x)|dx = \int_0^\infty N(x)\mathrm{sgn}\left(N\left(\frac{x}{2}\right)\right)dx.$$

En déduire que $N(x) > 0$ pour tout $x > 0$.

Indications : pour la première question, il suffit d'intégrer l'équation en espace contre la fonction constante égale à 1 d'une part, et contre la fonction x d'autre part. On obtient alors, en supposant que $n(t,x)$ et $xn(t,x)$ tendent vers 0 en $+\infty$,

$$\frac{d}{dt}\left(\int_0^\infty n(t,x)dx\right) = 2\int_0^\infty B(x)n(t,x)dx \geq 0,$$

$$\frac{d}{dt}\left(\int_0^\infty xn(t,x)dx\right) = \int_0^\infty n(t,x)dx \geq 0.$$

Dans la dernière question, on utilise la convention que $\mathrm{sgn}(a) = 1$ si $a > 0$, -1 si $a < 0$ et 0 si $a = 0$. On montre alors facilement qu'on peut majorer le membre de droite de l'égalité comme suit :

$$\int_0^\infty N(x)\mathrm{sgn}\left(N\left(\frac{x}{2}\right)\right)dx \leq \int_0^\infty |N(x)|dx,$$

avec égalité uniquement si $\mathrm{sgn}(N(x)) = \mathrm{sgn}(N(x/2))$. Cette dernière égalité implique que N ne peut pas changer de signe.

Exercice 6.15 (Un problème avec source) *Soit le problème à deux groupes de neutrons dans un domaine borné Ω*

$$\begin{cases} -\nabla\cdot(d_1\nabla\phi_1) + \sigma_{a1}\phi_1 = \nu\sigma_{f2}\phi_2 + s(x), \\ -\nabla\cdot(d_2\nabla\phi_2) + \sigma_{a2}\phi_2 = \sigma_r\phi_1. \end{cases}$$

On considère des conditions de Dirichlet homogènes sur le bord du domaine $\partial\Omega$. Les coefficients $\sigma_{a1}, \sigma_{f2}, \sigma_{a2}, \sigma_r$ sont constants et strictement positifs. Le nombre de neutrons rapides ϕ_1 créés lors d'une fission est $\nu > 0$. La source est positive ou nulle $s(x) \geq 0$.

1. *On prend $\phi_1^{(0)} = \phi_2^{(0)} = 0$. Montrer que la suite définie par*

$$\begin{cases} -\nabla\cdot(d_1\nabla\phi_1^{(p+1)}) + \sigma_{a1}\phi_1^{(p+1)} = \nu\sigma_{f2}\phi_2^{(p)} + s(x), \\ -\nabla\cdot(d_2\nabla\phi_2^{(p+1)}) + \sigma_{a2}\phi_2^{(p+1)} = \sigma_r\phi_1^{(p)} \end{cases}$$

   est positive.

2. *Montrer que la convergence géométrique de la suite vers une limite est possible, à condition que ν soit suffisamment petit. Quels sont les liens avec le calcul critique ?*

Indications : pour la deuxième question on poura étudier les relations de récurrence sur les différences $\varphi_1^{(p)} = \phi_1^{(p+1)} - \phi_1^{(p)}$ et $\varphi_2^{(p)} = \phi_2^{(p+1)} - \phi_2^{(p)}$. Montrer que l'on obtient

$$\|\varphi_1^{(p+2)}\|_{L^2(\Omega)} \leq \nu \frac{\sigma_{f2}\sigma_r}{\sigma_{a1}\sigma_{a2}} \|\varphi_1^{(p)}\|_{L^2(\Omega)}$$

et conclure.

Exercice 6.16 (Analyse de sensibilité) *On suppose que le réacteur (décrit à l'exercice 6.15) est légèrement sous-critique. Au cours du fonctionnement normal, les paramètres vont changer légèrement (les sources s'épuisent, les matériaux s'usent, ...). Il faut alors recaler le réacteur.*

1. *On note $(\sigma_{a_1}, \nu) \mapsto \lambda$ la fonction égale au facteur effectif. Montrer que*

$$\frac{\partial \lambda}{\partial \nu} = \frac{\lambda}{\nu}.$$

   Expliquer pourquoi cette relation est triviale et sans intérêt.

2. *Montrer que*

$$\frac{\partial \lambda}{\partial \sigma_{a_1}} < 0.$$

   Interpréter physiquement.

3. *Quelle action faut-il tenter de faire si λ est plus petit au bout d'un an en fonctionnement normal ?*

Chapitre 7

Homogénéisation

L'homogénéisation est la théorie qui étudie les méthodes de **moyennisation** dans les équations aux dérivées partielles. En d'autres termes, l'homogénéisation permet de trouver un modèle homogénéisé, ou macroscopique, ou moyenné, qui est une bonne approximation d'un problème originalement posé dans un milieu très hétérogène. L'intérêt de ce modèle homogénéisé est qu'il est plus facile à résoudre numériquement car posé dans un milieu homogène équivalent. Pour une présentation plus complète nous renvoyons à [3], [8], [29], [48].

Nous verrons aussi que l'homogénéisation permet encore de **faire le lien entre modèles de transport et modèles de diffusion**. En particulier, l'homogénéisation est à la base d'une approche numérique classique en neutronique ou physique des réacteurs nucléaires. En effet, le calcul de la distribution des neutrons dans un réacteur nucléaire, qui est un milieu très hétérogène (pour les réacteurs à eau pressurisée, les plus courants, de l'ordre de quelques centaines d'assemblages ou quelques dizaines de milliers de crayons de combustible entourés d'eau), nécessite de résoudre une équation de transport avec un maillage très fin, ce qui est très coûteux, voire impossible. Aussi, pour calculer économiquement une solution approchée, il est d'usage de faire un calcul local de transport pour chaque assemblage couplé avec un calcul global de diffusion pour l'ensemble du cœur du réacteur. L'homogénéisation permet de justifier cette approche [21], [43].

Dans ce chapitre nous nous contenterons d'exposer la théorie de l'homogénéisation pour des structures périodiques. Celles-ci sont très nombreuses dans la nature ou dans les applications industrielles et on dispose d'une méthode très simple et très puissante pour les homogénéiser : la méthode des développements asymptotiques à deux échelles que nous présentons ci-dessous. Néanmoins l'homogénéisation n'est pas réduite au cas périodique : il existe aussi des théories plus générales dont nous ne dirons rien ici par souci de simplicité.

Dans une structure périodique, nous notons ϵ le rapport de la période sur la taille caractéristique de la structure. En général, ce paramètre positif ϵ est petit, et l'homogénéisation consiste à effectuer une **analyse asymptotique** lorsque ϵ tend vers zéro. La limite ainsi obtenue sera dite homogénéisée, macroscopique,

ou effective. Dans le problème homogénéisé la forte hétérogénéité de la structure périodique d'origine est moyennée et remplacée par l'utilisation de coefficients effectifs.

7.1 Homogénéisation d'une équation de diffusion

7.1.1 Modèle de diffusion en milieu périodique

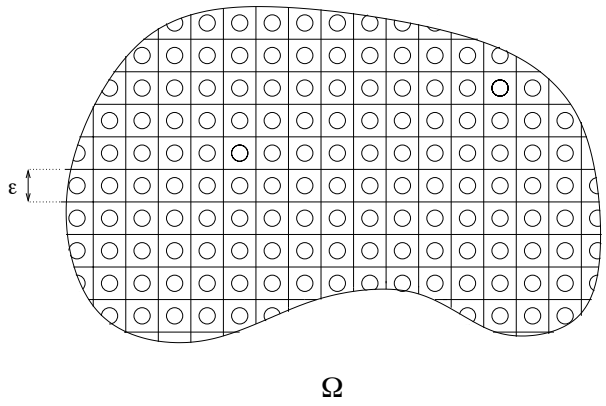

FIGURE 7.1 – Milieu hétérogène périodique

On considère un modèle de diffusion stationnaire dans un milieu périodique. On note Ω (un ouvert borné de $\mathbf{R}^N$ avec $N \geq 1$) le domaine périodique de période ϵ, avec $0 < \epsilon \ll 1$, et $Y = (0,1)^N$ la cellule unité de périodicité. Le tenseur de diffusion dans Ω n'est pas constant mais varie périodiquement avec la période ϵ dans chacune des directions de l'espace. Pour mettre en valeur cette périodicité de taille ϵ, on écrit ce tenseur sous la forme

$$D\left(\frac{x}{\epsilon}\right)$$

où $D(y)$ est un tenseur (une matrice), défini pour $y \in Y$, qui vérifie la propriété de Y-périodicité

$$D(y + e_i) = D(y) \ \forall i \in \{1, ..., N\}, \text{ avec } (e_i)_{1 \leq i \leq N} \text{ la base canonique de } \mathbf{R}^N, \tag{7.1}$$

c'est-à-dire que $D(y)$ est périodique de période 1 dans tous les directions principales $(e_i)_{1 \leq i \leq N}$ de l'espace. Ceci assure que $x \to D\left(\frac{x}{\epsilon}\right)$ est périodique de période ϵ pour tout $\epsilon > 0$. En toute généralité, $D(y)$ est une matrice symétrique non nécessairement isotrope. On suppose néanmoins que D est coercive et bornée, c'est-à-dire qu'il existe deux constantes $\beta \geq \alpha > 0$ telles que, pour

n'importe quel vecteur $\xi \in \mathbf{R}^N$ et en tout point $y \in Y$,

$$\alpha|\xi|^2 \leq \sum_{i,j=1}^{N} D_{ij}(y)\xi_i\xi_j \leq \beta|\xi|^2.$$

En toute généralité le tenseur $D(y)$ peut être discontinu en y pour modéliser les discontinuités des propriétés matérielles quand on passe d'une phase (ou matériau) à une autre.

Si l'on note $f(x)$ le terme source et si l'on impose des conditions aux limites de Dirichlet (par souci de simplicité), le problème modèle est

$$\begin{cases} -\operatorname{div}\left(D\left(\frac{x}{\epsilon}\right)\nabla u_\epsilon\right) = f & \text{dans } \Omega \\ u_\epsilon = 0 & \text{sur } \partial\Omega, \end{cases} \tag{7.2}$$

dont on sait qu'il admet une unique solution $u_\epsilon \in H_0^1(\Omega)$ si $f \in L^2(\Omega)$. En pratique, le domaine Ω avec son tenseur de diffusion $D\left(\frac{x}{\epsilon}\right)$ est très hétérogène à une petite échelle de l'ordre de ϵ. La connaissance des détails de la solution à cette si petite échelle n'est pas nécessaire pour une analyse globale du domaine : en général on se contente de déterminer son comportement moyen sous l'effet de la source f. D'un point de vue numérique, résoudre l'équation (7.2) par n'importe quelle méthode raisonnable nécessite un temps et une mémoire machine considérables si ϵ est petit, puisque le pas du maillage doit être au moins plus petit que ϵ, ce qui conduit à un nombre de degrés de liberté (ou de mailles) pour un niveau donné de précision au moins de l'ordre de $1/\epsilon^N$. Il est donc préférable de moyenner ou homogénéiser les propriétés matérielles de Ω et de calculer une approximation de u_ϵ sur un maillage grossier. Ce procédé de moyennisation de la solution de (7.2), et de détermination des propriétés effectives de Ω est précisément ce que l'on appelle **l'homogénéisation**.

Afin de trouver le comportement homogénéisé de Ω, on utilise une méthode de **développements asymptotiques à deux échelles**. Comme dans la Section 4.2 sur l'approximation du transport par la diffusion, l'idée de base est d'utiliser une série formelle, c'est-à-dire de postuler que la solution u_ϵ de (7.2) s'écrit, lorsque la période ϵ tend vers 0, comme une série en puissances de ϵ

$$u_\epsilon = \sum_{i=0}^{+\infty} \epsilon^i u_i.$$

Le premier terme u_0 de cette série sera identifié à la solution de l'équation, dite homogénéisée, dont le tenseur de diffusion D^* décrira les propriétés macroscopiques d'un milieu homogène équivalent. L'intérêt de cette approche est que les simulations numériques sur le modèle homogénéisé ne nécessitent qu'un maillage grossier puisque les hétérogénéités de taille ϵ ont été moyennées. Par ailleurs, cette méthode donnera une formule explicite pour calculer ce tenseur homogénéisé D^* qui, en général, n'est pas une moyenne usuelle de $D(y)$.

7.1.2 Développements asymptotiques à deux échelles

La méthode des développements asymptotiques à deux échelles est une méthode formelle d'homogénéisation qui permet de traiter un très grand nombre de problèmes posés dans des milieux périodiques. Comme nous l'avons déjà dit, l'hypothèse de départ est de supposer que la solution u_ϵ de l'équation (7.2) est donnée par un développement en série de ϵ, dit "à deux échelles", du type

$$u_\epsilon(x) = \sum_{i=0}^{+\infty} \epsilon^i u_i\left(x, \frac{x}{\epsilon}\right), \tag{7.3}$$

où chaque terme $u_i(x,y)$ est une fonction de deux variables $x \in \Omega$ et $y \in Y = (0,1)^N$, qui est périodique en y de période Y. La variable x est dite **lente ou macroscopique**, tandis que y est dite **rapide ou microscopique**. Cette série est injectée dans l'équation, et la règle de dérivation composée suivante est utilisée

$$\nabla\left(u_i\left(x, \frac{x}{\epsilon}\right)\right) = \left(\epsilon^{-1}\nabla_y u_i + \nabla_x u_i\right)\left(x, \frac{x}{\epsilon}\right),$$

où ∇_y et ∇_x désignent les dérivées partielles par rapport à la variable rapide y et à la variable lente x. Par exemple, on a

$$\nabla u_\epsilon(x) = \epsilon^{-1}\nabla_y u_0\left(x, \frac{x}{\epsilon}\right) + \sum_{i=0}^{+\infty} \epsilon^i \left(\nabla_y u_{i+1} + \nabla_x u_i\right)\left(x, \frac{x}{\epsilon}\right).$$

L'équation (7.2) devient une série en ϵ

$$\begin{aligned}
&-\epsilon^{-2}\Big(\operatorname{div}_y(D\nabla_y u_0)\Big)\left(x, \frac{x}{\epsilon}\right)\\
&-\epsilon^{-1}\Big(\operatorname{div}_y(D(\nabla_x u_0 + \nabla_y u_1)) + \operatorname{div}_x(D\nabla_y u_0)\Big)\left(x, \frac{x}{\epsilon}\right)\\
&-\sum_{i=0}^{+\infty}\epsilon^i\Big(\operatorname{div}_x\left(D(\nabla_x u_i + \nabla_y u_{i+1})\right)\\
&\qquad + \operatorname{div}_y\left(D(\nabla_x u_{i+1} + \nabla_y u_{i+2})\right)\Big)\left(x, \frac{x}{\epsilon}\right) \qquad = f(x).
\end{aligned}$$

En identifiant chaque puissance de ϵ dans la série ci-dessus comme une équation individuelle, on obtient une "cascade" d'équations (sur le principe qu'une série entière de ϵ est nulle si et seulement si tous ses coefficients sont nuls). En fait, seuls les trois premiers termes de cette série (en ϵ^{-2}, ϵ^{-1}, et ϵ^0) suffisent pour notre propos. Pour résoudre ces équations nous aurons besoin du résultat suivant d'existence et d'unicité.

Lemme 7.1.1 *Soit $H^1_\#(Y)$ l'espace de Sobolev des fonctions périodiques de période Y. Soit $g \in L^2(Y)$. Le problème aux limites*

$$\begin{cases} -\operatorname{div}_y\Big(D(y)\nabla_y v(y)\Big) = g(y) \text{ dans } Y \\ y \to v(y) \ Y\text{-périodique} \end{cases}$$

admet une unique solution $v \in H^1_\#(Y)/\mathbf{R}$ *(à une constante additive près) si et seulement si*

$$\int_Y g(y)\,dy = 0. \tag{7.4}$$

Démonstration. Vérifions que (7.4) (appelée alternative de Fredholm, voir le Théorème 4.2.2) est une condition nécessaire d'existence d'une solution. On intègre le membre de gauche de l'équation sur Y

$$\int_Y \mathrm{div}_y \Big(D(y)\nabla_y v(y)\Big)dy = \int_{\partial Y} D(y)\nabla_y v(y)\cdot n\,ds = 0$$

à cause des conditions aux limites de périodicité. En effet, la fonction $D(y)\nabla v(y)$, étant périodique, prend des valeurs égales sur des cotés opposés de Y, tandis que la normale n change de signe. Par conséquent, le membre de droite de l'équation a nécessairement une moyenne nulle sur Y : c'est (7.4). Nous laissons au lecteur le soin d'appliquer le Théorème de Lax-Milgram dans $H^1_\#(Y)/\mathbf{R}$ pour montrer que (7.4) est aussi suffisant. ■

Remarque 7.1.2 *L'espace de Sobolev* $H^1_\#(Y)$ *est défini comme l'ensemble des fonction définies sur* $\mathbf{R}^N$ *tout entier,* Y*-périodiques au sens de (7.1) et qui appartiennent à* $H^1(\Omega)$ *pour tout ouvert borné* Ω *de* $\mathbf{R}^N$*. C'est un espace de Hilbert muni du produit scalaire usuel de* $H^1(Y)$*. Nous renvoyons à la Remarque 2.2.9 pour plus de détails sur les problèmes aux limites périodiques. L'espace* $H^1_\#(Y)/\mathbf{R}$ *est l'ensemble des (classes de) fonctions de* $H^1_\#(Y)$ *définies à l'addition d'une constante près.*

L'équation en ϵ^{-2} est

$$-\mathrm{div}_y \Big(D(y)\nabla_y u_0(x,y)\Big) = 0,$$

qui s'interprète comme une équation dans la cellule unité Y avec des conditions aux limites de périodicité. Dans cette équation y est la variable et x n'est qu'un paramètre. En vertu du Lemme 7.1.1 il existe une unique solution de cette équation, à une constante additive près. On en déduit donc que u_0 est une fonction constante par rapport à y mais qui peut néanmoins dépendre de x, c'est-à-dire qu'il existe une fonction $u(x)$, qui dépend seulement de x, telle que

$$u_0(x,y) \equiv u(x).$$

Comme $\nabla_y u_0 = 0$, **l'équation en ϵ^{-1}** devient

$$-\mathrm{div}_y \Big(D(y)\nabla_y u_1(x,y)\Big) = \mathrm{div}_y \Big(D(y)\nabla_x u(x)\Big), \tag{7.5}$$

qui est une équation pour l'inconnue u_1 dans la cellule de périodicité Y. Si on moyenne le membre de droite dans (7.5) on obtient

$$\int_Y \mathrm{div}_y \Big(D(y)\nabla_x u(x)\Big)dy = \int_{\partial Y} D(y)\nabla_x u(x)\cdot n\,dS(y) = 0$$

car $D(y)$, étant périodique, prend des valeurs égales sur des faces opposées du cube Y alors que la normale n change de signe sur des faces opposées. On peut donc appliquer le Lemme 7.1.1 qui affirme que l'équation (7.5) admet une unique solution, à une constante additive près, ce qui nous permet de calculer $u_1(x,y)$ en fonction du gradient $\nabla_x u(x)$. On note $(e_i)_{1\leq i\leq N}$ la base canonique de $\mathbf{R}^N$. Pour chaque vecteur e_i, on appelle **problème de cellule** l'équation suivante avec condition aux limites de périodicité

$$\begin{cases} -\operatorname{div}_y \Big(D(y)\left(e_i + \nabla_y w_i(y)\right)\Big) = 0 & \text{dans } Y \\ y \to w_i(y) & Y\text{-périodique.} \end{cases} \tag{7.6}$$

En vertu du Lemme 7.1.1, (7.6) admet une unique solution w_i (à une constante additive près) que l'on interprète comme le flux local ou microscopique causé par le courant ou gradient moyen e_i. Par linéarité, on calcule facilement $u_1(x,y)$, solution de (7.5), en fonction des dérivées partielles de $u(x)$ et des $w_i(y)$

$$u_1(x,y) = \sum_{i=1}^{N} \frac{\partial u}{\partial x_i}(x) w_i(y). \tag{7.7}$$

En fait u_1 est défini à l'addition d'une fonction de x près, mais cela n'importe pas puisque seul le gradient $\nabla_y u_1$ joue un rôle dans la suite.

Finalement, **l'équation en ϵ^0** est

$$-\operatorname{div}_y \Big(D(y)\nabla_y u_2(x,y)\Big) = \operatorname{div}_y \Big(D(y)\nabla_x u_1\Big) + \operatorname{div}_x \Big(D(y)(\nabla_y u_1 + \nabla_x u)\Big) + f,$$

qui est une équation pour l'inconnue u_2 dans la cellule de périodicité Y. Selon le Lemme 7.1.1, cette équation admet une unique solution, à une constante additive près, si la condition de compatibilité suivante est vérifiée

$$\int_Y \Big[\operatorname{div}_y \Big(D(y)\nabla_x u_1\Big) + \operatorname{div}_x \Big(D(y)(\nabla_y u_1 + \nabla_x u)\Big) + f(x)\Big]\, dy = 0.$$

En utilisant la périodicité, le premier terme s'annule et la relation ci-dessus se simplifie en

$$-\operatorname{div}_x \left(\int_Y D(y)\left(\nabla_y u_1 + \nabla_x u\right) dy \right) = f(x)\,,$$

dans laquelle on insére l'expression (7.7) pour $u_1(x,y)$ (qui dépend linéairement de $\nabla_x u(x)$) pour obtenir finalement **l'équation homogénéisée** pour u

$$\begin{cases} -\operatorname{div}_x \Big(D^* \nabla_x u(x)\Big) = f(x) & \text{dans } \Omega, \\ u = 0 & \text{sur } \partial\Omega, \end{cases} \tag{7.8}$$

avec

$$D^* \nabla_x u = \sum_{j=1}^{N} \frac{\partial u}{\partial x_j} \int_Y D(y)\left(e_j + \nabla_y w_j\right) dy.$$

La condition aux limites de Dirichlet pour u provient du développement asymptotique appliqué à la même condition aux limites pour u_ϵ. Le **tenseur homogénéisé** D^* est donc défini par ses coefficients

$$D^*_{ij} = \int_Y D(y)\,(e_j + \nabla_y w_j)\cdot e_i dy. \tag{7.9}$$

De manière équivalente, D^* est défini par une formule plus symétrique

$$D^*_{ij} = \int_Y D(y)\,(e_j + \nabla_y w_j(y))\cdot (e_i + \nabla_y w_i(y))\,dy, \tag{7.10}$$

qui s'obtient en remarquant qu'à cause de la formulation variationnelle de (7.6)

$$\int_Y D(y)\,(e_j + \nabla_y w_j(y))\cdot \nabla_y w_i(y)\,dy = 0.$$

Les formules (7.9) ou (7.10) ne sont pas totalement explicites car elles dépendent des solutions w_i des problèmes de cellule que l'on ne peut pas résoudre analytiquement en général. Le tenseur constant D^* décrit les propriétés effectives ou homogénéisées du milieu hétérogène $D\left(\frac{x}{\epsilon}\right)$. Remarquons qu'il ne dépend pas du choix du domaine Ω, de la source f, ou des conditions aux limites sur $\partial\Omega$.

7.1.3 Convergence

La méthode des développements asymptotiques à deux échelles est seulement formelle d'un point de vue mathématique. En général, elle conduit heuristiquement à des résultats corrects, mais elle ne constitue pas une preuve du procédé d'homogénéisation. La raison en est double : d'une part, la série postulée (7.3) ne converge pas en général, d'autre part, elle n'est pas exacte après les deux premiers termes (ce sont les seuls que l'on peut pleinement justifier). En effet, cette série ne vérifie pas les conditions aux limites sur le bord $\partial\Omega$ et ne tient pas compte d'éventuels phénomènes de couches limites au voisinage du bord (qui sont pourtant présentes dans la plupart des cas). Nous renvoyons à [8], [29] pour plus de détails.

Néanmoins, il est possible de justifier rigoureusement que l'équation (7.8) est bien l'équation homogénéisée du problème d'origine (7.2), c'est-à-dire que u_ϵ est proche de la solution homogénéisée u lorsque ϵ est petit. Nous nous contentons ici d'énoncer ce résultat (voir [8], [29] pour une preuve).

Théorème 7.1.3 *Soit u_ϵ la solution de (7.2). Soit u la solution du problème homogénéisé (7.8), et $(w_i)_{1\le i\le N}$ les solutions des problèmes de cellule (7.6). On a*

$$u_\epsilon(x) = u(x) + \epsilon\sum_{i=1}^N \frac{\partial u}{\partial x_i}(x) w_i\left(\frac{x}{\epsilon}\right) + r_\epsilon \quad \textit{avec} \quad \|r_\epsilon\|_{H^1(\Omega)} \le C\sqrt{\epsilon}.$$

En particulier,

$$\|u_\epsilon - u\|_{L^2(\Omega)} + \left\|\nabla u_\epsilon(x) - \nabla u(x) - \sum_{i=1}^N \frac{\partial u}{\partial x_i}(x)(\nabla_y w_i)\left(\frac{x}{\epsilon}\right)\right\|_{L^2(\Omega)^N} \le C\sqrt{\epsilon}.$$

Il est bon de noter que si le terme correcteur, ϵu_1, est petit (de l'ordre de ϵ) en norme L^2, il n'en est pas de même en norme H^1 (ou de manière équivalente pour son gradient en norme L^2) où le terme correcteur est d'ordre 1. Le Théorème 7.1.3 se généralise facilement à d'autres types de conditions aux limites (Neumann par exemple) ou à d'autres types d'équations (diffusion multi-groupes par exemple).

7.2 Homogénéisation en transport

7.2.1 Homogénéisation d'un modèle stationnaire

Les premiers résultats mathématiques sur l'homogénéisation d'équations de transport remontent aux travaux de J. Keller et E. Larsen [32] (voir aussi [4], [49]) bien que des travaux plus formels en physique pré-existaient [21].

Nous allons commencer par considérer une équation stationnaire avec des sources dans un milieu légèrement **sous-critique**. Le domaine spatial, noté Ω, est supposé borné et l'ensemble des vitesses $\mathcal{V}$ est, pour simplifier, pris égal à la sphère unité $\mathbf{S}_{N-1}$. Pour simplifier les notations nous supposerons que la mesure dv sur la sphère unité $\mathcal{V} = \mathbf{S}_{N-1}$ est pondérée de manière à ce que l'intégrale corresponde à la moyenne, autrement dit $\int_{\mathcal{V}} dv = 1$. On cherche donc la solution $u_\epsilon(x,v)$ de

$$\left\{ \begin{array}{l} \epsilon^{-1} v \cdot \nabla u_\epsilon + \epsilon^{-2} \sigma(\frac{x}{\epsilon}) \left(u_\epsilon - \displaystyle\int_{\mathcal{V}} u_\epsilon \, dv \right) + \tilde{\sigma}(x, \frac{x}{\epsilon}) u_\epsilon \\ \qquad\qquad = S(x, \frac{x}{\epsilon}, v) \quad \text{dans } \Omega \times \mathcal{V} \\ u_\epsilon(x,v) = 0 \qquad\qquad \text{sur } \Gamma^- \end{array} \right. \tag{7.11}$$

avec la frontière rentrante $\Gamma^- = \{x \in \partial\Omega, v \in \mathcal{V}, v \cdot n(x) < 0\}$. Les coefficients ou sections efficaces $\sigma(y)$, $\tilde{\sigma}(x,y)$ sont des fonctions positives, bornées et Y-périodiques par rapport à la variable rapide $y \in Y = (0,1)^N$. Elles peuvent être éventuellement discontinues en y pour modéliser des discontinuités matérielles mais on suppose que $\tilde{\sigma}(x,y)$ est régulier par rapport à x. Remarquons que le choix de la "mise à l'échelle", c'est-à-dire des puissances de ϵ, dans (7.11) implique de facto que le milieu est à peine sous-critique car presque critique à ϵ^2 près. En fait, ce choix de mise à l'échelle garantit aussi que l'on obtiendra un modèle homogénéisé de type diffusion (de manière tout à fait similaire à ce que l'on a fait dans le modèle (4.1) du Chapitre 4). Physiquement, ce choix correspond à supposer que le libre parcours moyen d'une particule est précisément de l'ordre de la période ϵ.

Comme précédemment, l'hypothèse de départ est de supposer que la solution u_ϵ de l'équation (7.11) est donnée par un développement en série de ϵ, dit "à deux échelles", du type

$$u_\epsilon(x,v) = \sum_{i=0}^{+\infty} \epsilon^i u_i \left(x, \frac{x}{\epsilon}, v \right), \tag{7.12}$$

où chaque terme $u_i(x,y,v)$ est une fonction de trois variables $x \in \Omega$, $y \in Y = (0,1)^N$ et $v \in \mathcal{V}$, qui est périodique en y de période Y. En y injectant (7.12) l'équation (7.11) devient une série en ϵ dont les coefficients forment une "cascade" d'équations

$$
\begin{aligned}
&-\epsilon^{-2}\left[v\cdot\nabla_y u_0 + \sigma(y)\left(u_0 - \int_{\mathcal{V}} u_0\,dv\right)\right]\left(x,\frac{x}{\epsilon},v\right)\\
&-\epsilon^{-1}\left[v\cdot\nabla_y u_1 + v\cdot\nabla_x u_0 + \sigma(y)\left(u_1 - \int_{\mathcal{V}} u_1\,dv\right)\right]\left(x,\frac{x}{\epsilon},v\right)\\
&-\sum_{i=0}^{+\infty}\epsilon^i\left[v\cdot\nabla_y u_{i+2} + v\cdot\nabla_x u_{i+1} + \sigma(y)\left(u_{i+2} - \int_{\mathcal{V}} u_{i+2}\,dv\right)\right.\\
&\qquad\qquad\left. + \tilde{\sigma}(x,y)u_i\right]\left(x,\frac{x}{\epsilon},v\right) \;=\; S\left(x,\frac{x}{\epsilon},v\right).
\end{aligned}
$$

Pour résoudre ces équations nous aurons besoin du résultat suivant d'existence et d'unicité qui est l'équivalent pour le transport du Lemme 7.1.1 en diffusion.

Lemme 7.2.1 *Soit $g \in L^2(Y\times\mathcal{V})$. Le problème aux limites*

$$
\begin{cases}
v\cdot\nabla_y\phi + \sigma(y)\left(\phi - \displaystyle\int_{\mathcal{V}}\phi\,dv\right) = g(y,v) \quad \textit{dans } Y\times\mathcal{V}\\
y\to\phi(y,v)\ Y\textit{-périodique}
\end{cases}
$$

admet une unique solution $\phi \in L^2(Y\times\mathcal{V})/\mathbf{R}$ (à une constante additive près) si et seulement si

$$
\int_{\mathcal{V}}\int_Y g(y,v)\,dy\,dv = 0. \tag{7.13}
$$

Démonstration. Tout d'abord il est clair que la solution ϕ, si elle existe, n'est définie qu'à l'addition d'une constante près puisque $\int_{\mathcal{V}} dv = 1$. Vérifions que (7.13) est une condition nécessaire d'existence d'une solution. On intègre l'équation sur Y et le terme de transport disparaît car

$$
\int_Y v\cdot\nabla_y\phi\,dy = \int_{\partial Y} v\cdot n\phi\,ds = 0
$$

à cause des conditions aux limites de périodicité. On obtient donc

$$
\int_Y \sigma\left(\phi - \int_{\mathcal{V}}\phi\,dv\right)dy = \int_Y g\,dy
$$

que l'on intègre par rapport à v

$$
\int_{\mathcal{V}}\int_Y \sigma(y)\left(\phi - \int_{\mathcal{V}}\phi\,dv\right)dy\,dv = \int_{\mathcal{V}}\int_Y g\,dy\,dv.
$$

Comme la fonction $\left(\phi - \int_{\mathcal{V}}\phi\,dv\right)$ est de moyenne nulle en v et que σ ne dépend pas de v, on en déduit bien la condition (7.13). Nous laissons au lecteur le soin

d'appliquer les résultats d'existence du chapitre 2 (voir la Remarque 2.2.9) pour montrer que (7.13) est aussi suffisant (voir aussi le Théorème 4.2.2). ■

L'équation en ϵ^{-2} est

$$v \cdot \nabla_y u_0 + \sigma(y) \left(u_0 - \int_{\mathcal{V}} u_0 \, dv \right) = 0,$$

qui s'interprète comme une équation dans la cellule unité Y avec des conditions aux limites de périodicité. Dans cette équation y est la variable et x n'est qu'un paramètre. En vertu du Lemme 7.2.1 il existe une unique solution de cette équation, à une constante additive près. On en déduit donc que u_0 est une fonctions constante par rapport à (y, v) mais qui peut néanmoins dépendre de x, c'est-à-dire qu'il existe une fonction $u(x)$, qui dépend seulement de x, telle que

$$u_0(x, y, v) \equiv u(x).$$

L'équation en ϵ^{-1} est

$$v \cdot \nabla_y u_1 + \sigma(y) \left(u_1 - \int_{\mathcal{V}} u_1 \, dv \right) = -v \cdot \nabla_x u(x),$$

qui est une équation pour l'inconnue u_1 dans la cellule de périodicité Y. Comme $\mathcal{V} = \mathbf{S}_{N-1}$ est symétrique, on a

$$\int_{\mathcal{V}} v \cdot \nabla_x u(x) \, dv = 0,$$

et le Lemme 7.2.1 affirme que l'équation en ϵ^{-1} admet une unique solution, à une constante additive près, ce qui nous permet de calculer $u_1(x, y, v)$ en fonction du gradient $\nabla_x u(x)$. On note $(e_i)_{1 \le i \le N}$ la base canonique de $\mathbf{R}^N$. Pour chaque vecteur e_i, on appelle **problème de cellule** l'équation suivante avec condition aux limites de périodicité

$$\begin{cases} v \cdot \nabla_y w_i + \sigma(y) \left(w_i - \displaystyle\int_{\mathcal{V}} w_i \, dv \right) = -v \cdot e_i & \text{dans } Y \times \mathcal{V} \\ y \to w_i(y, v) & Y\text{-périodique.} \end{cases} \tag{7.14}$$

En vertu du Lemme 7.2.1, (7.14) admet une unique solution w_i (à une constante additive près). Par linéarité, on calcule facilement $u_1(x, y, v)$ en fonction des dérivées partielles de $u(x)$ et des $w_i(y, v)$

$$u_1(x, y, v) = \sum_{i=1}^{N} \frac{\partial u}{\partial x_i}(x) w_i(y, v). \tag{7.15}$$

En fait u_1 est défini à l'addition d'une fonction de x près, mais cela n'importera pas dans la suite.

Finalement, **l'équation en** ϵ^0 est

$$v \cdot \nabla_y u_2 + \sigma(y) \left(u_2 - \int_{\mathcal{V}} u_2 \, dv \right) = -v \cdot \nabla_x u_1 - \tilde{\sigma}(x,y) u + S,$$

qui est une équation pour l'inconnue u_2 dans la cellule de périodicité Y. Selon le Lemme 7.2.1, cette équation admet une unique solution, à une constante additive près, si la condition de compatibilité suivante est vérifiée

$$\int_Y \int_{\mathcal{V}} \left[-v \cdot \nabla_x u_1(x,y,v) - \tilde{\sigma}(x,y) u(x) + S(x,y,v) \right] dy \, dv = 0. \tag{7.16}$$

On introduit les moyennes

$$\sigma^*(x) = \int_Y \tilde{\sigma}(x,y) \, dy \quad \text{et} \quad S^*(x) = \int_Y \int_{\mathcal{V}} S(x,y,v) \, dy \, dv$$

et le tenseur homogénéisé D^* défini par ses composantes

$$D^*_{ij} = -\frac{1}{2} \left(\int_Y \int_{\mathcal{V}} v_j w_i(y,v) \, dy \, dv + \int_Y \int_{\mathcal{V}} v_i w_j(y,v) \, dy \, dv \right). \tag{7.17}$$

Remarquons que l'addition d'une constante à w_i ne change pas la valeur de D^*_{ij} car $\int_{\mathcal{V}} v_j dv = 0$. La définition (7.17) est parfois appelée formule de Kubo. En y insérant l'expression (7.15) pour $u_1(x,y,v)$ (qui dépend linéairement de $\nabla_x u(x)$) l'équation (7.16) est alors équivalente à **l'équation homogénéisée**

$$\begin{cases} -\operatorname{div}_x \left(D^* \nabla_x u(x) \right) + \sigma^*(x) u(x) = S^*(x) & \text{dans } \Omega, \\ u = 0 & \text{sur } \partial\Omega. \end{cases} \tag{7.18}$$

La condition aux limites de Dirichlet pour u provient du développement asymptotique appliqué à la même condition aux limites pour u_ϵ. En effet au premier ordre ϵ^0 on a

$$u_0(x,y,v) \equiv u(x) = 0 \text{ sur } \Gamma^- = \{ x \in \partial\Omega, v \in \mathcal{V}, v \cdot n(x) < 0 \}.$$

Comme $u(x)$ ne dépend pas de v, on en déduit que cette fonction doit être nulle sur tout le bord $\partial\Omega$. Remarquons qu'à l'ordre suivant ϵ^1 il n'est pas possible, en général, d'imposer que

$$u_1(x,y,v) \equiv \sum_{i=1}^N \frac{\partial u}{\partial x_i}(x) w_i(y,v) = 0 \text{ sur } \Gamma^-$$

car, d'une part w_i ne dépend pas de x et ne peut donc prendre une valeur particulière sur la frontière $\partial\Omega$ (différente d'à l'intérieur de Ω), et d'autre part on ne peut pas imposer que $\nabla u(x)$ s'annule sur $\partial\Omega$ puisqu'on a déjà imposé que $u(x)$ s'annule. Cela montre que le développement asymptotique à deux échelles postulé en (7.12) n'est pas correct tel quel, mais qu'il faut lui adjoindre des termes supplémentaires, dits de "couches limites", pour qu'il vérifie exactement les conditions aux limites. Pour plus de détails nous renvoyons le lecteur à [32].

En conclusion, on a formellement établi le résultat suivant.

Proposition 7.2.2 *La solution u_ϵ de l'équation de transport (7.11), quand ϵ tends vers zéro, est asymptotiquement donnée par*

$$u_\epsilon(x,v) \approx u(x) + \epsilon \sum_{i=1}^{N} \frac{\partial u}{\partial x_i}(x) w_i\left(\frac{x}{\epsilon}, v\right),$$

où w_i est solution du problème de cellule (7.14) et u est solution de l'équation homogénéisée de diffusion (7.18).

On vérifie que le problème homogénéisé (7.18), une équation de diffusion, est bien posé car son tenseur de diffusion est bien coercif. On pourra donc lui appliquer le théorème de Lax-Milgram pour démontrer l'existence et l'unicité d'une solution.

Lemme 7.2.3 *Le tenseur D^* est défini positif.*

Démonstration. Soit ξ un vecteur non nul de $\mathbf{R}^N$. Nous allons montrer que $D^*\xi \cdot \xi > 0$. Pour cela on introduit la fonction w_ξ définie par

$$w_\xi(y,v) = \sum_{i=1}^{N} \xi_i w_i(y,v)$$

qui est solution de l'équation

$$\begin{cases} v \cdot \nabla_y w_\xi + \sigma(y)\left(w_\xi - \int_{\mathcal{V}} w_\xi \, dv\right) = -v \cdot \xi & \text{dans } Y \times \mathcal{V} \\ y \to w_\xi(y,v) & Y\text{-périodique.} \end{cases} \tag{7.19}$$

On multiplie l'équation (7.19) par w_ξ et on l'intègre sur Y. Le terme de transport disparaît car

$$\int_Y v \cdot \nabla_y w_\xi \, w_\xi \, dy = \frac{1}{2} \int_{\partial Y} v \cdot n \, w_\xi^2 \, ds = 0$$

à cause des conditions aux limites de périodicité. On obtient donc

$$\int_Y \sigma \left(w_\xi - \int_{\mathcal{V}} w_\xi \, dv\right) w_\xi \, dy = -\int_Y v \cdot \xi \, w_\xi \, dy$$

que l'on intègre par rapport à v

$$\int_{\mathcal{V}} \int_Y \sigma \left(w_\xi - \int_{\mathcal{V}} w_\xi \, dv\right) w_\xi \, dy \, dv = -\int_{\mathcal{V}} \int_Y v \cdot \xi \, w_\xi \, dy \, dv.$$

Comme la fonction $\left(w_\xi - \int_{\mathcal{V}} w_\xi \, dv\right)$ est de moyenne nulle en v, on a aussi

$$\int_{\mathcal{V}} \int_Y \sigma \left(w_\xi - \int_{\mathcal{V}} w_\xi \, dv\right) \left(\int_{\mathcal{V}} w_\xi \, dv\right) dy \, dv = 0.$$

En combinant les deux on en déduit

$$\int_{\mathcal{V}} \int_Y \sigma \left(w_\xi - \int_{\mathcal{V}} w_\xi \, dv\right)^2 dy \, dv = -\int_{\mathcal{V}} \int_Y v \cdot \xi \, w_\xi \, dy \, dv = D^*\xi \cdot \xi$$

à cause de la définition (7.17) de D^*. On a donc $D^*\xi\cdot\xi \geq 0$. Montrons que cette inégalité est stricte. Si $D^*\xi\cdot\xi = 0$ pour un vecteur $\xi \neq 0$, alors on en déduit que $w_\xi \equiv \int_{\mathcal{V}} w_\xi\, dv$ est indépendant de v et en reportant dans l'équation (7.19) on obtient

$$v\cdot\nabla_y(w_\xi(y)+\xi\cdot y) = 0 \text{ dans } Y\times\mathcal{V}.$$

Comme v est quelconque et w_ξ ne dépend pas de v, cela implique que $w_\xi(y) = -\xi\cdot y + C$ où C est une constante quelconque. Le caractère affine de w_ξ contredit la condition aux limites de périodicité dans (7.19). Par conséquent, on doit avoir $D^*\xi\cdot\xi > 0$. ■

La formule (7.17) n'est pas totalement explicite car elle dépend des solutions w_i des problèmes de cellule que l'on ne peut pas résoudre analytiquement en général. Le tenseur constant D^*, qui décrit les propriétés effectives ou homogénéisées du milieu hétérogène, ne dépend pas du choix du domaine Ω, de la source S, ou des conditions aux limites sur $\partial\Omega$.

7.2.2 Homogénéisation d'un modèle instationnaire

Nous étudions maintenant le cas d'une équation de transport dépendant du temps (problème cinétique en neutronique). Une différence majeure avec le cas précédent est qu'il n'est plus nécessaire de supposer que le milieu est légèrement sous-critique. La **criticité**, ou non, sera automatiquement détectée par le processus d'homogénéisation qui est, par conséquent, un peu plus compliqué. Pour simplifier nous négligeons la présence éventuelle de sources et nous nous contentons d'étudier l'évolution d'une donnée initiale. On suppose toujours que le domaine spatial Ω est borné et que l'ensemble des vitesses est la sphère unité $\mathcal{V} = \mathbf{S}_{N-1}$ avec une mesure dv telle que $\int_{\mathcal{V}} dv = 1$. On cherche la solution $u_\epsilon(t,x,v)$ de

$$\begin{cases} \dfrac{\partial u_\epsilon}{\partial t} + \epsilon^{-1}v\cdot\nabla u_\epsilon + \epsilon^{-2}\sigma(\dfrac{x}{\epsilon},v)u_\epsilon = \epsilon^{-2}\displaystyle\int_{\mathcal{V}}\tilde\sigma(\frac{x}{\epsilon},v,v')u_\epsilon(v')\,dv' \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{dans } \Omega\times\mathcal{V}, \\ u_\epsilon(t=0,x,v) = u^0_\epsilon(x,v) \qquad \text{dans } \Omega\times\mathcal{V}, \\ u_\epsilon(x,v) = 0 \qquad\qquad\qquad \text{sur } \Gamma^-, \end{cases} \tag{7.20}$$

avec la frontière rentrante $\Gamma^- = \{x\in\partial\Omega, v\in\mathcal{V}, v\cdot n(x) < 0\}$. La donnée initiale est supposée être de la forme

$$u^0_\epsilon(x,v) = u^0(x,\frac{x}{\epsilon},v) \tag{7.21}$$

où $u^0(x,y,v)$ est Y-périodique par rapport à la variable y. Les sections efficaces $\sigma(y,v)$ et $\tilde\sigma(y,v,v')$ sont positives, bornées et Y-périodiques par rapport à la variable y mais ne sont pas supposées isotropes, c'est-à-dire qu'elles peuvent dépendre de la vitesse v. Néanmoins, nous allons supposer que ce sont des fonctions paires de la vitesse, ce qui correspond à un milieu "symétrique" ou réversible du point de vue des trajectoires,

$$\sigma(y,v) = \sigma(y,-v) \quad \text{et} \quad \tilde\sigma(y,v,v') = \tilde\sigma(y,-v,-v'). \tag{7.22}$$

Pour prendre en compte la possible non-criticité du problème (7.20) nous modifions notre postulat usuel de développement asymptotique à deux échelles en introduisant un paramètre supplémentaire $\lambda^* \in \mathbf{R}$, à déterminer, qui s'interprète comme l'inverse d'un taux de décroissance s'il est positif ou de croissance s'il est négatif. Autrement dit, on suppose que la solution u_ϵ de l'équation (7.20) s'écrit

$$u_\epsilon(t,x,v) = e^{-\frac{\lambda^* t}{\epsilon^2}} \sum_{i=0}^{+\infty} \epsilon^i u_i\left(t, x, \frac{x}{\epsilon}, v\right), \tag{7.23}$$

où chaque terme $u_i(t,x,y,v)$ est une fonction de quatre variables $t \in \mathbf{R}^+$, $x \in \Omega$, $y \in Y = (0,1)^N$ et $v \in \mathcal{V}$, qui est périodique en y de période Y. En y injectant (7.23) l'équation (7.20) devient une série en ϵ dont les coefficients forment une "cascade" d'équations. Pour résoudre ces équations nous aurons besoin d'un résultat d'existence et d'unicité pour les "problèmes de cellule" qui soit l'équivalent des Lemmes 7.1.1 et 7.2.1 : ça sera le Lemme 7.2.6 ci-dessous. Auparavant, nous devons donner (sans preuve) une version du théorème de Krein-Rutman ou du Théorème 6.4.1 qui corresponde au problème de cellule avec condition aux limites de périodicité.

Lemme 7.2.4 *Il existe une valeur propre $\lambda^* \in \mathbf{R}$ et une fonction propre strictement positive $\psi(y,v) > 0$ dans $Y \times \mathcal{V}$ telles que*

$$\begin{cases} -\lambda^*\psi + v \cdot \nabla_y \psi + \sigma(y,v)\psi = \displaystyle\int_{\mathcal{V}} \tilde{\sigma}(y,v,v')\psi(y,v')\,dv' \ \text{ dans } Y \times \mathcal{V} \\ y \to \psi(y,v) \qquad\qquad Y\text{-périodique.} \end{cases} \tag{7.24}$$

De plus, λ^* est aussi valeur propre du problème adjoint de (7.24) pour une fonction propre adjointe strictement positive $\psi^*(y,v) > 0$ dans $Y \times \mathcal{V}$

$$\begin{cases} -\lambda^*\psi^* - v \cdot \nabla_y \psi^* + \sigma(y,v)\psi^* = \displaystyle\int_{\mathcal{V}} \tilde{\sigma}^*(y,v,v')\psi^*(y,v')\,dv' \ \text{ dans } Y \times \mathcal{V} \\ y \to \psi^*(y,v) \qquad\qquad Y\text{-périodique,} \end{cases} \tag{7.25}$$

avec la section efficace adjointe $\tilde{\sigma}^*$ définie par $\tilde{\sigma}^*(y,v',v) = \tilde{\sigma}(y,v,v')$. La valeur propre λ^* est simple et de plus petit module, parmi toutes les valeurs propres, pour chacun de ces deux problèmes. Par ailleurs, parmi toutes les fonctions propres possibles, seules ψ et ψ^* sont positives dans tout le domaine $Y \times \mathcal{V}$.

Remarque 7.2.5 *A cause de la condition de symétrie en vitesse (7.22) sur les sections efficaces, il est facile de vérifier que la fonction propre adjointe ψ^* est simplement donnée par*

$$\psi^*(y,v) = \psi(y,-v).$$

Par ailleurs, puisque les fonctions propres sont définies à un coefficient multiplicatif près, on décide de les normaliser de manière à ce que

$$\int_Y \int_{\mathcal{V}} \psi(y,v)\psi^*(y,v)\,dy\,dv = 1. \tag{7.26}$$

Lemme 7.2.6 *Soit un terme source $S(y,v) \in L^2(Y \times \mathcal{V})$. Le problème*

$$\begin{cases} -\lambda^*\phi + v\cdot\nabla_y\phi + \sigma(y,v)\phi = \displaystyle\int_{\mathcal{V}} \tilde{\sigma}(y,v,v')\phi(y,v')\,dv' + S(y,v) \\ \qquad\qquad\qquad\qquad\qquad\qquad \textit{dans } Y\times\mathcal{V}, \\ y \to \phi(y,v) \qquad\qquad Y\textit{-périodique,} \end{cases} \tag{7.27}$$

admet une solution $\phi(y,v) \in L^2(Y\times\mathcal{V})$, unique à l'addition près d'un multiple de la première fonction propre $\psi(y,v)$, si et seulement si la source vérifie

$$\int_Y\int_{\mathcal{V}} S(y,v)\,\psi^*(y,v)\,dy\,dv = 0, \tag{7.28}$$

où $\psi^*(y,v)$ est la première fonction propre adjointe, solution de (7.25).

Démonstration. Pour vérifier que la condition (7.28) est nécessaire on multiplie l'équation (7.27) par ψ^* et on intègre par parties sur $Y\times\mathcal{V}$. Le terme de transport devient

$$\int_Y\int_{\mathcal{V}} v\cdot\nabla_y\phi\,\psi^*\,dy\,dv = -\int_Y\int_{\mathcal{V}} v\cdot\nabla_y\psi^*\,\phi\,dy\,dv.$$

Par ailleurs, en intervertissant l'ordre d'intégration en v et v' et en utilisant la section adjointe $\tilde{\sigma}^*$, le terme de collision devient

$$\int_Y\int_{\mathcal{V}}\int_{\mathcal{V}} \tilde{\sigma}(y,v,v')\phi(y,v')\,dv'\,\psi^*(y,v)\,dv\,dy = \int_Y\int_{\mathcal{V}}\int_{\mathcal{V}} \tilde{\sigma}^*(y,v',v)\psi^*(y,v)\,dv\,\phi(y,v')\,dv'\,dy.$$

On fait ainsi apparaître l'équation pour ψ^* en facteur de ϕ dans la somme de ces deux intégrales, qui s'annule donc, ce qui conduit à (7.28). On admettra que la condition (7.28) est aussi suffisante. ■

L'équation en ϵ^{-2} est

$$-\lambda^* u_0 + v\cdot\nabla_y u_0 + \sigma u_0 = \int_{\mathcal{V}} \tilde{\sigma} u_0\,dv', \tag{7.29}$$

qui s'interprète comme un problème aux valeurs propres dans la cellule unité Y avec des conditions aux limites de périodicité. Autrement dit, λ^* est une valeur propre et u_0 une fonction propre. D'un point de vue physique, nous calculons des solutions qui sont des densités de particules, c'est-à-dire qui ont la propriété d'être positives. Or, d'après le Lemme 7.2.4, seule la fonction propre ψ, associée à la plus petite valeur propre λ^*, est positive. Donc, dans l'équation (7.29), le paramètre λ^* qui vient du développement asymptotique (7.23) est bien la plus petite valeur propre donnée par le Lemme 7.2.4. De plus, comme cette valeur propre est simple, la fonction propre u_0, solution de (7.29), est nécessairement proportionnelle à ψ. Autrement dit, il existe une fonction $u(t,x)$, indépendante de (y,v), telle que

$$u_0(t,x,y,v) \equiv u(t,x)\,\psi(y,v).$$

L'équation en ϵ^{-1} est

$$-\lambda^* u_1 + v \cdot \nabla_y u_1 + \sigma u_1 = \int_{\mathcal{V}} \tilde{\sigma} u_1 \, dv' - v \cdot \nabla_x u_0, \tag{7.30}$$

qui est une équation pour l'inconnue u_1 dans la cellule de périodicité Y avec la source $S(y,v) = -v \cdot \nabla_x u_0$. En vertu du Lemme 7.2.6, on peut résoudre en u_1 si

$$\int_Y \int_{\mathcal{V}} v \cdot \nabla_x u_0(t,x) \, \psi^*(y,v) \, dy \, dv = 0,$$

ce qui est verifié, puisque $u_0 \equiv u \, \psi$, si

$$\int_Y \int_{\mathcal{V}} v \, \psi(y,v) \, \psi^*(y,v) \, dy \, dv = 0.$$

Or, dans la Remarque 7.2.5 on a montré que $\psi^*(y,v) = \psi(y,-v)$, et comme $\mathcal{V} = \mathbf{S}_{N-1}$ est symétrique, on a

$$\begin{aligned} \int_Y \int_{\mathcal{V}} v \, \psi(y,v) \, \psi^*(y,v) \, dy \, dv \quad &= \int_Y \int_{\mathcal{V}} v \, \psi(y,v) \, \psi(y,-v) \, dy \, dv = \\ &- \int_Y \int_{\mathcal{V}} v \, \psi(y,-v) \, \psi(y,v) \, dy \, dv = 0, \end{aligned} \tag{7.31}$$

c'est-à-dire que la condition (7.28) est bien satisfaite. L'équation (7.30) admet donc une solution ce qui nous permet de calculer $u_1(t,x,y,v)$ en fonction du gradient $\nabla_x u(t,x)$. On note $(e_i)_{1 \le i \le N}$ la base canonique de $\mathbf{R}^N$. Pour chaque vecteur e_i, on appelle **problème de cellule** l'équation suivante avec condition aux limites de périodicité

$$\begin{cases} -\lambda^* w_i + v \cdot \nabla_y w_i + \sigma w_i = \displaystyle\int_{\mathcal{V}} \tilde{\sigma} w_i \, dv' - v \cdot e_i \psi & \text{dans } Y \times \mathcal{V} \\ y \to w_i(y,v) & Y\text{-périodique.} \end{cases} \tag{7.32}$$

Par linéarité, on a

$$u_1(t,x,y,v) = \sum_{i=1}^{N} \frac{\partial u}{\partial x_i}(t,x) w_i(y,v) + C(t,x) \psi(y,v), \tag{7.33}$$

où $C(t,x)$ est n'importe quelle fonction indépendante de (y,v) (sa valeur n'aura aucune importance dans la suite).

Finalement, **l'équation en ϵ^0** est

$$-\lambda^* u_2 + v \cdot \nabla_y u_2 + \sigma u_2 = \int_{\mathcal{V}} \tilde{\sigma} u_2 \, dv' - v \cdot \nabla_x u_1 - \frac{\partial u_0}{\partial t},$$

qui est une équation pour l'inconnue u_2 dans la cellule de périodicité Y. Selon le Lemme 7.2.6, cette équation admet une solution, unique à l'addition près d'un multiple de ψ, si la condition de compatibilité suivante est vérifiée

$$\int_Y \int_{\mathcal{V}} \left(-v \cdot \nabla_x u_1 - \frac{\partial u_0}{\partial t} \right) \psi^* \, dy \, dv = 0. \tag{7.34}$$

Rappelons que, d'après la Remarque 7.2.5, l'intégrale du produit $\psi\psi^*$ est normalisée par (7.26), et introduisons le tenseur homogénéisé D^* défini par ses composantes

$$D^*_{ij} = -\int_Y \int_{\mathcal{V}} v_j w_i(y,v)\psi^*(y,v)\,dy\,dv. \tag{7.35}$$

Remarquons que l'addition d'une fonction $C(t,x)\psi(y,v)$ à w_i ne change pas la valeur de D^*_{ij} à cause de la relation (7.31). En y insérant l'expression (7.33) pour $u_1(t,x,y,v)$ (qui dépend linéairement de $\nabla_x u(t,x)$) l'équation (7.34) est alors équivalente à **l'équation homogénéisée**

$$\begin{cases} \dfrac{\partial u}{\partial t} - \mathrm{div}_x\left(D^*\nabla_x u\right) = 0 & \text{dans } \Omega\times\mathbf{R}^+, \\ u(t=0,x) = \tilde{u}^0(x) & \text{dans } \Omega, \\ u = 0 & \text{sur } \partial\Omega\times\mathbf{R}^+. \end{cases} \tag{7.36}$$

La condition aux limites de Dirichlet pour u provient du développement asymptotique appliqué à la même condition aux limites pour u_ϵ. La donnée initiale est définie par

$$\tilde{u}^0(x) = \int_{\mathcal{V}}\int_Y \psi^*(y,v)\,u^0(x,y,v)\,dy\,dv$$

où $u^0(x,y,v)$ provient de la donnée initiale (7.21) pour u_ϵ. On obtient cette définition en appliquant le développement asymptotique dans la donnée initiale (7.21) et en moyennant avec le poids ψ^* pour être compatible avec la normalisation (7.26).

En conclusion, on a formellement établi le résultat suivant.

Proposition 7.2.7 *La solution u_ϵ de l'équation de transport (7.20), quand ϵ tends vers zéro, est asymptotiquement donnée par*

$$u_\epsilon(t,x,v) \approx e^{-\frac{\lambda^* t}{\epsilon^2}}\left(\psi\left(\frac{x}{\epsilon},v\right)u(t,x) + \epsilon\sum_{i=1}^N \frac{\partial u}{\partial x_i}(t,x) w_i\left(\frac{x}{\epsilon},v\right)\right), \tag{7.37}$$

où (λ^*,ψ) est solution du problème spectral (7.24), w_i est solution du problème de cellule (7.32) et u est solution de l'équation homogénéisée de diffusion (7.36).

Une application typique de la Proposition 7.2.7 est le calcul de la **puissance neutronique dans un réacteur nucléaire**. Un réacteur est un milieu très hétérogène à structure périodique (les assemblages de combustibles ou bien la cellule composée d'un barreau de combustible entouré d'eau). Un calcul "exact" est soit impossible, soit très couteux en temps de calcul, aussi est-il d'usage de faire des calculs approchés basés sur le premier terme de la formule (7.37). Au lieu de calculer la solution "exacte" u_ϵ on calcule la solution "reconstruite par homogénéisation", c'est-à-dire le produit $e^{-\frac{\lambda^* t}{\epsilon^2}}\psi\left(\frac{x}{\epsilon},v\right)u(t,x)$. Autrement dit, on fait un calcul local en transport et un calcul global en diffusion. Une comparaison entre ces deux solutions, moyennées en vitesse, est visible à la Figure 7.2. On y remarque que la solution exacte ne vérifie pas la condition aux

limites de Dirichlet, contrairement à celle reconstruite par homogénéisation. C'est la raison pour laquelle on introduit parfois un terme correctif de type "couche limite".

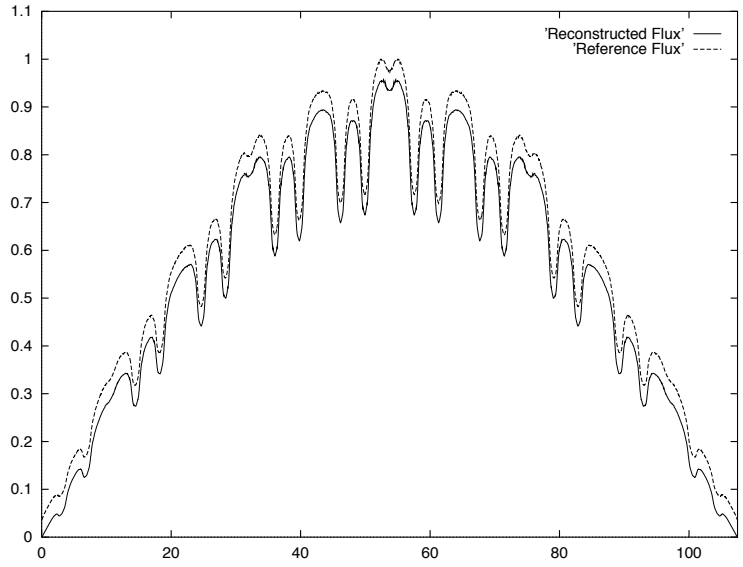

FIGURE 7.2 – Comparaison entre une solution exacte de l'équation du transport (trait pointillé) et une solution reconstruite par homogénéisation (trait plein), d'après [4].

Remarque 7.2.8 *On aurait pu ajouter un terme du type $\overline{\sigma}(x, \frac{x}{\epsilon}, v)u_\epsilon$ dans l'équation de transport (7.20) qui ne serait apparu que dans l'équation en ϵ^0 et aurait fait apparaître un terme supplémentaire $\overline{\sigma}^*(x)u$ dans l'équation homogénéisée (7.36), similaire au terme d'ordre zéro dans l'équation homogénéisée (7.18) de la section précédente. Cela ne change en rien les problèmes aux valeurs propres pour ψ et ψ^*, ni les problèmes de cellule pour w_i.*

Lemme 7.2.9 *Le tenseur D^* est défini positif.*

Démonstration. La preuve est parallèle à celle du Lemme 7.2.3 (voir [4]). Pour tout vecteur non nul $\xi \in \mathbf{R}^N$ nous allons montrer que $D^*\xi \cdot \xi > 0$. Pour cela on introduit la fonction θ_ξ définie par

$$\theta_\xi(y, v) = \sum_{i=1}^{N} \xi_i \frac{w_i(y, v)}{\psi(y, v)}$$

qui est solution de l'équation

$$\begin{cases} v\cdot\nabla_y\theta_\xi+\dfrac{\theta_\xi}{\psi}\displaystyle\int_{\mathcal{V}}\tilde\sigma\psi\,dv'-\frac{1}{\psi}\int_{\mathcal{V}}\tilde\sigma\theta_\xi\psi\,dv'=-v\cdot\xi & \text{dans } Y\times\mathcal{V}\\ y\to\theta_\xi(y,v) & Y\text{-périodique.}\end{cases}\tag{7.38}$$

On multiplie l'équation (7.38) par $\psi\psi^*\theta_\xi$ et on intègre par parties sur Y pour obtenir

$$\begin{aligned}&-\frac{1}{2}\int_{\mathcal{V}}\int_Y\theta_\xi^2 v\cdot(\psi\nabla\psi^*+\psi^*\nabla\psi)dy\,dv\\&+\int_{\mathcal{V}}\int_Y\psi^*\theta_\xi\left(\theta_\xi\int_{\mathcal{V}}\tilde\sigma\psi\,dv'-\int_{\mathcal{V}}\tilde\sigma\theta_\xi\psi\,dv'\right)dy\,dv\\&=-\int_{\mathcal{V}}\int_Y v\cdot\xi\,\psi\psi^*\theta_\xi\,dy\,dv=D^*\xi\cdot\xi\end{aligned}\tag{7.39}$$

à cause de la définition (7.35) de D^*. D'autre part, en soustrayant (7.25) multiplié par ψ à (7.24) multiplié par ψ^* on a

$$v\cdot(\psi\nabla\psi^*+\psi^*\nabla\psi)=\psi^*\int_{\mathcal{V}}\tilde\sigma\psi\,dv'-\psi\int_{\mathcal{V}}\tilde\sigma^*\psi^*\,dv'.$$

On en déduit que

$$\int_{\mathcal{V}}\int_Y\theta_\xi^2 v\cdot(\psi\nabla\psi^*+\psi^*\nabla\psi)dy\,dv=\int_{\mathcal{V}}\int_Y\theta_\xi^2\left(\psi^*\int_{\mathcal{V}}\tilde\sigma\psi\,dv'-\psi\int_{\mathcal{V}}\tilde\sigma^*\psi^*\,dv'\right)dy\,dv.$$

Or, en permutant l'ordre des intégrations en v et v' on a

$$\int_{\mathcal{V}}\int_Y\theta_\xi^2\psi\int_{\mathcal{V}}\tilde\sigma^*\psi^*\,dv'\,dy\,dv=\int_{\mathcal{V}}\int_Y\psi^*\int_{\mathcal{V}}\tilde\sigma\theta_\xi^2\psi\,dv\,dy\,dv'.$$

Au total l'équation (7.39) est donc équivalente à

$$\begin{aligned}&-\frac{1}{2}\int_{\mathcal{V}}\int_Y\theta_\xi^2\psi^*\left(\int_{\mathcal{V}}\tilde\sigma\psi\,dv'\right)dy\,dv+\frac{1}{2}\int_{\mathcal{V}}\int_Y\psi^*\left(\int_{\mathcal{V}}\tilde\sigma\theta_\xi^2\psi\,dv\right)dy\,dv'\\&+\int_{\mathcal{V}}\int_Y\psi^*\theta_\xi\left(\theta_\xi\int_{\mathcal{V}}\tilde\sigma\psi\,dv'-\int_{\mathcal{V}}\tilde\sigma\theta_\xi\psi\,dv'\right)dy\,dv=D^*\xi\cdot\xi\,,\end{aligned}$$

c'est-à-dire

$$\frac{1}{2}\int_{\mathcal{V}}\int_{\mathcal{V}}\int_Y\tilde\sigma\psi(v')\psi^*(v)\left(\theta_\xi(v)-\theta_\xi(v')\right)^2dv\,dy\,dv'=D^*\xi\cdot\xi.$$

On a donc $D^*\xi\cdot\xi\geq 0$. Montrons que cette inégalité est stricte. Si $D^*\xi\cdot\xi=0$ pour un vecteur $\xi\neq 0$, alors on en déduit que $\theta_\xi(v)$ est indépendant de v et en reportant dans l'équation (7.38) on obtient

$$v\cdot\nabla_y(\theta_\xi(y)+\xi\cdot y)=0\text{ dans }Y\times\mathcal{V}.$$

Comme v est quelconque et θ_ξ ne dépend pas de v, cela implique que $\theta_\xi(y)=-\xi\cdot y+C$ où C est une constante quelconque. Le caractère affine de θ_ξ contredit la condition aux limites de périodicité dans (7.38). Par conséquent, on doit avoir $D^*\xi\cdot\xi>0$. ■

7.3 Exercices

Exercice 7.1 *On considère une équation de diffusion avec convection et absorption*

$$\left\{ \begin{array}{ll} -\operatorname{div}\left(D\left(\frac{x}{\epsilon}\right)\nabla u_\epsilon\right) + V\left(\frac{x}{\epsilon}\right)\cdot\nabla u_\epsilon + \sigma\left(\frac{x}{\epsilon}\right)u_\epsilon = f & \text{dans } \Omega, \\ u_\epsilon = 0 & \text{sur } \partial\Omega, \end{array} \right.$$

où Ω est un ouvert borné de $\mathbf{R}^N$. On suppose que le tenseur D, la vitesse V et le coefficient d'absorption σ sont Y-périodiques, réguliers et que D est symétrique et coercif.

1. *Appliquer la méthode des développements asymptotiques à deux échelles pour l'homogénéisation de cette équation. Vérifier que le tenseur homogénéisé de diffusion D^* est donné par la formule "usuelle" (7.10) et que la vitesse homogénéisée V^* et le coefficient homogénéisé d'absorption σ^* sont des simples moyennes.*
2. *On se place en dimension $N = 1$. Calculer explicitement la solution du problème de cellule et montrer que*

$$D^\star = \left(\int_0^1 D^{-1}(y)\,dy\right)^{-1}.$$

3. *On se place maintenant en dimension $N = 2$ d'espace et on suppose que le coefficient de diffusion sur la cellule unité $Y = (0,1)^2$ est du type $D(y) = d(y)I$ avec*

$$d(y) = \left\{ \begin{array}{ll} d_1 & 0 < y_1 < 0.5, \\ d_2 & 0.5 < y_1 < 1. \end{array} \right.$$

Calculer explicitement les deux solutions des problèmes de cellule. Que vaut $D^\star$?

Indications : le cas mono-dimensionnel est un des très rares cas où on peut calculer explicitement des coefficients homogénéisés. Le cas de la question 3, appelé "structure laminée", est "quasi mono-dimensionnel" car $d(y)$ est indépendant de y_2. On se rappelera que les solutions des problèmes de cellule sont des solutions "faibles", ou variationnelles, ou au sens des distributions. Remarquons enfin que, pour la question 3, la matrice D^* n'est pas scalaire (ce qui correspond à un milieu effectif non-isotrope) alors même que la matrice D était scalaire (milieu isotrope).

Exercice 7.2 (Amortissement/amplification à moyenne nulle) *Dans ce problème (assez difficile) on va considérer l'équation instationnaire de diffusion avec un grand coefficient d'amortissement ou d'amplification*

$$\left\{ \begin{array}{ll} \dfrac{\partial u_\epsilon}{\partial t} + \dfrac{1}{\epsilon}\sigma\left(\dfrac{x}{\epsilon}\right)u_\epsilon - \operatorname{div}\left(D\left(\dfrac{x}{\epsilon}\right)\nabla u_\epsilon\right) = 0 & \text{dans } \Omega\times(0,T), \\ u_\epsilon = 0 & \text{sur } \partial\Omega\times(0,T), \\ u_\epsilon(x,0) = u_{in}(x) & \text{dans } \Omega, \end{array} \right.$$

où Ω est un ouvert borné de $\mathbf{R}^N$. On suppose que le tenseur D et le coefficient σ sont Y-périodiques, réguliers et que D est symétrique et coercif. On fait l'hypothèse supplémentaire

$$\int_Y \sigma(y)\, dy = 0. \tag{7.40}$$

1. *On étudie l'homogénéisation de cette équation à l'aide du développement asymptotique suivant*

$$u_\epsilon(t,x) = \sum_{i=0}^{+\infty} \epsilon^i u_i(t,x,\frac{x}{\epsilon})$$

où les fonctions $y \to u_i(t,x,y)$ sont Y-périodiques. Ecrire les équations satisfaites par u_0, u_1, et u_2 dans la cellule Y.

2. *En déduire que $u_0(t,x,y)$ ne dépend pas de y, et que $u_1(t,x,y)$ peut s'écrire comme*

$$u_1(t,x,y) = w_0(y)u_0(t,x) + \sum_{k=1}^{N} w_k(y)\frac{\partial u_0}{\partial x_k}(t,x)$$

où $(w_k)_{1\le k\le N}$ sont les solutions du problème de cellule habituel et w_0 est la solution d'un nouveau problème de cellule que l'on explicitera soigneusement. Montrer que (7.40) est nécessaire pour résoudre en w_0.

3. *Donner la condition de compatibilité nécessaire et suffisante pour pouvoir résoudre en $u_2(t,x,y)$. Montrer que l'équation homogénéisée est de la forme*

$$\frac{\partial u_0}{\partial t} + \sigma^* u_0 + V^* \cdot \nabla_x u_0 - \mathrm{div}_x\left(D^* \nabla_x u_0\right) = 0 \quad \text{dans } \Omega \times (0,T),$$

avec des formules précises pour σ^*, V^* et D^*. En utilisant les problèmes de cellule, montrer que $V^* = 0$ et $\sigma^* \le 0$ (ce qui montre en particulier qu'il n'y a pas de terme convectif dans l'équation homogénéisée).

Indications : pour calculer V^* et le signe de σ^* utiliser les formulations variationnelles pour w_0 et $(w_k)_{1\le k\le N}$.

Exercice 7.3 (Homogénéisation avec fort amortissement) *Nous étudierons dans cet exercice (difficile) l'équation instationnaire*

$$\begin{cases} \dfrac{\partial u_\epsilon}{\partial t} + \dfrac{1}{\epsilon^2}\sigma\left(\dfrac{x}{\epsilon}\right) u_\epsilon - \mathrm{div}\left(D\left(\dfrac{x}{\epsilon}\right)\nabla u_\epsilon\right) = 0 & \text{dans } \Omega \times (0,T), \\ u_\epsilon = 0 & \text{sur } \partial\Omega \times (0,T), \\ u_\epsilon(x,0) = u_{in}(x) & \text{dans } \Omega, \end{cases}$$

où Ω est un ouvert borné de $\mathbf{R}^N$. On suppose que le tenseur D et le coefficient σ sont Y-périodiques, réguliers et que D est symétrique et coercif.

1. *On étudie l'homogénéisation de cette équation à l'aide du développement asymptotique suivant*

$$u_\epsilon(t,x) = e^{-\frac{\lambda^* t}{\epsilon^2}} \sum_{i=0}^{+\infty} \epsilon^i u_i(t,x,\frac{x}{\epsilon})$$

où les fonctions $y \to u_i(t,x,y)$ sont Y-périodiques. Ecrire les équations satisfaites par u_0, u_1, et u_2 dans la cellule Y.

2. *En déduire que $u_0(t,x,y)$ est une fonction propre associée à la valeur propre λ^* d'un problème spectral de cellule que l'on déterminera. Désormais on admettra que λ^* est la première valeur propre de ce problème qui est simple et correspond à une fonction propre positive $\psi(y) > 0$ (par le théorème de Krein-Rutman ou Perron-Frobenius).*

3. *Pour $g(y) \in L^2(Y)$, on considère le problème*

$$\left\{ \begin{array}{l} \sigma(y)w - \mathrm{div}_y\left(D(y)\nabla_y w\right) - \lambda^* w = g \quad \textit{dans } Y, \\ y \to w(y) \ Y\textit{-périodique.} \end{array} \right. \tag{7.41}$$

Montrer que si $w \in H^1_\#(Y)$ est une solution, alors $w + C\psi$ est aussi une solution pour toute constante C. Montrer qu'une condition nécessaire pour pouvoir résoudre (7.41) est que

$$\int_Y g(y)\,\psi(y)\,dy = 0. \tag{7.42}$$

A partir de maintenant, on admet que (7.42) est une condition nécessaire et suffisante pour l'existence d'une solution $w \in H^1_\#(Y)$ de (7.41), qui est unique à addition près d'un multiple de ψ (c'est encore une alternative de Fredholm).

4. *Vérifier que le second membre de l'équation pour u_1 satisfait la condition de compatibilité (7.42) et montrer que*

$$u_1(t,x,y) = \sum_{k=1}^{N} z_k(y) \frac{\partial u}{\partial x_k}(t,x)$$

où $(z_k)_{1\le k\le N}$ sont des solutions de nouveaux problèmes de cellule à déterminer.

5. *Ecrire la condition de compatibilité nécessaire et suffisante pour résoudre en u_2. Montrer que l'équation homogénéisée est du type*

$$\frac{\partial u}{\partial t} - \mathrm{div}_x\left(D^*\nabla_x u\right) = 0 \quad \textit{dans } \Omega \times (0,T),$$

avec une expression explicite pour D^* en fonction des (z_k).

Indications : pour la question 3, multiplier l'équation (7.41) par ψ et intégrer par parties. On rappelle que l'espace $H^1_{\#}(Y)$ de fonctions périodiques est défini à la Remarque 7.1.2. De manière générale il faut suivre la méthode de la section 7.2.2, en plus simple, puisqu'il s'agit ici de diffusion et non de transport. En particulier le problème est auto-adjoint, autrement dit $\psi^* = \psi$.

Exercice 7.4 (Homogénéisation pour un problème de criticité) *On va étudier le problème de criticité pour calculer la première valeur propre λ_ϵ et la première fonction propre u_ϵ*

$$\begin{cases} \sigma\left(\dfrac{x}{\epsilon}\right) u_\epsilon - \operatorname{div}\left(D\left(\dfrac{x}{\epsilon}\right)\nabla u_\epsilon\right) = \lambda_\epsilon u_\epsilon & \text{dans } \Omega, \\ u_\epsilon = 0 & \text{sur } \partial\Omega, \end{cases}$$

où Ω est un ouvert borné de $\mathbf{R}^N$. On suppose que le tenseur D et le coefficient σ sont Y-périodiques, réguliers et que D est symétrique et coercif.

1. *On étudie l'homogénéisation de cette équation à l'aide du développement asymptotique suivant*

$$u_\epsilon(x) = \sum_{i=0}^{+\infty} \epsilon^i u_i(x, \frac{x}{\epsilon}) \quad \text{et} \quad \lambda_\epsilon = \sum_{i=0}^{+\infty} \epsilon^i \lambda_i$$

   où les fonctions $y \to u_i(x,y)$ sont Y-périodiques. Ecrire les équations satisfaites par u_0, u_1, et u_2 dans la cellule Y.

2. *En déduire que $u_0(x,y)$ ne dépend pas de y, et que $u_1(x,y)$ peut s'écrire comme*

$$u_1(x,y) = \sum_{k=1}^{N} w_k(y) \frac{\partial u_0}{\partial x_k}(x)$$

   où $(w_k)_{1\le k\le N}$ sont les solutions du problème de cellule habituel.

3. *Donner la condition de compatibilité nécessaire et suffisante pour pouvoir résoudre en $u_2(x,y)$. Montrer que l'équation homogénéisée est un problème de criticité où λ_0 est une valeur propre et u_0 une fonction propre correspondante. Donner un argument formel qui justifie que λ_0 est la première valeur propre du problème homogénéisé.*

Indication : pour la dernière question utiliser le théorème de Krein-Rutman ou Perron-Frobenius.

Exercice 7.5 (Modélisation physico-numérique) *(difficile) On va étudier dans cet exercice un calcul critique à un groupe de neutrons*

$$-\operatorname{div}(d(x)\nabla\phi(x)) + \sigma_a(x)\phi(x) = \frac{\nu\sigma_f(x)}{\lambda}\phi(x).$$

La structure en espace est hétérogène (en pratique 40 000 crayons "c" de combustibles)

$$\overline{\Omega} = \overline{\cup_c \Omega_c}$$

C'est pourquoi les fonctions d, σ_s et $s\sigma_f$ sont variables en espace. L'exercice qui suit est tiré de Méthodes mathématiques en neutronique *[43] (page 194) par J. Planchard.*

La méthode de factorisation consiste à écrire

$$\phi(x) = u(x)\psi(x)$$

dans lequel ψ est le flux neutronique obtenu par un calcul critique crayon par crayon (avec une condition de Neumann homogène au bord de chaque crayon). L'idée sous-jacente est que u varie peu (en espace) même si $d(x)$, $\sigma_a(x)$, $\sigma_f(x)$ et $\psi(x)$ varient beaucoup en espace.

1. *Ecrire les équations pour ψ en notant λ_c la valeur propre critique sur chaque crayon (c'est donc une fonction constante par morceaux sur chaque crayon). Expliquer pourquoi $\psi > 0$.*
2. *Montrer que l'équation réduite pour u s'écrit, dans chaque crayon,*

$$-\operatorname{div}(d\psi\nabla u) + \frac{\psi}{\lambda_c}\nu\sigma_f u - d\nabla\psi.\nabla u = \frac{\psi}{\lambda}\nu\sigma_f u.$$

3. *Montrer que, si on multiplie encore par ψ, l'équation réduite pour u devient*

$$-\operatorname{div}(d\psi^2\nabla u) + \frac{\psi^2}{\lambda_c}\nu\sigma_f u = \frac{\psi^2}{\lambda}\nu\sigma_f u.$$

   Quel est l'avantage de cette dernière équation sur la précédente ?

4. *Montrer que les conditions aux limites entre le crayon a et le crayon b s'écrivent à l'interface ($x \in \overline{\Omega_a} \cap \overline{\Omega_b}$)*

$$\psi_a(x)u_a(x) = \psi_b(x)u_b(x),$$

   et

$$d_a(x)\psi_a(x)(\partial_n u_a)(x) = d_b(x)\psi_b(x)(\partial_n u_b)(x).$$

5. *Proposer des formules $\sigma^\star$, $d^\star$ pour simplifier le problème pour u.*

Indications : la deuxième équation réduite pour u est plus avantageuse que la première car elle conduit à rechercher les valeurs propres d'un problème auto-adjoint, ce qui est beaucoup plus facile. D'autre part, on peut lui appliquer les formules de moyennes homogénéisées pour les coefficients obtenues à l'Exercice 7.4.

Bibliographie

[1] ABRAMOVITZ M., STEGUN I.A., *Handbook of mathemtical functions with formulas, graphs, and mathematical tables,* Number 55 in National Bureau of Standards applied mathemtical series, US Goverment printing office, Washngton DC, 1972.

[2] ALLAIRE G., *Analyse numérique et optimisation,* Editions de l'Ecole Polytechnique, Palaiseau (2012).

[3] ALLAIRE G., *Shape optimization by the homogenization method,* Springer Verlag, New York (2001).

[4] ALLAIRE G., BAL G., *Homogenization of the criticality spectral equation in neutron transport,* M2AN 33, pp.721-746 (1999).

[5] BARDOS C., SANTOS R., SENTIS R., *Diffusion approximation and computation of the critical size of a transport operator,* Trans. Amer. Math. Soc. 284 (1984), pp. 617–649.

[6] BASDEVANT J.-L., DALIBARD J., *Mécanique quantique,* Editions de l'Ecole Polytechnique, Palaiseau (2002).

[7] BENAIM M., EL KAROUI N., *Promenade aléatoire. Chaînes de Markov et simulations : martingales et stratégie,* Editions de l'Ecole Polytechnique, Palaiseau (2004).

[8] BENSOUSSAN A., LIONS J.L., PAPANICOLAOU G., *Asymptotic analysis for periodic structures,* North-Holland, Amsterdam (1978).

[9] BENSOUSSAN A., LIONS J.L., PAPANICOLAOU G., *Boundary layers and homogenization of transport processes,* Publ. R.I.M.S. Kyoto Univ. 15 (1979), pp. 53–115.

[10] BERMAN A., PLEMMONS R., *Nonnegative Matrices in the Mathematical Sciences,* Academic Press, New York (1979).

[11] BREZIS H., *Analyse fonctionnelle,* Masson, Paris (1983).

[12] BUSSAC J., REUSS P., *Traité de neutronique,* Hermann, Paris (1978).

[13] CASE K.M., ZWEIFEL P.F., *Linear transport theory* Addison Wesley, Reading (1967).

[14] CARLSON B., *The numerical theory of neutron transport,* dans Methods in computational Physics, vol. 1, Alder B. ed., pp.1-42, Academic Press, New York (1963).

[15] CHANDRASEKHAR S., *Radiative Transfer,* Oxford Univ. Press, London (1950).

[16] CIARLET P.G., *Introduction à l'analyse numérique matricielle et à l'optimisation,* Masson, Paris (1982).

[17] COCKBURN B., KARNIADAKIS G., SHU C.-W., *Discontinuous Galerkin methods,* (Newport, RI, 1999), Lecture Notes in Computational Science and Engineering, 11, Springer, Berlin (2000).

[18] COURANT R., HILBERT R., *Methods of mathematical physics,* John Wiley & Sons, New York (1989).

[19] DAUTRAY R., *Méthodes probabilistes pour les équations de la physique,* Eyrolles, Paris (1989).

[20] DAUTRAY R., LIONS J.-L., *Analyse mathématique et calcul numérique pour les sciences et les techniques,* Masson, Paris (1988).

[21] DENIZ V., *The theory of neutron leakage in reactor lattices,* in Handbook of nuclear reactor calculations, Y. Ronen ed., vol II, pp. 409–508 (1968).

[22] GKANTSIDIS D., MIHAIL M., ZEGURA E., *Spectral analysis of internet topologies,* IEEE Infocom (2003), http ://www.ieee-infocom.org/2003/papers/09_04.PDF.

[23] GOLSE F., *Distributions, analyse de Fourier, équations aux dérivées partielles,* Cours de 2ème année à l'Ecole Polytechnique (2010).

[24] GOLSE F., JIN S., LEVERMORE C.D., *The convergence of numerical transfer schemes in diffusive regimes I : discrete ordinate method,* SIAM J. Numer. Anal. 36 (1999), pp. 1333–1369.

[25] GOLSE F., JIN S., LEVERMORE C.D., *A domain decoposition analysis for a two-scale linear transport problem,* ESAIM M2AN Math. Model. and Numer. Anal. 37 (2003), pp. 869–892.

[26] GRAHAM C., TALAY D., *Simulations stochastiques et méthodes de Monte-Carlo,* Editions de l'Ecole Polytechnique, Palaiseau (2011).

[27] HILBERT D., *Begründung der kinetische Gastheorie,* Math. Annalen 72 (1912), pp. 562–577.

[28] ISTAS J., *Introduction aux modélisations mathématiques pour les sciences du vivant,* Collection Mathématiques & Applications, 34, Springer, Berlin (2000).

[29] JIKOV V., KOZLOV S., OLEINIK O., *Homogenization of differential operators and integral functionals,* Springer, Berlin, (1995).

[30] JOHNSON C., *Numerical solution of partial differential equations by the finite element method,* Cambridge University Press, Cambridge (1987).

[31] LAPEYRE B., PARDOUX E., SENTIS R., *Méthodes de Monte-Carlo pour les équations de transport et de diffusion,* Collection Mathématiques & Applications, 29, Springer, Berlin (1998).

[32] LARSEN E., KELLER J., *Asymptotic solution of neutron transport problems for small mean free paths*, J. Math. Phys., 15 (1974), pp. 75–81.

[33] LAUDENBACH F., *Calcul différentiel et intégral,* Editions de l'Ecole Polytechnique, Palaiseau (2005).

[34] LAX P., *Linear algebra,* John Wiley, New York (1997).

[35] LE GALL, J.-F., *Intégration, probabilités et processus aléatoires*, Cours à l'Ecole Normale Supérieure, 2006. (`http://www.dma.ens.fr/ legall/IPPA2.pdf`)

[36] LESAINT P., RAVIART P.-A., *On a finite element method for solving the neutron transport equation,* In : Mathematical aspects of finite elements in partial differential equations, pp. 89–123. Publication No. 33, Math. Res. Center, Univ. of Wisconsin-Madison, Academic Press, New York (1974).

[37] LEVERMORE C.D., POMRANING G.C., *A flux-limited diffusion theory,* Astrophys. J. 248 (1981), pp. 321–334.

[38] LEWIS E., MILLER W., *Computational methods of neutron transport*, Wiley, New York (1984).

[39] MELEARD S., *Modèles aléatoires en Ecologie et Evolution,* Cours de 3ème année à l'Ecole Polytechnique (2008).

[40] MIHALAS D., WEIBEL-MIHALAS B., *Foundations of radiation hydrodynamics,* Oxford Univ. Press, Oxford, New York (1984).

[41] NEDELEC J.-C., *Acoustic and electromagnetic equations : integral representations for harmonic problems,* Vol. 144. Springer, New York (2001).

[42] PERTHAME B., *Transport equations in biology,* Birkhäuser, Bâle (2007).

[43] PLANCHARD J., *Méthodes mathématiques en neutronique,* Collection de la Direction des Etudes et Recherches d'EDF, Eyrolles (1995).

[44] POMRANING G., *The equations of radiation hydrodynamics,* Pergamon Press, Oxford, New York (1973).

[45] REUSS P., *Précis de neutronique,* EDP Sciences, Collection génie atomique, Paris (2013).

[46] RUDIN W., *Real and Complex Analysis,* Mc Graw Hill, Singapore, 1986.

[47] SANCHEZ R., McCORMICK N., *A review of neutron transport approximations,* Nuclear Science and Engineering, 80, pp. 481-535 (1982).

[48] SANCHEZ-PALENCIA E., *Non homogeneous media and vibration theory,* Lecture notes in physics 127, Springer Verlag (1980).

[49] SENTIS R., *Study of the corrector of the eigenvalue of a transport operator,* SIAM J. Math. Anal. 16, pp. 151-166 (1985).

[50] SERRE D., *Les Matrices. Théorie et pratique,* Dunod, Paris (2001).

[51] SONNENDRUCKER E., *Approximation numérique des équations de Vlasov-Maxwell,* Notes de cours de Master M2, `http://www-irma.u-strasbg.fr/ sonnen/polyM2VM2010.pdf` (2010).

Index