



HAL
open science

Contributions à l'analyse numérique des méthodes quasi-Monte Carlo

Ibrahim Coulibaly

► **To cite this version:**

Ibrahim Coulibaly. Contributions à l'analyse numérique des méthodes quasi-Monte Carlo. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1997. Français. NNT: . tel-00004933

HAL Id: tel-00004933

<https://theses.hal.science/tel-00004933>

Submitted on 20 Feb 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

présentée par

IBRAHIM COULIBALY

pour obtenir le titre de

Docteur de l'Université Joseph Fourier - Grenoble 1

(Arrêtés ministériels du 5 Juillet 1984 et du 30 Mars 1992)

Spécialité:

Mathématiques Appliquées

Sujet de la thèse:

Contributions à l'analyse numérique des méthodes quasi-Monte Carlo

Soutenue le 3 novembre 1997 devant le jury composé de:

PIERRE BARAS

GEORGES-HENRI COTTET Président

HENRI FAURE

CHRISTIAN LÉCOT

A mes parents: Yayı et Papy.

A mes freres et sœurs: Ada, Bah, Méou, Ima, Papa, Matou et Assa.

Remerciements

Il me semble particulièrement important de citer en premier lieu dans ces remerciements mes parents qui m'ont toujours encouragé à aller le plus loin dans les études et m'ont donné les moyens de le faire.

Je tiens à exprimer toute ma reconnaissance à Christian Lécot qui a dirigé cette thèse. C'est lui qui a guidé mes premiers pas dans la recherche. En plus de ses qualités humaines, sa rigueur, sa bonne humeur, j'ai apprécié la liberté qu'il m'a accordée dans mes recherches; ses nombreux conseils m'ont été très utiles.

Je remercie Pierre Baras, Président de l'Université de Savoie qui a donné son agrément à la direction de ma thèse et qui me fait le grand honneur de participer à mon jury.

Henri Faure me fait l'honneur de rapporter sur cette thèse. Qu'il trouve ici l'expression de toute ma reconnaissance. Ses travaux sont une référence constante dans le domaine abordé ici.

Je remercie Ian Sloan qui a accepté de rapporter sur cette thèse. Ses remarques ont été très utiles pour l'élaboration de la version finale.

C'est un grand honneur pour moi d'avoir eu Georges-Henri Cottet comme président de mon jury de thèse. C'est lui qui m'a initié aux méthodes particulières.

J'exprime toute ma sympathie au LAMA pour m'avoir accueilli durant ma thèse.

Je ne saurais oublier dans mes remerciements le Centre Grenoblois de Calcul Vectoriel du Commissariat à l'Énergie Atomique qui m'a permis de faire mes essais numériques et les membres de l'équipe STAR qui m'ont souvent sorti de mes errements en informatique sur le CRAY C90 et le CRAY T3D.

Un grand merci à Séga, Mantala, Solo et aux frères chambériens.

Enfin, je remercie tous ceux qui ont contribué à ma formation en mathématiques.

Table des matières

Introduction	3
1 Les méthodes quasi-Monte Carlo	7
1.1 Introduction	7
1.2 Suites à faible discrédance	9
1.3 Réseaux- (t, m, s) et suites- (t, s)	13
1.4 Exemples de calcul approché d'intégrales	16
1.5 Estimations d'erreur	19
1.6 Quelques applications	22
2 Sur la résolution des EDO par des méthodes quasi-Monte Carlo	25
2.1 Introduction	25
2.2 Méthode de Runge-Kutta quasi-Monte Carlo	26
2.3 Analyse de l'erreur	27
2.4 Exemple numérique	31
2.5 Conclusion	32
3 Méthodes particulières pour l'équation de Boltzmann linéaire	33
3.1 Introduction	33
3.2 Algorithme quasi-Monte Carlo	34
3.3 Méthode de Nanbu	40
3.4 Bornes d'erreur pour la méthode quasi-Monte Carlo	41
3.5 Résultats numériques	47
3.6 Conclusion	59
4 Méthodes particulières de simulation du modèle de Kac	61
4.1 Introduction	61
4.2 Une méthode quasi-Monte Carlo pour la simulation de l'équation de Kac	62
4.3 Méthodes Monte Carlo pour la résolution de l'équation de Kac	66
4.4 Analyse d'erreur pour la méthode quasi-Monte Carlo de simulation	67
4.5 Études expérimentales	79
4.6 Conclusion	88

5	Simulation de la diffusion par des marches quasi-aléatoires	89
5.1	Introduction	89
5.2	Méthode des marches quasi-aléatoires	91
5.3	Convergence de la méthode	94
5.4	Exemples numériques	100
5.5	Conclusion	105
	Bibliographie	107

Introduction

Les méthodes de Monte Carlo sont des méthodes probabilistes basées sur l'utilisation de suites aléatoires. Elles sont généralement utilisées pour des problèmes que les méthodes classiques n'arrivent pas à résoudre. Par exemple les problèmes de calcul d'intégrales en grande dimension. Un autre exemple est celui de l'équation de Boltzmann. Des schémas de type différences finies ont été développés mais ces schémas se révèlent trop coûteux et les méthodes de type Monte Carlo sont les seules utilisées à l'heure actuelle dans la pratique. On a également recours aux méthodes de Monte Carlo pour des problèmes où il y a peu de régularité par exemple dans des calculs d'intégrales.

L'idée des méthodes quasi-Monte Carlo est de remplacer les suites aléatoires des méthodes de Monte Carlo par des suites quasi-aléatoires qui sont des suites déterministes également appelées suites à faible discrétance. Ces suites ont une meilleure répartition uniforme c'est-à-dire une discrétance plus faible que les suites aléatoires. Les méthodes quasi-Monte Carlo contrairement aux méthodes de Monte Carlo conduisent à des estimations d'erreur déterministes.

Le chapitre 1 est consacré aux outils des méthodes quasi-Monte Carlo et à leur utilisation pour l'intégration numérique. Des générateurs de suites à faible discrétance y sont présentés. Dans cette thèse on s'intéresse tout particulièrement à la famille des suites (t, s) et des réseaux (t, m, s) . Une majoration de l'erreur d'une quadrature quasi-Monte Carlo est fournie par l'inégalité de Koksma-Hlawka. Cette majoration dépend d'une part de la discrétance de la suite utilisée pour la quadrature et d'autre part de la variation au sens de Hardy-Krause de l'intégrand. Les dernières années ont vu une extension du domaine d'application des méthodes quasi-Monte Carlo. Le chapitre se termine par un tour d'horizon de ces applications.

Le chapitre 2 traite de la résolution de certains systèmes d'équations différentielles $y'(t) = f(t, y(t))$ avec condition initiale. Pour les méthodes numériques classiques de résolution d'un tel problème la fonction f doit être à la fois régulière en t et en y . Ici nous supposons une régularité en y et peu de régularité en t . La famille des méthodes RKMC (Runge-Kutta Monte Carlo) a été introduite par Stengle. Ces méthodes conduisent à formuler le problème avec un terme intégral multidimensionnel puis à effectuer une quadrature Monte Carlo. Dans [61] Stengle a analysé la méthode RKMC2 qui s'apparente à une méthode de Runge-Kutta d'ordre 2. Nous proposons une quadrature quasi-Monte Carlo pour ce problème à la place de la quadrature Monte Carlo. Les hypothèses en t sont un peu plus fortes que pour la quadrature Monte Carlo. Nous obtenons une estimation d'erreur qui dépend de la discrétance de la suite utilisée et du carré du maximum des pas

de temps. Les tests numériques montrent une amélioration spectaculaire des résultats par l'utilisation d'une quadrature quasi-Monte Carlo.

Dans les chapitres suivants on introduit des méthodes particulières quasi-Monte Carlo utilisant les réseaux- $(0, m, s)$ qui sont toutes construites sur le même principe. L'étape d'initialisation des particules consiste à approcher la distribution initiale par une somme de mesures de Dirac. Ces particules évoluent selon la dynamique décrite par le modèle. On fait un schéma d'Euler en temps. À chaque pas de temps on effectue une quadrature quasi-Monte Carlo et les particules sont renumérotées.

Au chapitre 3 on s'intéresse à la simulation de l'équation de Boltzmann linéaire en dimension quelconque par une méthode particulière quasi-Monte Carlo basée sur les réseaux- $(0, m, s)$. Ce type de méthodes a été développé par Lécot pour l'équation de Boltzmann isotrope en vitesse qui se ramène à un problème unidimensionnel. Le temps est discrétisé par un schéma d'Euler. À chaque pas de temps on effectue une quadrature quasi-Monte Carlo et les particules sont triées selon leurs vitesses. Ceci permet de démontrer la convergence de la méthode. Nous avons généralisé ce type de méthodes pour des problèmes en dimension quelconque. Nous utilisons une technique qui consiste à répartir les particules par paquets suivant les différentes directions. Il est ainsi possible de démontrer la convergence de la méthode. La borne d'erreur obtenue nous guide dans le choix des paramètres de la répartition par paquets. Ce choix de paramètres est très important dans la pratique puisqu'un mauvais choix peut conduire à des résultats catastrophiques. Les résultats expérimentaux indiquent que le nouveau schéma quasi-Monte Carlo est plus performant qu'un algorithme de Monte Carlo standard.

Au chapitre 4 le problème considéré est l'équation de Kac qui est un modèle cinétique unidimensionnel non-linéaire de l'équation de Boltzmann. C'est une équation qui conserve l'énergie cinétique mais pas la quantité de mouvement. On construit une méthode particulière qui conserve également l'énergie cinétique grâce à l'utilisation dans chaque quadrature de deux réseaux- $(0, m, 4)$ fabriqués par symétrie à partir d'un même réseau- $(0, m, 4)$. Une notion importante pour analyser l'erreur de la méthode est la propriété du quadrant généralisée. On évalue l'erreur provenant du calcul approché de la mesure d'un sous-intervalle de $[0, 1]^s$ ayant cette propriété. Ceci permet de démontrer la convergence. Les résultats expérimentaux montrent un gain de précision par rapport aux méthodes de simulation de type Monte Carlo.

Enfin au chapitre 5 nous présentons et analysons une méthode de simulation de la diffusion en dimension s quelconque par des marches quasi-aléatoires. Le point de départ est un travail de Morokoff et Caflish [38] qui se sont intéressés à la simulation de l'équation de la chaleur 1D avec des particules qui se déplacent sur une grille régulière. Ils ont posé le problème de l'extension de cette technique à la simulation de la diffusion en dimension quelconque. Cette extension est réalisée ici grâce à la technique de renumérotation des particules utilisée pour simuler l'équation de Boltzmann linéaire au chapitre 3. La méthode est construite à partir d'un schéma semi-discret en espace et à chaque pas de temps on effectue une quadrature quasi-Monte Carlo à l'aide d'un réseau- $(0, m, s + 1)$. La comparaison de la méthode des marches quasi-aléatoires avec la méthode des marches aléatoires a mis en évidence son efficacité par rapport à cette dernière.

Ces travaux ont été présentés dans différents colloques [8Γ35]Γet sont exposés dans plusieurs publications [9Γ10Γ11Γ36].

Chapitre 1

Les méthodes quasi-Monte Carlo

1.1 Introduction

Soit f une fonction réelle définie sur I^s où $I = [0, 1)$. On s'intéresse au calcul approché de l'intégrale s -dimensionnelle

$$E(f) = \int_{I^s} f(\mathbf{x}) d\mathbf{x}. \quad (1.1)$$

Pour $s = 1$ il existe des formules de quadrature bien connues (Simpson, trapèzes, rectangles, ...). L'utilisation de ces formules dans le cas multidimensionnel revient à faire s quadratures en dimension 1. Avec la formule des trapèzes par exemple, en faisant une quadrature à m points dans chaque dimension c'est à dire $N = m^s$ nœuds au total, nous obtenons une erreur en $O(N^{-2/s})$ (si f est régulière).

Le nombre de nœuds croît exponentiellement avec la dimension et la méthode atteint vite ses limites quand s augmente.

La méthode de *Monte Carlo* pour le calcul approché de l'intégrale s -dimensionnelle $E(f)$ consiste à faire la quadrature

$$\int_{I^s} f(\mathbf{x}) d\mathbf{x} \approx \frac{1}{N} \sum_{n=0}^{N-1} f(\mathbf{x}_n), \quad (1.2)$$

où $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N-1}$ sont N variables aléatoires indépendantes et de même distribution uniforme dans I^s . L'intégrale $E(f)$ est ainsi approchée par la valeur moyenne de la fonction f sur les points \mathbf{x}_n , $0 \leq n < N$.

La *loi forte des grands nombres* assure la convergence presque sûre de $\frac{1}{N} \sum_{n=0}^{N-1} f(\mathbf{x}_n)$ vers

$\int_{I^s} f(\mathbf{x}) d\mathbf{x}$ quand N tend vers l'infini.

Pour l'analyse probabiliste de l'erreur nous avons besoin de la variance de la fonction f définie par

$$\sigma^2(f) = \int_{I^s} f^2(\mathbf{x}) d\mathbf{x} - \left(\int_{I^s} f(\mathbf{x}) d\mathbf{x} \right)^2.$$

Cette variance n'existe que sous la condition $f \in L^2(I^s)$. L'espérance du carré de l'erreur vérifie pour tout entier $N \geq 1$ l'égalité

$$\int_{I^s \times \dots \times I^s} \left[\frac{1}{N} \sum_{n=0}^{N-1} f(\mathbf{x}_n) - E(f) \right]^2 d\mathbf{x}_0 \cdots d\mathbf{x}_{N-1} = \frac{\sigma^2(f)}{N}. \quad (1.3)$$

Ceci montre que l'erreur moyenne d'une quadrature Monte Carlo est $\sigma(f)N^{-\frac{1}{2}}$. Le *théorème de limite centrale* permet de retrouver la même estimation de l'erreur.

Ainsi l'erreur d'une quadrature Monte Carlo est indépendante de la dimension de l'intégrale. Un autre avantage de cette méthode est qu'elle demande peu de régularité sur la fonction à intégrer.

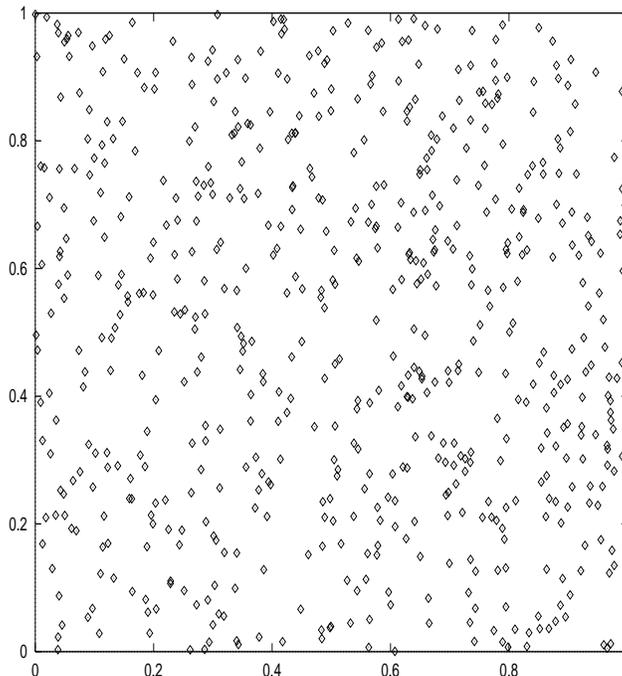


FIG. 1.1 – Suite pseudo-aléatoire de 625 points de I^2 calculée par NAG.

Le point crucial dans l'application d'une méthode de Monte Carlo est la génération de nombres aléatoires. De la qualité du générateur dépend en partie le succès de la méthode. Dans la pratique ces nombres sont générés de manière déterministe: on a des *nombres pseudo-aléatoires* (voir la figure 1.1).

La méthode de Monte Carlo comporte cependant quelques désavantages liés à son caractère stochastique. Elle ne fournit que des estimations d'erreurs probabilistes donc pas de résultats garantis. D'autre part dans les problèmes de simulation elle donne en général des résultats bruités avec une convergence assez lente.

Dans une quadrature Monte Carlo le caractère aléatoire des nombres (ou vecteurs) a moins d'importance que leur distribution uniforme. D'où l'idée de remplacer les suites

de nombres al eatoires par des suites de nombres d eterministes ayant une meilleure distribution. Ce sont les *suites quasi-al eatoires* ou *suites   faible discr epance* (voir la figure 1.2). Dans ce cas il est possible de trouver des estimations d'erreur d eterministes. C'est le principe fondamental des m ethodes de type *quasi-Monte Carlo* qui sont des versions d eterministes des m ethodes de Monte Carlo. Le principal but de ces m ethodes est de trouver de meilleures estimations d'erreur que les estimations d'erreur probabilistes des m ethodes de Monte Carlo et d'avoir une convergence plus rapide.

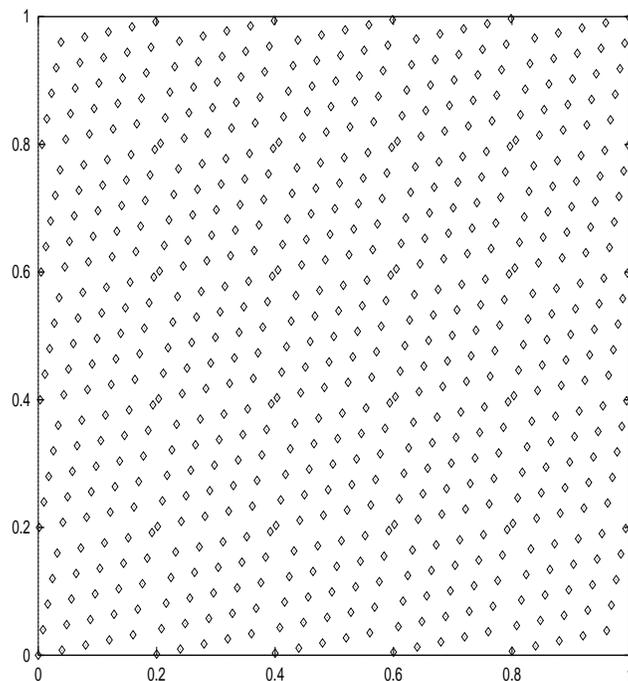


FIG. 1.2 – Suite quasi-al eatoire (suite de Hammersley) de 625 points de \overline{I}^2 .

Notons que le premier article publi e sur les m ethodes de Monte Carlo est celui de Metropolis et Ulam [37] en 1949. Ces m ethodes avaient  e utilis ees pendant plusieurs ann ees dans des projets secrets de la d efense am ericaine. Quant aux m ethodes quasi-Monte Carlo elles sont apparues pour la premi ere fois dans un rapport de Richtmyer [53] en 1951. Elles se sont par la suite consid erablement d evelopp ees (notamment dans les dix derni eres ann ees).

1.2 Suites   faible discr epance

1.2.1 La notion de discr epance

La discr epance est la d eviation par rapport   la distribution uniforme (ou en d'autres termes l'irr egularit e de la distribution).

Pour un ensemble $J \subseteq I^s$ et une suite finie $X = \{\mathbf{x}_n, 0 \leq n < N\}$ de N points de I^s

notons

$$D_N(J, X) = \frac{A(J, X)}{N} - \lambda_s(J), \quad (1.4)$$

la *discrédance locale* où $A(J, X)$ est le nombre d'indices n tels que $\mathbf{x}_n \in J$ et $\lambda_s(J)$ la mesure de Lebesgue (ou encore le volume) de J .

DÉFINITION 1.1 Si $X = \{\mathbf{x}_n, 0 \leq n < N\}$ est une suite de I^s , la *discrédance* $\mathbf{D}_N(X)$ de X est définie par

$$\mathbf{D}_N(X) = \sup_{J \in \mathcal{J}} |D_N(J, X)| \quad (1.5)$$

où \mathcal{J} est l'ensemble des parties de I^s de la forme $\prod_{i=1}^s [a_i, b_i)$.

Une autre notion de discrédance est la *discrédance-** ou *discrédance à l'origine*.

DÉFINITION 1.2 La *discrédance-** d'une suite $X = \{\mathbf{x}_n, 0 \leq n < N\}$ est définie par

$$\mathbf{D}_N^*(X) = \sup_{J \in \mathcal{J}^*} |D_N(J, X)| \quad (1.6)$$

\mathcal{J}^* étant l'ensemble des parties de I^s de la forme $\prod_{i=1}^s [0, b_i)$.

\mathbf{D}_N et \mathbf{D}_N^* sont des discrédances en norme L^∞ . Il existe également des *discrédances en norme L^2* [46].

DÉFINITION 1.3 La *discrédance L^2* d'une suite $X = \{\mathbf{x}_n, 0 \leq n < N\}$ est définie par

$$\mathbf{T}_N(X) = \left[\int_{J(\mathbf{a}, \mathbf{b}) \in I^{2s}, a_i < b_i} (D_N(J, X))^2 d\mathbf{a} d\mathbf{b} \right]^{\frac{1}{2}} \quad (1.7)$$

où $J = \prod_{i=1}^s [a_i, b_i)$ avec $\mathbf{a} = (a_1, a_2, \dots, a_s)$ et $\mathbf{b} = (b_1, b_2, \dots, b_s)$.

DÉFINITION 1.4 La *discrédance-** L^2 d'une suite $X = \{\mathbf{x}_n, 0 \leq n < N\}$ est définie par

$$\mathbf{T}_N^*(X) = \left[\int_{I^s} (D_N(J, X))^2 d\mathbf{b} \right]^{\frac{1}{2}} \quad (1.8)$$

avec $J = \prod_{i=1}^s [0, b_i)$ et $\mathbf{b} = (b_1, b_2, \dots, b_s)$.

Les relations entre \mathbf{D}_N et \mathbf{D}_N^* sont les suivantes:

$$\mathbf{D}_N^*(X) \leq \mathbf{D}_N(X) \leq 2^s \mathbf{D}_N^*(X).$$

Puisque la norme L^∞ est plus grande que la norme L^2 il vient

$$\mathbf{T}_N^*(X) \leq \mathbf{D}_N^*(X).$$

D'autre part la discr epance L^2 est li ee  a \mathbf{D}_N par

$$c_s (\mathbf{D}_N(X))^{\frac{s+2}{2}} \leq \mathbf{T}_N^*(X)$$

o u c_s est une constante positive d ependant de s .

Une suite infinie $X = \{\mathbf{x}_n, n \geq 0\}$ est uniform ement distribu ee si

$$\lim_{N \rightarrow \infty} \mathbf{D}_N(\{\mathbf{x}_n, 0 \leq n < N\}) = 0, \quad (1.9)$$

ou de fa con  equivalente

$$\lim_{N \rightarrow \infty} \mathbf{D}_N^*(\{\mathbf{x}_n, 0 \leq n < N\}) = 0. \quad (1.10)$$

En dimension un Γ nous avons le r esultat suivant d u  a Niederreiter [46].

TH EOR EME 1.1 Si $X = \{x_0, x_1, \dots, x_{N-1}\}$ avec $0 \leq x_0 \leq x_1 \leq \dots \leq x_{N-1} \leq 1$, alors

$$\mathbf{D}_N^*(X) = \frac{1}{2N} + \max_{0 \leq n < N} \left| x_n - \frac{2n+1}{2N} \right|, \quad (1.11)$$

et

$$(\mathbf{T}_N^*(X))^2 = \frac{1}{12N^2} + \frac{1}{N} \sum_{n=0}^{N-1} \left(x_n - \frac{2n+1}{2N} \right)^2. \quad (1.12)$$

Ce qui montre que la meilleure suite en dimension 1 est la suite des points milieux.

En dimension s quelconque Γ les calculs de $\mathbf{T}_N(X)$ et $\mathbf{T}_N^*(X)$ pour une suite $X = \{\mathbf{x}_n, 0 \leq n < N\}$ peuvent se faire par des formules explicites:

$$\begin{aligned} (\mathbf{T}_N(X))^2 &= \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \prod_{i=1}^s (1 - \max(x_{n,i}, x_{m,i})) \min(x_{n,i}, x_{m,i}) \\ &\quad - \frac{2^{1-s}}{N} \sum_{n=0}^{N-1} \prod_{i=1}^s x_{n,i} (1 - x_{n,i}) + 12^{-s}, \end{aligned} \quad (1.13)$$

et

$$(\mathbf{T}_N^*(X))^2 = \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \prod_{i=1}^s (1 - \max(x_{n,i}, x_{m,i})) - \frac{2^{1-s}}{N} \sum_{n=0}^{N-1} \prod_{i=1}^s (1 - (x_{n,i})^2) + 3^{-s}. \quad (1.14)$$

Diff erentes suites  a faible discr epance ont  et e  etudi ees d'un point de vue pratique par Morokoff et Caffisch [39]. Nous donnons quelques exemples classiques de suites  a faible discr epance. Une autre famille de suites  a faible discr epance Γ celle des suites (t, s) Γ fait l'objet d'un traitement  a part dans le prochain paragraphe.

1.2.2 Suite $\{n \cdot \alpha\}$

Soit $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s) \in \mathbb{R}^s$ tel que $1, \alpha_1, \alpha_2, \dots, \alpha_s$ soient linéairement indépendants sur \mathbb{Q} .

Les suites $\{n \cdot \alpha\}$ sont données sous la forme:

$$\mathbf{x}_n = (\{n \cdot \alpha_1\}, \{n \cdot \alpha_2\}, \dots, \{n \cdot \alpha_s\}), \quad n \geq 0,$$

où $\{x\}$ est la partie fractionnaire du nombre x . On peut choisir par exemple

$$\alpha = (\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_s}),$$

les p_i , $1 \leq i \leq s$ étant les s premiers entiers premiers.

Pour tout $\varepsilon > 0$ la discrédance d'une suite $\{n \cdot \alpha\}$ est en $O\left(\frac{(\log N)^{s+1+\varepsilon}}{N}\right)$ pour presque tout $\alpha \in \mathbb{R}^s$. Aucun α fournissant une discrédance en $O\left(\frac{(\log N)^{s+1}}{N}\right)$ n'est connu pour $s \geq 2$. Si les coordonnées de α sont des nombres algébriques alors la discrédance est en $O(N^{-1+\varepsilon})$ pour tout $\varepsilon > 0$ d'après un résultat de Niederreiter [44].

1.2.3 Suites de Van der Corput et de Halton

Etant donné un entier premier $b \geq 2$ on pose $F_b = \{0, 1, 2, \dots, b-1\}$. Tout entier positif n peut s'écrire de manière unique en base b :

$$n = \sum_{k=0}^{\infty} a_k(n+1)b^k, \quad (1.15)$$

où $a_k(n+1) \in F_b$ pour tout $k \geq 0$ et $a_k(n+1) = 0$ pour k suffisamment grand. C'est à dire que la somme est finie.

DÉFINITION 1.5 La fonction ϕ_b définie par

$$\phi_b(n) = \sum_{k=0}^{\infty} a_k(n+1)b^{-k-1}, \quad n \geq 0 \quad (1.16)$$

est appelée fonction radicale inverse.

La suite de Van der Corput [64] est donnée par

$$x_n = \phi_2(n), \quad n \geq 0. \quad (1.17)$$

La suite de Halton [16] est une généralisation de la suite de Van der Corput. Pour $s \geq 1$ on considère b_1, \dots, b_s des entiers premiers.

La suite de Halton dans les bases b_1, \dots, b_s est la suite infinie $\mathbf{x}_0, \mathbf{x}_1, \dots$ avec

$$\mathbf{x}_n = (\phi_{b_1}(n), \dots, \phi_{b_s}(n)) \in I^s, \quad \text{pour tout } n \geq 0. \quad (1.18)$$

La discrédance-* des N premiers termes d'une suite $X = \{\mathbf{x}_n, n \geq 0\}$ de Halton vérifie l'inégalité

$$\mathbf{D}_N^*(X) \leq C_s \frac{(\log N)^s}{N} + O\left(\frac{(\log N)^{s-1}}{N}\right), \quad (1.19)$$

où

$$C_s = \prod_{k=1}^s \frac{b_k - 1}{2 \log b_k}.$$

Il existe un algorithme récursif pour calculer cette suite [15Γ17]. Un algorithme vectorisable de calcul de la suite de Halton a été proposé par Lécot [32].

Dans la même famille il y a la *suite de Hammersley* [18] qui est une suite finie. Elle est définie pour $s \geq 2$ et $N \geq 1$ fixés par

$$\mathbf{x}_n = \left(\frac{n}{N}, \phi_{b_1}(n), \dots, \phi_{b_{s-1}}(n) \right) \in I^s, \quad \text{pour } n = 0, 1, \dots, N-1. \quad (1.20)$$

La discrédance-* d'une suite $X = \{\mathbf{x}_n, n = 0, 1, \dots, N-1\}$ de Hammersley en dimension s vérifie

$$\mathbf{D}_N^*(X) \leq C_{s-1} \frac{(\log N)^{s-1}}{N} + O\left(\frac{(\log N)^{s-2}}{N}\right), \quad (1.21)$$

avec C_{s-1} définie comme ci-dessus.

1.3 Réseaux- (t, m, s) et suites- (t, s)

1.3.1 Définitions

On rappelle quelques notions relatives aux suites- (t, s) .

DÉFINITION 1.6 *Pour une dimension s fixée et un entier $b \geq 2$, on appelle intervalle élémentaire en base b tout sous-intervalle E de I^s de la forme*

$$E = \prod_{i=1}^s \left[\frac{a_i}{b^{d_i}}, \frac{a_i + 1}{b^{d_i}} \right) \quad (1.22)$$

avec $a_i, d_i \in \mathbb{N}$ et $0 \leq a_i < b^{d_i}$ pour $1 \leq i \leq s$.

DÉFINITION 1.7 *Soit t et m deux entiers avec $0 \leq t \leq m$. Un réseau- (t, m, s) en base b est un ensemble X de b^m points de I^s tel que $A(E; X) = b^t$ pour tout intervalle élémentaire E en base b avec $\lambda_s(E) = b^{t-m}$.*

DÉFINITION 1.8 *Soit t un entier positif. Une suite $\mathbf{x}_0, \mathbf{x}_1, \dots$ de points de I^s est une suite- (t, s) en base b si, pour tous les entiers $k \geq 0$ et $m > t$, $\{\mathbf{x}_n, kb^m \leq n < (k+1)b^m\}$ est un réseau- (t, m, s) en base b .*

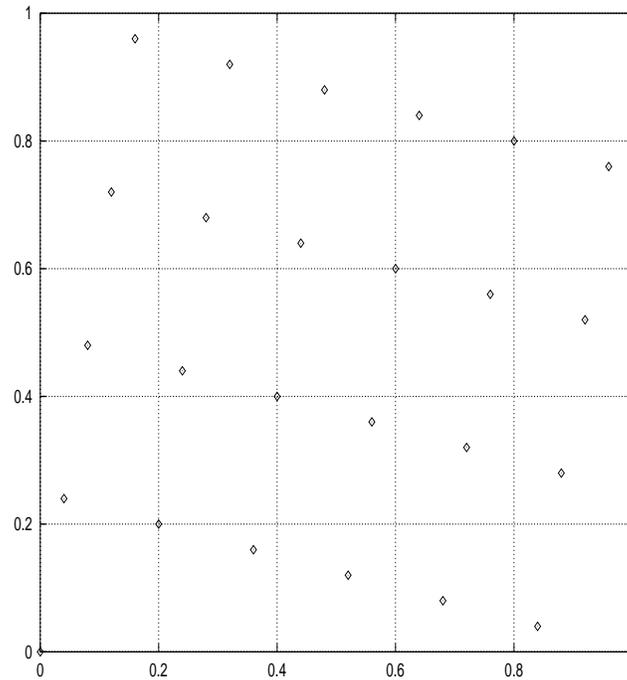


FIG. 1.3 – Un réseau-(0, 2, 2) en base $b = 5$.

Voir figure 1.3

Comme exemples de telles suites nous avons la suite de Sobol' [59] qui est une suite (t, s) en base $b = 2$ et la suite de Faure [14] qui est une suite $(0, s)$ en base b où b est un entier premier.

Dans la suite nous utiliserons surtout cette dernière qui est bien adaptée à la plupart des problèmes que nous étudierons.

Niederreiter [47] a développé une méthode de génération de suites (t, s) . Plus récemment Tezuka [63] a proposé des variantes aux suites de Faure et Niederreiter.

Une étude détaillée des propriétés ainsi que de la discrédance des réseaux (t, m, s) et des suites (t, s) est faite par Niederreiter dans [49 Chapitre 4].

1.3.2 Suite de Sobol'

La suite de Sobol' est une suite (t, s) en base $b = 2$. En dimension 1 la suite de Sobol' est calculée de la manière suivante.

Pour $k = 1, 2, \dots, w$ soit $v_k = \frac{m_k}{2^k}$ une suite de fractions binaires où $0 < m_k < 2^k$ sont des entiers impairs. Les nombres v_k sont appelés *nombre directionnels*. Alors

$$x_n = b_1 v_1 \oplus b_2 v_2 \oplus \dots \oplus b_w v_w, \quad n \geq 0, \quad (1.23)$$

où

$$n = \sum_{k=0}^{\infty} b_k 2^k$$

est la représentation binaire de n et \oplus l'opération *ou exclusif* bite par bite.

Par exemple $\Gamma n = 3$ s'écrit 11 en base 2. Si $v_1 = 0, 1$ et $v_2 = 0, 11$ alors

$$b_1 v_1 \oplus b_2 v_2 = 0, 1 \oplus 0, 11 = 0, 01.$$

La suite des nombres directionnels est calculée à l'aide d'un polynôme dont les coefficients sont dans $\{0, 1\}$ et qui est un *polynôme primitif* [27] dans le corps Z_2 . Considérons par exemple le polynôme primitif

$$P(x) = x^d + a_1 x^{d-1} + \dots + a_{d-1} x + 1$$

de degré d dans Z_2 . Alors les nombres directionnels sont obtenus par la formule de récurrence suivante:

$$v_k = a_1 v_{k-1} \oplus a_2 v_{k-2} \oplus \dots \oplus a_{d-1} v_{k-d+1} \oplus v_{k-d} \oplus \frac{v_{k-d}}{2^d}, \quad k > d, \quad (1.24)$$

où le dernier terme v_{k-d} est déplacé d fois à droite. Considérons maintenant une dimension $s \geq 1$ quelconque. Soit P_1, P_2, \dots, P_s Γs polynômes primitifs dans Z_2 distincts.

Notons $x_n^{(j)}$, $n \geq 0$ la suite unidimensionnelle de Sobol' calculée à partir du polynôme P_j . Alors la suite s -dimensionnelle de Sobol' est définie par

$$\mathbf{x}_n = \left(x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(s)} \right) \in I^s, \quad \text{pour tout } n \geq 0. \quad (1.25)$$

La discrédance d'une suite X de Sobol' vérifie l'inégalité

$$\mathbf{D}_N^*(X) \leq C_s \frac{(\log N)^s}{N} + O\left(\frac{(\log N)^{s-1}}{N}\right), \quad (1.26)$$

où

$$C_s = \frac{2^{t(s)}}{s!(\log 2)^s},$$

avec $t(s)$ vérifiant

$$k \frac{s \log s}{\log \log s} \leq t(s) \leq \frac{s \log s}{\log 2} + O(s \log \log s),$$

et $k > 0$.

Antonov et Saleev [1] ont suggéré une modification de l'algorithme original de Sobol' qui permet un calcul rapide de la suite en changeant l'ordre des termes Γs sans affecter ses bonnes propriétés.

Une implémentation sur ordinateur a été proposée par Bratley et Fox [3].

1.3.3 Suite de Faure

Soit b un entier premier supérieur ou égal à s (la dimension). La suite de Faure est une suite-($0, s$) en base b . Soit $F_b = \{0, 1, 2, \dots, b-1\}$.

Si n est un entier strictement positif Γ soit

$$n = \sum_{k=0}^{\infty} a_k(n+1)b^k, \quad a_k(n+1) \in F_b, \quad (1.27)$$

l'écriture de n en base b . Les $a_k(n+1)$ sont tous nuls sauf un nombre fini.

Notons b_1, b_2, \dots, b_s les éléments distincts de F_b .

Soit

$$x_n^{(i)} = \sum_{k=1}^{\infty} x_{n,k}^{(i)} b^{-k}, \quad n \geq 0, \quad 1 \leq i \leq s, \quad (1.28)$$

avec

$$x_{n,k}^{(i)} = \sum_{l=k-1}^{\infty} \binom{l}{k-1} b_i^{l-k+1} a_l(n+1) \pmod{b}. \quad (1.29)$$

La suite de Faure est définie par:

$$\mathbf{x}_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(s)}), \quad n \geq 0.$$

Si X est une suite de Faure en base b alors

$$\mathbf{D}_N^*(X) \leq C_s \frac{(\log N)^s}{N} + O\left(\frac{(\log N)^{s-1}}{N}\right), \quad (1.30)$$

avec

$$C_2 = \frac{3}{16(\log 2)^2},$$

et

$$C_s = \frac{1}{s!} \left(\frac{b-1}{2 \log b}\right)^s \quad \text{si } s \geq 3.$$

Une algorithmme de calcul de cette suite a été donné par Fox [15]. Dans [33] Lécot propose un algorithmme vectorisable de calcul de la suite de Faure.

REMARQUE 1.1 La constante C_s a un excellent comportement asymptotique pour la suite de Faure. Dans ce cas elle tend vers zéro quand la dimension s tend vers l'infini.

Ce n'est pas le cas pour les suites de Halton et Sobol' où l'on a asymptotiquement respectivement $\log C_s = O(s \log s)$ et $\log C_s = O(s \log \log s)$. Dans les deux cas la constante C_s a une croissance super-exponentielle. Elle croit plus vite pour la suite de Halton que pour la suite de Sobol' comme le confirment les tests effectués par Faure [14].

1.4 Exemples de calcul approché d'intégrales

Nous testons les suites décrites dans les deux paragraphes précédents à savoir les suites $\{n \cdot \alpha\}$ de Halton et Sobol' et Faure. Une quadrature Monte Carlo est également effectuée sur les mêmes exemples avec des suites pseudo-aléatoires données par NAG.

Les tests sont effectués sur les deux intégrales qui suivent:

$$K = \int_{I^3} \frac{|x_1 - x_2|}{1 + x_2 x_3} d\mathbf{x} \quad (1.31)$$

et

$$L_s = \frac{1}{s} \int_{I^s} \sum_{i=1}^s |4x_i - 2| d\mathbf{x}, \quad s \geq 1. \quad (1.32)$$

Notons que $K = \frac{5}{4} - 2 \log(2) + \frac{\pi^2}{24}$ et $L_s = 1$ pour tout $s \geq 1$.

Pour l'intégrale L_s deux cas sont considérés à savoir $s = 7$ et $s = 15$. Les suites $\{n \cdot \alpha\}$ ont été calculées avec $\alpha = (\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_s})$ les p_i , $1 \leq i \leq s$ étant les s premiers entiers premiers. Les suites de Halton ont été également construites à partir des nombres premiers consécutifs. Les suites de Faure ont été calculées pour la plus petite base b supérieure ou égale à la dimension de l'intégrale.

Le nombre de points pour chaque quadrature est de la forme b^m $m \geq 1$. Les quadratures dans le cas des suites de Faure ont ainsi été réalisées avec des réseaux $-(0, m, s)$ en base b .

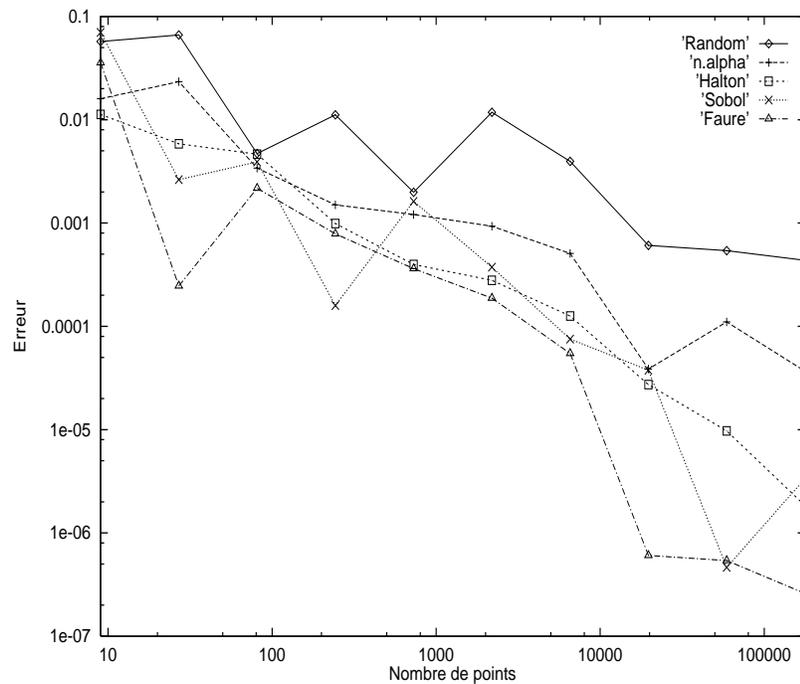


FIG. 1.4 – Erreurs dans le cas de l'intégrale K .

Les trois exemples (voir les figures 1.4Γ1.5 et 1.6) montrent de meilleurs résultats pour les quadratures quasi-Monte Carlo que pour les quadratures Monte Carlo. D'autre part l'utilisation de réseaux $-(0, m, s)$ dans le cas des suites de Faure donne des résultats très intéressants. Ces réseaux jouent un rôle essentiel dans cette thèse. La plupart des estimations d'erreur dans les problèmes que nous avons traités utilisent directement les propriétés des réseaux.

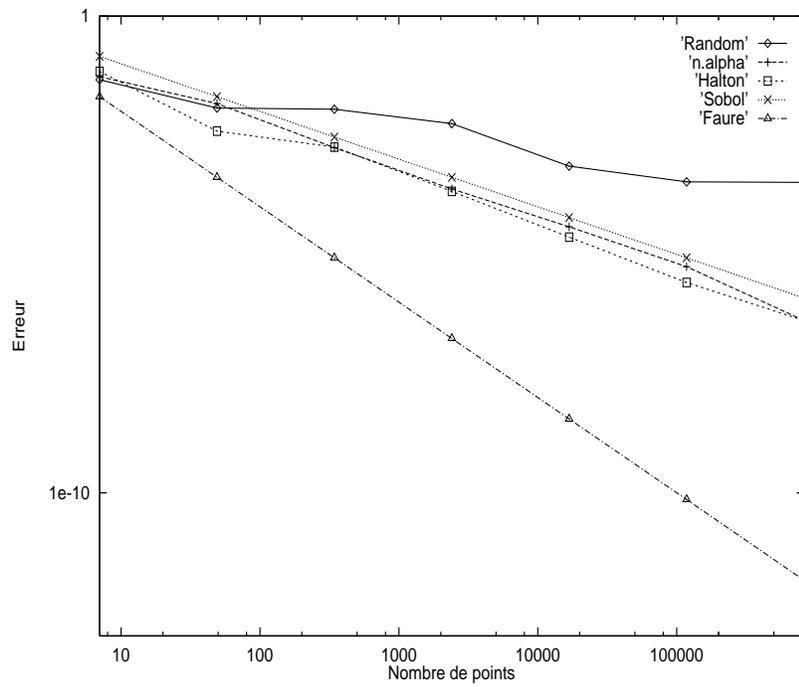


FIG. 1.5 – Erreurs dans le cas de l'intégrale L_7 ($s = 7$).

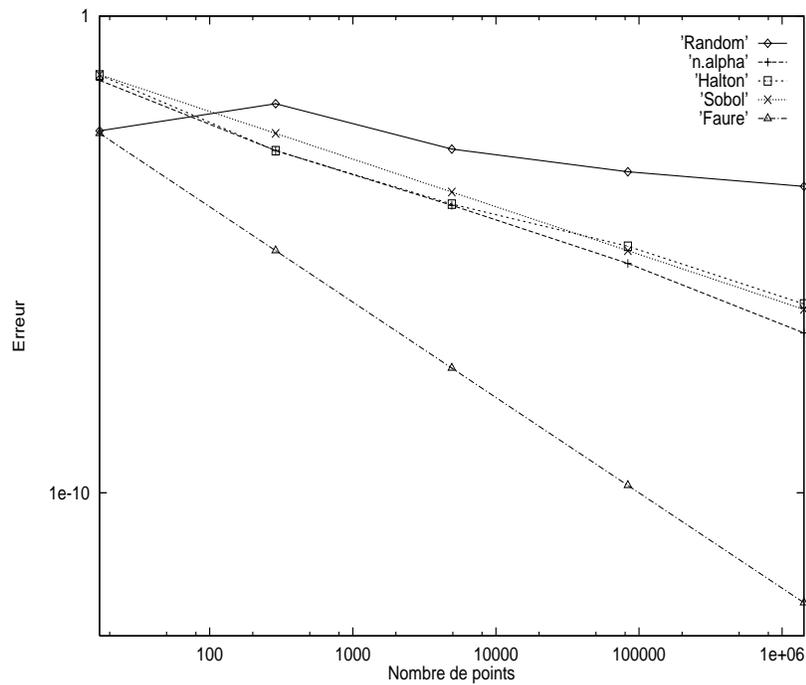


FIG. 1.6 – Erreurs dans le cas de l'intégrale L_{15} ($s = 15$).

1.5 Estimations d'erreur

Dans les estimations d'erreurs pour les méthodes quasi-Monte Carlo intervient souvent la notion de *fonction à variation bornée* au sens de Hardy-Krause que nous allons définir. En dimension un cette notion coïncide avec la notion classique de fonction à variation bornée.

Soit f une fonction définie sur \bar{I}^s et J une partie de \bar{I}^s de la forme

$$J = [a_1^{(1)}, a_2^{(1)}] \times \dots \times [a_1^{(s)}, a_2^{(s)}].$$

Notons

$$\Delta(f; J) = \sum_{l_1=1}^2 \dots \sum_{l_s=1}^2 (-1)^{l_1+\dots+l_s} f(a_{l_1}^{(1)}, \dots, a_{l_s}^{(s)}).$$

Si

$$V^{(s)}(f) = \sup_{\mathcal{P}} \sum_{J \in \mathcal{P}} |\Delta(f; J)| < \infty,$$

le sup étant pris sur toutes les partitions \mathcal{P} de \bar{I}^s alors la fonction f est dite à *variation bornée au sens de Vitali*.

Dans le cas d'une fonction suffisamment régulière $V^{(s)}(f)$ s'écrit

$$V^{(s)}(f) = \int_{I^s} \left| \frac{\partial^s f}{\partial x_1 \dots \partial x_s} \right| dx_1 \dots dx_s.$$

Pour $1 \leq k \leq s$ et $1 \leq i_1 < i_2 < \dots < i_k \leq s$ soit $V^{(k)}(f; i_1, i_2, \dots, i_k)$ la variation k -dimensionnelle au sens de Vitali de la restriction de f à

$$I_{i_1, \dots, i_k}^s = \left\{ (t_1, \dots, t_s) \in \bar{I}^s : t_j = 1 \text{ pour } j \neq i_1, \dots, i_k \right\}.$$

Si pour tout entier k $1 \leq k \leq s$ toutes les variations k -dimensionnelles sont finies la fonction f est dite à *variation bornée au sens de Hardy-Krause*.

Alors

$$V(f) = \sum_{k=1}^s \sum_{1 \leq i_1 < \dots < i_k \leq s} V^{(k)}(f; i_1, i_2, \dots, i_k) \quad (1.33)$$

est appelée la variation de f sur \bar{I}^s au sens de Hardy-Krause.

Un résultat fondamental de l'analyse de l'erreur des méthodes quasi-Monte Carlo est le théorème de Koksma-Hlawka.

THÉORÈME 1.2 *Si f est une fonction à variation bornée au sens de Hardy-Krause sur \bar{I}^s , alors pour tout ensemble $X = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N-1}\}$ de points de I^s , on a*

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} f(\mathbf{x}_n) - \int_{I^s} f(\mathbf{x}) d\mathbf{x} \right| \leq \mathbf{D}_N^*(X) V(f). \quad (1.34)$$

L'estimation (1.34) a été établie par Koksma [28] en dimension 1. L'extension aux fonctions de plusieurs variables est due à Hlawka [19].

L'inégalité de Koksma-Hlawka est en général la meilleure possible même pour des fonctions C^∞ comme le montre le théorème suivant dû à Niederreiter [49].

THÉORÈME 1.3 *Pour tout ensemble $X = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N-1}\}$ de points de I^s et tout réel $\varepsilon > 0$, il existe une fonction $f \in C^\infty(\bar{T}^s)$ avec $V(f) = 1$ et*

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} f(\mathbf{x}_n) - \int_{I^s} f(\mathbf{x}) d\mathbf{x} \right| > \mathbf{D}_N^*(X) - \varepsilon. \quad (1.35)$$

Pour une fonction f continue sur \bar{T}^s on définit son *module de continuité* par

$$\omega(f; t) = \sup_{u, v \in \bar{T}^s, \|u-v\| \leq t} |f(u) - f(v)|, \quad \text{pour } t \geq 0, \quad (1.36)$$

où $\|u\| = \max_{1 \leq i \leq s} |u_i|$ pour $u = (u_1, \dots, u_s) \in \mathbb{R}^s$.

Le résultat suivant a été établi par Niederreiter [44] en dimension 1 avant d'être généralisé par Proinov [52].

THÉORÈME 1.4 *Si f est une fonction continue sur \bar{T}^s , alors pour tout ensemble $X = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N-1}\}$ de points de \bar{T}^s , on a*

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} f(\mathbf{x}_n) - \int_{I^s} f(\mathbf{x}) d\mathbf{x} \right| \leq c_s \omega\left(f; \mathbf{D}_N^*(X)^{\frac{1}{s}}\right), \quad (1.37)$$

avec $c_1 = 1$ et $c_s = 4$ pour tout $s \geq 2$.

Woźniakowski [65] a démontré que l'erreur d'intégration moyenne est égale à la discrédance L^2 de la suite utilisée dans la quadrature.

THÉORÈME 1.5 *Soit C_s l'ensemble des fonctions continues sur I^s muni de la mesure de Wiener w . Alors pour toute suite $X = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N-1}\}$ de points de \bar{T}^s , on a*

$$\int_{C_s} \left[\frac{1}{N} \sum_{n=0}^{N-1} f(\mathbf{x}_n) - E(f) \right]^2 w(df) = (\mathbf{T}_N(X'))^2, \quad (1.38)$$

où $X' = \{\mathbf{x}'_0, \mathbf{x}'_1, \dots, \mathbf{x}'_{N-1}\}$ avec $\mathbf{x}'_n = (1 - x_n^1, 1 - x_n^2, \dots, 1 - x_n^s)$, $0 \leq n < N$.

Soit f une fonction de \bar{T}^{s-1} dans \bar{T} à variation bornée au sens de Hardy-Krause et E_f l'ensemble

$$E_f = \{\mathbf{x} = (\mathbf{x}', x_s) \in I^s : x_s < f(\mathbf{x}')\}.$$

Pour tout ensemble X de N points $\mathbf{x}_0, \dots, \mathbf{x}_{N-1}$ de I^s nous avons

$$\left| \frac{A(E_f, X)}{N} - \lambda_s(E_f) \right| = \left| \frac{1}{N} \sum_{n=0}^{N-1} c_{E_f}(\mathbf{x}_n) - \int_{I^s} c_{E_f}(\mathbf{x}) d\mathbf{x} \right|, \quad (1.39)$$

où c_{E_f} est la fonction caractéristique de E_f .

L'inégalité de Koksma-Hlawka ne peut pas s'appliquer ici car la fonction c_{E_f} n'est en général pas à variation bornée au sens de Hardy-Krause même si la fonction f est très régulière (par exemple $f(\mathbf{x}') = x_1$). On rencontre souvent des problèmes où il faut trouver une estimation pour (1.39). Des résultats pour de tels problèmes ont été établis par Lécot [34].

THÉORÈME 1.6 *Soit X un ensemble de N points de I^s et f une fonction de \overline{T}^{s-1} dans \overline{T} à variation bornée au sens de Hardy-Krause. Si $\mathbf{D}_N(X) \leq V(f)$, alors*

$$\left| \frac{A(E_f, X)}{N} - \lambda_s(E_f) \right| \leq sV(f) \left[\left(\frac{V(f)}{\mathbf{D}_N(X)} \right)^{\frac{1}{s}} \right]^{-1} \quad (1.40)$$

Dans le cas où on utilise un réseau- (t, m, s) nous avons le résultat suivant:

THÉORÈME 1.7 *Soit X un réseau- (t, m, s) en base b , et f une fonction de \overline{T}^{s-1} dans \overline{T} à variation bornée au sens de Hardy-Krause. Si $b^{t-m} \leq V(f)$, alors*

$$\left| \frac{A(E_f, X)}{b^m} - \lambda_s(E_f) \right| \leq sV(f)b^{-\lfloor \frac{m-t}{s} + \frac{\log V(f)}{s \log b} \rfloor}. \quad (1.41)$$

Un théorème plus général est celui de Niederreiter-Wills [45]:

THÉORÈME 1.8 *Soit E un sous espace mesurable de I^s . Pour $\varepsilon > 0$, soit*

$$E_\varepsilon = \{ \mathbf{x} \in I^s : \exists \mathbf{x}' \in E, \| \mathbf{x} - \mathbf{x}' \| < \varepsilon \},$$

et

$$E_{-\varepsilon} = \{ \mathbf{x} \in I^s : \forall \mathbf{x}' \in I^s \setminus E, \| \mathbf{x} - \mathbf{x}' \| \geq \varepsilon \},$$

où $\| \cdot \|$ est la norme euclidienne de \mathbb{R}^s . Si

$$\exists K > 0, \forall \varepsilon > 0, \max(\lambda_s(E_\varepsilon \setminus E), \lambda_s(E \setminus E_{-\varepsilon})) \leq K\varepsilon,$$

alors pour toute suite $X = \{ \mathbf{x}_i, i = 0, \dots, N-1 \}$ de points de I^s ,

$$\left| \frac{A(E, X)}{N} - \lambda_s(E) \right| \leq (4Ks^{\frac{1}{2}} + 2K + 1)\mathbf{D}_N(X)^{\frac{1}{s}}. \quad (1.42)$$

REMARQUE 1.2 Dans les quadratures quasi-Monte Carlo avec les suites à faible discrétion On a souvent utilisé l'hypothèse que la fonction est à variation bornée au sens de Hardy-Krause. Dans le cas d'une fonction plus régulière il n'y a pas d'amélioration de l'estimation d'erreur.

Il existe une autre technique de quadrature qui utilise les *bons treillis* [49Γ58]. Cette méthode s'applique dans le cas où la fonction que l'on intègre est périodique. Ce qui permet d'utiliser l'analyse de Fourier et on obtient des estimations d'erreur qui dépendent de la régularité de la fonction.

1.6 Quelques applications

Dans les paragraphes précédents nous avons surtout considéré les problèmes d'intégration numérique. Les méthodes quasi-Monte Carlo constituent en effet de bons outils pour l'intégration numérique. Elles ont aussi beaucoup d'applications dans la simulation de phénomènes physiques. Le plus souvent le problème aboutit à des calculs d'intégrales multidimensionnelles et on effectue des quadratures quasi-Monte Carlo.

Un problème classique en analyse numérique dans lequel les méthodes quasi-Monte Carlo peuvent être appliquées est l'optimisation globale. La méthode de Monte Carlo classique pour trouver les extréma globaux est la recherche aléatoire. Elle est utilisée quand la fonction à optimiser n'est pas différentiable et dans ce cas les méthodes classiques sont inopérantes. La version déterministe de la recherche aléatoire est la méthode quasi-Monte Carlo de recherche quasi-aléatoire [49].

Sarkar et Prasad [54] et Morokoff et Caffisch [40] ont appliqué les méthodes quasi-Monte Carlo à la résolution d'équations intégrales.

L'utilisation de méthodes quasi-Monte Carlo dans les problèmes de transfert radiatif [26] a donné des résultats meilleurs que ceux obtenus avec les méthodes de Monte Carlo.

Pour l'application des méthodes quasi-Monte Carlo à la résolution de problèmes aux limites pour certaines équations elliptiques et du problème de Cauchy pour certaines équations paraboliques on peut consulter le livre de Hua et Wang [23] Chapitre 10]. Le même livre traite des problèmes d'interpolation (Chapitre 9) et de la résolution de certaines équations intégrales (Chapitre 10).

Morokoff et Caffisch [38] ont développé une approche quasi-Monte Carlo pour la résolution de l'équation de la chaleur à partir de l'équation discrétisée en temps et en espace.

Dans un travail récent Hofmann et Mathé [22] ont montré comment les équations différentielles stochastiques peuvent être résolues par les méthodes quasi-Monte Carlo.

L'intégration quasi-Monte Carlo a été utilisée avec succès en statistique par Shaw [57].

Drmotič [12] a développé des applications de méthodes quasi-Monte Carlo à des techniques de mesure statistique.

Des quadratures quasi-Monte Carlo en grande dimension (jusqu'en dimension 360) ont été effectuées par Paskov [50] Paskov et Traub [51] pour des problèmes issus de la finance. Ils ont testé les suites de Halton et de Sobol' ainsi que deux suites pseudo-aléatoires. La suite de Sobol' s'est avérée la plus intéressante.

L'équation de Poisson a été résolue par la méthode quasi-Monte Carlo couplée à la méthode des solutions fondamentales (Chen [7]).

Dans les problèmes d'illumination on est amené à résoudre une équation intégrale de Fredholm de deuxième espèce avec un noyau très complexe. Keller [25] a montré que l'on pouvait obtenir des résultats plus précis en remplaçant la méthode de Monte Carlo par la méthode quasi-Monte Carlo pour ce problème.

La diffusion Monte Carlo est utilisée en mécanique quantique. Moskowitz [41] a montré que l'utilisation des suites quasi-aléatoires à la place des suites pseudo-aléatoires réduit l'erreur.

La simulation de processus stochastiques tels que le mouvement Brownien à l'aide des

suites quasi-aléatoires utiliserait des suites de très grande dimension. Moskowitz et Caflisch [42] ont suggéré une méthode de réduction de la dimension du problème.

Les méthodes quasi-Monte Carlo sont également utilisables dans les problèmes de transport de particules qui ont des applications dans la modélisation des réacteurs nucléaires et des dispositifs semiconducteurs (Spanier [60]).

Différentes approches quasi-Monte Carlo ont été développées par Lécot [30Γ31Γ33] pour la résolution de l'équation de Boltzmann homogène en espace et isotrope.

Chapitre 2

Sur la résolution des EDO par des méthodes quasi-Monte Carlo

Résumé: Nous analysons une méthode quasi-Monte Carlo pour la résolution d'un système différentiel $y'(t) = f(t, y(t))$ avec une condition initiale. La fonction f est régulière en y et on suppose que f et $D_y^1 f$ sont à variation bornée en t et que $D_y^2 f$ est bornée dans un voisinage du graphe de la solution. La méthode s'apparente à la méthode de Heun qui est une méthode de Runge-Kutta d'ordre deux. Elle utilise des quadratures quasi-Monte Carlo. La borne d'erreur dépend du carré du pas de temps et de la discrétisation de la suite utilisée dans l'approximation quasi-Monte Carlo. Les tests numériques montrent que la méthode quasi-aléatoire donne de meilleurs résultats qu'une méthode aléatoire récemment développée.

2.1 Introduction

La méthode de Monte Carlo est un outil assez général pour la résolution de problèmes variés issus de la physique mathématique et ses applications ne s'arrêtent pas à l'intégration numérique. C'est en fait toute méthode numérique utilisant des nombres aléatoires. Un grand effort a été fourni ces dernières années pour développer les méthodes quasi-Monte Carlo. Une méthode quasi-Monte Carlo peut être vue comme une version déterministe d'une méthode de Monte Carlo: les nombres aléatoires sont remplacés par des nombres déterministes judicieusement choisis. Par exemple dans le domaine de l'intégration numérique il importe peu que les points choisis soient aléatoires. Ce qui importe est qu'ils soient uniformément répartis. Une synthèse de ce domaine d'études est fournie dans la monographie de Niederreiter [49].

Un algorithme stochastique pour la résolution du système différentiel

$$y'(t) = f(t, y(t)), \quad 0 < t < T, \quad (2.1)$$

$$y(0) = y_0 \quad (2.2)$$

a été récemment proposé par Stengle [61]. La fonction f est supposée régulière en espace (y) mais seulement mesurable et bornée en temps (t). L'algorithme est un représentant d'une famille apparentée à la famille des méthodes de Runge-Kutta. Il engendre une suite

Y_n par la formule de récurrence

$$Y_{n+1} = Y_n + \frac{h_n}{2N} \sum_{0 \leq j < N} \left(f(u_{j,n}, Y_n) + f(U_{j,n}, Y_n + h_n f(u_{j,n}, Y_n)) \right),$$

$$U_{j,n} = \max(T_{1,j,n}, T_{2,j,n}), \quad u_{j,n} = \min(T_{1,j,n}, T_{2,j,n}),$$

où pour chaque pas de temps $\{T_{1,j,n} : 0 \leq j < N\}$ et $\{T_{2,j,n} : 0 \leq j < N\}$ sont de nouvelles suites aléatoires uniformément distribuées dans $[t_n, t_{n+1}]$. La famille de ces méthodes est appelée famille de méthodes de Runge-Kutta Monte Carlo (RKMC). Stengle a indiqué deux conditions sous lesquelles ces méthodes peuvent avoir des avantages par rapport aux méthodes classiques. D'une part la variation de f en y doit être très faible par rapport à la variation de f en t (une absence totale de régularité en t étant un cas extrême). D'autre part le gain de temps dû à la parallélisation des calculs de Monte Carlo doit être important.

On peut espérer également une amélioration grâce à l'utilisation de suites quasi-aléatoires pour remplacer les suites pseudo-aléatoires. Dans ce travail nous proposons une méthode de Runge-Kutta quasi-Monte Carlo (RKQMC) du second ordre.

Elle utilise les suites à faible discrédance. La méthode RKQMC utilise la même suite à discrédance faible pour toutes les quadratures quasi-Monte Carlo.

Le chapitre est organisé de la façon suivante. La méthode RKQMC est décrite au §2.2. Au §2.3 nous faisons une analyse de l'erreur de la méthode. Si \hat{h} est le maximum des pas de temps l'erreur est majorée pas une combinaison linéaire de \hat{h}^2 et de $\mathbf{D}_N(X)$. Au §2.4 nous considérons le problème modèle proposé par Stengle. Les résultats numériques indiquent une amélioration significative de l'erreur par rapport au schéma RKMC. Enfin au §2.5 nous résumons le chapitre et tirons quelques conclusions.

2.2 Méthode de Runge-Kutta quasi-Monte Carlo

Considérons le problème de Cauchy pour le système p -dimensionnel (2.1)–(2.2) où y_0 est donnée dans \mathbb{R}^p . Le problème est supposé avoir une unique solution $y(t)$ absolument continue dans $[0, T]$. Soit $\|\cdot\|$ une norme sur \mathbb{R}^p et $B(y, \rho)$ la boule ouverte correspondante centrée en y et de rayon ρ . On suppose que la fonction f vérifie l'hypothèse suivante.

Hypothèse 2.1 *Il existe $\tau > 0$ et $\rho > 0$ tels que*

– Pour tout $t \in [0, T]$ la fonction $y \rightarrow D_y^m f(t, y)$ est continue sur la boule ouverte $B(y(t), \rho)$, pour $0 \leq m \leq 2$.

– Soit

$$\Omega = \bigcup_{0 \leq t \leq T} [t, \min(t + \tau, T)] \times B(y(t), \rho).$$

Pour tout $t \in [0, T]$ et tout $y \in B(y(t), \rho)$,

1. la fonction $u \rightarrow D_y^m f(u, y)$ est définie sur $[t, \min(t + \tau, T)]$ et bornée par $\|D_y^m f\|_{\infty, \Omega}$, pour $0 \leq m \leq 2$,

2. la variation de la fonction $u \rightarrow D_y^m f(u, y)$ sur $[t, \min(t + \tau, T)]$ est bornée par $V_\Omega(D_y^m f)$, pour $0 \leq m \leq 1$.

La notation suivante sera utilisée: si \mathbf{v} est un vecteur de composantes (v_1, v_2) on écrit $\underline{v} = \min(v_1, v_2)$ et $\bar{v} = \max(v_1, v_2)$. Considérons une partition $0 = t_0 < t_1 < \dots < t_{\hat{n}} = T$ de $[0, T]$ en \hat{n} sous-intervalles de longueurs respectives $h_n = t_{n+1} - t_n$ et notons $\hat{h} = \max_{0 \leq n < \hat{n}} h_n$. Nous avons

$$y(t_{n+1}) = y(t_n) + \int_{t_n}^{t_{n+1}} f\left(t, y(t_n) + \int_{t_n}^t f(u, y(u)) du\right) dt. \quad (2.3)$$

En utilisant un développement de Taylor en y on a l'approximation

$$y(t_{n+1}) \approx y(t_n) + \frac{1}{2h_n} \int_{t_n}^{t_{n+1}} \int_{t_n}^{t_{n+1}} \left(f(\underline{s}, y(t_n)) + f(\bar{s}, y(t_n) + h_n f(\underline{s}, y(t_n))) \right) ds. \quad (2.4)$$

La méthode de Heun [62] correspond à l'approximation $\underline{s} \approx t_n$, $\bar{s} \approx t_{n+1}$. Soit X un ensemble de points $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N-1} \in \bar{T}^2$. La méthode RKQMC du second ordre engendre une suite (y_n) par

$$y_{n+1} = y_n + \frac{h_n}{2N} \sum_{0 \leq j < N} \left(f(t_n + h_n \underline{x}_j, y_n) + f(t_n + h_n \bar{x}_j, y_n + h_n f(t_n + h_n \underline{x}_j, y_n)) \right).$$

2.3 Analyse de l'erreur

L'analyse de l'erreur est assez voisine de celle du schéma de Heun. L'*erreur de discrétisation locale* est définie par

$$\begin{aligned} \varepsilon_n &= \frac{1}{h_n} (y(t_{n+1}) - y(t_n)) \\ &\quad - \frac{1}{2h_n^2} \int_{t_n}^{t_{n+1}} \int_{t_n}^{t_{n+1}} \left(f(\underline{s}, y(t_n)) + f(\bar{s}, y(t_n) + h_n f(\underline{s}, y(t_n))) \right) ds. \end{aligned}$$

Introduisons un *terme d'erreur*

$$\begin{aligned} E_n &= \frac{1}{2h_n^2} \int_{t_n}^{t_{n+1}} \int_{t_n}^{t_{n+1}} \left(f(\underline{s}, y_n) + f(\bar{s}, y_n + h_n f(\underline{s}, y_n)) \right) ds \\ &\quad - \frac{1}{2h_n^2} \int_{t_n}^{t_{n+1}} \int_{t_n}^{t_{n+1}} \left(f(\underline{s}, y(t_n)) + f(\bar{s}, y(t_n) + h_n f(\underline{s}, y(t_n))) \right) ds, \end{aligned}$$

et l'*erreur de l'approximation quasi-Monte Carlo*

$$\begin{aligned} \delta_n &= \frac{1}{2N} \sum_{0 \leq j < N} \left(f(t_n + h_n \underline{x}_j, y_n) + f(t_n + h_n \bar{x}_j, y_n + h_n f(t_n + h_n \underline{x}_j, y_n)) \right) \\ &\quad - \frac{1}{2} \int_{I^2} \left(f(t_n + h_n \underline{x}, y_n) + f(t_n + h_n \bar{x}, y_n + h_n f(t_n + h_n \underline{x}, y_n)) \right) d\mathbf{x}. \end{aligned}$$

L'erreur de discrétisation globale est $e_n = y_n - y(t_n)$. Nous avons la formule de récurrence

$$e_{n+1} = e_n - h_n \varepsilon_n + h_n E_n + h_n \delta_n. \quad (2.5)$$

Nous aurons à utiliser un concept approprié de variation totale pour les fonctions de plusieurs variables. Pour une fonction $\varphi : \bar{T}^s \rightarrow \mathbb{R}^p$ et pour $\mathbf{w}, \mathbf{w}' \in \bar{T}^s$ et $1 \leq i \leq s$ soit $T_{\mathbf{w}}^i \varphi$ la restriction de φ à l'hyperplan $x_i = w_i$ et $\Delta_{\mathbf{w}, \mathbf{w}'}^i \varphi = T_{\mathbf{w}'}^i \varphi - T_{\mathbf{w}}^i \varphi$. Si $K = \{i_1, \dots, i_k\} \subset [1, s]$ on pose

$$T_{\mathbf{w}}^K \varphi = T_{\mathbf{w}}^{i_1} \dots T_{\mathbf{w}}^{i_k} \varphi \quad \text{et} \quad \Delta_{\mathbf{w}, \mathbf{w}'}^K \varphi = \Delta_{\mathbf{w}, \mathbf{w}'}^{i_1} \dots \Delta_{\mathbf{w}, \mathbf{w}'}^{i_k} \varphi.$$

Notons $T_{\mathbf{w}} \varphi = T_{\mathbf{w}}^{[1, s]} \varphi$ et $\Delta_{\mathbf{w}, \mathbf{w}'} \varphi = \Delta_{\mathbf{w}, \mathbf{w}'}^{[1, s]} \varphi$. Si

$$0 = x_{0, i} < x_{1, i} < \dots < x_{n_i, i} = 1, \quad \text{pour} \quad 1 \leq i \leq s$$

définissent une partition de \bar{T}^s en sous-intervalles Γ et $\mathbf{a} = (a_1, \dots, a_s)$ avec des entiers a_i , $0 \leq a_i < n_i$ on écrit $\mathbf{x}_{\mathbf{a}} = (x_{a_1, 1}, \dots, x_{a_s, s})$ et $\mathbf{a}+ = (a_1 + 1, \dots, a_s + 1)$. La *variation de φ sur \bar{T}^s au sens de Vitali* est définie par

$$V^{(s)}(\varphi) = \sup_{\mathcal{P}} \sum_{\mathbf{a}} \|\Delta_{\mathbf{x}_{\mathbf{a}}, \mathbf{x}_{\mathbf{a}+}} \varphi\|,$$

où \mathcal{P} parcourt toutes les partitions de \bar{T}^s en sous-intervalles. Si $\mathbf{1} = (1, \dots, 1) \in \bar{T}^s$ alors

$$V(\varphi) = \sum_{k=1}^s \sum_{\substack{K \subset [1, s] \\ \#K=k}} V^{(k)}(T_{\mathbf{1}}^{K^c} \varphi)$$

est appelé la *variation de φ sur \bar{T}^s au sens de Hardy et Krause*. On renvoie à [29] pour plus d'information sur ce concept. L'ensemble de toutes les fonctions à variation bornée sur \bar{T}^s au sens de Hardy et Krause sera noté $BVHK^s$. Par [66] Proposition 2] toute fonction de $BVHK^s$ est intégrable au sens de Riemann sur \bar{T}^s . Etant donnée une fonction φ définie sur \bar{T}^2 on écrit $\varphi^\circ(\mathbf{x}) = \varphi(\underline{x}, \bar{x})$. On a besoin des résultats techniques suivants.

LEMME 2.1 *Si $\varphi \in BVHK^2$, alors $\varphi^\circ \in BVHK^2$.*

Démonstration. Nous avons $T_{\mathbf{1}}^1 \varphi^\circ = T_{\mathbf{1}}^2 \varphi^\circ = T_{\mathbf{1}}^2 \varphi$. Par un simple calcul nous avons

$$V^{(2)}(\varphi^\circ) \leq 3V^{(2)}(\varphi) + V^{(1)}(T_{\mathbf{1}}^1 \varphi) + V^{(1)}(T_{\mathbf{1}}^2 \varphi),$$

et le résultat du lemme suit. ■

Si $t \in [0, T]$ $0 \leq h$ et $y \in \mathbb{R}^p$ définissons

$$\begin{aligned} \phi_{t, h, y}(\mathbf{x}) &= f(t + hx_1, y), \\ \psi_{t, h, y}(\mathbf{x}) &= f(t + hx_2, y + hf(t + hx_1, y)), \end{aligned}$$

pour $\mathbf{x} \in \bar{T}^2$.

LEMME 2.2 *Si t et $t + h \in [0, T]$, $h \leq \tau$ et $\|y - y(t)\| + 2h\|f\|_{\infty, \Omega} < \rho$, alors $\phi_{t, h, y}$ et $\psi_{t, h, y} \in BVHK^2$.*

Démonstration. Des calculs directs montrent que

$$\begin{aligned} V^{(1)}(T_1^1 \phi_{t,h,y}) &= V^{(2)}(\phi_{t,h,y}) = 0, & V^{(1)}(T_1^2 \phi_{t,h,y}) &\leq V_\Omega(f), \\ V^{(1)}(T_1^1 \psi_{t,h,y}) &\leq V_\Omega(f). \end{aligned}$$

Avec un développement de Taylor en y on trouve

$$V^{(1)}(T_1^2 \psi_{t,h,y}) \leq hV_\Omega(f) \|D_y^1 f\|_{\infty, \Omega} \quad \text{et} \quad V^{(2)}(\psi_{t,h,y}) \leq hV_\Omega(f) V_\Omega(D_y^1 f).$$

Ceci termine la démonstration. ■

Posons

$$\phi_n = \phi_{t_n, h_n, y_n}, \quad \Phi_n = \phi_{t_n, h_n, y(t_n)}, \quad \psi_n = \psi_{t_n, h_n, y_n}, \quad \Psi_n = \psi_{t_n, h_n, y(t_n)}.$$

Alors

$$\varepsilon_n = \frac{1}{h_n} (y(t_{n+1}) - y(t_n)) - \frac{1}{2} \int_{I^2} (\Phi_n^\circ(\mathbf{x}) + \Psi_n^\circ(\mathbf{x})) d\mathbf{x}, \quad (2.6)$$

$$E_n = \frac{1}{2} \int_{I^2} (\phi_n^\circ(\mathbf{x}) + \psi_n^\circ(\mathbf{x})) d\mathbf{x} - \frac{1}{2} \int_{I^2} (\Phi_n^\circ(\mathbf{x}) + \Psi_n^\circ(\mathbf{x})) d\mathbf{x}, \quad (2.7)$$

$$\delta_n = \frac{1}{2N} \sum_{0 \leq j < N} (\phi_n^\circ(\mathbf{x}_j) + \psi_n^\circ(\mathbf{x}_j)) - \frac{1}{2} \int_{I^2} (\phi_n^\circ(\mathbf{x}) + \psi_n^\circ(\mathbf{x})) d\mathbf{x}. \quad (2.8)$$

Nous allons majorer chacun de ces termes Γ pour $0 \leq n < \hat{n}$.

PROPOSITION 2.1 *Si $h_n \leq \tau$ et $2h_n \|f\|_{\infty, \Omega} < \rho$, alors*

$$\|\varepsilon_n\| \leq \frac{h_n^2}{12} \|f\|_{\infty, \Omega} \left(5 \|f\|_{\infty, \Omega} \|D_y^2 f\|_{\infty, \Omega} + 2 \|D_y^1 f\|_{\infty, \Omega}^2 \right).$$

Démonstration. Grâce aux Lemmes 2.1 et 2.2 $\Gamma \varepsilon_n$ est bien défini. En utilisant des développements de Taylor en y nous obtenons $\varepsilon_n = -\varepsilon_{n,1} + \varepsilon_{n,2} + \varepsilon_{n,3}$ où

$$\begin{aligned} \varepsilon_{n,1} &= \frac{h_n^2}{2} \int_{I^2} \left(\int_I D_y^2 f(t_n + h_n \bar{x}, y(t_n) + \xi h_n \underline{x}, y(t_n)) \right. \\ &\quad \cdot \left. \left(f(t_n + h_n \underline{x}, y(t_n)) \right)^2 (1 - \xi) d\xi \right) d\mathbf{x}, \\ \varepsilon_{n,2} &= \frac{h_n}{2} \int_{I^2} D_y^1 f(t_n + h_n \bar{x}, y(t_n)) \\ &\quad \cdot \left(\int_I D_y^1 f(t_n + h_n \underline{x}, (1 - \xi)y(t_n) + \xi y(t_n + h_n \underline{x})) \right. \\ &\quad \cdot \left. (y(t_n + h_n \underline{x}) - y(t_n)) d\xi \right) d\mathbf{x}, \\ \varepsilon_{n,3} &= \int_I \left(\int_I D_y^2 f(t_n + h_n x, (1 - \xi)y(t_n) + \xi y(t)) \right. \\ &\quad \cdot \left. (y(t_n + h_n x) - y(t_n))^2 (1 - \xi) d\xi \right) dx. \end{aligned}$$

D'où le résultat de la proposition. ■

PROPOSITION 2.2 *Si $h_n \leq \tau$ et $\|e_n\| + 2h_n\|f\|_{\infty,\Omega} < \rho$, alors*

$$\|E_n\| \leq \|D_y^1 f\|_{\infty,\Omega} \left(1 + \frac{h_n}{2} \|D_y^1 f\|_{\infty,\Omega}\right) \|e_n\|.$$

Démonstration. Nous savons d'après les Lemmes 2.1 et 2.2 que chaque intégrale dans (2.7) existe. Par des développements de Taylor en y on trouve $E_n = E_{n,1} + E_{n,2}$ où

$$\begin{aligned} E_{n,1} &= \frac{1}{2} \int_{I^2} \left(\int_I D_y^1 f(t_n + h_n \underline{x}, (1-\xi)y(t_n) + \xi y_n) \cdot (y_n - y(t_n)) d\xi \right) dx, \\ E_{n,2} &= \frac{1}{2} \int_{I^2} \left(\int_I D_y^1 f \left(t_n + h_n \bar{x}, (1-\xi)(y(t_n) + h_n f(t_n + h_n \underline{x}, y(t_n))) \right. \right. \\ &\quad \left. \left. + \xi(y_n + h_n f(t_n + h_n \underline{x}, y_n)) \right) \right. \\ &\quad \left. \cdot (y_n + h_n f(t_n + h_n \underline{x}, y_n) - y(t_n) - h_n f(t_n + h_n \underline{x}, y(t_n))) d\xi \right) dx, \end{aligned}$$

ce qui conduit à l'estimation désirée sur $\|E_n\|$. ■

PROPOSITION 2.3 *Si $h_n \leq \tau$ et $\|e_n\| + 2h_n\|f\|_{\infty,\Omega} < \rho$, alors*

$$\|\delta_n\| \leq V_\Omega(f) \left(1 + \frac{h_n}{2} (\|D_y^1 f\|_{\infty,\Omega} + 2V_\Omega(D_y^1 f))\right) \mathbf{D}_N(X).$$

Démonstration. Considérons une double partition de \bar{I} .

$$\begin{aligned} 0 &= w_0 < w_1 < \dots < w_\ell = 1, \\ \xi_0 &= 0, \quad w_k \leq \xi_{k+1} < w_{k+1}, \text{ pour } 0 \leq k < \ell, \quad \xi_{\ell+1} = 1, \end{aligned}$$

telle que $X \subset \{(\xi_{k_1}, \xi_{k_2}) : 0 \leq k_1, k_2 \leq \ell + 1\}$. Posons

$$h(W) = \max_{0 \leq k < \ell} (w_{k+1} - w_k).$$

En utilisant la formule de sommation d'Abel multidimensionnelle avec la fonction auxiliaire

$$\zeta(\mathbf{x}) = D_N([0, x_1] \times [0, x_2], X),$$

on trouve pour $h(W)$ suffisamment petit

$$\begin{aligned} &\frac{1}{N} \sum_{0 \leq j < N} \phi_n^\circ(\mathbf{x}_j) - \sum_{0 \leq k_1, k_2 < \ell} (w_{k_1+1} - w_{k_1})(w_{k_2+1} - w_{k_2}) \phi_n^\circ(\xi_{k_1+1}, \xi_{k_2+1}) \\ &= \sum_{0 \leq k \leq \ell} (f(t_n + h_n \xi_{k+1}, y_n) - f(t_n + h_n \xi_k, y_n)) D_N([w_k, 1]^2, X), \end{aligned}$$

et si on pose $\xi_{\mathbf{k}} = (\xi_{k_1}, \xi_{k_2})\Gamma$

$$\begin{aligned} & \frac{1}{N} \sum_{0 \leq j < N} \psi_n^o(\mathbf{x}_j) - \sum_{0 \leq k_1, k_2 < \ell} (w_{k_1+1} - w_{k_1})(w_{k_2+1} - w_{k_2}) \psi_n^o(\xi_{k_1+1}, \xi_{k_2+1}) \\ &= - \sum_{0 \leq k_1 < k_2 \leq \ell} \Delta_{\xi_{\mathbf{k}}, \xi_{\mathbf{k}+}} \psi_n D_N([w_{k_1}, w_{k_2}]^2, X) + \\ & \quad \sum_{0 \leq k_1, k_2 \leq \ell} \Delta_{\xi_{\mathbf{k}}, \xi_{\mathbf{k}+}} \psi_n D_N([0, w_{k_2}]^2, X) - \\ & \quad \sum_{0 \leq k \leq \ell} \Delta_{\xi_{\mathbf{k}}, \xi_{\mathbf{k}+}} T_1^1 \psi_n D_N([0, w_k]^2, X) + \sum_{0 \leq k \leq \ell} \Delta_{\xi_{\mathbf{k}}, \xi_{\mathbf{k}+}} T_1^2 \psi_n D_N([w_k, 1]^2, X). \end{aligned}$$

En passant à la limite $h(W) \rightarrow 0$ et en utilisant le Lemme 2.1 et les estimations du Lemme 2.2 on arrive à l'estimation désirée. ■

Nous sommes maintenant en mesure de démontrer le résultat de convergence. Posons

$$\begin{aligned} c_1(f) &= \frac{\|f\|_{\infty, \Omega}}{12} \left(5\|f\|_{\infty, \Omega} \|D_y^2 f\|_{\infty, \Omega} + 2\|D_y^1 f\|_{\infty, \Omega}^2 \right), \\ c_2(f, \hat{h}) &= \|D_y^1 f\|_{\infty, \Omega} \left(1 + \frac{\hat{h}}{2} \|D_y^1 f\|_{\infty, \Omega} \right), \\ c_3(f, \hat{h}) &= V_{\Omega}(f) \left(1 + \frac{\hat{h}}{2} \left(\|D_y^1 f\|_{\infty, \Omega} + 2V_{\Omega}(D_y^1 f) \right) \right). \end{aligned}$$

PROPOSITION 2.4 Si $\hat{h} \leq \tau$ et

$$e^{c_2(f, \hat{h})T} \|e_0\| + \frac{e^{c_2(f, \hat{h})T} - 1}{c_2(f, \hat{h})} \left(c_1(f) \hat{h}^2 + c_3(f, \hat{h}) \mathbf{D}_N(X) \right) + 2\hat{h} \|f\|_{\infty, \Omega} < \rho,$$

alors, pour $0 \leq n \leq \hat{n}$,

$$\|e_n\| \leq e^{c_2(f, \hat{h})tn} \|e_0\| + \frac{e^{c_2(f, \hat{h})tn} - 1}{c_2(f, \hat{h})} \left(c_1(f) \hat{h}^2 + c_3(f, \hat{h}) \mathbf{D}_N(X) \right).$$

Démonstration. Il suffit d'utiliser l'équation (2.5) avec les Propositions 2.1, 2.2 et 2.3. ■

2.4 Exemple numérique

Pour illustrer ce qui précède nous allons utiliser différentes méthodes d'ordre deux pour résoudre un problème modèle proposé par Stengle.

$$y'(t) = y(t) + \mu \sin(\cos(\lambda t)), \quad 0 < t < 1, \quad (2.9)$$

$$y(0) = 1, \quad (2.10)$$

où $\lambda = 1023$ et $\mu = 5$. La Figure 2.1 compare les erreurs obtenues par en résolvant ce problème par les schémas suivants.

- La méthode de Heun avec 10 (RK10) et 100 (RK100) pas de temps égaux.

- La méthode RKMC de Stengle avec 10 pas de temps égaux. Elle utilise une nouvelles suite aléatoire uniforme de 100 éléments à chaque pas de temps.
- La méthode RKQMC avec 10 pas de temps égaux. Elle utilise l'ensemble de Hammersley $\Gamma X = \{(\phi_2(j), \frac{j}{N}) : 0 \leq j < N\}$ (où ϕ_2 est la fonction radicale inverse en base 2: voir [49]) avec $N = 100$ pour l'intégration quasi-Monte Carlo .

Le graphique de gauche montre les résultats pour les méthodes RK10 et RKMC. Le graphique de droite contient les résultats pour les méthodes RK100 (seules les erreurs aux instants $t_n = n/10$, $0 \leq n \leq 10$ ont été reportées pour des raisons de clarté) et RKQMC. On notera la différence d'échelle. Les paramètres sont choisis de telle manière que la méthode numérique avec le plus grand pas de temps est à la limite de sa capacité d'approcher la solution. Les résultats de l'algorithme RKMC avec 10 pas de temps sont assez proches de ceux de la méthode RK la plus précise. Mais l'utilisation de l'ensemble de Hammersley conduit à une amélioration spectaculaire de l'erreur par rapport à la méthode RKMC.

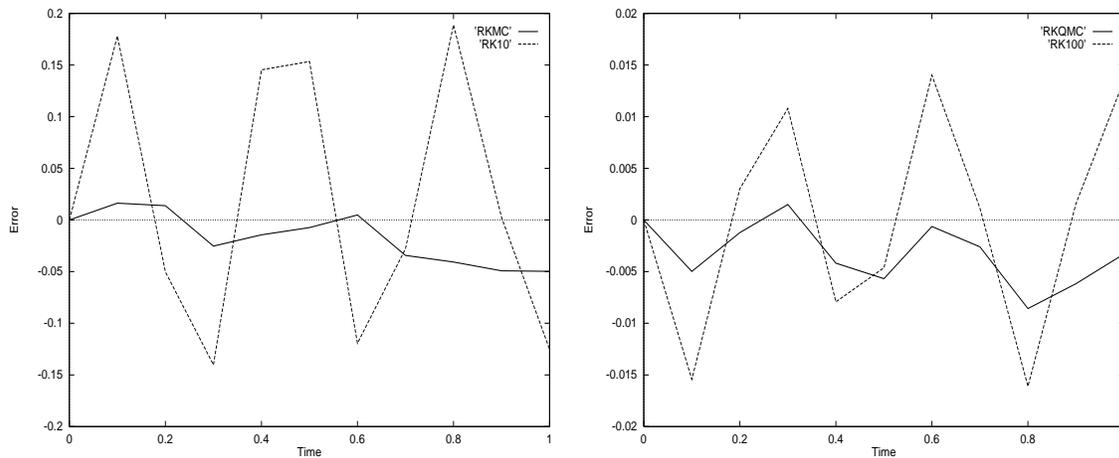


FIG. 2.1 – *Comparaison de méthodes Runge-Kutta d'ordre deux.*

2.5 Conclusion

Ce travail montre qu'une méthode de Runge-Kutta quasi-aléatoire est faisable et peut donner des résultats plus précis qu'une méthode aléatoire. Il apparait que le caractère aléatoire des points n'est pas important. L'analyse de la convergence montre que si les suites à faible discrèpances sont utilisées il est possible d'obtenir de meilleurs résultats de convergence qu'avec les suites aléatoires. Ceci est confirmé par les tests numériques. Il existe des méthodes de type quasi-Monte Carlo pour de nombreux autres problèmes. En fait pour de nombreuses méthodes Monte Carlo il est possible de développer des méthodes quasi-Monte Carlo correspondantes comme des versions déterministes. Ceci est fait dans les chapitres suivants.

Chapitre 3

Méthodes particulières pour l'équation de Boltzmann linéaire

Résumé: Nous décrivons une méthode particulière déterministe pour simuler l'équation de Boltzmann linéaire homogène en espace. Elle combine un schéma d'Euler en temps et une intégration quasi-Monte Carlo utilisant une suite $(0, 2s + 1)$. Les particules sont renumérotées par rapport à leurs composantes de vitesse à chaque pas de temps. Une preuve de la convergence de la méthode est donnée. Des tests numériques sont présentés pour un problème test pour lequel la solution exacte est connue. Le schéma quasi-Monte Carlo est comparé du point de vue de la précision à un schéma standard de Monte Carlo. Les résultats montrent une amélioration de l'ordre de l'erreur et du taux de convergence par rapport à l'approche Monte Carlo.

3.1 Introduction

Les méthodes quasi-Monte Carlo peuvent être décrites comme une version déterministe des méthodes Monte Carlo : les échantillons aléatoires sont remplacés par des points uniformément distribués ; de plus une majoration d'erreur est obtenue au lieu d'une convergence stochastique. Dans le cas d'une intégration numérique la méthode quasi-Monte Carlo permet en général d'obtenir une plus grande précision que la méthode Monte Carlo. Des études numériques d'intégration multidimensionnelle par des méthodes quasi-Monte Carlo sont présentées dans [40].

On peut se demander si l'amélioration fournie par l'utilisation de suites quasi-aléatoires à la place des nombres aléatoires peut être aussi obtenue dans la simulation de problèmes physiques.

Plusieurs méthodes de types quasi-Monte Carlo ont été proposées ces dernières années pour différents problèmes [5, 7, 25, 26, 30, 31, 33, 38, 40, 41, 54, 60].

Quelques résultats de convergence sont établis et les tests numériques ont montré une amélioration de convergence par rapport aux méthodes de type Monte Carlo. Dans [30, 31, 33] Lécot a proposé un schéma déterministe pour résoudre une équation de Boltzmann non-linéaire unidimensionnelle et spatialement homogène. Il a fourni des résultats de convergence dans le cas où les particules sont réordonnées à chaque pas de

temps. La méthode est basée sur le tri des particules selon leur vitesse. Les résultats numériques montrent une nette supériorité des méthodes quasi-Monte Carlo sur celles de Monte Carlo pour ce problème. Ici nous nous sommes intéressés à l'extension des méthodes quasi-Monte Carlo à la simulation particulière d'un problème multidimensionnel: l'équation de Boltzmann linéaire homogène en espace et en dimension s de vitesse. Cette méthode est basée sur les réseaux- (t, m, s) et les suites- (t, s) . Dans ce schéma les particules sont réordonnées par paquets à chaque pas de temps comme suit. Le nombre N de particules simulées est pris sous la forme $b^{d_1} \dots b^{d_s}$ où b et d_1, \dots, d_s sont des entiers. Les particules sont triées en b^{d_1} sous-ensembles de niveau 1 selon la première composante de leur vitesse. Ensuite les $b^{d_2} \dots b^{d_s}$ particules de chaque sous-ensemble de niveau 1 sont triées en b^{d_2} sous-ensembles de niveau 2 selon la deuxième composante de leur vitesse et ainsi de suite. Finalement les b^{d_s} particules de chaque sous ensemble de niveau $s - 1$ sont triées selon la s -ième composante de leur vitesse.

L'organisation de ce chapitre est la suivante. Dans le §3.2 l'équation linéaire intégral-différentielle de Boltzmann est introduite. Le domaine de la vitesse est \mathbb{R}^s . Après quelques changements de variables le domaine devient I^s . Ensuite on décrit la méthode quasi-Monte Carlo pour l'équation: elle combine un schéma d'Euler explicite en temps avec une intégration numérique sur I^{2s+1} . Les particules simulées sont renumérotées à chaque pas de temps. Les quadratures utilisent une suite- $(0, 2s + 1)$ en base b .

Le §3.3 expose la méthode de Nanbu qui est une approche de type Monte Carlo.

Dans le §3.4 la preuve de la convergence est donnée. L'erreur du schéma est définie comme la discrétisation des vitesses discrètes relativement à la distribution de vitesse exacte.

Dans le §3.5 les erreurs et les taux de convergence expérimentaux sont calculés. On considère un cas test où la solution exacte est connue pour les dimensions $1 \leq s \leq 5$. Les erreurs et les taux de convergence sont comparés à ceux de la méthode de Nanbu.

Les exemples indiquent que le nouveau schéma utilise moins de particules pour obtenir le même niveau d'erreur que la simulation aléatoire. Il permet aussi d'obtenir une amélioration du taux de convergence. Les conclusions sont enfin données dans le §3.6.

3.2 Algorithme quasi-Monte Carlo

Pour l'obtention de l'équation de Boltzmann on renvoie à la monographie de Cercignani [6]. Considérons un mélange spatialement homogène de deux gaz. Un des composants de ce mélange que l'on appellera L a une densité très faible. L'autre constituant appelé M est macroscopiquement au repos et en équilibre.

Soit ρ_M la masse par unité de volume et m_M la masse moléculaire de l'espèce M et ρ_L et m_L les mêmes quantités pour l'espèce L . Le gaz M a une distribution de vitesse Maxwellienne notée F_M

$$F_M(\mathbf{v}) = \frac{\rho_M}{(2\pi R_M T)^{s/2}} \exp\left(-\frac{|\mathbf{v}|^2}{2R_M T}\right), \quad \mathbf{v} \in \mathbb{R}^s,$$

où R_M est la constante de Boltzmann pour le gaz M T est la température et $|\cdot|$ la norme

euclidienne. Soit $f_{L,0}$ la distribution de vitesse initiale pour le gaz L . On a

$$\int_{\mathbb{R}^s} f_{L,0}(\mathbf{v}) d\mathbf{v} = \rho_L.$$

La distribution de vitesse pour le gaz L vérifie l'équation de Boltzmann linéaire

$$\begin{aligned} & \frac{\partial f_L}{\partial t}(\mathbf{v}, t) \\ &= \frac{1}{m_M} \int_{\mathcal{D}_{\mathbf{v}}} \left(f_L(\mathbf{v}', t) F_M(\mathbf{w}') - f_L(\mathbf{v}, t) F_M(\mathbf{w}) \right) |\mathbf{v} - \mathbf{w}| s_{sc}(\theta, |\mathbf{v} - \mathbf{w}|) d\mathbf{w} d\mathbf{n}, \end{aligned}$$

$$\mathbf{v} \in \mathbb{R}^s, \quad t > 0,$$

où

$$\mathcal{D}_{\mathbf{v}} = \left\{ (\mathbf{w}, \mathbf{n}) \in \mathbb{R}^{2s} : |\mathbf{n}| = 1, \mathbf{n} \cdot (\mathbf{v} - \mathbf{w}) > 0 \right\},$$

$$\mathbf{v}' = \mathbf{v} - \frac{2m_M}{m_L + m_M} \mathbf{n} \cdot (\mathbf{v} - \mathbf{w}) \mathbf{n},$$

$$\mathbf{w}' = \mathbf{w} + \frac{2m_L}{m_L + m_M} \mathbf{n} \cdot (\mathbf{v} - \mathbf{w}) \mathbf{n},$$

s_{sc} est la section efficace de diffusion et θ l'angle entre \mathbf{n} et $\mathbf{v} - \mathbf{w}$.

Avant de commencer toute discrétisation Γ nous faisons des changements de variables de manière à ramener le domaine d'intégration à I^s . Ces changements de variables sont motivés par le fait que la quadrature quasi-Monte Carlo est effectuée avec des suites à faible discrédance uniformément réparties dans le cube unité.

Notons

$$F_L(\mathbf{v}) = \left(\frac{m_L}{m_M} \right)^{s/2} \frac{\rho_L}{\rho_M} F_M \left(\left(\frac{m_L}{m_M} \right)^{1/2} \mathbf{v} \right).$$

En utilisant les nouvelles variables \mathbf{v}' et $\mathbf{n}' = \frac{\mathbf{v} - \mathbf{w}}{|\mathbf{v} - \mathbf{w}|}$ dans l'intégrale Γ on arrive à l'équation de Boltzmann linéaire décrite par Cercignani dans [6] Chapitre 4 Section 3]

$$\frac{\partial f_L}{\partial t}(\mathbf{v}, t) = \int_{\mathbb{R}^s} \left(f_L(\mathbf{v}', t) F_L(\mathbf{v}) - f_L(\mathbf{v}, t) F_L(\mathbf{v}') \right) \kappa_{L,M}(\mathbf{v}', \mathbf{v}) d\mathbf{v}', \quad (3.1)$$

où $\kappa_{L,M}$ est un noyau symétrique positif ou nul défini par

$$\begin{aligned} \kappa_{L,M}(\mathbf{v}', \mathbf{v}) &= \frac{1}{F_L(\mathbf{v})} \frac{1}{m_M} \left(\frac{m_L + m_M}{2m_M} \right)^{s+1} \\ &\cdot \int_{\mathbf{n}' \cdot (\mathbf{v} - \mathbf{v}') > 0} F_M \left(\mathbf{v} + \frac{m_L}{m_M} (\mathbf{v} - \mathbf{v}') - \frac{m_L + m_M}{2m_M} \frac{|\mathbf{v} - \mathbf{v}'|}{\cos \theta} \mathbf{n}' \right) \\ &\cdot |\mathbf{v} - \mathbf{v}'| s_{sc} \left(\theta, \frac{m_L + m_M}{2m_M} \frac{|\mathbf{v} - \mathbf{v}'|}{\cos \theta} \right) \frac{d\mathbf{n}'}{(\cos \theta)^{s+1}}, \end{aligned}$$

et θ est l'angle entre \mathbf{n}' et $\mathbf{v} - \mathbf{v}'$. Alors Γ si

$$\begin{aligned} f_0(\mathbf{v}) &= \frac{1}{\rho_L} \left(\frac{2R_M T m_M}{m_L} \right)^{s/2} f_{L,0} \left(\left(\frac{2R_M T m_M}{m_L} \right)^{1/2} \mathbf{v} \right), \\ f(\mathbf{v}, t) &= \frac{1}{\rho_L} \left(\frac{2R_M T m_M}{m_L} \right)^{s/2} f_L \left(\left(\frac{2R_M T m_M}{m_L} \right)^{1/2} \mathbf{v}, t \right), \\ F(v) &= \frac{1}{\pi^{s/2}} e^{-|v|^2}, \\ \kappa(\mathbf{v}, \mathbf{v}') &= \rho_L \kappa_{L,M} \left(\left(\frac{2R_M T m_M}{m_L} \right)^{1/2} \mathbf{v}, \left(\frac{2R_M T m_M}{m_L} \right)^{1/2} \mathbf{v}' \right), \end{aligned}$$

on obtient

$$\frac{\partial f}{\partial t}(\mathbf{v}, t) = \int_{\mathbb{R}^s} (f(\mathbf{v}', t)F(\mathbf{v}) - f(\mathbf{v}, t)F(\mathbf{v}')) \kappa(\mathbf{v}, \mathbf{v}') d\mathbf{v}', \quad \mathbf{v} \in \mathbb{R}^s, \quad t > 0, \quad (3.2)$$

$$f(\mathbf{v}, 0) = f_0(\mathbf{v}), \quad \mathbf{v} \in \mathbb{R}^s,$$

avec

$$\int_{\mathbb{R}^s} f_0(\mathbf{v}) d\mathbf{v} = 1.$$

On ramène le domaine de vitesse à I^s par un nouveau changement de variables. Soit D la fonction de \mathbb{R} dans I définie par

$$D(v) = \frac{1}{2} (1 + \operatorname{erf}(v)), \quad v \in \mathbb{R}.$$

L'utilisation des nouvelles variables définies par $x_i = D(v_i)$ conduit à une nouvelle équation Γ

$$\frac{\partial u}{\partial t}(\mathbf{x}, t) = \int_{I^s} (u(\mathbf{x}', t) - u(\mathbf{x}, t)) \gamma(\mathbf{x}, \mathbf{x}') d\mathbf{x}', \quad \mathbf{x} \in I^s, \quad t > 0, \quad (3.3)$$

$$u(\mathbf{x}, 0) = u_0(\mathbf{x}), \quad \mathbf{x} \in I^s,$$

où

$$u(\mathbf{x}, t) = \frac{f(\mathbf{v}, t)}{F(\mathbf{v})}, \quad \gamma(\mathbf{x}, \mathbf{x}') = \kappa(\mathbf{v}, \mathbf{v}'), \quad u_0(\mathbf{x}) = \frac{f_0(\mathbf{v})}{F(\mathbf{v})}.$$

On supposera que le noyau γ est symétrique Γ positif ou nul et borné. Posons

$$\|\gamma\|_\infty = \sup_{\mathbf{x}, \mathbf{x}' \in I^s} \gamma(\mathbf{x}, \mathbf{x}').$$

Pour construire une méthode particulière quasi-Monte Carlo de résolution du problème (3.3) il est nécessaire de passer par une formulation faible.

DÉFINITION 3.1 Une fonction σ définie dans $A \subset I^s$ dont l'image est constituée d'un nombre fini de points dans $[0, +\infty)$ sera appelée une fonction simple. On notera $\mathcal{S}(A)$ l'ensemble des fonctions simples mesurables au sens de Borel dans A .

En multipliant l'équation (3.3) par $\sigma \in \mathcal{S}(\bar{I}^s)$ et en intégrant sur I^s on a la formulation faible suivante

$$\frac{d}{dt} \int_{I^s} \sigma(\mathbf{x}) u(\mathbf{x}, t) d\mathbf{x} = \int_{I^{2s}} (\sigma(\mathbf{x}') - \sigma(\mathbf{x})) \gamma(\mathbf{x}, \mathbf{x}') u(\mathbf{x}, t) d\mathbf{x} d\mathbf{x}', \quad \sigma \in \mathcal{S}(\bar{I}^s). \quad (3.4)$$

Notons $\delta(\mathbf{x} - \mathbf{y})$ la mesure de Dirac concentrée au point $\mathbf{y} \in I^s$. Soit $b \geq 2$ et $m \geq 0$ deux entiers. Posons $N = b^m$. Soit d_1, \dots, d_s des entiers tels que $d_1 + \dots + d_s = m$. Étant donné des entiers a_i avec $0 \leq a_i < b^{d_i}$ posons $\mathbf{a} = (a_1, \dots, a_s)$. Un ensemble $\mathbf{X}^{(0)}$ de N points $\mathbf{x}_j^{(0)}$ dans I^s est choisi de sorte que

$$u^{(0)}(\mathbf{x}) = \frac{1}{N} \sum_{j=0}^{N-1} \delta(\mathbf{x} - \mathbf{x}_j^{(0)})$$

“approche” la distribution initiale $u_0(\mathbf{x}) d\mathbf{x}$ (dans un sens qui sera précisé dans le §3.4). Soit Δt un pas de temps avec

$$\Delta t \|\gamma\|_\infty \leq 1.$$

Avec cette condition le schéma explicite en temps est faisable.

Introduisons alors les instants discrets $t_n = n\Delta t$ et posons $u_n(\mathbf{x}) = u(\mathbf{x}, t_n)$. Les ensembles $\mathbf{X}^{(n)}$ de N points $\mathbf{x}_j^{(n)}$ de I^s sont calculés de telle sorte que

$$u^{(n)}(\mathbf{x}) = \frac{1}{N} \sum_{j=0}^{N-1} \delta(\mathbf{x} - \mathbf{x}_j^{(n)})$$

“approche” la distribution $u_n(\mathbf{x}) d\mathbf{x}$.

On a besoin d'une suite $(0, 2s + 1)$ en base b : $\mathbf{y}_0, \mathbf{y}_1, \dots$ pour effectuer les quadratures quasi-Monte Carlo. La simulation est construite comme suit.

(i) Renumérotation des particules:

Cette technique nous permettra dans le §3.4 de démontrer la convergence de la méthode. Dans le §3.5 on verra que la renumérotation des particules permet effectivement de réduire l'erreur. Les N points de $\mathbf{X}^{(n)}$ sont renumérotés $\mathbf{x}_\mathbf{a}^{(n)} = (x_{\mathbf{a},1}^{(n)}, \dots, x_{\mathbf{a},s}^{(n)})$ avec $0 \leq a_i < b^{d_i}$ $i = 1, \dots, s$ de telle manière que

- si $a_1 < b_1$ alors $x_{\mathbf{a},1}^{(n)} \leq x_{\mathbf{b},1}^{(n)}$,
- si $a_1 = b_1$ et $a_2 < b_2$ alors $x_{\mathbf{a},s}^{(n)} \leq x_{\mathbf{b},s}^{(n)}$
.....
- si $a_1 = b_1, \dots, a_{s-1} = b_{s-1}$ et $a_s < b_s$ alors $x_{\mathbf{a},s}^{(n)} \leq x_{\mathbf{b},s}^{(n)}$.

Voir la Figure 3.1.

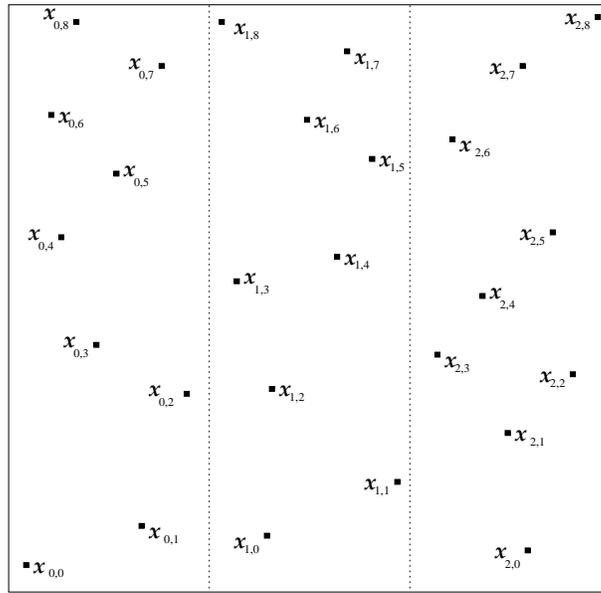


FIG. 3.1 – Renumérotation des particules ($s = 2$, $b = 3$, $d_1 = 1$, $d_2 = 2$).

(ii) Schéma d'Euler:

En utilisant un schéma d'Euler explicite en temps dans l'équation (3.4) une approximation $v^{(n+1)}$ de la distribution exacte $u_{n+1}(\mathbf{x})d\mathbf{x}$ est définie par

$$\begin{aligned} & \frac{1}{\Delta t} \int_{I^s} \sigma(\mathbf{x}) \left(v^{(n+1)}(\mathbf{x}) - u^{(n)}(\mathbf{x}) \right) \\ &= \int_{I^{2s}} \left(\sigma(\mathbf{x}') - \sigma(\mathbf{x}) \right) \gamma(\mathbf{x}, \mathbf{x}') u^{(n)}(\mathbf{x}) d\mathbf{x}', \quad \sigma \in \mathcal{S}(\bar{I}^s), \end{aligned}$$

c'est à dire

$$\begin{aligned} \int_{I^s} \sigma(\mathbf{x}) v^{(n+1)}(\mathbf{x}) &= \frac{1}{N} \sum_{\mathbf{a}} \left(1 - \Delta t \int_{I^s} \gamma(\mathbf{x}_{\mathbf{a}}^{(n)}, \mathbf{x}') d\mathbf{x}' \right) \sigma(\mathbf{x}_{\mathbf{a}}^{(n)}) \\ &+ \frac{\Delta t}{N} \sum_{\mathbf{a}} \int_{I^s} \sigma(\mathbf{x}') \gamma(\mathbf{x}_{\mathbf{a}}^{(n)}, \mathbf{x}') d\mathbf{x}'. \end{aligned} \tag{3.5}$$

(iii) Intégration quasi-Monte Carlo:

$v^{(n+1)}$ est une mesure continue. Or nous cherchons la solution sous la forme d'une somme de N mesures de Dirac; d'où la nécessité de procéder à une approximation au moyen d'une quadrature quasi-Monte Carlo. Pour cela nous allons mettre le second membre de l'équation (3.5) sous la forme d'une intégrale multidimensionnelle. Soit $c_{\mathbf{a}}$ la fonction caractéristique de $I_{\mathbf{a}} = \prod_{i=1}^s \left[\frac{a_i}{b^{d_i}}, \frac{a_i + 1}{b^{d_i}} \right)$ et $\chi_{\mathbf{a}}^{(n)}$ celle de

$$\left\{ \mathbf{x} = (\mathbf{x}', x_{s+1}) \in I^{s+1} : x_{s+1} < \Delta t \gamma(\mathbf{x}_{\mathbf{a}}^{(n)}, \mathbf{x}') \right\}.$$

Etant donné $\sigma \in \mathcal{S}(\bar{I}^s)$ on définit $\Sigma^{(n)} \in \mathcal{S}(\bar{I}^{2s+1})$ par

$$\Sigma^{(n)}(\mathbf{x}) = \sum_{\mathbf{a}} c_{\mathbf{a}}(\mathbf{x}') \left(\left(1 - \chi_{\mathbf{a}}^{(n)}(\mathbf{x}'', x_{2s+1})\right) \sigma(\mathbf{x}_{\mathbf{a}}^{(n)}) + \chi_{\mathbf{a}}^{(n)}(\mathbf{x}'', x_{2s+1}) \sigma(\mathbf{x}'') \right),$$

$$\mathbf{x} = (\mathbf{x}', \mathbf{x}'', x_{2s+1}) \in \bar{I}^{2s+1}.$$

Puisque $\Delta t \|\gamma\|_{\infty} \leq 1$ on obtient

$$\int_{I^s} \sigma(\mathbf{x}) v^{(n+1)}(\mathbf{x}) = \int_{I^{2s+1}} \Sigma^{(n)}(\mathbf{x}) d\mathbf{x}. \quad (3.6)$$

On définit maintenant l'approximation $u^{(n+1)}$ de la distribution exacte $u_{n+1}(\mathbf{x}) d\mathbf{x}$ grâce à une quadrature quasi-Monte Carlo avec un réseau $(0, m, 2s+1)$ en base $b \Gamma\{\mathbf{y}_{nN+j} : 0 \leq j < N\}$.

$$\int_{I^s} \sigma(\mathbf{x}) u^{(n+1)}(\mathbf{x}) = \frac{1}{N} \sum_{j=0}^{N-1} \Sigma^{(n)}(\mathbf{y}_{nN+j}), \quad \sigma \in \mathcal{S}(\bar{I}^s). \quad (3.7)$$

Soit

$$\mathbf{a}^{(n)}(j) = ([b^{d_1} y_{nN+j,1}], \dots, [b^{d_s} y_{nN+j,s}]),$$

où $[c]$ est le plus grand entier inférieur ou égal à c . Le choix d'un réseau $(0, m, 2s+1)$ dans la quadrature quasi-Monte Carlo implique que chaque cellule $I_{\mathbf{a}}$ contient exactement un point $\mathbf{y}'_{nN+j} = (y_{nN+j,1}, \dots, y_{nN+j,s})$. Ainsi l'application $\mathbf{a}^{(n)}(j)$ est une bijection. L'équation (3.7) se réécrit

$$\frac{1}{N} \sum_{j=0}^{N-1} \sigma(\mathbf{x}_j^{(n+1)}) = \frac{1}{N} \sum_{j=0}^{N-1} \left(\left(1 - \chi_{\mathbf{a}^{(n)}(j)}^{(n)}(\mathbf{y}''_{nN+j}, y_{nN+j,2s+1})\right) \sigma(\mathbf{x}_{\mathbf{a}^{(n)}(j)}^{(n)}) + \chi_{\mathbf{a}^{(n)}(j)}^{(n)}(\mathbf{y}''_{nN+j}, y_{nN+j,2s+1}) \sigma(\mathbf{y}''_{nN+j}) \right)$$

où $\mathbf{y}''_{nN+j} = (y_{nN+j,s+1}, \dots, y_{nN+j,2s})$.

Donc les nouvelles vitesses sont telles que

$$\begin{aligned} \mathbf{x}_j^{(n+1)} &= \mathbf{y}''_{nN+j} \quad \text{si } y_{nN+j,2s+1} < \Delta t \gamma(\mathbf{x}_{\mathbf{a}^{(n)}(j)}^{(n)}, \mathbf{y}''_{nN+j}), \\ \mathbf{x}_j^{(n+1)} &= \mathbf{x}_{\mathbf{a}^{(n)}(j)}^{(n)} \quad \text{sinon.} \end{aligned} \quad (3.8)$$

REMARQUE 3.1 Par l'inégalité de Koksma-Hlawka de faibles erreurs de quadratures quasi-Monte Carlo sont garanties si on utilise des suites à faible discrèpance. Si $\mathbf{y}_0, \mathbf{y}_1, \dots$ est une suite $(0, 2s+1)$ en base $b \Gamma$ alors les réseaux $(0, m, 2s+1)$ que l'on utilise sont les ensembles de points ayant les plus faibles discrèpances connues.

3.3 Méthode de Nanbu

Cette méthode a été décrite par Nanbu [43] pour l'équation Master. Etant donné un pas de temps Δt on commence par faire un schéma d'Euler explicite en temps pour l'équation (3.3):

$$\frac{u^{(n+1)}(\mathbf{x}) - u^{(n)}(\mathbf{x})}{\Delta t} = \int_{I^s} (u^{(n)}(\mathbf{x}') - u^{(n)}(\mathbf{x})) \gamma(\mathbf{x}, \mathbf{x}') d\mathbf{x}',$$

c'est à dire

$$u^{(n+1)}(\mathbf{x}) = \left(1 - \Delta t \int_{I^s} \gamma(\mathbf{x}, \mathbf{x}') d\mathbf{x}'\right) u^{(n)}(\mathbf{x}) + \Delta t \int_{I^s} u^{(n)}(\mathbf{x}') \gamma(\mathbf{x}, \mathbf{x}') d\mathbf{x}',$$

où $u^{(n)}$ est la solution à l'instant $t_n = n\Delta t$.

Le pas de temps est choisi de manière à satisfaire la condition

$$\Delta t \|\gamma\|_{\infty} \leq 1.$$

Soit N le nombre de particule simulées. Comme pour la méthode quasi-Monte Carlo la distribution initiale est approchée par une somme de mesures de Dirac. La méthode de Nanbu permet de calculer la solution approchée à l'instant t_{n+1}

$$u^{(n+1)}(\mathbf{x}) = \frac{1}{N} \sum_{j=0}^{N-1} \delta(\mathbf{x} - \mathbf{x}_j^{(n+1)}),$$

à partir de la solution approchée

$$u^{(n)}(\mathbf{x}) = \frac{1}{N} \sum_{j=0}^{N-1} \delta(\mathbf{x} - \mathbf{x}_j^{(n)})$$

à l'instant t_n .

En utilisant l'approximation particulière et compte tenu du fait que le noyau γ est symétrique on arrive à

$$u^{(n+1)}(x) = \frac{1}{N} \sum_{j=0}^{N-1} \left[\left(1 - \Delta t \int_{I^s} \gamma(\mathbf{x}_j^{(n)}, \mathbf{x}) d\mathbf{x}\right) \delta(\mathbf{x} - \mathbf{x}_j^{(n)}) + \Delta t \gamma(\mathbf{x}_j^{(n)}, \mathbf{x}) \right].$$

Notons

$$p_j^{(n)} = \Delta t \int_{I^s} \gamma(\mathbf{x}_j^{(n)}, \mathbf{x}) d\mathbf{x},$$

et

$$\tilde{\gamma}(\mathbf{x}_j^{(n)}, \mathbf{x}) = \frac{\gamma(\mathbf{x}_j^{(n)}, \mathbf{x})}{\int_{I^s} \gamma(\mathbf{x}_j^{(n)}, \mathbf{x}) d\mathbf{x}}.$$

La fonction $\tilde{\gamma}(\mathbf{x}_j^{(n)}, \mathbf{x})$ est une densité de probabilité

$$\int_{I^s} \tilde{\gamma}(\mathbf{x}_j^{(n)}, \mathbf{x}) d\mathbf{x} = 1,$$

et

$$u^{(n+1)}(x) = \frac{1}{N} \sum_{j=0}^{N-1} \left[(1 - p_j^{(n)}) \delta(\mathbf{x} - \mathbf{x}_j^{(n)}) + p_j^{(n)} \tilde{\gamma}(\mathbf{x}_j^{(n)}, \mathbf{x}) \right]. \quad (3.9)$$

Soit $\tilde{\mathbf{x}}_j^{(n)}$ un échantillon de la variable aléatoire ayant $\tilde{\gamma}(\mathbf{x}_j^{(n)}, \mathbf{x})$ pour densité. L'équation (3.9) se réécrit

$$u^{(n+1)}(x) = \frac{1}{N} \sum_{j=0}^{N-1} \left[(1 - p_j^{(n)}) \delta(\mathbf{x} - \mathbf{x}_j^{(n)}) + p_j^{(n)} \delta(\mathbf{x} - \tilde{\mathbf{x}}_j^{(n)}) \right]. \quad (3.10)$$

Finalement pour tout $j \in \{0, \dots, N-1\}$ on tire uniformément un nombre p dans $[0, 1]$. Il y a deux cas:

– Si $0 \leq p \leq p_j^{(n)}$: alors la nouvelle vitesse est

$$\mathbf{x}_j^{(n+1)} = \tilde{\mathbf{x}}_j^{(n)}.$$

– Si $p_j^{(n)} < p \leq 1$: alors

$$\mathbf{x}_j^{(n+1)} = \mathbf{x}_j^{(n)}.$$

3.4 Bornes d'erreur pour la méthode quasi-Monte Carlo

L'erreur du schéma à l'instant t_n est définie par

$$\mathbf{D}_N^* (\mathbf{X}^{(n)}, u_n) = \sup_{\mathbf{z} \in \bar{I}^s} |d_N^{(n)}(\mathbf{z})|,$$

où

$$d_N^{(n)}(\mathbf{z}) = \frac{1}{N} \sum_{\mathbf{a}} \sigma_{\mathbf{z}}(\mathbf{x}_{\mathbf{a}}^{(n)}) - \int_{I^s} \sigma_{\mathbf{z}}(\mathbf{x}) u_n(\mathbf{x}) d\mathbf{x},$$

$\sigma_{\mathbf{z}}$ étant la fonction caractéristique de $\prod_{i=1}^s [0, z_i]$.

Ainsi $\mathbf{D}_N^* (\mathbf{X}^{(n)}, u_n)$ est la discrédance* de l'ensemble des points $\mathbf{X}^{(n)}$ par rapport à la distribution u_n . Cette notion est définie dans le livre de Hlawka, Firneis et Zinterhof [21, Chapitre 1, Section 5]. L'analyse de l'erreur est celle de la méthode d'Euler.

On introduit l'erreur de troncature

$$\begin{aligned} \varepsilon^{(n)}(\mathbf{z}) &= \frac{1}{\Delta t} \int_{I^s} \sigma_{\mathbf{z}}(\mathbf{x}) (u_{n+1}(\mathbf{x}) - u_n(\mathbf{x})) d\mathbf{x} \\ &\quad - \int_{I^{2s}} (\sigma_{\mathbf{z}}(\mathbf{x}') - \sigma_{\mathbf{z}}(\mathbf{x})) \gamma(\mathbf{x}, \mathbf{x}') u_n(\mathbf{x}) d\mathbf{x} d\mathbf{x}', \end{aligned}$$

et un terme d'erreur $e_N^{(n)}(\mathbf{z}) = e_{N,1}^{(n)}(\mathbf{z}) - e_{N,2}^{(n)}(\mathbf{z})\Gamma$ où

$$e_{N,1}^{(n)}(\mathbf{z}) = \frac{1}{N} \sum_{\mathbf{a}} \int_{I^s} \gamma(\mathbf{x}_{\mathbf{a}}^{(n)}, \mathbf{x}') d\mathbf{x}' \sigma_{\mathbf{z}}(\mathbf{x}_{\mathbf{a}}^{(n)}) - \int_{I^{2s}} \gamma(\mathbf{x}, \mathbf{x}') \sigma_{\mathbf{z}}(\mathbf{x}) u_n(\mathbf{x}) d\mathbf{x} d\mathbf{x}',$$

$$e_{N,2}^{(n)}(\mathbf{z}) = \frac{1}{N} \sum_{\mathbf{a}} \int_{I^s} \gamma(\mathbf{x}_{\mathbf{a}}^{(n)}, \mathbf{x}') \sigma_{\mathbf{z}}(\mathbf{x}') d\mathbf{x}' - \int_{I^{2s}} \gamma(\mathbf{x}, \mathbf{x}') \sigma_{\mathbf{z}}(\mathbf{x}') u_n(\mathbf{x}) d\mathbf{x} d\mathbf{x}'.$$

De plus on a besoin de l'erreur d'intégration quasi-Monte Carlo

$$\delta_N^{(n)}(\mathbf{z}) = \frac{1}{N} \sum_{j=0}^{N-1} \Sigma_{\mathbf{z}}^{(n)}(\mathbf{y}_{nN+j}) - \int_{I^{2s+1}} \Sigma_{\mathbf{z}}^{(n)}(\mathbf{x}) d\mathbf{x},$$

où $\Sigma_{\mathbf{z}}^{(n)} \in \mathcal{S}(\bar{I}^{2s+1})$ correspond à $\sigma_{\mathbf{z}} \in \mathcal{S}(\bar{I}^s)$ par

$$\Sigma_{\mathbf{z}}^{(n)}(\mathbf{x}) = \sum_{\mathbf{a}} c_{\mathbf{a}}(\mathbf{x}') \left(\left(1 - \chi_{\mathbf{a}}^{(n)}(\mathbf{x}'', x_{2s+1})\right) \sigma_{\mathbf{z}}(\mathbf{x}_{\mathbf{a}}^{(n)}) + \chi_{\mathbf{a}}^{(n)}(\mathbf{x}'', x_{2s+1}) \sigma_{\mathbf{z}}(\mathbf{x}'') \right),$$

$$\mathbf{x} = (\mathbf{x}', \mathbf{x}'', x_{2s+1}) \in \bar{I}^{2s+1}.$$

On a alors la formule de récurrence

$$d_N^{(n)}(\mathbf{z}) = d_N^{(n-1)}(\mathbf{z}) - \Delta t e_N^{(n-1)}(\mathbf{z}) - \Delta t \varepsilon^{(n-1)}(\mathbf{z}) + \delta_N^{(n-1)}(\mathbf{z}). \quad (3.11)$$

La généralisation suivante de l'inégalité de Koksma-Hlawka est donnée dans le livre de Hlawka, Firneis et Zinterhof [21]. La démonstration est assez voisine de celle de l'inégalité de Koksma-Hlawka donnée par Zaremba dans [66].

LEMME 3.1 *Soit ρ une fonction positive ou nulle et intégrable au sens de Riemann sur \bar{I}^s avec*

$$\int_{I^s} \rho(\mathbf{x}) d\mathbf{x} = 1.$$

Si f est à variation bornée $V(f)$ sur \bar{I}^s au sens de Hardy et Krause, alors, pour tout ensemble de points $\mathbf{x}_0, \dots, \mathbf{x}_{N-1} \in I^s$, on a

$$\left| \frac{1}{N} \sum_{j=0}^{N-1} f(\mathbf{x}_j) - \int_{I^s} f(\mathbf{x}) \rho(\mathbf{x}) d\mathbf{x} \right| \leq V(f) \mathbf{D}_N^*(X, \rho). \quad (3.12)$$

Le lemme suivant est utilisé pour estimer les termes d'erreur $e_{N,1}^{(n)}(\mathbf{z})$ et $e_{N,2}^{(n)}(\mathbf{z})$.

LEMME 3.2 *Pour $\mathbf{x} \in \bar{I}^s$, soit $\sigma_{\mathbf{x}}$ la fonction caractéristique de $\prod_{i=1}^s [0, x_i]$. Si f a une variation bornée $V(f)$ sur \bar{I}^s au sens de Hardy et Krause, alors pour tout $\mathbf{x} \in \bar{I}^s$, $f\sigma_{\mathbf{x}}$ a une variation bornée $V(f\sigma_{\mathbf{x}})$ sur \bar{I}^s au sens de Hardy et Krause et on a la majoration*

$$V(f\sigma_{\mathbf{x}}) \leq V(f) + |f(1, \dots, 1)|. \quad (3.13)$$

Preuve. On peut supposer que $x_i > 0$ pour $1 \leq i \leq s$ (sinon $f\sigma_{\mathbf{x}} = 0$). On a $V(f\sigma_{\mathbf{x}}) = V^{(s)}(f\sigma_{\mathbf{x}})$.

Soit

$$0 = x_{0,i} < x_{1,i} < \dots < x_{n_i,i} = 1, \quad \text{pour } 1 \leq i \leq s$$

qui définissent une partition de \bar{T}^s en sous-intervalles. Soit \mathbf{a}^* tel que $x_{a_i^*,i} < x_i \leq x_{a_i^*+1,i}$ pour $1 \leq i \leq s$. On a

$$\prod_{i=1}^s [0, n_i] = \bigcup_{k=0}^s \bigcup_{\substack{K \subset [1,s] \\ \#K=k}} \prod_{i=1}^s I_{\mathbf{a}^*,K,i}, \quad (3.14)$$

où $I_{\mathbf{a}^*,K,i} = [0, a_i^*]$ if $i \in K$; sinon $I_{\mathbf{a}^*,K,i} = [a_i^*, n_i]$. Alors

$$\sum_{\mathbf{a}} |\Delta_{\mathbf{x}_{\mathbf{a}}, \mathbf{x}_{\mathbf{a}^+}}(f\sigma_{\mathbf{x}})| = \sum_{k=0}^s \sum_{\substack{K \subset [1,s] \\ \#K=k}} \sum_{\substack{0 \leq a_i < a_i^* \\ i \in K}} \left| \Delta_{\mathbf{x}_{\mathbf{a}}, \mathbf{x}_{\mathbf{a}^+}}^K T_{\mathbf{x}_{\mathbf{a}^*}}^{K^c} f \right|. \quad (3.15)$$

Par ailleurs pour une fonction g définie sur \bar{T}^t $t \leq s$ on a

$$T_{\mathbf{x}_{\mathbf{a}^*}} g = \sum_{\ell=0}^t (-1)^\ell \sum_{\substack{L \subset [1,t] \\ \#L=\ell}} \sum_{\substack{a_i^* \leq a_i < n_i \\ i \in L}} \Delta_{\mathbf{x}_{\mathbf{a}}, \mathbf{x}_{\mathbf{a}^+}}^L T_{\mathbf{1}}^{L^c} g. \quad (3.16)$$

La combinaison de (3.15) et (3.16) (avec $g = \Delta_{\mathbf{x}_{\mathbf{a}}, \mathbf{x}_{\mathbf{a}^+}}^K f$) et (3.14) donne

$$\sum_{\mathbf{a}} |\Delta_{\mathbf{x}_{\mathbf{a}^+}, \mathbf{x}_{\mathbf{a}}}(f\sigma_{\mathbf{x}})| \leq \sum_{m=0}^s \sum_{\substack{M \subset [1,s] \\ \#M=m}} \sum_{\substack{0 \leq a_i < n_i \\ i \in M}} \left| \Delta_{\mathbf{x}_{\mathbf{a}}, \mathbf{x}_{\mathbf{a}^+}}^M T_{\mathbf{1}}^{M^c} f \right| \leq V(f) + |f(\mathbf{1})|. \quad (3.17)$$

En maximisant sur toutes les partitions en sous-intervalles de \bar{T}^s on obtient le résultat du lemme. ■

Avec les deux lemmes précédents l'estimation des termes $e_{N,1}^{(n)}(\mathbf{z})$ et $e_{N,2}^{(n)}(\mathbf{z})$ est immédiate.

PROPOSITION 3.1 Si, pour tout $\mathbf{z} \in \bar{T}^s$, la fonction $\bar{\gamma}_{\mathbf{z}}(\mathbf{x}) = \int_{I^s} \gamma(\mathbf{x}, \mathbf{x}') \sigma_{\mathbf{z}}(\mathbf{x}') d\mathbf{x}'$ a une variation bornée $V(\bar{\gamma}_{\mathbf{z}})$ sur \bar{T}^s au sens de Hardy et Krause, et si

$$\bar{V}(\gamma) = \sup_{\mathbf{z} \in \bar{T}^s} V(\bar{\gamma}_{\mathbf{z}}) < +\infty,$$

alors

$$\sup_{\mathbf{z} \in \bar{T}^s} |e_{N,1}^{(n)}(\mathbf{z})| \leq (V(\bar{\gamma}_{\mathbf{1}}) + \|\gamma\|_{\infty}) \mathbf{D}_N^* (\mathbf{X}^{(n)}, u_n), \quad (3.18)$$

$$\sup_{\mathbf{z} \in \bar{T}^s} |e_{N,2}^{(n)}(\mathbf{z})| \leq \bar{V}(\gamma) \mathbf{D}_N^* (\mathbf{X}^{(n)}, u_n). \quad (3.19)$$

Preuve. Par le Lemme 3.1 on trouve $|e_{N,2}^{(n)}(\mathbf{z})| \leq V(\bar{\gamma}_{\mathbf{z}}) \mathbf{D}_N^* (\mathbf{X}^{(n)}, u_n)$, d'où l'estimation (3.19) et $|e_{N,1}^{(n)}(\mathbf{z})| \leq V(\bar{\gamma}_{\mathbf{1}} \sigma_{\mathbf{z}}) \mathbf{D}_N^* (\mathbf{X}^{(n)}, u_n)$. Ainsi (3.18) résulte du Lemme 3.2. ■

REMARQUE 3.2 Si γ a une variation bornée $V(\gamma)$ sur \bar{T}^{2s} au sens de Hardy et Krause alors la condition de la Proposition 3.1 est satisfaite et $\bar{V}(\gamma) \leq V(\gamma)$.

La majoration

$$|\varepsilon^{(n)}(\mathbf{z})| \leq \int_{t_n}^{t_{n+1}} \int_{I^s} \left| \frac{\partial^2 u}{\partial t^2}(\mathbf{x}, t) \right| dx dt \quad (3.20)$$

est immédiate. Il reste maintenant à analyser l'erreur d'intégration $\delta_N^{(n)}(\mathbf{z})$. Le résultat suivant est dû à Niederreiter [47].

LEMME 3.3 *Soit X un réseau- (t, m, s) en base b . Pour tout intervalle élémentaire $J' \subset I^{s-1}$ en base b et pour tout $x_s \in \bar{I}$ on a*

$$\left| \frac{A(J' \times [0, x_s), X)}{b^m} - \lambda_s(J' \times [0, x_s]) \right| \leq b^{t-m}. \quad (3.21)$$

L'ensemble de points $Y^{(n)} = \{\mathbf{y}_{nN+j} : 0 \leq j < N\}$ est un réseau- $(0, m, 2s+1)$ en base b . La quantité $\delta_N^{(n)}(\mathbf{z})$ peut s'exprimer

$$\begin{aligned} \delta_N^{(n)}(\mathbf{z}) &= \frac{A(F_{\mathbf{z}}^{(n)}, Y^{(n)})}{N} - \lambda_{2s+1}(F_{\mathbf{z}}^{(n)}) - \frac{A(E_{\mathbf{z},1}^{(n)}, Y^{(n)})}{N} + \lambda_{2s+1}(E_{\mathbf{z},1}^{(n)}) \\ &\quad + \frac{A(E_{\mathbf{z},2}^{(n)}, Y^{(n)})}{N} - \lambda_{2s+1}(E_{\mathbf{z},2}^{(n)}), \end{aligned} \quad (3.22)$$

où

$$\begin{aligned} F_{\mathbf{z}}^{(n)} &= \left\{ \mathbf{x} = (\mathbf{x}', \mathbf{x}'', x_{2s+1}) \in I^{2s+1} : \sum_{\mathbf{a}} c_{\mathbf{a}}(\mathbf{x}') \sigma_{\mathbf{z}}(\mathbf{x}_{\mathbf{a}}^{(n)}) = 1 \right\}, \\ E_{\mathbf{z},k}^{(n)} &= \left\{ \mathbf{x} = (\mathbf{x}', \mathbf{x}'', x_{2s+1}) \in I^{2s+1} : x_{2s+1} < \gamma_{\mathbf{z},k}^{(n)}(\mathbf{x}', \mathbf{x}'') \right\}, \quad k = 1, 2, \\ \gamma_{\mathbf{z},1}^{(n)}(\mathbf{x}', \mathbf{x}'') &= \Delta t \sum_{\mathbf{a}} c_{\mathbf{a}}(\mathbf{x}') \gamma(\mathbf{x}_{\mathbf{a}}^{(n)}, \mathbf{x}'') \sigma_{\mathbf{z}}(\mathbf{x}_{\mathbf{a}}^{(n)}), \quad (\mathbf{x}', \mathbf{x}'') \in \bar{I}^{2s}, \\ \gamma_{\mathbf{z},2}^{(n)}(\mathbf{x}', \mathbf{x}'') &= \Delta t \sum_{\mathbf{a}} c_{\mathbf{a}}(\mathbf{x}') \gamma(\mathbf{x}_{\mathbf{a}}^{(n)}, \mathbf{x}'') \sigma_{\mathbf{z}}(\mathbf{x}''), \quad (\mathbf{x}', \mathbf{x}'') \in \bar{I}^{2s}. \end{aligned}$$

L'ensemble $F_{\mathbf{z}}^{(n)}$ est une union disjointe d'intervalles élémentaires $I_{\mathbf{a}} \times I^{s+1}$ en base b avec $\lambda_{2s+1}(I_{\mathbf{a}} \times I^{s+1}) = b^{-m} \Gamma$ donc

$$\frac{A(F_{\mathbf{z}}^{(n)}, Y^{(n)})}{N} - \lambda_{2s+1}(F_{\mathbf{z}}^{(n)}) = 0. \quad (3.23)$$

Nous donnons maintenant une généralisation du Lemme 1 de [34] qui est utile pour la majoration des autres termes.

LEMME 3.4 *Soit f une fonction à variation bornée $V(f)$ sur \bar{I}^s au sens de Hardy et Krause. Soit*

$$\begin{aligned} 0 &= x_{0,1} \leq x_{1,1} \leq \dots \leq x_{n_1,1} = 1, \\ 0 &= x_{a_1,0,2} \leq x_{a_1,1,2} \leq \dots \leq x_{a_1,n_2,2} = 1, \quad \text{pour } 0 \leq a_1 < n_1, \\ &\dots \end{aligned}$$

$0 = x_{\mathbf{a}',0,s} \leq x_{\mathbf{a}',1,s} \leq \dots \leq x_{\mathbf{a}',n_s,s} = 1$, pour $\mathbf{a}' = (a_1, \dots, a_{s-1})$, $0 \leq a_i < n_i$ qui définissent des partitions de \bar{T} en sous-intervalles. Pour $\mathbf{a} = (a_1, \dots, a_s)$ avec des entiers a_i , $0 \leq a_i < n_i$, soit $\bar{T}_{\mathbf{a}} = [x_{a_1,1}, x_{a_1+1,1}] \times [x_{a_1,a_2,2}, x_{a_1,a_2+1,2}] \times \dots \times [x_{\mathbf{a}',a_s,s}, x_{\mathbf{a}',a_s+1,s}]$ et $\mathbf{y}_{\mathbf{a}}, \mathbf{z}_{\mathbf{a}} \in \bar{T}_{\mathbf{a}}$. Alors

$$\sum_{\mathbf{a}} |f(\mathbf{z}_{\mathbf{a}}) - f(\mathbf{y}_{\mathbf{a}})| \leq V(f) \sum_{i=1}^s \frac{1}{n_i} \prod_{i=1}^s n_i. \quad (3.24)$$

Preuve. On a

$$\sum_{\mathbf{a}} |f(\mathbf{z}_{\mathbf{a}}) - f(\mathbf{y}_{\mathbf{a}})| \leq \sum_{i=1}^s \sum_{\mathbf{a}^{*,i}} S(i, \mathbf{a}^{*,i}), \quad (3.25)$$

où $\mathbf{a}^{*,i} = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_s)$ et

$$S(i, \mathbf{a}^{*,i}) = \sum_{a_i < n_i} |f(y_{\mathbf{a},1}, \dots, y_{\mathbf{a},i-1}, z_{\mathbf{a},i}, \dots, z_{\mathbf{a},s}) - f(y_{\mathbf{a},1}, \dots, y_{\mathbf{a},i}, z_{\mathbf{a},i+1}, \dots, z_{\mathbf{a},s})|.$$

Définissons une nouvelle partition de \bar{T}^s en sous-intervalles

$$0 = w_{0,j} < w_{1,j} < \dots < w_{m_j,j} = 1, \quad \text{pour } 1 \leq j \leq s,$$

par

$$\begin{aligned} \{w_{\alpha_j,j} : 0 \leq \alpha_j \leq m_j\} &= \{0, 1\} \cup \{y_{\mathbf{a},j} : 0 \leq a_i < n_i\}, \quad \text{pour } j < i, \\ \{w_{\alpha_i,i} : 0 \leq \alpha_i \leq m_i\} &= \{0, 1\} \cup \{y_{\mathbf{a},i} : 0 \leq a_i < n_i\} \cup \{z_{\mathbf{a},i} : 0 \leq a_i < n_i\}, \\ \{w_{\alpha_j,j} : 0 \leq \alpha_j \leq m_j\} &= \{0, 1\} \cup \{z_{\mathbf{a},j} : 0 \leq a_i < n_i\}, \quad \text{pour } j > i. \end{aligned}$$

Soit $\alpha(\mathbf{a}) = (\alpha_1(\mathbf{a}), \dots, \alpha_s(\mathbf{a}))$ tel que

$$\begin{aligned} w_{\alpha_j(\mathbf{a}),j} &= y_{\mathbf{a},j}, \quad \text{pour } j < i, \quad w_{\alpha_j(\mathbf{a}),j} = z_{\mathbf{a},j}, \quad \text{pour } j > i, \\ \{w_{\alpha_i(\mathbf{a}),i}, w_{\alpha_i(\mathbf{a})+1,i}\} &= \{y_{\mathbf{a},i}, z_{\mathbf{a},i}\}. \end{aligned}$$

Alors

$$S(i, \mathbf{a}^{*,i}) = \sum_{a_i < n_i} \left| \Delta_{(\mathbf{w}_{\alpha(\mathbf{a})+}, \mathbf{w}_{\alpha(\mathbf{a})})}^{(i)} T_{\mathbf{w}_{\alpha(\mathbf{a})}}^{[1,s]^{*,i}} f \right|, \quad (3.26)$$

où $[1, s]^{*,i} = [1, s] \setminus \{i\}$. Puisque

$$T_{\mathbf{w}_{\alpha(\mathbf{a})}}^{[1,s]^{*,i}} f = \sum_{k=0}^{s-1} (-1)^k \sum_{\substack{K \subset [1,s]^{*,i} \\ \#K=k}} \sum_{\substack{\alpha_j(\mathbf{a}) \leq \alpha_j < m_j \\ j \in K}} \Delta_{(\mathbf{w}_{\alpha+}, \mathbf{w}_{\alpha})}^K T_{\mathbf{1}}^{K^c} f, \quad (3.27)$$

on obtient $S(i, \mathbf{a}^{*,i}) \leq V(f)$ et le résultat du lemme s'ensuit. ■

PROPOSITION 3.2 *Si la fonction γ a une variation bornée $V(\gamma)$ sur \bar{T}^{2s} au sens de Hardy et Krause, alors, pour $k = 1, 2$,*

$$\begin{aligned} \left| \frac{A(E_{\mathbf{z},k}^{(n)}, Y^{(n)})}{N} - \lambda_{2s+1}(E_{\mathbf{z},k}^{(n)}) \right| &\leq \frac{1}{b^{d_s - (s+1) \lfloor \frac{d_s}{s+2} \rfloor}} \\ &+ (V(\gamma) + \|\gamma\|_{\infty}) \Delta t \left(\sum_{i=1}^{s-1} \frac{1}{b^{d_i}} + \frac{s+1}{b \lfloor \frac{d_s}{s+2} \rfloor} \right). \end{aligned} \quad (3.28)$$

Preuve. Soit $\delta_1, \dots, \delta_{2s}$ des entiers. Pour $\alpha = (\alpha', \alpha'') = (\alpha_1, \dots, \alpha_{2s})$ avec des entiers $0 \leq \alpha_i < b^{\delta_i}$ définissons

$$I'_{\alpha'} = \prod_{i=1}^s \left[\frac{\alpha_i}{b^{\delta_i}}, \frac{\alpha_i + 1}{b^{\delta_i}} \right), \quad I''_{\alpha''} = \prod_{i=s+1}^{2s} \left[\frac{\alpha_i}{b^{\delta_i}}, \frac{\alpha_i + 1}{b^{\delta_i}} \right), \quad I_{\alpha} = I'_{\alpha'} \times I''_{\alpha''}$$

et notons

$$\underline{E}_{\mathbf{z},k}^{(n)} = \bigcup_{\alpha} I_{\alpha} \times \left[0, \inf_{I_{\alpha}} \gamma_{\mathbf{z},k}^{(n)} \right), \quad \overline{E}_{\mathbf{z},k}^{(n)} = \bigcup_{\alpha} I_{\alpha} \times \left[0, \sup_{I_{\alpha}} \gamma_{\mathbf{z},k}^{(n)} \right),$$

$$\partial E_{\mathbf{z},k}^{(n)} = \bigcup_{\alpha} I_{\alpha} \times \left[\inf_{I_{\alpha}} \gamma_{\mathbf{z},k}^{(n)}, \sup_{I_{\alpha}} \gamma_{\mathbf{z},k}^{(n)} \right].$$

Puisque $\underline{E}_{\mathbf{z},k}^{(n)} \subset E_{\mathbf{z},k}^{(n)} \subset \overline{E}_{\mathbf{z},k}^{(n)}$ on a

$$\frac{A(\underline{E}_{\mathbf{z},k}^{(n)}, Y^{(n)})}{N} - \lambda_{2s+1}(\underline{E}_{\mathbf{z},k}^{(n)}) - \lambda_{2s+1}(\partial E_{\mathbf{z},k}^{(n)}) \leq \frac{A(E_{\mathbf{z},k}^{(n)}, Y^{(n)})}{N} - \lambda_{2s+1}(E_{\mathbf{z},k}^{(n)}) \quad (3.29)$$

$$\frac{A(E_{\mathbf{z},k}^{(n)}, Y^{(n)})}{N} - \lambda_{2s+1}(E_{\mathbf{z},k}^{(n)}) \leq \frac{A(\overline{E}_{\mathbf{z},k}^{(n)}, Y^{(n)})}{N} - \lambda_{2s+1}(\overline{E}_{\mathbf{z},k}^{(n)}) + \lambda_{2s+1}(\partial E_{\mathbf{z},k}^{(n)}).$$

Par le Lemme 3.3 on trouve

$$\left| \frac{A(E_{\mathbf{z},k}^{(n)}, Y^{(n)})}{N} - \lambda_{2s+1}(E_{\mathbf{z},k}^{(n)}) \right| \leq \frac{b^{\delta_1 + \dots + \delta_{2s}}}{b^m} \quad (3.30)$$

$$+ \frac{1}{b^{\delta_1 + \dots + \delta_{2s}}} \sum_{\alpha} \left(\sup_{I_{\alpha}} \gamma_{\mathbf{z},k}^{(n)} - \inf_{I_{\alpha}} \gamma_{\mathbf{z},k}^{(n)} \right).$$

Soit

$$\sigma_{\mathbf{z},1}(\mathbf{x}', \mathbf{x}'') = \sigma_{\mathbf{z}}(\mathbf{x}'), \quad \sigma_{\mathbf{z},2}(\mathbf{x}', \mathbf{x}'') = \sigma_{\mathbf{z}}(\mathbf{x}''), \quad (\mathbf{x}', \mathbf{x}'') \in \overline{I}^{2s}.$$

Si $\delta_1 = d_1, \dots, \delta_{s-1} = d_{s-1}$ et $\delta_s \leq d_s$ on obtient

$$\sup_{I_{\alpha}} \gamma_{\mathbf{z},k}^{(n)} = \Delta t \sup \left\{ (\gamma \sigma_{\mathbf{z},k})(\mathbf{x}_{\alpha_1, \dots, \alpha_{s-1}, a_s}^{(n)}, \mathbf{x}'') : \quad (3.31)$$

$$\alpha_s b^{d_s - \delta_s} \leq a_s < (\alpha_s + 1) b^{d_s - \delta_s}, \quad \mathbf{x}'' \in I''_{\alpha''} \right\},$$

et une égalité analogue pour l'infimum. La renumérotation des particules nous permet de définir de nouvelles partitions de \overline{I} en sous intervalles

$$0 = w_{0,1}^{(n)} \leq w_{1,1}^{(n)} \leq \dots \leq w_{b^{d_1},1}^{(n)} = 1,$$

$$0 = w_{\alpha_1,0,2}^{(n)} \leq w_{\alpha_1,1,2}^{(n)} \leq \dots \leq w_{\alpha_1, b^{d_2},2}^{(n)} = 1, \quad \text{pour } 0 \leq \alpha_1 < b^{d_1},$$

$$\dots$$

$$0 = w_{\alpha_1, \dots, \alpha_{s-1}, 0, s}^{(n)} \leq w_{\alpha_1, \dots, \alpha_{s-1}, 1, s}^{(n)} \leq \dots \leq w_{\alpha_1, \dots, \alpha_{s-1}, b^{\delta_s}, s}^{(n)} = 1,$$

$$\text{pour } 0 \leq \alpha_1 < b^{d_1}, \dots, 0 \leq \alpha_{s-1} < b^{d_{s-1}}$$

telles que Γ pour tout entier a_s avec $\alpha_s b^{d_s - \delta_s} \leq a_s < (\alpha_s + 1) b^{d_s - \delta_s}$ Γ on a

$$\mathbf{x}_{\alpha_1, \dots, \alpha_{s-1}, a_s}^{(n)} \in \left[w_{\alpha_1, 1}^{(n)}, w_{\alpha_1 + 1, 1}^{(n)} \right] \times \dots \times \left[w_{\alpha_1, \dots, \alpha_s, s}^{(n)}, w_{\alpha_1, \dots, \alpha_s + 1, s}^{(n)} \right]. \quad (3.32)$$

Ainsi Γ par les Lemmes 3.2 et 3.4 Γ

$$\frac{1}{b^{\delta_1 + \dots + \delta_{2s}}} \sum_{\alpha} \left(\sup_{I_{\alpha}} \gamma_{\mathbf{z}, k}^{(n)} - \inf_{I_{\alpha}} \gamma_{\mathbf{z}, k}^{(n)} \right) \leq (V(\gamma) + \|\gamma\|_{\infty}) \Delta t \sum_{i=1}^{2s} \frac{1}{b^{\delta_i}}. \quad (3.33)$$

D'où l'inégalité (3.28) en choisissant $\delta_s = \dots = \delta_{2s} = \left\lfloor \frac{d_s}{s+2} \right\rfloor$. ■

Grâce aux différentes inégalités établies ci-dessus Γ on trouve une majoration de l'erreur du schéma quasi-Monte Carlo.

PROPOSITION 3.3 *Supposons que γ a une variation bornée $V(\gamma)$ sur \bar{I}^{2s} au sens de Hardy et Krause, et soit $|\gamma| = 2V(\gamma) + \|\gamma\|_{\infty}$. Alors*

$$\begin{aligned} \mathbf{D}_N^* (\mathbf{X}^{(n)}, u_n) &\leq e^{|\gamma| t_n} \mathbf{D}_N^* (\mathbf{X}^{(0)}, u_0) \\ &+ \Delta t \int_{(0, t_n) \times I^s} e^{|\gamma|(t_n - t)} \left| \frac{\partial^2 u}{\partial t^2} (\mathbf{x}, t) \right| d\mathbf{x} dt \\ &+ \frac{2e^{|\gamma| t_n}}{|\gamma| \Delta t} \left(\frac{1}{b^{d_s - (s+1) \lfloor \frac{d_s}{s+2} \rfloor}} + (V(\gamma) + \|\gamma\|_{\infty}) \Delta t \left(\sum_{i=1}^{s-1} \frac{1}{b^{d_i}} + \frac{s+1}{b^{\lfloor \frac{d_s}{s+2} \rfloor}} \right) \right). \end{aligned}$$

Preuve. Cette estimation résulte des relations (3.11) Γ (3.22) et (3.23) Γ utilisées avec l'inégalité (3.20) et les estimations des Propositions 3.1 et 3.2. ■

REMARQUE 3.3 La borne d'erreur de la Proposition 3.3 augmente quand Δt décroît. D'autre part Γ la Proposition 3.3 reste valable quand on fait toutes les quadratures quasi-Monte Carlo avec le même réseau $(0, m, 2s+1) \Gamma Y^*$ en base b ($\forall n \geq 0 \Gamma Y^{(n)} = Y^*$). Des tests numériques effectués dans ce cas sur un problème monodimensionnel (voir [30]) ont montré une augmentation effective de l'erreur quand Δt tend vers 0. D'autres travaux (voir [31] Γ [33]) ont montré qu'on pouvait éliminer ce comportement en utilisant un nouveau réseau à chaque pas de temps.

3.5 Résultats numériques

Nous allons comparer les performances de l'algorithme quasi-Monte Carlo (QMC) et de la méthode de simulation Monte Carlo (MC) proposée par Nanbu. Les erreurs effectives sont calculées dans un cas où la solution exacte de (3.3) est connue. Quand on utilise les nombres pseudo-aléatoires Γ le théorème de limite centrale indique le taux de convergence

de la simulation Monte Carlo : il est de $N^{-0.5}\Gamma$ où N est le nombre de particules. Le calcul de la discrédance $\mathbf{D}_N^*(\mathbf{X}^{(n)}, u_n)$ est très coûteux si la dimension s est supérieure ou égale à 2. Aussi les erreurs des deux schémas sont calculées comme suit :

$$\widetilde{\mathbf{D}}_N^*(\mathbf{X}^{(n)}, u_n) = \max_{1 \leq k \leq s} \sup_{r \in \overline{I}} |d_{N,k}^{(n)}(r)|,$$

où

$$d_{N,k}^{(n)}(r) = \frac{1}{N} \sum_{i=0}^{N-1} \sigma_r(x_{i,k}^{(n)}) - \int_{I^s} \sigma_r(x_k) u_n(\mathbf{x}) d\mathbf{x}, \quad 1 \leq k \leq s,$$

et σ_r est la fonction caractéristique de $[0, r)$.

Si on suppose $\widetilde{\mathbf{D}}_N^*(\mathbf{X}^{(n)}, u_n) = N^{-\alpha}\Gamma$ l'exposant α peut être estimé à partir de

$$\alpha_N^{(n)} = \frac{-\log(\widetilde{\mathbf{D}}_N^*(\mathbf{X}^{(n)}, u_n))}{\log N}.$$

On appelle $\alpha_N^{(n)}$ l'exposant de l'erreur. Toutes les simulations sont effectuées avec un échantillon de N particules en partant d'une distribution initiale u_0 . Si

$$u_0(\mathbf{x}) = \prod_{i=1}^s u_{0,i}(x_i), \quad \mathbf{x} = (x_1, \dots, x_s) \in \overline{I}^s,$$

les positions initiales des particules sont calculées en transformant un réseau $-(0, m, s)$ en base $b\Gamma\Xi\Gamma$ par la fonction inverse de

$$U_0(\mathbf{x}) = \left(\int_0^{x_1} u_{0,1}(z) dz, \dots, \int_0^{x_s} u_{0,s}(z) dz \right) \quad \mathbf{x} = (x_1, \dots, x_s) \in \overline{I}^s.$$

Alors

$$\mathbf{D}_N^*(\mathbf{X}^{(0)}, u_0) = \mathbf{D}_N^*(\Xi). \quad (3.34)$$

L'algorithme quasi-Monte Carlo utilise une suite $-(0, 2s + 1)$ en base b . Le résultat suivant a été établi par Niederreiter [47].

LEMME 3.5 Une suite $-(0, s)$ en base b ne peut exister que si $s \leq b$.

Puisque la construction quand b est un nombre premier est plus simple que dans le cas général on choisit pour b le plus petit entier premier plus grand ou égal à $2s + 1$. Un algorithme pour construire des suites $-(0, s)$ et des réseaux $-(0, m, s)$ en base $b\Gamma b$ premier Γ a été proposé par Faure [14]. Si

$$\forall \mathbf{x} \in \overline{I}^s \quad \int_{I^s} \gamma(\mathbf{x}, \mathbf{x}') d\mathbf{x}' = \overline{\gamma},$$

la transformée de Laplace de l'équation (3.3) est une équation intégrale de Fredholm de seconde espèce. La solution est obtenue en utilisant la méthode des approximations successives. On choisit le noyau

$$\gamma(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^s \left(1 - \frac{1}{4} \left(1 - 12 \left(x_i - \frac{1}{2} \right)^2 \right) \left(1 - 12 \left(x'_i - \frac{1}{2} \right)^2 \right) \right),$$

$$(\mathbf{x}, \mathbf{x}') = (x_1, \dots, x_s, x'_1, \dots, x'_s) \in \bar{T}^{2s}$$

et la distribution initiale

$$u_0(\mathbf{x}) = \prod_{i=1}^s \frac{1}{\pi \sqrt{x_i(1-x_i)}}.$$

Alors la solution exacte est

$$u(\mathbf{x}, t) = e^{-t} u_0(\mathbf{x}) + \sum_{r=0}^s (-5)^r \left(\exp\left(-\left(1 - \left(-\frac{1}{5}\right)^r\right)t\right) - \exp(-t) \right) \cdot \Pi_r \left(\frac{1}{8} - \frac{3}{2} \left(x_1 - \frac{1}{2}\right)^2, \dots, \frac{1}{8} - \frac{3}{2} \left(x_s - \frac{1}{2}\right)^2 \right), \quad (3.35)$$

$$\mathbf{x} = (x_1, \dots, x_s) \in \bar{T}^s, t \geq 0,$$

où Π_1, \dots, Π_s sont les polynômes symétriques élémentaires de x_1, \dots, x_s .

Pour les simulations le pas de temps est $\Delta t = 0.01$. La taille de l'échantillon est de $N = b^{d_1 + \dots + d_s}$ particules. Les erreurs $\widetilde{\mathbf{D}}_N^*(\mathbf{X}^{(n)}, u_n)$ et les exposants de l'erreur $\alpha_N^{(n)}$ sont représentés sur les Figures 3.2-3.3-3.4-3.5 ($s=1$)-3.6-3.7-3.8-3.9-3.10-3.11-3.12-3.13 ($s=2$)-3.14-3.15-3.16-3.17-3.18-3.19 ($s=3$)-3.20-3.21-3.22 ($s=4$) et 3.23-3.24-3.25-3.26 ($s=5$).

Nous voyons sur les Figures 3.2-3.3-3.4-3.5-3.6-3.7-3.8-3.9-3.14-3.15-3.16 et 3.17 que l'algorithme QMC est plus performant que le schéma MC si $s \leq 3$ et $d_i \approx d_s/(s+2)$ pour $i < s$. Par ailleurs le taux de convergence de l'algorithme quasi-Monte Carlo est nettement meilleur que la borne théorique $N^{-1/(2s+1)}$ donnée par la Proposition 3.3. Un autre choix des paramètres d_i peut conduire à de très mauvais résultats : voir les Figures 3.10-3.11-3.12-3.13-3.18 et 3.19. D'autre part les Figures 3.20-3.21-3.22-3.23-3.24-3.25 et 3.26 montrent clairement une diminution de l'efficacité de la méthode QMC quand la dimension augmente. Pour $s \geq 4$ les tests ont été faits avec un nombre de particules trop petit pour assurer l'efficacité de la méthode quasi-Monte Carlo.

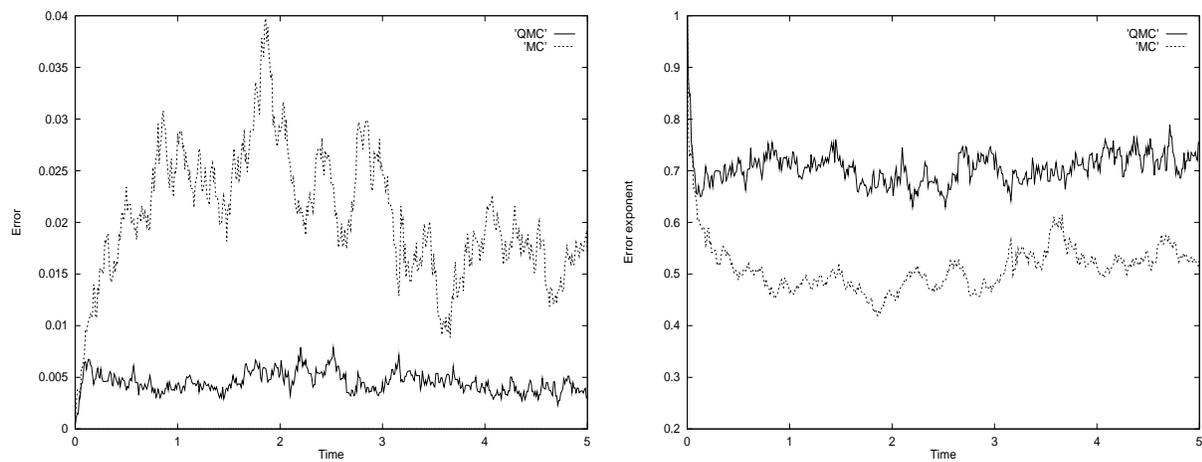


FIG. 3.2 – Dimension $s = 1$. Schémas MC et QMC: $N = 3^7 = 2\,187$ particules.

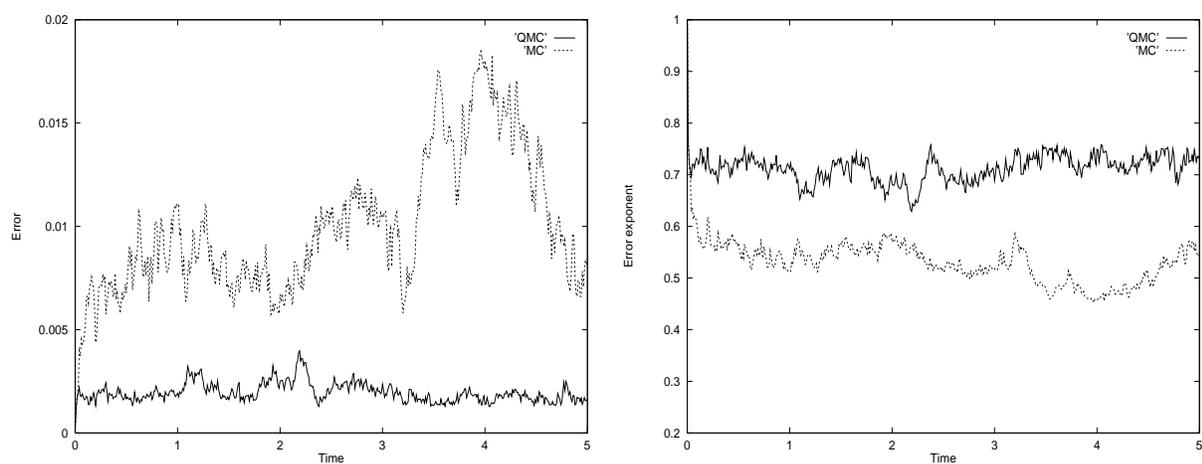


FIG. 3.3 – Dimension $s = 1$. Schémas MC et QMC: $N = 3^8 = 6\,561$ particules.

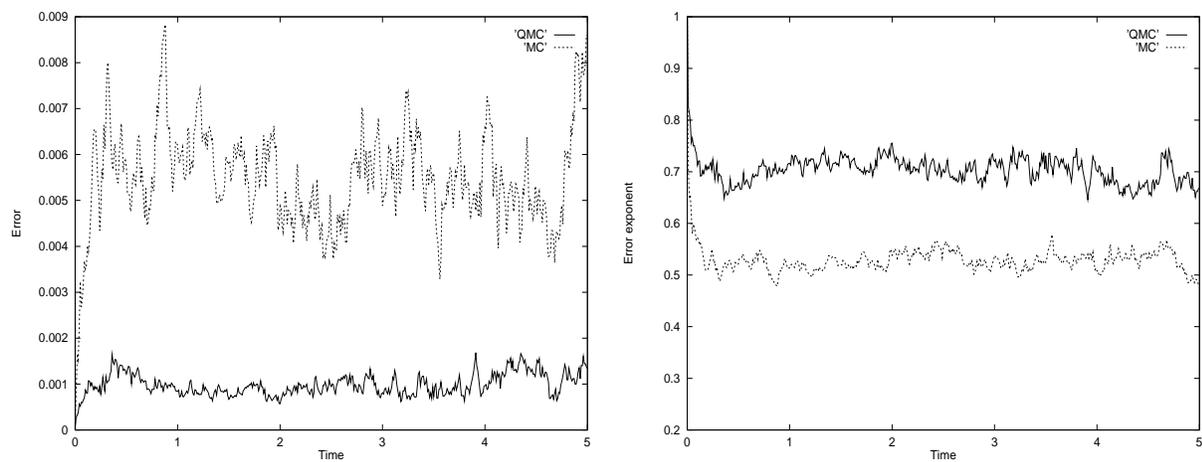


FIG. 3.4 – Dimension $s = 1$. Schémas MC et QMC: $N = 3^9 = 19\,683$ particules.

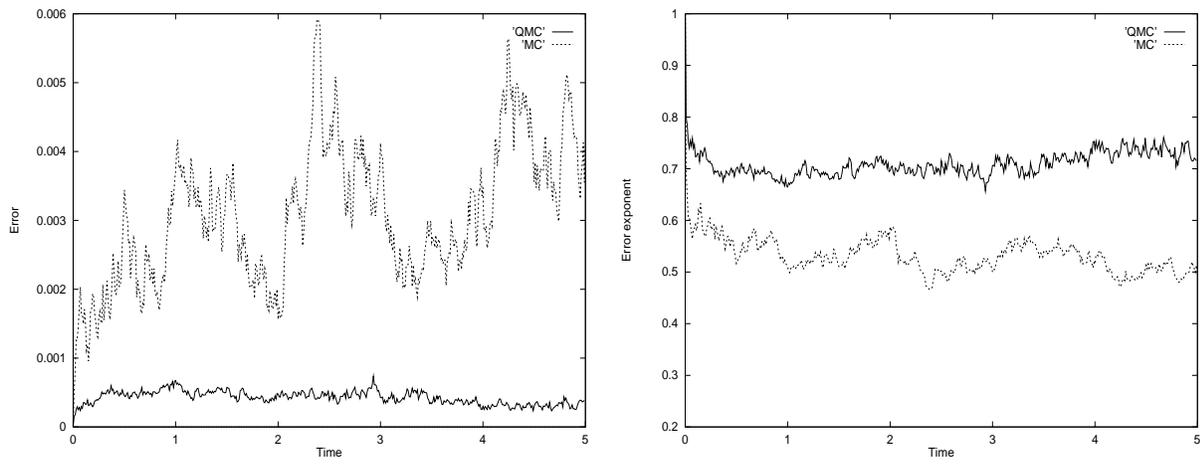


FIG. 3.5 – Dimension $s = 1$. Schémas MC et QMC: $N = 3^{10} = 59\,049$ particules.

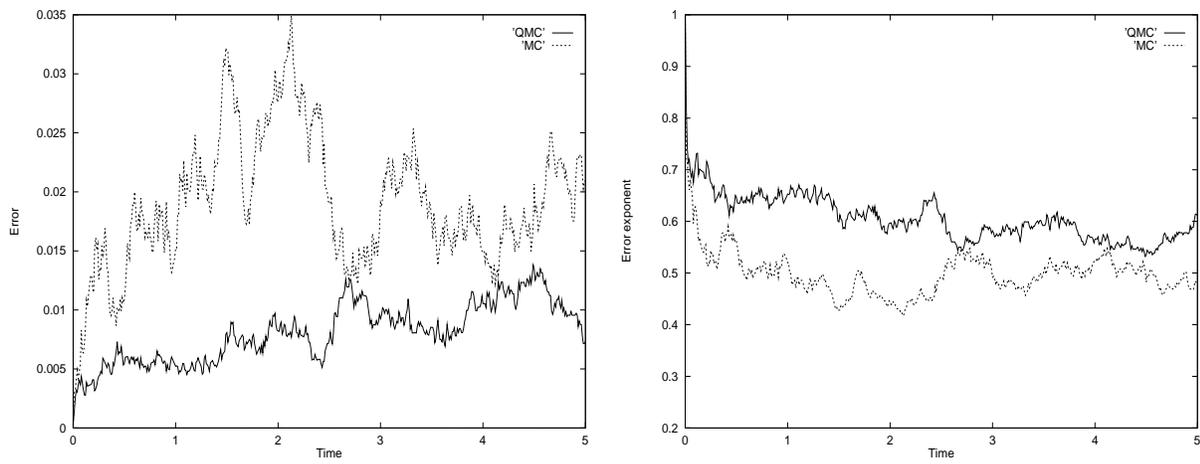


FIG. 3.6 – Dimension $s = 2$. Schémas MC et QMC: $N = 5^2 5^3 = 3\,125$ particules.

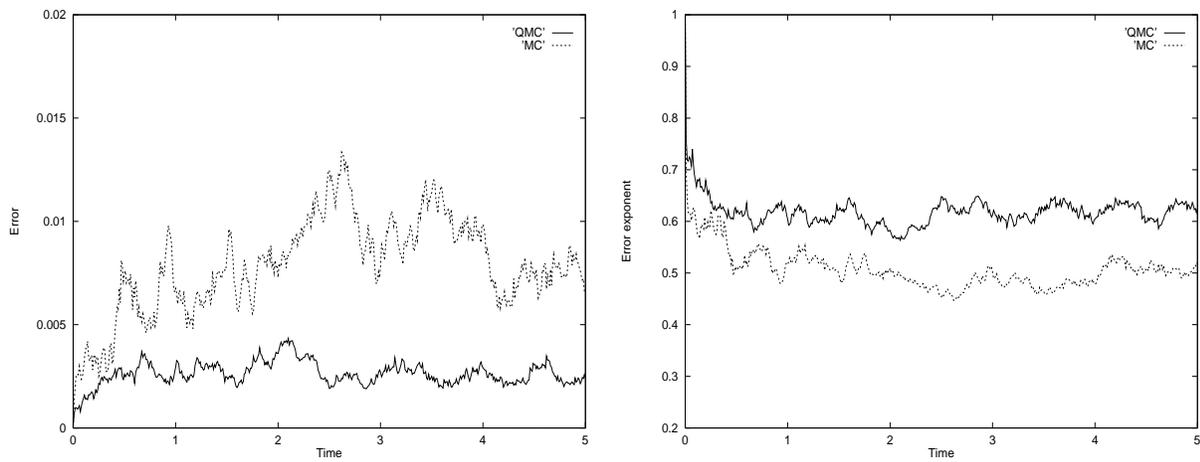


FIG. 3.7 – Dimension $s = 2$. Schémas MC et QMC: $N = 5^2 5^4 = 15\,625$ particules.

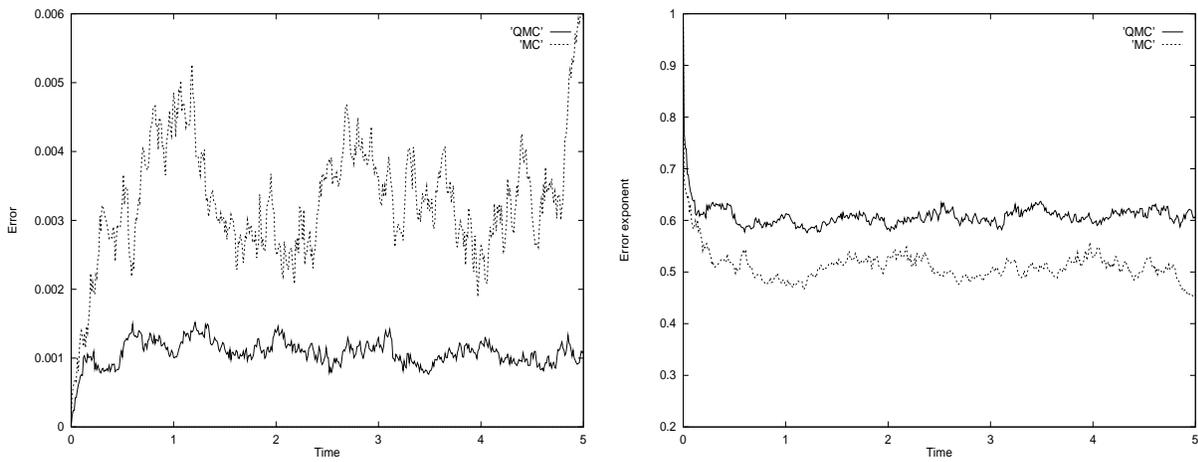


FIG. 3.8 – Dimension $s = 2$. Schémas MC et QMC: $N = 5^2 5^5 = 78\,125$ particules.

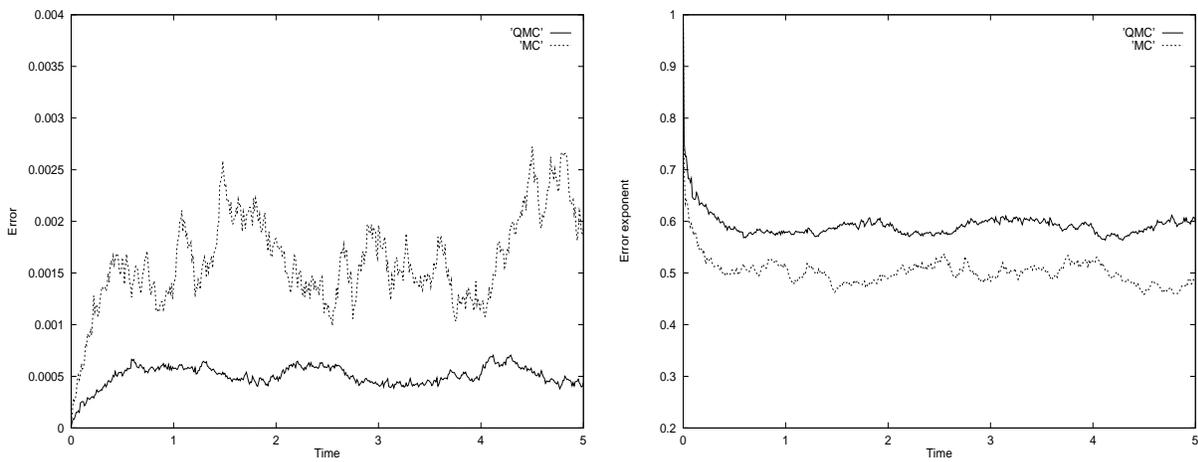


FIG. 3.9 – Dimension $s = 2$. Schémas MC et QMC: $N = 5^2 5^6 = 390\,625$ particules.

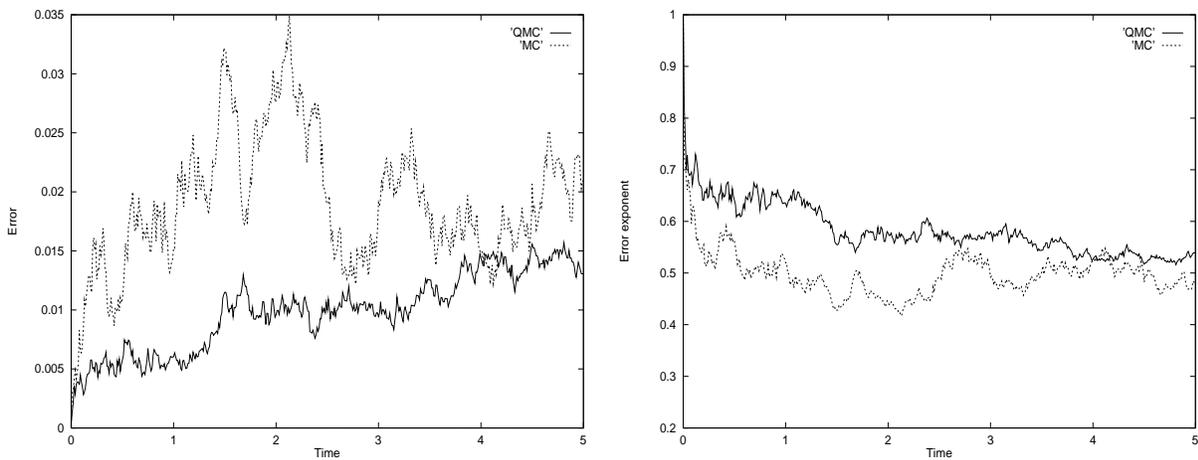


FIG. 3.10 – Dimension $s = 2$. Schémas MC et QMC: $N = 5^3 5^2 = 3\,125$ particules.

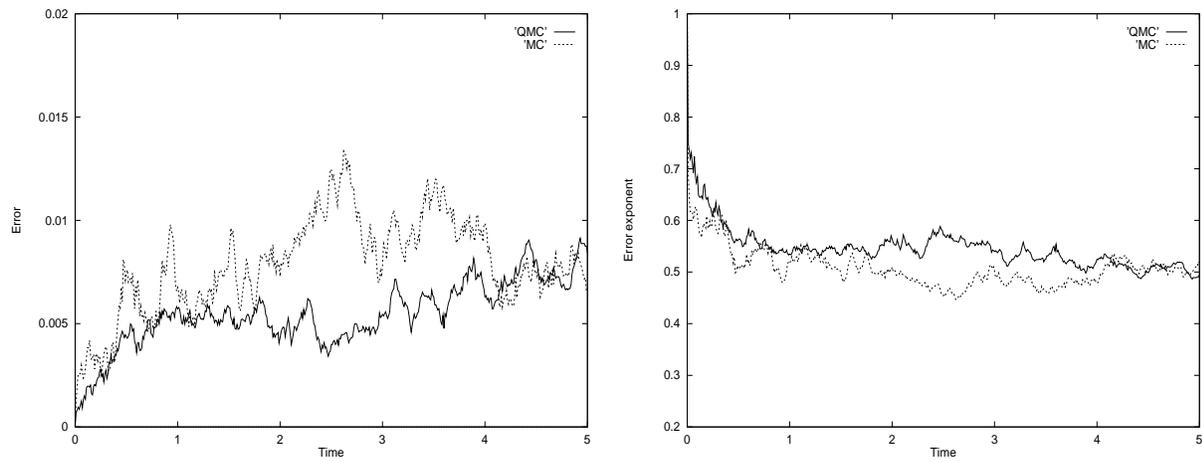


FIG. 3.11 – Dimension $s = 2$. Schémas MC et QMC: $N = 5^4 5^2 = 15\,625$ particules.

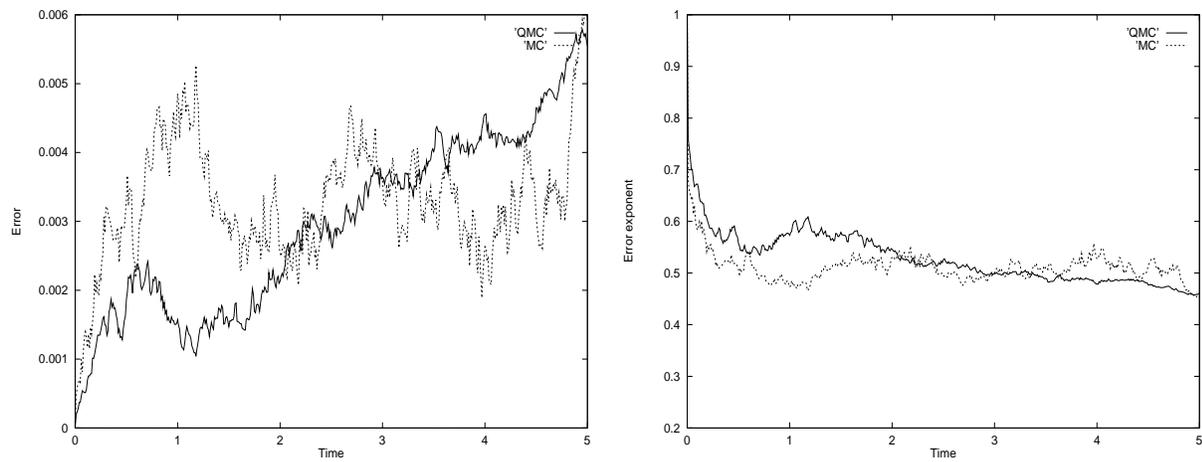


FIG. 3.12 – Dimension $s = 2$. Schémas MC et QMC: $N = 5^5 5^2 = 78\,125$ particules.

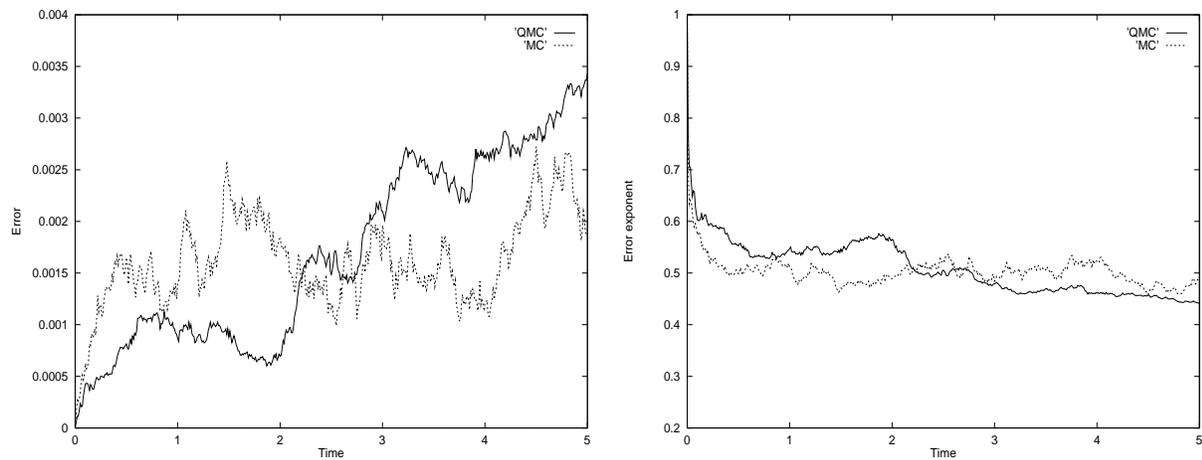


FIG. 3.13 – Dimension $s = 2$. Schémas MC et QMC: $N = 5^6 5^2 = 390\,625$ particules.

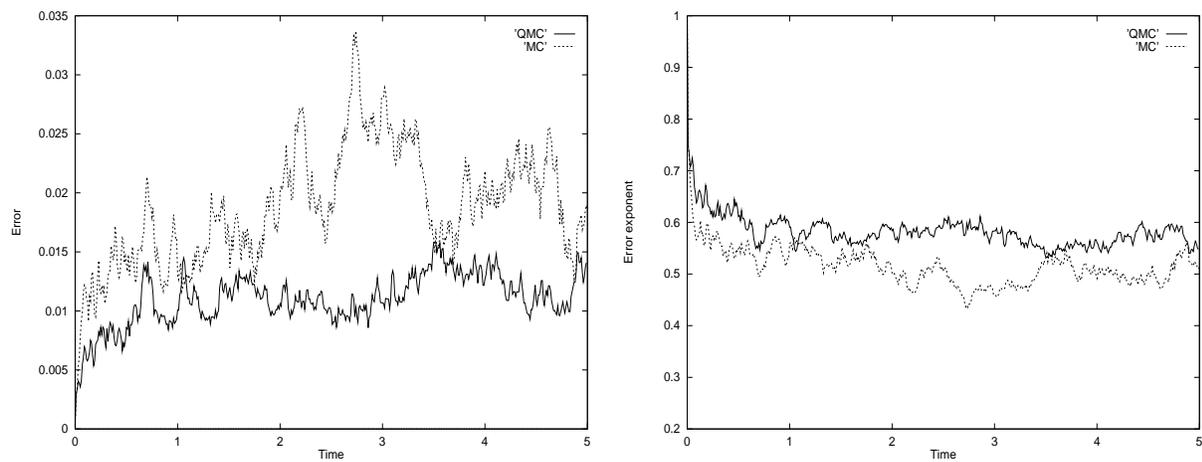


FIG. 3.14 – Dimension $s = 3$. Schémas MC et QMC: $N = 7^1 7^1 7^2 = 2\,401$ particules.

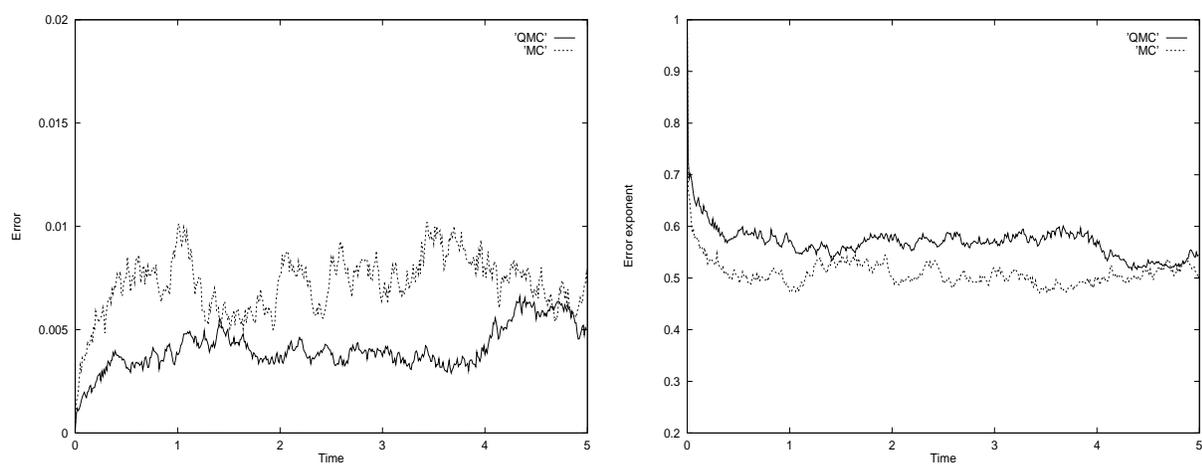


FIG. 3.15 – Dimension $s = 3$. Schémas MC et QMC: $N = 7^1 7^1 7^3 = 16\,807$ particules.

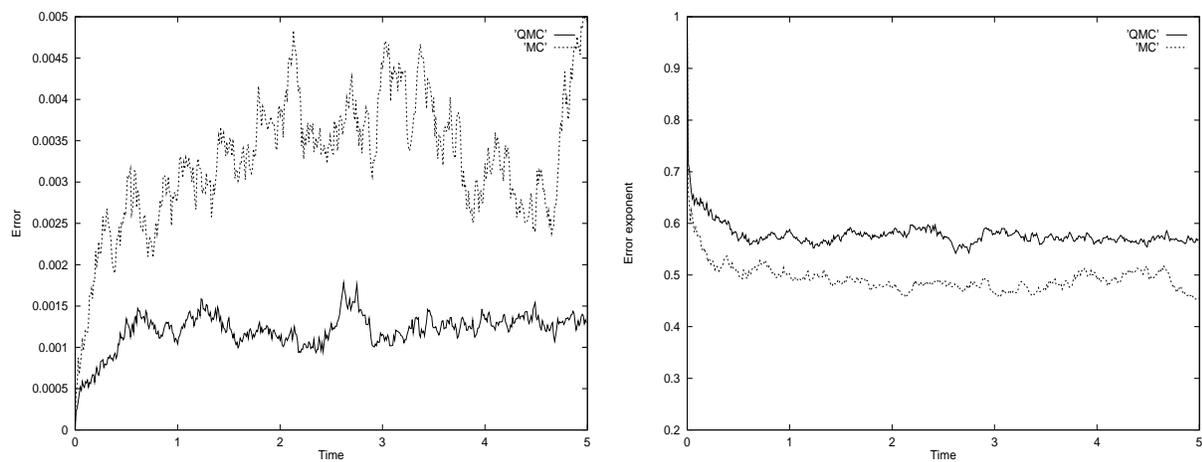


FIG. 3.16 – Dimension $s = 3$. Schémas MC et QMC: $N = 7^1 7^1 7^4 = 117\,649$ particules.

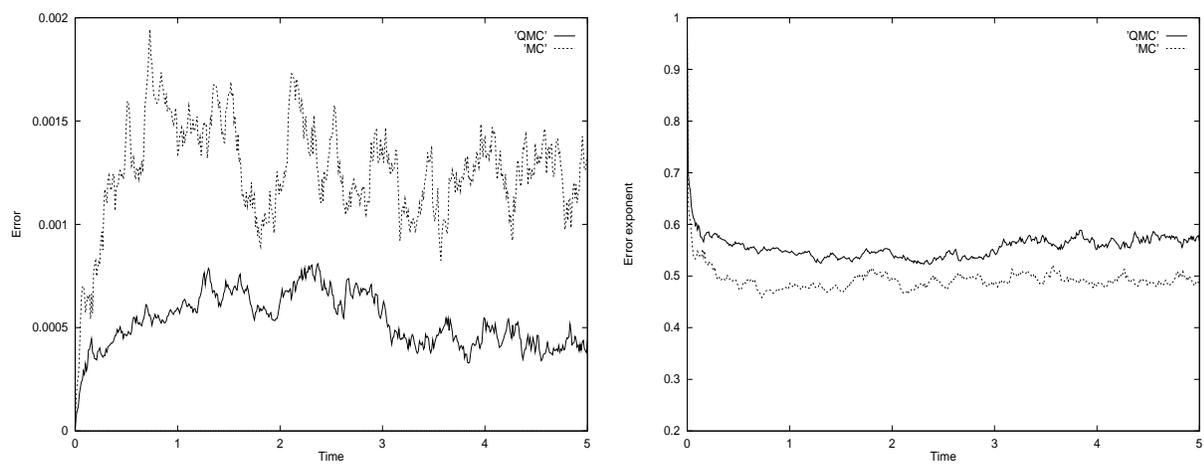


FIG. 3.17 – Dimension $s = 3$. Schémas MC et QMC: $N = 7^1 7^1 7^5 = 823\,543$ particules.

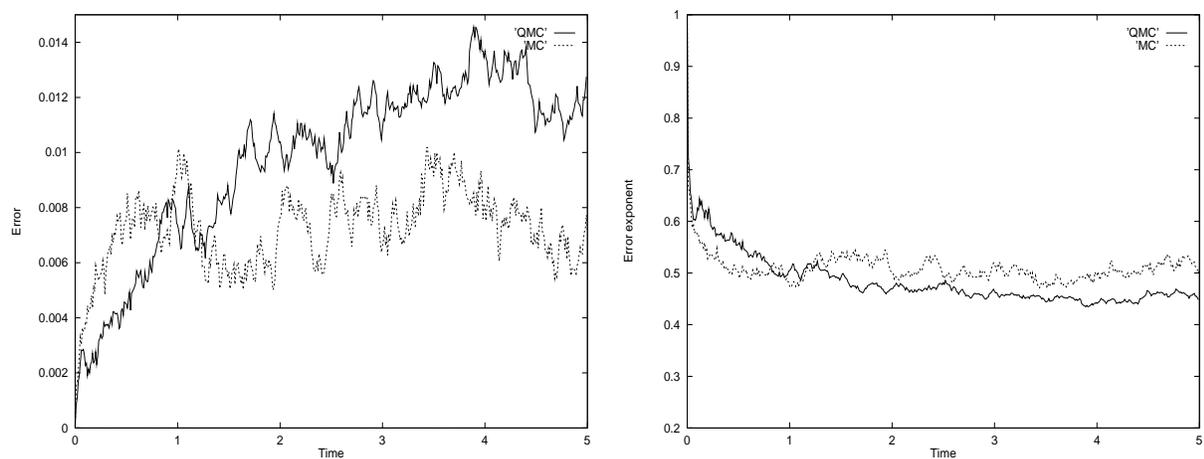


FIG. 3.18 – Dimension $s = 3$. Schémas MC et QMC: $N = 7^2 7^2 7^1 = 16\,807$ particules.

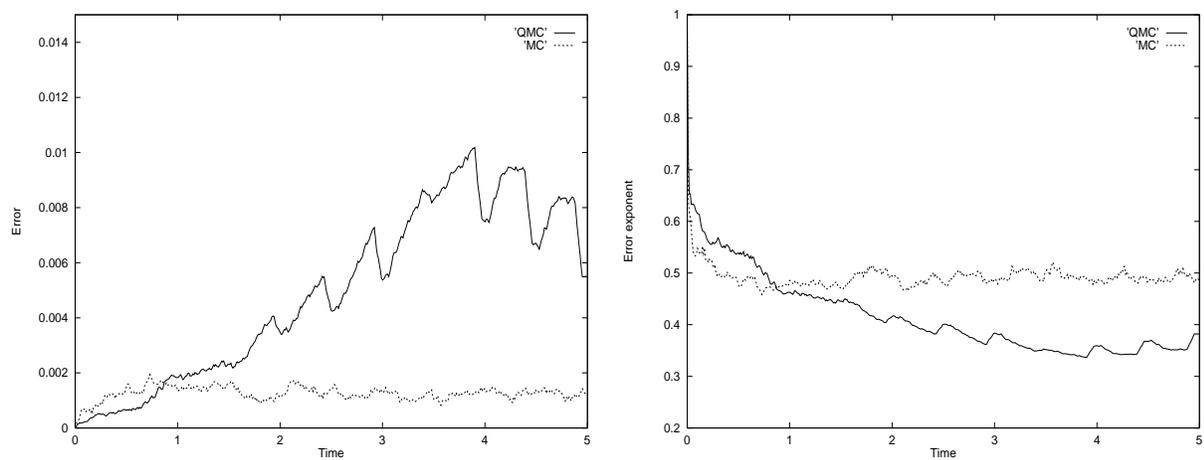


FIG. 3.19 – Dimension $s = 3$. Schémas MC et QMC: $N = 7^3 7^3 7^1 = 823\,543$ particules.

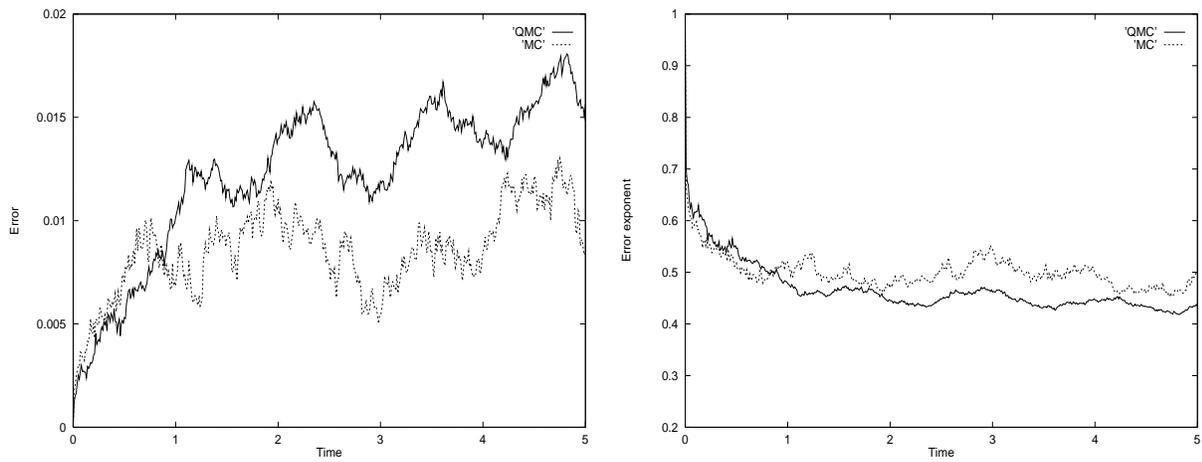


FIG. 3.20 – Dimension $s = 4$. Schémas MC et QMC: $N = 11^1 11^1 11^1 11^1 = 14\,641$ particules.

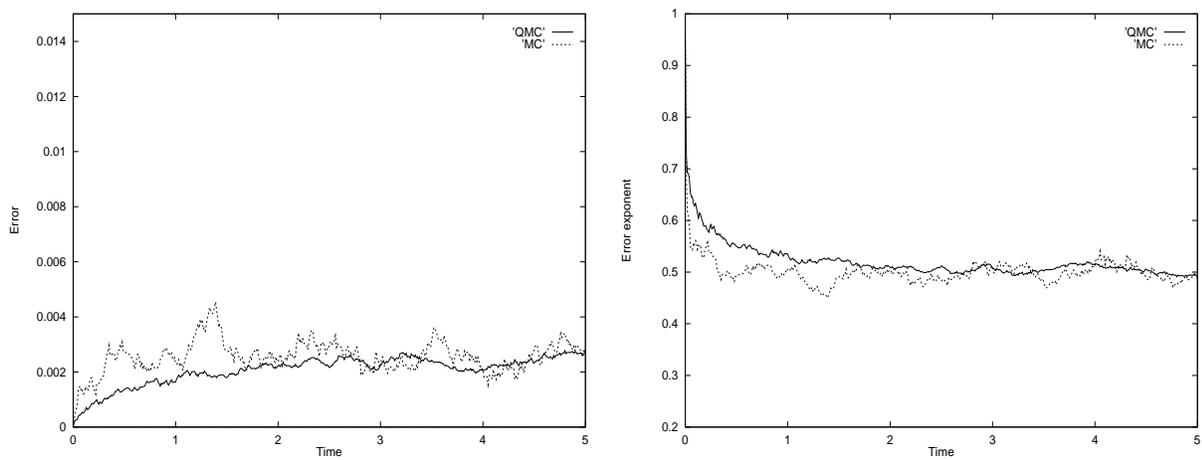


FIG. 3.21 – Dimension $s = 4$. Schémas MC et QMC: $N = 11^1 11^1 11^1 11^2 = 161\,051$ particules.

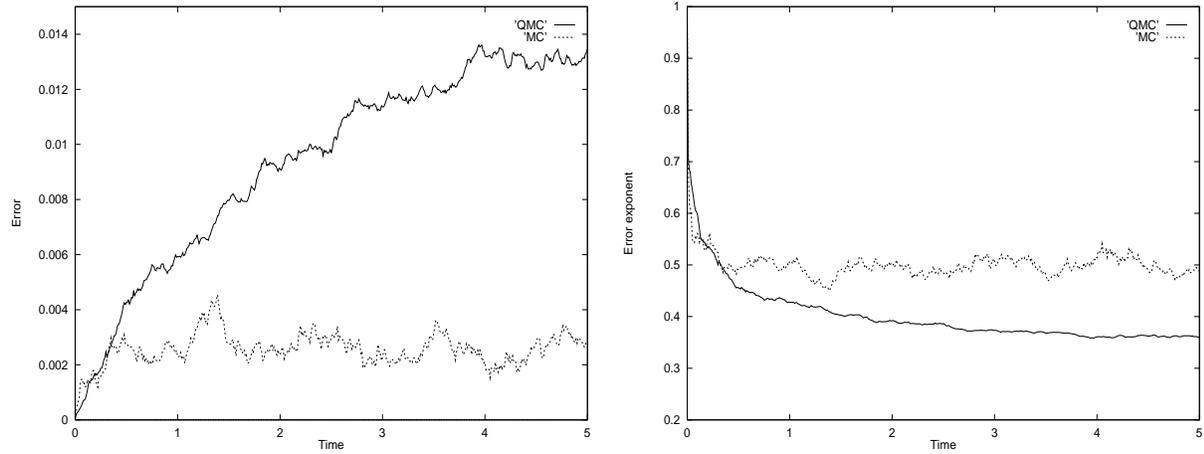


FIG. 3.22 – Dimension $s = 4$. Schémas MC et QMC: $N = 11^2 11^1 11^1 11^1 = 161\,051$ particules.

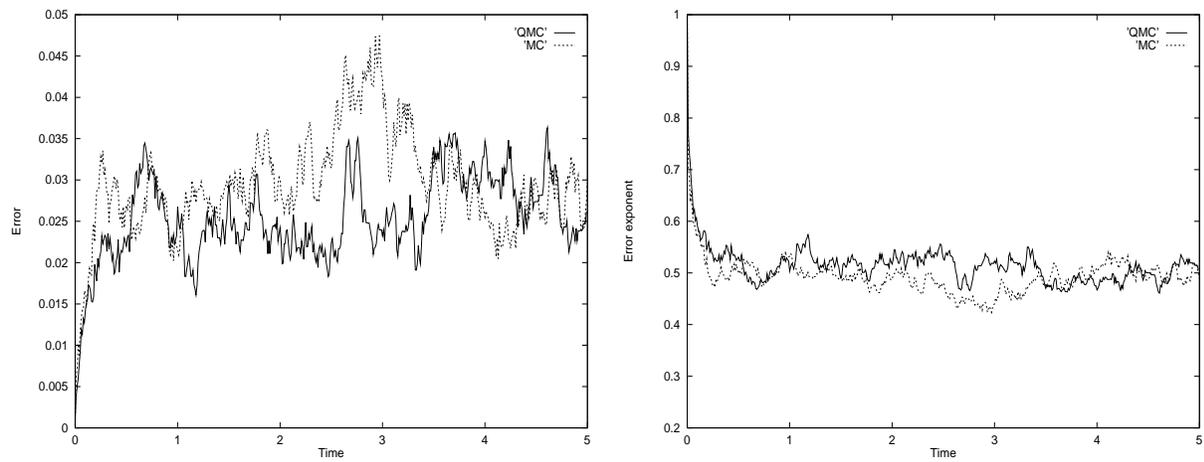


FIG. 3.23 – Dimension $s = 5$. Schémas MC et QMC: $N = 11^0 11^0 11^0 11^0 11^3 = 1\,331$ particules.

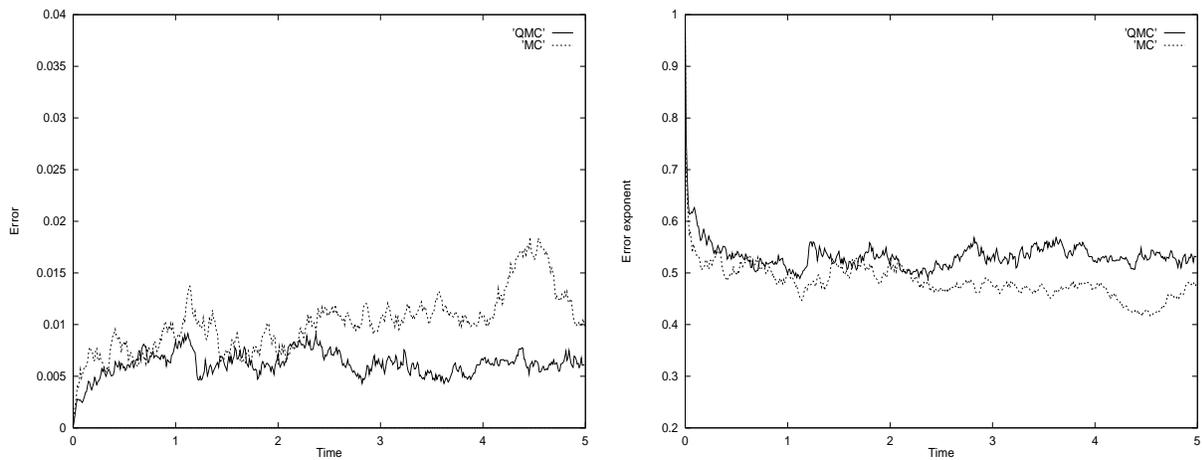


FIG. 3.24 – Dimension $s = 5$. Schémas MC et QMC: $N = 11^0 11^0 11^0 11^0 11^4 = 14\,641$ particules.

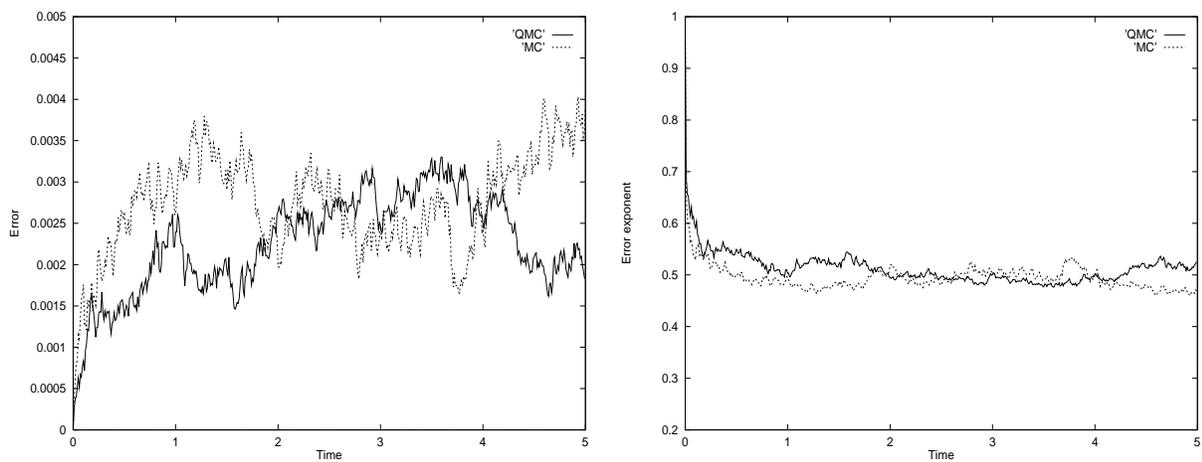


FIG. 3.25 – Dimension $s = 5$. Schémas MC et QMC: $N = 11^0 11^0 11^0 11^0 11^5 = 161\,051$ particules.

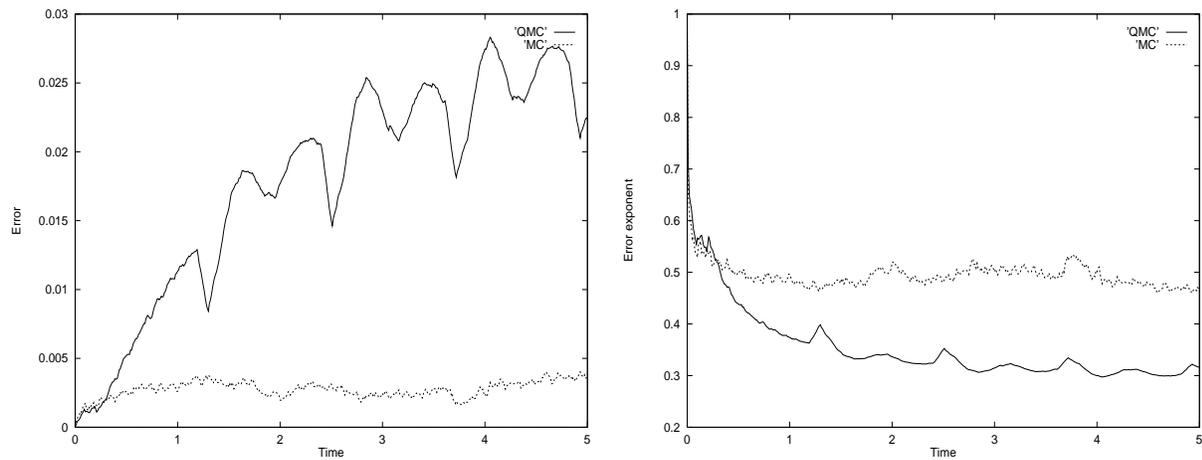


FIG. 3.26 – Dimension $s = 5$. Schémas MC et QMC: $N = 11^1 11^1 11^1 11^1 11^1 = 161\ 051$ particules.

3.6 Conclusion

Un nouvel algorithme quasi-Monte Carlo pour la simulation numérique d'une équation intégral-différentielle linéaire a été analysé dans ce chapitre. Le temps a été discrétisé et la solution a été approchée par une somme de mesures de Dirac (ou particules). Quand les particules sont renumérotées à chaque pas de temps on a prouvé la convergence de la méthode. Les résultats expérimentaux indiquent que le nouveau schéma est plus performant qu'un algorithme de Monte Carlo standard.

Les erreurs obtenues dans les tests numériques sont nettement meilleures que les bornes théoriques.

Cependant l'analyse théorique est une aide pour le choix des différents paramètres Γ et permet de développer un schéma quasi-Monte Carlo efficace.

Cette étude nous amène à nous poser un certain nombre de questions nécessitant de nouvelles recherches. La principale est la possibilité d'application de la méthode quasi-Monte Carlo à des modèles non-linéaires tels que l'équation de Boltzmann. L'extension de la méthode au modèle de Kac (unidimensionnel et non-linéaire) est traitée dans le chapitre suivant. Une autre question intéressante est la généralisation aux problèmes non homogènes en espace.

Chapitre 4

Méthodes particulières de simulation du modèle de Kac

Résumé: On considère un modèle mathématique simplifié de l'équation de Boltzmann, qui a été introduit par Kac. On décrit une méthode particulière quasi-Monte Carlo de simulation utilisant les réseaux- (t, m, s) . Le mouvement des particules est ramené à l'évaluation du volume d'un sous-ensemble de $I^4 = [0, 1]^4$. On démontre une estimation d'erreur pour une approximation quasi-Monte Carlo du volume d'un ensemble ayant la propriété du quadrant généralisée lorsqu'on utilise les réseaux- (t, m, s) . On prouve la convergence de la méthode dans le cas où les particules sont renumérotées à chaque pas de temps. Enfin on compare les méthodes particulières quasi-Monte Carlo et Monte Carlo de simulation dans des expériences numériques. Les résultats indiquent une convergence plus rapide pour la méthode quasi-Monte Carlo par rapport à des méthodes classiques de simulation Monte Carlo.

4.1 Introduction

Les méthodes particulières de simulation de type Monte Carlo sont couramment utilisées pour la résolution d'équations cinétiques intégral-différentielles. Des particules sont initialisées selon la distribution initiale. Elles agissent entre elles selon la dynamique décrite par l'équation. Des nombres pseudo-aléatoires sont utilisés pour choisir les particules entrant en collision et le résultat des collisions. Un gros effort a été entrepris au cours des dernières années [57, 25, 26, 31, 33, 38, 40, 41, 54, 60] pour développer des méthodes de type quasi-Monte Carlo. Ce sont des méthodes dans lesquelles les nombres aléatoires sont remplacés par des suites déterministes ayant une meilleure distribution. Une utilisation soigneuse de ces suites permet en général d'avoir des résultats meilleurs que ceux fournis par les suites pseudo-aléatoires.

Dans ce chapitre un algorithme quasi-Monte Carlo est proposé pour la simulation d'un processus non-linéaire. Un modèle simplifié de l'équation de Boltzmann a été introduit par Kac [24]. Dans ce modèle un gaz unidimensionnel est caractérisé par une densité de probabilité $f(v, t)$ des molécules ayant la vitesse v à l'instant t . Pour la simulation on fixe le nombre N de particules ainsi que le pas de temps Δt . La solution exacte $f(v, t_n)$ à

l'instant $t_n = n\Delta t$ est approchée par une somme de N mesures de Dirac

$$f^{(n)}(v) = \frac{1}{N} \sum_{j=0}^{N-1} \delta(v - v_j^{(n)}).$$

La méthode quasi-Monte Carlo de simulation est basée sur la théorie des réseaux- (t, m, s) et des suites- (t, s) . Les réseaux- (t, m, s) et les suites- (t, s) ont les meilleures propriétés d'équidistribution dans le cas où $t = 0$. Pour la transformation des réseaux- (t, m, s) On a les résultats suivants établis par Niederreiter [47] Lemmes 2.7 et 2.8].

LEMME 4.1 *Pour $1 \leq r \leq s$ soit i_1, \dots, i_r , des éléments distincts de l'ensemble $\{1, \dots, s\}$. L'application P définie par*

$$P(\mathbf{x}) = (x_{i_1}, \dots, x_{i_r}) \quad \text{pour} \quad \mathbf{x} = (x_1, \dots, x_s) \in I^s,$$

transforme tout réseau- (t, m, s) en base b en un réseau- (t, m, r) en base b .

LEMME 4.2 *Soit X un réseau- (t, m, s) en base b , J un intervalle élémentaire en base b avec $\lambda_s(J) = b^{-u}$, où $0 \leq u \leq m - t$, et T la transformation affine de J dans I^s . Alors les points de X appartenant à J sont transformés par T en un réseau- $(t, m - u, s)$ en base b .*

La méthode quasi-Monte Carlo consiste à ramener le problème de simulation à l'évaluation de volumes de sous-ensembles de I^4 . Si les particules sont réordonnées à chaque pas de temps de telle manière que $v_0^{(n)} \leq v_1^{(n)} \leq \dots \leq v_{N-1}^{(n)}$ On obtient une faible erreur d'intégration. On établit des bornes d'erreur pour des sous-ensembles particuliers de I^s . Pour cela On généralise une notion introduite par Schmidt [55] (voir aussi Schmidt [56] Chapitre 2] Section 13]). Un *ensemble ayant la propriété du quadrant* est un sous-ensemble E de I^s tel que pour $\mathbf{x} = (x_1, \dots, x_s) \in E$ l'intervalle $\prod_{i=1}^s [0, x_i]$ est contenu dans E .

Dans la Section 4.2 de ce chapitre On décrit la méthode particulière quasi-Monte Carlo de simulation pour le modèle de Kac de l'équation de Boltzmann. La Section 4.3 est consacrée aux méthodes de type Monte Carlo pour la simulation de l'équation de Kac. Dans la Section 4.4 On établit une borne d'erreur pour l'approximation du volume d'un ensemble ayant la propriété du quadrant généralisée puis on démontre la convergence de la méthode de simulation quasi-Monte Carlo. Dans la Section 4.5 On étudie un problème modèle dont la solution analytique est connue. Une étude comparative est faite sur ce problème entre la méthode quasi-Monte Carlo utilisant la renumérotation et les deux méthodes Monte Carlo décrites dans la Section 4.3. Enfin On indique dans la Section 4.6 des thèmes d'études pour les méthodes particulières quasi-Monte Carlo de simulation.

4.2 Une méthode quasi-Monte Carlo pour la simulation de l'équation de Kac

Dans [24] Chapitre 3] Section 16] Kac a introduit un modèle mathématique pour un gaz de molécules Maxwelliennes spatialement homogène et unidimensionnel en vitesse.

Ce modèle est gouverné par une équation de Boltzmann simplifiée: la distribution des vitesses est donnée par une fonction positive ou nulle définie sur $\mathbb{R} \times \mathbb{R}_+$ solution de

$$\frac{\partial f}{\partial t}(v, t) = \frac{\nu}{2\pi} \int_{-\infty}^{+\infty} \int_0^{2\pi} (f(v', t)f(w', t) - f(v, t)f(w, t)) dw d\theta, \quad v \in \mathbb{R}, t > 0, \quad (4.1)$$

$$f(v, 0) = f_0(v), \quad v \in \mathbb{R}, \quad (4.2)$$

où ν est une constante positive Γ

$$v' = v \cos \theta + w \sin \theta, \quad w' = -v \sin \theta + w \cos \theta,$$

et f_0 est une fonction positive ou nulle définie sur \mathbb{R} telle que

$$\int_{-\infty}^{+\infty} f_0(v) dv = 1. \quad (4.3)$$

Dans ce modèle la masse et l'énergie cinétique sont conservées mais pas la quantité de mouvement:

$$\forall t > 0 \quad \int_{-\infty}^{+\infty} f(v, t) dv = 1, \quad \int_{-\infty}^{+\infty} v^2 f(v, t) dv = \int_{-\infty}^{+\infty} v^2 f_0(v) dv. \quad (4.4)$$

Il est nécessaire de commencer par établir une formulation faible de l'équation 4.1. Soit $\mathcal{S}(\mathbb{R})$ l'ensemble des fonctions simples mesurables au sens de Borel sur \mathbb{R} . Multiplions l'équation 4.1 par $s \in \mathcal{S}(\mathbb{R})$ et intégrons sur \mathbb{R} . On a

$$\begin{aligned} & \forall s \in \mathcal{S}(\mathbb{R}) \quad \frac{d}{dt} \int_{-\infty}^{+\infty} s(v) f(v, t) dv \\ &= \nu \int_{\mathbb{R}^2 \times I} (s(v \cos 2\pi x - w \sin 2\pi x) - s(v)) f(v, t) f(w, t) dv dw dx. \end{aligned} \quad (4.5)$$

Soit N un entier. On choisit un ensemble $V^{(0)}$ de N points $v_0^{(0)}, \dots, v_{N-1}^{(0)}$ de \mathbb{R} tel que

$$f^{(0)}(v) = \frac{1}{N} \sum_{j=0}^{N-1} \delta(v - v_j^{(0)})$$

“approche” la mesure $f_0(v) dv$ (une technique de construction est donnée dans la Section 4.5). On fixe un pas de temps $\Delta t > 0$ vérifiant

$$\nu \Delta t \leq 1.$$

Cette condition assure la faisabilité du schéma (voir (iii) ci-dessous). On a besoin d'une suite $Y = \{\mathbf{y}_n, n \geq 0\}$ de points de I^4 pour l'intégration quasi-Monte Carlo. Pour tout entier $n \geq 1$ on note $t_n = n\Delta t$ et $f_n(v) = f(v, t_n)$. Les ensembles de points $V^{(n)} \subset \mathbb{R}$ et les mesures

$$f^{(n)}(v) = \frac{1}{N} \sum_{j=0}^{N-1} \delta(v - v_j^{(n)})$$

sont construits par la procédure suivante.

(i) Renumerotation des particules:

Les vitesses sont réordonnées de telle manière que

$$i < j \Rightarrow v_i^{(n)} \leq v_j^{(n)}.$$

Cette procédure joue un rôle important dans la démonstration de la convergence.

(ii) Schéma d'Euler:

En remplaçant la dérivée en temps par une différence finie dans la formulation faible Γ on obtient une mesure $g^{(n+1)}(v)$ qui est une approximation de la distribution exacte $f_{n+1}(v)dv$:

$$\begin{aligned} \forall s \in \mathcal{S}(\mathbb{R}) \quad & \frac{1}{\Delta t} \int_{-\infty}^{+\infty} s(v) \left(g^{(n+1)}(v) - f^{(n)}(v) \right) \\ & = \nu \int_{\mathbb{R}^2 \times I} \left(s(v \cos 2\pi x - w \sin 2\pi x) - s(v) \right) f^{(n)}(v) f^{(n)}(w) dx, \end{aligned}$$

d'où

$$\begin{aligned} \int_{-\infty}^{+\infty} s(v) g^{(n+1)}(v) \quad & = \frac{1 - \nu \Delta t}{N} \sum_{j=0}^{N-1} s(v_j^{(n)}) \\ & + \frac{\nu \Delta t}{N^2} \sum_{j,k=0}^{N-1} \int_0^1 s(v_j^{(n)} \cos 2\pi x - v_k^{(n)} \sin 2\pi x) dx. \end{aligned} \tag{4.6}$$

La mesure $g^{(n+1)}(v)$ n'étant pas sous la forme d'une somme de mesure de Dirac Γ on a besoin d'une étape supplémentaire qui consiste à faire une quadrature quasi-Monte Carlo.

(iii) Approximation quasi-Monte Carlo:

On réécrit le second membre de l'équation (4.6) sous la forme d'une intégrale en dimension 4. Pour $0 \leq j, k < N$ on note $c_{j,k}$ la fonction caractéristique de l'intervalle

$$\left[\frac{j}{N}, \frac{j+1}{N} \right) \times \left[\frac{k}{N}, \frac{k+1}{N} \right)$$

et χ la fonction caractéristique de $[0, \nu \Delta t)$. A chaque fonction simple s sur \mathbb{R} correspond une fonction simple $S^{(n)}$ sur \bar{T}^4 par

$$\begin{aligned} \forall \mathbf{x} = (x_1, x_2, x_3, x_4) \in \bar{T}^4 \quad S^{(n)}(\mathbf{x}) & = \sum_{j,k=0}^{N-1} c_{j,k}(x_1, x_2) \\ & \cdot \left((1 - \chi(x_3)) s(v_j^{(n)}) + \chi(x_3) s(v_j^{(n)} \cos 2\pi x_4 - v_k^{(n)} \sin 2\pi x_4) \right). \end{aligned} \tag{4.7}$$

On a

$$\int_{-\infty}^{+\infty} s(v)g^{(n+1)}(v) = \int_{I^4} S^{(n)}(\mathbf{x})d\mathbf{x}. \quad (4.8)$$

Alors on définit une approximation $f^{(n+1)}(v)$ de $f_{n+1}(v)dv$ par une quadrature quasi-Monte Carlo avec l'ensemble de points $\{\mathbf{y}_{nN+j} : 0 \leq j < N\} \subset I^4$.

$$\int_{-\infty}^{+\infty} s(v)f^{(n+1)}(v) = \frac{1}{N} \sum_{j=0}^{N-1} S^{(n)}(\mathbf{y}_{nN+j}), \quad s \in \mathcal{S}(\mathbb{R}). \quad (4.9)$$

Notons $\lfloor x \rfloor$ la partie entière du nombre réel x et

$$k_1^{(n)}(j) = \lfloor Ny_{nN+j,1} \rfloor, \quad k_2^{(n)}(j) = \lfloor Ny_{nN+j,2} \rfloor.$$

Si chaque cellule $\left[\frac{j}{N}, \frac{j+1}{N}\right) \times \left[\frac{k}{N}, \frac{k+1}{N}\right) \times I^2$ contient un seul élément de l'ensemble de points utilisé dans la quadrature quasi-Monte Carlo l'équation (4.9) devient

$$\begin{aligned} \frac{1}{N} \sum_{j=0}^{N-1} s(v_j^{(n+1)}) &= \frac{1}{N} \sum_{j=0}^{N-1} \\ &\left((1 - \chi(y_{nN+j,3}))s(v_{k_1^{(n)}(j)}^{(n)}) + \chi(y_{nN+j,3})s(v_{k_1^{(n)}(j)}^{(n)}) \cos 2\pi y_{nN+j,4} - v_{k_2^{(n)}(j)}^{(n)} \sin 2\pi y_{nN+j,4} \right). \end{aligned}$$

Alors on a l'un des deux cas suivants pour chaque nouvelle vitesse:

$$\begin{aligned} v_j^{(n+1)} &= v_{k_1^{(n)}(j)}^{(n)} \cos 2\pi y_{nN+j,4} - v_{k_2^{(n)}(j)}^{(n)} \sin 2\pi y_{nN+j,4}, \quad \text{si } y_{nN+j,3} < \nu \Delta t, \\ v_j^{(n+1)} &= v_{k_1^{(n)}(j)}^{(n)}, \quad \text{sinon.} \end{aligned} \quad (4.10)$$

La masse est automatiquement conservée. L'énergie cinétique est conservée et on a de faibles erreurs pour les quadratures quasi-Monte Carlo si Y possède des propriétés de symétrie assurant la conservation de l'énergie pour les particules entrant en collision. Soit $b \geq 2$ et $m \geq 0$ des entiers et prenons le nombre de particules de la forme $N = 2b^m$. Soit $X = \{\mathbf{x}_n, n \geq 0\}$ une suite $(0, 4)$ en base b . Pour $nb^m \leq \ell < (n+1)b^m$ on pose

$$\begin{aligned} \mathbf{y}_{nb^m+\ell} &= \left(\frac{1}{2} \left(x_{\ell,1} + \frac{\lfloor b^m x_{\ell,1} \rfloor + 1}{b^m} \right), \frac{1}{2} \left(x_{\ell,2} + \frac{\lfloor b^m x_{\ell,2} \rfloor}{b^m} \right), x_{\ell,3}, x_{\ell,4} \right), \\ \mathbf{y}_{(n+1)b^m+\ell} &= \left(\frac{1}{2} \left(x_{\ell,2} + \frac{\lfloor b^m x_{\ell,2} \rfloor}{b^m} \right), \frac{1}{2} \left(x_{\ell,1} + \frac{\lfloor b^m x_{\ell,1} \rfloor + 1}{b^m} \right), x_{\ell,3}, 1 - x_{\ell,4} \right). \end{aligned}$$

Les ensembles de points $Y_1^{(n)} = \{\mathbf{y}_{nb^m+\ell} : nb^m \leq \ell < (n+1)b^m\}$ et $Y_2^{(n)} = \{\mathbf{y}_{(n+1)b^m+\ell} : nb^m \leq \ell < (n+1)b^m\}$ sont des réseaux $(0, m, 4)$ en base b .

Les deux premières composantes de la suite $Y_1^{(n)}$ permettent de choisir les particules $k_1^{(n)}(j)$ et $k_2^{(n)}(j)$ qui vont entrer en collision. La première composante choisit un numéro

de particule impair tandis que la deuxième composante choisit un numéro de particule pair. A ce niveau les nouvelles vitesses des particules de numéro impair sont calculées. Avec la suite $Y_2^{(n)}$ les mêmes particules sont choisies mais en ordre inverse c'est à dire un numéro pair puis un numéro impair.

Ainsi aucune particule ne peut entrer en collision avec elle même. D'autre part chaque particule peut faire tout au plus une collision durant un pas de temps.

4.3 Méthodes Monte Carlo pour la résolution de l'équation de Kac

4.3.1 Méthode de Bird

Cette méthode basée sur la physique du problème et non pas sur une discrétisation de l'équation de Boltzmann a été introduite par Bird (voir [2]).

Soit N le nombre de particules simulées et Δt le pas de temps choisi de manière à satisfaire

$$\nu \Delta t < 1.$$

La probabilité de collision pour une particule durant un intervalle de temps (t_n, t_{n+1}) est

$$P_c = \nu \Delta t.$$

La méthode consiste à

- tirer au hasard $\lfloor \frac{N}{2} P_c \rfloor$ paires de particules qui entrent en collision.
- tirer au hasard dans $[0, 2\pi]$ un angle $\theta_{i,j}$ pour chaque paire de particules (i, j) entrant en collision. Les nouvelles vitesses sont alors:

$$\begin{aligned} v_i^{(n+1)} &= v_i^{(n)} \cos \theta_{i,j} + v_j^{(n)} \sin \theta_{i,j} \\ v_j^{(n+1)} &= -v_i^{(n)} \sin \theta_{i,j} + v_j^{(n)} \cos \theta_{i,j}. \end{aligned}$$

La méthode de Bird conserve l'énergie cinétique.

4.3.2 Méthode de Nanbu

C'est une méthode basée sur une discrétisation de l'équation de Boltzmann (voir [43]). Soit N le nombre de particules simulées. Tout d'abord on utilise un schéma d'Euler en temps

$$f^{(n+1)}(v) = f^{(n)}(v) + \frac{\nu \Delta t}{2\pi} \int_{-\infty}^{+\infty} \int_0^{2\pi} (f^{(n)}(v') f^{(n)}(w') - f^{(n)}(v) f^{(n)}(w)) dw d\theta. \quad (4.11)$$

avec un pas de temps Δt vérifiant

$$\nu \Delta t \leq 1.$$

Grâce à l'approximation particulière

$$f^{(n)}(v) = \frac{1}{N} \sum_{i=0}^{N-1} \delta(v - v_i^{(n)}),$$

on a

$$f^{(n+1)}(v) = \frac{1}{N} \sum_{i=0}^{N-1} \left((1 - \nu \Delta t) \delta(v - v_i^{(n)}) + \nu \Delta t g_i^{(n)}(v) \right), \quad (4.12)$$

avec

$$g_i^{(n)}(v) = \frac{1}{2\pi N} \sum_{j=0}^{N-1} \int_{-\infty}^{+\infty} \int_0^{2\pi} \delta(v' - v_i^{(n)}) \delta(w' - v_j^{(n)}) dw' d\theta.$$

Cette distribution $g_i^{(n)}$ se réécrit

$$g_i^{(n)}(v) = \frac{1}{N} \sum_{j=0}^{N-1} \int_0^1 \delta(v - v_i^{(n)} \cos 2\pi x - v_j^{(n)} \sin 2\pi x) dx.$$

On vérifie que $g_i^{(n)}$ est une densité de probabilité:

$$\int_{-\infty}^{+\infty} g_i^{(n)}(v) dv = 1.$$

Soit $\tilde{v}_i^{(n)}$ une valeur de la variable aléatoire ayant pour densité de probabilité $g_i^{(n)}$. Alors $\tilde{v}_i^{(n)}$ est de la forme

$$\tilde{v}_i^{(n)} = v_i^{(n)} \cos 2\pi x - v_j^{(n)} \sin 2\pi x,$$

j et x étant des variables uniformes respectivement à valeurs dans $\{0, 1, \dots, N-1\}$ et $[0, 1]$. Par conséquent

$$f^{(n+1)}(v) = \frac{1}{N} \sum_{i=0}^{N-1} \left((1 - p) \delta(v - v_i^{(n)}) + p \delta(v - \tilde{v}_i^{(n)}) \right) \quad (4.13)$$

où $p = \nu \Delta t$ est la probabilité que la particule i entre en collision avec la particule j dans l'intervalle de temps (t_n, t_{n+1}) ; sa nouvelle vitesse est alors $\tilde{v}_i^{(n)}$.

Avec cette méthode l'énergie cinétique n'est pas conservée.

4.4 Analyse d'erreur pour la méthode quasi-Monte Carlo de simulation

La discrédance d'un ensemble de points est une mesure de la déviation par rapport à la distribution uniforme. On peut également considérer la discrédance par rapport à une distribution autre que la distribution uniforme (voir [20]). Pour $z \in \mathbb{R}$ soit s_z la fonction caractéristique de $(-\infty, z)$ et

$$d_N^{(n)}(z) = \frac{1}{N} \sum_{j=0}^{N-1} s_z(v_j^{(n)}) - \int_{-\infty}^{+\infty} s_z(v) f_n(v) dv.$$

La *discrédance à l'origine* de $V^{(n)}$ par rapport à f_n est le nombre

$$\mathbf{D}_N^*(V^{(n)}, f_n) = \sup_{z \in \mathbb{R}} |d_N^{(n)}(z)|.$$

C'est une mesure de la différence entre la distribution des particules de $V^{(n)}$ et la solution exacte à l'instant t_n . L'analyse de l'erreur est semblable à celle de l'erreur d'un schéma d'Euler. L'*erreur de troncature locale* est définie par

$$\begin{aligned} \varepsilon^{(n)}(z) &= \frac{1}{\Delta t} \int_{-\infty}^{+\infty} s_z(v) (f_{n+1}(v) - f_n(v)) dv \\ &\quad - \nu \int_{\mathbb{R}^2 \times I} (s_z(v \cos 2\pi x - w \sin 2\pi x) - s_z(v)) f_n(v) f_n(w) dv dw dx. \end{aligned}$$

On a également besoin d'un *terme d'erreur*

$$\begin{aligned} e_N^{(n)}(z) &= \frac{1}{N^2} \sum_{j,k=0}^{N-1} \int_I s_z(v_j^{(n)} \cos 2\pi x - v_k^{(n)} \sin 2\pi x) dx \\ &\quad - \int_{\mathbb{R}^2 \times I} s_z(v \cos 2\pi x - w \sin 2\pi x) f_n(v) f_n(w) dv dw dx. \end{aligned}$$

Soit $S_z^{(n)}$ la fonction simple sur $\bar{I}^4 \Gamma$ qui correspond à s_z . L'*erreur de la quadrature quasi-Monte Carlo* est donnée par

$$\delta_N^{(n)}(z) = \frac{1}{N} \sum_{j=0}^{N-1} S_z^{(n)}(\mathbf{y}_{nN+j}) - \int_{I^4} S_z^{(n)}(\mathbf{x}) d\mathbf{x}.$$

On a la formule de récurrence

$$d_N^{(n)}(z) = (1 - \nu \Delta t) d_N^{(n-1)}(z) + \nu \Delta t e_N^{(n-1)}(z) - \Delta t \varepsilon^{(n-1)}(z) + \delta_N^{(n-1)}(z). \quad (4.14)$$

PROPOSITION 4.1 *Le terme d'erreur $e_N^{(n)}(z)$ vérifie*

$$|e_N^{(n)}(z)| \leq 2\mathbf{D}_N^*(V^{(n)}, f_n). \quad (4.15)$$

Démonstration. Le terme $e_N^{(n)}(z)$ peut s'écrire

$$e_N^{(n)}(z) = \frac{1}{N} \sum_{k=0}^{N-1} \int_I d_N^{(n)}(z; x, v_k^{(n)}) dx + \int_{\mathbb{R} \times I} d_N^{(n)}(z; v, x) f_n(v) dv dx, \quad (4.16)$$

où

$$\begin{aligned} d_N^{(n)}(z; x, w) &= \frac{1}{N} \sum_{j=0}^{N-1} s_z(v_j^{(n)} \cos 2\pi x - w \sin 2\pi x) \\ &\quad - \int_{-\infty}^{+\infty} s_z(v \cos 2\pi x - w \sin 2\pi x) f_n(v) dv, \end{aligned}$$

$$d_N^{(n)}(z; v, x) = \frac{1}{N} \sum_{k=0}^{N-1} s_z(v \cos 2\pi x - v_k^{(n)} \sin 2\pi x) - \int_{-\infty}^{+\infty} s_z(v \cos 2\pi x - w \sin 2\pi x) f_n(w) dw.$$

Puisque

$$|d_N^{(n)}(z; x, w)| \leq \mathbf{D}_N^*(V^{(n)}, f_n), \quad |d_N^{(n)}(z; v, x)| \leq \mathbf{D}_N^*(V^{(n)}, f_n), \quad (4.17)$$

il vient $|e_N^{(n)}(z)| \leq 2\mathbf{D}_N^*(V^{(n)}, f_n)$. ■

L'inégalité suivante est immédiate

$$|\varepsilon^{(n)}(z)| \leq \int_{t_n}^{t_{n+1}} \int_{-\infty}^{+\infty} \left| \frac{\partial^2 f}{\partial t^2}(v, t) \right| dv dt. \quad (4.18)$$

Considérons maintenant l'erreur $\delta_N^{(n)}(z)$ introduite par la quadrature quasi-Monte Carlo. On fait des évaluations d'intégrales de fonctions caractéristiques. Si l'on pose

$$X^{(n)} = \{\mathbf{x}_{nb^{m+j}} : 0 \leq j < b^m\} \quad \text{et} \quad Y^{(n)} = \{\mathbf{y}_{nN+j} : 0 \leq j < N\}.$$

alors $X^{(n)}$ est un réseau-(0, m , 4) en base b et on peut écrire

$$\begin{aligned} \delta_N^{(n)}(z) &= \frac{A(E_z^{(n)}, Y^{(n)})}{N} - \lambda_4(E_z^{(n)}) - \frac{A(F_z^{(n)}, Y^{(n)})}{N} + \lambda_4(F_z^{(n)}) \\ &\quad + \frac{A(G_z^{(n)}, Y^{(n)})}{N} - \lambda_4(G_z^{(n)}), \end{aligned} \quad (4.19)$$

où

$$\begin{aligned} E_z^{(n)} &= \left\{ \mathbf{x} \in \bar{T}^4 : \sum_{j,k=0}^{N-1} c_{j,k}(x_1, x_2) s_z(v_j^{(n)}) = 1 \right\}, \\ F_z^{(n)} &= \left\{ \mathbf{x} \in \bar{T}^4 : \sum_{j,k=0}^{N-1} c_{j,k}(x_1, x_2) \chi(x_3) s_z(v_j^{(n)}) = 1 \right\}, \\ G_z^{(n)} &= \left\{ \mathbf{x} \in \bar{T}^4 : \sum_{j,k=0}^{N-1} c_{j,k}(x_1, x_2) \chi(x_3) s_z(v_j^{(n)}) \cos 2\pi x_4 - v_k^{(n)} \sin 2\pi x_4 = 1 \right\}. \end{aligned}$$

La quantité $\delta_N^{(n)}(z)$ est une somme d'erreurs introduites par le calcul approché de mesures de sous-ensembles de \bar{T}^4 . L'inégalité classique de Koksma-Hlawka ne peut pas être utilisée pour majorer $|\delta_N^{(n)}(z)|$. Soit E un sous-ensemble de \bar{T}^s et X un ensemble de N points de I^s . Une technique d'analyse de l'erreur

$$D(E, X) = \frac{A(E, X)}{N} - \lambda_s(E)$$

à été proposée par Niederreiter et Wills (voir le théorème 1.8). Si E a une forme régulière $|\delta_N^{(n)}(z)|$ est majoré par $\mathbf{D}_N(X)^{\frac{1}{s}}$. Le résultat de Niederreiter et Wills a été utilisé par Lécot

dans [33] pour obtenir une borne d'erreur après des calculs assez complexes. Dans le cas où X est un réseau- (t, m, s) en base $b\Gamma$ $|D(E, X)|$ peut être borné par $b^{\lfloor \frac{m-t}{s} \rfloor}$ (voir le théorème 1.7). Une méthode directe est utilisée ici pour majorer $|\delta_N^{(n)}(z)|$. Un réseau- (t, m, s) étant un ensemble de points pour lequel la discrédance locale $D(E, X)$ est nulle sur beaucoup de sous-intervalles de I^s l'utilisation des réseaux est essentielle. Par exemple $D(E_z^{(n)}, Y^{(n)}) = 0$. Une autre notion fondamentale est celle des ensembles ayant la propriété du quadrant généralisée (voir ci-dessous). Nous allons borner $D(E, X)$ dans le cas où E est un ensemble ayant la propriété du quadrant généralisée. Ce résultat nous servira ensuite à majorer $D(G_z^{(n)}, Y^{(n)})$ après avoir décomposé $G_z^{(n)}$.

L'ensemble $X^{(n)}$ étant un réseau- $(0, m, 4)$ en base $b\Gamma$ on a $A\left(\left[\frac{j}{N}, \frac{j+1}{N}\right) \times \bar{I}^3, Y^{(n)}\right) = 1\Gamma$ pour tout entier j tel que $0 \leq j < N$. Par conséquent

$$\frac{A(E_z^{(n)}, Y^{(n)})}{N} = \lambda_4(E_z^{(n)}). \quad (4.20)$$

Les estimations suivantes sont dues à Niederreiter [47] Théorèmes 3.6 et 3.7.

LEMME 4.3 *La discrédance à l'origine d'un réseau- (t, m, s) X en base b vérifie*

$$\mathbf{D}_N^*(X) \leq \left\lfloor \frac{b-1}{2}(m-t) + \frac{3}{2} \right\rfloor b^{t-m}, \quad \text{si } s = 2, \quad (4.21)$$

$$\mathbf{D}_N^*(X) \leq \left\lfloor \left(\frac{b-1}{2}\right)^2 (m-t)^2 + \frac{b-1}{2}(m-t) + \frac{9}{4} \right\rfloor b^{t-m}, \quad \text{si } s = 3. \quad (4.22)$$

Ce qui entraîne

PROPOSITION 4.2 *On a*

$$\left| \frac{A(F_z^{(n)}, Y^{(n)})}{N} - \lambda_4(F_z^{(n)}) \right| \leq \left\lfloor \frac{b-1}{2}m + \frac{3}{2} \right\rfloor b^{-m}. \quad (4.23)$$

Démonstration. Notons $j^{(n)}(z) = \max\{j : 0 \leq j < N, v_j^{(n)} < z\}$. Les particules étant réordonnées de telle manière que $i < j \Rightarrow v_i^{(n)} \leq v_j^{(n)}$ on obtient

$$F_z^{(n)} = \left[0, \frac{j^{(n)}(z) + 1}{N} \right) \times I \times [0, \nu\Delta t) \times \bar{I}. \quad (4.24)$$

On peut écrire $A(F_z^{(n)}, Y^{(n)}) = A(\tilde{F}_z^{(n)}, X^{(n)}) + A(\hat{F}_z^{(n)}, X^{(n)})\Gamma$ où

$$\begin{aligned} \tilde{F}_z^{(n)} &= \left[0, \frac{1}{b^m} \left\lfloor \frac{j^{(n)}(z) + 1}{2} \right\rfloor \right) \times I \times [0, \nu\Delta t) \times I, \\ \hat{F}_z^{(n)} &= I \times \left[0, \frac{1}{b^m} \left\lfloor \frac{j^{(n)}(z) + 2}{2} \right\rfloor \right) \times [0, \nu\Delta t) \times I. \end{aligned}$$

Puisque $\lambda_4(F_z^{(n)}) = \frac{1}{2} (\lambda_4(\tilde{F}_z^{(n)}) + \lambda_4(\hat{F}_z^{(n)}))$ il vient

$$\begin{aligned} \frac{A(F_z^{(n)}, Y^{(n)})}{N} - \lambda_4(F_z^{(n)}) &= \frac{1}{2} \left(\frac{A(P_{1,3}\tilde{F}_z^{(n)}, P_{1,3}X^{(n)})}{b^m} - \lambda_2(P_{1,3}\tilde{F}_z^{(n)}) \right) \\ &\quad + \frac{1}{2} \left(\frac{A(P_{2,3}\hat{F}_z^{(n)}, P_{2,3}X^{(n)})}{b^m} - \lambda_2(P_{2,3}\hat{F}_z^{(n)}) \right). \end{aligned} \quad (4.25)$$

Le résultat désiré est obtenu en utilisant les lemmes 4.1 et 4.3. ■

Finalement considérons la différence $\frac{A(G_z^{(n)}, Y^{(n)})}{N} - \lambda_4(G_z^{(n)})$ qui est l'erreur introduite par le calcul approché sur le terme d'interaction non-linéaire. L'analyse est assez proche de celle de $\frac{A(F_z^{(n)}, Y^{(n)})}{N} - \lambda_4(F_z^{(n)})$.
On part d'un concept général.

DÉFINITION 4.1 *Un sous-ensemble E de I^s a la propriété du quadrant généralisée s'il existe un $\mathbf{v} = (v_1, \dots, v_s)$, avec $v_i \in \{0, 1\}$ pour $1 \leq i \leq s$, tel que*

$$\mathbf{x} = (x_1, \dots, x_s) \in E \Rightarrow I^s \cap \prod_{i=1}^s [\min(x_i, v_i), \max(x_i, v_i)] \subset E.$$

La définition d'un ensemble ayant la *propriété du quadrant* (ce qui correspond à $\mathbf{v} = \mathbf{0}$) a été introduite par Schmidt [55]. Avec cette notion on a l'inégalité suivante.

LEMME 4.4 *Soit X un réseau- (t, m, s) en base b , et $E \subset I^s$ un ensemble ayant la propriété du quadrant généralisée. Alors, pour $m \geq t + s$,*

$$\left| \frac{A(E, X)}{b^m} - \lambda_s(E) \right| \leq sb^{-\lfloor \frac{m-t}{s} \rfloor}. \quad (4.26)$$

Démonstration. Soit \mathbf{v} associé à E comme dans la définition 4.1. On peut supposer que $v_s = 1$. Soit $d = \lfloor \frac{m-t}{s} \rfloor$. Pour $a = (a_1, \dots, a_s)$ avec des entiers a_i tels que $0 \leq a_i < b^d$ notons $I_a = \prod_{i=1}^s \left[\frac{a_i}{b^d}, \frac{a_i + 1}{b^d} \right)$. Si on pose

$$E_- = \bigcup_{I_a \subset E} I_a \quad \text{et} \quad E_+ = \bigcup_{I_a \cap E \neq \emptyset} I_a,$$

alors $E_- \subset E \subset E_+$ et

$$\begin{aligned} \frac{A(E_-, X)}{b^m} - \lambda_s(E_-) - \lambda_s(E \setminus E_-) &\leq \frac{A(E, X)}{b^m} - \lambda_s(E), \\ \frac{A(E, X)}{b^m} - \lambda_s(E) &\leq \frac{A(E_+, X)}{b^m} - \lambda_s(E_+) + \lambda_s(E_+ \setminus E). \end{aligned} \quad (4.27)$$

Les sous-ensembles E_- et E_+ sont des unions disjointes d'intervalles élémentaires I_a avec $\lambda_s(I_a) \geq b^{t-m}$ donc $\frac{A(E_-, X)}{b^m} - \lambda_s(E_-) = \frac{A(E_+, X)}{b^m} - \lambda_s(E_+) = 0$. De plus $E \setminus E_-$ et $E_+ \setminus E$ sont des sous-ensembles de $E_+ \setminus E_-$ d'où

$$\left| \frac{A(E, X)}{b^m} - \lambda_s(E) \right| \leq \lambda_s(E_+ \setminus E_-). \quad (4.28)$$

Pour $a' = (a_1, \dots, a_{s-1})$ notons $I'_{a'} = \prod_{i=1}^{s-1} \left[\frac{a_i}{b^d}, \frac{a_i + 1}{b^d} \right)$ et

$$\begin{aligned} \underline{a}_s(a') &= \min\{a_s : I_{(a', a_s)} \cap E \neq \emptyset \text{ et } I_{(a', a_s)} \not\subset E\}, \\ \bar{a}_s(a') &= \max\{a_s : I_{(a', a_s)} \cap E \neq \emptyset \text{ et } I_{(a', a_s)} \not\subset E\}, \end{aligned}$$

$$a'^+ = (a_1^+, \dots, a_{s-1}^+), \text{ où } a_i^+ = a_i + 1 \text{ si } v_i = 0, \quad a_i^+ = a_i - 1 \text{ si } v_i = 1.$$

Puisque E est un ensemble ayant la propriété du quadrant généralisée il vient

$$E_+ \setminus E_- = \bigcup_{a'} I'_{a'} \times \left[\frac{\underline{a}_s(a')}{b^d}, \frac{\bar{a}_s(a') + 1}{b^d} \right) \quad (4.29)$$

et

$$\bar{a}_s(a') \leq \underline{a}_s(a'^+).$$

Définissons par récurrence $a'^{\alpha+} = (a'^{(\alpha-1)+})^+$. Alors en utilisant la convention

$$\bar{a}_s(a') = \underline{a}_s(a') = 0 \quad \text{si } \exists i \quad a_i < 0 \text{ ou } a_i \geq b^d,$$

on obtient

$$\begin{aligned} \sum_{a'} (\bar{a}_s(a') - \underline{a}_s(a')) &= \sum_{i=1}^{s-1} \sum_{a_1=0}^{b^d-1} \cdots \sum_{a_{i-1}=0}^{b^d-1} \sum_{a_{i+1}=0}^{b^d-1} \cdots \sum_{a_{s-1}=0}^{b^d-1} \sum_{\alpha=0}^{b^d-1} \\ &\quad \cdot (\bar{a}_s((a_1, \dots, a_{i-1}, (b^d - 1)v_i, a_{i+1}, \dots, a_{s-1})^{\alpha+}) \\ &\quad - \underline{a}_s((a_1, \dots, a_{i-1}, (b^d - 1)v_i, a_{i+1}, \dots, a_{s-1})^{\alpha+})) \\ &\leq (s-1)b^{d(s-1)}, \end{aligned}$$

qui conduit au résultat du lemme. ■

REMARQUE 4.1 D'après un exemple donné dans [34] Proposition 4] l'ordre de l'estimation (4.26) est le meilleur possible pour $t = 0$.

LEMME 4.5 Soit X un réseau- (t, m, s) en base b et $E' \subset I^{s-1}$ un ensemble ayant la propriété du quadrant généralisée. Soit $0 \leq \xi_s \leq 1$ et $E = E' \times [0, \xi_s)$. Alors, pour tout $m \geq t + s - 1$,

$$\left| \frac{A(E, X)}{b^m} - \lambda_s(E) \right| \leq (s-1)(m-t-s+2)b^{-\lfloor \frac{m-t-1}{s-1} \rfloor}. \quad (4.30)$$

Démonstration. On procède par récurrence sur m . L'inégalité est triviale pour $m = t + s - 1$. Supposons $m \geq t + s$ et (4.30) vraie pour $m - 1$. Si $\xi_s = 1$ le résultat se déduit des lemmes 4.1 et 4.4. Si $\xi_s < 1$ on introduit l'entier $\ell = \lfloor b\xi_s \rfloor$ et on écrit E sous la forme d'une union disjointe d'intervalles

$$E_h = E' \times \left[\frac{h}{b}, \frac{h+1}{b} \right), \quad 0 \leq h < \ell \quad \text{et} \quad E_\ell = E' \times \left[\frac{\ell}{b}, \xi_s \right).$$

Pour $0 \leq h \leq \ell$ notons X_h l'ensemble de points $X \cap I^{s-1} \times \left[\frac{h}{b}, \frac{h+1}{b} \right)$.

Si $0 \leq h < \ell$ alors

$$\frac{A(E_h, X)}{b^m} - \lambda_s(E_h) = \frac{1}{b} \left(\frac{A(E', PX_h)}{b^{m-1}} - \lambda_{s-1}(E') \right), \quad (4.31)$$

avec $P = P_{1, \dots, s-1}$. D'après le lemme 4.1 PX_h est un réseau $(t, m-1, s-1)$ en base b . Donc

$$\left| \frac{A(E_h, X)}{b^m} - \lambda_s(E_h) \right| \leq \frac{1}{b} (s-1) b^{-\lfloor \frac{m-t-1}{s-1} \rfloor}, \quad (4.32)$$

d'après le lemme 4.4.

D'autre part soit T la transformation affine de $I^{s-1} \times \left[\frac{\ell}{b}, \frac{\ell+1}{b} \right)$ dans I^s . D'après le lemme 4.2 TX_ℓ est un réseau $(t, m-1, s)$ en base b . Par l'hypothèse de récurrence

$$\left| \frac{A(E_\ell, X)}{b^m} - \lambda_s(E_\ell) \right| \leq \frac{1}{b} (s-1)(m-t-s+1) b^{-\lfloor \frac{m-t-2}{s-1} \rfloor}, \quad (4.33)$$

et ainsi (4.30) est établie. ■

Nous rappelons un résultat qui a été démontré par Niederreiter [47] Lemme 3.4.(ii)].

LEMME 4.6 Soit X un réseau (t, m, s) en base b . Pour tout intervalle élémentaire $J' \subset I^{s-1}$ en base b et pour tout $\xi_s \in \bar{I}$, on a

$$\left| \frac{A(J' \times [0, \xi_s), X)}{b^m} - \lambda_s(J' \times [0, \xi_s)) \right| \leq b^{t-m}. \quad (4.34)$$

Ceci conduit à la

PROPOSITION 4.3 On a l'estimation

$$\begin{aligned} \left| \frac{A(G_z^{(n)}, Y^{(n)})}{N} - \lambda_4(G_z^{(n)}) \right| &\leq 48(m-2) b^{-\lfloor \frac{m-1}{3} \rfloor} + 26(m-1) b^{-\lfloor \frac{m-1}{2} \rfloor} \\ &+ \left(11 \left[\left(\frac{b-1}{2} \right)^2 m^2 + \frac{b-1}{2} m + \frac{9}{4} \right] + 1 + \nu \Delta t \right) b^{-m}. \end{aligned} \quad (4.35)$$

Démonstration. On suppose d'abord $z > 0$. Pour $(x, y) \neq (0, 0)$ et $0 \leq \frac{z}{(x^2 + y^2)^{1/2}} \leq 1\Gamma$ notons

$$\begin{aligned}\phi_z(x, y) &= \frac{1}{2\pi} \left(\arccos \frac{x}{(x^2 + y^2)^{1/2}} - \arccos \frac{z}{(x^2 + y^2)^{1/2}} \right), \\ \psi_z(x, y) &= \frac{1}{2\pi} \left(\arccos \frac{x}{(x^2 + y^2)^{1/2}} + \arccos \frac{z}{(x^2 + y^2)^{1/2}} \right).\end{aligned}$$

Etant donnés deux entiers $0 \leq j, k < N\Gamma$ on note

$$K_{j,k} = \left[\frac{j}{N}, \frac{j+1}{N} \right) \times \left[\frac{k}{N}, \frac{k+1}{N} \right), \quad K_{j,k}^{\Delta t} = K_{j,k} \times [0, \nu \Delta t).$$

Alors $\Gamma G_z^{(n)}$ s'écrit comme l'union disjointe des ensembles $G_{z,g}^{(n)}$ pour $1 \leq g \leq 5\Gamma$ avec

$$\begin{aligned}G_{z,1}^{(n)} &= \bigcup_{v_j^{(n)2} + v_k^{(n)2} < z^2} K_{j,k}^{\Delta t} \times \bar{I}, \\ G_{z,2}^{(n)} &= \bigcup_{\substack{v_j^{(n)} < z, v_k^{(n)} < 0 \\ v_j^{(n)2} + v_k^{(n)2} \geq z^2}} K_{j,k}^{\Delta t} \times \left[0, \phi_z(v_j^{(n)}, v_k^{(n)}) \right) \cup \left(\psi_z(v_j^{(n)}, v_k^{(n)}), 1 \right], \\ G_{z,3}^{(n)} &= \bigcup_{\substack{v_j^{(n)} < z, v_k^{(n)} \geq 0 \\ v_j^{(n)2} + v_k^{(n)2} \geq z^2}} K_{j,k}^{\Delta t} \times \left[0, 1 - \psi_z(v_j^{(n)}, v_k^{(n)}) \right) \cup \left(1 - \phi_z(v_j^{(n)}, v_k^{(n)}), 1 \right], \\ G_{z,4}^{(n)} &= \bigcup_{\substack{v_j^{(n)} \geq z, v_k^{(n)} < 0 \\ v_j^{(n)2} + v_k^{(n)2} \geq z^2}} K_{j,k}^{\Delta t} \times \left(\psi_z(v_j^{(n)}, v_k^{(n)}), 1 + \phi_z(v_j^{(n)}, v_k^{(n)}) \right), \\ G_{z,5}^{(n)} &= \bigcup_{\substack{v_j^{(n)} \geq z, v_k^{(n)} \geq 0 \\ v_j^{(n)2} + v_k^{(n)2} \geq z^2}} K_{j,k}^{\Delta t} \times \left(-\phi_z(v_j^{(n)}, v_k^{(n)}), 1 - \psi_z(v_j^{(n)}, v_k^{(n)}) \right).\end{aligned}$$

Pour des entiers $0 \leq j, k < b^m\Gamma$ on pose

$$J_{j,k} = \left[\frac{j}{b^m}, \frac{j+1}{b^m} \right) \times \left[\frac{k}{b^m}, \frac{k+1}{b^m} \right), \quad J_{j,k}^{\Delta t} = J_{j,k} \times [0, \nu \Delta t).$$

On peut écrire $A(G_{z,g}^{(n)}, Y^{(n)}) = A(\tilde{G}_{z,g}^{(n)}, X^{(n)}) + A(\hat{G}_{z,g}^{(n)}, X^{(n)})\Gamma$ où

$$\tilde{G}_{z,1}^{(n)} = \bigcup_{v_{2j+1}^{(n)2} + v_{2k}^{(n)2} < z^2} J_{j,k}^{\Delta t} \times I, \quad \hat{G}_{z,1}^{(n)} = \bigcup_{v_{2j}^{(n)2} + v_{2k+1}^{(n)2} < z^2} J_{k,j}^{\Delta t} \times I,$$

et

$$\begin{aligned}\tilde{G}_{z,2}^{(n)} &= \bigcup_{\substack{v_{2j+1}^{(n)} < z, v_{2k}^{(n)} < 0 \\ v_{2j+1}^{(n)2} + v_{2k}^{(n)2} \geq z^2}} J_{j,k}^{\Delta t} \times \left[0, \phi_z(v_{2j+1}^{(n)}, v_{2k}^{(n)})\right) \cup \left(\psi_z(v_{2j+1}^{(n)}, v_{2k}^{(n)}), 1\right), \\ \hat{G}_{z,2}^{(n)} &= \bigcup_{\substack{v_{2j}^{(n)} < z, v_{2k+1}^{(n)} < 0 \\ v_{2j}^{(n)2} + v_{2k+1}^{(n)2} \geq z^2}} J_{k,j}^{\Delta t} \times \left[0, 1 - \psi_z(v_{2j}^{(n)}, v_{2k+1}^{(n)})\right) \cup \left(1 - \phi_z(v_{2j}^{(n)}, v_{2k+1}^{(n)}), 1\right).\end{aligned}$$

Les ensembles $\tilde{G}_{z,g}^{(n)}$ et $\hat{G}_{z,g}^{(n)}$ $\Gamma 3 \leq g \leq 5$ se définissent de façon analogue. Alors

$$\begin{aligned}\frac{A(G_z^{(n)}, Y^{(n)})}{N} - \lambda_4(G_z^{(n)}) &= -\sum_{g=1}^5 \left(\lambda_4(G_{z,g}^{(n)}) - \frac{1}{2} \left(\lambda_4(\tilde{G}_{z,g}^{(n)}) + \lambda_4(\hat{G}_{z,g}^{(n)}) \right) \right) \\ &+ \frac{1}{2} \sum_{g=1}^5 \left(\frac{A(\tilde{G}_{z,g}^{(n)}, X^{(n)})}{b^m} - \lambda_4(\tilde{G}_{z,g}^{(n)}) + \frac{A(\hat{G}_{z,g}^{(n)}, X^{(n)})}{b^m} - \lambda_4(\hat{G}_{z,g}^{(n)}) \right).\end{aligned}\quad (4.36)$$

On a

$$\begin{aligned}& \left| \lambda_4(G_{z,1}^{(n)}) - \frac{1}{2} \left(\lambda_4(\tilde{G}_{z,1}^{(n)}) + \lambda_4(\hat{G}_{z,1}^{(n)}) \right) \right| \\ & \leq \frac{\nu \Delta t}{N^2} \left| \#\{(j, k) : 0 \leq j, k < N, j, k \text{ impairs}, v_j^{(n)2} + v_k^{(n)2} < z^2\} \right. \\ & \quad \left. - \#\{(j, k) : 0 \leq j, k < N, j \text{ impair}, k \text{ pair}, v_j^{(n)2} + v_k^{(n)2} < z^2\} \right| \\ & \quad + \frac{\nu \Delta t}{N^2} \left| \#\{(j, k) : 0 \leq j, k < N, j, k \text{ pairs}, v_j^{(n)2} + v_k^{(n)2} < z^2\} \right. \\ & \quad \left. - \#\{(j, k) : 0 \leq j, k < N, j \text{ pair}, k \text{ impair}, v_j^{(n)2} + v_k^{(n)2} < z^2\} \right|\end{aligned}\quad (4.37)$$

et

$$\begin{aligned}& \left| \lambda_4(G_{z,2}^{(n)}) - \frac{1}{2} \left(\lambda_4(\tilde{G}_{z,2}^{(n)}) + \lambda_4(\hat{G}_{z,2}^{(n)}) \right) \right| \\ & \leq \frac{\nu \Delta t}{N^2} \sum_{j \text{ impair}, v_j^{(n)} < z} \left| \sum_{\substack{k \text{ impair}, v_k^{(n)} < 0 \\ v_j^{(n)2} + v_k^{(n)2} \geq z^2}} \left(1 - \frac{1}{\pi} \arccos \frac{z}{(v_j^{(n)2} + v_k^{(n)2})^{1/2}} \right) \right. \\ & \quad \left. - \sum_{\substack{k \text{ pair}, v_k^{(n)} < 0 \\ v_j^{(n)2} + v_k^{(n)2} \geq z^2}} \left(1 - \frac{1}{\pi} \arccos \frac{z}{(v_j^{(n)2} + v_k^{(n)2})^{1/2}} \right) \right| \\ & \quad + \frac{\nu \Delta t}{N^2} \sum_{j \text{ pair}, v_j^{(n)} < z} \left| \sum_{\substack{k \text{ pair}, v_k^{(n)} < 0 \\ v_j^{(n)2} + v_k^{(n)2} \geq z^2}} \left(1 - \frac{1}{\pi} \arccos \frac{z}{(v_j^{(n)2} + v_k^{(n)2})^{1/2}} \right) \right.\end{aligned}$$

$$- \sum_{\substack{k \text{ impair, } v_k^{(n)} < 0 \\ v_j^{(n)2} + v_k^{(n)2} \geq z^2}} \left(1 - \frac{1}{\pi} \arccos \frac{z}{(v_j^{(n)2} + v_k^{(n)2})^{1/2}} \right) \Big|, \quad (4.38)$$

et des bornes similaires pour $g = 3, 4, 5$. Les particules étant réordonnées il vient

$$\left| \lambda_4(G_{z,1}^{(n)}) - \frac{1}{2} (\lambda_4(\tilde{G}_{z,1}^{(n)}) + \lambda_4(\hat{G}_{z,1}^{(n)})) \right| \leq \frac{\nu \Delta t}{N^2} \cdot (\#\{j : 0 \leq j < N, j \text{ impair}\} + \#\{j : 0 \leq j < N, j \text{ pair}\}) \leq \frac{\nu \Delta t}{N} \quad (4.39)$$

et

$$\begin{aligned} & \left| \lambda_4(G_{z,2}^{(n)}) - \frac{1}{2} (\lambda_4(\tilde{G}_{z,2}^{(n)}) + \lambda_4(\hat{G}_{z,2}^{(n)})) \right| \leq \\ & \frac{\nu \Delta t}{\pi N^2} \sum_{\substack{v_k^{(n)} < 0, v_j^{(n)} < z \\ v_j^{(n)2} + v_k^{(n)2} \geq z^2}} \left| \arccos \frac{z}{(v_j^{(n)2} + v_k^{(n)2})^{1/2}} - \arccos \frac{z}{(v_j^{(n)2} + v_{k-1}^{(n)2})^{1/2}} \right| \\ & \leq \frac{1}{2} \frac{\nu \Delta t}{N^2} \#\{j : 0 \leq j < N, v_j^{(n)} < z\}. \end{aligned} \quad (4.40)$$

Les ensembles $G_{z,g}^{(n)} \Gamma$ pour $3 \leq g \leq 5$ sont traités de la même manière que l'ensemble $G_{z,2}^{(n)}$. Par conséquent

$$\sum_{g=2}^5 \left| \lambda_4(G_{z,g}^{(n)}) - \frac{1}{2} (\lambda_4(\tilde{G}_{z,g}^{(n)}) + \lambda_4(\hat{G}_{z,g}^{(n)})) \right| \leq \frac{\nu \Delta t}{N}. \quad (4.41)$$

D'autre part $\Gamma \tilde{G}_{z,1}^{(n)} = \hat{G}_{z,1}^{(n)}$. Soit

$$\begin{aligned} \Gamma_{z,1,1}^{(n)} &= \bigcup_{\substack{v_{2j+1}^{(n)} < 0, v_{2k}^{(n)} < 0 \\ v_{2j+1}^{(n)2} + v_{2k}^{(n)2} \geq z^2}} J_{j,k}^{\Delta t}, & \Gamma_{z,1,2}^{(n)} &= \bigcup_{\substack{v_{2j+1}^{(n)} < 0, v_{2k}^{(n)} \geq 0 \\ v_{2j+1}^{(n)2} + v_{2k}^{(n)2} \geq z^2}} J_{j,k}^{\Delta t}, \\ \Gamma_{z,1,3}^{(n)} &= \bigcup_{\substack{v_{2j+1}^{(n)} \geq 0, v_{2k}^{(n)} < 0 \\ v_{2j+1}^{(n)2} + v_{2k}^{(n)2} \geq z^2}} J_{j,k}^{\Delta t}, & \Gamma_{z,1,4}^{(n)} &= \bigcup_{\substack{v_{2j+1}^{(n)} \geq 0, v_{2k}^{(n)} \geq 0 \\ v_{2j+1}^{(n)2} + v_{2k}^{(n)2} \geq z^2}} J_{j,k}^{\Delta t}. \end{aligned}$$

On a une union disjointe

$$I^2 \times [0, \nu \Delta t) = \bigcup_{h=1}^4 \Gamma_{z,1,h}^{(n)} \cup P_{1,2,3} \tilde{G}_{z,1}^{(n)}. \quad (4.42)$$

Quand les particules sont réordonnées Γ chaque ensemble $\Gamma_{z,1,h}^{(n)}$ est un ensemble ayant la propriété du quadrant généralisée. D'où il vient d'après les lemmes 4.1, 4.5 et 4.6 que

$$\left| \frac{A(P_{1,2,3} \tilde{G}_{z,1}^{(n)}, P_{1,2,3} X^{(n)})}{b^m} - \lambda_3(P_{1,2,3} \tilde{G}_{z,1}^{(n)}) \right| \leq 8(m-1)b^{-\lfloor \frac{m-1}{2} \rfloor} + b^{-m}, \quad (4.43)$$

et alors

$$\frac{1}{2} \left| \frac{A(\tilde{G}_{z,1}^{(n)}, X^{(n)})}{b^m} - \lambda_4(\tilde{G}_{z,1}^{(n)}) + \frac{A(\hat{G}_{z,1}^{(n)}, X^{(n)})}{b^m} - \lambda_4(\hat{G}_{z,1}^{(n)}) \right| \leq 8(m-1)b^{-\lfloor \frac{m-1}{2} \rfloor} + b^{-m}. \quad (4.44)$$

Maintenant Γ définissons

$$\begin{aligned} \tilde{G}_{z,2,1}^{(n)} = & \bigcup_{v_{2j+1}^{(n)} < z, v_{2k}^{(n)} < -z} J_{j,k}^{\Delta t} \times [0, \phi_z(v_{2j+1}^{(n)}, v_{2k}^{(n)})] \\ & \cup \bigcup_{\substack{0 \leq v_{2j+1}^{(n)} < z, -z \leq v_{2k}^{(n)} < 0 \\ v_{2j+1}^{(n)2} + v_{2k}^{(n)2} \geq z^2}} J_{j,k}^{\Delta t} \times [0, \phi_z(v_{2j+1}^{(n)}, v_{2k}^{(n)})], \end{aligned}$$

$$\tilde{G}_{z,2,2}^{(n)} = \bigcup_{\substack{v_{2j+1}^{(n)} < 0, -z \leq v_{2k}^{(n)} < 0 \\ v_{2j+1}^{(n)2} + v_{2k}^{(n)2} \geq z^2}} J_{j,k}^{\Delta t} \times [0, \phi_z(v_{2j+1}^{(n)}, v_{2k}^{(n)})],$$

$$\tilde{G}_{z,2,3}^{(n)} = \bigcup_{v_{2j+1}^{(n)} < -z, v_{2k}^{(n)} < 0} J_{j,k}^{\Delta t} \times (\psi_z(v_{2j+1}^{(n)}, v_{2k}^{(n)}), 1),$$

$$\begin{aligned} \tilde{G}_{z,2,4}^{(n)} = & \bigcup_{\substack{-z \leq v_{2j+1}^{(n)} < 0, v_{2k}^{(n)} < 0 \\ v_{2j+1}^{(n)2} + v_{2k}^{(n)2} \geq z^2}} J_{j,k}^{\Delta t} \times (\psi_z(v_{2j+1}^{(n)}, v_{2k}^{(n)}), 1) \\ & \cup \bigcup_{0 \leq v_{2j+1}^{(n)} < z, v_{2k}^{(n)} < -z} J_{j,k}^{\Delta t} \times (\psi_z(v_{2j+1}^{(n)}, v_{2k}^{(n)}), 1), \end{aligned}$$

$$\tilde{G}_{z,2,5}^{(n)} = \bigcup_{\substack{0 \leq v_{2j+1}^{(n)} < z, -z \leq v_{2k}^{(n)} < 0 \\ v_{2j+1}^{(n)2} + v_{2k}^{(n)2} \geq z^2}} J_{j,k}^{\Delta t} \times (\psi_z(v_{2j+1}^{(n)}, v_{2k}^{(n)}), 1),$$

et

$$\begin{aligned} \tilde{G}_{z,2,1}^{(n)'} = \tilde{G}_{z,2,5}^{(n)'} = & \bigcup_{v_{2j+1}^{(n)} < 0, -z \leq v_{2k}^{(n)}} J_{j,k}^{\Delta t} \times I \cup \bigcup_{\substack{0 \leq v_{2j+1}^{(n)} < z, v_{2k}^{(n)} < 0 \\ v_{2j+1}^{(n)2} + v_{2k}^{(n)2} < z^2}} J_{j,k}^{\Delta t} \times I \\ & \cup \bigcup_{0 \leq v_{2j+1}^{(n)} < z, 0 \leq v_{2k}^{(n)}} J_{j,k}^{\Delta t} \times I, \end{aligned}$$

$$\tilde{G}_{z,2,2}^{(n)'} = \bigcup_{\substack{v_{2j+1}^{(n)} < 0, -z \leq v_{2k}^{(n)} < 0 \\ v_{2j+1}^{(n)2} + v_{2k}^{(n)2} < z^2}} J_{j,k}^{\Delta t} \times I \cup \bigcup_{\substack{0 \leq v_{2j+1}^{(n)} \\ -z \leq v_{2k}^{(n)} < 0}} J_{j,k}^{\Delta t} \times I \cup \bigcup_{0 \leq v_{2k}^{(n)}} J_{j,k}^{\Delta t} \times I,$$

$$\tilde{G}_{z,2,3}^{(n)'} = \bigcup_{-z \leq v_{2j+1}^{(n)}, v_{2k}^{(n)} < 0} J_{j,k}^{\Delta t} \times I,$$

$$\begin{aligned} \tilde{G}_{z,2,4}^{(n)'} = & \bigcup_{\substack{v_{2j+1}^{(n)} < 0, -z \leq v_{2k}^{(n)} < 0 \\ v_{2j+1}^{(n)2} + v_{2k}^{(n)2} < z^2}} J_{j,k}^{\Delta t} \times I \cup \bigcup_{-z \leq v_{2j+1}^{(n)} < 0, 0 \leq v_{2k}^{(n)}} J_{j,k}^{\Delta t} \times I \\ & \cup \bigcup_{0 \leq v_{2j+1}^{(n)} < z, -z \leq v_{2k}^{(n)}} J_{j,k}^{\Delta t} \times I \cup \bigcup_{z \leq v_{2j+1}^{(n)}} J_{j,k}^{\Delta t} \times I. \end{aligned}$$

Alors

$$\tilde{G}_{z,2,h}^{(n)} \cap \tilde{G}_{z,2,h}^{(n)'} = \emptyset \quad \text{for } 1 \leq h \leq 5. \quad (4.45)$$

Les particules étant réordonnées l'ensemble $P_{1,2} \tilde{G}_{z,2,h}^{(n)'}$ est un ensemble ayant la propriété du quadrant généralisée et c'est un intervalle de la forme $[\xi_1, 1) \times [0, \xi_2)$ si $h = 3$. L'ensemble $P_{1,2,4}(\tilde{G}_{z,2,h}^{(n)} \cup \tilde{G}_{z,2,h}^{(n)'})$ a aussi la propriété du quadrant généralisée. Dès lors d'après les lemmes 4.1 et 4.3 et 4.5 il vient

$$\begin{aligned} \left| \frac{A(\tilde{G}_{z,2}^{(n)}, X^{(n)})}{b^m} - \lambda_4(\tilde{G}_{z,2}^{(n)}) \right| & \leq 8(m-1)b^{-\lfloor \frac{m-1}{2} \rfloor} + 15(m-2)b^{-\lfloor \frac{m-1}{3} \rfloor} \\ & + \left[\left(\frac{b-1}{2} \right)^2 m^2 + \frac{b-1}{2} m + \frac{9}{4} \right] b^{-m}. \end{aligned} \quad (4.46)$$

Les ensembles $\tilde{G}_{z,g}^{(n)} \Gamma$ pour $3 \leq g \leq 5$ ainsi que les ensembles $\hat{G}_{z,g}^{(n)} \Gamma$ pour $2 \leq g \leq 5$ sont traités comme $\tilde{G}_{z,2}^{(n)}$. On trouve

$$\begin{aligned} \frac{1}{2} \sum_{g=2}^5 \left| \frac{A(\tilde{G}_{z,g}^{(n)}, X^{(n)})}{b^m} - \lambda_4(\tilde{G}_{z,g}^{(n)}) + \frac{A(\hat{G}_{z,g}^{(n)}, X^{(n)})}{b^m} - \lambda_4(\hat{G}_{z,g}^{(n)}) \right| \\ \leq 48(m-2)b^{-\lfloor \frac{m-1}{3} \rfloor} + 18(m-1)b^{-\lfloor \frac{m-1}{2} \rfloor} \\ + 7 \left[\left(\frac{b-1}{2} \right)^2 m^2 + \frac{b-1}{2} m + \frac{9}{4} \right] b^{-m}. \end{aligned} \quad (4.47)$$

A partir de (4.36) et (4.39) et (4.41) et (4.44) et (4.47) il vient

$$\begin{aligned} \left| \frac{A(G_z^{(n)}, Y^{(n)})}{N} - \lambda_4(G_z^{(n)}) \right| & \leq 48(m-2)b^{-\lfloor \frac{m-1}{3} \rfloor} + 26(m-1)b^{-\lfloor \frac{m-1}{2} \rfloor} \\ & + \left(7 \left[\left(\frac{b-1}{2} \right)^2 m^2 + \frac{b-1}{2} m + \frac{9}{4} \right] + 1 + \nu \Delta t \right) b^{-m}. \end{aligned} \quad (4.48)$$

Si $z \leq 0$ la démonstration conduisant à (4.48) donne

$$\begin{aligned} \left| \frac{A(G_z^{(n)}, Y^{(n)})}{N} - \lambda_4(G_z^{(n)}) \right| &\leq 48(m-2)b^{-\lfloor \frac{m-1}{3} \rfloor} + 10(m-1)b^{-\lfloor \frac{m-1}{2} \rfloor} \\ &+ 11 \left[\left(\frac{b-1}{2} \right)^2 m^2 + \frac{b-1}{2} m + \frac{9}{4} \right] b^{-m}. \end{aligned} \quad (4.49)$$

D'où le résultat de la proposition. ■

On combine les estimations précédentes pour avoir une borne d'erreur pour la méthode quasi-Monte Carlo de simulation.

PROPOSITION 4.4 *On a*

$$\begin{aligned} \mathbf{D}_N^*(V^{(n)}, f_n) &\leq e^{\nu t_n} \mathbf{D}_N^*(V^{(0)}, f_0) + \Delta t \int_0^{t_n} \int_{-\infty}^{+\infty} e^{\nu(t_n-t)} \left| \frac{\partial^2 f}{\partial t^2}(v, t) \right| dv dt \\ &+ \frac{e^{\nu t_n}}{\nu \Delta t} \left(48(m-2)b^{-\lfloor \frac{m-1}{3} \rfloor} + 26(m-1)b^{-\lfloor \frac{m-1}{2} \rfloor} \right. \\ &\left. + \left(11 \left[\left(\frac{b-1}{2} \right)^2 m^2 + \frac{b-1}{2} m + \frac{9}{4} \right] + \left[\frac{b-1}{2} m + \frac{3}{2} \right] + 1 + \nu \Delta t \right) b^{-m} \right). \end{aligned} \quad (4.50)$$

Démonstration. Le résultat vient des relations (4.14)Γ(4.19) et (4.20)Γavec l'inégalité (4.18) et les propositions 4.1Γ4.2 et 4.3. ■

REMARQUE 4.2 La Proposition 4.4 indique que l'on a un taux de convergence de l'ordre de $N^{-\frac{1}{3}}$ qui est plus mauvais que celui d'une méthode de Monte Carlo. Les expériences numériques montrent que cette estimation est pessimiste.

4.5 Études expérimentales

Pour valider l'utilisation des réseaux- (t, m, s) Γune expérimentation est faite dans un cas où la solution exacte est connue. Si

$$\nu = \frac{\sqrt{\pi}}{2} \quad \text{et} \quad f_0(v) = \frac{2}{\sqrt{\pi}} v^2 e^{-v^2},$$

alors

$$f(v, t) = \frac{1}{\sqrt{\pi}} \left(\frac{3}{2} (1 - c(t)) c(t)^{1/2} + (3c(t) - 1) c(t)^{3/2} v^2 \right) e^{-c(t)v^2}, \quad (4.51)$$

où

$$c(t) = \frac{1}{3 - 2e^{-\frac{\sqrt{\pi}t}{16}}}.$$

Une représentation tridimensionnelle de la densité $f(v, t)$ est donnée sur la Figure 4.1. On voit que le système est pratiquement en équilibre à l'instant $t = 10$.

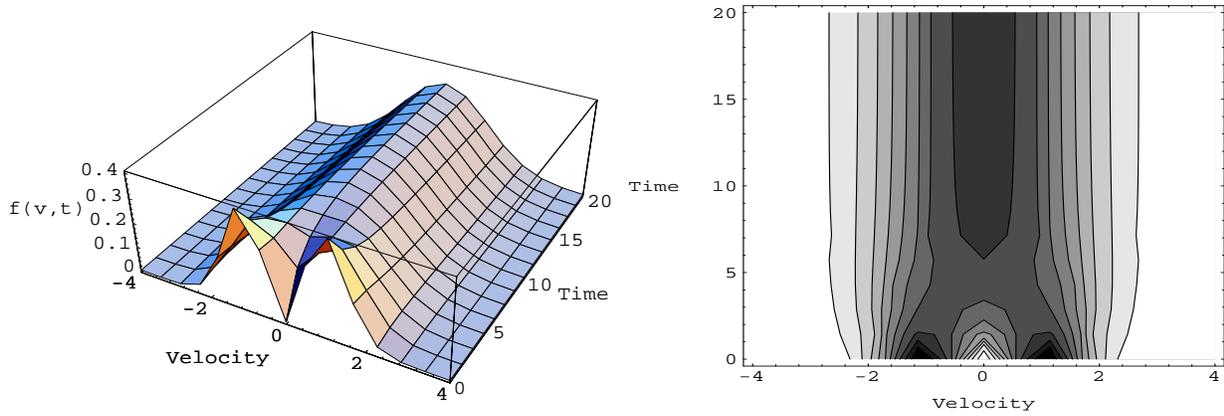


FIG. 4.1 – Solution exacte du modèle de Kac.

La précision de la simulation particulaire est mesurée au moyen de l'erreur (discrédance-*) $\mathbf{D}_N^*(V^{(n)}, f_n)$. On compare les erreurs obtenues en résolvant le modèle de Kac par trois méthodes différentes:

1. la méthode quasi-Monte Carlo utilisant les réseaux: elle sera notée QMCF
2. la méthode DSMC (direct simulation Monte Carlo method) de Bird Γ
3. la méthode de simulation stochastique de Nanbu que l'on notera SS.

Toutes les simulations sont faites en partant de N particules représentant la distribution initiale de vitesse. En dimension un Γ il est facile de trouver la borne inférieure de la discrédance-* $\mathbf{D}_N^*(X)$ pour N fixé. On a toujours $\mathbf{D}_N^*(X) \geq 1/2N$ et cette inégalité devient une égalité si $X = \Xi\Gamma$ où

$$\Xi = \left\{ \frac{2j+1}{2N} : 0 \leq j < N \right\}.$$

L'initialisation des particules est faite en résolvant les N équations non-linéaires suivantes

$$\int_{-\infty}^{v_j^{(0)}} f_0(w)dw = \frac{2j+1}{2N} : 0 \leq j < N.$$

Alors on a

$$\mathbf{D}_N^*(V^{(0)}, f_0) = \mathbf{D}_N^*(\Xi) = \frac{1}{2N}. \quad (4.52)$$

La méthode quasi-Monte Carlo de simulation utilise une suite-(0, 4) en base b .

La construction d'une suite-(0, s) en une base b qui est un nombre premier étant plus simple que la construction dans une base quelconque $b\Gamma$ la base $b = 5$ est choisie. Une méthode pour construire une telle suite-(0, s) a été développée par Faure [14] et implémentée par Lécot dans [33]. Toutes les simulations sont faites jusqu'à l'instant $T =$

20. On a représenté les erreurs $\mathbf{D}_N^*(V^{(n)}, f_n)$ en fonction du temps. D'autre part un taux empirique de convergence N^{-p} est déterminé par la représentation de l'exposant de l'erreur $p_N^{(n)} = -\log(\mathbf{D}_N^*(V^{(n)}, f_n)) / \log N$ en fonction du temps. Les figures 4.2Γ4.3Γ4.4 et 4.5 contiennent les résultats de simulations pour les trois méthodes: QMC ΓDSMC et SS. Les tests ont été effectués avec $N = 2 \cdot 5^4, 2 \cdot 5^5, 2 \cdot 5^6$ et $2 \cdot 5^7$ particules. Le pas de temps utilisé est $\Delta t = 0.1$. Les résultats pour $\Delta t = 0.05$ sont donnés dans les figures 4.6Γ4.7Γ4.8 et 4.9. Les figures 4.10Γ4.11Γ4.12 et 4.13 contiennent les résultats correspondant à $\Delta t = 0.01$. Nous avons également représenté les moments d'ordre 1 et 4 avec $N = 6250$ particules et $\Delta t = 0.1$ (voir figure 4.14). Les deux tableaux contiennent les erreurs minimalesΓles erreurs maximalesΓles erreurs moyennes ainsi que les exposants d'erreurs moyens pour les 3 schémas.

Les résultats montrent clairement de meilleures performances pour la méthode QMC que pour les méthodes DSMC et SS. Pour un niveau d'erreur fixéΓon a une réduction du nombre de particules simulées par un facteur 5 pour la simulation QMC. Le taux de convergence passe de $N^{-0.5}$ pour les méthodes de type Monte Carlo à $N^{-0.6}$ pour le schéma quasi-Monte Carlo. Dans la pratique on a de meilleurs résultats de convergence que ceux fournis par l'analyse théorique. D'autre part on observe que l'erreur n'est améliorée en régime transitoire que si le pas de temps est suffisamment petit.

Dans le but de souligner l'importance de l'étape de tri dans la simulation QMCΓnous avons testé cette méthode en omettant de trier à chaque pas de temps les particules (voir la figure 4.15). Les résultats de la méthode QMC se détériorent et se retrouvent au même niveau que ceux des méthodes DSMC et SS.

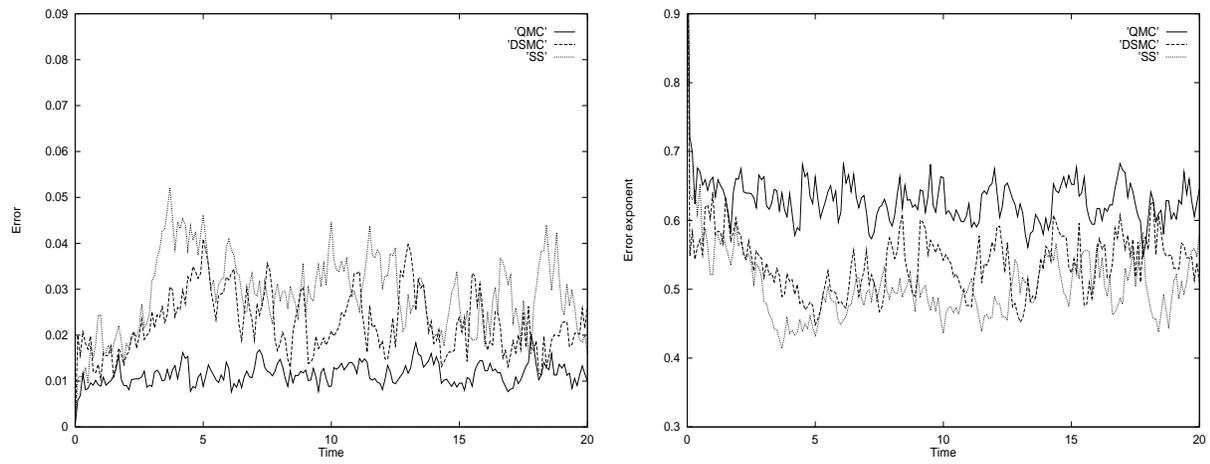


FIG. 4.2 – $N = 2.5^4 = 1\,250$ *particules*, $\Delta t = 0.1$.

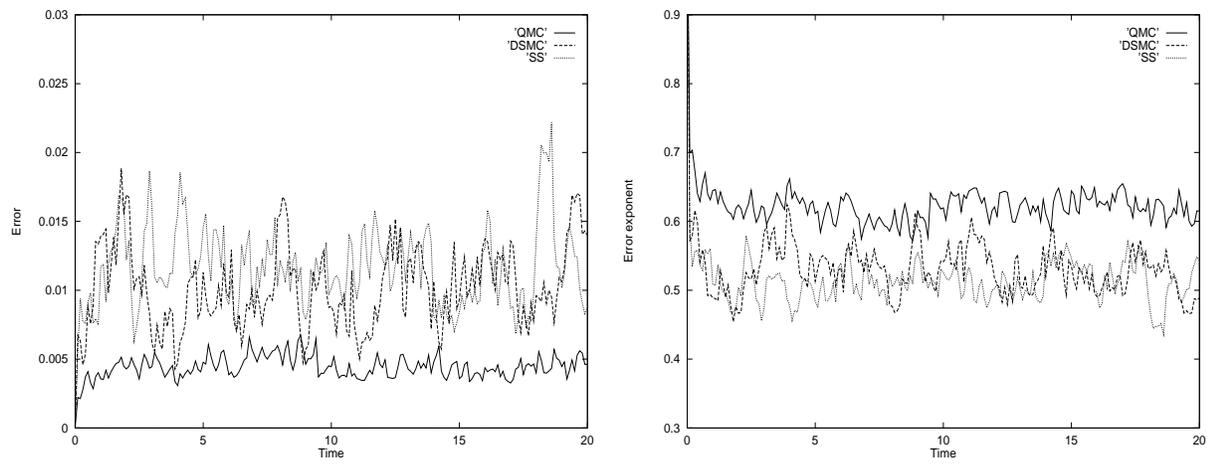


FIG. 4.3 – $N = 2.5^5 = 6\,250$ *particules*, $\Delta t = 0.1$.

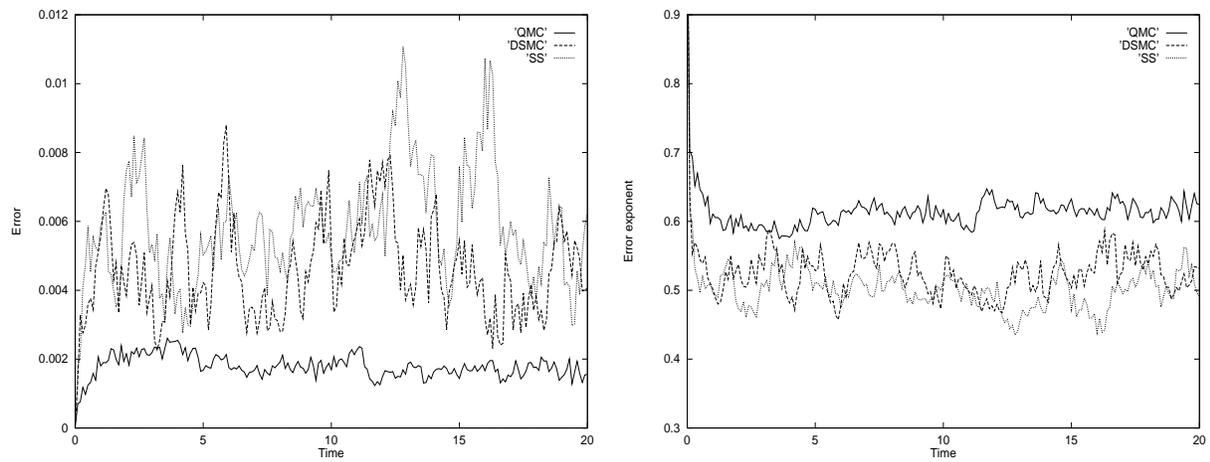
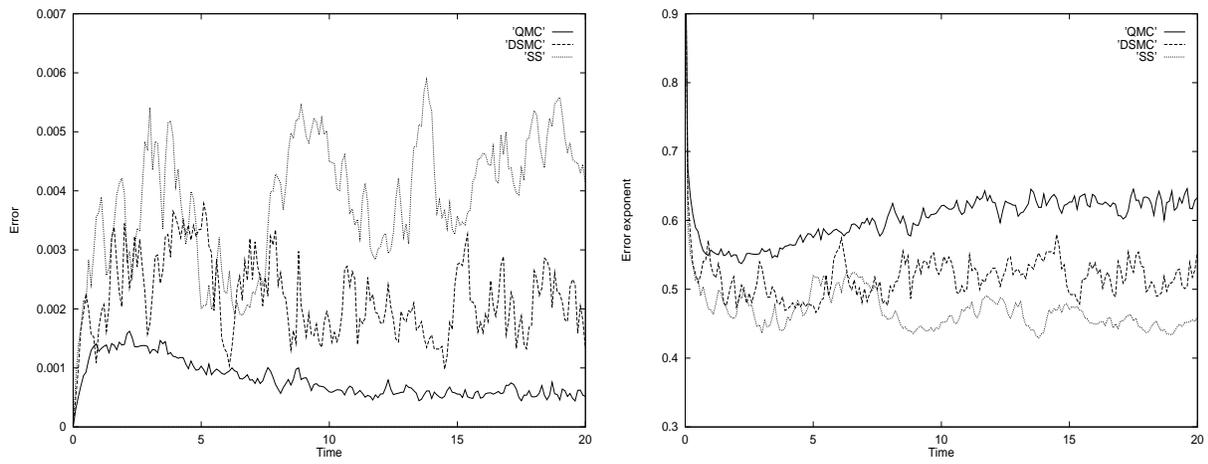
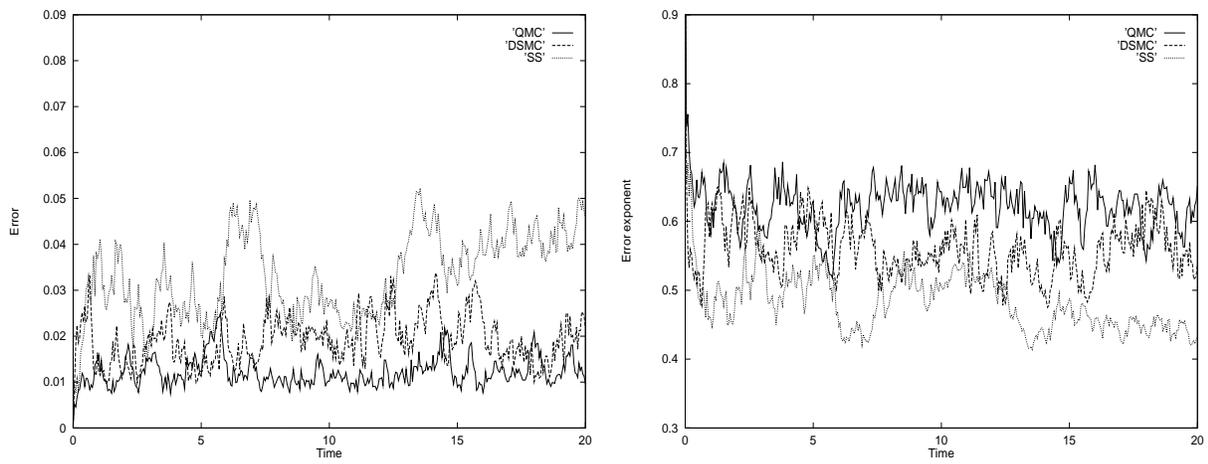
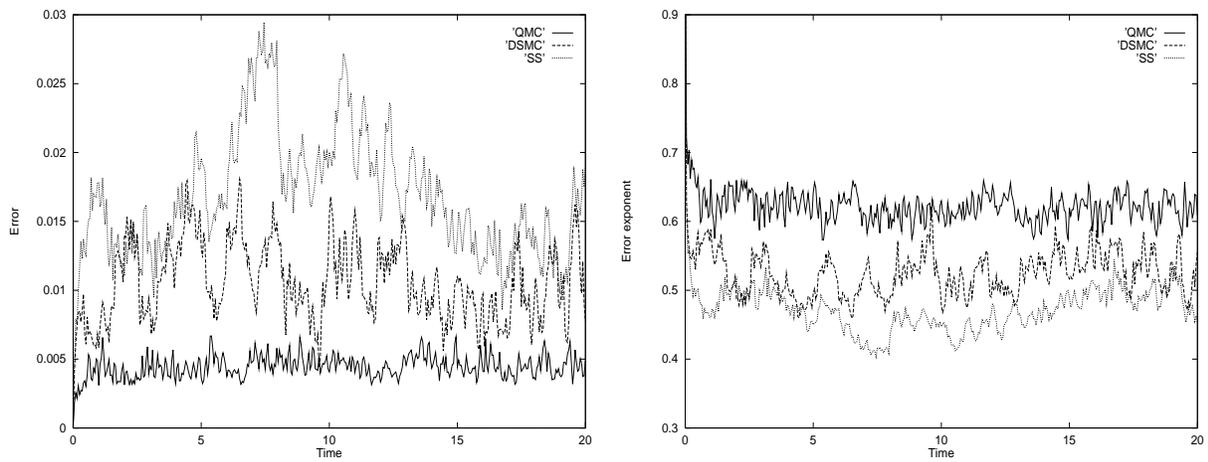


FIG. 4.4 – $N = 2.5^6 = 31\,250$ *particules*, $\Delta t = 0.1$.

FIG. 4.5 – $N = 2.5^7 = 156\,250$ *particules*, $\Delta t = 0.1$.FIG. 4.6 – $N = 2.5^4 = 1\,250$ *particules*, $\Delta t = 0.05$.FIG. 4.7 – $N = 2.5^5 = 6\,250$ *particules*, $\Delta t = 0.05$.

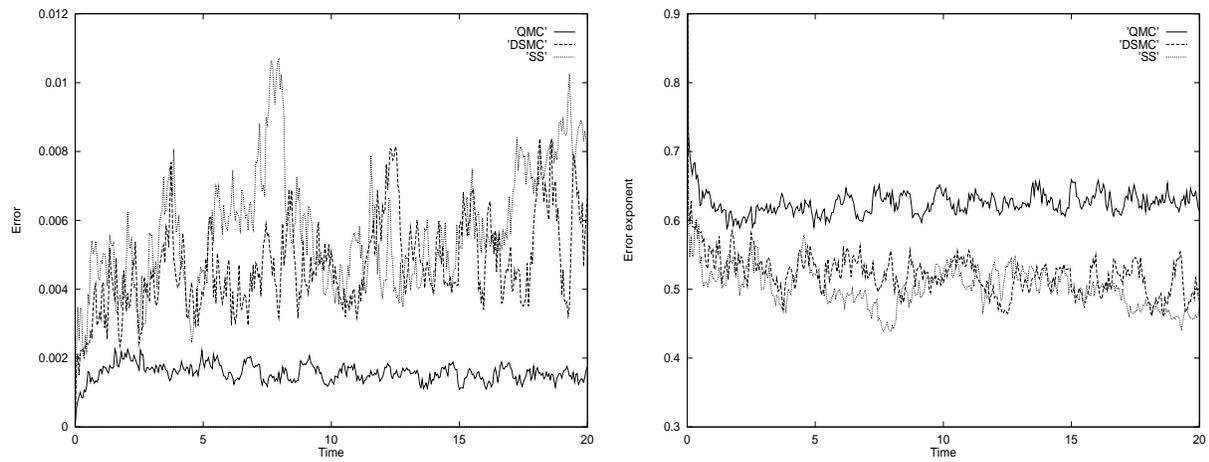


FIG. 4.8 – $N = 2.5^6 = 31\,250$ *particules*, $\Delta t = 0.05$.

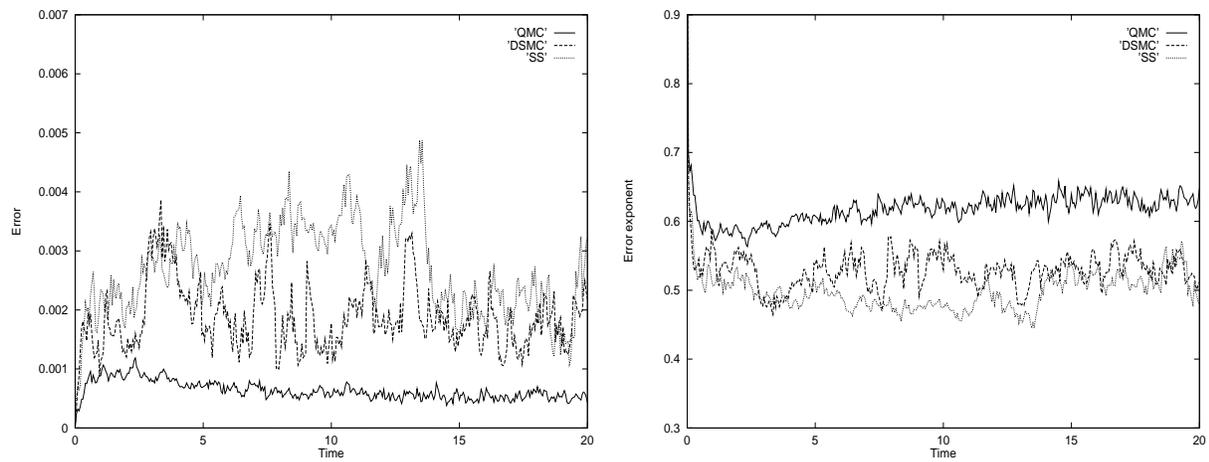


FIG. 4.9 – $N = 2.5^7 = 156\,250$ *particules*, $\Delta t = 0.05$.

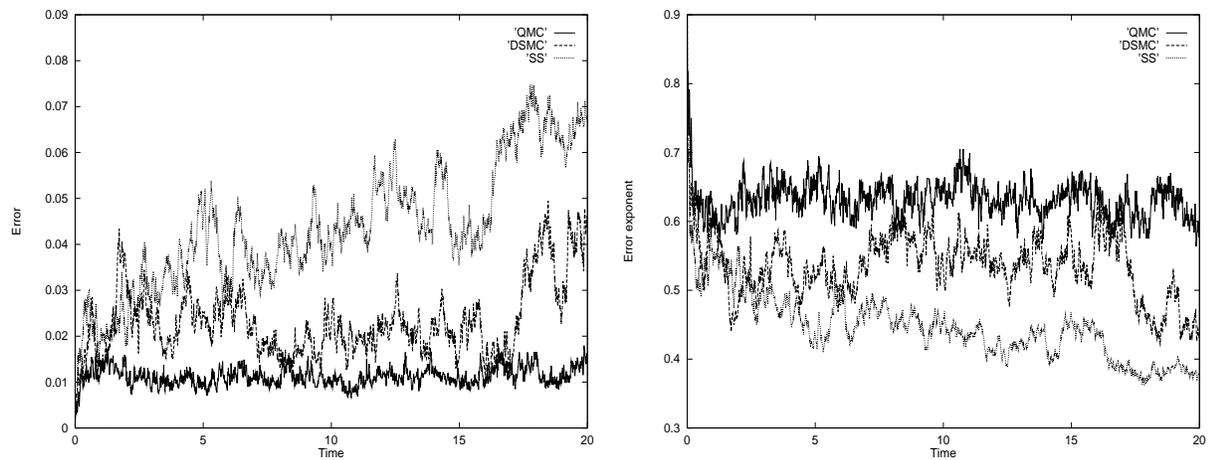
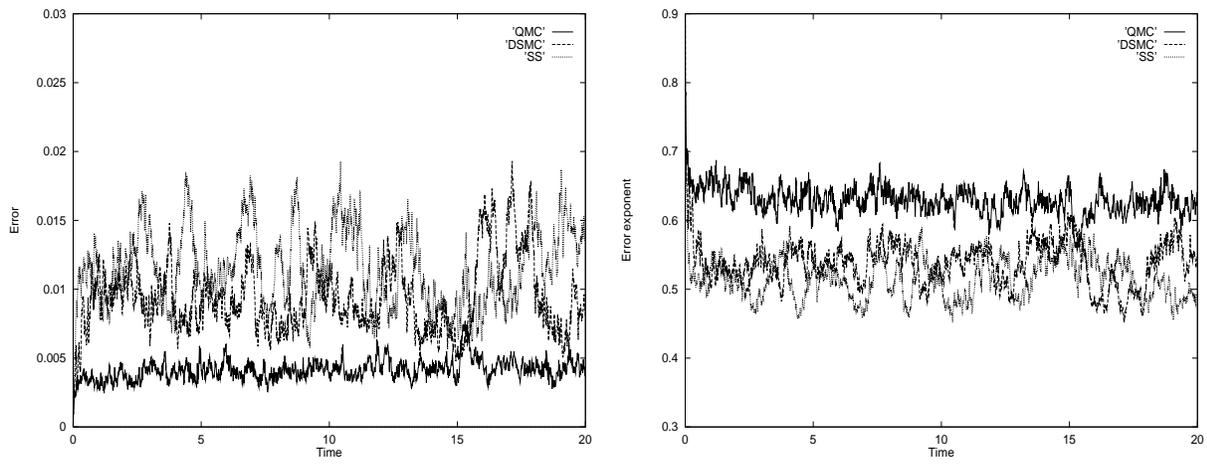
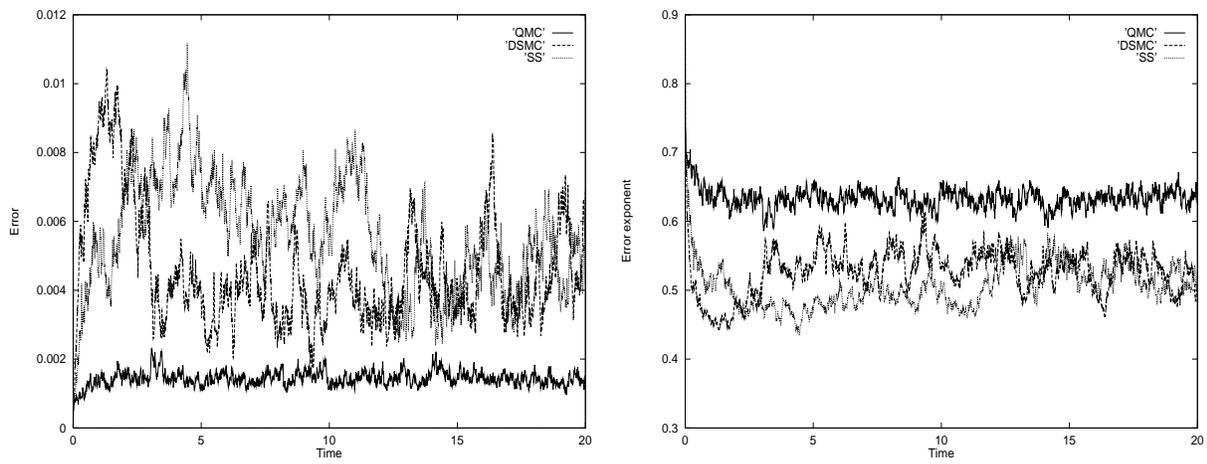
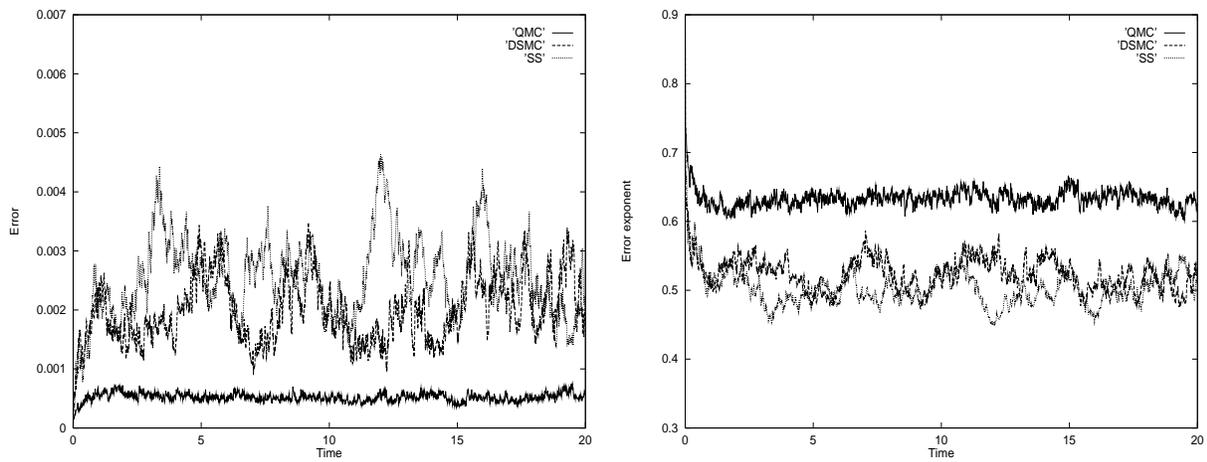


FIG. 4.10 – $N = 2.5^4 = 1\,250$ *particules*, $\Delta t = 0.01$.

FIG. 4.11 – $N = 2.5^5 = 6\,250$ *particules*, $\Delta t = 0.01$.FIG. 4.12 – $N = 2.5^6 = 31\,250$ *particules*, $\Delta t = 0.01$.FIG. 4.13 – $N = 2.5^7 = 156\,250$ *particules*, $\Delta t = 0.01$.

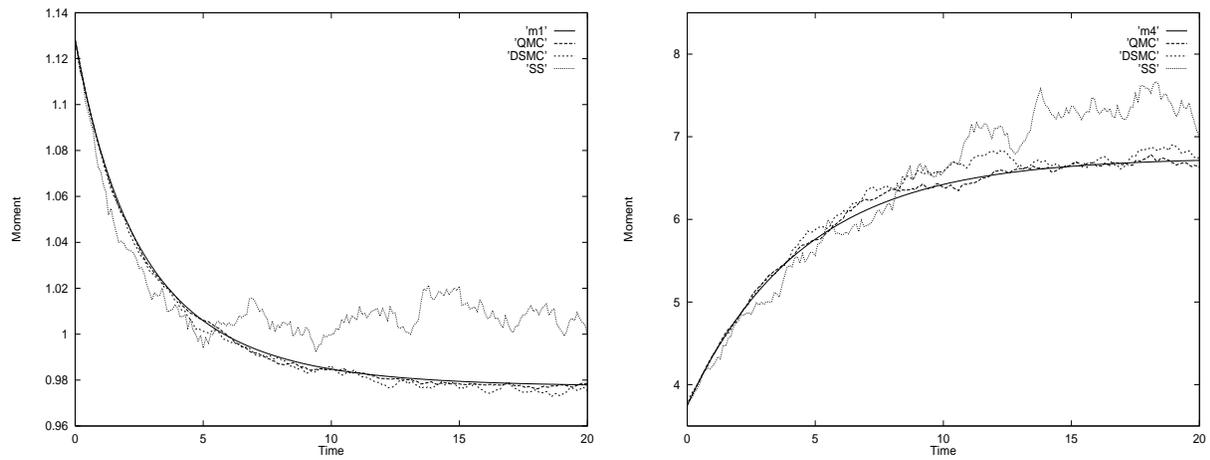


FIG. 4.14 – Moments d'ordre 1 et 4 avec $N = 2.5^5 = 6\,250$ particules et $\Delta t = 0.1$.

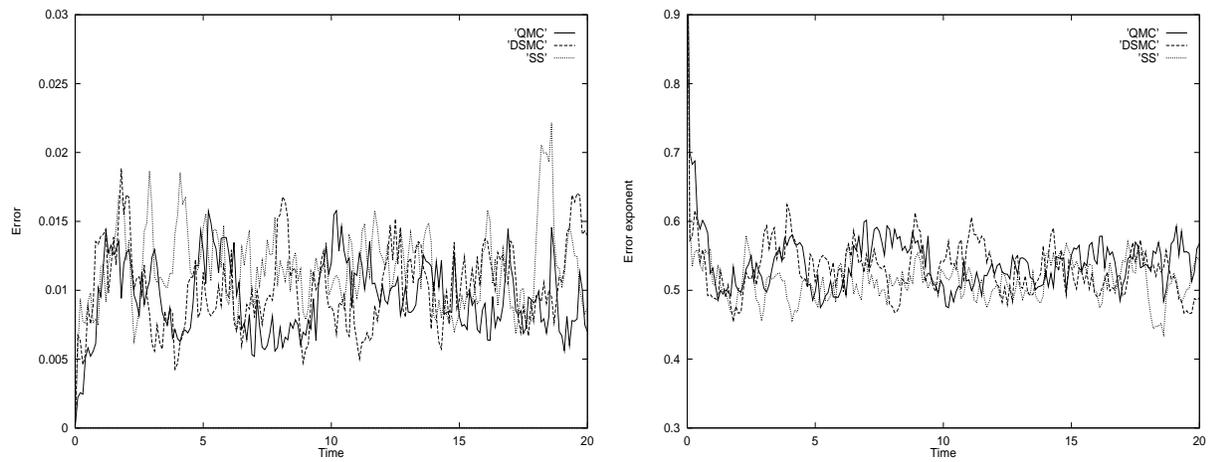


FIG. 4.15 – Méthode QMC sans tri avec $N = 2.5^5 = 6\,250$ particules et $\Delta t = 0.1$.

$\Delta t = 0.1$	erreur minimum	erreur maximum	erreur moyenne	exposant d'erreur moyen
QMC				
N=250	1.78E-02	5.50E-02	3.27E-02	0.619
N=1250	5.82E-03	2.00E-02	1.15E-02	0.625
N=6250	2.13E-03	6.72E-03	4.46E-03	0.619
N=31250	6.97E-04	2.60E-03	1.80E-03	0.610
N=156250	3.05E-04	1.62E-03	8.05E-04	0.595
SS				
N=250	1.94E-02	1.32E-01	5.74E-02	0.517
N=1250	9.52E-03	5.21E-02	2.90E-02	0.496
N=6250	5.45E-03	2.21E-02	1.16E-02	0.509
N=31250	1.77E-03	1.10E-02	5.82E-03	0.497
N=156250	8.18E-04	5.88E-03	3.87E-03	0.464
DSMC				
N=250	1.91E-02	1.28E-01	5.53E-02	0.524
N=1250	1.03E-02	4.09E-02	2.27E-02	0.530
N=6250	4.27E-03	1.88E-02	1.02E-02	0.524
N=31250	1.32E-03	8.79E-03	4.63E-03	0.519
N=156250	5.06E-04	3.78E-03	2.20E-03	0.511

$\Delta t = 0.01$	erreur minimum	erreur maximum	erreur moyenne	exposant d'erreur moyen
QMC				
N=250	6.60E-03	5.40E-02	3.23E-02	0.621
N=1250	2.92E-03	1.79E-02	1.12E-02	0.629
N=6250	1.03E-03	7.69E-03	4.13E-03	0.627
N=31250	4.80E-04	2.30E-03	1.43E-03	0.632
N=156250	1.53E-04	7.54E-04	5.19E-04	0.632
SS				
N=250	1.01E-02	1.38E-01	7.82E-02	0.461
N=1250	2.50E-03	7.47E-02	4.36E-02	0.439
N=6250	1.01E-03	1.92E-02	1.15E-02	0.510
N=31250	7.19E-04	1.11E-02	5.55E-03	0.501
N=156250	2.61E-04	4.63E-03	2.55E-03	0.499
DSMC				
N=250	6.06E-03	1.03E-01	5.60E-02	0.521
N=1250	3.04E-03	4.94E-02	2.31E-02	0.527
N=6250	1.44E-03	1.93E-02	9.38E-03	0.534
N=31250	6.40E-04	1.03E-02	4.54E-03	0.521
N=156250	1.97E-04	3.48E-03	1.98E-03	0.520

4.6 Conclusion

L'analyse théorique et l'étude expérimentale montrent qu'il est possible d'appliquer les méthodes particulières quasi-Monte Carlo de simulation à des modèles cinétiques non-linéaires. Les réseaux- (t, m, s) garantissent de faibles erreurs d'intégration à condition de réordonner les particules à chaque pas de temps. La convergence de la simulation peut être démontrée. L'outil principal pour démontrer cette convergence est l'analyse de l'erreur introduite par le calcul approché de la mesure d'un ensemble ayant la propriété du quadrant généralisée. Les résultats expérimentaux montrent une amélioration de l'erreur et du taux de convergence par rapport à ceux des simulations de type Monte Carlo. Ce travail constitue un préliminaire à la résolution de l'équation de Boltzmann en utilisant les réseaux- (t, m, s) . Il y a plusieurs directions possibles pour les travaux futurs. La technique utilisée par Morokoff et Caflish [40] Lemme 1] pourrait servir à améliorer la borne d'erreur. L'extension de l'algorithme à l'équation de Boltzmann non linéaire et homogène en espace est à l'étude.

Chapitre 5

Simulation de la diffusion par des marches quasi-aléatoires

Résumé: Nous présentons une méthode de simulation de la diffusion. Elle utilise des marches quasi-aléatoires de particules. Le problème considéré est le problème de Cauchy pour l'équation de la diffusion de dimension s en espace. Nous introduisons s pas de discrétisation en espace Δx_i de manière à effectuer une semi-discrétisation consistant à remplacer les dérivées spatiales par des différences finies. Ensuite, une formulation faible est construite à partir de cette équation semi-discrétisée en espace. N particules sont utilisées pour approcher la distribution initiale. L'intervalle de temps est découpé en sous-intervalles de longueur Δt . Puis nous faisons une discrétisation en temps par un schéma d'Euler. À chaque pas de temps, le mouvement des particules est décrit à l'aide d'une intégration en dimension $s + 1$. Une quadrature quasi-Monte Carlo est alors effectuée avec une suite $(0, s + 1)$. Un élément clé de l'utilisation réussie de la suite à faible discrèpance est une technique de renumérotation des particules à chaque pas de temps. Nous démontrons la convergence de la solution approchée vers la solution de l'équation semi-discrétisée quand $N \rightarrow \infty$ et $\Delta t \rightarrow 0$. Les tests numériques montrent un gain de précision lorsque les suites quasi-aléatoires sont utilisées avec la renumérotation des particules, par rapport à la méthode des marches aléatoires.

5.1 Introduction

Les méthodes particulières aléatoires sont utilisées pour la résolution numérique des équations de diffusion et convection-diffusion [13] Chapitre 2 Section 7]. Les méthodes avec transport de la fonction semblent avoir un intérêt surtout théorique à cause des erreurs statistiques. Une voie pour améliorer la précision est l'utilisation des méthodes de transport du gradient. D'un autre côté des modifications des méthodes de Monte Carlo utilisant des suites quasi-aléatoires ont été récemment proposées [5] 7] 25] 26] 33] 38] 40] 41] 60]. Ces différents travaux ont montré une amélioration de la convergence par rapport aux méthodes Monte Carlo standards.

Ici on considère une méthode particulière dans le cadre d'un problème de diffusion s -dimensionnel. On fixe s pas d'espace Δx_i puis on remplace les dérivées spatiales par des différences finies. Ensuite on intègre en espace l'équation semi-discrétisée après l'avoir

multipliée par une fonction test. On obtient alors une formulation faible. On choisit le pas de temps Δt ainsi que le nombre N de particules. La solution du problème faible à l'instant $t_n = n\Delta t$ est approchée par une expression de la forme

$$u^{(n)}(\mathbf{x}) = \frac{1}{N} \sum_{j=0}^{N-1} \delta(\mathbf{x} - \mathbf{x}_j^{(n)}),$$

où $\mathbf{x}_j^{(n)}$ est la position de la $j^{\text{ème}}$ particule à l'instant t_n et $\delta(\mathbf{x} - \mathbf{y})$ la mesure de Dirac au point $\mathbf{y} \in \mathbb{R}^s$. Les positions des N particules à l'instant $t_0 = 0$ approchent la distribution initiale. On remplace la dérivée en temps par une différence finie explicite pour ensuite calculer la solution à l'instant $t_{n+1} = t_n + \Delta t$ grâce à une quadrature Monte Carlo.

L'idée de base d'une méthode quasi-Monte Carlo est de remplacer les suites aléatoires par des suites déterministes judicieusement choisies. Des constructions de suites à faible discrédance ont été proposées par Halton [16] Sobol' [59] Faure [14] et Niederreiter [47]. La construction de Sobol' fournit des suites (t, s) en base 2. Celle de Faure fournit des suites $(0, s)$ avec des bases qui sont des entiers premiers et la construction de Niederreiter donne des suites $(0, s)$ avec des bases égales à des puissances de nombres premiers. Les réseaux (t, m, s) et les suites (t, s) avec $t = 0$ sont ceux qui ont les meilleures propriétés d'équidistribution.

L'approche de ce travail consiste à utiliser une approximation quasi-Monte Carlo pour calculer une intégrale $(s + 1)$ -dimensionnelle à chaque pas de temps. Si le nombre de particules est $N = b^m$ avec des entiers b et m les nœuds de quadrature sont les points d'un réseau $(0, m, s + 1)$ en base b garantissant ainsi une faible erreur d'approximation. On augmente l'efficacité des suites quasi-aléatoires en augmentant la régularité de l'intégrand. Une technique consistant à renuméroter les particules à chaque pas de temps a été développée dans le but de surmonter le manque de régularité [33, 38, 41, 60]. Ici cette méthode consiste à répartir les particules en sous-ensembles de niveau 1 par rapport à leur coordonnée x_1 puis à trier les particules de chaque sous-ensemble de niveau 1 en sous-ensembles de niveau 2 par rapport à leur coordonnée x_2 et ainsi de suite. Une borne d'erreur pour la méthode des marches quasi-aléatoires est obtenue en utilisant cette technique de renumérotation des particules.

Le plan du chapitre est le suivant. Dans la section 5.2 l'approximation semi-discrète de l'équation de diffusion est introduite et sa simulation à partir des marches quasi-aléatoires est présentée. Dans la section 5.3 nous prouvons un théorème de convergence pour la méthode. Le traitement de l'erreur provenant du calcul approché de la mesure d'un sous-intervalle de I^{s+1} de la forme

$$\prod_{i=1}^{s-1} \left[\frac{a_i}{b^{d_i}}, \frac{a_i + 1}{b^{d_i}} \right) \times \left[0, \frac{a_s}{b^{d_s}} \right) \times [0, \xi_{s+1})$$

constitue la partie majeure de l'analyse. La Section 5.4 est consacrée à des tests numériques illustrant les résultats théoriques. La précision de la méthode des marches quasi-aléatoires est comparée à celle de la méthode standard des marches aléatoires. Enfin dans la dernière section nous faisons un bilan des résultats obtenus et nous parlons des futures directions de recherche.

5.2 Méthode des marches quasi-aléatoires

Nous nous intéressons à la résolution numérique du problème de Cauchy pour une équation de diffusion

$$\frac{\partial \tilde{u}}{\partial t}(\mathbf{x}, t) = \sum_{i=1}^s D_i \frac{\partial^2 \tilde{u}}{\partial x_i^2}(\mathbf{x}, t), \quad \mathbf{x} \in \mathbb{R}^s, t > 0, \quad (5.1)$$

$$\tilde{u}(\mathbf{x}, 0) = u_0(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^s, \quad (5.2)$$

où les D_i sont des constantes positives et la condition initiale u_0 est choisie de telle sorte que

$$u_0 \in L^1(\mathbb{R}^s), \quad u_0 \geq 0 \quad \text{et} \quad \int_{\mathbb{R}^s} u_0(\mathbf{x}) d\mathbf{x} = 1. \quad (5.3)$$

Dans un premier temps le problème (5.1)–(5.2) est discrétisé uniquement en espace. Soit $\mathbf{e}^{(i)}$, $1 \leq i \leq s$ la base canonique de \mathbb{R}^s . Si Δx_i , $1 \leq i \leq s$ sont s pas d'espace et $\tau_{\pm \Delta x_i}^{(i)}$ l'opérateur défini par

$$\tau_{\pm \Delta x_i}^{(i)} v(\mathbf{x}) = v(\mathbf{x} \pm \Delta x_i \mathbf{e}^{(i)}),$$

l'approximation semi-discrète de (5.1)–(5.2) est alors

$$\frac{\partial u}{\partial t}(\mathbf{x}, t) = \sum_{i=1}^s \frac{D_i}{\Delta x_i^2} \left(\tau_{-\Delta x_i}^{(i)} u - 2u + \tau_{\Delta x_i}^{(i)} u \right) (\mathbf{x}, t), \quad \mathbf{x} \in \mathbb{R}^s, t > 0, \quad (5.4)$$

$$u(\mathbf{x}, 0) = u_0(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^s. \quad (5.5)$$

Il est facile de vérifier que la masse totale est conservée

$$\forall t > 0 \quad \int_{\mathbb{R}^s} u(\mathbf{x}, t) d\mathbf{x} = 1. \quad (5.6)$$

Il est nécessaire de construire d'abord une formulation faible de l'équation. Soit $\mathcal{S}(\mathbb{R}^s)$ l'ensemble de toutes les fonctions simples mesurables au sens de Borel sur \mathbb{R}^s . En multipliant par $\sigma \in \mathcal{S}(\mathbb{R}^s)$ l'équation semi-discrète (5.4) et en intégrant sur \mathbb{R}^s on trouve

$$\frac{d}{dt} \int_{\mathbb{R}^s} \sigma(\mathbf{x}) u(\mathbf{x}, t) d\mathbf{x} = \sum_{i=1}^s \frac{D_i}{\Delta x_i^2} \int_{\mathbb{R}^s} \left(\tau_{-\Delta x_i}^{(i)} \sigma - 2\sigma + \tau_{\Delta x_i}^{(i)} \sigma \right) (\mathbf{x}) u(\mathbf{x}, t) d\mathbf{x}. \quad (5.7)$$

La simulation particulière se fait comme suit. Soit d_1, \dots, d_s et b des entiers et posons $d_1 + \dots + d_s = m$ et $b^m = N$. Nous introduisons N particules de positions $\mathbf{x}_j^{(0)}$ telles que

$$u^{(0)}(\mathbf{x}) = \frac{1}{N} \sum_{j=0}^{N-1} \delta(\mathbf{x} - \mathbf{x}_j^{(0)})$$

"approche" la distribution initiale $u_0(\mathbf{x})$. Maintenant considérons l'évolution de ces particules dans le temps. Le temps est discrétisé en intervalles de longueur Δt la condition de stabilité suivante étant requise

$$\sum_{i=1}^s D_i \frac{\Delta t}{\Delta x_i^2} \leq \frac{1}{2}. \quad (5.8)$$

Soit $\mathbf{y}_0, \mathbf{y}_1, \dots$ une suite $(0, s+1)$ en base b . Les positions $\mathbf{x}_j^{(n)}$ des particules à l'instant $t_n = n\Delta t$ et l'approximation

$$u^{(n)}(\mathbf{x}) = \frac{1}{N} \sum_{j=0}^{N-1} \delta(\mathbf{x} - \mathbf{x}_j^{(n)})$$

de la solution à l'instant t_n sont construites par une méthode à trois étapes.

(i) Renumérotation des particules:

Comme dans les chapitres 3 et 4 cette étape joue un rôle primordial. Les points $\mathbf{x}_j^{(n)}$ sont notés $\mathbf{x}_{\mathbf{a}}^{(n)}$ avec $\mathbf{a} = (a_1, \dots, a_s)$ et $0 \leq a_i < b^{d_i} \Gamma$ de telle manière que

$$a_1 = b_1, \dots, a_{i-1} = b_{i-1}, a_i < b_i \Rightarrow x_{\mathbf{a},i}^{(n)} \leq x_{\mathbf{b},i}^{(n)}.$$

(ii) Schéma d'Euler explicite:

On cherche une approximation de la solution de l'équation (5.7) en résolvant un problème dans lequel la dérivée en temps est remplacée par une différence finie. Une mesure $v^{(n+1)}(\mathbf{x})$ approchant la solution à l'instant t_{n+1} est définie par

$$\begin{aligned} \forall \sigma \in \mathcal{S}(\mathbb{R}^s) \quad & \frac{1}{\Delta t} \int_{\mathbb{R}^s} \sigma(\mathbf{x}) (v^{(n+1)}(\mathbf{x}) - u^{(n)}(\mathbf{x})) \\ & = \sum_{i=1}^s \frac{D_i}{\Delta x_i^2} \int_{\mathbb{R}^s} (\tau_{-\Delta x_i}^{(i)} \sigma - 2\sigma + \tau_{\Delta x_i}^{(i)} \sigma)(\mathbf{x}) u^{(n)}(\mathbf{x}). \end{aligned}$$

Nous obtenons

$$\begin{aligned} \int_{\mathbb{R}^s} \sigma(\mathbf{x}) v^{(n+1)}(\mathbf{x}) & = \frac{1}{N} \sum_{\mathbf{a}} \left(\sum_{i=1}^s D_i \frac{\Delta t}{\Delta x_i^2} \sigma(\mathbf{x}_{\mathbf{a}}^{(n)} - \Delta x_i \mathbf{e}^{(i)}) \right. \\ & \quad \left. + \left(1 - 2 \sum_{i=1}^s D_i \frac{\Delta t}{\Delta x_i^2} \right) \sigma(\mathbf{x}_{\mathbf{a}}^{(n)}) + \sum_{i=1}^s D_i \frac{\Delta t}{\Delta x_i^2} \sigma(\mathbf{x}_{\mathbf{a}}^{(n)} + \Delta x_i \mathbf{e}^{(i)}) \right). \end{aligned} \quad (5.9)$$

(iii) Intégration quasi-Monte Carlo:

L'approximation résultante est une somme de $(2s+1)N$ mesures de Dirac pondérées sur \mathbb{R}^s . L'intégrale d'une fonction simple $\sigma(\mathbf{x})$ sur \mathbb{R}^s par rapport à cette mesure peut s'écrire comme l'intégrale d'une fonction $\Sigma^{(n)}(\mathbf{y})$ sur le cube unité I^{s+1} . La solution approchée à l'instant t_{n+1} est obtenue en faisant une intégration quasi-Monte Carlo. Notons $c_{\mathbf{a}}$ la fonction caractéristique de

$$I'_{\mathbf{a}} = \prod_{i=1}^s \left[\frac{a_i}{b^{d_i}}, \frac{a_i + 1}{b^{d_i}} \right).$$

Pour $1 \leq i \leq s$ soit χ_{-i} la fonction caractéristique de l'intervalle

$$I_{-i} = \left[\sum_{h=i+1}^s D_h \frac{\Delta t}{\Delta x_h^2}, \sum_{h=i}^s D_h \frac{\Delta t}{\Delta x_h^2} \right),$$

et χ_i celle de l'intervalle

$$I_i = \left[1 - \sum_{h=i}^s D_h \frac{\Delta t}{\Delta x_h^2}, 1 - \sum_{h=i+1}^s D_h \frac{\Delta t}{\Delta x_h^2} \right).$$

De même soit χ_0 la fonction caractéristique de l'intervalle

$$I_0 = \left[\sum_{h=1}^s D_h \frac{\Delta t}{\Delta x_h^2}, 1 - \sum_{h=1}^s D_h \frac{\Delta t}{\Delta x_h^2} \right).$$

Notons $\varepsilon(i)$ le signe de i avec $\varepsilon(0) = 0$. Etant donné $\sigma \in \mathcal{S}(\mathbb{R}^s)$ posons

$$\Sigma^{(n)}(\mathbf{y}) = \sum_{\mathbf{a}} c_{\mathbf{a}}(\mathbf{y}') \sum_{i=-s}^s \chi_i(y_{s+1}) \sigma \left(\mathbf{x}_{\mathbf{a}}^{(n)} + \varepsilon(i) \Delta x_{|i|} \mathbf{e}^{(|i|)} \right), \quad (5.10)$$

pour $\mathbf{y} = (\mathbf{y}', y_{s+1}) \in I^{s+1}$. Il est facile de vérifier que

$$\int_{\mathbb{R}^s} \sigma(\mathbf{x}) v^{(n+1)}(\mathbf{x}) = \int_{I^{s+1}} \Sigma^{(n)}(\mathbf{y}) d\mathbf{y}. \quad (5.11)$$

Nous définissons une approximation $u^{(n+1)}$ de la distribution exacte $u_{n+1}(\mathbf{x}) d\mathbf{x}$ par

$$\int_{\mathbb{R}^s} \sigma(\mathbf{x}) u^{(n+1)}(\mathbf{x}) = \frac{1}{N} \sum_{j=0}^{N-1} \Sigma^{(n)}(\mathbf{y}_{nN+j}) \quad (5.12)$$

où $\{\mathbf{y}_{nN+j} \mid 0 \leq j < N\}$ est un réseau $(0, m, s+1)$ en base b . Alors chaque intervalle $I_{\mathbf{a}} \times I$ contient exactement un point du réseau $\{\mathbf{y}_{nN+j} \mid 0 \leq j < N\}$.

Pour $0 \leq j < N$ soit

$$\mathbf{a}^{(n)}(j) = \left(\lfloor b^{d_1} y_{nN+j,1} \rfloor, \dots, \lfloor b^{d_s} y_{nN+j,s} \rfloor \right).$$

L'équation (5.12) devient

$$\frac{1}{N} \sum_{j=0}^{N-1} \sigma(\mathbf{x}_j^{(n+1)}) = \frac{1}{N} \sum_{j=0}^{N-1} \sum_{i=-s}^s \chi_i(y_{nN+j,s+1}) \sigma \left(\mathbf{x}_{\mathbf{a}^{(n)}(j)}^{(n)} + \varepsilon(i) \Delta x_{|i|} \mathbf{e}^{(|i|)} \right).$$

La composante $y_{nN+j,s+1}$ permet soit de maintenir la particule sur place soit de l'envoyer dans l'une des $2s$ directions possibles. Les nouvelles positions des particules sont alors

$$\begin{aligned} \mathbf{x}_j^{(n+1)} &= \mathbf{x}_{\mathbf{a}^{(n)}(j)}^{(n)} - \Delta x_i \mathbf{e}^{(i)}, \text{ si } \sum_{h=i+1}^s D_h \frac{\Delta t}{\Delta x_h^2} \leq y_{nN+j,s+1} < \sum_{h=i}^s D_h \frac{\Delta t}{\Delta x_h^2}, \\ \mathbf{x}_j^{(n+1)} &= \mathbf{x}_{\mathbf{a}^{(n)}(j)}^{(n)}, \text{ si } \sum_{h=1}^s D_h \frac{\Delta t}{\Delta x_h^2} \leq y_{nN+j,s+1} < 1 - \sum_{h=1}^s D_h \frac{\Delta t}{\Delta x_h^2}, \\ \mathbf{x}_j^{(n+1)} &= \mathbf{x}_{\mathbf{a}^{(n)}(j)}^{(n)} + \Delta x_i \mathbf{e}^{(i)}, \text{ si } 1 - \sum_{h=i}^s D_h \frac{\Delta t}{\Delta x_h^2} \leq y_{nN+j,s+1} < 1 - \sum_{h=i+1}^s D_h \frac{\Delta t}{\Delta x_h^2}. \end{aligned} \quad (5.13)$$

5.3 Convergence de la méthode

Cette section est consacrée à l'analyse de la convergence de la mesure $u^{(n)}(\mathbf{x})$ vers la solution $u_n(\mathbf{x}) = u(\mathbf{x}, t_n)$ du problème semi-discret (5.4)–(5.5). On mesure l'erreur entre la distribution des particules et la solution exacte de la manière suivante. Désignons par $X^{(n)}$ l'ensemble des éléments $\mathbf{x}_a^{(n)}$. Un concept de discrédance par rapport à une fonction continue peut être défini (voir [29] Chapitre 2 Section 1]). Pour $\mathbf{z} \in \mathbb{R}^s$ notons $\sigma_{\mathbf{z}}$ la fonction caractéristique de $\prod_{i=1}^s (-\infty, z_i)$. Nous appelons

$$d_N^{(n)}(\mathbf{z}) = \frac{1}{N} \sum_{\mathbf{a}} \sigma_{\mathbf{z}}(\mathbf{x}_a^{(n)}) - \int_{\mathbb{R}^s} \sigma_{\mathbf{z}}(\mathbf{x}) u_n(\mathbf{x}) d\mathbf{x}$$

la *discrédance locale* de $X^{(n)}$ par rapport à u_n . L'erreur est mesurée par le maximum sur tous les $\mathbf{z} \in \mathbb{R}^s$

$$\mathbf{D}_N^*(X^{(n)}, u_n) = \sup_{\mathbf{z} \in \mathbb{R}^s} |d_N^{(n)}(\mathbf{z})|.$$

L'analyse de la convergence est semblable à celle du schéma d'Euler explicite. La quantité

$$\begin{aligned} \epsilon^{(n)}(\mathbf{z}) &= \frac{1}{\Delta t} \int_{\mathbb{R}^s} \sigma_{\mathbf{z}}(\mathbf{x}) (u_{n+1}(\mathbf{x}) - u_n(\mathbf{x})) d\mathbf{x} \\ &\quad - \sum_{i=1}^s \frac{D_i}{\Delta x_i^2} \int_{\mathbb{R}^s} \left(\tau_{-\Delta x_i}^{(i)} \sigma_{\mathbf{z}} - 2\sigma_{\mathbf{z}} + \tau_{\Delta x_i}^{(i)} \sigma_{\mathbf{z}} \right) (\mathbf{x}) u_n(\mathbf{x}) d\mathbf{x} \end{aligned}$$

représente l'*erreur de troncature*. Introduisons un *terme d'erreur*

$$\begin{aligned} e_N^{(n)}(\mathbf{z}) &= \sum_{i=1}^s \frac{D_i}{\Delta x_i^2} \frac{1}{N} \sum_{\mathbf{a}} \left(\tau_{-\Delta x_i}^{(i)} \sigma_{\mathbf{z}} - 2\sigma_{\mathbf{z}} + \tau_{\Delta x_i}^{(i)} \sigma_{\mathbf{z}} \right) (\mathbf{x}_a^{(n)}) \\ &\quad - \sum_{i=1}^s \frac{D_i}{\Delta x_i^2} \int_{\mathbb{R}^s} \left(\tau_{-\Delta x_i}^{(i)} \sigma_{\mathbf{z}} - 2\sigma_{\mathbf{z}} + \tau_{\Delta x_i}^{(i)} \sigma_{\mathbf{z}} \right) (\mathbf{x}) u_n(\mathbf{x}) d\mathbf{x}, \end{aligned}$$

et l'*erreur d'intégration*

$$\delta_N^{(n)}(\mathbf{z}) = \frac{1}{N} \sum_{j=0}^{N-1} \Sigma_{\mathbf{z}}^{(n)}(\mathbf{y}_{nN+j}) - \int_{I^{s+1}} \Sigma_{\mathbf{z}}^{(n)}(\mathbf{y}) d\mathbf{y},$$

où $\Sigma_{\mathbf{z}}^{(n)}$ est définie par (5.10) avec $\sigma = \sigma_{\mathbf{z}}$. Il est facile de voir que

$$d_N^{(n)}(\mathbf{z}) = d_N^{(n-1)}(\mathbf{z}) + \Delta t e_N^{(n-1)}(\mathbf{z}) - \Delta t \epsilon^{(n-1)}(\mathbf{z}) + \delta_N^{(n-1)}(\mathbf{z}). \quad (5.14)$$

Nous avons

$$e_N^{(n)}(\mathbf{z}) = \sum_{i=1}^s \frac{D_i}{\Delta x_i^2} \left(d_N^{(n)}(\mathbf{z} - \Delta x_i \mathbf{e}^{(i)}) - 2d_N^{(n)}(\mathbf{z}) + d_N^{(n)}(\mathbf{z} + \Delta x_i \mathbf{e}^{(i)}) \right). \quad (5.15)$$

L'estimation

$$|\epsilon^{(n)}(\mathbf{z})| \leq \int_{t_n}^{t_{n+1}} \int_{\mathbb{R}^s} \left| \frac{\partial^2 u}{\partial t^2}(\mathbf{x}, t) \right| d\mathbf{x} dt \quad (5.16)$$

est immédiate.

Considérons maintenant le terme $\delta_N^{(n)}(\mathbf{z})$. C'est l'erreur introduite par le calcul approché de la mesure d'une partie Borélienne de I^{s+1} . Nous écrivons $\mathbf{x} < \mathbf{y}$ si $\forall i \ x_i < y_i$. Définissons

$$\mathcal{E}_{\mathbf{z}}^{(n)} = \bigcup_{\mathbf{x}_{\mathbf{a}}^{(n)} < \mathbf{z} + \Delta x_i \mathbf{e}^{(i)}} I'_{\mathbf{a}} \times I_{-i} \cup \bigcup_{\mathbf{x}_{\mathbf{a}}^{(n)} < \mathbf{z}} I'_{\mathbf{a}} \times I_0 \cup \bigcup_{\mathbf{x}_{\mathbf{a}}^{(n)} < \mathbf{z} - \Delta x_i \mathbf{e}^{(i)}} I'_{\mathbf{a}} \times I_i,$$

et $Y^{(n)} = \{\mathbf{y}_{nN+j} : 0 \leq j < N\}$. Alors $\Sigma_{\mathbf{z}}^{(n)}$ est la fonction caractéristique de $\mathcal{E}_{\mathbf{z}}^{(n)} \Gamma Y^{(n)}$ est un réseau- $(0, m, s+1)$ en base b et

$$\delta_N^{(n)}(\mathbf{z}) = \frac{A(\mathcal{E}_{\mathbf{z}}^{(n)}, Y^{(n)})}{N} - \lambda_{s+1}(\mathcal{E}_{\mathbf{z}}^{(n)}). \quad (5.17)$$

Partitionnons $\mathcal{E}_{\mathbf{z}}^{(n)}$ en ensembles disjoints

$$E_{\mathbf{z},i}^{(n)} = \bigcup_{\mathbf{x}_{\mathbf{a}}^{(n)} < \mathbf{z}^{(i)}} I'_{\mathbf{a}} \times I_i, \quad -s \leq i \leq s,$$

où $\mathbf{z}^{(i)} = \mathbf{z} - \varepsilon(i) \Delta x_{|i|} \mathbf{e}^{(|i|)}$. Si $J_{-s-1} = \emptyset \Gamma$

$$J_{-i} = \left[0, \sum_{h=i}^s D_h \frac{\Delta t}{\Delta x_h^2} \right), \quad 1 \leq i \leq s, \quad J_i = \left[0, 1 - \sum_{h=i+1}^s D_h \frac{\Delta t}{\Delta x_h^2} \right), \quad 0 \leq i \leq s,$$

$E_{\mathbf{z},i}^{(n)}$ est la différence entre les ensembles

$$F_{\mathbf{z},i}^{(n)} = \bigcup_{\mathbf{x}_{\mathbf{a}}^{(n)} < \mathbf{z}^{(i)}} I'_{\mathbf{a}} \times J_i \quad \text{et} \quad G_{\mathbf{z},i}^{(n)} = \bigcup_{\mathbf{x}_{\mathbf{a}}^{(n)} < \mathbf{z}^{(i)}} I'_{\mathbf{a}} \times J_{i-1}.$$

Par conséquent

$$\begin{aligned} \delta_N^{(n)}(\mathbf{z}) &= \sum_{i=-s}^s \left(\frac{A(F_{\mathbf{z},i}^{(n)}, Y^{(n)})}{N} - \lambda_{s+1}(F_{\mathbf{z},i}^{(n)}) \right) \\ &\quad - \sum_{i=-s+1}^s \left(\frac{A(G_{\mathbf{z},i}^{(n)}, Y^{(n)})}{N} - \lambda_{s+1}(G_{\mathbf{z},i}^{(n)}) \right). \end{aligned} \quad (5.18)$$

Puisque $F_{\mathbf{z},s}^{(n)}$ peut s'écrire comme une union disjointe d'intervalles élémentaires en base b de volume b^{-m} il vient

$$\frac{A(F_{\mathbf{z},s}^{(n)}, Y^{(n)})}{N} - \lambda_{s+1}(F_{\mathbf{z},s}^{(n)}) = 0. \quad (5.19)$$

Il reste à borner les autres termes de $\delta_N^{(n)}(\mathbf{z})$. Les estimations qui suivent sont dues à Niederreiter [47ΓLemme 3.4].

LEMME 5.1 *Soit X un réseau- (t, m, s) en base b .*

(i) *Pour tout intervalle J contenu dans un intervalle élémentaire en base b de mesure b^{t-m} , on a*

$$\left| \frac{1}{b^m} A(J, X) - \lambda_s(J) \right| \leq b^{t-m}. \quad (5.20)$$

(ii) *Pour tout intervalle élémentaire $J' \subset I^{s-1}$ en base b et pour tout $\xi_s \in \bar{I}$, on a*

$$\left| \frac{1}{b^m} A(J' \times [0, \xi_s), X) - \lambda_s(J' \times [0, \xi_s)) \right| \leq b^{t-m}. \quad (5.21)$$

Le résultat suivant nous est également utile. La démonstration est semblable à celle de [47ΓThéorème 3.6].

LEMME 5.2 *Soit X un réseau- $(t, m, s+1)$ en base b et $d \geq t$ un entier. Pour tout intervalle élémentaire $J \subset I^{s-1}$ en base b avec $\lambda_{s-1}(J) = b^{d-m}$, pour tout entier a avec $0 \leq a \leq b^d$ et pour tout $\xi_{s+1} \in \bar{I}$, on a*

$$\begin{aligned} & \left| \frac{1}{b^m} A \left(J \times \left[0, \frac{a}{b^d} \right) \times [0, \xi_{s+1}), X \right) - \lambda_{s+1} \left(J \times \left[0, \frac{a}{b^d} \right) \times [0, \xi_{s+1}) \right) \right| \\ & \leq \left[\frac{b-1}{2}(d-t) + \frac{3}{2} \right] b^{t-m}. \end{aligned} \quad (5.22)$$

Démonstration. Pour $d = t$ (5.22) se déduit de (5.20) et si $a = b^d$ alors (5.22) se déduit de (5.21). Supposons que $d > t$ et $a < b^d$ et posons

$$K = J \times \left[0, \frac{a}{b^d} \right) \times [0, \xi_{s+1}).$$

L'écriture du nombre a dans la base b est

$$a = \sum_{j=0}^{d-1} \alpha_j b^j.$$

L'intervalle $\left[0, \frac{a}{b^d} \right)$ peut s'écrire comme une union disjointe

$$\left[0, \frac{a}{b^d} \right) = I_{t-1} \cup \bigcup_{j=t}^{d-1} \bigcup_{k=0}^{\alpha_j - 1} I_{j,k}, \quad (5.23)$$

où

$$\begin{aligned} I_{t-1} &= \left[\sum_{j=t}^{d-1} \alpha_j b^{j-d}, ab^{-d} \right), \\ I_{j,k} &= \left[\sum_{\ell=j+1}^{d-1} \alpha_\ell b^{\ell-d} + kb^{j-d}, \sum_{\ell=j+1}^{d-1} \alpha_\ell b^{\ell-d} + (k+1)b^{j-d} \right). \end{aligned}$$

Ainsi

$$K = J \times I_{t-1} \times [0, \xi_{s+1}) \cup \bigcup_{j=t}^{d-1} \bigcup_{k=0}^{\alpha_j-1} J \times I_{j,k} \times [0, \xi_{s+1}). \quad (5.24)$$

Par le Lemme 5.1 (i) Γ

$$\left| \frac{1}{b^m} A(J \times I_{t-1} \times [0, \xi_{s+1}), X) - \lambda_{s+1}(J \times I_{t-1} \times [0, \xi_{s+1})) \right| \leq b^{t-m}, \quad (5.25)$$

et par le Lemme 5.1 (ii) Γ

$$\left| \frac{1}{b^m} A(J \times I_{j,k} \times [0, \xi_{s+1}), X) - \lambda_{s+1}(J \times I_{j,k} \times [0, \xi_{s+1})) \right| \leq b^{t-m}, \quad (5.26)$$

D'où

$$\left| \frac{1}{b^m} A(K, X) - \lambda_{s+1}(K) \right| \leq (\alpha + 1)b^{t-m}, \quad (5.27)$$

avec

$$\alpha = \sum_{j=t}^{d-1} \alpha_j.$$

Un traitement analogue de

$$L = J \times \left[\frac{a}{b^d}, 1 \right) \times [0, \xi_{s+1})$$

conduit à

$$\left| \frac{1}{b^m} A(L, X) - \lambda_{s+1}(L) \right| \leq ((b-1)(d-t) - \alpha + 1)b^{t-m}. \quad (5.28)$$

De plus $\Gamma K \cup L = J \times I \times [0, \xi_{s+1}) \Gamma$ alors par le Lemme 5.1 (ii) Γ

$$\left| \frac{1}{b^m} A(K \cup L, X) - \lambda_{s+1}(K \cup L) \right| \leq b^{t-m}, \quad (5.29)$$

et par suite

$$\left| \frac{1}{b^m} A(K, X) - \lambda_{s+1}(K) \right| \leq ((b-1)(d-t) - \alpha + 2)b^{t-m}. \quad (5.30)$$

En prenant le maximum sur $0 \leq \alpha \leq (b-1)(d-t)$ du minimum des membres de droite de (5.27) et (5.30) on obtient le résultat du lemme. ■

Nous sommes maintenant en mesure de démontrer le résultat qui suit.

PROPOSITION 5.1 *Pour $-s \leq i < s$ et $-s < j \leq s$, on a les estimations*

$$\begin{aligned} \left| \frac{A(F_{\mathbf{z},i}^{(n)}, Y^{(n)})}{N} - \lambda_{s+1}(F_{\mathbf{z},i}^{(n)}) \right| &\leq \lambda_1(J_i) \sum_{h=1}^s \frac{1}{b^{d_h}} + \left[\frac{b-1}{2} d_s + \frac{3}{2} \right] \frac{1}{b^{d_s}}, \\ \left| \frac{A(G_{\mathbf{z},j}^{(n)}, Y^{(n)})}{N} - \lambda_{s+1}(G_{\mathbf{z},j}^{(n)}) \right| &\leq \lambda_1(J_{j-1}) \sum_{h=1}^s \frac{1}{b^{d_h}} + \left[\frac{b-1}{2} d_s + \frac{3}{2} \right] \frac{1}{b^{d_s}}. \end{aligned}$$

Démonstration. Pour $z_1 \in \mathbb{R}\Gamma$ définissons

$$\begin{aligned} \underline{A}_1^{(n)}(z_1) &= \{\alpha_1 : \forall(\alpha_2, \dots, \alpha_s) \quad x_{\alpha_1, \dots, \alpha_s, 1}^{(n)} < z_1\}, \\ \underline{a}_1^{(n)}(z_1) &= \max \underline{A}_1^{(n)}(z_1) + 1, \\ \overline{A}_1^{(n)}(z_1) &= \{\alpha_1 : \exists(\alpha_2, \dots, \alpha_s) \quad x_{\alpha_1, \dots, \alpha_s, 1}^{(n)} < z_1\}, \\ \overline{a}_1^{(n)}(z_1) &= \max \overline{A}_1^{(n)}(z_1) + 1, \end{aligned}$$

où $\max \mathcal{N} = -1\Gamma$ si $\mathcal{N} = \emptyset$. Les points $\mathbf{x}_{\mathbf{a}}^{(n)}$ étant ordonnés de telle manière que

$$a_1 < b_1 \Rightarrow x_{\mathbf{a}, 1}^{(n)} \leq x_{\mathbf{b}, 1}^{(n)},$$

on a

$$a_1 < \underline{a}_1^{(n)}(z_1) \Rightarrow x_{\mathbf{a}, 1}^{(n)} < z_1 \Rightarrow a_1 < \overline{a}_1^{(n)}(z_1), \quad (5.31)$$

$$\overline{a}_1^{(n)}(z_1) \leq \underline{a}_1^{(n)}(z_1) + 1. \quad (5.32)$$

Pour $0 \leq a_1 < b^{d_1}, \dots, 0 \leq a_{i-1} < b^{d_{i-1}}$ et $z_i \in \mathbb{R}\Gamma$ soit

$$\begin{aligned} \underline{A}_i^{(n)}(a_1, \dots, a_{i-1}, z_i) &= \{\alpha_i : \forall(\alpha_{i+1}, \dots, \alpha_s) \quad x_{a_1, \dots, a_{i-1}, \alpha_i, \dots, \alpha_s, i}^{(n)} < z_i\}, \\ \underline{a}_i^{(n)}(a_1, \dots, a_{i-1}, z_i) &= \max \underline{A}_i^{(n)}(a_1, \dots, a_{i-1}, z_i) + 1, \\ \overline{A}_i^{(n)}(a_1, \dots, a_{i-1}, z_i) &= \{\alpha_i : \exists(\alpha_{i+1}, \dots, \alpha_s) \quad x_{a_1, \dots, a_{i-1}, \alpha_i, \dots, \alpha_s, i}^{(n)} < z_i\}, \\ \overline{a}_i^{(n)}(a_1, \dots, a_{i-1}, z_i) &= \max \overline{A}_i^{(n)}(a_1, \dots, a_{i-1}, z_i) + 1. \end{aligned}$$

Puisque les points $\mathbf{x}_{\mathbf{a}}^{(n)}$ sont ordonnés comme suit Γ

$$a_1 = b_1, \dots, a_{i-1} = b_{i-1}, a_i < b_i \Rightarrow x_{\mathbf{a}, i}^{(n)} \leq x_{\mathbf{b}, i}^{(n)},$$

on arrive à

$$a_i < \underline{a}_i^{(n)}(a_1, \dots, a_{i-1}, z_i) \Rightarrow x_{\mathbf{a}, i}^{(n)} < z_i \Rightarrow a_i < \overline{a}_i^{(n)}(a_1, \dots, a_{i-1}, z_i), \quad (5.33)$$

$$\overline{a}_i^{(n)}(a_1, \dots, a_{i-1}, z_i) \leq \underline{a}_i^{(n)}(a_1, \dots, a_{i-1}, z_i) + 1. \quad (5.34)$$

La notation $\mathbf{a} \prec \underline{\mathbf{a}}^{(n)}(\mathbf{z})$ est utilisée dans la suite pour indiquer

$$a_1 < \underline{a}_1^{(n)}(z_1), \dots, a_s < \underline{a}_s^{(n)}(a_1, \dots, a_{s-1}, z_s),$$

et de même pour $\mathbf{a} \prec \overline{\mathbf{a}}^{(n)}(\mathbf{z})$. Définissons

$$\begin{aligned} \underline{F}_{\mathbf{z}, i}^{(n)} &= \bigcup_{\mathbf{a} \prec \underline{\mathbf{a}}^{(n)}(\mathbf{z}^{(i)})} I'_{\mathbf{a}} \times J_i, & \overline{F}_{\mathbf{z}, i}^{(n)} &= \bigcup_{\mathbf{a} \prec \overline{\mathbf{a}}^{(n)}(\mathbf{z}^{(i)})} I'_{\mathbf{a}} \times J_i, \\ \partial F_{\mathbf{z}, i}^{(n)} &= \bigcup_{\underline{a}_1^{(n)}(z_1^{(i)}) \leq a_1 < \overline{a}_1^{(n)}(z_1^{(i)})} I'_{\mathbf{a}} \times J_i \cup \dots \cup \\ &\quad \bigcup_{\underline{a}_s^{(n)}(a_1, \dots, a_{s-1}, z_s^{(i)}) \leq a_s < \overline{a}_s^{(n)}(a_1, \dots, a_{s-1}, z_s^{(i)})} I'_{\mathbf{a}} \times J_i. \end{aligned}$$

De (5.31) et (5.33) se déduit $\underline{F}_{\mathbf{z},i}^{(n)} \subset F_{\mathbf{z},i}^{(n)} \subset \overline{F}_{\mathbf{z},i}^{(n)}$. De plus $\Gamma \overline{F}_{\mathbf{z},i}^{(n)} \setminus \underline{F}_{\mathbf{z},i}^{(n)} \subset \partial F_{\mathbf{z},i}^{(n)}$. Ainsi

$$\begin{aligned} \frac{A(\underline{F}_{\mathbf{z},i}^{(n)}, Y^{(n)})}{N} - \lambda_{s+1}(\underline{F}_{\mathbf{z},i}^{(n)}) - \lambda_{s+1}(\partial F_{\mathbf{z},i}^{(n)}) &\leq \frac{A(F_{\mathbf{z},i}^{(n)}, Y^{(n)})}{N} - \lambda_{s+1}(F_{\mathbf{z},i}^{(n)}), \\ \frac{A(F_{\mathbf{z},i}^{(n)}, Y^{(n)})}{N} - \lambda_{s+1}(F_{\mathbf{z},i}^{(n)}) &\leq \frac{A(\overline{F}_{\mathbf{z},i}^{(n)}, Y^{(n)})}{N} - \lambda_{s+1}(\overline{F}_{\mathbf{z},i}^{(n)}) + \lambda_{s+1}(\partial F_{\mathbf{z},i}^{(n)}). \end{aligned} \quad (5.35)$$

L'ensemble $\underline{F}_{\mathbf{z},i}^{(n)}$ peut s'écrire comme une union disjointe d'au plus $b^{d_1} \dots b^{d_{s-1}}$ intervalles de la forme

$$\prod_{h=1}^{s-1} \left[\frac{a_h}{b^{d_h}}, \frac{a_h + 1}{b^{d_h}} \right) \times \left[0, \frac{a_s^{(n)}(a_1, \dots, a_{s-1}, z_s^{(i)})}{b^{d_s}} \right) \times J_i.$$

Du Lemme 5.2 on déduit alors

$$\left| \frac{A(\underline{F}_{\mathbf{z},i}^{(n)}, Y^{(n)})}{N} - \lambda_{s+1}(\underline{F}_{\mathbf{z},i}^{(n)}) \right| \leq \left[\frac{b-1}{2} d_s + \frac{3}{2} \right] \frac{1}{b^{d_s}}. \quad (5.36)$$

En remplaçant $\underline{F}_{\mathbf{z},i}^{(n)}$ par $\overline{F}_{\mathbf{z},i}^{(n)}$ on arrive au même résultat. D'autre part (5.32) et (5.34) montrent que

$$\lambda_{s+1}(\partial F_{\mathbf{z},i}^{(n)}) \leq \lambda_1(J_i) \sum_{h=1}^s \frac{1}{b^{d_h}}. \quad (5.37)$$

La première inégalité de la proposition découle de (5.36) et de l'estimation précédente. La démonstration de la seconde inégalité se fait de façon analogue. ■

Nous arrivons au résultat de convergence.

PROPOSITION 5.2 *L'erreur de la méthode des marches quasi-aléatoires est estimée par*

$$\begin{aligned} \mathbf{D}_N^*(X^{(n)}, u_n) &\leq \mathbf{D}_N^*(X^{(0)}, u_0) + \Delta t \int_0^{t_n} \int_{\mathbb{R}^s} \left| \frac{\partial^2 u}{\partial t^2}(\mathbf{x}, t) \right| d\mathbf{x} dt \\ &\quad + 2ns \left(\sum_{i=1}^s \frac{1}{b^{d_i}} + \left[\frac{b-1}{2} d_s + \frac{3}{2} \right] \frac{2}{b^{d_s}} \right). \end{aligned}$$

Démonstration. De la Proposition 5.1 et de (5.18) et (5.19) il vient

$$\left| \delta_N^{(n)}(\mathbf{z}) \right| \leq 2s \sum_{i=1}^s \frac{1}{b^{d_i}} + 4s \left[\frac{b-1}{2} d_s + \frac{3}{2} \right] \frac{1}{b^{d_s}}. \quad (5.38)$$

Ceci combiné avec (5.14) et (5.15) et (5.16) donne

$$\mathbf{D}_N^*(X^{(n)}, u_n) \leq \mathbf{D}_N^*(X^{(n-1)}, u_{n-1}) + \Delta t \int_{t_{n-1}}^{t_n} \int_{\mathbb{R}^s} \left| \frac{\partial^2 u}{\partial t^2}(\mathbf{x}, t) \right| d\mathbf{x} dt + \delta_N, \quad (5.39)$$

où δ_N désigne le second membre de (5.38). Le résultat de la proposition suit par induction. ■

5.4 Exemples numériques

Dans cette section nous présentons des résultats numériques illustrant la section précédente. Une méthode standard de marches aléatoires (RW) et la méthode de marches quasi-aléatoires (QRW) sont appliquées au problème (5.1)–(5.2) avec $D_i = 1$, $1 \leq i \leq s$ et

$$u_0(\mathbf{x}) = \prod_{i=1}^s \chi_{[-\frac{1}{2}, +\frac{1}{2}]}(x_i),$$

où $\chi_{[-\frac{1}{2}, +\frac{1}{2}]}$ est la fonction caractéristique de l'intervalle $[-\frac{1}{2}, +\frac{1}{2}]$. Ce qui conduit à la solution

$$\tilde{u}(\mathbf{x}, t) = \frac{1}{2^s} \prod_{i=1}^s \left(\operatorname{erf} \left(\frac{x_i + 1/2}{2\sqrt{t}} \right) - \operatorname{erf} \left(\frac{x_i - 1/2}{2\sqrt{t}} \right) \right), \quad (5.40)$$

erf étant la fonction erreur. Pour toutes les simulations les positions initiales des particules sont calculées en translatant un réseau $(0, m, s)$ en base b noté Ξ dans $[-\frac{1}{2}, +\frac{1}{2}]^s$. Alors

$$\mathbf{D}_N^*(X^{(0)}, u_0) = \mathbf{D}_N^*(\Xi). \quad (5.41)$$

Une suite $(0, s + 1)$ en base b est utilisée pour assurer les déplacements des particules.

La construction de Faure [14] a été utilisée pour produire de telles suites Γb étant le plus petit nombre premier $\geq s + 1$. L'approximation semi-discrète de l'équation de la diffusion ayant fait l'objet de beaucoup d'études en analyse numérique nous mesurons ici uniquement la discrédance par rapport à la solution \tilde{u} . D'autre part le calcul de la discrédance est très coûteux dès que l'on est en dimension $s \geq 2$. Par conséquent l'erreur sera mesurée par

$$\widetilde{\mathbf{D}}_N^*(X^{(n)}, \tilde{u}_n) = \max_{1 \leq i \leq s} \sup_{z_i \in \mathbb{R}} |\tilde{d}_N^{(n)}(z_i)|,$$

où

$$\begin{aligned} \tilde{d}_N^{(n)}(z_i) &= \frac{1}{N} \sum_{\mathbf{a}} \chi_{(-\infty, z_i)}(x_{\mathbf{a}, i}) \\ &\quad - \int_{\mathbb{R}^{s-1}} \left(\int_{-\infty}^{z_i} \tilde{u}(\mathbf{x}, t_n) dx_i \right) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_s, \end{aligned}$$

et $\chi_{(-\infty, z_i)}$ est la fonction caractéristique de l'intervalle $(-\infty, z_i)$. Une estimation du taux de convergence N^{-r} est donnée par l'*exposant de l'erreur*

$$r_N^{(n)} = -\log \left(\widetilde{\mathbf{D}}_N^*(X^{(n)}, u_n) \right) / \log N.$$

Les calculs ont été faits jusqu'à $T = 0.1$. Les Figures 5.1–5.2–5.3 et 5.4 contiennent les erreurs et les exposants d'erreur pour les deux méthodes en dimension $s = 1$ avec $\Delta x = 0.05$ et $\Delta t = 6.25E - 4$. Les résultats en dimension $s = 2$ avec $\Delta x_1 = \Delta x_2 = 0.05$ et $\Delta t = 3.125E - 4$ sont donnés sur les Figures 5.5–5.6–5.7 et 5.8. L'efficacité des suites quasi-aléatoires avec renumérotation est évident.

On observe en dimension $s = 1$ que l'erreur pour la méthode QRW avec $N = 2^{10}$ particules est nettement plus faible que celle de la méthode RW avec $N = 2^{13}$ particules.

Le taux de convergence passe de $N^{-0.55}$ pour la méthode RW à $N^{-0.85}$ pour l'approche QRW. En dimension $s = 2\Gamma$ la méthode QRW utilise 9 fois moins de particules pour atteindre le même niveau d'erreur que la méthode RW. Le taux de convergence passe de $N^{-0.50}$ pour la méthode RW à $N^{-0.65}$ pour la méthode QRW.

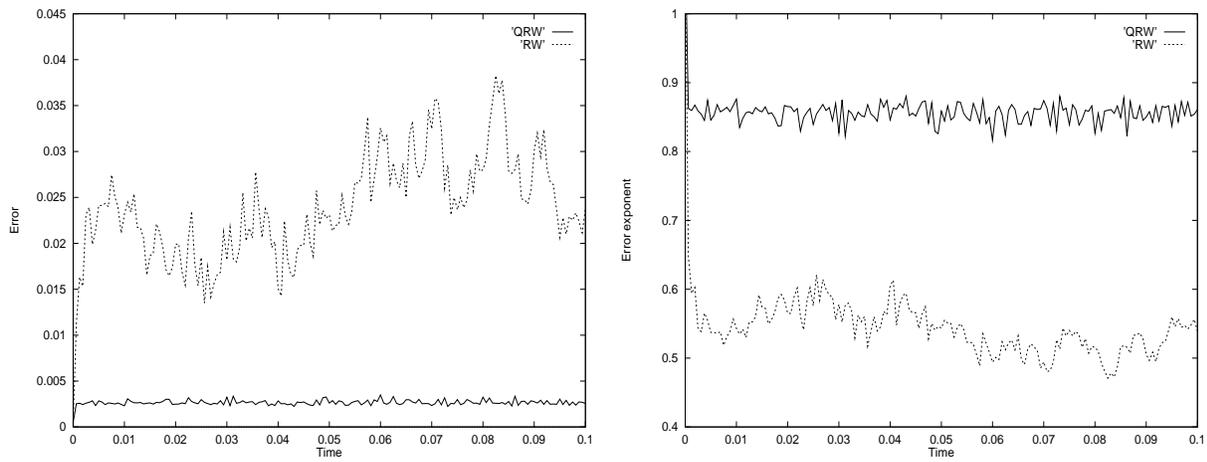


FIG. 5.1 – Dimension $s = 1$, $N = 2^{10} = 1\,024$ particules.

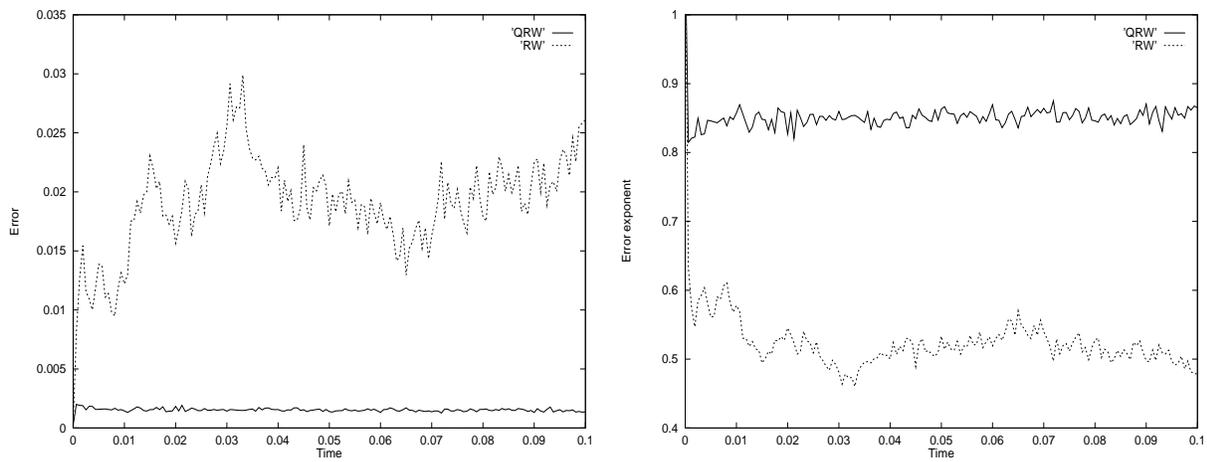


FIG. 5.2 – Dimension $s = 1$, $N = 2^{11} = 2\,048$ particules.

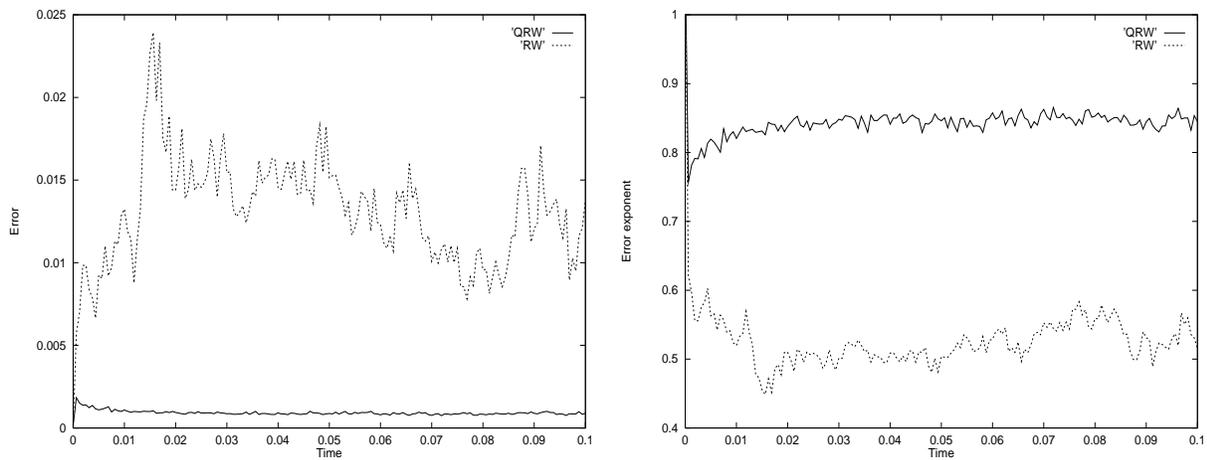


FIG. 5.3 – Dimension $s = 1$, $N = 2^{12} = 4\,096$ particules.

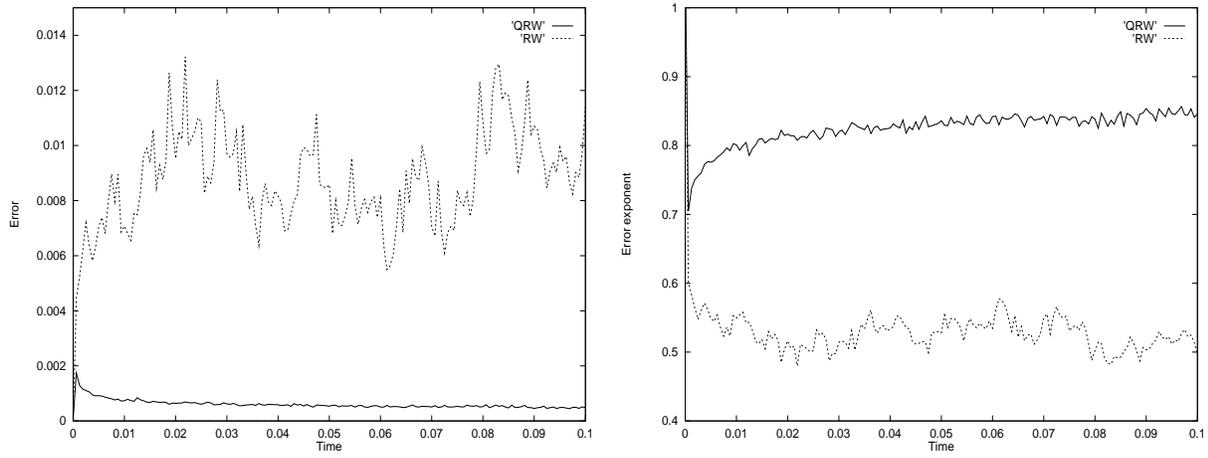


FIG. 5.4 – Dimension $s = 1$, $N = 2^{13} = 8\,192$ particles.

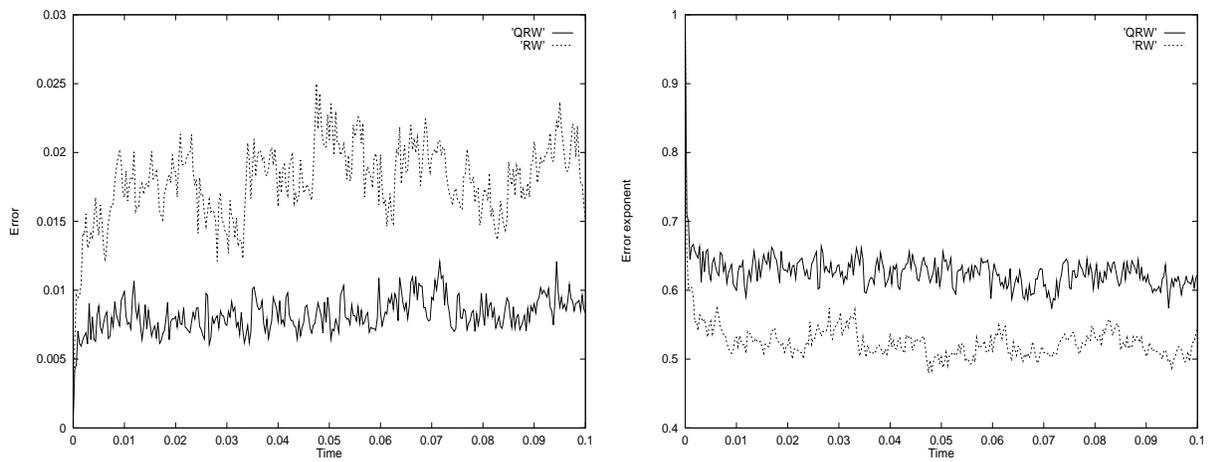


FIG. 5.5 – Dimension $s = 2$, $N = 3^{2+5} = 2\,187$ particles.

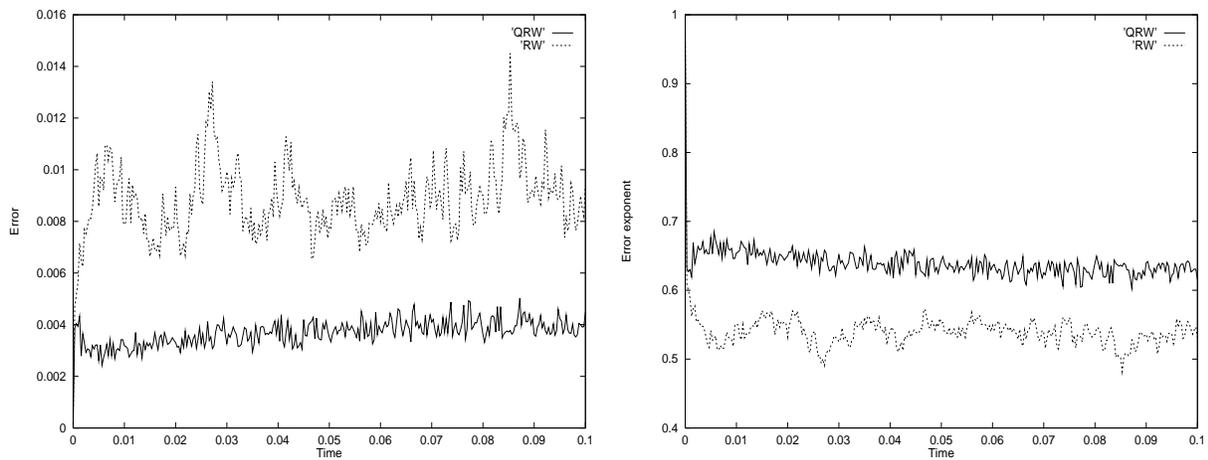


FIG. 5.6 – Dimension $s = 2$, $N = 3^{3+5} = 6\,561$ particles.

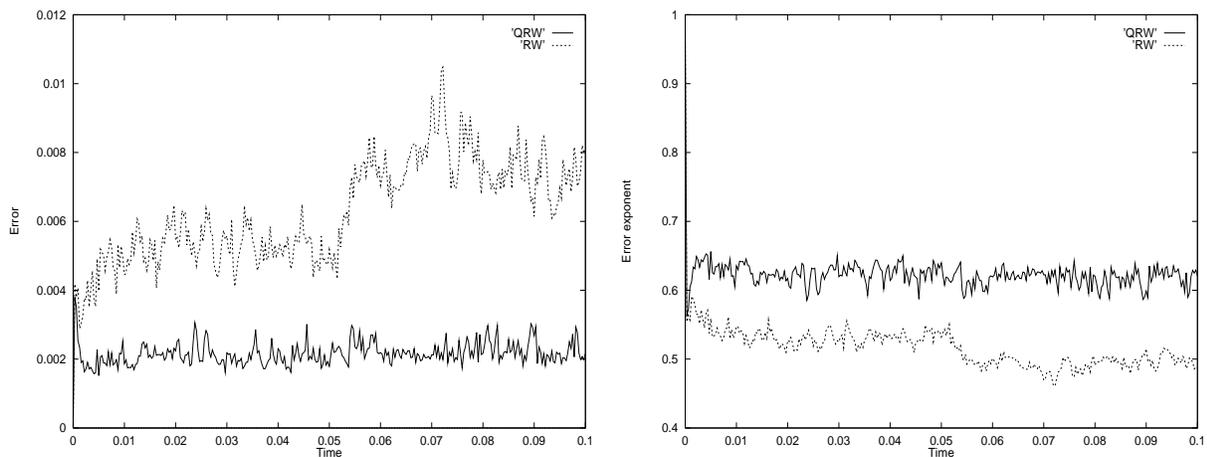


FIG. 5.7 – Dimension $s = 2$, $N = 3^{3+6} = 19\,683$ particules.

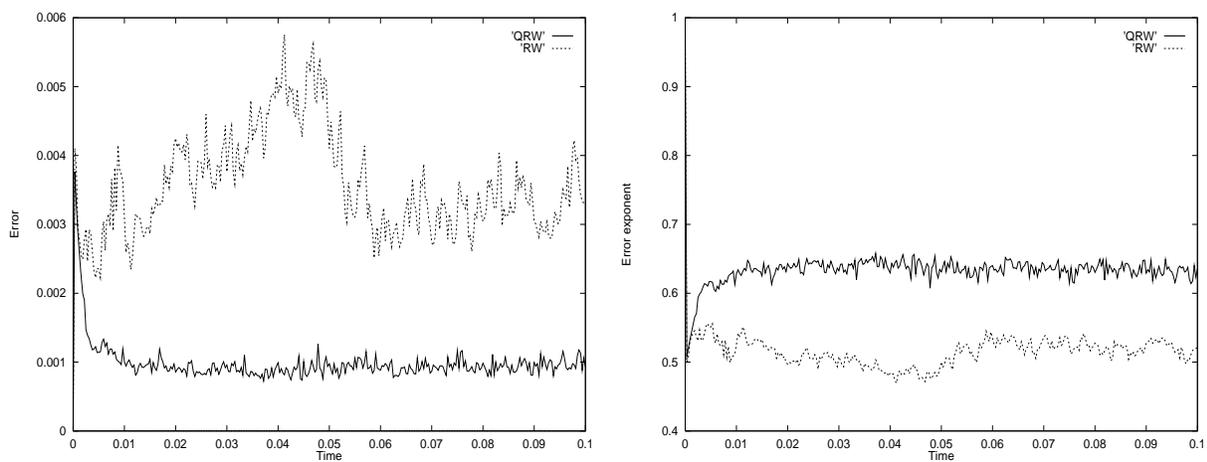


FIG. 5.8 – Dimension $s = 2$, $N = 3^{4+6} = 59\,049$ particules.

5.5 Conclusion

Nous avons présenté une méthode de marches quasi-aléatoires pour la simulation de la diffusion. Dans le cas où les particules sont renumérotées à chaque pas de temps nous avons prouvé la convergence de la solution obtenue vers celle du problème semi-discret lorsque $N \rightarrow \infty$ et $\Delta t \rightarrow 0$. La méthode a été comparée à la méthode des marches aléatoires sur un problème modèle et nous avons mis en évidence son efficacité par rapport à la méthode des marches aléatoires sur ce problème. Nous explorons actuellement l'extension de la méthodes des marches quasi-aléatoires à des équations paraboliques générales.

Bibliographie

- [1] I. A. ANTONOV, V. M. SALEEV Γ *An economic method of computing LP_τ -sequences* Γ U.S.S.R. Comp. Math. and Math. Phys. **19** Π 979 Π 252-256.
- [2] G.A. BIRD Γ *Molecular Gas Dynamics and the Direct Simulation of Gas Flows* (Clarendon Press Γ Oxford Γ 1994).
- [3] P. BRATLEY, B. L. FOX Γ *Algorithm 659: Implementing Sobol's quasirandom sequence generators* Γ ACM Trans. Math. Software **14** Π 988 Π 88-100.
- [4] P. BRATLEY, B. L. FOX, H. NIEDERREITER Γ *Implementation and tests of low-discrepancy sequences* Γ ACM Trans. on Mod. and Comp. Simulation **2** Π 992 Π 95-213.
- [5] R. E. CAFLISCH AND B. MOSKOWITZ Γ *Modified Monte Carlo methods using quasi-random sequences* Γ dans *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (H. Niederreiter and P. J.S. Shiue eds.) Π Lecture Notes in Stat. **106** Π Springer-Verlag Γ New York Γ 1995 Π 1-16.
- [6] C. CERCIGNANI Γ *The Boltzmann equation and its applications* Γ Springer-Verlag Γ New York Π 1988.
- [7] C. S. CHEN Γ *The method of fundamental solutions and the quasi-Monte Carlo method for Poisson's equation* Γ dans *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (H. Niederreiter and P. J.S. Shiue eds.) Π Lecture Notes in Stat. **106** Π Springer-Verlag Γ New York Γ 1995 Π 58-167.
- [8] I. COULIBALY, C. LÉCOT Γ *Monte Carlo and quasi-Monte Carlo algorithms for a linear integro-differential equation* Γ dans *Monte Carlo and Quasi-Monte Carlo Methods 1996* (H. Niederreiter Π P. Hellekalek Π G. Larcher Γ and P. Zinterhof Π eds.) Π Lecture Notes in Stat. **127** Π Springer-Verlag Γ New York Γ 1997.
- [9] I. COULIBALY, C. LÉCOT Γ *Simulation of diffusion using quasi-random walk methods* Γ Math. Comput. Simul. Γ à paraître.
- [10] I. COULIBALY, C. LÉCOT Γ *Particle simulations of the Kac model of the Boltzmann equation* Γ J. Comp. Appl. Math. **87** Π 997 Π 69-193.

- [11] I. COULIBALY, C. LÉCOT Γ *quasi-randomized Runge-Kutta method* Γ *Math. Comput* Γ à paraître.
- [12] M. DRMOTA Γ *Applications of uniformly distributed functions and sequences in statistical ergodic measuring techniques* Γ *J. Comp. Appl. Math.* **131** Π 990 \S 57-62.
- [13] R. DAUTRAY Γ *Méthodes probabilistes pour les équations de la physique* (Eyrolles Γ Paris Γ 1989).
- [14] H. FAURE Γ *Discrépance de suites associées à un système de numération (en dimension s)* Γ *Acta Arith.* **41** Π 982 \S 37-351.
- [15] B. L. FOX Γ *Algorithm 647: Implementation and relative efficiency of quasirandom sequence generators* Γ *ACM Trans. Math. Soft.* **12** Π 986 \S 62-376.
- [16] J. H. HALTON Γ *On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals* Γ *Numer. Math.* **12** Π 960 \S 4-90; *Berichtigung* Γ *ibid.* **12** Π 960 Π 96.
- [17] J. H. HALTON, G. B. SMITH Γ *Algorithm 247: Radical-inverse quasi-random point sequence* Γ *Comm. ACM* **17** Π 964 Γ 701-702.
- [18] J. M. HAMMERSLEY Γ *Monte Carlo methods for solving multivariable problems* Γ *Ann. New York Acad. Sci.* **186** Γ 1960 \S 44-874.
- [19] E. HLAWKA Γ *Funktionen von beschränkter Variation in der Theorie der Gleichverteilung* Γ *Ann. Mat. Pura Appl.* **154** Π 961 \S 25-333.
- [20] E. HLAWKA, R. MÜCK Γ *A transformation of equidistributed sequences* Γ dans *Applications of number theory to numerical analysis* (S. K. Zaremba ed.) Γ Academic Press Γ New York Π 972.
- [21] E. HLAWKA, F. FIRNEIS, P. ZINTERHOF Γ *Zahlentheoretische Methoden in der Numerischen Mathematik* Γ R. Oldenburg Verlag Γ Wien Γ 1981.
- [22] N. HOFMANN, P. MATHÉ Γ *On quasi-Monte Carlo simulation of stochastic differential equations* Γ *Math. Comp.* **166** Π 997 \S 73-589.
- [23] L. K. HUA, Y. WANG Γ *Applications of Number Theory to Numerical Analysis* Γ Springer Γ Berlin Π 981.
- [24] M. KAC Γ *Probability and Related Topics in Physical Sciences* Γ Interscience Publishers Γ New York Γ 1959.
- [25] A. KELLER Γ *A quasi-Monte Carlo algorithm for the global illumination problem in the radiosity setting* Γ dans *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (H. Niederreiter and P. J.S. Shiue eds.) Γ *Lecture Notes in Stat.* **106** \S Springer-Verlag Γ New York Γ 1995 \S 239-251.

- [26] A. KERSCH, W. MOROKOFF, A. SCHUSTER Γ *Radiative heat transfer with quasi-Monte Carlo methods* Γ Transp. Theory Stat. Phys. **123** Π 994 Π 001-1021.
- [27] D. E. KNUTH Γ *The art of computer programming, vol.2: Seminumerical Algorithms*. 2nd ed. Γ Addison-Wesley Γ Reading Γ Mass. Π 1981.
- [28] J. F. KOKSMA Γ *Een algemeene stelling uit de theorie der gelijkmatige verdeling modulo 1* Γ Mathematica B (Zutphen) **11** Π 942/43 Π 7-11.
- [29] L. KUIPERS AND H. NIEDERREITER Γ *Uniform Distribution of Sequences* Γ John Wiley Γ New York Γ 1974.
- [30] C. LÉCOT Γ *A direct simulation Monte Carlo scheme and uniformly distributed sequences for solving the Boltzmann equation* Γ Computing **41** Π 989 Π 41-57.
- [31] C. LÉCOT Γ *Low discrepancy sequences for solving the Boltzmann equation* Γ J. Comp. Appl. Math. **125** Π 989 Π 237-249.
- [32] C. LÉCOT Γ *An algorithm for generating low discrepancy sequences on vector computers* Γ Parallel Comput **11** Π 989 Π 13-116.
- [33] C. LÉCOT Γ *A quasi-Monte Carlo method for the Boltzmann equation* Γ Math. Comput. **156** Π 991 Π 621-644.
- [34] C. LÉCOT Γ *Error bounds for quasi-Monte Carlo integration with nets* Γ Math. Comput. **165** Π 996 Π 79-187.
- [35] C. LÉCOT, I. COULIBALY Γ *Quasi-Monte Carlo methods for kinetic equation* Γ Proceedings of the 5th International Colloquium on Numerical Analysis (E. Minchev ed.) Γ Academic Publications Γ Sofia Π 997 Π 77-86.
- [36] C. LÉCOT, I. COULIBALY Γ *A quasi-Monte Carlo scheme using nets for a linear Boltzmann equation* Γ SIAM J. Numer. Anal. **135** Π 998.
- [37] N. METROPOLIS, S. ULAM Γ *The Monte Carlo method* Γ J. Amer. Statist. Assoc. **44** Π 949 Π 335-341.
- [38] W. J. MOROKOFF, R. E. CAFLISCH Γ *A quasi-Monte Carlo approach to particle simulation of the heat equation* Γ SIAM J. Numer. Anal. **130** Π 993 Π 558-1573.
- [39] W. J. MOROKOFF, R. E. CAFLISCH Γ *Quasi-random sequences and their discrepancies* Γ SIAM J. Sci. Comput. **16** Π 994 Π 251-1279.
- [40] W. J. MOROKOFF, R. E. CAFLISCH Γ *Quasi-Monte Carlo integration* Γ J. Comput. Phys. **1122** Π 995 Π 218-230.
- [41] B. MOSKOWITZ Γ *Quasirandom diffusion Monte Carlo* Γ dans *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (H. Niederreiter and P. J.S. Shiue eds.) Γ Lecture Notes in Stat. **106** Γ Springer-Verlag Γ New York Γ 1995 Π 278-298.

- [42] B. MOSKOWITZ, R. E. CAFLISCH Γ *Modified Monte Carlo methods using quasi-random* Γ dans *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (H. Niederreiter and P. J.S. Shiue eds.) Γ Lecture Notes in Stat. **106** Γ Springer-Verlag Γ New York Γ 1995 Γ 1-16.
- [43] K. NANBU Γ *Stochastic solution method of the master equation and the model Boltzmann equation* Γ J. Phys. Soc. Japan **52** Γ 1983 Γ 2654-2658.
- [44] H. NIEDERREITER Γ *Methods for estimating discrepancy* Γ dans *Applications of number theory to numerical analysis* (S. K. Zaremba ed.) Γ Academic Press Γ New York Γ 1972 Γ 203-236.
- [45] H. NIEDERREITER, J. M. WILLS Γ *Diskrepanz und Distanz von Maßen bezüglich konvexer und Jordanscher Mengen* Γ Math. Z **144** Γ 1975 Γ 25-134; Berichtigung Γ *ibid.* **148** Γ 1976 Γ 99.
- [46] H. NIEDERREITER Γ *Quasi-Monte Carlo methods and pseudo-random numbers* Γ Bull. Amer. Math. Soc. **184** Γ 1978 Γ 957-1041.
- [47] H. NIEDERREITER Γ *Point sets and sequences with small discrepancy* Γ Monatsh. Math. **104** Γ 1987 Γ 273-337.
- [48] H. NIEDERREITER Γ *Low-discrepancy and low-dispersion sequences* Γ J. Number Theory **30** (1988) 51-70.
- [49] H. NIEDERREITER Γ *Random Number Generation and Quasi-Monte Carlo Methods* Γ SIAM Γ Philadelphia Γ 1992.
- [50] S. H. PASKOV Γ *New methodologies for valuing derivatives* Γ *Mathematics of Derivatives Securities* (S. Pliska and M. Dempster eds.) Γ Isaac Newton Institute Γ Cambridge University Press Γ 1996.
- [51] S. H. PASKOV, J. F. TRAUB Γ *Faster valuation of financial derivatives* Γ J. Portfolio Management **122** Γ 1995 Γ 113-120.
- [52] P. D. PROINOV Γ *Discrepancy and integration of continuous fonctions* Γ J. Approx. Theory **52** Γ 1988 Γ 21-131.
- [53] R. D. RICHTMYER Γ *The evaluation of definite integrals and a quasi-Monte Carlo method based on the properties of algebraic numbers* Γ Report LA-1342 Γ Los Alamos Scientific Laboratory Γ Los Alamos Γ INM Γ 1951.
- [54] P. K. SARKAR, M. A. PRASAD Γ *A comparative study of pseudo and quasi random sequences for the solution of integral equations* Γ J. Comput. Phys. **168** Γ 1987 Γ 166-88.
- [55] W. M. SCHMIDT Γ *Irregularities of distribution IX* Γ Acta Arith. **27** (1975) 385-396.

- [56] W. M. SCHMIDT Γ *Lectures on Irregularities of Distribution* Γ Tata Institute of Fundamental Research Γ Bombay Γ 1977.
- [57] J. E. H. SHAW Γ *A quasirandom approach to integration in Bayesian statistics* Γ Ann. Statist. **16** Π 988 \S 95-914.
- [58] I. H. SLOAN, S. JOE Γ *Lattice Methods for Multiple Integration* Γ Oxford University Press Γ Oxford Π 1994.
- [59] I. M. SOBOLOV Γ *On the distribution of points in a cube and the approximate evaluation of integrals* Γ U.S.S.R. Comput. Math. and Math. Phys. **17** Π 967 \S 6-112.
- [60] J. SPANIER Γ *Quasi-Monte Carlo methods for particle transport problems* Γ dans *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (H. Niederreiter and P. J.S. Shiue eds.) Π Lecture Notes in Stat. **106** \S Springer-Verlag Π New York Γ 1995 Π 21-148.
- [61] G. STENGLE Γ *Error analysis of a randomized numerical method* Γ Numer. Math. Γ 70 Π 995 Π 19-128.
- [62] J. STOER AND R. BULIRSCH Γ *Introduction to Numerical Analysis* Γ 2nd edition Γ (Springer Verlag Γ New York Γ 1993).
- [63] S. TEZUKA Γ *Uniform Random Numbers: Theory and Practice* Γ Kluwer Academic Publishers Π 1995.
- [64] J. G. VAN DER CORPUT Γ *Verteilungsfunktionen I, III* Π Nederl. Akad. Wetensch. Proc. Ser. B **38** Π 935 \S 13-821 Π 058-1066.
- [65] H. WOŹNIAKOWSKI Γ *Average case complexity of multivariate integration* Γ Bull. Amer. Math. Soc. **24** Π 991 Π 85-194.
- [66] S.K. ZAREMBA Γ *Some applications of multidimensional integration by parts* Γ Ann. Pol. Math. **21** Π 968 \S 5-96.

Résumé

Les méthodes de type quasi-Monte Carlo sont des versions déterministes des méthodes de Monte Carlo. Les nombres aléatoires sont remplacés par des nombres déterministes qui forment des ensembles ou des suites à faible discrédance ayant une meilleure distribution uniforme. L'erreur d'une méthode quasi-Monte Carlo dépend de la discrédance de la suite utilisée la discrédance étant une mesure de la déviation par rapport à la distribution uniforme.

Dans un premier temps nous nous intéressons à la résolution par des méthodes quasi-Monte Carlo d'équations différentielles pour lesquelles il y a peu de régularité en temps. Ces méthodes consistent à formuler le problème avec un terme intégral pour effectuer ensuite une quadrature quasi-Monte Carlo.

Ensuite des méthodes particulières quasi-Monte Carlo sont proposées pour résoudre les équations cinétiques suivantes: l'équation de Boltzmann linéaire et le modèle de Kac. Enfin nous nous intéressons à la résolution de l'équation de la diffusion à l'aide de méthodes particulières utilisant des marches quasi-aléatoires. Ces méthodes comportent trois étapes: un schéma d'Euler en temps une approximation particulière et une quadrature quasi-Monte Carlo à l'aide de réseaux $(0, m, s)$. À chaque pas de temps les particules sont réparties par paquets dans le cas des problèmes multi-dimensionnels ou triées si le problème est uni-dimensionnel. Ceci permet de démontrer la convergence.

Les tests numériques montrent pour les méthodes de type quasi-Monte Carlo de meilleurs résultats que ceux fournis par les méthodes de type Monte Carlo.

Mots clés: méthodes quasi-Monte Carlo méthodes particulières suites à faible discrédance suites (t, s) et réseaux (t, m, s) équations cinétiques équations différentielles équation de la diffusion.

Abstract

Quasi-Monte Carlo methods are deterministic versions of Monte Carlo methods. Random numbers are replaced by low discrepancy point sets. The error of a quasi-Monte Carlo method is estimated by means of the discrepancy of the sequence used.

First we are interested in solving ODE's with non-smooth coefficients by quasi-Monte Carlo methods. This is accomplished by treating the problem as being equivalent to the evaluation of an integral.

Next quasi-Monte Carlo particle methods are described for solving the following kinetic equations: linear Boltzmann equation and Kac equation. Finally we propose a particle scheme using quasi-random walks for diffusion equations. These methods contain three steps: an Euler scheme a particle approximation and a quasi-Monte Carlo quadrature using $(0, m, s)$ -nets. The particles are relabeled at each time step. Then convergence is proved.

The numerical experiments show that better results are obtained with quasi-Monte Carlo methods than with Monte Carlo methods.

Key words: quasi-Monte Carlo methods particle methods low discrepancy sequences (t, s) -sequences and (t, m, s) -nets kinetic equations ordinary differential equations diffusion equation.