

# Introduction à l'analyse de l'algorithme de Metropolis

GILLES LEBEAU

Département de Mathématiques,  
Université de Nice Sophia-Antipolis  
Parc Valrose 06108 Nice Cedex 02, France

[lebeau@unice.fr](mailto:lebeau@unice.fr)

1<sup>er</sup> septembre 2009

## Résumé

Les méthodes de type Monté-Carlo et les algorithmes MCMC (Markov Chain Monte Carlo) sont utilisés en pratique pour effectuer des calculs inaccessibles aux méthodes déterministes. Ces stratégies de calcul stochastique ont des applications nombreuses en physique, chimie, biologie, sciences de l'ingénieur, économie...

L'algorithme de Metropolis (et ses variantes) est un des algorithmes MCMC les plus utilisés (peut être le plus utilisé), mais son analyse mathématique en est sur bien des aspects à ses débuts.

Ce mini-cours propose une introduction à l'analyse de l'algorithme de Metropolis destiné à des analystes sans connaissances particulières en probabilités. On y expose certains liens avec l'analyse des EDP.

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Quelques rappels de probabilités</b>	<b>4</b>
<b>3</b>	<b>Chaînes de Markov</b>	<b>11</b>
3.1	Opérateur de Markov . . . . .	11
3.2	Convergence . . . . .	14
3.3	Le cas $\Omega$ ensemble fini . . . . .	21
3.4	Marche aléatoire . . . . .	24
<b>4</b>	<b>Algorithme de Metropolis</b>	<b>27</b>
4.1	Description de l'algorithme . . . . .	27
4.2	Algorithme de Metropolis de type $N$ -corps . . . . .	29
4.3	Exemples . . . . .	30
4.4	Quelques problèmes ouverts . . . . .	34
<b>5</b>	<b>Marche aléatoire sur une variété</b>	<b>35</b>
5.1	Marche aléatoire et laplacien . . . . .	35
5.2	Calcul h-pseudo-différentiel . . . . .	37
5.3	Quelques propriétés de $K_h$ . . . . .	40
5.4	Convergence vers le mouvement brownien . . . . .	42
	<b>Bibliographie</b>	<b>45</b>

# 1 Introduction

Un des problèmes les plus importants en mathématique est de savoir calculer certaines quantités, même si prouver l'existence de ces quantités peut déjà être bien difficile. L'archétype de cette problématique est de savoir calculer des intégrales, ou même simplement des moyennes :

$$A = \frac{a_1 + \dots + a_N}{N}$$

Si on peut calculer chacun des  $a_j$ , mais si on ne dispose pas de formule analytique ou de méthodes d'approximations pour la somme, dès que  $N$  est trop grand, ce calcul est simplement infaisable pour des raisons de temps de calcul.

Mais supposons que chaque  $a_j$  appartienne à  $[0, 1[$ , soit connu avec une incertitude  $10^{-3}$ , et qu'on souhaite calculer la moyenne avec la même incertitude. Alors tout se passe comme si les  $a_j$  ne pouvaient prendre que au plus  $10^3$  valeurs, et si  $N$  est très grand, disons  $N = 10^{100}$ , la redondance dans la suite des valeurs prises par les  $a_j$  va être gigantesque, de sorte qu'on peut espérer n'avoir à calculer les  $a_j$  que pour un nombre de valeurs de  $j$  raisonnable. C'est le fondement des techniques dites probabilistes qui consistent à poser le problème sous la forme : pour chaque  $k < 1000$ , quelle est la proportion  $p_k$  des  $j$  pour lesquels  $a_j \in [\frac{k}{1000}, \frac{k+1}{1000}[$  ?

Evidemment, on n'a fait que déplacer le problème, puisque si on a bien

$$A = \frac{1}{1000} \sum_{k=0}^{999} k p_k \text{ à } 10^{-3} \text{ près}$$

et que faire la somme de  $10^3$  termes est moins effrayant que faire la somme de  $10^{100}$  termes, il faut quand même être capable de calculer les  $p_k$ , i.e savoir calculer pratiquement une probabilité, ou encore, savoir choisir parmi les  $10^{100}$  valeurs de  $j$  possibles, par exemple  $10^9$  valeurs qui vont donner une bonne approximation des  $p_k$ .

L'algorithme de Metropolis est justement une méthode permettant de choisir astucieusement les  $j$  en question, c'est à dire dans un cadre plus abstrait, et avec des mots du vocabulaire courant, choisir un point au hasard pour une probabilité bien définie, mais a-priori difficilement "calculable".

Ce mini-cours est organisé comme suit :

Après un bref rappel de calcul des probabilités, on expose dans la section 3 les bases de la théorie des chaînes de Markov, en mettant l'accent sur les propriétés des opérateurs markoviens réversibles, puisque ce sera le cas pour l'algorithme de Metropolis. Le paragraphe sur la convergence vers l'équilibre contient des définitions basiques et une proposition (3.16) un peu technique mais qui illustre l'usage des inégalités de Nash. Le petit paragraphe sur les matrices stochastiques est surtout là pour rappeler le théorème classique de Perron-Frobenius sur les matrices irréductibles et apériodiques. Enfin, le dernier paragraphe de la section 3 explique ce qu'est une marche aléatoire associée à un noyau markovien, par exemple la marche aléatoire canonique sur un espace métrique mesuré.

La section 4 est surtout descriptive. On y donne la définition de l'algorithme de Metropolis, et de sa variante pour les problèmes à  $N$ -corps. Suivent quelques exemples : d'abord l'exemple historique de N. Metropolis et al. [MRR<sup>+</sup>53] pour le problème des sphères dures, puis celui relié au calcul des quantités macroscopiques associées aux mesures de Gibbs, et enfin le problème d'échantillonnage d'une densité de probabilité sur un ouvert borné de  $\mathbb{R}^d$ . La section 4 s'achève par une liste non exhaustive de problèmes ouverts.

La section 5 contient une étude plus détaillée de ce type d'algorithme dans un cas très particulier de marche aléatoire sur une variété. Les résultats exposés sont extraits d'un article (avec L. Michel), [LM] à paraître à Annals of Probability. On y expose en particulier les bases du calcul  $h$ -pseudo-différentiel, outil qui s'avère très commode pour l'analyse des opérateurs de Markov de type Metropolis local, contenant un petit paramètre  $h$  qui quantifie la taille du déplacement local à chaque pas de l'algorithme. On y prouve aussi la convergence, quand  $h \rightarrow 0$ , de la marche aléatoire vers le mouvement brownien.

Garder à l'esprit que ce mini-cours n'est pas rédigé par un probabiliste. En particulier, il ne contient rien sur les techniques purement probabilistes d'analyse des chaînes de Markov, (couplage, récurrence de Harris, ), ni les résultats abstraits sur l'ergodicité des chaînes de Markov. L'auteur n'a pas non plus de connaissances sur les finesses des applications pratiques... Par contre, il espère que le lecteur y trouvera de bonnes raisons de s'intéresser à ce type de problèmes d'analyse mathématique.

Voici enfin quelques références bibliographiques.

Pour une introduction générale aux chaînes de Markov le livre de E. Nummelin [Num84] et pour une analyse plus détaillée sur les ensembles finis ou discrets, les livres de J.-S. Liu [Liu01], de L. Brémaud [Bré99] et de L. Saloff-Coste [SC97].

Pour l'analyse des formes de Dirichlet et des processus de Markov réversibles, le livre de M. Fukushima, Y. Oshima et M. Takeda [FOT94].

Pour des résultats probabilistes sur l'analyse de l'algorithme de Metropolis, les articles [BD01], [DSC98] et leurs références. Pour l'utilisation des techniques microlocales les articles [DL09], [LM]. L'article [DLM08] pour des résultats sur le problème à  $N$  corps. Et bien sûr l'article fondateur de Metropolis et al. [MRR<sup>+</sup>53] et sa généralisation par W.-K. Hastings [Has70].

Pour l'étude du mouvement brownien et l'analyse stochastique sur les variétés, les livres de E. Hsu [Hsu03], de I. Karatzas et S.-E. Shreve [KS88], de M. Emery et P.-A. Meyer [EM89].

Pour une introduction à l'analyse semi-classique et au calcul pseudo-différentiel, les livres de A. Martinez [Mar02] et de M. Dimassi et J. Sjöstrand [DS99].

Enfin, on trouvera dans l'article de survey de P. Diaconis, [Dia], une bibliographie très complète sur le sujet, beaucoup d'exemples (cryptographie, théorie des groupes, statistique sur les variétés) et des références sur l'usage des techniques MCMC en physique, biologie, chimie, calcul scientifique...

## 2 Quelques rappels de probabilités

Soit  $\Omega$  un espace métrique, localement compact, dénombrable à l'infini et  $\mathcal{B}$  la tribu des boreliens sur  $\Omega$ . Une probabilité sur  $\Omega$  est une mesure positive  $\pi$  sur  $(\Omega, \mathcal{B})$  de masse totale 1, i.e  $\pi(\Omega) = 1$ . L'ensemble des probabilités sur  $\Omega$  est un convexe dans l'espace des mesures. Si  $A \in \mathcal{B}$  est un borelien, on appelle  $A$  un événement et on note

$$\pi(A) = \int_A d\pi \quad (2.1)$$

la probabilité de  $A$ . Deux événements  $A$  et  $B$  sont indépendants ssi  $\pi(A \cap B) = \pi(A)\pi(B)$ .

La forme linéaire sur  $C_b^0(\Omega)$  (espace des fonctions continues bornées muni de la norme sup)  $f \mapsto \int f d\pi$  est continue, positive, de norme 1, et d'après le théorème de Riesz, si  $\Phi$  est une forme linéaire sur  $C_b^0(\Omega)$ , continue, positive, telle que  $\Phi(1) = 1$ , il existe une unique probabilité  $\pi$  telle que  $\Phi(f) = \int f d\pi$ . On note, pour  $p \in [1, \infty[$ ,  $L^p(\Omega, d\pi)$  l'espace des fonctions mesurables sur  $\Omega$  à valeurs réelles telles que  $\int |f|^p d\pi < \infty$ .  $L^\infty(\Omega, d\pi)$  est l'espace des fonctions  $f$  mesurables sur  $\Omega$  à valeurs réelles, et telles qu'il existe un borelien  $A$  avec  $\pi(A) = 1$  et  $\sup_A |f| < \infty$ .  $L^\infty(\Omega)$  est l'espace des fonctions mesurables et bornées sur  $\Omega$ . On a alors les injections pour  $p > q$

$$L^\infty(\Omega) \subset L^\infty(\Omega, d\pi) \subset L^p(\Omega, d\pi) \subset L^q(\Omega, d\pi) \subset L^1(\Omega, d\pi) \quad (2.2)$$

Une fonction mesurable  $f$  de  $\Omega$  dans  $\mathbb{R}$  s'appelle une variable aléatoire (v.a). Une probabilité  $\pi$  étant donnée, on définit pour  $f \in L^1(\Omega, d\pi)$  l'espérance de  $f$  par

$$E_\pi(f) = \int f d\pi \quad (2.3)$$

et pour  $f \in L^2(\Omega, d\pi)$  la variance de  $f$  par

$$var_\pi(f) = \int f^2 d\pi - \left(\int f d\pi\right)^2 = \frac{1}{2} \int \int |f(x) - f(y)|^2 d\pi(x) d\pi(y) \quad (2.4)$$

de sorte que  $var_\pi(f)$  est le carré de la distance de  $f$  au sous-espace des fonctions constantes dans  $L^2(\Omega, d\pi)$ . L'écart type  $\sigma > 0$  est défini par  $\sigma^2 = var_\pi(f)$ .

Si  $F = (f_1, \dots, f_m)$  est une v.a à valeurs dans  $\mathbb{R}^m$ , la loi  $\mu_F$  de  $F$  est la probabilité sur  $\mathbb{R}^m$  image de  $\pi$  par  $F$ , i.e pour tout borelien  $B$  de  $\mathbb{R}^m$

$$\mu_F(B) = \pi(F^{-1}(B)) \quad (2.5)$$

et la fonction caractéristique de  $F$  est la transformée de Fourier de sa loi, i.e

$$\chi_F(\xi) = \hat{\mu}_F(\xi) = \int_{\mathbb{R}^m} e^{-ix \cdot \xi} d\mu_F = E_\pi(e^{-iF(\omega) \cdot \xi}) \quad (2.6)$$

Deux v.a à valeurs réelles  $f$  et  $g$  sont indépendantes ssi pour tous boreliens  $A, B$  de  $\mathbb{R}$ , les événements  $f^{-1}(A)$  et  $g^{-1}(B)$  sont indépendants. On a alors

$$\mu_{(f,g)} = \mu(f) \otimes \mu(g) \quad (2.7)$$

et il en résulte  $\mu(f + g) = \mu_f * \mu_g$  et  $\chi_{f+g} = \chi_f \chi_g$ .

On dispose de plusieurs notions de convergence d'une suite de v.a  $f_n$  vers une v.a  $f$ , parmi lesquelles :

- 1.  $f_n$  converge faiblement vers  $f$  ssi la suite de probabilités sur  $\mathbb{R}$   $d\mu_{f_n}$  converge faiblement vers  $d\mu_f$ , i.e pour toute fonction  $g \in C_b^0(\mathbb{R})$  on a

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} g d\mu_{f_n} = \int_{\mathbb{R}} g d\mu_f \Leftrightarrow \lim_{n \rightarrow \infty} \int_{\Omega} g(f_n) d\pi = \int_{\Omega} g(f) d\pi \quad (2.8)$$

- 2.  $f_n$  converge en probabilité vers  $f$  ssi pour tout  $\varepsilon > 0$  on a

$$\lim_{n \rightarrow \infty} \pi(|f_n - f| > \varepsilon) = 0 \quad (2.9)$$

- 3.  $f_n$  converge presque sûrement vers  $f$  ssi

$$\pi(\lim_{n \rightarrow \infty} |f_n - f| = 0) = 1 \quad (2.10)$$

La convergence presque sûre implique la convergence en probabilité qui elle même implique la convergence faible.

Les deux premiers théorèmes basiques de la théorie des probabilités sont la loi forte des grands nombres et le théorème de la limite centrale.

**Théorème 2.1 (Loi forte des grands nombres)** *Si  $f_n$  est une suite de v.a indépendantes, de même loi, avec  $E(|f_1|) < \infty$  alors la suite de v.a*

$$I_N = \frac{1}{N}(f_1 + \dots + f_N) \quad (2.11)$$

*converge presque sûrement vers la constante  $E(f_1)$ .*

**Théorème 2.2 (Théorème de la limite centrale)** *Sous les hypothèses du théorème (2.1) et si de plus  $\sigma^2 = \text{var}(f_1) < \infty$ , alors la loi de la v.a*

$$\sqrt{N}(I_N - E(f_1)) \quad (2.12)$$

*converge faiblement vers la gaussienne  $N(0, \sigma^2)$ , i.e la probabilité sur  $\mathbb{R}$  de densité  $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$ .*

La loi forte des grands nombres est à la base de la méthode de Monte Carlo. Soit  $\pi$  une probabilité sur  $\Omega$  et  $f \in L^1(\Omega, d\pi)$  une v.a. On souhaite évidemment pouvoir calculer  $E_\pi(f) = \int f d\pi$ . On introduit l'espace produit  $X = \Omega^{\mathbb{N}^*}$  dont les éléments sont les suites infinies  $\underline{x} = (x_1, \dots, x_j, \dots)$  avec  $x_j \in \Omega$ . Soit  $p_j : X \rightarrow \Omega$  l'application coordonnée  $p_j(\underline{x}) = x_j$ . On munit  $X$  de la topologie produit (la topologie la moins fine pour laquelle les applications  $p_j$  sont continues) et de la tribu produit  $\mathbb{B}$  (la tribu engendrée par les  $p_j^{-1}(\mathcal{B})$ ). Un cylindre  $C$  est un sous-ensemble de  $X$  de la forme

$$C = A_1 \times \dots \times A_N \times \Omega \times \Omega \times \dots, \quad A_j \in \mathcal{B} \quad (2.13)$$

et  $\mathbb{B}$  est la plus petite tribu qui contient tous les cylindres. Il existe alors une unique probabilité  $P = \pi^\infty$  sur  $(X, \mathbb{B})$  telle que pour tout cylindre  $C$  de la forme (5.31) on ait

$$P(C) = \prod_{j=1}^N \pi(A_j) \quad (2.14)$$

Soit  $f_n$  la v.a sur  $X$  définie par  $f_n(\underline{x}) = f(x_n)$ . Les  $f_n$  forment une suite de v.a sur  $X$  indépendantes, de même loi  $\mu_f$ , et intégrables. Par la loi forte des grands nombres, on a donc

### Théorème 2.3 (Monte Carlo)

Soit  $f \in L^1(\Omega, d\pi)$ . Pour presque toute suite  $\underline{x} \in X$  on a

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^{j=N} f(x_j) = \int_{\Omega} f d\pi \quad (2.15)$$

Le théorème précédent est de type ergodique : on a égalité entre une moyenne spatiale ( $\int_{\Omega} f d\pi$ ) et une moyenne temporelle ( $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^{j=N} f(x_j)$ ). De plus, dans le cas  $\int |f|^2 d\pi < \infty$ , le théorème de la limite centrale dit qu'avec  $I_N(\underline{x}) = \frac{1}{N} \sum_{j=1}^{j=N} f(x_j)$ , la distribution de la v.a  $\sqrt{N}(I_N - E_{\pi}(f))$  est proche de la gaussienne  $N(0, \sigma^2)$ , et donc la vitesse de convergence de  $I_N$  vers sa limite est d'ordre  $N^{-1/2}$ .

**Exemple 2.4** Soit  $\Omega = [0, 1]^d$  muni de la probabilité  $dx = dx_1 \dots dx_d$ , et  $f$  une fonction régulière sur  $\Omega$ . Si on veut estimer numériquement l'intégrale  $\int_{[0,1]^d} f(x) dx$  par une méthode de type Riemann avec  $p$  points pour discrétiser l'intervalle  $[0, 1]$ , il faudra évaluer  $f$  en  $p^d$  points, ce qui est hors de portée en dimension  $d$  grande ( $2^{50} \simeq 1.126 \times 10^{15}$ ). Malheureusement, dans de nombreuses applications,  $d$  est grand (plusieurs centaines à plusieurs milliers pour des calculs sur des molécules). Le calcul déterministe étant infaisable, on est réduit à utiliser un calcul probabiliste. Voici un exemple, avec  $d = 50$  et  $f = \sin(2\pi(x_1 + \dots + x_d))$ . Dans ce cas, on a évidemment par un calcul analytique et élémentaire  $E(f) = \int_{[0,1]^d} f(x) dx = 0$ . La figure (2.1) donne le comportement de la suite  $I_N(\underline{x})$  pour  $N \in [1, 50000]$ , pour un choix aléatoire de  $(x_1, \dots, x_{5 \cdot 10^4}) \in ([0, 1]^d)^{5 \cdot 10^4}$  qui nécessite  $50 \times 50000 = 2.5 \times 10^6$  tirages d'un point au hasard dans l'intervalle  $[0, 1]$ . La figure (2.2) illustre le comportement de la convergence en  $N^{-1/2}$  et la figure (2.3) représente l'histogramme pour 8000 expériences du type précédent et illustre le théorème de la limite centrale.

Plusieurs remarques s'imposent concernant le théorème (2.3).

D'abord, la fonction  $f$  étant donnée, la moyenne temporelle  $I_N(\underline{x}) = \frac{1}{N} \sum_{j=1}^{j=N} f(x_j)$  ne dépend que de la suite  $\underline{x} \in \Omega^{\mathbb{N}^*}$ , alors que  $\int_{\Omega} f d\pi$  dépend de la probabilité  $\pi$ . La phrase sibylline 'Pour presque toute suite  $\underline{x} \in X$ ' signifie qu'il existe un ensemble  $A \in \mathbb{B}$  de suites dans  $X = \Omega^{\mathbb{N}^*}$  avec  $P(A) = 1$  tel que 2.15 soit vrai pour toute suite  $\underline{x} \in A$ . Or la probabilité produit  $P = \pi^{\infty}$  sur  $X$  dépend très fortement de  $\pi$ . Le théorème (2.3) implique en particulier que si  $\pi_1$  et  $\pi_2$  sont deux probabilités différentes, les probabilités produit  $P_1 = \pi_1^{\infty}$  et  $P_2 = \pi_2^{\infty}$  sur  $X$  sont étrangères : Il existe  $A \in \mathbb{B}$  tel que  $P_1(A) = 1$  et  $P_2(A) = 0$ . Le fait de constater la convergence numérique de  $I_N(\underline{x})$  n'entraîne donc pas qu'on a bien estimé  $\int_{\Omega} f d\pi$ . Il est essentiel pour cela de disposer d'une méthode qui permette de choisir correctement un point au hasard pour  $\pi$  dans  $\Omega$ , ce qui est loin d'être simple en général, et est justement la raison d'être des méthodes MCMC.

De plus, la convergence dans (2.15) peut être très lente, la convergence en  $N^{-1/2}$  n'étant qu'une indication asymptotique. Reprenons l'exemple précédent ; en divisant chaque intervalle  $[0, 1]$  en deux, on écrit  $[0, 1]^{50}$  comme la réunion de  $2^{50}$  cubes de volume  $2^{-50}$ .

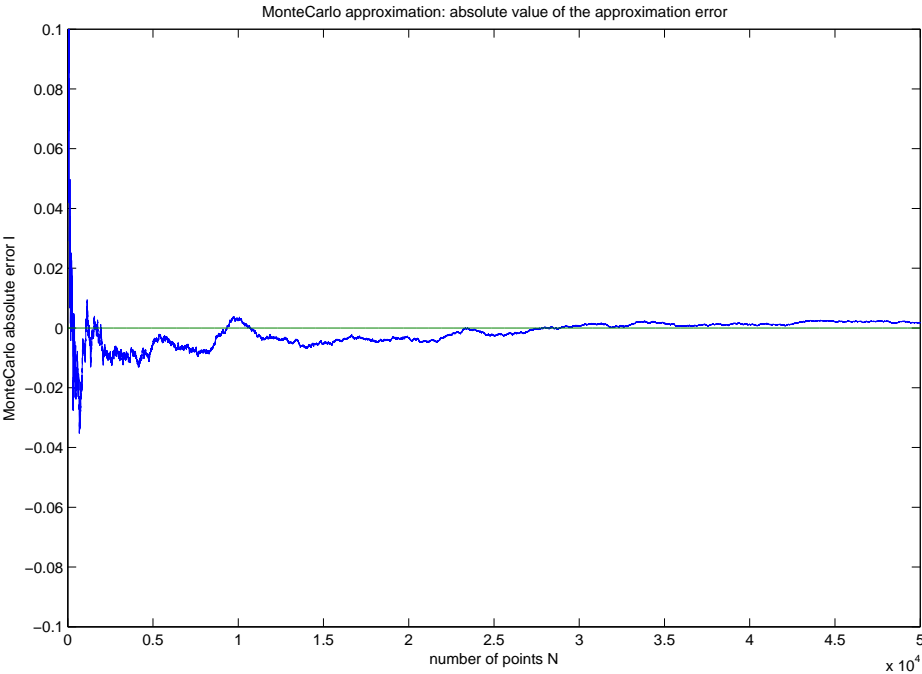


FIGURE 2.1 –

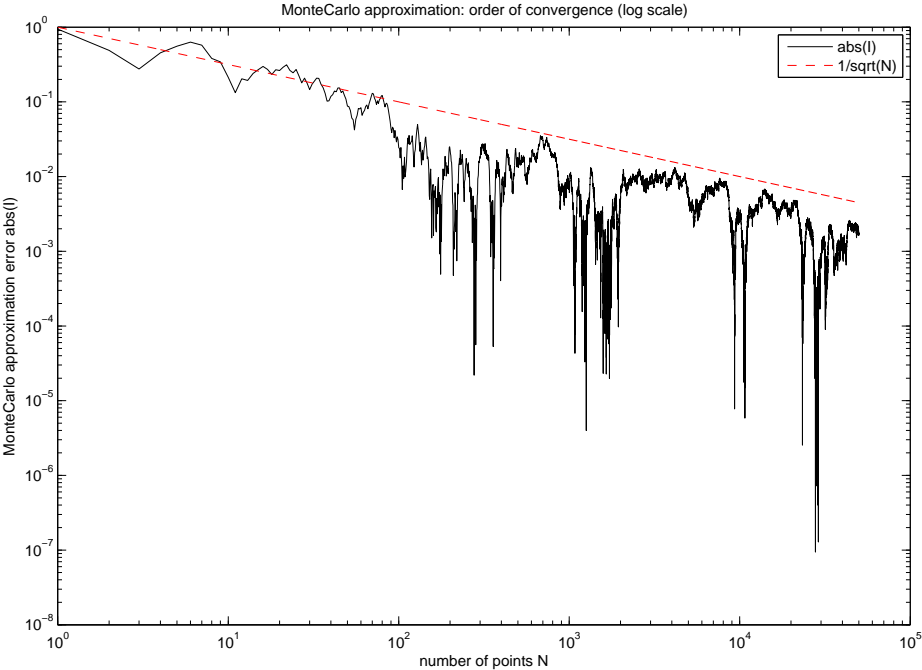


FIGURE 2.2 –



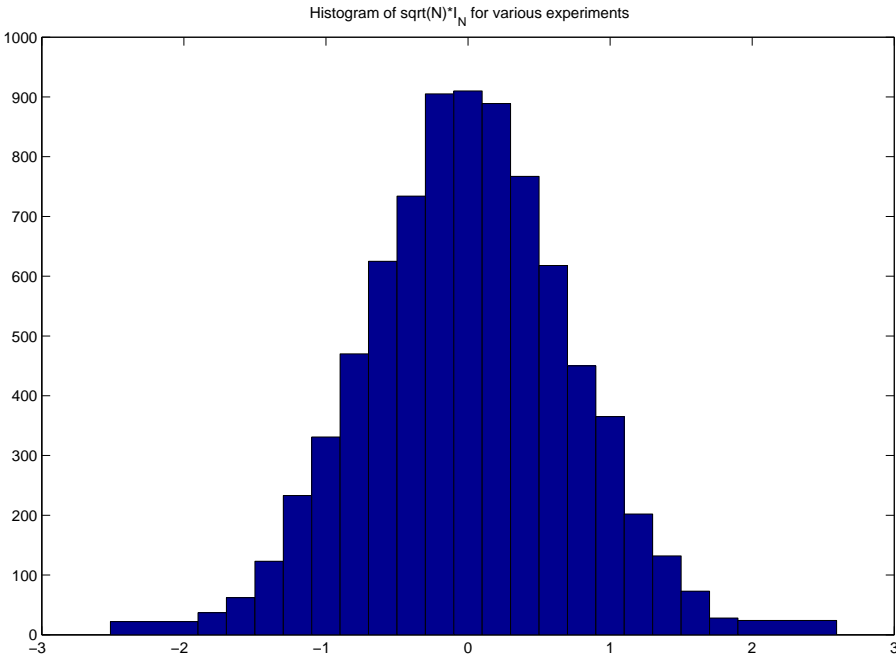


FIGURE 2.3 –

Si la fonction  $f$  est nulle sauf sur un de ces cubes, il faudra bien qu'au moins un des points  $x_j$  tombe dans ce cube si on veut estimer l'intégrale, donc on ne pourra pas faire mieux que la complexité déterministe... Il n'y a pas de miracle! Si  $f$  est très petite sur des ensembles de grande probabilité et très grande sur des ensembles de très petite probabilité, Monte Carlo ne marchera pas. On peut contourner cette difficulté en écrivant  $\int_{\Omega} f d\pi = \int_{\Omega} (f/g) g d\pi$  avec  $\int_{\Omega} g d\pi = 1$  avec un choix de  $g > 0$  qui rend  $f/g$  raisonnable, mais il faut alors savoir choisir des points au hasard pour la probabilité  $g\pi$ ...

On a évidemment  $\int f d\pi = \int y d\mu_f(y)$  où  $d\mu_f$  est la loi de  $f$ . Si on dispose de  $N$  expériences  $x_1, \dots, x_N$ , l'histogramme de  $f(x_1), \dots, f(x_N)$  fournit une loi empirique qui est bien souvent la seule information dont on dispose en pratique, de sorte que le théorème 2.3 peut être vu comme une justification de type auto-cohérence du formalisme mathématique de la théorie des probabilités.

Le théorème 2.3 implique qu'il *existe au moins une suite*  $x_1, \dots, x_n, \dots$  telle que (2.15) soit vrai, mais *il n'en exhibe aucune*... Tout le problème est de *construire une suite* telle que (2.15) soit vrai. Dans chaque situation concrète, rien n'exclut que cela puisse être fait par des techniques non-probabilistes, avec une vitesse de convergence bien plus rapide que celle qui intervient pour l'ensemble de suites de probabilité 1 donné par ce théorème, et ceci d'autant plus que les suites qui donnent une convergence rapide forment un ensemble de probabilité 0! Le médiocre est presque sûr, pas le talent...

### 3 Chaînes de Markov

#### 3.1 Opérateur de Markov

Soit  $\Omega$  un espace métrique, localement compact, dénombrable à l'infini, muni de la tribu borelienne  $\mathcal{B}$ . On note  $L^\infty(\Omega)$  l'espace des fonctions  $f$  mesurables et bornées de  $\Omega$  dans  $\mathbb{R}$  muni de la norme  $\sup_{x \in \Omega} |f(x)|$ . On note  $\mathbf{1}$  la fonction constante égale à 1 sur  $\Omega$ .

**Définition 3.1** *Un noyau de Markov sur  $\Omega$  est la donnée, pour tout  $x \in \Omega$ , d'une probabilité  $K(x, dy)$  sur  $\Omega$  telle que*

$$\forall A \in \mathcal{B}, \quad x \rightarrow K(x, A) = \int_A K(x, dy) \text{ est mesurable.} \quad (3.1)$$

Pour  $f \in L^\infty(\Omega)$ , on note  $K(f)$  la fonction sur  $\Omega$

$$K(f)(x) = \int f(y)K(x, dy) \quad (3.2)$$

Alors  $K(f)$  est mesurable, bornée, et on a

$$\begin{aligned} K(\mathbf{1}) &= \mathbf{1} \\ f \geq 0 &\implies K(f) \geq 0 \\ \|K(f)\|_{L^\infty} &\leq \|f\|_{L^\infty} \end{aligned} \quad (3.3)$$

Un opérateur  $K$  défini par (3.2) s'appelle un opérateur markovien. Un opérateur markovien  $K$  a la propriété de Feller ssi  $K$  envoie  $C_b^0(\Omega)$  dans  $C_b^0(\Omega)$ .

Si  $A$  est un borelien,  $\int_A K(x, dy)$  s'interprète comme la probabilité de transition de  $x$  vers un élément de  $A$ . Le produit  $K = K_1 K_2$  de deux opérateurs markoviens (resp markoviens et Feller) est markovien (resp markovien et Feller). En effet, si on pose pour  $A$  borelien,  $K(x, A) = K_1 K_2(\mathbf{1}_A)(x)$ ,  $x \rightarrow K(x, A)$  est mesurable, l'opérateur  $K$  défini par (3.2) est égal à  $K_1 K_2$  sur les fonctions étagées, donc aussi sur  $L^\infty(\Omega)$ , et on a  $K(x, \Omega) = 1$ . Il reste à vérifier que  $A \rightarrow K(x, A)$  est une mesure. Si  $A$  est la réunion disjointe des  $A_n \in \mathcal{B}, n \in \mathbb{N}^*$ , posons  $f_N(y) = \sum_{n=1}^N K_2(y, A_n)$ . On a  $f_N(y) \in [0, 1]$ , la suite  $f_N$  est croissante et  $\lim_{N \rightarrow \infty} f_N(y) = K_2(\mathbf{1}_A)(y)$  pour tout  $y$ . Par le théorème de convergence dominée, il en résulte

$$K(x, A) = \int K_2(\mathbf{1}_A)(y)K_1(x, dy) = \lim_{N \rightarrow \infty} \int f_N(y)K_1(x, dy) = \lim_{N \rightarrow \infty} \sum_{n=1}^N K(x, A_n)$$

**Exemple 3.2** *Soit  $F$  une application mesurable de  $\Omega$  dans  $\Omega$ . Alors  $K(g)(x) = g(F(x))$  est markovien (et Feller si  $F$  est continue). C'est le cas déterministe : on a  $K(x, dy) = \delta_{y=F(x)}$ . Les opérateurs markoviens généralisent donc les applications, et l'étude du comportement asymptotique quand  $n \rightarrow \infty$  de  $K^n$  contient en particulier la théorie des systèmes dynamiques... Sans hypothèse supplémentaire sur  $K$ , on ne pourra pas dire grand chose.*

**Exemple 3.3** *Soit  $\pi$  une probabilité sur  $\Omega$  et*

$$K(g) = \left( \int g d\pi \right) \mathbf{1} \quad (3.4)$$

Alors  $K$  est un projecteur de rang 1, markovien et Feller.

Dans toute la suite, si  $K$  est un opérateur markovien où markovien et Feller, on dira simplement que  $K$  est un opérateur de Markov, et  $K(x,dy)$  un noyau de Markov. Lorsque cela sera utile, on précisera si  $K$  est Feller.

Soit  $K$  un opérateur de Markov et  $\mathcal{M}(\Omega)$  l'espace des probabilités sur  $\Omega$ . Alors  $\mathcal{M}(\Omega)$  et  $C_b^0(\Omega)$  sont en dualité par

$$\langle \pi, g \rangle = \int g d\pi \quad (3.5)$$

et par dualité,  $K$  opère sur  $\mathcal{M}(\Omega)$  par la formule

$$\forall g \in C_b^0(\Omega), \quad \langle {}^tK(\pi), g \rangle = \int_{\Omega} K(g) d\pi \quad (3.6)$$

où on a utilisé le fait que  $K(g) \in L^\infty(\Omega)$  est intégrable pour la probabilité  $\pi$ .

**Définition 3.4** Une probabilité  $\pi$  est invariante par  $K$  ssi  ${}^tK(\pi) = \pi$ .

**Définition 3.5** Soit  $\pi$  une probabilité et  $K$  un opérateur de Markov. On dit que  $K$  est  $\pi$ -réversible ssi  $K$  est autoadjoint sur  $L^2(\Omega, d\pi)$ , c'est à dire

$$\int K(f)g d\pi = \int fK(g) d\pi, \quad \forall f, g \in L^\infty(\Omega) \quad (3.7)$$

**Lemme 3.6** Soit  $K$  un opérateur de Markov  $\pi$ -réversible. Alors  $\pi$  est invariante par  $K$  et pour tout  $p \in [1, \infty]$ ,  $K$  s'étend en opérateur continu et de norme 1 sur  $L^p(\Omega, d\pi)$ .

*Démonstration.* D'après (3.7) on a pour tout  $g \in C_b^0(\Omega)$

$$\langle {}^tK(\pi), g \rangle = \int_{\Omega} K(g) d\pi = \int_{\Omega} gK(\mathbf{1}) d\pi = \langle \pi, g \rangle \quad (3.8)$$

donc  $\pi$  est invariante. De plus, pour  $f \in L^\infty(\Omega)$ , on a

$$\|K(f)\|_{L^1(\Omega, d\pi)} = \sup_{\|g\|_{L^\infty} \leq 1} \left| \int K(f)g d\pi \right| \leq \|f\|_{L^1(\Omega, d\pi)}$$

Comme  $L^\infty(\Omega)$  est dense dans  $L^1(\Omega, d\pi)$ , ceci prouve que  $K$  s'étend de manière unique en opérateur continu sur  $L^1(\Omega, d\pi)$  et qu'on a  $\|K(f)\|_{L^1} \leq \|f\|_{L^1}$ . En particulier, si  $\pi(A) = 0$ ,  $\|K(\mathbf{1}_A)\|_{L^1} = 0$ , donc  $K(\mathbf{1}_A)$  est nul  $\pi$  presque partout, donc  $K$  opère sur  $L^\infty(\Omega, d\pi)$ . Le cas  $p \in ]1, \infty[$  s'obtient par interpolation, et le fait que  $K$  soit de norme exactement égale à 1 sur  $L^p(\Omega, d\pi)$  résulte de  $K(\mathbf{1}) = \mathbf{1}$ .  $\square$

Lorsque  $K$  est  $\pi$ -réversible,  $K$  définit donc un opérateur auto-adjoint sur  $L^2(\Omega, d\pi)$ , de norme 1. On notera  $\text{Spec}(K)$  le spectre de  $K$  agissant sur  $L^2(\Omega, d\pi)$ . On a  $\text{Spec}(K) \subset [-1, 1]$ , et  $1 \in \text{Spec}(K)$  puisque  $K(\mathbf{1}) = \mathbf{1}$ . On notera  $(f|g)_\pi = \int f\bar{g} d\pi$  le produit scalaire de deux fonctions  $f, g$  à valeurs complexes.

**Définition 3.7** Soit  $K$  un opérateur de Markov  $\pi$ -réversible. Sa forme de Dirichlet  $\mathcal{E}(f, g)$  est

$$\mathcal{E}(f, g) = ((Id - K)f|g)_\pi \quad (3.9)$$

On a

$$\mathcal{E}(f, f) = \frac{1}{2} \int \int |f(x) - f(y)|^2 K(x, dy) d\pi(x) \quad (3.10)$$

Comme  $Id - K$  est auto-adjoint positif sur  $L^2(\Omega, d\pi)$ , on a évidemment  $\mathcal{E}(f, f) \geq 0$ , et comme  $K(\mathbf{1}) = \mathbf{1}$ ,  $\mathcal{E}(\mathbf{1}, \mathbf{1}) = 0$ . L'opérateur  $Id - K$  s'appelle le laplacien (positif) associé à  $K$ .

**Définition 3.8** Soit  $K$  un opérateur de Markov  $\pi$ -réversible. On dit que  $K$  vérifie l'inégalité de Poincaré avec constante  $A$  ssi on a pour tout  $f \in L^2(\Omega, d\pi)$

$$\text{var}_\pi(f) \leq A \mathcal{E}(f, f) \quad (3.11)$$

L'inégalité (3.11) équivaut à

$$\int \int |f(x) - f(y)|^2 d\pi(y) d\pi(x) \leq A \int \int |f(x) - f(y)|^2 K(x, dy) d\pi(x) \quad (3.12)$$

Soit  $L_{\pi,0}^2 = \{f \in L^2(\Omega, d\pi), \int f d\pi = 0\}$  le sous espace de  $L^2(\Omega, d\pi)$  orthogonal à  $\mathbf{1}$ . Comme on a  $\mathcal{E}(\mathbf{1}, g) = 0$  pour tout  $g \in L^2(\Omega, d\pi)$ , l'inégalité (3.11) équivaut aussi à

$$\forall f \in L_{\pi,0}^2, \quad 2 \int |f|^2 d\pi \leq A \int \int |f(x) - f(y)|^2 K(x, dy) d\pi(x) \quad (3.13)$$

**Lemme 3.9** Soit  $K$  un opérateur de Markov  $\pi$ -réversible. Les deux propriétés suivantes sont équivalentes.

- (i)  $K$  vérifie l'inégalité de Poincaré avec constante  $A$ .
- (ii)  $1$  est valeur propre simple de  $K$  et  $\text{Spec}(K) \setminus \{1\} \subset [-1, 1 - 1/A]$ .

*Démonstration.* Immédiat par la théorie spectrale des opérateurs bornés auto-adjoints.  $\square$

Dans la pratique, il peut être plus facile de prouver une inégalité de Poincaré pour un itéré  $K^n$  de  $K$ , et d'utiliser le lemme suivant.

**Lemme 3.10** Soit  $K$  un opérateur de Markov  $\pi$ -réversible. Si  $K^n$  vérifie l'inégalité de Poincaré avec constante  $A$ , alors  $K$  vérifie l'inégalité de Poincaré avec constante  $nA$ .

*Démonstration.*  $K^n$  est  $\pi$ -réversible, et il s'agit de vérifier qu'on a

$$((Id - K^n)f|f)_\pi = \mathcal{E}_{K^n}(f, f) \leq n\mathcal{E}_K(f, f) = n((Id - K)f|f)_\pi \quad (3.14)$$

Si on pose  $|\Delta| = (Id - K)$ , l'opérateur auto-adjoint  $|\Delta|$  est positif,  $\text{Spec}(|\Delta|) \in [0, 2]$ , et (3.14) résulte de

$$nx - 1 + (1 - x)^n \geq 0, \quad \forall x \in [0, 2]$$

$\square$

## 3.2 Convergence

Lorsque l'opérateur de Markov  $K$  possède une unique probabilité invariante  $\pi$ , on appelle  $\pi$  la mesure d'équilibre de  $K$ . On espère alors que pour toute probabilité  $\nu$ , la suite de probabilités  $({}^tK)^n(\nu)$  converge vers l'unique probabilité invariante  $\pi$ . Pour donner un sens à cette convergence, il faut disposer d'une notion de distance entre deux probabilités. On travaillera ici avec la distance en variation totale.

**Définition 3.11** Soient  $\pi_1$  et  $\pi_2$  deux probabilités. Leur distance en variation totale est

$$\|\pi_1 - \pi_2\|_{VT} = \sup_{A \in \mathcal{B}} |\pi_1(A) - \pi_2(A)| = \frac{1}{2} \sup_{f \in L^\infty, |f| \leq 1} \left| \int f d\pi_1 - \int f d\pi_2 \right| \quad (3.15)$$

On remarquera qu'on a toujours

$$\|\pi_1 - \pi_2\|_{VT} \leq 1 \quad (3.16)$$

Soit  $K$  un opérateur de Markov. Pour  $n \geq 1$  soit  $K^n(x, dy)$  le noyau de l'opérateur de Markov  $K^n$ . On a

$$K^n(x, dy) = ({}^tK)^n(\delta_{y=x}) \quad (3.17)$$

D'après (3.15), pour toute probabilité  $\pi$  on a pour tout  $x \in \Omega$

$$\|K^n(x, dy) - \pi\|_{VT} = \frac{1}{2} \sup_{f \in L^\infty, |f| \leq 1} |K^n(f)(x) - \int f d\pi| \quad (3.18)$$

Si  $\pi$  une probabilité, on notera  $\Pi_\pi$  le projecteur orthogonal sur le sous espace des constantes dans  $L^2(\Omega, d\pi)$  :

$$\Pi_\pi(f) = \left( \int f d\pi \right) \mathbf{1} = E_\pi(f) \mathbf{1} \quad (3.19)$$

Soit  $K$  un opérateur de Markov tel que  $\pi$  soit invariante par  $K$ . On a alors  $K\Pi_\pi = \Pi_\pi$  et aussi  $\Pi_\pi K = \Pi_\pi$  car  $\Pi_\pi K(f) = \left( \int K(f) d\pi \right) \mathbf{1} = \langle {}^tK(\pi), f \rangle \mathbf{1} = \langle \pi, f \rangle \mathbf{1} = \Pi_\pi(f)$ . Il en résulte pour tout entier  $n \geq 1$

$$K^n - \Pi_\pi = (K - \Pi_\pi)^n \quad (3.20)$$

Il résulte alors de (3.18) et (3.20) qu'on a :

**Lemme 3.12** Soit  $K$  un opérateur de Markov et  $\pi$  une probabilité invariante par  $K$ . Alors

$$\sup_{x \in \Omega} \|({}^tK)^n(\delta_{y=x}) - \pi\|_{VT} = \frac{1}{2} \|(K - \Pi_\pi)^n\|_{L^\infty} \quad (3.21)$$

Pour toute probabilité  $\nu$ , on a

$$\|({}^tK)^n(\nu) - \pi\|_{VT} \leq \frac{1}{2} \|(K - \Pi_\pi)^n\|_{L^\infty} \quad (3.22)$$

D'après le lemme précédent, la vitesse de convergence vers la mesure d'équilibre est contrôlée par  $\|(K - \Pi_\pi)^n\|_{L^\infty}$ .

**Lemme 3.13** La suite  $n \geq 1 \rightarrow \|(K - \Pi_\pi)^n\|_{L^\infty}$  est décroissante, et il existe  $r \in [0, 1]$  tel que

$$\lim_{n \rightarrow \infty} (\|(K - \Pi_\pi)^n\|_{L^\infty})^{1/n} = r \quad (3.23)$$

*Démonstration.* On a pour  $n \geq 1$ ,  $\|(K - \Pi_\pi)^{n+1}\|_{L^\infty} = \|K(K - \Pi_\pi)^n\|_{L^\infty} \leq \|(K - \Pi_\pi)^n\|_{L^\infty}$ . L'opérateur  $K - \Pi_\pi$  étant borné sur  $L^\infty(\Omega)$ , d'après le théorème du rayon spectral, il existe  $r \leq \|K - \Pi_\pi\|_{L^\infty} \leq 2$ , tel que  $\lim_{n \rightarrow \infty} (\|(K - \Pi_\pi)^n\|_{L^\infty})^{1/n} = r$ . Le fait qu'on a  $r \leq 1$  résulte de (3.21) qui implique  $\|(K - \Pi_\pi)^n\|_{L^\infty} \leq 2$ .  $\square$

Voici deux définitions, usuelles en probabilité, d'ergodicité pour un noyau de Markov.

**Définition 3.14** Soit  $K$  un opérateur de Markov et  $\pi$  une probabilité invariante par  $K$ . On dit que  $K$  est géométriquement ergodique s'il existe  $r < 1$  et  $c(x) \in L^1(\Omega, d\pi)$  tels que

$$\|({}^tK)^n(\delta_{y=x}) - \pi\|_{VT} \leq c(x)r^n, \quad \forall x \in \Omega, \quad \forall n \geq 1 \quad (3.24)$$

On dit que  $K$  est uniformément ergodique s'il existe  $r < 1$  et une constante  $C$  tels que

$$\sup_{x \in \Omega} \|({}^tK)^n(\delta_{y=x}) - \pi\|_{VT} \leq Cr^n, \quad \forall n \geq 1 \quad (3.25)$$

Si  $K$  est géométriquement ergodique,  $\pi$  est évidemment l'unique probabilité invariante par  $K$ . L'ergodicité uniforme implique trivialement  $\lim_{n \rightarrow \infty} (\|(K - \Pi_\pi)^n\|_{L^\infty})^{1/n} < 1$ , et réciproquement, si on a  $\lim_{n \rightarrow \infty} (\|(K - \Pi_\pi)^n\|_{L^\infty})^{1/n} = r < 1$ , il résulte de (3.21) que pour tout  $r' \in ]r, 1[$ , il existe  $C$  tel que

$$\forall n, \quad \sup_{x \in \Omega} \|({}^tK)^n(\delta_{y=x}) - \pi\|_{VT} \leq \min(1, Cr'^n) \quad (3.26)$$

donc  $K$  est uniformément ergodique. Toutefois, si on ne sait pas estimer les constantes  $C$  et  $r$  dans (3.25), (3.26) ne dit rien d'utilisable en pratique.

Il n'est pas simple en général de pouvoir estimer (et a fortiori calculer...) la norme  $\|(K - \Pi_\pi)^n\|_{L^\infty}$ . Lorsque  $K$  est  $\pi$ -réversible, une méthode naturelle consiste à utiliser des estimations  $L^2$  couplées avec des inégalités qui assurent que  $K$  ou un de ses itérés  $K^p$  est régularisant. On supposera dans la suite de ce paragraphe que l'opérateur de Markov  $K$  est  $\pi$ -réversible.

La constante  $\gamma = \|K - \Pi_\pi\|_{L^2}$  joue un rôle essentiel dans l'analyse de la vitesse de convergence vers l'équilibre. Pour avoir  $\gamma < 1$ , il faut qu'on ait  $-1 \notin \text{Spec}(K)$ , ce qui équivaut à l'existence d'une constante  $c > 0$  telle que pour tout  $f \in L^2(\Omega, d\pi)$  on ait

$$((Id + K)f|f)_\pi = \frac{1}{2} \int \int |f(x) + f(y)|^2 K(x, dy) d\pi(x) \geq c\|f\|^2 \quad (3.27)$$

et que 1 soit valeur propre simple isolée dans le spectre de  $K$ . D'après les lemmes (3.9) et (3.10), on a  $\gamma < 1$  ssi  $K^2$  vérifie une inégalité de Poincaré.

**Définition 3.15** Soit  $K$  un noyau de Markov  $\pi$ -réversible. On définit le trou spectral de  $K$  (le gap) par

$$\begin{aligned} \text{gap} &= 0 \text{ si } 1 \text{ n'est pas valeur propre simple isolée dans le spectre de } K, \\ \text{gap} &= \text{dist}(\text{Spec}(K) \setminus \{1\}, 1) \text{ sinon.} \end{aligned} \quad (3.28)$$

Soit  $K$  un opérateur de Markov  $\pi$ -réversible, vérifiant une inégalité de Poincaré, et tel que  $-1 \notin \text{Spec}(K)$ . On a alors  $\|K - \Pi_\pi\|_{L^2} = \gamma < 1$ . Comme  $K - \Pi_\pi$  est auto-adjoint, il en résulte pour tout  $n \geq 1$

$$\|(K - \Pi_\pi)^n\|_{L^2} = \gamma^n \quad (3.29)$$

Supposons de plus qu'il existe un entier  $p$  tel que  $K^p$  soit borné de  $L^2(\Omega, d\pi)$  dans  $L^\infty(\Omega)$ . Comme on a  $(K - \Pi_\pi)^n = K^n - \Pi_\pi = K^p(K^{n-p} - \Pi_\pi) = K^p(K - \Pi_\pi)^{n-p}$  pour  $n > p$ , on obtient alors en utilisant aussi  $\|f\|_{L^2} \leq \|f\|_{L^\infty}$

$$\|(K - \Pi_\pi)^n\|_{L^\infty} \leq \gamma^{n-p} \|K^p\|_{L^2 \rightarrow L^\infty} \quad (3.30)$$

Le fait qu'il existe un entier  $p$  tel que  $K^p$  soit borné de  $L^2(\Omega, d\pi)$  dans  $L^\infty(\Omega)$  dépend évidemment de la structure du noyau  $K$ . Ce ne sera pas le cas pour un noyau de type Metropolis agissant sur un ouvert de  $\mathbb{R}^d$ . Même quand cela a lieu, la constante  $\|K^p\|_{L^2 \rightarrow L^\infty}$  peut être trop grande pour que l'estimation (3.30) soit utilisable.

### Une technique d'estimation de la vitesse de convergence adaptée aux opérateurs de type Metropolis.

On explicite dans ce paragraphe une méthode de type  $L^2$  pour estimer la vitesse de convergence vers l'équilibre pour un noyau de Markov  $K$   $\pi$ -réversible. On suppose d'abord que  $K$  s'écrit sous la forme  $K = m + \mathcal{K}_1$ , avec  $\mathcal{K}_1$  opérateur borné auto-adjoint sur  $L^2(\Omega, d\pi)$ ,  $m$  un multiplicateur  $m(f)(x) = m(x)f(x)$  avec  $m \in L^\infty(\Omega)$ , et qu'on a

$$\begin{aligned} \sup_{x \in \Omega} |m(x)| &= \alpha < 1 \\ \|\mathcal{K}_1\|_{L^2 \rightarrow L^\infty} &= C_{2,\infty} < \infty \end{aligned} \quad (3.31)$$

On a alors pour tout  $p \geq 1$ , en utilisant  $\|K\|_{L^2 \rightarrow L^2} = 1$ ,

$$\begin{aligned} K^p &= m^p + \mathcal{K}_p, \quad \mathcal{K}_p = m\mathcal{K}_{p-1} + \mathcal{K}_1 K^{p-1} \\ \|m^p\|_{L^\infty \rightarrow L^\infty} &\leq \alpha^p \\ \|\mathcal{K}_p\|_{L^2 \rightarrow L^\infty} &\leq C_{2,\infty}(1 + \alpha + \dots + \alpha^{p-1}) \leq \frac{C_{2,\infty}}{1 - \alpha} \end{aligned} \quad (3.32)$$

On suppose ensuite qu'il existe des constantes  $\alpha < 1 - \delta < \gamma < 1$  telles qu'on ait

$$\begin{aligned} \text{gap} &= 1 - \gamma \\ \text{Spec}(K) &\subset [-1 + \delta, 1] \\ \text{Spec}(K) &\text{ est discret dans } [1 - \delta, 1] \end{aligned} \quad (3.33)$$

On note alors

$$\lambda_1 = 1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_M \geq 1 - \delta > 0 \quad (3.34)$$

les valeurs propres avec multiplicité de  $K$  dans l'intervalle  $[1 - \delta, 1]$ . On notera  $(e_k)_{1 \leq k \leq M}$ , avec  $K(e_k) = \lambda_k e_k$ , une base orthonormale associée de vecteurs propres dans  $L^2(\Omega, d\pi)$ . On a  $e_1 = \mathbf{1}$ ,  $\lambda_2 = \gamma$ , et le spectre essentiel de  $K$  est contenu dans  $[-1 + \delta, 1 - \delta]$ . On a aussi

$$\|K - \Pi_\pi\|_{L^2} = \gamma \quad (3.35)$$



Pour tout  $k$ , on a  $\lambda_k > \alpha$ , donc  $\lambda_k - m$  est inversible sur  $L^\infty$ , et on a  $(\lambda_k - m)^{-1} \mathcal{K}_1(e_k) = e_k$ , d'où en utilisant (3.31),  $e_k \in L^\infty$  et

$$\|e_k\|_{L^\infty} \leq \frac{1}{\lambda_k - \alpha} C_{2,\infty} \quad (3.36)$$

Pour  $2 \leq N \leq M$ , soit  $E_N$  l'espace vectoriel engendré par les  $e_k$  pour  $2 \leq k \leq N$ . L'espace  $E_N$  est stable sous l'action de  $K$ . Pour  $f = \sum f_k e_k \in E_N$ , on a  $(K(f)|f) = \sum \lambda_k |f_k|^2 \geq \sum \lambda_k^2 |f_k|^2 = \|K(f)\|_{L^2}^2$ , d'où

$$\forall f \in E_N, \quad \mathcal{E}(f, f) = ((Id - K)f|f)_\pi \leq \|f\|_{L^2}^2 - \|K(f)\|_{L^2}^2 \quad (3.37)$$

On suppose enfin qu'il existe  $p > 2$ ,  $N \in [2, M]$ , et des constantes  $A > 0, h > 0$  tels qu'on ait l'inégalité de Sobolev suivante,

$$\forall f \in E_N, \quad \|f\|_{L^p}^2 \leq Ah^{-2} \left( \mathcal{E}(f, f) + h^2 \|f\|_{L^2}^2 \right) \quad (3.38)$$

D'après l'inégalité de Holder, on a  $\|f\|_{L^2}^2 \leq \|f\|_{L^p}^{\frac{p}{p-1}} \|f\|_{L^1}^{\frac{p}{p-1}}$ , d'où avec  $1/D = 2 - 4/p > 0$

$$\forall f \in E_N, \quad \|f\|_{L^2}^{2+1/D} \leq Ah^{-2} \left( \mathcal{E}(f, f) + h^2 \|f\|_{L^2}^2 \right) \|f\|_{L^1}^{1/D} \quad (3.39)$$

En utilisant (3.37), on obtient donc avec  $D \in ]1/2, \infty[$

$$\forall f \in E_N, \quad \|f\|_{L^2}^{2+1/D} \leq Ah^{-2} \left( \|f\|_{L^2}^2 - \|K(f)\|_{L^2}^2 + h^2 \|f\|_{L^2}^2 \right) \|f\|_{L^1}^{1/D} \quad (3.40)$$

On appelle (3.40) une inégalité de Nash.

On a  $(K - \Pi_\pi)^n(e_2) = K^n(e_2) - \int e_2 d\pi = \gamma^n e_2$  et  $e_2 \in L^\infty$  d'après (3.36), donc

$$\|(K - \Pi_\pi)^n\|_{L^\infty} \geq \gamma^n \quad (3.41)$$

**Proposition 3.16** *Soit  $K$  un opérateur de Markov  $\pi$ -réversible, vérifiant les hypothèses (3.31), (3.33) et (3.38). Alors on a*

$$\|(K - \Pi_\pi)^n\|_{L^\infty} \leq C_n = C_{1,n} + C_{2,n} + C_{3,n}, \quad \forall n \geq 2 \quad (3.42)$$

avec

$$\begin{aligned} C_{1,n} &= \min_{m+p+j=n, p \geq 1} \gamma^m A^D (2 + C_{2,p} + C_{3,p}) (1 + 2D)^D \max\left(1, \left(\frac{1+h^{-2}}{1+j}\right)^D\right) \\ C_{2,n} &= \frac{C_{2,\infty}}{1 - \delta - \alpha} \sum_{k=N+1}^M \lambda_k^n \\ C_{3,n} &= \min_{p+j=n, p \geq 1, j \geq 1} \left( \alpha^p \left( 2 + \frac{C_{2,\infty}}{1 - \delta - \alpha} \sum_{k=2}^M \lambda_k^j \right) + \frac{C_{2,\infty}}{1 - \alpha} (1 - \delta)^j \right) \end{aligned} \quad (3.43)$$

**Remarque 3.17** Voyons comment utiliser cette proposition technique. Supposons les constantes  $\alpha, \delta, A, D$  données, que  $h$  est un petit paramètre, et qu'on a

$$\begin{aligned} c_0 h^2 &\leq 1 - \gamma \leq c_1 h^2 \\ M &\leq c_3 h^{-b} \\ C_{2,\infty} &\leq c_4 h^{-a} \\ \lambda_N &\leq 1 - c_5 h^{2-c} \end{aligned} \quad (3.44)$$

avec des constantes  $c_i, a, b, c > 0$  indépendantes de  $h$ . Alors pour  $n \geq h^{-2+c/2}$ , on obtient des deux dernières lignes de (3.43) qu'il existe  $C_0$  indépendant de  $h$  tel que  $C_{2,n} + C_{3,n} \leq C_0$ , et en choisissant  $j \simeq h^{-2}$ ,  $p \simeq h^{-2+c/2}$  dans la première ligne, on obtient qu'il existe  $C_1, C_2$  indépendants de  $h$  tel que pour  $n \geq 2h^{-2}$ , on a  $C_{1,n} \leq C_1 \gamma^{n-2h^{-2}} \leq C_1 (1 - c_1 h^2)^{-2h^{-2}} \gamma^n \leq C_2 \gamma^n$ . Comme  $\gamma \leq 1 - c_0 h^2$ , on obtient qu'il existe  $C$  indépendant de  $h$  tel que

$$\|(K - \Pi_\pi)^n\|_{L^\infty} \leq C \gamma^n \leq C (1 - c_0 h^2)^n, \quad \forall n \quad (3.45)$$

En comparant (3.41) et (3.45), on voit qu'il faut attendre  $n \simeq h^{-2}$  itération pour obtenir la convergence, et que ceci est bien donnée par la valeur du gap. Obtenir les estimations  $M \leq c_3 h^{-b}$  et  $\lambda_N \leq 1 - c_5 h^{2-c}$  correspond à obtenir des estimations de Weyl sur la fonction de comptage des valeurs propres d'un opérateur elliptique. Noter toutefois que sur un espace de grande dimension, la constante  $p$  d'injection de Sobolev sera très proche de 2, donc  $D$  très grand, de sorte que l'estimation de  $C_{1,n}$  devient très mauvaise. On a alors intérêt à remplacer l'estimation de Sobolev (3.38) par une estimation de type Log-Sobolev

$$\forall f \in E_N, \int |f(x)|^2 \log \left( \frac{|f(x)|^2}{\|f\|_{L^2}^2} \right) d\pi(x) \leq A h^{-2} \left( \mathcal{E}(f, f) + h^2 \|f\|_{L^2}^2 \right) \quad (3.46)$$

*Démonstration.* Pour prouver la proposition 3.16, on décompose  $K - \Pi_\pi$  en somme de trois opérateurs

$$\begin{aligned} K - \Pi_\pi &= K_1 + K_2 + K_3 \\ K_1(f) &= \sum_{k=2}^N \lambda_k (f|e_k)_\pi e_k, \quad K_2(f) = \sum_{k=N}^M \lambda_k (f|e_k)_\pi e_k \end{aligned} \quad (3.47)$$

Alors  $K_1$  (resp  $K_2$ , resp  $K_3$ ) est la restriction de  $K$  au sous espace spectral associé à l'intervalle  $[\lambda_N, 1[$  (resp  $[1 - \delta, \lambda_N[$ , resp  $[-1 + \delta, 1 - \delta]$ ). Les opérateurs  $K_j$  vérifient

$$\|K_1\|_{L^2} \leq \gamma, \quad \|K_2\|_{L^2} \leq \lambda_{N+1}, \quad \|K_3\|_{L^2} \leq 1 - \delta \quad (3.48)$$

On a  $K_i K_j = 0$  pour  $i \neq j$ , donc  $K K_j = K_j K = K_j^2$  et pour tout  $n \geq 1$

$$(K - \Pi_\pi)^n = K_1^n + K_2^n + K_3^n \quad (3.49)$$

On estime séparément les trois contributions dans (3.49). On a en utilisant (3.47), (3.36) et  $|(f|e_k)_\pi| \leq \|f\|_{L^2} \leq \|f\|_{L^\infty}$

$$\|(K_1 + K_2)^n(f)\|_{L^\infty} \leq \sum_{k=2}^M \lambda_k^n \|f\|_{L^\infty} \frac{C_{2,\infty}}{\lambda_k - \alpha} \quad (3.50)$$

d'où en utilisant (3.49) et  $\|(K - \Pi_\pi)^n\|_{L^\infty} = \|K^n - \Pi_\pi\|_{L^\infty} \leq 2$

$$\|K_3^n\|_{L^\infty} \leq 2 + \frac{C_{2,\infty}}{1 - \delta - \alpha} \sum_{k=2}^M \lambda_k^n \quad (3.51)$$

On obtient de même

$$\|K_2^n\|_{L^\infty} \leq \frac{C_{2,\infty}}{1 - \delta - \alpha} \sum_{k=N+1}^M \lambda_k^n \quad (3.52)$$

□

d'où l'estimation de la constante  $C_{2,n}$  dans (3.43). On va maintenant améliorer l'estimation (3.51). En utilisant (3.32), (3.48),  $KK_3 = K_3^2$ , et pour tout opérateur  $E$ ,  $\|E\|_{L^\infty \rightarrow L^\infty} \leq \|E\|_{L^2 \rightarrow L^\infty}$ , on obtient

$$\begin{aligned} \|K_3^{p+n}\|_{L^\infty \rightarrow L^\infty} &= \|K^p K_3^n\|_{L^\infty \rightarrow L^\infty} \leq \|m^p K_3^n\|_{L^\infty \rightarrow L^\infty} + \|\mathcal{K}_p K_3^n\|_{L^2 \rightarrow L^\infty} \\ &\leq \alpha^p \left( 2 + \frac{C_{2,\infty}}{1 - \delta - \alpha} \sum_{k=2}^M \lambda_k^n \right) + \|\mathcal{K}_p\|_{L^2 \rightarrow L^\infty} \|K_3^n\|_{L^2 \rightarrow L^2} \\ &\leq \alpha^p \left( 2 + \frac{C_{2,\infty}}{1 - \delta - \alpha} \sum_{k=2}^M \lambda_k^n \right) + \frac{C_{2,\infty}}{1 - \alpha} (1 - \delta)^n \end{aligned} \quad (3.53)$$

d'où l'estimation de la constante  $C_{3,n}$ . Pour estimer  $C_{1,n}$ , on utilise l'estimation de Nash (3.40). On a d'après (3.49)

$$\|K_1^n\|_{L^\infty} \leq \|K^n - \Pi_\pi\|_{L^\infty} + \|K_2^n\|_{L^\infty} + \|K_3^n\|_{L^\infty} \leq 2 + C_{2,n} + C_{3,n} \quad (3.54)$$

donc aussi, puisque  $K_1$  est auto-adjoint sur  $L^2$ ,

$$\|K_1^n\|_{L^1} \leq 2 + C_{2,n} + C_{3,n} \quad (3.55)$$

Soit alors  $g \in L^2(\Omega, d\pi)$  tel que  $\|g\|_{L^1} = 1$ . Fixons  $p \geq 1$  et considérons la suite  $(c_n)_{n \geq 0}$  définie par

$$c_n = \|K_1^{n+p}(g)\|_{L^2}^2 = \|K^n K_1^p(g)\|_{L^2}^2 \quad (3.56)$$

Comme  $\|K\|_{L^2} = 1$ , la suite  $c_n \geq 0$  est décroissante, et en utilisant (3.40) avec  $f = K_1^{n+p}(g) \in E_N$ , ainsi que  $\|f\|_{L^1} = \|K^n K_1^p(g)\|_{L^1} \leq \|K_1^p(g)\|_{L^1}$  et (3.55) on obtient

$$c_n^{1+1/2D} \leq Ah^{-2} (c_n - c_{n+1} + h^2 c_n) (2 + C_{2,p} + C_{3,p})^{1/D} \quad (3.57)$$

**Lemme 3.18** Soient  $D > 1/2$  et  $h > 0$  donnés. Soit  $B > 0$  et  $c_n \geq 0$  une suite décroissante vérifiant

$$\forall n \geq 0, \quad c_n^{1+1/2D} \leq Bh^{-2} (c_n - c_{n+1} + h^2 c_n) \quad (3.58)$$

Alors on a

$$\forall n \geq 0, \quad c_n \leq B^{2D} (1 + 2D)^{2D} \max\left(1, \left(\frac{1 + h^{-2}}{1 + n}\right)^{2D}\right) \quad (3.59)$$

*Démonstration.* En posant  $c_n = B^{2D} c'_n$ , on se ramène à  $B = 1$ . D'après (3.58), on a alors

$$0 \leq c_{n+1} \leq c_n(1 + h^2 - h^2 c_n^{1/2D}) \quad (3.60)$$

En particulier,  $c_0 \leq (1 + h^{-2})^{2D}$ , donc puisque  $C = 1 + 2D > 1$ , (3.59) est vrai pour  $n = 0$ . Supposons d'abord que (3.59) est vrai au rang  $n$ , avec  $n + 1 \leq h^{-2}$ . Si on avait  $c_{n+1} > (\frac{C(1+h^{-2})}{2+n})^{2D}$ , alors on aurait

$$\left(\frac{C(1+h^{-2})}{2+n}\right)^{2D} < c_{n+1} \leq c_n \leq \left(\frac{C(1+h^{-2})}{1+n}\right)^{2D} \quad (3.61)$$

donc en utilisant (3.60)

$$\left(\frac{C(1+h^{-2})}{2+n}\right)^{2D} < \left(\frac{C(1+h^{-2})}{1+n}\right)^{2D} \left(1 + h^2 - h^2 \frac{C(1+h^{-2})}{2+n}\right) \quad (3.62)$$

ce qui équivaut à

$$C(1+h^2) < (n+2) \left(h^2 + 1 - \left(\frac{n+1}{n+2}\right)^{2D}\right) \quad (3.63)$$

Or la fonction  $x \geq 2 \rightarrow x(1 - (1 - \frac{1}{x})^{2D})$  est croissante et tend vers  $2D$  à l'infini et on a pour  $n + 1 \leq h^{-2}$ ,  $(2+n)h^2 \leq 1 + h^2$ . Donc (3.63) contredit  $C = 1 + 2D$ . Il en résulte que (3.59) est vrai pour tout  $n \leq h^{-2}$ . Maintenant, pour  $n \geq h^{-2}$ , on a  $c_{n+1} \leq c_n \leq C^{2D}$ , donc (3.59) est vrai au rang  $n + 1$ .

Reste à étudier le cas  $h^{-2} - 1 < n < h^{-2}$ . Si on avait  $c_{n+1} > C^{2D}$ , on aurait d'après (3.60)

$$C^{2D} < \left(\frac{C(1+h^{-2})}{1+n}\right)^{2D} (1 + h^2 - h^2 C) \quad (3.64)$$

donc  $1 < (1 + h^2)^{2D}(1 + h^2 - h^2 C)$ , d'où  $C < \frac{(1+h^2)^{2D+1}-1}{h^2(1+h^2)^{2D}} \leq 1 + 2D$  ce qui contredit à nouveau la définition de  $C$ .  $\square$

D'après le lemme précédent, (3.56) et (3.57), on a donc pour tout  $n \geq 0$

$$\|K_1^{n+p}(g)\|_{L^2} = c_n^{1/2} \leq A^D(2 + C_{2,p} + C_{3,p})(1 + 2D)^D \max\left(1, \left(\frac{1+h^{-2}}{1+n}\right)^D\right) \quad (3.65)$$

donc

$$\|K_1^{n+p}\|_{L^1 \rightarrow L^2} \leq A^D(2 + C_{2,p} + C_{3,p})(1 + 2D)^D \max\left(1, \left(\frac{1+h^{-2}}{1+n}\right)^D\right) \quad (3.66)$$

donc aussi, puisque  $K_1^{n+p}$  est auto-adjoint sur  $L^2$

$$\|K_1^{n+p}\|_{L^2 \rightarrow L^\infty} \leq A^D(2 + C_{2,p} + C_{3,p})(1 + 2D)^D \max\left(1, \left(\frac{1+h^{-2}}{1+n}\right)^D\right) \quad (3.67)$$

et l'estimation de  $C_{1,n}$  de (3.43) résulte de  $\|K_1\|_{L^2} \leq \gamma$ , de (3.66) et de

$$\|K_1^{p+j+m}\|_{L^\infty \rightarrow L^\infty} \leq \|K_1^{p+j+m}\|_{L^2 \rightarrow L^\infty} \leq \|K_1^m\|_{L^2 \rightarrow L^2} \|K_1^{p+j}\|_{L^2 \rightarrow L^\infty} \leq \gamma^m \|K_1^{p+j}\|_{L^2 \rightarrow L^\infty} \quad (3.68)$$

Ceci achève la preuve de la proposition 3.16.

### 3.3 Le cas $\Omega$ ensemble fini

Dans cette section,  $\Omega$  désigne un ensemble fini à  $N$  éléments, muni de la topologie discrète et de la tribu des parties de  $\Omega$ .

$$\Omega = \{x_1, \dots, x_N\} \quad (3.69)$$

Une probabilité  $\pi$  sur  $\Omega$  est simplement la donnée de  $N$  nombres réels  $\pi(x_i) \geq 0$  tels que

$$\sum_i \pi(x_i) = 1 \quad (3.70)$$

et un opérateur de Markov est une matrice  $N \times N$ ,  $K = K_{i,j}$ , telle que

$$\forall i, j \quad K_{i,j} \geq 0 \quad \text{et} \quad \forall i \quad \sum_{j=1}^N K_{i,j} = 1 \quad (3.71)$$

Une matrice qui vérifie (3.71) s'appelle une matrice stochastique. L'ensemble des matrices stochastiques est un convexe fermé dont les points extrémaux sont les matrices  $K_{i,j}$  telles que pour tout  $i$ ,  $K_{i,j}$  n'est non nul que pour un seul indice  $j = F(i)$  et  $K_{i,F(i)} = 1$ , i.e la matrice associée à l'application  $F$  de  $\Omega$  dans  $\Omega$ .

$K$  opère sur les fonctions par

$$K(f)(x_i) = \sum_j K_{i,j} f(x_j) \quad (3.72)$$

et sur les probabilités par

$${}^t K(\pi)(x_j) = \sum_i \pi(x_i) K_{i,j} \quad (3.73)$$

On a

$$(f|g)_\pi = \sum f(x_i) \bar{g}(x_i) \pi(x_i) \quad (3.74)$$

L'opérateur  $K$  est  $\pi$ -réversible (i.e autoadjoint sur  $L^2(\Omega, d\pi)$ ) ssi on a

$$\forall i, j \quad \pi(x_i) K_{i,j} = \pi(x_j) K_{j,i} \quad (3.75)$$

On identifiera les fonctions  $f$  à des vecteurs colonnes  ${}^t(f_1, \dots, f_N)$  avec  $f_i = f(x_i)$  et les probabilités  $\nu$  à des vecteurs lignes  $(\nu_1, \dots, \nu_N)$  avec  $\nu_i = \nu(x_i)$  et  $\sum \nu_i = 1$ .

**Lemme 3.19** *Soit  $K$  une matrice stochastique.*

(i) 1 est valeur propre de  $K$  de vecteur propre associé  ${}^t(1, \dots, 1) = \mathbf{1}$ ,  $\text{Spec}(K) \subset \{z \in \mathbb{C}, |z| \leq 1\}$ , et si  $\lambda \in \text{Spec}(K)$  alors  $\bar{\lambda} \in \text{Spec}(K)$ .

(ii) L'équation  $(K - Id)a = \mathbf{1}$  n'a pas de solution.

*Démonstration.* On a évidemment  $K(\mathbf{1}) = \mathbf{1}$ . Soit  $\lambda \in \text{Spec}(K)$  et  $a \neq 0$  tel que  $\sum_j K_{i,j} a_j = \lambda a_i$ . Soit  $i_0$  tel que  $|a_{i_0}| = \sup_j |a_j|$ . On a  $|a_{i_0}| > 0$  et

$$|\lambda| |a_{i_0}| \leq \sum_j K_{i_0,j} |a_j| \leq |a_{i_0}| \sum_j K_{i_0,j} = |a_{i_0}|$$

ce qui prouve  $|\lambda| \leq 1$ . L'invariance du spectre par conjugaison complexe résulte du fait que les coefficients  $K_{i,j}$  sont réels. Enfin, si  $(K - Id)a = \mathbf{1}$ , on a pour tout  $n \geq 1$   $K^n(a) - a = n\mathbf{1}$ . Or la suite de vecteurs  $K^n(a) - a$  est bornée puisque  $\sup_j |K^n(a)_j| \leq \sup_j |a_j|$  : contradiction.  $\square$

**Exemple 3.20** Soit  $K$  la matrice stochastique (associée à la bijection  $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ )

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad (3.76)$$

On a  $K^3 = Id$  et  $Spec(K) = \{1, e^{2i\pi/3}, e^{4i\pi/3}\}$

**Exemple 3.21** Soit  $K$  la matrice stochastique avec  $a > 0, c > 0, a + c < 1$

$$\begin{pmatrix} a & 0 & 1 - a \\ c & a & 1 - a - c \\ 0 & 0 & 1 \end{pmatrix} \quad (3.77)$$

On a  $\det(K - \lambda Id) = (1 - \lambda)(a - \lambda)^2$ ,  $Spec(K) = \{1, a, a\}$  et  $Ker(K - aId) = \{^t(x, y, z), x = z = 0\}$  qui est de dimension 1 de sorte que  $K$  n'est pas diagonalisable.

On remarquera qu'une matrice stochastique admet toujours une probabilité invariante. En effet, comme 1 est valeur propre de  $K$ , 1 valeur propre de  ${}^tK$ . Il existe donc une mesure  $\mu$  telle que  ${}^tK(\mu) = \mu$ ; si  $\mu = \mu_+ - \mu_-$  est la décomposition de  $\mu$  en parties positives et négatives, on a  $\mu_+ - \mu_- = {}^tK(\mu_+) - {}^tK(\mu_-)$ , donc puisque  ${}^tK(\mu_{\pm})$  sont des mesures positives,  ${}^tK(\mu_{\pm}) = \mu_{\pm} + \theta$ , où  $\theta$  est une mesure positive, et comme  ${}^tK$  conserve la masse totale, on obtient  $\theta = 0$ . Les mesures positives  $\mu_{\pm}$  sont donc invariantes. Si  $\Omega$  est infini,  $K$  peut n'avoir aucune probabilité invariante; par exemple, avec  $\Omega = \mathbb{Z}$  l'opérateur de Markov  $K(g)(n+1) = g(n)$  n'admet pas de probabilité invariante. On a bien  $Spec(K) = Spec({}^tK)$ , mais le fait que  $1 \in Spec({}^tK)$  n'implique pas que 1 est valeur propre de  ${}^tK$  lorsque  $K$  opère sur un espace de dimension infinie.

A toute matrice stochastique  $K$ , on associe un graphe orienté  $G = G(K)$ . Les sommets de  $G$  sont les éléments  $\{1, \dots, N\}$  de  $\Omega$  et on trace une arête orientée de  $i$  vers  $j$  ssi  $K_{i,j} > 0$ .

**Définition 3.22** Soit  $K = K_{i,j}$  une matrice stochastique.

- (i)  $K$  est irréductible ssi  $\forall i, j, \exists n \geq 1 \quad K_{i,j}^n > 0$
- (ii)  $K$  est apériodique ssi  $\forall i, \text{pgcd}\{n \geq 1, K_{i,i}^n > 0\} = 1$

Le fait que  $K$  soit irréductible a une interprétation simple en terme du graphe  $G(K)$  : Pour tout couple  $(i, j)$ , on peut passer du sommet  $i$  au sommet  $j$  du graphe en suivant des arêtes (orientées) du graphe.

**Proposition 3.23 (Perron-Frobenius)** Soit  $K$  une matrice stochastique.

- (i) Si  $K$  est irréductible,  $K$  admet une unique probabilité invariante et le sous espace caractéristique de  $K$  associé à la valeur propre 1 est de dimension 1, engendré par  $\mathbf{1}$ .
- (ii) Si de plus  $K$  est apériodique, on a  $Spec(K) \setminus \{1\} \subset \{z \in \mathbb{C}, |z| < 1\}$ .

*Démonstration.* Pour (i), il s'agit juste de vérifier que 1 est valeur propre simple de  $K$ . Soit  $a$  tel que  $K(a) = a$ . On a  $K^n(a) = a$  pour tout  $n \geq 1$ , donc  $\sum_j K_{i_0,j}^n a_j = a_{i_0}$ . Soit  $i_0$  tel que  $|a_{i_0}| = \sup |a_j|$ . On peut supposer  $a_{i_0} = 1$ . On a en utilisant  $K^n(\mathbf{1}) = \mathbf{1}$ ,

$$1 = a_{i_0} = \operatorname{Re}\left(\sum_j K_{i_0,j}^n a_j\right) = \sum_j K_{i_0,j}^n (\operatorname{Re}(a_j) - 1) + 1 \quad (3.78)$$

Puisque  $K_{i_0,j}^n \geq 0$  et  $\operatorname{Re}(a_j) \leq 1$ , on obtient pour tout  $n \geq 1, j$ ,  $K_{i_0,j}^n (1 - \operatorname{Re}(a_j)) = 0$ . En fixant  $j$  et en choisissant  $n$  tel que  $K_{i_0,j}^n > 0$  on obtient  $\operatorname{Re}(a_j) = 1$ , et comme  $|a_j| \leq 1$ , cela prouve  $a_j = 1$ , donc  $a = \mathbf{1}$ .

Soit pour  $p \geq 1$ ,  $E_p = \operatorname{Ker}(K - Id)^p$ . Pour  $p \geq 2$  et  $b \in E_p$ , on a  $(K - Id)^{p-1}(b) \in E_1$ , donc d'après ce qui précède  $(K - Id)^{p-1}(b) = \mu \mathbf{1}$ , donc d'après le point (ii) du lemme (3.19),  $\mu = 0$ , donc  $b \in E_{p-1}$ . On a donc  $\cup_{p \geq 1} E_p = E_1$ .

Pour (ii), supposons de plus que  $K$  est apériodique. Soit  $\lambda \in \operatorname{Spec}(K)$  avec  $|\lambda| = 1$  et  $\lambda \neq 1$ . Soit  $\theta \in ]0, 2\pi[$  tel que  $\lambda = e^{i\theta}$  et  $a \neq 0$  tel que  $Ka = \lambda a$ . Soit à nouveau  $i_0$  tel que  $|a_{i_0}| = \sup |a_j|$ . On peut toujours supposer  $a_{i_0} = 1$ . On a  $K^n(a) = e^{in\theta} a$  pour  $n \geq 1$ , d'où

$$1 = |e^{in\theta} a_{i_0}| \leq \sum_j K_{i_0,j}^n |a_j| = \sum_j K_{i_0,j}^n (|a_j| - 1) + 1 \quad (3.79)$$

et comme  $K$  est irréductible, on obtient comme précédemment  $|a_j| = 1$  pour tout  $j$ . Posons  $a_j = e^{i\phi_j}$ . On a pour tout  $n \geq 1$

$$\sum_{j \neq i_0} K_{i_0,j}^n e^{i\phi_j} = e^{in\theta} - K_{i_0,i_0}^n \quad (3.80)$$

Soit  $A = \{n \geq 1, K_{i_0,i_0}^n > 0\}$ . Comme  $K$  est irréductible,  $A$  est non vide, et de  $K_{i_0,i_0}^{n+m} = \sum_j K_{i_0,j}^n K_{j,i_0}^m \geq K_{i_0,i_0}^n K_{i_0,i_0}^m$ , on déduit que  $n + m \in A$  pour  $n, m \in A$ .

Pour  $n \in A$ , on a  $\sum_{j \neq i_0} K_{i_0,j}^n < 1$ , donc d'après (3.80)  $|e^{in\theta} - K_{i_0,i_0}^n| < 1$ . Il en résulte

$$n \in A \implies \operatorname{Re}(e^{in\theta}) > 0 \quad (3.81)$$

Soit  $l = \inf\{n, n \in A\}$ . Pour tout  $k \geq 1$  on  $kl \in A$  et donc d'après (3.81),  $kl\theta \in ]-\pi/2, \pi/2[ \pmod{2\pi}$ . Il en résulte

$$\theta = 2\pi j/l \quad \text{avec } j \in \{1, \dots, l-1\} \quad (3.82)$$

Soit  $l = p_1^{b_1} \dots p_k^{b_k}$  la décomposition en facteurs premiers de  $l$ . Comme  $K$  est apériodique, pour tout  $j \in \{1, \dots, k\}$ , il existe  $n_j \in A$  non divisible par  $p_j$ . Alors  $\sum_j (\prod_{e \neq j} p_e) n_j = m \in A$  n'est divisible par aucun des  $p_j$ , donc est premier à  $l$ . Or en appliquant (3.81) aux  $km \in A$ , on obtient comme précédemment  $m\theta = 2\pi j'$  avec  $j' \in \{1, \dots, m-1\}$ , d'où  $2\pi/\theta = m/j' = l/j$ , donc  $mj = lj'$  donc  $m$  qui est premier à  $l$  divise  $j'$  : contradiction.  $\square$

**Remarque 3.24** La matrice stochastique (3.76) de l'exemple (3.20) est irréductible, mais pas apériodique : on a  $\operatorname{pgcd}\{n \geq 1, K_{i,i}^n > 0\} = 3$  pour tout  $i$ .

Le théorème suivant est une conséquence immédiate de la proposition (3.23) et de la réduction de Jordan des matrices.

**Théorème 3.25** Soit  $\Omega = \{1, \dots, N\}$  et  $K$  une matrice stochastique sur  $\Omega$ , irréductible et apériodique. Alors  $K$  admet une unique probabilité  $\pi$  invariante. Le spectre de  $K$  avec multiplicité est  $\text{Spec}(K) = \{1, \lambda_2, \dots, \lambda_N\}$  avec  $1 > |\lambda_2| \geq \dots \geq |\lambda_N|$  et il existe des constantes  $C, A$  telles que pour tout  $n$  on ait

$$\sup_x \|K^n(x, dy) - \pi\|_{VT} \leq C|\lambda_2|^n n^A \quad (3.83)$$

En particulier, si la matrice stochastique  $K$  est irréductible et apériodique,  $K$  est uniformément ergodique. La constante  $A$  qui intervient dans (3.83) se calcule à partir de la structure des blocs de Jordan de  $K$  associés aux valeurs propres de module  $|\lambda_2|$ .

### 3.4 Marche aléatoire

Rappelons qu'on note  $X = \Omega^{\mathbb{N}^*}$  l'espace produit dont les éléments sont les suites infinies  $\underline{x} = (x_1, \dots, x_j, \dots)$  avec  $x_j \in \Omega$ , et  $p_j : X \rightarrow \Omega$  l'application coordonnée  $p_j(\underline{x}) = x_j$ . On munit  $X$  de la tribu produit  $\mathbb{B}$  (la tribu engendrée par les  $p_j^{-1}(\mathcal{B})$ ).

Soit  $K$  un opérateur de Markov sur  $\Omega$  de noyau  $K(x, dy)$ , et  $x_0$  un point de  $\Omega$ . Pour tout cylindre  $C$  de  $X$  de la forme

$$C = A_1 \times \dots \times A_N \times \Omega \times \Omega \times \dots, \quad A_j \in \mathcal{B} \quad (3.84)$$

on pose

$$P_{x_0}(C) = K(\mathbf{1}_{A_1} K(\mathbf{1}_{A_2} K(\dots K(\mathbf{1}_{A_N}))))(x_0) \quad (3.85)$$

qu'on peut écrire aussi sous la forme

$$P_{x_0}(C) = \int \dots \int_{A_1 \times A_2 \times \dots \times A_n} K(x_0, dx_1) K(x_1, dx_2) \dots K(x_{N-1}, dx_N) \quad (3.86)$$

On vérifie aisément que  $C \rightarrow P_{x_0}(C)$  est une mesure positive additive sur l'algèbre des cylindres, et bien sur que  $P_{x_0}(X) = 1$ .

Pour tout élément  $(x_1, \dots, x_j, \dots)$  de  $X$ , on dit que  $(x_0, x_1, \dots, x_j, \dots)$  est une marche aléatoire issue de  $x_0$ , et on interprète  $P_{x_0}(C)$  comme la probabilité pour qu'on ait  $x_1 \in A_1, \dots, x_N \in A_N$ , qu'on note

$$P_{x_0}(x_1 \in A_1, \dots, x_N \in A_N) = P_{x_0}(C) \quad (3.87)$$

Par exemple, on a d'après (3.85)

$$P_{x_0}(x_n \in A) = \int_A K^n(x_0, dy) \quad (3.88)$$

qui est la probabilité pour que le nième terme de la marche appartienne à  $A$ . Il existe alors une unique probabilité  $P_{x_0}$  sur  $(X, \mathbb{B})$  qui vérifie (3.85) pour tout cylindre  $C$ . **référence.** On remarquera qu'une marche aléatoire n'est pas une marche, mais une probabilité sur l'espace des trajectoires. Les probabilistes appellent chaîne de Markov la suite  $x_1, \dots, x_n, \dots$  de variables aléatoires sur  $X$ .

Lorsque  $K(f) = (\int f d\pi)\mathbf{1}$  pour une probabilité  $\pi$  sur  $\Omega$ , on retrouve simplement  $P_{x_0} = \pi^\infty$ , la probabilité produit.



**Théorème 3.26 (Markov Chain Monte Carlo)** Soit  $K$  un opérateur de Markov sur  $\Omega$ ,  $x_0$  un point de  $\Omega$  et  $P_{x_0}$  la probabilité associée sur  $X$ . On suppose que  $K$  admet une unique probabilité invariante  $\pi$  et qu'on a  $\lim_{n \rightarrow \infty} \|(K - \Pi_\pi)^n\|_{L^\infty} = 0$ , i.e on suppose que  $K$  est uniformément ergodique.

Soit  $f \in L^1(\Omega, d\pi)$ . Pour  $P_{x_0}$  presque toute suite  $\underline{x} \in X$  on a

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^{j=N} f(x_j) = \int_{\Omega} f d\pi \quad (3.89)$$

On remarquera que les variables aléatoires  $f_j(\underline{x}) = f(x_j)$  ne sont pas en général indépendantes sur  $X$ , mais que les corrélations entre  $f_j$  et  $f_{j+n}$  tendent vers zéro quand  $n \rightarrow \infty$ . Ce théorème est valable sous des conditions bien plus faibles que l'uniforme ergodicité et il existe aussi dans ce cadre des versions du théorème de la limite centrale. Pour une preuve de ce théorème, voir [Num84] et aussi [Tie94].

Un exemple typique de marche aléatoire associée à un noyau de Markov est le suivant.

Soit  $\Omega$  un espace métrique,  $d(x, y)$  la distance sur  $\Omega$ ,  $\mathcal{B}$  la tribu engendrée par les ouverts de  $\Omega$ , et  $\mu$  une mesure positive sur  $(\Omega, \mathcal{B})$ . On suppose que pour tout  $x \in \Omega$  et tout  $h > 0$ , on a

$$0 < \mu(B(x, h)) < \infty \quad (3.90)$$

où  $B(x, h) = \{y \in \Omega, d(y, x) < h\}$  est la boule ouverte de centre  $x$  et de rayon  $h$ . On fixe  $h > 0$  et considère le noyau de Markov

$$K_h(x, dy) = \frac{\mathbf{1}_{B(x, h)}(y)}{\mu(B(x, h))} d\mu(y) \quad (3.91)$$

associé à l'opérateur

$$K_h(f)(x) = \frac{1}{\mu(B(x, h))} \int_{d(y, x) < h} f(y) d\mu(y) \quad (3.92)$$

La marche aléatoire correspondante est simplissime : si la marche est au point  $x_n$ , on choisit un point  $y \in B(x, h)$  uniforme pour  $\mu$ , et on pose  $x_{n+1} = y$ .

Lorsque  $\mu$  est une probabilité sur  $\Omega$ ,  $K_h$  est  $\pi$ -réversible pour la probabilité (qui dépend de  $h$ )

$$\begin{aligned} d\pi(x) &= \frac{\mu(B(x, h))}{Z_h} d\mu(x) \\ Z_h &= \int \int_{d(x, y) < h} d\mu(x) d\mu(y) > 0 \end{aligned} \quad (3.93)$$

La forme de Dirichlet et le laplacien associés sont

$$\begin{aligned} \mathcal{E}_h(f, f) &= \frac{1}{2Z_h} \int \int_{d(x, y) < h} |f(x) - f(y)|^2 d\mu(x) d\mu(y) \\ |\Delta_h|(f)(x) &= \frac{1}{\mu(B(x, h))} \int_{d(y, x) < h} (f(x) - f(y)) d\mu(y) \end{aligned} \quad (3.94)$$

et la formule d'intégration par parties s'écrit

$$(|\Delta_h|f|g)_\pi = \mathcal{E}_h(f, g) \quad (3.95)$$

Comprendre la théorie spectrale de  $K_h$ , la structure de l'algèbre des opérateurs engendrée par  $K_h$  et les multiplicateurs par des fonctions, le comportement de cette marche quand  $h \rightarrow 0$ , ..., autant de questions qui peuvent être simples ou difficiles en fonction de la nature de l'espace métrique mesuré  $\Omega$ .

Voici un exemple simple associé à la discrétisation par différences finies de l'opérateur  $f \rightarrow f''$ . On étudiera dans la section 5 un exemple analogue dans le cadre plus général du laplacien sur une variété riemannienne.

Soit  $U$  le disque unité complexe  $U = \{z \in \mathbb{C}, |z| = 1\}$  qu'on identifie à  $\mathbb{R}/\mathbb{Z}$  par  $\theta \in \mathbb{R}/\mathbb{Z} \mapsto z = e^{2i\pi\theta} \in U$ . Pour  $N \geq 2$  soit  $\Omega_N = \Omega = \{z \in \mathbb{C}, z^N = 1\} = \{x_1, \dots, x_N\}$  avec  $x_j = e^{2i\pi j/N}$ . On pose  $h = 1/N$ . On note, pour  $k \in \mathbb{Z}$ ,  $e_k$  la fonction sur  $U$ ,  $e_k(z) = z^k$ . Les restrictions des  $e_k$  à  $\Omega$  pour  $0 \leq k \leq N-1$  forment une base des fonctions sur  $\Omega$ , et  $e_k|_\Omega = e_{k+N}|_\Omega$  pour tout  $k \in \mathbb{Z}$ . Soit  $\pi$  la probabilité uniforme sur  $\Omega_N$ , i.e  $\pi(x_j) = 1/N$  pour tout  $j$ , et  $K_h$  le noyau de Markov défini par

$$K_h(f)(x_j) = \frac{1}{3}(f(x_{j-1}) + f(x_j) + f(x_{j+1})) \quad (3.96)$$

Alors  $K_h$  est  $\pi$ -réversible, et on a

$$K_h(e_k) = \left(\frac{1 + 2 \cos(2\pi k/N)}{3}\right)e_k \quad (3.97)$$

de sorte que

$$\text{Spec}(K_h) = \left\{ \frac{1 + 2 \cos(2\pi hk)}{3}, k = 0, \dots, N-1 \right\}.$$

En particulier, on a

$$\text{Spec}(K) \subset [-1/3, 1] \quad \text{et} \quad \text{gap} = \frac{2}{3}(1 - \cos(2h\pi)) \quad (3.98)$$

On a pour toute fonction régulière  $f(\theta)$ ,  $\theta \in \mathbb{R}/\mathbb{Z}$ , en notant aussi  $f$  par abus sa restriction à  $\Omega_N$

$$|\Delta_h|(f) = (Id - K)(f) = -\frac{1}{3}h^2 f'' + O(h^3) \quad (3.99)$$

Pour tout  $t \in [0, \infty[$  on a, en désignant par  $\text{Ent}(x) \in ]x-1, x]$  la partie entière de  $x$

$$\lim_{h \rightarrow 0} K_h^{\text{Ent}(t/h^2)}(e_k) = \lim_{h \rightarrow 0} \left(\frac{1 + 2 \cos(2\pi hk)}{3}\right)^{\text{Ent}(t/h^2)} e_k = e^{-\frac{t}{3}4\pi^2 k^2} e_k \quad (3.100)$$

Lorsque  $h \rightarrow 0$  l'opérateur  $K_h^{\text{Ent}(t/h^2)}$  converge donc vers le noyau de la chaleur  $e^{\frac{t}{3}\Delta}$  sur  $U \simeq \mathbb{R}/\mathbb{Z}$ . On prouve aussi facilement que la marche aléatoire associée converge vers le mouvement brownien sur le cercle.

## 4 Algorithme de Metropolis

### 4.1 Description de l'algorithme

Soit  $\Omega$  un espace métrique localement compact et dénombrable à l'infini, et  $\mathcal{B}$  la tribu des boreliens sur  $\Omega$ . Soit  $\mu$  une mesure de Borel positive sur  $(\Omega, \mathcal{B})$ . Soit enfin  $\rho(x) \in [0, \infty[$  une densité sur  $\Omega$  telle que  $d\pi = \rho(x)d\mu$  soit une probabilité, i.e

$$\int_{\Omega} \rho(x)d\mu = 1 \quad (4.1)$$

L'algorithme de Metropolis construit une chaîne de Markov qui admet  $\pi$  comme mesure d'équilibre. Il s'agit donc d'une méthode MCMC dont le but est de construire un algorithme qui permette de choisir un point au hasard selon  $\pi$ .

Pour ce faire, on se donne d'abord une fonction  $k(x, y) \geq 0$  mesurable sur  $\Omega \times \Omega$  telle que  $k(x, y)d\mu(y)$  soit un noyau de Markov sur  $\Omega$ . Dans la pratique, on devra savoir choisir pour tout  $x$  un point  $y$  selon la probabilité  $k(x, y)d\mu(y)$ . On verra plus loin des variantes de type  $N$ -corps pour lesquelles on travaille avec un noyau de Markov  $k(x, dy)$  qui n'admet pas de densité par rapport à  $\mu$ .

On définit ensuite le taux d'acceptation  $A(x, y) \geq 0$  comme suit :

$$\begin{aligned} A(x, y) &= 1 \quad \text{si } \rho(x)k(x, y) = 0 \\ A(x, y) &= \frac{\rho(y)k(y, x)}{\rho(x)k(x, y)} \quad \text{sinon} \end{aligned} \quad (4.2)$$

On définit alors l'opérateur de Metropolis  $M$  par

$$\begin{aligned} M(f)(x) &= m(x)f(x) + \int_{\Omega} k(x, y) \min(1, A(x, y))f(y)d\mu(y) \\ m(x) &= 1 - \int_{\Omega} k(x, y) \min(1, A(x, y))d\mu(y) \in [0, 1] \end{aligned} \quad (4.3)$$

Par définition, on a pour  $\rho(x) > 0$ ,  $k(x, y) \min(1, A(x, y)) = \min(k(x, y), \frac{\rho(y)k(y, x)}{\rho(x)})$  et lorsque  $\rho(x) = 0$ ,  $k(x, y) \min(1, A(x, y)) = k(x, y)$ .

On a par construction  $M(\mathbf{1}) = \mathbf{1}$ ,  $M(f) \geq 0$  si  $f \geq 0$  et le noyau associé à  $M$  est

$$M(x, dy) = m(x)\delta_{y=x} + k(x, y) \min(1, A(x, y))d\mu(y) \quad (4.4)$$

Pour tout borelien  $B$ ,  $x \rightarrow m(x, B)$  est mesurable, et on a

$$M(x, B) = m(x)\mathbf{1}_B(x) + \int_B k(x, y) \min(1, A(x, y))d\mu(y) \quad (4.5)$$

le fait que l'opérateur markovien  $M$  soit de type Feller, c'est à dire vérifie  $M(f) \in C_b^0(\Omega)$  pour  $f \in C_b^0(\Omega)$ , dépend évidemment des données  $\mu, k(x, y)$ .

L'interprétation algorithmique en terme de marche aléatoire de l'opérateur markovien  $M$  est simple : supposons que la marche soit au point  $x_n$ .

- 1. On choisit  $y$  selon  $k(x_n, y)d\mu$ .
- 2. Si  $A(x_n, y) \geq 1$ , on pose  $x_{n+1} = y$ .
- 3. Si  $A(x_n, y) < 1$ , on tire à pile ou face avec probabilité de gagner égale à  $A(x_n, y)$ .
  - Si on gagne, on pose  $x_{n+1} = y$ .
  - Si on perd, on pose  $x_{n+1} = x_n$ .

Il est donc facile d'implémenter sur machine cet algorithme, dès qu'on dispose d'une méthode qui permette de réaliser le point 1 : choisir  $y$  selon  $k(x_n, y)d\mu$ . Le point clé de cet algorithme est que même si le noyau de Markov  $k(x_n, y)d\mu$  utilisé dans l'étape 1. n'a rien à voir avec la probabilité  $\pi$  qu'on souhaite échantillonner, l'introduction du taux d'acceptation donné par (4.2) va permettre d'assurer que  $\pi$  est une probabilité d'équilibre pour le noyau de Markov  $M(x, dy)$ . Un point fondamental de cet algorithme est qu'à chaque étape, on a, en général, une probabilité non nulle de ne pas se déplacer (des variantes de l'algorithme de Metropolis, de type dynamique moléculaire, sont telles qu'à chaque étape de l'algorithme, on peut suivre un champ de vecteur déterministe  $X$  pendant un certain temps). Cela introduit des difficultés notables, puisqu'on perd le phénomène de régularisation locale des diffusions elliptiques ou hypo-elliptiques. D'une certaine manière, cet algorithme interpole entre mouvement déterminisme (ici, rester sur place), et mouvement stochastique (se déplacer aléatoirement). Cela se voit très bien sur la formule (4.3) qui écrit l'opérateur de Metropolis comme la somme de deux termes : l'opérateur de multiplication par la fonction  $m$  (composante déterministe), et l'opérateur de moyenne pondérée sur  $\Omega$  (composante stochastique). C'est le point remarquable de cet algorithme.

**Lemme 4.1**  *$M$  est  $\pi$ -réversible, et donc  $\pi$  est invariante par  $M$ .*

*Démonstration.* On a

$$(M(f)|g)_\pi = \int f(x)\bar{g}(x)m(x)\rho(x)d\mu(x) + \int \int k(x, y) \min(1, A(x, y))f(y)\bar{g}(x)\rho(x)d\mu(y)d\mu(x) \quad (4.6)$$

et par construction  $k(x, y) \min(1, A(x, y))\rho(x) = \min(\rho(x)k(x, y), \rho(y)k(y, x))$  qui est réel et symétrique en  $x, y$ , donc  $(M(f)|g)_\pi = (f|M(g))_\pi$ .  $\square$

**Lemme 4.2** *La forme de Dirichlet associée à  $M$  est*

$$\begin{aligned} \mathcal{E}(f, f) &= ((Id - M)f|f)_\pi \\ &= \frac{1}{2} \int \int |f(x) - f(y)|^2 \min(\rho(x)k(x, y), \rho(y)k(y, x))d\mu(y)d\mu(x) \end{aligned} \quad (4.7)$$

*Démonstration.* Immédiat par ce qui précède.  $\square$

Le choix du noyau de Markov  $k(x, dy)$  va évidemment influencer les performances de l'algorithme. On parle d'algorithme local lorsque  $k(x, dy)$  n'autorise que des déplacements en des points  $y$  voisins de  $x$ , et d'algorithme global sinon. La première chose à faire pour

étudier cet algorithme est de savoir s'il vérifie une inégalité de Poincaré, et d'estimer le gap. L'inégalité de Poincaré s'écrit

$$\begin{aligned} & \int \int |f(x) - f(y)|^2 \rho(y) \rho(x) d\mu(y) d\mu(x) \\ & \leq A \int \int |f(x) - f(y)|^2 \min(\rho(x)k(x, y), \rho(y)k(y, x)) d\mu(y) d\mu(x) \end{aligned} \quad (4.8)$$

En particulier, si on sait minorer  $k(x, y)$  par une fonction  $p(x, y)$  qui est telle que l'inégalité (4.8) est vraie avec  $p$  en place de  $k$ , alors (4.8) est trivialement vrai.

Plus généralement, pour démontrer une telle inégalité, on peut d'abord utiliser le lemme 3.10, i.e montrer que  $M^n$  vérifie une inégalité de Poincaré pour un certain  $n \geq 1$ . En effet, il y a de bonnes chances que le noyau de  $M^n$  admette une minoration plus intéressante que le noyau de  $M$ . En minorant le noyau de Markov  $M^n(x, dy) \geq p(x, y) d\mu(y)$  avec une fonction  $p(x, y) \geq 0$  telle que  $\int p(x, y) d\mu(y) \leq 1$ , on se ramène en utilisant (3.10), à prouver une inégalité de la forme

$$\int \int |f(x) - f(y)|^2 \rho(y) \rho(x) d\mu(y) d\mu(x) \leq C \int \int |f(x) - f(y)|^2 p(x, y) \rho(x) d\mu(y) d\mu(x) \quad (4.9)$$

Si la fonction  $p(x, y)$  est nulle (ou trop petite) lorsque les points  $x$  et  $y$  ne sont pas proches, on peut enfin essayer de prouver cette dernière inégalité par un argument de chemins : pour tout couple  $(x, y)$  on choisit des points intermédiaires  $z_0(x, y) = x, z_1(x, y), \dots, z_N(x, y) = y$ , avec  $z_j$  proche de  $z_{j+1}$ , et on utilise

$$|f(x) - f(y)|^2 \leq \sqrt{N} \sum_j |f(z_j) - f(z_{j+1})|^2 \quad (4.10)$$

ce qui ramène à vérifier

$$\int \int |f(z_j(x, y)) - f(z_{j+1}(x, y))|^2 d\pi(y) d\pi(x) \leq C \int \int |f(x) - f(y)|^2 p(x, y) \rho(x) d\mu(y) d\mu(x) \quad (4.11)$$

Si  $(x, y) \mapsto (z_j(x, y), z_{j+1}(x, y))$  est une application raisonnable, on pourra alors effectuer le changement de variables  $(x, y) \mapsto (z_j(x, y), z_{j+1}(x, y))$  dans le membre de gauche, et donc relocaliser le problème sur le support de  $p$ . Evidemment, rien n'assure que ce type de méthode donne autre chose qu'une borne inférieure sur le gap.

## 4.2 Algorithme de Metropolis de type $N$ -corps

Dans les applications, on rencontre souvent la situation suivante, avec  $N$  entier :

- $Y$  est un espace métrique localement compacte et dénombrable à l'infini.
- $Z$  est un fermé de  $Y^N$ .
- $\Omega$  est un ouvert de  $Z$ .

Pour  $j \in \{1, \dots, N\}$  et  $y = (y_1, \dots, y_N) \in Y^N$ , on note  $y^j = (y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_N) \in Y^{N-1}$ , et aussi  $y = (y^j, y_j)$ . Pour  $x \in \Omega$ , on pose

$$\Omega_{x,j} = \{y \in Y, (x^j, y) \in \Omega\} \subset Y \quad (4.12)$$

Evidemment, si  $x = (x^j, x_j)$ , on a  $x_j \in \Omega_{x,j}$ . On suppose que pour tout  $j \in \{1, \dots, N\}$ , il existe une probabilité  $d\nu^j$  sur  $Y^{N-1}$ , et pour tout  $x \in \Omega$  une probabilité  $d\pi_{x^j}$  sur  $\Omega_{x,j}$ , dépendante de  $x^j$ , telles qu'on ait

$$\int_{\Omega} f(x) d\pi(x) = \int_{Y^{N-1}} \left( \int_{\Omega_{x,j}} f(x^j, z) d\pi_{x^j}(z) \right) d\nu^j(x^j) \quad (4.13)$$

En d'autres termes,  $d\pi_{x^j}$  est une probabilité conditionnelle, toutes les coordonnées de  $x \in \Omega$  étant fixées, sauf la  $j^{\text{ième}}$ . L'opérateur de Metropolis de type  $N$ -corps est un opérateur de la forme

$$M(f)(x) = \frac{1}{N} \sum_{j=1}^N M_{x^j}(f)(x^j, x_j) \quad (4.14)$$

où  $M_{x^j}$  est un opérateur de Metropolis sur  $\Omega_{x,j}$ , qui échantillonne la probabilité  $d\pi_{x^j}$ .

La marche associée est donc la suivante :

Si la marche est au point  $x \in \Omega$ , on choisit un  $j \in \{1, \dots, N\}$  au hasard pour la probabilité uniforme. On a alors  $x = (x^j, x_j)$ , et on ne modifie que la coordonnée  $x_j$  en utilisant un algorithme de Metropolis sur la tranche  $\Omega_{x,j}$  de  $\Omega$ .

### 4.3 Exemples

#### A. L'exemple historique.

Soit  $Q = ]0, 1[^2$  le carré unité ouvert dans le plan. On souhaite placer au hasard  $N$  disques fermés de rayon  $\varepsilon > 0$  dans  $Q$ , de sorte que 2 disques quelconques ne s'intersectent pas. L'espace de configuration est l'ouvert  $\mathcal{C}$  de  $\mathbb{R}^{2N}$

$$\mathcal{C} = \mathcal{C}_{N,\varepsilon} = \{ \underline{x} = (x_1, \dots, x_N), x_j \in ]\varepsilon, 1 - \varepsilon[^2, \quad \forall i \neq j \quad \text{dist}(x_i, x_j) > 2\varepsilon \} \quad (4.15)$$

Les  $x_j$  sont les centres des disques, et on munit  $\mathcal{C}$  de la mesure de Lebesgue. Dans cet exemple, il faut garder à l'esprit que  $\varepsilon$  peut être très petit et  $N$  très grand. Choisir un point au hasard dans  $\mathcal{C}$  est alors hautement non trivial. On peut bien sûr travailler aussi en dimension 3 avec  $Q = ]0, 1[^3$ , et avec des sphères à la place des disques, mais si on pense ainsi modéliser une répartition de molécules dans une boîte, le nombre d'Avogadro, i.e le nombre d'atomes dans 12 grammes de carbone-12 étant  $N_A = 6.022141 \times 10^{23} \text{ mol}^{-1}$ , la dimension  $d = 3N$  de  $\mathcal{C}$  est gigantesque. Même pour des valeurs beaucoup plus petites de  $N$ , la complexité numérique de ce problème reste hors de portée des méthodes déterministes.

La structure topologique et métrique de l'espace de configuration  $\mathcal{C}$  est très mal connue. C'est un ouvert borné semi-algébrique de  $\mathbb{R}^{2N}$ , évidemment non vide dès que la densité de disques  $N\pi\varepsilon^2 = \delta$  n'est pas trop grande, dont l'adhérence est contenue dans

$$\mathcal{C}^f = \mathcal{C}_{N,\varepsilon}^f = \{ \underline{x} = (x_1, \dots, x_N), x_j \in [\varepsilon, 1 - \varepsilon]^2, \quad \forall i \neq j \quad \text{dist}(x_i, x_j) \geq 2\varepsilon \} \quad (4.16)$$

On peut avoir  $\bar{\mathcal{C}} \neq \mathcal{C}^f$ , et  $\mathcal{C}^f$  peut avoir des composantes connexes d'intérieur vide, dès que  $2N\varepsilon = \mu$  est grand avec  $\delta$  arbitrairement petit (voir [Kah09]). Comme on a  $\mathcal{C}_{N,\varepsilon}^f \subset \mathcal{C}_{N,\varepsilon'}$  pour  $\varepsilon' < \varepsilon$ , il en résulte qu'il est faux (!) que  $\mathcal{C}$  soit un ouvert connexe si la densité  $\delta$

est assez petite. Dans [DLM08] il est prouvé que le bord de  $\mathcal{C}$  est lipschitz lorsque  $\mu$  est assez petit, mais il y a des singularités de type cusp dans le bord de  $\mathcal{C}$  par exemple pour des valeurs de  $\mu$  proches de 1.

L'intérêt de cet exemple est qu'il est censé être un modèle qui permettrait de comprendre la nature des transitions de phase liquide-solide (voir [Rad08]). Beaucoup de tests numériques ont été effectués sur ce modèle (voir [Kra09]). Ils semblent confirmer une transition de phase pour des valeurs de la densité  $\delta$  des disques de l'ordre de 0.7, la densité maximale de remplissage étant  $\delta = \frac{\pi}{2\sqrt{3}} \simeq 0.9$ , lorsque les disques de rayon  $\varepsilon$  petit sont centrés sur le réseau hexagonal. Mais en fait, malgré 50 ans d'effort, on ne sait pratiquement pas grand chose...

L'algorithme de Metropolis et al. [MRR+53] est le suivant :  
On part d'une configuration admissible de disques  $\underline{x} = (x_1, \dots, x_N) \in \mathcal{C}$ . On choisit un  $j$  au hasard pour la mesure uniforme sur  $\{1, \dots, N\}$  et on choisit un  $y \in ]\varepsilon, 1 - \varepsilon[$  au hasard pour la mesure de Lebesgue (on sait faire). On remplace  $x_j$  par  $y$ . Si le nouveau point de  $\mathbb{R}^{2N}$ ,  $\underline{x}' = (x_1, \dots, x_{j-1}, y, x_{j+1}, \dots, x_N)$  est dans  $\mathcal{C}$ , on va en  $\underline{x}'$ , sinon, on reste en  $\underline{x}$ .

Il s'agit donc d'un algorithme de Metropolis global et de type  $N$ -corps.

### B. Mesure de Gibbs.

Soit  $\Omega = \mathbb{R}^d$  et  $d\mu = dx$  la mesure de Lebesgue. Soit  $U(x)$  une fonction sur  $\mathbb{R}^d$  telle que  $e^{-U(x)}$  soit intégrable. Posons

$$Z = \int e^{-U(x)} dx \quad (4.17)$$

On appelle  $Z$  la fonction de partition (lorsque  $U$  dépend de paramètres auxiliaires, par exemple la température,  $Z$  est fonction de ces paramètres). En pratique,  $Z$  est souvent incalculable. On définit la densité de probabilité  $\pi(x)$  par

$$\pi(x) = \frac{1}{Z} e^{-U(x)} \quad (4.18)$$

Enfin, on choisit une fonction  $\varphi(x) \geq 0$  sur  $\mathbb{R}^d$ , symétrique ( $\varphi(x) = \varphi(-x)$ ) et d'intégrale 1,  $\int \varphi(x) dx = 1$ . On définit alors  $k(x, y)$  par

$$k(x, y) = \varphi(x - y) \quad (4.19)$$

Le taux d'acceptation de l'algorithme de Metropolis est

$$A(x, y) = \frac{\pi(y)}{\pi(x)} = e^{-(U(y)-U(x))} \quad \text{si } \varphi(x - y) \neq 0, \quad A(x, y) = 1 \quad \text{si } \varphi(x - y) = 0 \quad (4.20)$$

et le point fondamental est que ce taux d'acceptation ne dépend pas de  $Z$ , mais seulement de la fonction  $U$  qu'on sait calculer en pratique.

Dans les applications à la physique statistique, la variable sans dimension  $x$  repère les états possibles du système physique. On a

$$U(x) = H(x)/kT \quad (4.21)$$

où  $H(x)$  est l'énergie du système dans l'état  $x$ ,  $k = 1.3806503 \cdot 10^{-23} \text{kgm}^2 \text{s}^{-2} \text{K}^{-1}$  est la constante de Boltzmann et  $T$  la température mesurée en Kelvin. Il est illusoire de vouloir calculer naïvement la fonction de partition  $Z(T)$  pour un système à  $N$  particules.

La probabilité  $\pi(x)dx = \frac{1}{Z(T)}e^{-H(x)/kT}dx$  s'appelle la distribution de Gibbs ou de Boltzmann. L'énergie interne du système est

$$\langle H \rangle = E_\pi(H) = \frac{1}{Z(T)} \int H(x)e^{-H(x)/kT} dx \quad (4.22)$$

L'énergie libre  $F$  est définie par

$$F = -kT \log(Z(T)) \quad (4.23)$$

et l'entropie du système est par définition

$$S = (\langle H \rangle - F)/T \quad (4.24)$$

La chaleur spécifique  $C$  est la quantité macroscopique

$$C = \partial_T \langle H \rangle = \frac{1}{kT^2} \text{var}_\pi(H) \quad (4.25)$$

Les transitions de phases correspondent à des variations très rapides de la fonction de partition  $T \rightarrow Z(T)$ . En posant  $\beta = 1/kT$ , on a trivialement

$$\partial_\beta \log Z = - \langle H \rangle \quad (4.26)$$

de sorte que le calcul de l'énergie interne  $\langle H \rangle = E_\pi(H)$  du système permet de calculer les variations de  $Z(T)$ . Or d'après le théorème 3.26, le calcul de l'espérance  $E_\pi(H)$  peut être effectué par une méthode de type Monte-Carlo fondée sur un algorithme de type Metropolis. On remarquera que pour un système à  $N$  particules, même si à  $N$  fixé,  $Z(T)$  est analytique en  $T$ , son comportement pour  $N$  grand peut ressembler à celui d'une fonction discontinue pour des valeurs dites critiques de la température.

### C. Métropolis sur un ouvert borné.

Soit  $\Omega$  un ouvert borné de  $\mathbb{R}^d$ ,  $\rho(x) \geq 0$  une densité de probabilité sur  $\Omega$ , et  $d\pi = \rho(x)dx$  la probabilité associée.

Lorsque  $\rho(x) = 1_\Omega(x)/\text{Vol}(\Omega)$  est la densité uniforme, une méthode simple pour choisir un point au hasard dans  $\Omega$  consiste à choisir une boîte  $Q = \prod_{i=1}^d [a_i, b_i]$  qui contienne  $\Omega$  et à implémenter l'algorithme de Metropolis correspondant au choix  $k(x, y) = \frac{1_Q}{\text{Vol}(Q)}$ . Essentiellement, cela revient à faire tomber des points au hasard dans  $Q$ , et à ne conserver que ceux qui tombent dans  $\Omega$ . Le taux d'acceptation est alors

$$A(x, y) = 1_{Q \setminus \Omega}(x) + 1_\Omega(x)1_\Omega(y) \quad (4.27)$$

L'algorithme s'écrit :

Si la marche aléatoire est en  $x_n \in Q$ , on choisit  $y$  pour la mesure uniforme dans le cube  $Q$  (on sait faire),



- si  $x_n \notin \Omega$  ou  $y \in \Omega$  on va en  $y$ ,  $x_{n+1} = y$
- si  $x_n \in \Omega$  et  $y \notin \Omega$  on reste en  $x_n$ ,  $x_{n+1} = x_n$ .

On remarquera que dès que la marche est rentrée dans  $\Omega$ , elle ne peut plus en sortir. Pour faire tourner cet algorithme, il n'est pas nécessaire de connaître le volume de  $\Omega$ . Le noyau markovien associé est de type Metropolis

$$M(f)(x) = 1_\Omega(x) \left(1 - \frac{\text{Vol}(\Omega)}{\text{Vol}(Q)}\right) f(x) + \frac{1_\Omega(x)}{\text{Vol}(Q)} \int_\Omega f(y) dy + \frac{1_{Q \setminus \Omega}(x)}{\text{Vol}(Q)} \int_Q f(y) dy \quad (4.28)$$

En décomposant toute fonction  $f$  sur  $Q$  en  $f = g + h$ , avec  $g = 1_\Omega f$  et  $h = 1_{Q \setminus \Omega} f$ , on écrit  $M$  sous forme matricielle

$$M = \begin{pmatrix} A & 0 \\ B & D \end{pmatrix} \quad (4.29)$$

Avec  $C = \frac{\text{Vol}(\Omega)}{\text{Vol}(Q)}$ , on a

$$A(g) = (1 - C)g + C \int_\Omega g d\pi, \quad B(g) = C \int_\Omega g d\pi, \quad D(h) = \frac{1}{\text{Vol}(Q)} \int_{Q \setminus \Omega} h dx \quad (4.30)$$

Si on se restreint à choisir le point de départ de la marche dans  $\Omega$ , seul le noyau de Markov  $A$  opérant sur  $L^\infty(\Omega)$  intervient. Comme  $A = (1 - C)Id + C\Pi_\pi$ , on a  $\text{gap} = C$ ,  $\text{Spec}(A) = \{1 - C, 1\}$  et on obtient d'après (3.21), en utilisant  $(Id - \Pi_\pi)^n = Id - \Pi_\pi$

$$\sup_{x \in \Omega} \|(A)^n(\delta_{y=x}) - \pi\|_{VT} = \frac{1}{2} \|(1 - C)^n (Id - \Pi_\pi)\|_{L^\infty(\Omega)} = (1 - C)^n \quad (4.31)$$

Il est donc clair que cette méthode est désastreuse si  $C = \frac{\text{Vol}(\Omega)}{\text{Vol}(Q)}$  est trop petit. Evidemment, si  $\text{Vol}(\Omega)$  est très petit devant  $\text{Vol}(Q)$ , en faisant tomber des points au hasard dans  $Q$ , il y a peu de chances qu'ils tombent dans  $\Omega$ ! Par exemple, si  $\Omega$  est la boule unité de  $\mathbb{R}^{2m}$ , on a  $\text{Vol}(\Omega) = \frac{\pi^m}{m!}$ , le volume du cube standard qui contient la boule unité est  $2^{2m}$ , et  $C = \frac{\pi^m}{4^m m!}$  est bien trop petit dès que  $m$  est grand.

L'algorithme précédent est de type global. Voici un algorithme local pour échantillonner une probabilité  $d\pi = \rho(x)dx$  sur l'ouvert borné  $\Omega$  de  $\mathbb{R}^d$ , avec  $\rho > 0$  sur  $\Omega$ .

On se donne une fonction  $\varphi(x) \geq 0$  sur  $\mathbb{R}^d$ , radiale, telle que  $\int \varphi(x) dx = 1$ . Ce peut être une gaussienne, la fonction caractéristique d'une boule... On pose  $\varphi_h(x) = h^{-d} \varphi(\frac{x}{h})$  et on définit alors l'opérateur de Metropolis  $M_h$  par

$$\begin{aligned} M_h(f)(x) &= m_h(x) f(x) + \int_\Omega \varphi_h(x - y) \min(1, \rho(y)/\rho(x)) f(y) dy \\ m_h(x) &= 1 - \int_\Omega \varphi_h(x - y) \min(1, \rho(y)/\rho(x)) dy \end{aligned} \quad (4.32)$$

L'algorithme pour construire le point  $x_{n+1}$  de la marche, la marche étant en  $x_n \in \Omega$ , est le suivant.

1. On choisit un point  $y$  de  $\mathbb{R}^d$  pour la probabilité  $h^{-d} \varphi(\frac{x_n - y}{h}) dy$ .
2. Si  $y \notin \Omega$ , on pose  $x_{n+1} = x_n$ .

3. Si  $y \in \Omega$  et  $\rho(y) \geq \rho(x)$ , on pose  $x_{n+1} = y$ .
4. Si  $y \in \Omega$  et  $\rho(y) < \rho(x)$ , on tire à pile ou face avec probabilité de gagner  $\rho(y)/\rho(x)$ ; si on gagne, on pose  $x_{n+1} = y$ , si on perd on pose  $x_{n+1} = x_n$ .

On trouvera dans [DLM08] des résultats sur cet algorithme, lorsque le bord de  $\Omega$  est lipschitz, et la fonction  $\rho$  lisse sur  $\bar{\Omega}$ . En particulier, on a le fait que pour  $h$  petit, la convergence vers l'équilibre est contrôlée par la première valeur propre non triviale de l'opérateur elliptique

$$-\Delta + \frac{\nabla(\rho)}{\rho} \nabla \quad (4.33)$$

avec conditions de Neumann au bord, qui est le laplacien associé à la forme de Dirichlet

$$\int_{\Omega} |\nabla f|^2 \rho(x) dx \quad (4.34)$$

On remarquera que le fait que le bord de  $\Omega$  soit lipschitz assure qu'il existe  $c < 1$  tel que  $m_h(x) \leq c$  pour tout  $x \in \Omega$  et tout  $h$  assez petit. On trouvera aussi dans [DL09] des résultats beaucoup plus précis dans le cas  $\Omega = ]-1, 1[$ ,  $\rho = 1/2$ , et où on constate que les phénomènes de bord en  $x = \pm 1$  ne sont pas triviaux.

#### D. Marche aléatoire sur un espace métrique mesuré.

Reprenons l'exemple de la marche aléatoire sur un espace métrique mesuré  $(\Omega, \mathcal{B}, \mu)$  avec  $\mu$  une probabilité. D'après (3.93),  $\mu$  n'est pas en général invariante. On peut appliquer à cette marche la stratégie de Metropolis pour rendre  $\mu$  invariante et  $K_h$   $\mu$ -réversible. Il suffit de poser

$$\begin{aligned} M_h(f)(x) &= m_h(x)f(x) + \int_{d(y,x) < h} \min\left(\frac{1}{\mu(B(x,h))}, \frac{1}{\mu(B(y,h))}\right) f(y) d\mu(y) \\ m_h(x) &= 1 - \int_{d(y,x) < h} \min\left(\frac{1}{\mu(B(x,h))}, \frac{1}{\mu(B(y,h))}\right) d\mu(y) \end{aligned} \quad (4.35)$$

La marche associée est la suivante : si la marche est en  $x_n$ , on choisit  $y \in B(x, h)$  uniforme pour  $\mu$ . Si  $\mu(B(y, h)) \geq \mu(B(x, h))$ , on pose  $x_{n+1} = y$ , sinon, on tire à pile ou face avec probabilité de gagner  $\mu(B(y, h))/\mu(B(x, h))$ . Si on gagne, on pose  $x_{n+1} = y$ , si on perd, on pose  $x_{n+1} = x_n$ .

## 4.4 Quelques problèmes ouverts

- 1. La théorie spectrale des opérateurs de type Metropolis n'est pas bien connue : structure du spectre loin de 1, formules de Weyl précises près de 1...
- 2. Pour un algorithme de Metropolis dans un ouvert borné  $\Omega$ , le comportement des fonctions propres près du bord de  $\Omega$  fait apparaître des phénomènes de couches limites qui restent à analyser.
- 3. La régularité des fonctions propres.
- 4. Que peut-on dire de la convergence vers la mesure invariante lorsque 1 est dans le spectre essentiel de  $M$  ?
- 5. Last but not least, pour les algorithmes à  $N$  corps, étudier l'asymptotique  $N$  grand.

## 5 Marche aléatoire sur une variété

### 5.1 Marche aléatoire et laplacien

Soit  $(M, g)$  une variété riemannienne compacte, connexe, de dimension  $d$ ,  $d_g x$  sa forme volume canonique, et  $d_g(x, y)$  la distance riemannienne sur  $M \times M$ . Pour  $x \in M$  et  $h > 0$ , soit  $B(x, h) = \{y, d_g(x, y) < h\}$  la boule de rayon  $h$  centrée en  $x$ , et  $|B(x, h)| = \int_{B(x, h)} d_g y$  son volume. Pour tout  $h > 0$ , on note  $K_h$  l'opérateur Markovien

$$(K_h f)(x) = \frac{1}{|B(x, h)|} \int_{B(x, h)} f(y) d_g y \quad (5.1)$$

Alors  $K_h$  est réversible pour la probabilité, qui dépend de  $h$ ,

$$d\nu_h = \frac{|B(x, h)|}{Z_h c_d h^d} d_g x \quad (5.2)$$

où  $c_d$  est le volume de la boule unité de  $\mathbb{R}^d$ . On a pour  $h$  petit

$$|B(x, h)| = c_d h^d \left( 1 - \frac{S(x)}{6(d+2)} h^2 + o(h^3) \right) \quad (5.3)$$

où  $S(x)$  est la courbure scalaire en  $x$ , i.e la trace du tenseur de Ricci. La constante de normalisation  $Z_h$  est telle que  $d\nu_h(M) = 1$ . Pour  $h$  petit,  $d\nu_h$  est proche de  $d_g x / \text{Vol}(M)$  et  $Z_h$  proche de  $\text{Vol}(M)$ . Les espaces de Hilbert  $L^2(M, d\nu_h)$  et  $L^2(M, d_g x)$  sont égaux en tant qu'espaces vectoriels, et leurs normes sont uniformément en  $h$  équivalentes. L'opérateur  $K_h$  est compact sur  $L^2(M, d\nu_h)$ . On note

$$0 < \dots \leq \mu_{k+1}(h) \leq \mu_k(h) \leq \dots \mu_1(h) < \mu_0(h) = 1 \quad (5.4)$$

la suite décroissante des valeurs propres positives, avec multiplicitée, de  $K_h$  ( $K_h$  a aussi des valeurs propres négatives si  $h$  est assez petit). L'inégalité  $\mu_1(h) < \mu_0(h)$ , qui équivaut au fait que, à  $h$  fixé,  $K_h$  vérifie une inégalité de Poincaré, est trivialement conséquence de la compacité de  $K_h$  et de la connexité de  $M$  : Si  $K_h(f) = f$ , on a  $\mathcal{E}(f, f) = 0$ , donc d'après (3.10),  $f(x) = f(y)$  p.p pour  $d_g(x, y) < h$ , donc  $f$  est constante.

On note  $G_d(\xi)$  la fonction de  $\xi \in \mathbb{R}^d$ ,

$$G_d(\xi) = \frac{1}{c_d} \int_{|y| \leq 1} e^{iy\xi} dy \quad (5.5)$$

Au facteur  $\frac{1}{c_d}$  près, la fonction  $G_d$  est la transformée de Fourier de la fonction caractéristique de la boule unité de  $\mathbb{R}^d$ , et ne dépend que de  $|\xi|^2$ . On notera  $\Gamma_d(s)$  la fonction définie sur  $[0, \infty[$  par

$$G_d(\xi) = \Gamma_d(|\xi|^2). \quad (5.6)$$

La fonction  $\Gamma_d$  est analytique réelle,  $|\Gamma_d(s)| \leq 1$ , et  $\lim_{s \rightarrow \infty} \Gamma_d(s) = 0$ , puisque  $G_d(\xi)$  est la transformée de Fourier d'une fonction réelle paire de  $L^1$  de masse totale 1. On a près de  $s = 0$

$$\Gamma_d(s) = 1 - \frac{s}{2(d+2)} + \mathcal{O}(s^2) \quad (5.7)$$

On notera aussi  $\Delta_g$  l'opérateur de Laplace-Beltrami sur  $(M, g)$ , et par  $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \leq \dots$  le spectre de l'opérateur  $-\Delta_g$  sur  $L^2(M, d_g x)$ .

Le théorème suivant est démontré dans [LM].

**Théorème 5.1** *Soit  $h_0 > 0$  petit. 1 est valeur propre simple de  $K_h$  et il existe  $\gamma < 1$  tel que pour tout  $h \in ]0, h_0]$  on a  $\text{Spec}(K_h) \subset [-\gamma, 1]$ . Il existe  $c_0 > 0, c_1 > 0$  tels que*

$$c_0 h^2 \leq \text{gap} \leq c_1 h^2 \quad (5.8)$$

Plus précisément, pour tout  $L > 0$ , il existe  $C$  tel que pour tout  $h \in ]0, h_0]$  et tout  $k \leq L$ , on a

$$\left| \frac{1 - \mu_k(h)}{h^2} - \frac{\lambda_k}{2(d+2)} \right| \leq C h^2 \quad (5.9)$$

Soit  $|\Delta_h|$  l'opérateur autoadjoint sur  $L^2(M, d\nu_h)$  défini par

$$1 - K_h = \frac{h^2}{2(d+2)} |\Delta_h| \quad (5.10)$$

Soit  $F_1$  et  $F_2$  les deux fermés de  $\mathbb{C}$ ,  $F_1 = \{z, \text{dist}(z, \text{spec}(-\Delta_g)) \leq \varepsilon\}$ ,  $F_2 = \{z, \text{Re}(z) \geq A, |\text{Im}(z)| \leq \varepsilon \text{Re}(z)\}$ . Soit  $\varepsilon > 0$  petit,  $A > 0$  grand,  $F = F_1 \cup F_2$  et  $U = \mathbb{C} \setminus F$ . Il existe  $C, h_0 > 0$  tels que pour tout  $h \in ]0, h_0]$ , et tout  $z \in U$ , on a

$$\|(z - |\Delta_h|)^{-1} - (z + \Delta_g)^{-1}\|_{L^2} \leq C h^2 \quad (5.11)$$

Soit pour  $a > 0$ ,  $N(a, h)$  le nombre de valeurs propres de  $K_h$  comptées avec multiplicité dans l'intervalle  $[a, 1]$ . Pour tout  $\delta \in ]0, 1[$ , il existe  $C_{\delta, i}$  indépendant de  $h \in ]0, h_0]$ , tel que : Pour tout  $\tau \in [0, (1 - \delta)h^{-2}]$ ,  $N(1 - \tau h^2, h)$  vérifie la loi de Weyl

$$|N(1 - \tau h^2, h) - (2\pi h)^{-d} \int_{\Gamma_d(|\xi|_x^2) \in [1 - \tau h^2, 1]} dx d\xi| \leq C_{\delta, 1} (1 + \tau)^{\frac{d-1}{2}} \quad (5.12)$$

où  $dx d\xi$  est la forme volume canonique de la variété symplectique  $T^*M$ , et  $|\xi|_x$  la longueur riemannienne du vecteur cotangent  $\xi$  au point  $x$ . En particulier, on a

$$N(1 - \tau h^2, h) \leq C_{\delta, 2} (1 + \tau)^{d/2} \quad (5.13)$$

De plus, pour toute fonction propre  $e_k^h$  de  $K_h$  associée à la valeur propre  $\mu_k(h) \in [\delta, 1]$ , on a avec  $\tau_k(h) = h^{-2}(1 - \mu_k(h))$

$$\|e_k^h\|_{L^\infty} \leq C_{\delta, 3} (1 + \tau_k(h))^{d/4} \|e_k^h\|_{L^2}. \quad (5.14)$$

On remarquera que dans le cas  $M = (\mathbb{R}/2\pi\mathbb{Z})^d$ , on a l'égalité

$$K_h = \Gamma_d(-h^2 \Delta_g) \quad (5.15)$$

de sorte que les opérateurs  $K_h$  and  $\Delta_g$  ont exactement les mêmes fonctions propres  $e^{ikx}$ . Dans ce cas, (et plus généralement si  $(M, g)$  est un espace symétrique), le théorème 5.1 est élémentaire. Sur une variété générale  $(M, g)$ , les deux opérateurs  $K_h$  et  $\Delta_g$  ne commutent

pas et la formule 5.15 est fautive. Toutefois, on montre en utilisant des techniques pseudo-différentielles que la formule 5.15 est ‘presque’ vraie pour  $h$  petit, modulo des termes d’ordre inférieur qui font apparaître les variations de la courbure de  $M$ .

La stratégie de Metropolis permet de modifier la chaîne de Markov précédente, de sorte que la nouvelle chaîne admette comme probabilité invariante  $d\mu_M = d_g x / \text{Vol}(M)$ , qui ne dépend pas de  $h$ . Il suffit de poser (voir (4.35))

$$\begin{aligned} M_h(f)(x) &= m_h(x)f(x) + \int_{d_g(y,x) < h} \min\left(\frac{1}{|B(x,h)|}, \frac{1}{|B(y,h)|}\right) f(y) d_g(y) \\ m_h(x) &= 1 - \int_{d_g(y,x) < h} \min\left(\frac{1}{|B(x,h)|}, \frac{1}{|B(y,h)|}\right) d_g(y) \in [0, 1[ \end{aligned} \quad (5.16)$$

L’opérateur  $M_h$  est un opérateur de Metropolis local. Il n’est évidemment pas compact, à cause de la présence du multiplicateur  $m_h(x)$ , mais est auto-adjoint sur  $L^2(M, d_g x)$ , et donc le volume riemannien  $d_g x$  est une mesure stationnaire pour  $M_h$  quel que soit  $h$ . On remarquera que pour  $h$  petit,  $M_h$  est une petite perturbation de  $K_h$ , puisque d’après (5.3) et (5.16) on a  $\sup_{x \in M} m_h(x) \in O(h^3)$ . Le théorème suivant (voir [LM]) dit que la vitesse de convergence vers l’équilibre est estimée pour  $h$  petit, par la première valeur propre non triviale  $\lambda_1$  du laplacien  $-\Delta_g$ .

**Théorème 5.2** *Soit  $h_0 > 0$  petit. Il existe  $A$  tel que pour tout  $h \in ]0, h_0]$  on a :*

$$e^{-\gamma'(h)nh^2} \leq 2 \sup_{x \in M} \|M_h^n(x, dy) - d\mu_M\|_{TV} \leq Ae^{-\gamma(h)nh^2} \quad \text{pour tout } n \quad (5.17)$$

où  $\gamma(h), \gamma'(h)$  sont deux fonctions positives telles que  $\gamma(h) \simeq \gamma'(h) \simeq \frac{\lambda_1}{2(d+2)}$  quand  $h \rightarrow 0$ .

## 5.2 Calcul h-pseudo-différentiel

L’étude d’une famille d’opérateurs de type  $K_h$  ou  $M_h$  fait naturellement intervenir le calcul h-pseudo-différentiel. Ici  $h$  est une longueur caractéristique. Les analystes qui travaillent sur l’équation de Schrödinger de la mécanique quantique ont pris l’habitude de désigner par la lettre  $h$  leur petit paramètre. La vraie constante de Planck,  $h \simeq 6.626 \times 10^{-34} \text{kgm}^2 \text{s}^{-1}$  est une action, l’énergie d’un photon de fréquence  $\nu$  étant  $E = h\nu$ .

La limite semi-classique consiste à étudier le régime  $h \rightarrow 0$ . En mécanique quantique, on retrouve ainsi la mécanique classique. Dans le cadre qui est le nôtre, on trouve que dans la limite  $h \rightarrow 0$ , la marche aléatoire associée à  $M_h$  (ou à  $K_h$ ) converge vers le mouvement brownien sur  $M$ , ce qui est facile à vérifier dans le cas plat  $M = (\mathbb{R}/2\pi\mathbb{Z})^d$ .

Pour  $m \in \mathbb{R}$ , on note  $S^m$  l’ensemble des fonctions  $a(x, \xi, h)$  de classe  $C^\infty$  en  $(x, \xi) \in \mathbb{R}^{2d}$ , qui dépendent de plus du paramètre  $h \in ]0, 1]$ , et qui vérifient que pour tout  $\alpha, \beta$ , il existe une constante  $C_{\alpha, \beta}$  telle que pour tout  $(x, \xi) \in \mathbb{R}^{2d}$  et tout  $h \in ]0, 1]$  on a

$$|\partial_x^\alpha \partial_\xi^\beta a(x, \xi, h)| \leq C_{\alpha, \beta} (1 + |\xi|)^{m - |\beta|} \quad (5.18)$$

Pour  $a \in S^m$ , on note  $Op(a)$  l'opérateur agissant sur l'espace de Schwartz  $\mathcal{S}(\mathbb{R}^d)$

$$Op(a)(f)(x) = (2\pi h)^{-d} \int e^{i(x-y)\xi/h} a(x, \xi, h) f(y) dy d\xi \quad (5.19)$$

Un opérateur de la forme (5.19) s'appelle un opérateur  $h$ -pseudo-différentiel.

**Théorème 5.3** *Pour tout  $a \in S^0$ , l'opérateur  $Op(a)$  est uniformément borné en  $h$  sur l'espace  $L^2(\mathbb{R}^d)$ . De plus, on a  $\|Op(a)\|_{L^2} \leq C$  où  $C$  ne dépend que d'un nombre fini de constantes  $C_{\alpha,\beta}$  intervenant dans (5.18)*

Pour  $a \in S^m, b \in S^k$ , on a  $Op(a)Op(b) = Op(c)$  où  $c = a\#b \in S^{m+k}$  est donné par l'intégrale oscillante

$$c(x, \xi, h) = (2\pi h)^{-d} \int e^{-iz\theta/h} a(x, \xi + \theta, h) b(x + z, \xi, h) dz d\theta \quad (5.20)$$

et admet le développement asymptotique

$$c(x, \xi, h) = \sum_{|\alpha| < N} \frac{h^{|\alpha|}}{i^{|\alpha|} \alpha!} \partial_\xi^\alpha a(x, \xi, h) \partial_x^\alpha b(x, \xi, h) + h^N r_N(x, \xi, h), \quad r_N \in S^{m+l-N} \quad (5.21)$$

On note  $S_{cl}^m$  le sous ensemble de  $S^m$  des  $a(x, \xi, h) \in S^m$  tels qu'il existe une suite  $a_n(x, \xi) \in S^{m-n}, n \geq 0$ , telle que pour tout  $N$ , on a

$$a(x, \xi, h) = \sum_{0 \leq n < N} (h/i)^n a_n(x, \xi) + h^N r_N(x, \xi, h), \quad r_n \in S^{m-N} \quad (5.22)$$

D'après 2.6, on a  $a\#b \in S_{cl}^{m+k}$  pour  $a \in S_{cl}^m$  et  $b \in S_{cl}^k$ .

Soit à présent  $(M, g)$  une variété riemannienne compacte de dimension  $d$ . Dans un système de coordonnées locales, un vecteur tangent  $u \in T_x M$  en  $x$  à  $M$  s'écrit  $u = \sum_i u_i \partial_{x_i}$  et sa longueur est

$$|u|_x^2 = \sum_{i,j} g_{i,j}(x) u_i u_j \quad (5.23)$$

L'élément de volume riemannien est donné par

$$d_g(x) = \sqrt{\det(g)(x)} dx = \sqrt{\det(g_{i,j})(x)} dx \quad (5.24)$$

Rappelons que le laplacien  $\Delta_g$  sur  $M$  est l'opérateur différentiel d'ordre 2 qui s'écrit

$$\Delta_g(f) = \frac{1}{\sqrt{\det(g)(x)}} \sum_{i,j} \partial_{x_i} (g^{i,j}(x) \sqrt{\det(g)(x)} \partial_{x_j}(f)) \quad (5.25)$$

où  $(g^{i,j})(x)$  est la matrice inverse de la matrice  $(g_{i,j})(x)$ .

Soit  $e_j(x) \in C^\infty(M), j \geq 0$  une base orthonormale dans  $L^2(M, d_g x)$  de fonctions propres réelles de  $-\Delta_g$  avec  $-\Delta_g e_j = \lambda_j e_j$ . Pour toute distribution  $f \in \mathcal{D}'(M)$ , les coefficients de Fourier de  $f$  sont définis par  $f_j = \int f e_j d_g x$ , et on a  $f(x) = \sum_j f_j e_j(x)$

où la série converge dans  $\mathcal{D}'(M)$ . Pour  $s \in \mathbb{R}$ , soit  $H^s(M) = (1 - \Delta_g)^{-s/2} L^2(M, d_g x)$  l'espace de Sobolev usuel sur  $M$ . pour  $f \in \mathcal{D}'(M)$ , on a  $f \in H^s(M)$  ssi  $\|f\|_{H^s(M)}^2 = \sum_j (1 + \lambda_j)^s |f_j|^2 < \infty$ . On utilise aussi les normes semi-classique définies par

$$\|f\|_{h,s}^2 = \sum_j (1 + h^2 \lambda_j)^s |f_j|^2 \quad (5.26)$$

Une famille d'opérateurs  $R_h$ ,  $h \in ]0, 1]$ , agissant sur l'espace des distributions  $\mathcal{D}'(M)$  est dite régularisante si pour tout  $s, t, N$ ,  $R_h$  envoie  $H^s(M)$  dans  $H^t(M)$  et il existe des constantes  $C_{s,t,N}$  telles que pour tout  $h \in ]0, 1]$  on a

$$\|R_h(f)\|_{H^t(M)} \leq C_{s,t,N} h^N \|R_h(f)\|_{H^s(M)} \quad (5.27)$$

**Définition 5.4** On dit qu'une famille d'opérateurs  $A_h$ ,  $h \in ]0, 1]$ , agissant sur l'espace des distributions  $\mathcal{D}'(M)$ , appartient à l'espace  $\mathcal{E}_{cl}^m$  des opérateurs  $h$ -pseudo-différentiels classiques d'ordre  $m$ , ssi pour tout  $x_0 \in M$ , il existe un ouvert de carte  $U$  centré en  $x_0$  et deux fonctions  $\varphi, \psi \in C_0^\infty(U)$  égales à 1 près de  $x_0$ , avec  $\psi$  égale à 1 au voisinage du support de  $\varphi$ , tels qu'on a

$$A_h \varphi = \psi A_h \varphi + R_h$$

avec  $R_h$  régularisant et il existe  $a \simeq \sum_{n \geq 0} (h/i)^n a_n(x, \xi) \in S_{cl}^m$ , tel que dans la carte locale  $U$ , on a  $\psi A_h \varphi = Op(a)$ .

Le symbole principal de  $A_h$ ,  $\sigma_0(A_h)(x, \xi)$ , est par définition le premier terme  $a_0(x, \xi)$  du développement de  $a(x, \xi, h)$ . C'est une fonction  $C^\infty$  bien définie sur l'espace cotangent à  $M$ ,  $T^*M$ , caractérisée par le fait que pour toute fonction  $\varphi \in C^\infty(M)$ , on a

$$e^{-i\varphi(x)/h} A_h (e^{i\varphi(x)/h}) = \sigma_0(A_h)(x, d\varphi(x)) + \mathcal{O}(h) \quad (5.28)$$

On note  $\mathcal{E}_{cl} = \cup_{m \in \mathbb{R}} \mathcal{E}_{cl}^m$  l'algèbre des opérateurs  $h$ -pseudo-différentiels classiques sur  $M$ . Pour  $A_h \in \mathcal{E}_{cl}^m$  et  $B_h \in \mathcal{E}_{cl}^k$ , on a  $A_h B_h \in \mathcal{E}_{cl}^{m+k}$ ,  $\sigma_0(A_h B_h) = \sigma_0(A_h) \sigma_0(B_h)$ . Le commutateur  $[A_h, B_h] = A_h B_h - B_h A_h$  vérifie  $[A_h, B_h] \in h \mathcal{E}_{cl}^{m+k-1}$ . On a

$$\sigma_0\left(\frac{i}{h} [A_h, B_h]\right) = \{\sigma_0(A_h), \sigma_0(B_h)\}$$

où  $\{f, g\}(x) = \sum_i \frac{\partial f}{\partial \xi_i} \frac{\partial g}{\partial x_i} - \frac{\partial f}{\partial x_i} \frac{\partial g}{\partial \xi_i}$  est le crochet de Poisson. Si  $A_h \in \mathcal{E}_{cl}^m$ , son adjoint  $A_h^*$  vérifie  $A_h^* \in \mathcal{E}_{cl}^m$ , et on a  $\sigma_0(A_h^*) = \overline{\sigma_0(A_h)}$ . Enfin si  $A_h \in \mathcal{E}_{cl}^m$ , pour tout  $s \in \mathbb{R}$ , il existe  $C_s$  indépendant de  $h \in ]0, 1]$  tel que

$$\|A_h f\|_{h,s-m} \leq C_s \|f\|_{h,s} \quad \forall f \in H^s(M) \quad (5.29)$$

**Lemme 5.5** Soit  $A_h \in \mathcal{E}_{cl}^0$  vérifiant pour un  $c > 0$ ,  $\inf_{(x,\xi) \in T^*M} |\sigma_0(A_h(x, \xi))| \geq c$ . Alors il existe  $h_0 > 0$  tel que pour tout  $h \in ]0, h_0]$  l'opérateur  $A_h$  est inversible sur  $L^2(M)$  et son inverse  $A_h^{-1}$  appartient à  $\mathcal{E}_{cl}^0$ .

L'opérateur  $-h^2\Delta_g$  appartient à  $\mathcal{E}_{cl}^2$  et son symbole principal est

$$\sigma_0(-h^2\Delta_g) = |\xi|_x^2 = \sum_{i,j} g^{i,j}(x)\xi_i\xi_j \quad (5.30)$$

où  $|\xi|_x$  est la longueur riemannienne du covecteur  $\xi = \sum_i \xi_j dx_j \in T_x^*M$ . Rappelons aussi que pour toute fonction  $\Phi \in C_0^\infty([0, \infty[)$ , l'opérateur  $\Phi(-h^2\Delta_g)$  défini par

$$\Phi(-h^2\Delta_g)(f) = \sum_j \Phi(h^2\lambda_j) f_j e_j(x) \quad (5.31)$$

appartient à  $\mathcal{E}_{cl}^{-\infty} = \cap_m \mathcal{E}_{cl}^m$ , et que son symbole principal est donné par

$$\sigma_0(\Phi(-h^2\Delta_g))(x, \xi) = \Phi(|\xi|_x^2) \quad (5.32)$$

Pour l'étude d'opérateurs de type  $K_h$ , on a besoin d'étendre un peu la définition des opérateurs  $h$ -pseudo-différentiel. En effet un opérateur tel que  $K_h$  ne se comporte pas comme un opérateur  $h$ -pseudo-différentiel à des échelles très petites devant  $h$  (il se comporte comme une somme d'opérateurs intégraux de Fourier). Techniquement, cela entraîne qu'on a pas le contrôle des hautes fréquences donné par (5.18) pour  $|\xi| \rightarrow \infty$ . Ce point n'a de conséquences que pour l'étude de la théorie spectrale de  $K_h$  près de la valeur spectrale 0, et donc est sans importance dans l'analyse de la chaîne de Markov associée.

**Définition 5.6** *On dira qu'une famille d'opérateurs  $C_h$ ,  $h \in ]0, 1]$  agissant sur l'espace des distributions  $\mathcal{D}'(M)$ , appartient  $\tilde{\mathcal{E}}_{cl}^0$  ssi  $C_h$  est borné uniformément en  $h$  sur  $L^2(M)$ , et pour tout  $\Phi_0 \in C_0^\infty([0, \infty[)$ , on a*

$$\Phi_0(-h^2\Delta_g)C_h \quad \text{et} \quad C_h\Phi_0(-h^2\Delta_g) \quad \text{appartient à} \quad \mathcal{E}_{cl}^{-\infty} \quad (5.33)$$

Par exemple, soit  $\Gamma_{d,h}$  l'opérateur  $\Gamma_{d,h} = \Gamma_d(-h^2\Delta_g)$ , défini par

$$\Gamma_{d,h}(f)(x) = \sum_j \Gamma_d(h^2\lambda_j) f_j e_j(x). \quad (5.34)$$

Comme  $\Phi_0\Gamma_d \in C_0^\infty([0, \infty[)$ , on a évidemment  $\Gamma_{d,h} \in \tilde{\mathcal{E}}_{cl}^0$ . On a bien sur  $\mathcal{E}_{cl}^0 \subset \tilde{\mathcal{E}}_{cl}^0$ .

### 5.3 Quelques propriétés de $K_h$ .

Soit  $U \subset M$  un ouvert de carte muni du système de coordonnées  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ . Pour  $x \in U$  et  $r > 0$  petit, la boule géodésique de rayon  $r$  centrée en  $x$  est de la forme

$$B(x, r) = \{x + u, \sum k_{i,j}(x, u) u_i u_j < r^2\} \quad (5.35)$$

où  $(k_{i,j}(x, u))$  est une matrice symétrique définie positive,  $C^\infty$  en  $(x, u)$  telle que  $k_{i,j}(x, 0) = g_{i,j}(x)$ . Pour toute fonction  $f$  à support compact dans  $U$  et  $h$  petit,  $K_h f$  est à support dans  $U$  et on a

$$K_h f(x) = \frac{1}{|B(x, h)|} \int_{t_{uk}(x, u) u < h^2} f(x + u) \sqrt{\det(g(x + u))} du \quad (5.36)$$



En utilisant la variable  $hv = w = k^{1/2}(x, u)u$  dans (5.36), on obtient

$$K_h f(x) = \frac{h^d}{|B(x, h)|} \int_{|v| < 1} f(x + hm(x, hv)v) \rho(x, hv) dv \quad (5.37)$$

où  $m(x, w)$  est une matrice symétrique définie positive,  $C^\infty$  en  $x, w$ , et telle que près de  $u = 0$  on a  $w = k^{1/2}(x, u)u \Leftrightarrow u = m(x, w)w$ , donc  $m(x, 0) = g^{-1/2}(x)$ , et  $\rho(x, w) = \sqrt{\det(g(x+u))} |\det \frac{\partial u}{\partial w}|$  est  $C^\infty$  en  $(x, w)$  et  $\rho(x, 0) = 1$ .

**Lemme 5.7** *Pour  $h_0 > 0$  petit et tout  $k$ ,  $K_h$  est un opérateur borné sur  $C^k(M)$  uniformément en  $h \in ]0, h_0]$ . De plus, il existe  $C$  indépendant de  $h$  tel que, avec  $|\Delta_h|$  défini en (5.10), on a pour tout  $f \in C^2(M)$*

$$\| |\Delta_h| f \|_{L^\infty} \leq C \| f \|_{C^2} \quad (5.38)$$

Dans l'ouvert de carte  $U$  précédent, on peut définir le symbole de  $K_h$ ,  $\sigma(K_h)$  par

$$\sigma(K_h)(x, \xi, h) = e^{-ix\xi/h} K_h(e^{ix\xi/h}) \quad (5.39)$$

Pour tout compact  $K \subset U$ , il existe  $h_K > 0$  tel que  $\sigma(K_h)(x, \xi, h)$  est bien défini pour  $x \in K, \xi \in \mathbb{R}^d$  et  $h \in ]0, h_K]$ . D'après (5.39) on a

$$\sigma(K_h)(x, \xi, h) = \frac{h^d}{|B(x, h)|} \int_{|v| \leq 1} e^{it\xi \cdot m(x, hv)v} \rho(x, hv) dv \quad (5.40)$$

et donc, pour tout  $\alpha, \beta$ , il existe  $C_{\alpha, \beta}$  indépendant de  $h$  tel que

$$|\partial_x^\alpha \partial_\xi^\beta \sigma(K_h)(x, \xi, h)| \leq C_{\alpha, \beta} (1 + |\xi|)^{|\alpha|} \quad (5.41)$$

Comme  $m(x, 0) = g^{-1/2}(x)$  et  $\rho(x, 0) = 1$ , on a

$$\sigma(K_h)(x, \xi, 0) = \Gamma_d(|\xi|_x^2) \quad (5.42)$$

Le lemme suivant est conséquence de la formule (5.37).

**Lemme 5.8** *Soit  $h_0$  petit. Pour  $h \in ]0, h_0]$ , l'opérateur  $K_h$  appartient à  $\tilde{\mathcal{E}}_{cl}^0$ .*

Le lemme suivant est le point clé de l'analyse de  $K_h$ .

**Lemme 5.9** *Soit  $\Phi_0 \in C_0^\infty([0, \infty[)$ , et  $A_h = h^{-2}(K_h - \Gamma_{d,h})\Phi_0(-h^2\Delta_g)$ . Alors  $A_h$  appartient à  $\mathcal{E}_{cl}^{-\infty}$ . Son symbole principal,  $\sigma_0(A_h)$ , vérifie près de  $\xi = 0$ ,*

$$\sigma_0(A_h)(x, \xi) = \left( \frac{S(x)}{3} |\xi|_x^2 (\Gamma_d''(0) - \Gamma_d'(0)^2) + \frac{\Gamma_d''(0)}{3} Ric(x)(\xi, \xi) \right) \Phi_0(|\xi|_x^2) + \mathcal{O}(\xi^3) \quad (5.43)$$

où  $Ric(x)$  est le tenseur de Ricci et  $S(x)$  la courbure scalaire en  $x$ . Soit  $U$  un ouvert de carte,  $K$  un compact de  $U$ , et  $\varphi \in C_0^\infty(U)$  tel que  $\varphi(x) = 1$  au voisinage de  $K$ . Soit  $a(x, \xi, h) \simeq \sum (h/i)^k a_k(x, \xi) \in S_{cl}^{-\infty}$  tel que dans cette carte on a  $A_h \varphi = Op(a) + R_h$  avec  $R_h$  régularisant. Alors, pour tout  $k$  et tout  $x \in K$  on a  $a_k(x, 0) = 0$ .

## 5.4 Convergence vers le mouvement brownien

On expose ici les étapes de la preuve de la convergence de la marche de Metropolis associée à  $M_h$  vers le mouvement Brownien sur la variété  $M$  (on a le même résultat pour la marche associée à  $K_h$ ). Cette preuve ne fait pas usage des équations différentielles stochastiques (voir ([EM89]) et ([Hsu03])). Elle est l'analogie de la preuve classique du fait que la marche aléatoire standard sur le réseau  $h\mathbb{Z}^d$  converge vers le mouvement brownien standard sur  $\mathbb{R}^d$  quand  $h \rightarrow 0$  (voir [KS88], chapter 2.4). Il y a essentiellement deux étapes : la première consiste à vérifier que la famille de mesures associées à la marche est compact dans l'espace de mesures sur  $C^0([0, \infty[, M)$  ; la deuxième consiste à vérifier que n'importe quelle valeur d'adhérence de cette famille est la mesure de Wiener associée à l'équation de la chaleur sur  $M$ . Cette preuve donne en particulier l'existence du mouvement brownien sur  $M$ .

Pour  $x_0 \in M$ , on note  $X_{x_0} = \{\omega \in C^0([0, \infty[, M), \omega(0) = x_0\}$  l'ensemble des trajectoires continues de  $[0, \infty[$  dans  $M$ , partant de  $x_0$ . On munit  $X_{x_0}$  de la topologie de la convergence uniforme sur les compacts de  $[0, \infty[$ , et on note  $\mathcal{B}$  la tribu engendrée par les ouverts de  $X_{x_0}$ .

Soit  $p_t(x, y)d_g y$  le noyau de la chaleur sur  $M$ , i.e le noyau de l'opérateur auto-adjoint  $e^{t\Delta_g/2}$ . La mesure de Wiener,  $W_{x_0}$ , est une probabilité sur  $(X_{x_0}, \mathcal{B})$ , telle que pour tout  $0 < t_1 < t_2 < \dots < t_k$  et tout boréliens  $A_1, \dots, A_k$  de  $M$ , on ait

$$W_{x_0}(\omega(t_1) \in A_1, \omega(t_2) \in A_2, \dots, \omega(t_k) \in A_k) = \int_{A_1 \times A_2 \times \dots \times A_k} p_{t_k - t_{k-1}}(x_k, x_{k-1}) \dots p_{t_2 - t_1}(x_2, x_1) p_{t_1}(x_1, x_0) d_g x_1 d_g x_2 \dots d_g x_k \quad (5.44)$$

D'après le théorème 4.15 of [KS88], si cette mesure existe, elle est uniquement déterminée par la condition (5.44).

Pour  $h \in ]0, 1]$ , soit  $\mathcal{M}_{h, x_0}^{\mathbb{N}}$  le sous espace fermé de l'espace produit  $M^{\mathbb{N}}$ ,

$$\mathcal{M}_{h, x_0}^{\mathbb{N}} = \{\underline{x} = (x_1, x_2, \dots, x_n, \dots), \forall j \geq 0, d_g(x_j, x_{j+1}) \leq h\} \quad (5.45)$$

Equippé de la topologie produit,  $M^{\mathbb{N}}$  est un espace compact métrisable, et la chaîne de Metropolis issue de  $x_0$  définit une probabilité  $\mathcal{P}_{x_0, h}$  sur  $M^{\mathbb{N}}$ , telle que  $\mathcal{P}_{x_0, h}(\mathcal{M}_{h, x_0}^{\mathbb{N}}) = 1$ , en posant pour tout  $k$  et tout boréliens  $A_1, \dots, A_k$  de  $M$ ,

$$\mathcal{P}_{x_0, h}(x_1 \in A_1, x_2 \in A_2, \dots, x_k \in A_k) = \int_{A_1 \times A_2 \times \dots \times A_k} M_h(x_{k-1}, dx_k) \dots M_h(x_1, dx_2) M_h(x_0, dx_1) \quad (5.46)$$

Soit  $j_{x_0, h}$  l'application de  $\mathcal{M}_{h, x_0}^{\mathbb{N}}$  dans  $X_{x_0}$  définie par

$$j_{x_0, h}(\underline{x}) = \omega \iff \forall j \geq 0 \quad \omega(jh^2/(d+2)) = x_j, \quad \text{et} \quad \forall t \in \left[ \frac{jh^2}{d+2}, \frac{(j+1)h^2}{d+2} \right] \\ \omega(t) \quad \text{est la géodésique reliant } x_j \text{ to } x_{j+1} \quad \text{à la vitesse } h^{-2}(d+2)d_g(x_j, x_{j+1}). \quad (5.47)$$

Pour  $h > 0$  fixé et plus petit que le rayon d'injectivité de la variété riemannienne  $M$ , l'application  $j_{x_0,h}$  est bien définie et continue, donc mesurable. Soit  $P_{x_0,h}$  la probabilité sur  $X_{x_0}$  image de  $\mathcal{P}_{x_0,h}$  par l'application mesurable  $j_{x_0,h}$ . Remarquer que dans (5.47), le temps macroscopique  $t$  et le temps microscopique  $n$  sont reliés par  $t \simeq nh^2$ , ce qui correspond bien à l'homogénéité de l'équation de la chaleur, et est le seul régime intéressant dans ce cadre.

**Théorème 5.10** *La limite faible  $W_{x_0} = \lim_{h \rightarrow 0} P_{x_0,h}$  existe, i.e il existe une mesure  $W_{x_0}$  sur  $X_{x_0}$  telle que pour toute fonction continue et bornée  $\omega \mapsto f(\omega)$  sur  $X_{x_0}$ , on a*

$$\lim_{h \rightarrow 0} \int f dP_{x_0,h} = \int f dW_{x_0} \quad (5.48)$$

De plus,  $W_{x_0}$  vérifie (5.44).

La première étape de la preuve consiste à vérifier la proposition suivante. Essentiellement, cette proposition dit que les trajectoires qui sont vues par la probabilité  $\mathcal{P}_{x_0,h}$  restent uniformément dans la classe de Holder  $C^{1/2-0}$ . La régularité  $C^{1/2-0}$  est une propriété classique des trajectoires browniennes.

**Proposition 5.11** *Soit  $\varepsilon_0 > 0$  plus petit que le rayon d'injectivité de  $M$ . Il existe des constantes  $C, A, a, c_0, h_0 > 0$  telles que pour tout  $\varepsilon \in ]0, \varepsilon_0]$ , tout  $\delta \in ]0, c_0\varepsilon^2]$ , et tout  $h \in ]0, h_0]$ , on a*

$$\sup_{x_0 \in M, nh^2 \leq \delta} \mathcal{P}_{x_0,h}(d_g(X_{h,x_0}^n, x_0) > \varepsilon) \leq C\varepsilon^{-A} e^{-a\varepsilon^2/\delta} \quad (5.49)$$

On renvoie à [LM] pour la preuve détaillée de cette proposition. Le lemme simple 5.12 suivant, qui fait intervenir des estimations  $L^2$  à poids exponentiels, et qui est une estimation de type vitesse finie de propagation, y joue un rôle clé.

Si  $\varphi$  est une fonction lipschitzienne sur  $M$ , on note  $M_{h,\varphi}$  l'opérateur borné sur  $L^2$

$$M_{h,\varphi} = e^{\varphi/h} M_h e^{-\varphi/h}$$

**Lemme 5.12** *Soit  $\psi$  une fonction lipschitzienne et à valeurs réelles sur  $M$ ,  $\rho > 0$  et  $0 < \theta < 2\pi$ . On suppose que l'on a*

$$\rho \sin(\theta/2) - \sum_{k=2}^{\infty} \frac{\rho^{k/2} \|\psi\|_{Lips}^k}{k!} |\sin((k-1)\theta/2)| = c > 0 \quad (5.50)$$

Alors, avec  $w = \rho e^{i\theta} \in \mathbb{C} \setminus [0, \infty[$  et  $\varphi = i\rho^{1/2} e^{i\theta/2} \psi$ , on a

$$\|(1 - M_{h,\varphi} - w)^{-1}\|_{L^2} \leq 1/c \quad (5.51)$$

*Démonstration.* Pour  $k(x, y) \in \mathbb{C}$  fonction mesurable et bornée sur  $M \times M$ , on note  $\mathcal{A}_{k,h}$  l'opérateur borné sur  $L^2$

$$\mathcal{A}_{k,h}(f)(x) = \int_{d_g(x,y) < h} \min\left(\frac{1}{|B(x,h)|}, \frac{1}{|B(y,h)|}\right) k(x,y) f(y) d_g y \quad (5.52)$$

Avec  $k^*(x, y) = \bar{k}(y, x)$ , l'adjoint sur  $L^2$  de  $\mathcal{A}_{k,h}$  est égal à  $\mathcal{A}_{k^*,h}$ , et on a les estimations

$$\|\mathcal{A}_{k,h}\|_{L^2} \leq \|k\|_{L^\infty(M \times M)} \quad (5.53)$$

D'après (5.16), on a  $M_h = m_h + \mathcal{A}_{1,h}$ , et un calcul facile donne

$$M_{h,\varphi} = m_h + \mathcal{A}_{k_\varphi,h} \quad k_\varphi(x, y) = 1_{d_g(x,y) \leq h} e^{\frac{\varphi(x) - \varphi(y)}{h}} \quad (5.54)$$

Soit  $\tau(x, y) = 1_{d_g(x,y) \leq h} i(\psi(x) - \psi(y))/h$ . Avec  $\varphi = i\rho^{1/2}e^{i\theta/2}\psi$ , on obtient donc

$$M_{h,\varphi} = m_h + \sum_{k=0}^{\infty} \frac{(\rho^{1/2}e^{i\theta/2})^k}{k!} \mathcal{A}_{\tau^k,h} \quad (5.55)$$

En utilisant 5.55 et  $w = \rho e^{i\theta}$ , on obtient, avec  $S = -e^{-i\theta/2}(1 - M_{h,\varphi} - w)$

$$\begin{aligned} S &= -e^{-i\theta/2}(1 - M_h) + \rho^{1/2} \mathcal{A}_{\tau,h} + \rho e^{i\theta/2} Id + N \\ N &= e^{-i\theta/2} \sum_{k=2}^{\infty} \frac{(\rho^{1/2}e^{i\theta/2})^k}{k!} \mathcal{A}_{\tau^k,h} \end{aligned} \quad (5.56)$$

Puisque  $\tau^* = \tau$ , le deuxième terme de la première ligne de (5.56) est auto-adjoint, d'où

$$\begin{aligned} Im(S) &= \sin(\theta/2)(1 - M_h) + \rho \sin(\theta/2) Id + Im(N) \\ Im(N) &= \sum_{k=2}^{\infty} \frac{\rho^{k/2}}{k!} \sin((k-1)\theta/2) \mathcal{A}_{\tau^k,h} \end{aligned} \quad (5.57)$$

Comme  $\sin(\theta/2)(1 - M_h) \geq 0$ , et puisque d'après (5.53) l'opérateur auto-adjoint  $\mathcal{A}_{\tau^k,h}$  a une norme  $\leq \|\psi\|_{Lips}^k$ , (5.50) et (5.57) impliquent  $Im(S) \geq cId$ , qui est le résultat voulu.  $\square$

Soit à présent  $T > 0$  donné. On a pour  $0 < \delta \leq c_0\varepsilon^2$  et  $h \in ]0, h_0]$

$$\begin{aligned} &\mathcal{P}_{x_0,h}(\exists j < l \leq h^{-2}T, (l-j)h^2 \leq \delta, \quad d_g(X_{x_0}^j, X_{x_0}^l) > 4\varepsilon) \\ &\leq \frac{C}{\delta} \sup_{y_0 \in M} \mathcal{P}_{y_0,h}(\exists j < l \leq h^{-2}\delta, \quad d_g(X_{y_0}^j, X_{y_0}^l) > 4\varepsilon) \\ &\leq \frac{C}{\delta} \sup_{y_0 \in M} \mathcal{P}_{y_0,h}(\exists j \leq h^{-2}\delta, \quad d_g(X_{y_0}^j, y_0) > 2\varepsilon) \\ &\leq \frac{2C}{\delta} \sup_{z_0 \in M, nh^2 \leq \delta} \mathcal{P}_{z_0,h}(d_g(X_{z_0}^n, z_0) > \varepsilon) \end{aligned} \quad (5.58)$$

(d'après 5.49)  $\leq C' \delta^{-(1+A/2)} e^{-a\varepsilon^2/\delta}$

Pour la première inégalité dans (5.58), on utilise juste le fait que l'intervalle  $[0, T]$  est union de  $\simeq C/\delta$  intervalles de longueur  $\delta/2$ . La deuxième inégalité est évidente puisque l'événement  $\{\exists j < l \leq h^{-2}\delta, \quad d_g(X_{y_0}^j, X_{y_0}^l) > 4\varepsilon\}$  est un sous-ensemble de  $\{\exists j \leq h^{-2}\delta, \quad d_g(X_{y_0}^j, y_0) > 2\varepsilon\}$ . Pour la troisième, on utilise le fait que l'événement  $A = \{\exists j \leq h^{-2}\delta, \quad d_g(X_{y_0}^j, y_0) > 2\varepsilon\}$  est contenu dans  $B \cup_{j < k} (C_j \cap D_j)$ , avec  $B = \{d_g(X_{y_0}^k, y_0) > \varepsilon\}$  ( $k$  est le plus grand entier  $\leq \delta h^{-2}$ ),  $C_j = \{d_g(X_{y_0}^j, X_{y_0}^k) > \varepsilon\}$ ,  $D_j = \{d_g(X_{y_0}^j, y_0) > 2\varepsilon$  et  $d_g(X_{y_0}^l, y_0) \leq 2\varepsilon$  pour  $l < j\}$ , et le fait que les événements  $C_j$  and  $D_j$  sont indépendants.

En utilisant la définition (5.47) de l'application  $j_{x_0,h}$ , (5.58) implique pour  $T > 0$  et  $\varepsilon > 0$

$$\lim_{\delta \rightarrow 0} \left( \limsup_{h \rightarrow 0} P_{x_0,h} \left( \max_{|s-t| \leq \delta, 0 \leq s,t \leq T} d_g(\omega(s), \omega(t)) > \varepsilon \right) \right) = 0 \quad (5.59)$$

Donc la famille de probabilités  $P_{x_0,h}$  est "tendue", donc compacte par le théorème de Prohorov.

Il reste à vérifier que toute limite faible  $P_{x_0}$  d'une suite  $P_{x_0,h_k}$ ,  $h_k \rightarrow 0$  vérifie la caractérisation (5.44) de la mesure de Wiener  $W_{x_0}$ . Il suffit pour cela de montrer que pour tout  $m$ , tout  $0 < t_1 < \dots < t_m$ , et toute fonction continue  $f(x_1, \dots, x_m)$ , on a

$$\begin{aligned} \lim_{k \rightarrow \infty} \int f(\omega(t_1), \dots, \omega(t_m)) dP_{x_0,h_k} = \\ \int f(x_1, \dots, x_m) p_{t_m-t_{m-1}}(x_m, x_{m-1}) \dots p_{t_2-t_1}(x_2, x_1) p_{t_1}(x_1, x_0) d_g x_1 d_g x_2 \dots d_g x_m \end{aligned} \quad (5.60)$$

Comme dans [KS88], on peut supposer  $m = 2$  pour alléger les notations. Pour  $t \geq 0$  donné, soit  $n(t, h) \in \mathbb{N}$  le plus grand entier tel que  $h^2 n(t, h) \leq (d+2)t$ . D'après (5.47), on a  $\text{dist}(\omega(t), X_{h,x_0}^{n(t,h)}) \leq h$  et donc  $P_{x_0,h}(\text{dist}(\omega(t), X_{h,x_0}^{n(t,h)}) > \varepsilon) = 0$  pour  $h \leq \varepsilon$ . Il s'agit donc de vérifier

$$\begin{aligned} \lim_{h \rightarrow 0} \int f(X_{h,x_0}^{n(t_1,h)}, X_{h,x_0}^{n(t_2,h)}) dP_{x_0,h} \\ = \int f(x_1, x_2) p_{t_2-t_1}(x_2, x_1) p_{t_1}(x_1, x_0) d_g x_1 d_g x_2 \end{aligned} \quad (5.61)$$

On obtient d'après (5.46)

$$\begin{aligned} \int f(X_{h,x_0}^{n(t_1,h)}, X_{h,x_0}^{n(t_2,h)}) dP_{x_0,h} \\ = \int f(x_1, x_2) M_h^{n(t_2,h)-n(t_1,h)}(x_1, dx_2) M_h^{n(t_1,h)}(x_0, dx_1) \end{aligned} \quad (5.62)$$

D'après (5.61), et (5.62), on doit vérifier que pour toute fonction continue  $f(x_1, x_2)$  sur  $M \times M$ , on a

$$\begin{aligned} \lim_{h \rightarrow 0} \int_{M \times M} f(x_1, x_2) M_h^{n(t_2,h)-n(t_1,h)}(x_1, dx_2) M_h^{n(t_1,h)}(x_0, dx_1) \\ = \int_{M \times M} f(x_1, x_2) p_{t_2-t_1}(x_2, x_1) p_{t_1}(x_1, x_0) d_g x_1 d_g x_2 \end{aligned} \quad (5.63)$$

ou de manière équivalente

$$\begin{aligned} \lim_{h \rightarrow 0} M_h^{n(t_1,h)} \left( M_h^{n(t_2,h)-n(t_1,h)}(f(x_1, \cdot))(x_1) \right) (x_0) \\ = e^{t_1 \Delta_g / 2} \left( e^{(t_2-t_1) \Delta_g / 2} (f(x_1, \cdot))(x_1) \right) (x_0) \end{aligned} \quad (5.64)$$

Puisque  $\|M_h^{n(t,h)}\|_{L^\infty} \leq 1$  et  $\|e^{t \Delta_g / 2}\|_{L^\infty} \leq 1$ , la proposition suivante conclut la preuve du théorème 5.10.

**Proposition 5.13** *Pour tout  $f \in C^0(M)$ , et tout  $t > 0$ , on a*

$$\lim_{h \rightarrow 0} \|e^{t \Delta_g / 2}(f) - M_h^{n(t,h)}(f)\|_{L^\infty} = 0 \quad (5.65)$$

## Bibliographie

- [BD01] L. Billera and P. Diaconis. A geometric interpretation of the metropolis algorithm. *Stat. Sci.*, 20 :1–5, 2001.
- [Bré99] L. Brémaud. Markov chains. Texts in Applied Mathematics 31. Springer Verlag, New York, 1999.
- [Dia] P. Diaconis. The markov chain monte carlo revolution. *Bull AMS*, to appear.
- [DL09] P. Diaconis and G. Lebeau. Spectral analysis for the Metropolis algorithm in Euclidian space. *Mathematische Zeitschrift*, 262(2) :411–442, 2009.
- [DLM08] P. Diaconis, G. Lebeau, and L. Michel. Geometric analysis for the Metropolis algorithm on lipschitz domains. *preprint Nice university*, 2008.
- [DS99] M. Dimassi and J. Sjöstrand. *Spectral Asymptotics in the Semi-Classical Limit*. Lecture Note Series. 268. Cambridge University Press, London, 1999.
- [DSC98] P. Diaconis and L. Saloff-Coste. What do we know about the metropolis algorithm. *Jour. Comp. and Syst. Sci.*, 57 :20–36, 1998.
- [EM89] M. Emery and P.A. Meyer. *Stochastic calculus in manifolds*. Universitext. Springer-Verlag, Berlin, 1989.
- [FOT94] M. Fukushima, Y. Oshima, and M. Takeda. *Dirichlet Forms and symmetric Markov Processes*. Studies in Mathematics n.19. De Gruyter, Berlin, 1994.
- [Has70] W. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57 :97–109, 1970.
- [Hsu03] E. Hsu. *Stochastic analysis on manifolds*. Graduate Studies in Mathematics. Amer. Math. Soc, Providence, 2003.
- [Kah09] M. Kahle. Asymptotically zero-density stable configuration of hard discs. *preprint, Stanford university*, 2009.
- [Kra09] Werner Krauth. Four lectures on computational statistical physics. *arXiv :0901.2496v1*, pages 1–31, 2009.
- [KS88] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate texts in Mathematics. Springer-Verlag, 1988.
- [Liu01] J. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, 2001.
- [LM] G. Lebeau and L. Michel. Semi-classical analysis of a random walk on a manifold. *Annals of Probability*, to appear.
- [Mar02] A. Martinez. *An introduction to Semiclassical and Microlocal Analysis*. Springer, New York, 2002.
- [MRR<sup>+</sup>53] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21 :1087–1092, 1953.
- [Num84] E. Nummelin. *General irreducible Marov Chains and Non-Negative operators*. Cambridge Univ. Press, 1984.
- [Rad08] Charles Radin. Random close packing of granular matter. *J. Stat. Phys.*, 131(4) :567–573, 2008.

- [SC97] L. Saloff-Coste. Lectures on finite markov chains. Lectures notes in math. 1665. Springer, New York, 1997.
- [Tie94] L. Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, 22(4) :1701–1728, 1994.