

Limites singulières d'opérateurs de Schrödinger et de processus de Markov

Y. Colin de Verdière* Y. Pan** B.Ycart**

* Institut Fourier, UMR 5582 CNRS-UJF, BP 74, 38402 Saint Martin d'Hères Cedex France

** LMC/IMAG, BP 53, 38041 Grenoble Cedex 9 France

Abstract

After introducing the Γ -convergence of a sequence of symmetric matrices, we study the limits in that sense, of Schrödinger operators on a finite graph. The main result is that any such limit can be interpreted as a Schrödinger operator on a new graph, the construction of which is described explicitly. The operators to which the construction is applied are reversible, almost reducible Markov generators. An explicit method for computing an equivalent of the spectrum is described. Among possible applications, quasi-decomposable processes, low-temperature simulated annealing and Ising processes are studied.

Résumé

Après avoir introduit la Γ -convergence d'une suite de matrices symétriques, on étudie les Γ -limites possibles d'un opérateur de type Schrödinger sur un graphe fini. Le résultat principal est que toute limite de ce type peut être naturellement interprétée comme un nouvel opérateur de Schrödinger sur un autre graphe, dont la construction est décrite explicitement. Les opérateurs auxquels nous appliquons cette construction sont des générateurs de Markov réversibles presque réductibles, pour lesquels nous donnons une méthode de calcul de l'équivalent du spectre. Parmi les applications, nous étudions les processus quasi-décomposables, l'algorithme du recuit simulé à basse température, et les processus de Ising.

Mots clés : Opérateurs de Schrödinger, Spectres, Processus de Markov

AMS Subject Classification : 47 A 55 , 60 J 27

1 Introduction

Les opérateurs de Schrödinger sur un graphe fini s'interprètent, à des transformations élémentaires près, comme des générateurs de Markov réversibles sur l'ensemble des sommets du graphe.

Après avoir introduit la Γ -convergence d'une suite de matrices symétriques, on étudie les Γ -limites possibles d'un opérateur de type Schrödinger sur un graphe fini. Le résultat principal est que toute limite de ce type peut être naturellement interprétée comme un nouvel opérateur de Schrödinger sur un autre graphe, dont la construction est décrite explicitement.

Les processus de Markov auxquels nous appliquons ce résultat sont presque réductibles, au sens où leur générateur Λ_ε dépend d'un paramètre ε destiné à tendre vers 0. Pour $\varepsilon = 0$, le générateur est réductible et la valeur propre 0 a une multiplicité supérieure à 1. Donc pour $\varepsilon > 0$, certaines valeurs propres sont d'ordre $O(\varepsilon)$. Or ce sont ces valeurs propres "petites" qui contrôlent la vitesse d'accès à l'équilibre du processus et il est important d'en avoir une estimation précise, c'est-à-dire un équivalent. Nous proposons une méthode générale de calcul explicite des équivalents des valeurs propres petites de Λ_ε quand ε tend vers 0. L'idée est la suivante. Certains parmi les taux de transition de Λ_ε sont faibles. Ils correspondent à des transitions lentes. Pour ε suffisamment petit, et sur un intervalle de temps suffisamment court par rapport à $1/\varepsilon$, aucune de ces transitions ne se produira, et l'évolution du processus sera apparemment gouvernée par le générateur Λ_0 . Si on souhaite observer des transitions dont le taux est de l'ordre de ε , il faut se placer sur des intervalles de temps beaucoup plus longs (de l'ordre de $1/\varepsilon$), c'est-à-dire changer l'échelle de temps d'un facteur ε . Ceci revient à multiplier le générateur Λ_ε par $1/\varepsilon$. Le nouveau générateur converge alors au sens de la Γ -convergence vers un générateur sur les classes récurrentes de Λ_0 , dont les valeurs propres, une fois multipliées par ε sont les équivalents de celles des valeurs propres de Λ_ε qui sont d'ordre $O(\varepsilon)$.

La procédure ci-dessus peut être itérée sur des échelles de temps successives de plus en plus lentes ($\varepsilon^2, \varepsilon^3, \dots$). On obtient ainsi une hiérarchie de processus de Markov, définis sur des ensembles de classes de plus en plus réduits. La description de cette hiérarchisation est donnée dans la section 3, et les expressions explicites des générateurs successifs font l'objet du théorème 3.1.

Nous appliquons cette méthode à deux types de processus presque réductibles, les processus quasi-décomposables (section 4) et l'algorithme de recuit simulé à basse température (section 5). Comme cas particulier des résultats de la section 5, nous étudierons dans la section 6 le spectre de la dynamique de Metropolis pour le modèle de Ising.

2 La Γ -convergence

2.1 Définitions et convergence spectrale

Soit $(X, \langle \cdot | \cdot \rangle)$ un espace euclidien de dimension finie n . Soit A un opérateur symétrique sur X . On note Q_A la forme quadratique associée, que l'on écrit aussi :

$$Q_A(x) = \langle x | A | x \rangle .$$

Soit $\Gamma(A) = \{(x, Ax) \mid x \in X\} \subset X \oplus X$ le graphe de A .

Soit $Y \subset X$ un sous-espace euclidien de X et $B : Y \rightarrow Y$ un opérateur symétrique sur Y , on dira que le couple (Y, B) est un opérateur non borné sur X si $Y \neq X$. On étend la définition du graphe aux opérateurs non bornés de la façon suivante.

$$\Gamma(Y, B) = \{(y, z) \in Y \oplus X \mid \forall w \in Y, \langle z | w \rangle = \langle B y | w \rangle\} .$$

Comme cas particulier, $\Gamma(A) = \Gamma(X, A)$. Il est clair que $\Gamma(Y, B)$ est toujours un sous-espace de dimension n de $X \oplus X$.

Soit ω la forme symplectique définie par

$$\omega((x, y), (x', y')) = \langle y' | x \rangle - \langle x' | y \rangle .$$

Un sous-espace de $X \oplus X$ de dimension n , qui est isotrope pour la forme symplectique ω est dit *lagrangien*. On vérifie que tout sous-espace lagrangien est le graphe d'un unique opérateur symétrique, borné ou non, et que réciproquement tout graphe $\Gamma(Y, B)$ est lagrangien. L'ensemble des opérateurs symétriques avec domaine sur X est ainsi une compactification de l'espace des endomorphismes symétriques de X en une variété compacte de classe C^∞ .

On dira que (Y, B) est *inversible* si $\sigma(\Gamma(Y, B))$, où $\sigma(x, y) = (y, x)$, est le graphe d'une application bornée de X dans $Y \subset X$.

Pour $\lambda \in \mathcal{C}$, on définit

$$(Y, \lambda I - B)$$

par son graphe

$$\Gamma(Y, \lambda I - B) = \{(y, \lambda y - \xi) , (y, \xi) \in \Gamma(Y, B)\} .$$

Ceci définit du même coup les deux notions de résolvante (inverse de $\lambda I - B$) et de spectre (valeurs singulières de λ). Ainsi le spectre de (Y, B) est réunion du spectre de B et de la valeur propre ∞ répétée $(n - \dim Y)$ fois.

Définition 2.1 *On dira qu'une famille A_ε , $\varepsilon > 0$ d'opérateurs symétriques sur X Γ -converge vers (Y, B) si $\Gamma(A_\varepsilon)$ converge vers $\Gamma(Y, B)$ quand $\varepsilon \rightarrow 0$ au sens de la topologie naturelle de la grassmannienne des sous-espaces de dimension n de $X \oplus X$. On note*

$$A_\varepsilon \xrightarrow{\Gamma} (Y, B) .$$

Le résultat suivant étend au cadre de la Γ -convergence les résultats classiques de Kato [19] sur la convergence des spectres de formes quadratiques. Il justifie l'utilisation que nous ferons de la Γ -convergence pour l'étude des spectres de processus de Markov quasi-réductibles dans la section 3.

Théorème 2.1 *Supposons $A_\varepsilon \xrightarrow{\Gamma} (Y, B)$. Soient*

$$\lambda_1(\varepsilon) \leq \dots \leq \lambda_n(\varepsilon)$$

les valeurs propres de A_ε . Alors le spectre de A_ε converge vers celui de (Y, B) au sens de la topologie de $P^1(\mathbb{R}) = \mathbb{R} \cup \infty$.

Démonstration

Il suffit de remarquer que la Γ -convergence équivaut à la convergence des résolvantes $(\lambda I - B)$, pour tout $\lambda \notin \mathbb{R}$, comme opérateurs bornés de X dans X .

□

L'existence de Γ -limites n'est pas exceptionnelle.

Proposition 2.1 *Si $A_\varepsilon = \sum_{i=0}^N A_i \varepsilon^{-i}$, où les A_i sont symétriques, la famille A_ε admet une Γ -limite.*

Si $N = 1$, on identifie facilement la limite (voir par exemple la proposition 4.1). Le cas général est beaucoup moins simple (voir section 3, ainsi que [1]).

On a le résultat général suivant :

Théorème 2.2 *Soient $A_i : X \rightarrow X$ des applications linéaires symétriques, $\alpha_1 < \dots < \alpha_N$ et, pour $\varepsilon > 0$:*

$$A_\varepsilon = \sum_{i=1}^N A_i \varepsilon^{\alpha_i} .$$

Soit Z la grassmannienne des sous-espaces lagrangiens de dimension n de $X \oplus X$ et $C_\varepsilon \in Z$ le graphe de A_ε .

Alors l'application $\varepsilon \rightarrow C_\varepsilon$ de $\mathbb{R}^+ \setminus 0$ dans Z se prolonge par continuité en $\varepsilon = 0$.

De plus, si les α_i sont entiers, le prolongement est analytique réel.

Démonstration

Soit $\lambda_0 \notin \mathbb{R}$, alors la résolvante

$$R_\varepsilon = (\lambda_0 I - A_\varepsilon)^{-1}$$

est majorée en norme par $1/|\text{Im}\lambda_0|$ et est une série asymptotique de la forme

$$R_\varepsilon = \sum B_j \varepsilon^{\rho_j} ,$$

avec les B_j non nuls (cela résulte de l'existence de formules explicites pour l'inverse d'une matrice). Cela impose que les ρ_j sont ≥ 0 .

□

La Γ -convergence étant assurée dans de nombreux cas d'intérêt pratique, on cherchera à caractériser une Γ -limite (Y, B) en décrivant le domaine Y et les coefficients de la matrice B .

Dans toute la suite, les opérateurs symétriques sont toujours supposés positifs.

Définition 2.2 On note $x_\varepsilon \xrightarrow{s} x_0$ si x_ε converge vers x_0 avec $Q_\varepsilon(x_\varepsilon) = O(1)$.

Lemme 2.1 Si $A_\varepsilon \xrightarrow{\Gamma} (Y, B)$, $x_0 \in Y$ ssi il existe x_ε tel que $x_\varepsilon \xrightarrow{s} x_0$.

Démonstration

Soit $X = Y_\varepsilon \oplus Z_\varepsilon$ la décomposition de X en somme d'espaces propres correspondant aux valeurs propres d'ordre $O(1)$ et l'orthogonal. On a : $x_\varepsilon = y_\varepsilon + z_\varepsilon$. On voit que $z_\varepsilon \rightarrow 0$ et $\|A_\varepsilon y_\varepsilon\| = O(Q_\varepsilon(x_\varepsilon)) = O(1)$.

Un argument de suite extraite finit la preuve.

□

Lemme 2.2 Soit Y_ε un sous-espace vectoriel de dimension p de X tel que Y_ε tend vers Y et que, pour $\varepsilon \neq 0$, la norme de la restriction de A_ε à Y_ε est bornée.

Soit ϕ_ε un isomorphisme de Y dans Y_ε tel que pour tout $y \in Y$, lorsque $\varepsilon \rightarrow 0$, $\phi_\varepsilon(y)$ tend vers y et

$$Q_\varepsilon(y) = \langle \phi_\varepsilon(y) | A_\varepsilon | \phi_\varepsilon(y) \rangle .$$

Alors Q_ε admet une limite Q_0 comme forme quadratique sur Y et si B est l'opérateur associé, on a :

$$A_\varepsilon \xrightarrow{\Gamma} (Y, B) .$$

Remarquons que l'hypothèse que la norme de la restriction de A_ε est bornée entraîne que $\phi_\varepsilon(y) \xrightarrow{s} y$. Une version pratique de ce lemme est la suivante : soit (y_i) une base de Y et $y_{i,\varepsilon}$ convergents vers y_i avec $A_\varepsilon y_{i,\varepsilon} = O(1)$, alors la matrice de $b_{i,j} = \langle y_i | B | y_j \rangle$ de B est donnée par :

$$b_{i,j} = \lim \langle y_{i,\varepsilon} | A_\varepsilon | y_{j,\varepsilon} \rangle .$$

Attention : il ne faut pas prendre pour définition de Q_ε

$$Q_\varepsilon(y) = \langle y | A_\varepsilon | y \rangle ,$$

qui pourrait ne pas avoir de limite ou une limite incorrecte !

Exemple 2.1 *Plaçons-nous sur $X = \mathbb{R}^4$, avec la forme quadratique*

$$\langle x|A_\varepsilon|x \rangle = \sum_{i=1}^3 (x_i - \frac{x_0}{\varepsilon})^2.$$

Prenons $p = 3$ et $Y_\varepsilon = \{y_0 = (\varepsilon/3)(y_1 + y_2 + y_3)\}$.

Alors $Q_0(y) = \frac{1}{3} \sum_{i=1}^3 (y_i - y_{i+1})^2$ et non $\sum_{i=1}^3 y_i^2$.

Démonstration

1er cas : on suppose que Y_ε est indépendant de ε et donc égal à Y . On décompose X en $X = Y \oplus Z$ (somme orthogonale) et on écrit alors A_ε matriciellement relativement à la décomposition précédente :

$$A_\varepsilon = \begin{pmatrix} K_\varepsilon & C_\varepsilon \\ C_\varepsilon^t & D_\varepsilon \end{pmatrix}.$$

Il résulte des hypothèses et du minimax que $D^{-1} = o(1)$ et que K_ε et C_ε sont uniformément bornés.

On peut écrire le graphe de A_ε sous la forme :

$$\begin{aligned} z &= D_\varepsilon^{-1}(z' - C_\varepsilon^t y) \\ x' &= K_\varepsilon y + C_\varepsilon z. \end{aligned}$$

L'existence de la limite au sens de la Γ -convergence montre que K_ε a une limite K_0 et que la limite des graphes est $z = 0$, $y' = K_0 y$ qui est bien la conclusion du théorème dans ce cas.

2ème cas :

soit U_ε une famille continue d'isométries de X telle que $U_0 = Id$ et $U_\varepsilon(Y) = Y_\varepsilon$. On pose $\tilde{A}_\varepsilon = U_\varepsilon^t A_\varepsilon U_\varepsilon$, ce qui nous ramène au 1er cas. Alors la limite au sens des graphes de \tilde{A}_ε est la même que celle de A_ε . On obtient comme forme quadratique limite la limite sur Y de

$$q_\varepsilon(y) = \langle U_\varepsilon y | A_\varepsilon | U_\varepsilon y \rangle$$

qui est la même qu'en remplaçant dans la formule précédente $U_\varepsilon y$ par des y_ε qui vérifient $\|U_\varepsilon y - y_\varepsilon\| = o(1)$ alors que A_ε est uniformément borné.

□

Nous terminons par le cas particulier d'une suite croissante de formes quadratiques positives, pour lequel la notion de Γ -limite se ramène à celle de limite simple.

Théorème 2.3 Soit A_n une suite croissante de formes quadratiques positives définies sur X et $q_n(x) = \langle x|A_n|x \rangle$ la suite croissante des formes quadratiques associées. Soit

$$q_\infty : X \rightarrow \mathbb{R} \cup +\infty$$

la limite simple des q_n , et

$$Y = \{x \in X \mid q_\infty(x) < \infty\}.$$

Soit (Y, B) la forme quadratique avec domaine associée (si $y \in Y$, $q_\infty(y) = \langle y|B|y \rangle$). Alors

$$A_n \xrightarrow{\Gamma} (Y, B).$$

La démonstration du théorème 2.3 passe par les deux lemmes suivants.

Lemme 2.3 Soit (Y, B) une forme quadratique avec domaine, alors $(x_0, z_0) \in \Gamma(Y, B)$ si et seulement si x_0 est un point critique de l'application F_{z_0} définie sur Y par

$$F_{z_0}(y) = \langle y|B|y \rangle - 2 \langle z_0|y \rangle.$$

Ce lemme est évident ; en particulier, si K est un compact de X et B est positive, un point x_0 de l'intérieur de K est un point de minimum de F_{z_0} si et seulement si $(x_0, z_0) \in \Gamma(Y, B)$.

Lemme 2.4 Soit K un compact et f_n une suite de fonctions semi-continues inférieurement (sci) définies sur K et à valeurs dans $[0, +\infty]$, telles que f_n converge simplement vers une fonction f_∞ et qu'il existe des fonctions sci sur K , $g_n \leq f_n$ telles que la suite g_n soit croissante et ait même limite simple que la suite f_n . Alors si m_n (resp. m_∞) = $\inf_{x \in K} f_n(x)$ (resp. $\inf_{x \in K} f_\infty(x)$), $m_\infty = \lim m_n$ et, si $f_n(x_n) = m_n$ et x_n tend vers a , alors $f_\infty(a) = m_\infty$.

Démonstration

(du lemme 2.4)

1) f_∞ est sci comme limite croissante de fonctions sci et donc il existe $x_\infty \in K$ tel que $f_\infty(x_\infty) = m_\infty$. On a évidemment $\limsup m_n \leq m_\infty$. Supposons donc que $\liminf m_n = m_\infty - \alpha$, $\alpha > 0$; Soit

$$K_n = \{x \in K \mid g_n(x) \leq m_\infty - \frac{\alpha}{2}\}.$$

Il est clair que la suite des compacts K_n est décroissante et d'intersection vide. Donc, il existe $K_{n_0} = \emptyset$ et donc $g_n(x) > m_\infty - \frac{\alpha}{2}$ pour $n \geq n_0$ et $x \in K$. On a ainsi une contradiction car $f_n \geq g_n$.

2.a) Si $f_n(x_n) = m_n$ et $x_n \rightarrow a$, alors $m_n = f_n(x_n) \leq f_n(a)$ et en prenant les limites des 2 membres :

$$m_\infty \leq f_\infty(a).$$

2.b) Supposons donc $f_\infty(a) = m_\infty + \beta$ avec $\beta > 0$. Soit U un voisinage compact de a où $f_\infty \geq m_\infty + \frac{\beta}{2}$. On applique le 1) aux restrictions à U . On a donc, pour n assez grand, $\inf_U f_n = m_n$ et $\inf_U f_\infty \geq m_\infty + \frac{\beta}{2}$. D'où contradiction.

□

Démonstration

(du théorème 2.3).

On suppose donc que q_∞ associée à (B, Y) est la limite simple des q_n . Soit $(x_n, z_n) \in \Gamma(A_n)$ et supposons que (x_n, z_n) tend vers (a, b) . Il suffit de montrer que $(a, b) \in \Gamma(Y, B)$. Pour cela, on applique le lemme 2.4 à la suite $q_n - 2 < z_n | . >$, restreinte à un compact K de X contenant a comme point intérieur. On conclut alors par le lemme 2.3.

□

2.2 Le résultat principal

Soit $G = (V, E)$ un graphe fini, on munit \mathbb{R}^V de la structure euclidienne canonique.

Définition 2.3 *Un opérateur symétrique $A : \mathbb{R}^V \rightarrow \mathbb{R}^V$ est un opérateur de Schrödinger sur G (noté $A \in O_G$) si les conditions suivantes sont satisfaites :*

- (i) *si $\{i, j\} \in E$, $a_{i,j} < 0$.*
- (ii) *si $\{i, j\} \notin E$ et $i \neq j$, $a_{i,j} = 0$.*
- (iii) *L'opérateur A est positif.*

Théorème 2.4 *Soit $A_\epsilon \in O_G$. Supposons que $A_\epsilon \xrightarrow{\Gamma} (Y, B)$ avec $p = \dim Y$.*

Description de Y

Il existe une partition (définie par A_ϵ)

$$V = W_0 \cup V_1 \cup \dots \cup V_p ,$$

où les V_i , $i \geq 1$ sont connexes, telle que si $F_i = \{\varphi \in Y \text{ t.q. } \text{Supp}(\varphi) \subset V_i\}$, les F_i , $i \geq 1$ sont de dimension 1 et $Y = \bigoplus_{i=1}^p F_i$.

Chaque F_i est engendré par une fonction $\varphi_i \geq 0$ de norme 1. Ces φ_i forment donc une base orthonormée de Y .

Description de B

La matrice $(b_{i,j})$ de B dans la base des ϕ_i est un élément de O_{G_0} où $V(G_0) = \{1, \dots, p\}$ et, si $\{i, j\} \in E(G_0)$, il existe un chemin γ de G qui joint V_i à V_j sans rencontrer les autres V_l .

Démonstration

L'espace Y

Le lemme 2.1 implique que, si $\phi \in Y$, alors $|\phi| \in Y$; en effet

$$Q_\varepsilon(|\phi|) \leq Q_\varepsilon(\phi) .$$

Soit $Z \subset V$ l'ensemble des sommets i tels qu'il existe $\phi \in Y$ avec $\phi(i) \neq 0$. Pour $i \in Z$, soit $\varepsilon_i \in Y' \setminus 0$ définie par $\varepsilon_i(\phi) = \phi(i)$. Introduisons la relation d'équivalence dans Z donnée par la dépendance linéaire de ε_i et ε_j . Soient W_1, \dots, W_q les classes d'équivalence et soit $a_i \in W_i$. Alors l'application

$$Y \rightarrow \mathbb{R}^q$$

qui à ϕ associe $(\phi(a_i))$ est un isomorphisme. En effet, elle est visiblement injective. Pour la surjectivité on peut utiliser le lemme 12-1 p. 141 de [4]. L'argument précédent montre aussi que les ϕ_i peuvent être choisies strictement positives sur W_i .

Les V_i

Les W_i sont disjoints, mais pas toujours connexes. On va définir V_i tels que $W_i \subset V_i$. On définit en fait les V_i comme l'ensemble des sommets $a \in V$ tels qu'il n'existe pas $\phi_\varepsilon \xrightarrow{s} \phi_i$ avec $\phi_\varepsilon(a) = 0$.

Il existe ainsi $\phi_\varepsilon \xrightarrow{s} \phi_i$, avec $\forall a \notin V_i, \phi_\varepsilon(a) = 0$, et V_i est minimal pour cette propriété. En effet, il suffit de prendre, pour chaque $a \notin V_i, \phi_{a,\varepsilon} \xrightarrow{s} \phi$ avec $\phi_{a,\varepsilon}(a) = 0$ et $\phi_{a,\varepsilon} \geq 0$, puis $\phi_\varepsilon = \inf \phi_{a,\varepsilon}$.

Il est clair que les V_i sont connexes (si $V_i = V_i' \cup V_i''$ non connectés et $W_i \cap V_i' \neq \emptyset$, on peut remplacer ϕ_ε par sa restriction à V_i' et contredire la minimalité).

Soit $a \in V_i \cap V_j$ et supposons qu'on puisse trouver $0 \leq \phi_{i,\varepsilon} \xrightarrow{s} \phi_i, 0 \leq \phi_{j,\varepsilon} \xrightarrow{s} \phi_j$ et que $\frac{\phi_{i,\varepsilon}(a)}{\phi_{j,\varepsilon}(a)} \rightarrow l, 0 \leq l < \infty$.

On peut alors trouver l_ε tel que $\Phi_\varepsilon(a) = \phi_{i,\varepsilon}(a) - l_\varepsilon \phi_{j,\varepsilon}(a) = 0$ et $\Phi_\varepsilon \xrightarrow{s} \phi_i - l \phi_j$. On prend alors $(\Phi_\varepsilon)_+ \xrightarrow{s} \phi_i$ et nulle en a .

Si l n'existe pas, il suffit de permuter les rôles de i et j .

Le graphe G_0

On choisit $a_i \in W_i$, on pose $m_i = \phi_i(a_i) > 0$ et $\Phi_{i,\varepsilon}$ telle que $\Phi_{i,\varepsilon}(a_j) = \delta_{i,j} m_i$ et $A_\varepsilon \Phi_{i,\varepsilon}(x) = 0, \forall x \neq a_1, \dots, a_p$. Cela revient à minimiser $Q_\varepsilon(\Phi)$ avec les valeurs aux a_j imposées. Il est clair que les $\Phi_{i,\varepsilon}$ sont ≥ 0 partout (les remplacer par leurs valeurs absolues garde les conditions limites et diminue la forme quadratique).

Alors les $\Phi_{i,\varepsilon}$ satisfont les hypothèses du lemme 2.2. En effet, si $\psi_{i,\varepsilon} \xrightarrow{s} \phi_i$ sont telles que $\psi_{i,\varepsilon}(a_j) = \delta_{i,j} m_i$, on a :

$$Q_\varepsilon(\Phi_{i,\varepsilon}) \leq Q_\varepsilon(\psi_{i,\varepsilon}) = O(1) .$$

Donc $\Phi_{i,\varepsilon} \xrightarrow{s} \phi_i$, car les $\phi_{i,\varepsilon}$ ont même valeur aux a_j que $\psi_{i,\varepsilon}$. Puis $A_\varepsilon \Phi_{i,\varepsilon}(x) = 0$ si $x \neq a_1, \dots, a_p$, et

$$A_\varepsilon \Phi_{i,\varepsilon}(a_j) = \frac{1}{m_j} \langle \Phi_{i,\varepsilon} | A_\varepsilon | \Phi_{j,\varepsilon} \rangle = O(1) ,$$

par Cauchy-Schwarz.

Donc les éléments de matrice $b_{i,j}$ de la limite B sont donnés par

$$b_{i,j} = \lim \langle \Phi_{i,\varepsilon} | A_\varepsilon | \Phi_{j,\varepsilon} \rangle .$$

Montrons que, si $i \neq j$, $b_{i,j} \leq 0$: la positivité de $\Phi_{i,\varepsilon}$ au voisinage de a_j le donne immédiatement.

Montrons que, s'il n'existe pas de chemin γ de V_i à V_j sans rencontrer les autres V_l , alors $b_{i,j} = 0$.

On montre d'abord que, pour $a \in V_k$, $\Phi_{i,\varepsilon}(a) = o(\Phi_{k,\varepsilon}(a))$.

Sinon quitte à extraire des sous-suites, on aurait $\Phi_{k,\varepsilon}(a) - c_\varepsilon \Phi_{i,\varepsilon}(a) = 0$ avec c_ε borné. Donc

$$|\Phi_{k,\varepsilon} - c_\varepsilon \Phi_{i,\varepsilon}| \xrightarrow{s} \phi_k + |c| \phi_i .$$

Puis, en prenant le inf avec $|c_\varepsilon| \Phi_{i,\varepsilon}$, on obtient que $a \notin V_k$.

On a donc

$$\Phi_{i,\varepsilon}(a) = o\left(\sum_{k \neq i} \Phi_{k,\varepsilon}(a)\right)$$

pour tout $a \in \cup_{k \neq i} V_k$. On en déduit la même estimation sur l'ensemble des a qui ne sont pas joints à V_i sans rencontrer un des V_k (positivité des solutions du problème de Dirichlet avec données au bord positives). En particulier, c'est vrai des voisins de a_j . On en déduit

$$\langle \Phi_{i,\varepsilon} | A_\varepsilon | \Phi_{j,\varepsilon} \rangle = o\left(\sum_{k \neq i} \langle \Phi_{k,\varepsilon} | A_\varepsilon | \Phi_{j,\varepsilon} \rangle\right) ,$$

et donc $b_{i,j} = 0$.

□

En général le graphe G_0 n'est pas un mineur de G , comme nous l'avons constaté dans le cas de l'exemple 2.1 qui correspond à une transformation étoile-triangle (voir aussi [1]). Il est intéressant de rechercher des conditions suffisantes sur la famille A_ε pour que G_0 soit un mineur de G . Une telle condition est $A_\varepsilon \mathbb{1}_V = 0$, car alors $W_0 = \emptyset$. Nous retrouverons ce cas particulier pour les processus quasi-décomposables (section 4). Une autre condition suffisante est conséquence du théorème 2.3.

Proposition 2.2 *Soit A_n une suite croissante d'opérateurs de O_G et $A_\infty \in O_{G_0}$ sa Γ -limite, alors G_0 est un mineur de G .*

Démonstration

Dans ce cas, d'après le théorème 2.3, il suffit de regarder la convergence simple sur Y : on en déduit que les W_i sont déjà connexes et qu'il ne peut y avoir d'arête de G_0 que si il y a des sommets $a_i \in W_i$ et $a_j \in W_j$ tels que $\{a_i, a_j\} \in E(G)$.

□

3 Processus de Markov réversibles

3.1 Définitions

Soit $\Lambda = (\lambda_{ij})_{i,j \in V}$ le générateur d'un processus de Markov sur un espace d'états V fini. C'est une matrice carrée dont les éléments non diagonaux sont positifs ou nuls, la somme des éléments d'une même ligne étant nulle.

$$\forall i \in V, \quad \lambda_{ii} = - \sum_{j \neq i} \lambda_{ij} .$$

Soit p une mesure de probabilité strictement positive sur V . On dit que Λ admet p comme mesure réversible, ou plus simplement que Λ est p -réversible, si la condition suivante, dite de *bilan détaillé*, est vérifiée (voir par exemple [20]).

$$\forall i, j \in V, \quad p(i)\lambda_{ij} = p(j)\lambda_{ji} . \quad (3.1)$$

Un générateur est donc p -réversible si et seulement si il est auto-adjoint dans $L_2(V, p)$. On associe à un tel générateur son *graphe réduit*, $G = (V, E)$ dont l'ensemble des sommets est V et l'ensemble des arêtes est

$$E = \{\{i, j\}, \lambda_{ij} > 0\} .$$

Le générateur est dit irréductible si son graphe réduit est connexe.

Soit D la matrice diagonale

$$D = \text{diag}(\sqrt{p(1)}, \dots, \sqrt{p(n)}) ,$$

et D^{-1} son inverse. Dire que Λ est p -réversible équivaut à dire que la matrice $D\Lambda D^{-1}$ est symétrique. Son opposé est donc un opérateur de Schrödinger sur le graphe G , au sens de la définition 2.3. Sa seule particularité par rapport à la définition générale 2.3, est d'admettre 0 comme valeur propre, associée au vecteur propre $D\mathbb{1}_V$. Tout générateur réversible est diagonalisable, ses valeurs propres étant réelles, négatives ou

nulles. Si Λ est p -réversible, ses valeurs propres seront classées par ordre décroissant et nous appellerons spectre de Λ le vecteur de ces valeurs propres ordonnées. Avec cette convention, la première coordonnée du spectre est nulle et l'opposé de la seconde est le *trou spectral* de Λ , noté $gap(\Lambda)$.

On étudie généralement la vitesse d'accès à l'équilibre d'un processus de générateur Λ à l'aide de sa forme de Dirichlet (voir Saloff-Coste [29]).

$$\mathcal{E}(\varphi, \varphi) = \langle \varphi | -\Lambda | \varphi \rangle_p = \frac{1}{2} \sum_{i,j \in E} (\varphi(i) - \varphi(j))^2 p(i) \lambda_{ij} .$$

Le point de vue adopté dans la section 2 revient à utiliser la forme quadratique positive Q de matrice $-D\Lambda D^{-1}$.

$$Q(\varphi, \varphi) = \langle \varphi | -D\Lambda D^{-1} | \varphi \rangle_p = \frac{1}{2} \sum_{i,j \in E} \left(\frac{\varphi(i)}{\sqrt{p(i)}} - \frac{\varphi(j)}{\sqrt{p(j)}} \right)^2 p(i) \lambda_{ij} .$$

Bien que la réversibilité de Λ soit une hypothèse très restrictive par rapport au cas général des mesures stationnaires non réversibles, c'est sous cette condition que se placent la plupart des références du domaine [11, 13, 15, 31]. Ceci tient à plusieurs raisons. L'une est bien sûr la simplification du traitement mathématique. Mais il est également vrai que dans la plupart des algorithmes utilisant des chaînes de Markov c'est une chaîne réversible qui est simulée. De plus, il est possible de contrôler la vitesse d'accès à l'équilibre d'un processus de Markov quelconque par le gap d'un générateur réversible (voir [29]).

Rappelons ici la notion de troncature, et le résultat classique sur la conservation de la réversibilité par cette opération (corollaire 1.10 p. 26 de Kelly [20]).

Proposition 3.1 *Soit Λ un générateur irréductible p -réversible sur V , et α un sous-ensemble de V . Considérons les taux de transition λ_{ij}^α définis par*

$$\lambda_{ij}^\alpha = \begin{cases} \lambda_{ij} & \text{si } i, j \in \alpha, \\ 0 & \text{sinon} . \end{cases}$$

Soit Λ^α le générateur correspondant aux taux λ_{ij}^α et supposons qu'il est irréductible sur α . Alors Λ^α est p_α -réversible, où p_α est la loi conditionnelle de p sur α :

$$p_\alpha(i) = \begin{cases} p(i)/p(\alpha) & \text{si } i \in \alpha, \\ 0 & \text{sinon} . \end{cases}$$

3.2 Perturbations analytiques

Désormais ε désigne un paramètre réel positif et Λ_ε un générateur p_ε -réversible et irréductible sur V , tel que les taux de transition λ_{ij}^ε dépendent analytiquement de ε au

voisinage de 0. Il est aisé de déduire de (3.1) que la mesure p_ε est également analytique au voisinage de 0. Les générateurs que nous étudions sont presque réductibles, au sens où la limite de Λ_ε notée Λ_0 , est réductible. D'après la théorie des perturbations des opérateurs symétriques (voir [19]), les valeurs propres de Λ_ε ainsi que les sous-espaces propres associés convergent vers les valeurs propres et les sous-espaces propres correspondants de Λ_0 , autrement dit, le spectre de Λ_0 donne l'équivalent des valeurs propres d'ordre $O(1)$ de Λ_ε . Mais Λ_0 étant réductible, la valeur propre 0 a en général une multiplicité strictement supérieure à 1. Ce qui signifie que certaines valeurs propres de Λ_ε sont d'ordre $O(\varepsilon)$. Il est difficile voire impossible de calculer les valeurs propres, même si la taille de l'espace d'états est modérée. Nous proposons une méthode générale de calcul explicite des équivalents des valeurs propres d'ordre $O(\varepsilon^k)$ pour tout $k \geq 1$ qui vise à calculer les valeurs propres sur un espace d'états largement réduit. L'idée est la suivante. Les valeurs propres d'ordre $O(\varepsilon^k)$ correspondent aux mouvements du processus à l'échelle de temps ε^{-k} . Pour calculer leurs équivalents, il faut donc déterminer la Γ -limite de $\varepsilon^{-k}\Lambda_\varepsilon$ dont le spectre fournira les coefficients de ces équivalents, d'après le théorème 2.1. La détermination explicite de cette Γ -limite, et son interprétation en termes probabilistes fait l'objet du théorème 3.1.

3.3 Résultat principal

Notons $S_k(\Lambda)$ le vecteur composé des valeurs propres d'ordre $O(\varepsilon^k)$ du générateur Λ avec multiplicité éventuelle. La Γ -limite de $\varepsilon^{-k}\Lambda_\varepsilon$ est un générateur Λ_k sur un espace d'états de taille $\#S_k$, issu d'une partition de l'ensemble V des états, comme dans le théorème 2.4. Cette partition s'interprète de manière intuitive dans le cadre dynamique des processus de Markov.

Voici l'interprétation pour $k = 1$. Les *classes irréductibles* de Λ_0 sont des sous-ensembles d'états maximaux tels que à l'intérieur de chacun de ces sous-ensembles, tout état i est relié à tout autre état j par une suite de transitions, de taux strictement positifs. Une classe irréductible est *transiente* s'il existe une transition de taux strictement positif qui en sort, elle est *récurrente* sinon (voir par exemple [5] pour ces définitions classiques). Notons \mathcal{E}_1 l'ensemble des classes irréductibles de Λ_0 et \mathcal{F}_1 l'ensemble des classes récurrentes, qui est donc un sous-ensemble de \mathcal{E}_1 . En fait, on peut voir les classes de \mathcal{E}_1 d'une autre façon en considérant le processus Λ_ε . Pour ε suffisamment petit, partant d'une classe transiente élément de $\mathcal{E}_1 \setminus \mathcal{F}_1$, le processus en sort et va vers une classe récurrente dans un temps en moyenne d'ordre $O(1)$, et dans une même échelle de temps, partant d'une classe récurrente de \mathcal{F}_1 , le processus peut atteindre n'importe quel état de cette classe. Or la sortie des classes récurrentes se produira dans un temps en moyenne d'ordre $O(1/\varepsilon)$. Donc si on se place sur une échelle de temps d'ordre $O(1)$, on voit que le processus atteint son équilibre à l'intérieur de chaque classe récurrente de \mathcal{F}_1 , mais on ne voit pas de communication entre les classes de \mathcal{F}_1 . Si on souhaite observer des transitions entre les classes de \mathcal{F}_1 , il faut se placer sur des intervalles de temps beaucoup plus longs, d'ordre $O(1/\varepsilon)$, c'est-à-dire changer

l'échelle de temps d'un facteur ε . A cette échelle, le processus atteint son équilibre à l'intérieur de chaque classe de \mathcal{F}_1 presque immédiatement. On peut alors ignorer la façon dont le processus atteint son équilibre à l'intérieur des classes et identifier les états d'une même classe récurrente en les agrégeant en un seul état. Les classes de \mathcal{F}_1 sont les sous-ensembles de sommets V_1, \dots, V_p du théorème 2.4. La classe W_0 est constituée de tous les autres états. En itérant cette procédure, nous définissons une partition \mathcal{E}_k de l'espace d'états V et un sous-ensemble \mathcal{F}_k de \mathcal{E}_k ayant les propriétés suivantes. Dans un temps d'ordre $O(1/\varepsilon^{k-1})$ en moyenne, partant d'une classe de \mathcal{E}_k le processus peut atteindre n'importe quel état de cette classe et en sortir en allant vers une classe de \mathcal{F}_k si la classe de départ est une classe de $\mathcal{E}_k \setminus \mathcal{F}_k$. Par contre il faut un temps d'ordre $O(1/\varepsilon^k)$ en moyenne pour sortir d'une classe de \mathcal{F}_k . Alors on trouve que non seulement le nombre des valeurs propres d'ordre $O(\varepsilon^k)$ est égal au cardinal de \mathcal{F}_k mais aussi que les équivalents de ces valeurs propres sont le produit de ε^k par les valeurs propres de la Γ -limite de $\varepsilon^{-k}\Lambda_\varepsilon$. La construction récursive des partitions $\{\mathcal{E}_k\}_{0 \leq k \leq K}$ et $\{\mathcal{F}_k\}_{0 \leq k \leq K}$ est assez longue et est décrite explicitement dans [28]. Nous ne la reprendrons pas ici, d'autant qu'elle n'est pas essentiellement nouvelle. On peut la voir comme une décomposition en cycles au sens de Freidlin et Wentzell [14]. Elle a été obtenue par des moyens différents, pour des chaînes de Markov à temps discret avec une interprétation en termes d'échelles de temps analogue à la nôtre, par Delebecque [9].

Les partitions \mathcal{E}_k , $0 \leq k \leq K$ sont emboîtées, de sorte que $\#\mathcal{F}_k$ est décroissant. Soit K le plus petit entier tel que

$$\#\mathcal{F}_K = 1 .$$

Les valeurs propres d'ordre ε^{K-1} sont les plus proches de zéro de tout le spectre, et ε^{K-1} est l'ordre de grandeur du gap.

Posons

$$U_{k+1} = V - \bigcup_{\alpha_k \in \mathcal{F}_k} \alpha_k$$

et notons $\Lambda_{U_{k+1}}$ le générateur restreint de Λ_ε sur U_{k+1} . Pour toute classe α_k de \mathcal{F}_k , notons p_{α_k} la mesure de probabilité réversible du générateur restreint de Λ_ε sur α_k (loi conditionnelle de p_ε sur α_k , voir proposition 3.1). Pour tous sous-ensembles S_1, S_2 de E avec $S_1 \cap S_2 = \emptyset$, notons

$$G_{S_1}^{S_2} = (\lambda_{ab})_{a \in S_1, b \in S_2}$$

la matrice composée des taux de transition de S_1 à S_2 . Posons $D_{U_{k+1}}$ la matrice diagonale telle que

$$D_{U_{k+1}} \mathbb{1}_{U_{k+1}} = \sum_{\alpha_k \in \mathcal{F}_k} G_{U_{k+1}}^{\alpha_k} \mathbb{1}_{\alpha_k}$$

où $\mathbb{1}_{U_{k+1}}$ et $\mathbb{1}_{\alpha_k}$ désignent les vecteurs constants de valeur 1 sur U_{k+1} et α_k respectivement.

Notre résultat principal est le suivant.

Théorème 3.1 *Soit k un entier compris entre 1 et K .*

(1) *Pour tous $\alpha_k \neq \beta_k \in \mathcal{F}_k$, définissons le taux de transition de α_k à β_k comme suit.*

$$\lambda_{\alpha_k \beta_k} = \frac{1}{\varepsilon^k} p_{\alpha_k} \left[G_{\alpha_k}^{\beta_k} \mathbb{1}_{\beta_k} - G_{\alpha_k}^{U_{k+1}} \left(\Lambda_{U_{k+1}} - D_{U_{k+1}} \right)^{-1} G_{U_{k+1}}^{\beta_k} \mathbb{1}_{\beta_k} \right]$$

Alors $\lim_{\varepsilon \searrow 0} \lambda_{\alpha_k \beta_k}$ existe. Soit $l \geq k + 1$ l'entier le plus petit tel que α_k soit contenu dans un élément de \mathcal{F}_l , noté α_l . Alors si $\beta_k \not\subseteq \alpha_l$ on a en plus

$$\lim_{\varepsilon \searrow 0} \lambda_{\alpha_k \beta_k} = 0 .$$

(2) *Pour tout a de U_{k+1} , notons $p_a^{\beta_k}$ la probabilité que le processus Λ_ε , partant de a , atteigne β_k pour la première fois sans visiter les autres classes de \mathcal{F}_k . Soit $P_{U_{k+1}}^{\beta_k}$ le vecteur colonne composé des $p_a^{\beta_k}$ pour a de U_{k+1} . Alors*

$$\lambda_{\alpha_k \beta_k} = \frac{1}{\varepsilon^k} p_{\alpha_k} \left(G_{\alpha_k}^{\beta_k} \mathbb{1}_{\beta_k} + G_{\alpha_k}^{U_{k+1}} P_{U_{k+1}}^{\beta_k} \right) .$$

(3) *Soit Λ_k le générateur sur \mathcal{F}_k composé des taux de transition $\lambda_{\alpha_k \beta_k}$. Alors à une permutation près,*

$$\lim_{\varepsilon \searrow 0} S_0(\Lambda_k) = \lim_{\varepsilon \searrow 0} \frac{1}{\varepsilon^k} S_k(\Lambda) .$$

Le démonstration de ce résultat, assez technique, est due à Y. Pan (voir [28]).

4 Processus quasi-décomposables

La notion de quasi-décomposabilité semble avoir été introduite par les économistes Simon et Ando [30] dans le cadre de l'analyse hiérarchique de systèmes complexes. L'idée, fort naturelle, a ensuite été appliquée en sociologie, biologie, et de manière très poussée pour l'analyse de performance des réseaux informatiques (pour un historique complet, voir l'introduction de Courtois [8]). Cette idée est la suivante. Dans un système dynamique complexe coexistent plusieurs échelles de temps. Pour fixer les idées, parlons de secondes, heures et mois. A l'échelle de quelques secondes, le système peut être décomposé en sous-systèmes isolés qui n'ont que des interactions internes. Ce n'est qu'à l'échelle de plusieurs heures que des interactions sensibles apparaissent entre des composantes jusque là isolées. Ceci induit un regroupement des anciennes composantes en de nouvelles composantes qui constituent une partition du système global. Certaines de ces composantes interagissent à l'échelle du mois, pour constituer par regroupement une partition plus grossière. On obtient ainsi une hiérarchie de partitions emboîtées dont chacune correspond à une échelle de temps différente.

L'idée est formalisée à l'aide de chaînes de Markov dans le premier chapitre de Courtois [8]. Nous l'exprimons à l'aide des processus à temps continu. Les résultats ci-dessous s'adaptent de manière immédiate aux matrices de transition des chaînes à temps discret.

Soit Λ un générateur irréductible sur $V = \{i, j, \dots\}$, correspondant aux taux de transition $(\lambda_{ij})_{i,j \in V}$. Le générateur Λ est supposé p -réversible, où $p = (p(i))_{i \in V}$. Sans perte de généralité, nous supposons que Λ est irréductible, ce qui entraîne que la mesure p est strictement positive. Considérons une partition de V , $\mathcal{F} = \{\alpha, \beta, \dots\}$.

$$\forall \alpha \neq \beta \in \mathcal{F}, \quad \alpha \cap \beta = \emptyset \quad \text{et} \quad \bigcup_{\alpha \in \mathcal{F}} \alpha = V.$$

On note

$$q(\alpha) = \sum_{i \in \alpha} p(i)$$

la mesure de l'ensemble α . La mesure conditionnelle de i sur α sera notée p_α .

$$\forall i \in \alpha, \quad p_\alpha(i) = p(i)/q(\alpha).$$

Pour tout $\varepsilon > 0$, nous définissons le générateur Λ_ε , correspondant aux taux λ_{ij}^ε à partir du générateur Λ en multipliant par ε les taux de transitions entre sous-ensembles différents de la partition \mathcal{F} . Ceci revient à ralentir les échanges entre classes distinctes, ou encore à distinguer l'échelle de temps $1/\varepsilon$ où se produisent les transitions entre classes distinctes, de l'échelle de temps 1 qui reste celle des échanges à l'intérieur de chaque classe.

$$\forall \alpha \in \mathcal{F}, \forall i \neq j \in \alpha, \quad \lambda_{ij}^\varepsilon = \lambda_{ij},$$

$$\forall \alpha \neq \beta \in \mathcal{F}, \forall i \in \alpha, j \in \beta, \quad \lambda_{ij}^\varepsilon = \varepsilon \lambda_{ij}.$$

Les résultats du paragraphe précédent s'appliquent immédiatement à ce cas particulier. Par rapport aux notations de la section 2, l'ensemble W_0 est vide, les ensembles de sommets V_k sont les éléments de la partition \mathcal{F} . Quand ε tend vers zéro, le spectre de Λ_ε , noté s_ε , converge vers le spectre de Λ_0 , qui est la concaténation des spectres des troncatures de Λ aux sous-ensembles de la partition.

Nous notons Λ^α la troncature de Λ à α , qui est p_α -réversible (proposition 3.1), et s_α le spectre de Λ^α (vecteur de dimension $|\alpha|$). Nous désignons par \oplus l'opération de concaténation des spectres, consistant à mettre les vecteurs bout à bout puis à les ordonner par ordre décroissant.

$$\lim_{\varepsilon \rightarrow 0} s_\varepsilon = s_0 = \bigoplus_{\alpha \in \mathcal{F}} s_\alpha.$$

Le résultat ci-dessus indique que parmi les valeurs propres de Λ_ε , $|\mathcal{F}|$ convergent vers 0 (l'une d'entre elles est constamment nulle). Notons s'_ε leur vecteur ($|\mathcal{F}|$ premières coordonnées de s_ε). Une application directe du théorème 3.1 montre que $s'_\varepsilon/\varepsilon$ converge vers le spectre d'un générateur de Markov M sur \mathcal{F} , dit "agrégé".

Proposition 4.1 Soit M le générateur défini par les taux de transition suivants.

$$\forall \alpha \neq \beta \in \mathcal{F}, \quad \mu_{\alpha\beta} = \sum_{\substack{i \in \alpha \\ j \in \beta}} \lambda_{ij} p_{\alpha}(i).$$

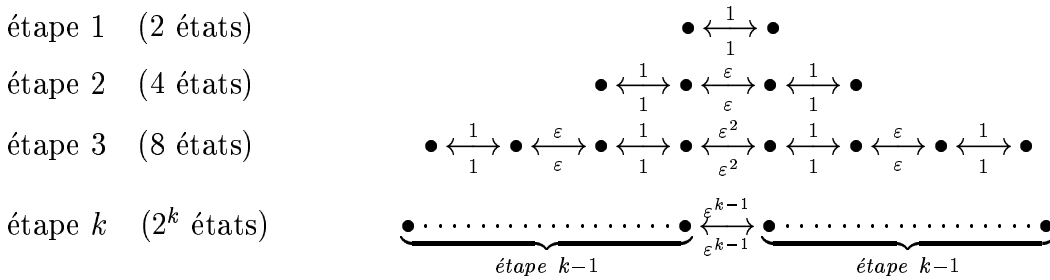
Soit t son spectre.

$$\lim_{\varepsilon \rightarrow 0} s'_{\varepsilon} / \varepsilon = t.$$

On vérifie immédiatement que le générateur M est q -réversible. Le taux $\mu_{\alpha\beta}$ s'interprète comme le flux d'échange de α vers β pour le processus de générateur Λ . L'idée est la suivante. Pour ε petit, les transitions se limitent aux échanges à l'intérieur des classes de la partition. L'état d'équilibre est atteint à l'intérieur de chaque classe avec une très bonne approximation, au bout d'un temps petit devant $1/\varepsilon$. En divisant par ε le générateur Λ_{ε} , on change d'échelle de temps. On peut donc considérer que chacun des processus gouvernés par les Λ_{α} a atteint son état d'équilibre. Seules comptent alors les transitions entre deux classes distinctes. Le taux de transition entre les classes α et β est la moyenne, relative à la mesure d'équilibre sur α des taux de transition depuis les éléments de la classe α vers la classe β .

Le résultat ci-dessus peut évidemment être étendu à d'autres échelles de temps. Il faudrait pour cela supposer l'existence d'une partition plus grossière pour laquelle les taux de transition entre classes distinctes sont d'ordre ε^2 . On obtiendrait alors un certain nombre de valeurs propres d'ordre également ε^2 , le coefficient de l'équivalent étant valeur propre d'un générateur de Markov décrivant les transitions à l'échelle de temps $1/\varepsilon^2$ entre les classes du deuxième niveau. Il n'est pas utile de décrire formellement cette construction, cas particulier du théorème 3.1 qui est essentiellement déjà contenu dans la proposition 4.1.

A titre d'exemple, nous décrivons une famille de processus de naissance et de mort sur $\{1, \dots, 2^k\}$ présentant une hiérarchisation des échelles de temps. La famille est construite de manière itérative selon les diagrammes de transition suivants.



Plus précisément, soit Λ_2 le générateur symétrique sur $\{1, 2\}$,

$$\Lambda_2 = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Pour tout $k \geq 1$, si Λ_{2^k} est défini sur $\{1, \dots, 2^k\}$, définissons le générateur $\Lambda_{2^{k+1}}$ sur $\{1, \dots, 2^{k+1}\}$ par

$$\Lambda_{2^{k+1}} = \begin{pmatrix} \Lambda_{2^k} & (0) \\ (0) & \Lambda_{2^k} \end{pmatrix} + \varepsilon^k B_{2^{k+1}}$$

où $B_{2^{k+1}}$ est le générateur dont les taux de transition entre 2^k et $2^k + 1$ valent 1, les autres étant nuls.

$$B_{2^{k+1}} = (b_{ij})_{1 \leq i, j \leq 2^{k+1}} = \left(\begin{array}{c|c} (0) & (0) \\ \hline -1 & 1 \\ \hline (0) & (0) \end{array} \right) \quad (4.1)$$

avec

$$b_{ij} = \begin{cases} -1 & \text{si } i = j = 2^k \text{ ou } 2^k + 1 \\ 1 & \text{si } (i, j) = (2^k, 2^k + 1) \text{ ou } (i, j) = (2^k + 1, 2^k) \\ 0 & \text{sinon} \end{cases}$$

Pour tout $k \geq 1$, le générateur Λ_{2^k} est irréductible et symétrique. Il est donc réversible par rapport à la mesure uniforme sur $\{1, \dots, 2^k\}$. Pour décrire le comportement des processus correspondants, supposons que l'unité de temps est la seconde et fixons $\varepsilon = 1/3600$. Considérons le processus à 8 états (étape 3) et supposons qu'il parte de 1 à l'instant 0. A l'échelle de la seconde, le processus va osciller entre 1 et 2, et donc se comporter comme le processus de l'étape 1. Ce n'est qu'au bout d'une heure en moyenne qu'il passera de l'état 2 à l'état 3, et se mettra à osciller entre 3 et 4 toutes les secondes environ. A l'échelle de l'heure, le processus se comportera comme celui de l'étape 2. Il faudra attendre environ 3600 heures (150 jours) pour le voir quitter les états $\{1, 2, 3, 4\}$ et visiter l'autre moitié de l'espace d'états.

Les générateurs ainsi construits sont quasi-décomposables. A l'échelle de temps $1/\varepsilon^k$, pour le générateur $\Lambda_{2^{k+1}}$ l'espace d'états est divisé en deux moitiés pour laquelle la dynamique locale est gouvernée par le générateur Λ_{2^k} . En utilisant la proposition 4.1, on vérifie par récurrence le résultat suivant.

Proposition 4.2 *Soient $\lambda_1 = 0 > \lambda_2 > \dots > \lambda_{2^{k+1}}$ les valeurs propres de $\Lambda_{2^{k+1}}$, rangées par ordre décroissant. On a*

$$\lambda_1 = 0 \quad , \quad \lambda_{2^k} = -2$$

et $\forall i = 0, \dots, k-1, \quad \forall j = 2^i + 1, \dots, 2^{i+1}$

$$\lambda_j = -\frac{\varepsilon^{k-i}}{2^{k-i-1}} + o(\varepsilon^{k-i}) \quad , \quad \lambda_{2^k+j} = -2 - \frac{\varepsilon^{k-i}}{2^{k-i-1}} + o(\varepsilon^{k-i}) \quad .$$

Il est naturel de se demander quelle est la qualité de l'approximation obtenue par la décomposition hiérarchique décrite ici. Cette question est discutée dans [8], chapitre

II. Pour l'exemple du générateur Λ_{2^k} construit ci-dessus, nous présentons quelques résultats de comparaison portant sur le gap.

$$\text{gap}(\Lambda_{2^{k+1}}) = \varepsilon^k / 2^{k-1} + o(\varepsilon^k) \quad .$$

Sur des exemples numériques, on constate que cet équivalent fournit une approximation correcte du gap, même pour des valeurs de ε de l'ordre de plusieurs dixièmes. Ces vérifications numériques ont été faites à l'aide de Mathematica [33]. Les tables ci-dessous donnent des valeurs de $\text{gap}(\Lambda_{2^{k+1}})$ et de son estimation par $\varepsilon^k / 2^{k-1}$, pour $k = 2$ et 3 (8 et 16 états respectivement).

ε	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
$\text{gap}(\Lambda_8)$	0,0012	0,0047	0,0101	0,0172	0,0255	0,0347	0,0444	0,0544	0,0645	0,0744
$\varepsilon^2/2$	0,0013	0,0050	0,0113	0,0200	0,0312	0,0450	0,0612	0,0800	0,1012	0,1250

ε	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
$\text{gap}(\Lambda_{16})$	0,0000	0,0002	0,0008	0,0017	0,0031	0,0049	0,0071	0,0096	0,0123	0,0151
$\varepsilon^3/4$	0,0000	0,0003	0,0008	0,0020	0,0039	0,0068	0,0107	0,0160	0,0228	0,0313

5 Recuit simulé à basse température

L'algorithme de Metropolis [25] a suscité une abondante littérature ces dix dernières années (voir [12] et les références qu'il contient). L'intérêt pour cet algorithme provient de ses nombreuses applications, du recuit simulé à l'échantillonnage de Gibbs ([16, 22, 17, 18, 2]). Le problème théorique le plus important posé par ce type de technique est la détermination précise de la vitesse de convergence. Une présentation générale de l'ensemble des résultats rigoureux connus à ce jour est donnée par Diaconis et Saloff-Coste [12]. La plupart portent sur des minoration du trou spectral des chaînes de Markov simulées ou sur des inégalités de Sobolev logarithmiques. Ces résultats ne sont évidemment que des pis-allers, à défaut de pouvoir connaître l'ensemble du spectre de la chaîne, ce qu'interdit dans le cas général la taille de l'espace d'états. Une description complète du spectre a cependant été obtenue dans des cas particuliers par Diaconis et Hanlon [10] et Liu [23]. C'est un résultat de ce type que nous proposons ici, dans le cas où certaines transitions sont très faibles. Un cas typique d'application est celui de l'algorithme du recuit simulé à basse température. Nous nous placerons donc dans ce cadre, en décrivant exactement les équivalents des différentes valeurs propres à basse température.

Rappelons tout d'abord la description classique de l'algorithme du recuit simulé. Soit $V = \{i, j, \dots\}$ un ensemble fini, et H une fonction de V dans \mathbb{R} que l'on souhaite minimiser. Cette fonction est interprétée comme une énergie et la mesure de Gibbs correspondante, notée p est définie par

$$\forall i \in V, \quad p(i) = \frac{1}{Z} e^{-\frac{1}{T} H(i)},$$

où T désigne un paramètre de température, et Z la constante de normalisation (fonction de partition). Quand T tend vers 0, la famille de mesures ainsi définie converge vers la loi uniforme sur l'ensemble des minima globaux de la fonction H . L'algorithme du recuit simulé est une variante de l'algorithme de Metropolis qui simule une chaîne de Markov pour laquelle p est une mesure réversible. En toute généralité, l'algorithme de Metropolis est une méthode de simulation par rejet à partir d'un noyau markovien K , qui n'est pas nécessairement symétrique (voir [12]). Dans la plupart des applications, le noyau K est celui de la marche aléatoire symétrique sur un graphe non orienté dont V est l'ensemble des sommets. C'est dans ce cas que nous nous placerons. L'ensemble des arêtes de ce graphe est noté E , et la relation de voisinage est notée \sim .

$$i \sim j \iff \{i, j\} \in E .$$

L'algorithme est vu ici comme un processus de Markov à temps continu, et décrit par ses taux de transition, qui sont non nuls seulement sur les arêtes du graphe.

$$\forall i \sim j \in E , \quad \lambda_{ij} = \begin{cases} 1 & \text{si } H(j) \leq H(i) \\ e^{-\frac{1}{T}(H(j)-H(i))} & \text{si } H(j) > H(i) . \end{cases}$$

Il est immédiat de vérifier que le générateur ainsi défini est p -réversible. Notre but est d'étudier le spectre de ce générateur à basse température. Nous poserons donc $\varepsilon = e^{-\frac{1}{T}}$, et désignerons par Λ_ε le générateur correspondant. Afin de simplifier les écritures, nous supposons que la fonction H ne peut prendre que des valeurs entières à partir de 0, et que les différences d'énergies entre sommets voisins valent au plus 1.

$$\forall i \in E , H(i) \in \mathbb{N} \quad \inf_{i \in E} H(i) = 0 \quad \forall i \sim j \in E , H(i) - H(j) \in \{-1, 0, 1\} .$$

Cette hypothèse n'est pas vraiment restrictive. Ramener les valeurs d'une fonction quelconque à des valeurs dans \mathbb{R}^+ correspond à une translation de H qui ne change ni la mesure p ni le générateur. D'autre part pour traiter le cas général à partir du cas particulier envisagé ici, il suffit de remplacer les puissances entières de ε qui apparaissent dans les développements ci-dessous par des exposants correspondant aux différences d'énergie entre voisins $H(j) - H(i)$. Nous notons donc désormais Λ_ε le générateur défini sur E par

$$\forall i, j \in E , \quad \lambda_{ij}^\varepsilon = \begin{cases} 0 & \text{si } i \not\sim j \\ 1 & \text{si } i \sim j \text{ et } H(j) \leq H(i) \\ \varepsilon & \text{si } i \sim j \text{ et } H(j) > H(i) . \end{cases}$$

Ce générateur est réversible par rapport à la mesure p_ε , qui est maintenant définie par

$$\forall i \in E , \quad p_\varepsilon(i) = \varepsilon^{H(i)} .$$

Nous étudions donc à nouveau un cas particulier de la section 3. La structure du problème permet une description agréable et naturelle de la décomposition hiérarchique

des classes. Nous serons amenés à effectuer des agrégations d'états, correspondant à des classes connexes au sens du graphe, et d'énergie constante. On notera \mathcal{R} la relation d'équivalence correspondante.

$$\forall i, j \in E, \quad i\mathcal{R}j \iff \exists k_1 = i \sim k_2 \sim \dots \sim k_\ell = j \text{ et } H(k_1) = \dots = H(k_\ell).$$

Les classes de cette relation seront notées $\alpha, \beta \dots$. On notera encore $H(\alpha)$ la valeur commune de la fonction d'énergie sur la classe α . On dira que deux classes α et β *communiquent* s'il existe une arête du graphe qui les relie (la différence $|H(\alpha) - H(\beta)|$ ne peut valoir que 1 dans ce cas). Une classe α est dite minimale pour la fonction d'énergie H si elle ne communique qu'avec des classes de niveau d'énergie strictement supérieur.

La proposition suivante décrit les ordres de grandeur des équivalents des valeurs propres en fonction des différentes puissances de ε .

Proposition 5.1 *Notons $H_0 = H$. Soit \mathcal{F}_0 l'ensemble des classes minimales de H_0 . Pour k entier positif, supposons définie une fonction d'énergie H_k sur E . Soit \mathcal{F}_k l'ensemble de ses classes minimales. La fonction H_{k+1} est définie sur E par*

$$H_{k+1}(i) = \begin{cases} H_k(i) + 1 & \text{si } \exists \alpha \in \mathcal{F}_k, i \in \alpha \\ H_k(i) & \text{sinon} \end{cases}$$

Notons K le premier indice tel que $|\mathcal{F}_k| = 1$.

Alors parmi les valeurs propres de Λ_ε , $|V| - |\mathcal{F}_0|$ convergent vers une limite strictement positive. Pour tout k entre 1 et \overline{K} , $|\mathcal{F}_{i-1}| - |\mathcal{F}_i|$ valeurs propres de Λ_ε sont équivalentes, à une constante près, à ε^i .

Les ordres de grandeur des valeurs propres se décrivent donc comme suit. A l'ordre ε^0 , les valeurs propres de Λ_0 convergent vers les valeurs propres de Λ_0 , par la théorie classique des perturbations (Kato [19]). Mais pour Λ_0 , les classes minimales (éléments de \mathcal{F}_0) sont absorbantes. La valeur propre 0 a donc pour multiplicité $|\mathcal{F}_0|$. En d'autres termes, parmi les valeurs propres de Λ_ε , $|V| - |\mathcal{F}_0|$ convergent vers une limite non nulle, les autres convergent vers 0. Déterminer les valeurs propres à l'ordre ε^1 , revient à changer l'échelle de temps, de manière à rendre possibles les transitions de taux ε . Ceci revient à augmenter de 1 exactement le niveau d'énergie de chacune des classes minimales, c'est-à-dire à remplacer H_0 par H_1 . Le nombre de classes minimales ne peut que rester constant ou diminuer. S'il est resté constant, toutes les valeurs propres "petites" sont d'ordre ε^2 au moins. S'il a diminué, la différence entre $|\mathcal{F}_0|$ et $|\mathcal{F}_1|$ est le nombre de valeurs propres d'ordre ε^1 exactement. Cette description se poursuit jusqu'à ce que toutes les classes minimales aient été fondues en une seule. Ceci arrive pour la valeur K , qui représente la hauteur minimale qu'il faut franchir dans le paysage d'énergie pour pouvoir sortir de n'importe quel puits. Toute valeur propre est donc

d'ordre ε^i où $i \leq K$. La plus petite d'entre elles, ou le trou spectral du générateur est d'ordre ε^K exactement

Cette description n'est pas nouvelle. On la déduit classiquement de la théorie de Freidlin et Wentzell [14]. Elle a été obtenue par Catoni [3], et généralisée par Trouvé [32] (voir aussi dans un cadre différent Miclo [26, 27]).

Plutôt que de répéter dans le cas particulier du recuit simulé l'expression explicite des générateurs successifs donnée par le théorème 3.1, nous nous contenterons de l'illustrer par un exemple sur un graphe ligne à 17 sommets. Voici tout d'abord la fonction d'énergie H (figure 1).

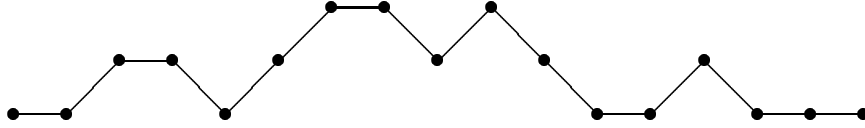


Figure 1: Fonction d'énergie sur un graphe ligne à 17 sommets.

La proposition 5.1 prévoit

- 12 valeurs propres d'ordre 1
- 3 valeurs propres d'ordre ε
- 1 valeur propre d'ordre ε^2
- 1 valeur propre égale à 0 .

Le générateur Λ_0 a

- 3 valeurs propres égales à -3
- 4 valeurs propres égales à -2
- 5 valeurs propres égales à -1
- 5 valeurs propres égales à 0.

Ceci donne les limites des valeurs propres d'ordre 1. On obtient les valeurs propres d'ordre ε à l'aide du générateur Λ_1 . Il est défini sur les classes minimales qui sont

$$\alpha_1 = \{1, 2\} , \alpha_2 = \{5\} , \alpha_3 = \{9\} , \alpha_4 = \{12, 13\} , \alpha_5 = \{15, 16, 17\}$$

Ce générateur est le suivant

$$\begin{bmatrix} -\frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 \\ \frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & -\frac{5}{6} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & -\frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{6} & -\frac{1}{6} \end{bmatrix}$$

Il a pour spectre

$$0, 0, -\frac{5}{12}, -\frac{1}{2}, -\frac{5}{6}.$$

Ceci donne les équivalents des valeurs propres d'ordre ε^1 . Il reste à calculer l'équivalent de la valeur propre d'ordre ε^2 , qui est le trou spectral. On l'obtient en contruisant la partition \mathcal{E}_2 , et le générateur Λ_2 . L'agrégation des classes α_1, α_2 d'une part et des classes α_4, α_5 d'autre part conduit à un processus à deux états. Le taux de transition de la première classe vers la seconde est $1/15$, l'autre vaut $1/25$. Le mesure réversible est proportionnelle aux poids cumulés des classes qui sont 3 et 5 respectivement. Le générateur correspondant a pour valeurs propres 0 et $-\frac{8}{75}$. Le trou spectral de Λ_ε est donc équivalent à $\frac{8}{75}\varepsilon^2$.

En résumé les coefficients des équivalents des 16 valeurs propres non nulles de Λ_ε sont

- valeurs propres d'ordre 0 : $-3, -3, -3, -2, -2, -2, -1, -1, -1, -1, -1$
- valeurs propres d'ordre 1 : $-\frac{5}{12}, -\frac{1}{2}, -\frac{5}{6}$.
- valeur propre d'ordre 2 : $-\frac{8}{75}$.

Des essais numériques ont été effectués sur cet exemple à l'aide de *mathematica*. Voici, pour 4 valeurs de ε , l'erreur relative maximale entre les valeurs propres calculées numériquement et leurs équivalents calculés ci-dessus.

$$\Delta = \max \frac{|\text{numérique} - \text{équivalent}|}{\text{numérique}}.$$

ε	10^{-1}	10^{-2}	10^{-3}	10^{-4}
Δ	0.253	0.076	0.023	0.007

Il faut remarquer ici que le cas du recuit simulé n'est vraiment particulier que au premier niveau. Dès le premier regroupement, on se retrouve pratiquement dans le cas général traité dans la section 3.

6 Dynamique de Metropolis pour le modèle de Ising

Le modèle de Ising (voir par exemple [21]) est un des modèles de physique statistique les plus étudiés par les probabilistes. La dynamique de Metropolis est l'algorithme de base qui permet de le simuler, et le problème de sa vitesse de convergence a déjà suscité une abondante littérature (voir par exemple [24]). Le problème est un cas particulier de celui du recuit simulé, étudié au paragraphe précédent.

Soit S un ensemble fini de sites, muni d'une structure de graphe U non orienté. L'espace d'états V est l'ensemble des configurations $\{0, 1\}^S$, muni de sa structure de graphe naturelle d'hypercube. Deux états (ou configurations) sont voisins s'ils diffèrent en exactement une coordonnée (ou site). La fonction d'énergie H est le nombre de paires de sites voisins (au sens de U) joignant deux sites de valeurs opposées. Cette fonction a clairement deux classes minimales qui correspondent aux deux configurations constantes $\mathbf{0}$ et $\mathbf{1}$. Ce ne sont pas les seules en général. Les techniques décrites dans la section 5 permettent d'obtenir aisément l'ordre de grandeur, et dans les cas les plus simples un équivalent, du trou spectral. Voici quelques résultats pour différents types de graphe U .

- ligne à n sommets L_n

$$Gap = \frac{2}{n}\varepsilon + o(\varepsilon) .$$

- cycle à n sommets C_n

$$Gap = \frac{n}{n+2}\varepsilon^2 + o(\varepsilon^2) .$$

- grille $L_n \times L_m$

$$Gap = O(\varepsilon^{n+m-2}) .$$

- tore $C_n \times C_m$

$$Gap = O(\varepsilon^{2n+2m-4}) .$$

- arbre de Cayley régulier de degré d de rayon r

$$Gap = O(\varepsilon^{d(d-1)^{r-1}}) .$$

References

- [1] R. Bacher et Y. Colin de Verdière. Multiplicités des valeurs propres et transformations étoile-triangle des graphes. *Bull. Soc. Math. Fr.*, 123:101–117, 1995.
- [2] D. Bertsimas and J. Tsitsiklis. Simulated annealing. *Statistical Science*, 8:10–15, 1993.

- [3] O. Catoni. Rough large deviation estimates for simulated annealing: Application to exponential schedules. *Ann. Probab.*, **20**,3:1109–1146, 1992.
- [4] G. Choquet. *Cours d'Analyse, Tome II Topologie* Masson, Paris, 1969.
- [5] E. Çinlar. *Introduction to stochastic processes*. Prentice Hall, New York, 1975.
- [6] Y. Colin de Verdière. Multiplicities of eigenvalues and tree-width of graphs. *Prépublication de l'Institut Fourier* 366:1-21, 1996.
- [7] Y. Colin de Verdière. *Spectres de graphes*. cours de DEA à l'ENS de Lyon, Polycoché de l'Institut Fourier, 1995.
- [8] P.J. Courtois. *Decomposability, queuing and computer systems applications*. Academic Press, New York, 1977.
- [9] F. Delebecque. A reduction process for perturbed Markov chains. *SIAM J. Appl. Math.*, **43**,2:325–350, 1983.
- [10] P. Diaconis and P. Hanlon. Eigenanalysis for some examples of the Metropolis algorithm. *Contemp. Math.*, 138:99–117, 1992.
- [11] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.*, 3(3):696–730, 1993.
- [12] P. Diaconis and L. Saloff-Coste. What do we know about the Metropolis algorithm. To appear in *J. Comp. Sci. Syst.*, 1996.
- [13] P. Diaconis and D.W. Strook. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.
- [14] M.I. Freidlin and A.D. Wentzell. *Random perturbations of dynamical systems*. Springer Verlag, Berlin, 1984.
- [15] A. Frigessi, F. Martinelli, and J. Sander. Computational complexity of MCMC methods. *preprint*, 1994.
- [16] A. Frigessi and P. Stefano. Convergence rates of the Gibbs sampler, the Metropolis algorithm and other single-site updating dynamics. *J.R. Statist. Soc.*, 55:205–219, 1993.
- [17] D. Geman and S. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741, 1984.
- [18] S. Ingrassia. On the rates of convergence of Metropolis algorithm and Gibbs sampler. *Ann. Appl. Probab.*, 1994.

- [19] T. Kato. *Perturbation theory for linear operators*. Springer Verlag, New York, 1978.
- [20] F.P. Kelly. *Reversibility and stochastic networks*. Wiley, London, 1979.
- [21] R. Kindermann and J.L. Snell. *Markov Random Fields and their Applications*. AMS Contemporary Math, Providence R.I., 1980.
- [22] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimisation by simulated annealing. *Science*, 220:671–680, 1983.
- [23] J. Liu. Eigenanalysis for a Metropolis sampling scheme with comparisons to rejection sampling and importance sampling. To appear, 1995.
- [24] F. Martinelli. On the two-dimensional dynamical Ising model in phase coexistence region. *J. Statist. Phys.*, 76:1179–1246, 1994.
- [25] M. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculation by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.
- [26] L. Miclo. Comportement de spectres d’opérateurs de Schrödinger à basse température. *Bull. Sc. math.*, 119:529–553, 1995.
- [27] L. Miclo. Une étude des algorithmes de recuit simulé sous-admissibles. *Ann. Fac. Sci. Toulouse*, 4(4):819–877, 1995.
- [28] Y. Pan. *Spectres de processus de Markov*. Thèse UJF, Grenoble 1997.
- [29] L. Saloff-Coste. Lectures on finite Markov chains. Ecole d’été de probabilités de Saint-Flour, 1996.
- [30] H.A. Simon and A. Ando. Aggregation of variables in dynamic systems. *Econometrica*, 29:111–138, 1961.
- [31] A. Sinclair and M. Jerrum. Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.*, 22:1087–1116, 1993.
- [32] A. Trounev. Cycle decompositions and simulated annealing. *SIAM J. Control Optimization*, **34**,3:966–986, 1996.
- [33] S. Wolfram. *Mathematica*. Wolfram Research, Inc., 1992.
- [34] B. Ycart. *Méthodes de Monte-Carlo*. Cours de DEA polycopié, Grenoble 1996.