J. OPERATOR THEORY
41(1999), 151–173

SINGULAR LIMITS OF SCHRÖDINGER OPERATORS AND MARKOV PROCESSES

Y. COLIN de VERDIÈRE, Y. PAN and B. YCART

Communicated by Nikolai K. Nikolskii

Abstract. After introducing the Γ-convergence of a family 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 and low-temperature simulated annealing are studied.

Keywords: *Schrödinger operator, perturbations, spectrum, Markov generators.*

MSC (2000): 47A55, 60J27.

1. INTRODUCTION

Schrödinger operators on a finite graph can be interpreted, up to elementary transformations, as reversible Markov generators on the set of vertices of the graph. For a general introduction one can consult [6], [7], [8], and [10]. After introducing the Γ-convergence of a family of symmetric matrices, the possible Γ-limits of a Schrödinger operator on a finite graph are studied. The main result (Theorem 2.4) states that any limit of that type can naturally be interpreted as a Schrödinger operator on another graph, the construction of which is explicitly described. The spectrum of a Γ-convergent family is proved to converge to the spectrum of the Γ-limit, thus extending the classical perturbation results of Kato ([21]).

The Markov processes to which this result is applied are almost reducible, in the sense that their generator Λ_ε depends on a parameter ε, tending to 0. For $\varepsilon = 0$, the generator is reducible and the eigenvalue 0 has multiplicity strictly larger than 1. Thus for $\varepsilon > 0$, some eigenvalues are of order $\mathrm{O}(\varepsilon)$. These "small" eigenvalues control the access to equilibrium of the process and it is important to estimate them precisely, by giving an equivalent as ε tends to 0. We propose a general method for computing the equivalents of the eigenvalues of Λ_ε.

The idea is the following. In order to observe the effect of small eigenvalues, one has to change the scale of time. This amounts to multiplying the generator Λ_ε by ε^{-1}. The new generator Γ-converges to a generator on the recurrent classes of Λ_0, the eigenvalues of which, once multiplied by ε are the equivalents of those among the eigenvalues of Λ_ε which are of order $\mathrm{O}(\varepsilon)$. This procedure can be iterated on slower time scales (ε^{-2}, $\varepsilon^{-3}, \ldots$). It creates a hierarchy of Markov processes on nested sets of classes. That hierarchy is described in Section 3, and the explicit expressions of the successive generators are given by Theorem 3.1.

As examples of applications, we shall treat two cases of almost reducible processes, the quasi-decomposable processes (Section 4) and the simulated annealing algorithm at low temperature (Section 5). Our application results are similar to those obtained in the context of discrete time Markov chains by Delebecque ([11]). They could also be obtained as an application of the perturbation theory for Markov processes of Freidlin and Wentzell ([16]) (see for instance [3] and [34]). However our technique is purely algebraic and much simpler to apply than Freidlin and Wentzell's theory.

2. THE Γ-CONVERGENCE

2.1. Definitions and spectral convergence. Let $(X, \langle \cdot \mid \cdot \rangle)$ be a Euclidean space with finite dimension n. Let A be a symmetric operator on X. The associated quadratic form is denoted by Q_A, also written as

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

Let $\Gamma(A) = \{(x, Ax) \mid x \in X\} \subset X \oplus X$ be the graph of A.

Let $Y \subset X$ be a Euclidean subspace of X and $B : Y \to Y$ a symmetric operator on Y. The pair (Y, B) is called an *unbounded operator* on X if $Y \neq X$. The definition of the graph is extended to unbounded operators in the following way.

$$\Gamma(Y, B) = \{(y, z) \in Y \oplus X \text{ such that for every } w \in Y, \langle z \mid w \rangle = \langle By \mid w \rangle\}.$$

As a particular case, $\Gamma(A) = \Gamma(X, A)$. It is clear that $\Gamma(Y, B)$ is still an n-dimensional subspace of $X \oplus X$.

Let ω be the symplectic form defined by

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

An n-dimensional subspace of $X \oplus X$, which is isotropic for the symplectic form ω is said to be *Lagrangian.* It is easy to check that any Lagrangian subspace is the graph of a unique symmetric operator, bounded or not. Conversely, any graph $\Gamma(Y, B)$ is Lagrangian. The set of symmetric operators with domain on X can thus be seen as a compactification of the set of symmetric endomorphisms of X, which makes it a C^∞ compact manifold.

In order to define the spectrum, we need the notion of invertibility. The operator (Y, B) is said to be *invertible* if the space $\sigma(\Gamma(Y, B))$, where $\sigma(x, y) = (y, x)$, is the graph of a bounded mapping from X into $Y \subset X$.

For $\lambda \in \mathbb{C}$, define the operator

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

by its graph

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

This definition yields the two notions of resolvent (inverse of $\lambda I - B$) and spectrum (set of singular values of λ). Thus the spectrum of (Y, B) is the union of the spectrum of B and of the eigenvalue ∞, repeated $(n - \dim Y)$ times.

DEFINITION 2.1. A family A_ε, $\varepsilon > 0$ of symmetric operators on X is said to Γ-*converge* to (Y, B) if $\Gamma(A_\varepsilon)$ converges to $\Gamma(Y, B)$ as $\varepsilon \to 0$, in the sense of the natural topology of the Grassmannian of n-dimensional subspaces of $X \oplus X$.

The notation is

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

The following result extends to the case of Γ-convergence the classical results of Kato ([21]) on the spectral convergence of perturbed quadratic forms. It will justify in Section 3 our use of Γ-convergence to study the spectrum of almost reducible Markov processes.

THEOREM 2.2. *Assume $A_\varepsilon \xrightarrow{\Gamma} (Y, B)$. Then the spectrum of A_ε converges to that of (Y, B) in the sense of the topology of $P^1(\mathbb{R}) = \mathbb{R} \cup \infty$.*

Proof. It suffices to notice that the Γ-convergence is equivalent to the convergence of the resolvents $(\lambda I - A_\varepsilon)$, for all $\lambda \notin \mathbb{R}$, as bounded operators from X into X. ∎

The existence of Γ-limits is not uncommon.

Proposition 2.3. *If* $A_\varepsilon = \sum_{i=0}^{N} A_i \varepsilon^{-i}$*, where the* A_i *are symmetric, then the family* A_ε Γ*-converges.*

If $N = 1$, the Γ-limit is easily identified (see for instance Proposition 4.1). The general case is much less simple (see Section 3, and [1]).

Proposition 2.3 can be generalized as follows.

Theorem 2.4. *Let* $A_i : X \to X$ *be symmetric operators,* $\alpha_1 < \cdots < \alpha_N$ *a sequence of reals. For* $\varepsilon > 0$*, define*

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

Let Z *be the Grassmannian of Lagrangian* n*-dimensional subspaces of* $X \oplus X$*, and* $C_\varepsilon \in Z$ *the graph of* A_ε*. Then the mapping* $\varepsilon \to C_\varepsilon$ *from* $\mathbb{R}^+ \setminus 0$ *into* Z *can be continuously extended at* $\varepsilon = 0$*. Moreover, if the* α_i *are integers, the extension is analytic in* ε *real.*

Proof. Let $\lambda_0 \notin \mathbb{R}$. The resolvent

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

is bounded above in norm by $1/|\mathrm{Im}\,\lambda_0|$ and can be written as an asymptotic series of the form

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

with non null B_j's. This comes from explicit formulae for the inverse of a matrix. Notice that the ρ_j's are necessarily $\geqslant 0$. ∎

In many cases of practical interest, Γ-convergence holds, and the Γ-limit will be characterized by its domain Y and the coefficients of the matrix B.

2.2. Criteria for Γ-convergence. From now on, symmetric operators are always assumed to be nonnegative.

Definition 2.5. Let (x_ε) be a family of elements of X. We write $x_\varepsilon \xrightarrow{\mathrm{s}} x_0$ if x_ε converges to x_0 with $Q_{A_\varepsilon}(x_\varepsilon) = \mathrm{O}(1)$.

Lemma 2.6. *If* $A_\varepsilon \xrightarrow{\Gamma} (Y, B)$ *then* $x_0 \in Y$ *iff there exists* x_ε *such that* $x_\varepsilon \xrightarrow{\mathrm{s}} x_0$.

Proof. Let $X = Y_\varepsilon \oplus Z_\varepsilon$ be the decomposition of X into the sum Y_ε of eigenspaces associated to eigenvalues of order $\mathrm{O}(1)$, and its orthogonal. Let us write the decomposition of x_ε as $x_\varepsilon = y_\varepsilon + z_\varepsilon$. It can be easily seen that $z_\varepsilon \to 0$ and $\|A_\varepsilon y_\varepsilon\| = \mathrm{O}(Q_{A_\varepsilon}(x_\varepsilon)) = \mathrm{O}(1)$. The proof is completed by extracting a subsequence. ∎

LEMMA 2.7. *Let Y_ε be a family of p-dimensional subspaces of X such that Y_ε tends to Y. Assume that for $\varepsilon \neq 0$, the norm of the restriction of A_ε to Y_ε remains uniformly bounded. Let ϕ_ε be an isomorphism from Y into Y_ε such that for all $y \in Y$, when $\varepsilon \to 0$, $\phi_\varepsilon(y)$ tends to y. Define*

$$Q_\varepsilon^\phi(y) = \langle \phi_\varepsilon(y) \mid A_\varepsilon \mid \phi_\varepsilon(y) \rangle.$$

Then Q_ε^ϕ admits a limit denoted by Q_0^ϕ as a quadratic form on Y and if B denotes the associated operator, one has:

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

Notice that the boundedness hypothesis for the restriction of A_ε implies that $\phi_\varepsilon(y) \xrightarrow{\mathrm{s}} y$.

A practical version of this lemma is the following:

Let (y_i) be a base of Y. For each y_i choose a family $y_{i,\varepsilon}$ converging to y_i with $A_\varepsilon y_{i,\varepsilon} = \mathrm{O}(1)$. Then the matrix of the limit B is given by

$$b_{i,j} = \langle y_i \mid B \mid y_j \rangle = \lim \langle y_{i,\varepsilon} \mid A_\varepsilon \mid y_{j,\varepsilon} \rangle.$$

Beware that Q_ε^ϕ is not the restriction to Y of Q_{A_ε}, that could have an incorrect limit or no limit at all. Here is a concrete example.

EXAMPLE 2.8. Consider $X = \mathbb{R}^4$, with the quadratic form

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

Take $p = 3$ and $Y_\varepsilon = \{y_0 = (\varepsilon/3)(y_1 + y_2 + y_3)\}$. Then $Q_0(y) = \frac{1}{3} \sum_{i=1}^{3} (y_i - y_{i+1})^2$ and not $\sum_{i=1}^{3} y_i^2$.

Proof. First case: Let us suppose that Y_ε does not depend on ε thus being equal to Y. Let us decompose X into $X = Y \oplus Z$ (orthogonal sum) and write A_ε as a matrix according to that decomposition:

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

It comes from the hypotheses and the minimax that $D_\varepsilon^{-1} = \mathrm{o}(1)$ whereas K_ε and C_ε are uniformly bounded.

The graph of A_ε is the set of couples $((y,z),(y',z'))$ such that:

$$z = D_\varepsilon^{-1}(z' - C_\varepsilon^{\mathrm{t}} y)$$
$$y' = K_\varepsilon y + C_\varepsilon z.$$

The existence of a limit in the sense of Γ-convergence shows that K_ε has a limit K_0 and that the limit of graphs is the set of pairs (z,y') such that $z=0, y' = K_0 y$, which is indeed the conclusion of the theorem in this case.

Second case: Let U_ε be a continuous family of isometries of X such that U_0 is the identity and $U_\varepsilon(Y) = Y_\varepsilon$. Setting $\widetilde{A}_\varepsilon = U_\varepsilon^{\mathrm{t}} A_\varepsilon U_\varepsilon$ takes us back to the first case. Thus the Γ-limit of $\widetilde{A}_\varepsilon$ is the same as that of A_ε. The limit quadratic form is the limit on Y of

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

The same limit is obtained by replacing in the previous formula $U_\varepsilon y$ by vectors y_ε satisfying $\|U_\varepsilon y - y_\varepsilon\| = \mathrm{o}(1)$ whereas A_ε stays uniformly bounded. ∎

We shall end this section by the particular case of an increasing sequence of nonnegative quadratic forms, for which the notion of Γ-convergence is equivalent to pointwise convergence.

Theorem 2.9. *Let A_n be an increasing sequence of nonnegative symmetric operators defined on X and $Q_n(x) = \langle x \mid A_n \mid x\rangle$ the sequence of associated quadratic forms. Let*

$$Q_\infty : X \to \mathbb{R} \cup +\infty$$

be the pointwise limit of Q_n and

$$Y = \{x \in X \text{ such that } Q_\infty(x) < \infty\}.$$

Let B be defined on Y by $Q_\infty(y) = \langle y \mid B \mid y\rangle$. Then

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

The proof of Theorem 2.9 uses the following two lemmas.

Lemma 2.10. *Let (Y,B) be a (possibly unbounded) operator, then $(x_0,z_0) \in \Gamma(Y,B)$ if and only if x_0 is a critical point of the mapping F_{z_0} defined on Y by*

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

This lemma is obvious. In particular, if K is a compact subset of X and B is nonnegative, the mapping F_{z_0} is minimal at a point x_0 in the interior of K if and only if $(x_0,z_0) \in \Gamma(Y,B)$.

LEMMA 2.11. *Let K be a compact subset of X and f_n be a sequence of lower semi-continuous* (lsc) *functions defined on K, with values in $[0,+\infty]$, such that f_n converges pointwise to a function f_∞. Assume moreover that there exist* lsc *functions on K, $g_n \leqslant f_n$ such that the sequence g_n is increasing and has the same pointwise limit as the sequence f_n. Denote by m_n and m_∞ the minimal values of f_n and f_∞ over K. Let x_n be such that $f_n(x_n) = m_n$ and assume that the sequence (x_n) converges to a. Then $m_\infty = \lim m_n$ and $f_\infty(a) = m_\infty$.*

Proof. (of Lemma 2.11.) (1) The function f_∞, being an increasing limit of lsc functions, is lsc as well. Thus there exists $x_\infty \in K$ such that $f_\infty(x_\infty) = m_\infty$. Obviously one has $\limsup m_n \leqslant m_\infty$. Assume that $\liminf m_n = m_\infty - \alpha$, for some positive α. Let

$$K_n = \left\{ x \in K \text{ such that } g_n(x) \leqslant m_\infty - \frac{\alpha}{2} \right\}.$$

It is clear that the sequence of compact sets K_n is decreasing, with empty intersection. So for some $n_0 > 0$, $K_{n_0} = \emptyset$ and thus $g_n(x) > m_\infty - \alpha/2$ for $n \geqslant n_0$ and $x \in K$. This leads to a contradiction since $f_n \geqslant g_n$.

(2) (a) If $f_n(x_n) = m_n$ and $x_n \to a$, then $m_n = f_n(x_n) \leqslant f_n(a)$. Taking limits in both sides:

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

(2) (b) Suppose now that $f_\infty(a) = m_\infty + \beta$ with $\beta > 0$. Let U be a compact neighborhood of a over which $f_\infty \geqslant m_\infty + \beta/2$, for some positive β. Let us apply (1) to restrictions to U. For n large enough, $\inf_U f_n = m_n$ and $\inf_U f_\infty \geqslant m_\infty + \beta/2$. Hence the contradiction. ∎

Proof. (of Theorem 2.9.) Assume that the quadratic form Q_∞ associated to (B, Y) is the pointwise limit of q_n. Let $(x_n, z_n) \in \Gamma(A_n)$ and assume that (x_n, z_n) tends to (a, b). It is enough to show that $(a, b) \in \Gamma(Y, B)$. This is done by applying Lemma 2.11 to the sequence $q_n - 2\langle z_n \mid \cdot \rangle$, restricted to some compact subset K of X, containing a as an interior point. The conclusion is given by Lemma 2.10. ∎

2.3. THE MAIN RESULT. Let $G = (V, E)$ be a finite graph. The vector space $\mathbb{R}^V$ is endowed with the canonical Euclidean structure. Schrödinger operators on graphs are defined by analogy with the continuous case (see [10] or the introduction of [7]).

Definition 2.12. A symmetric operator $A : \mathbb{R}^V \to \mathbb{R}^V$ is a *Schrödinger operator* on G (notation $A \in \mathrm{O}_G$) if the following conditions are satisfied:

(i) if $\{i,j\} \in E$, then $a_{i,j} < 0$;
(ii) if $\{i,j\} \notin E$ and $i \neq j$, then $a_{i,j} = 0$;
(iii) the operator A is nonnegative.

We are going to show that any Γ-limit of a family of Schrödinger operators on the graph G can naturally be identified to a Schrödinger operator on another graph G', where G' is a *weak-minor* of G in the following sense.

Definition 2.13. The graph $G' = (V' = \{1,\dots,p\}, E')$ is said to be a *weak-minor* of $G = (V,E)$ if and only if there exist G-connected disjoint subsets $V_1,\dots,V_p$ of V such that ($\{i,j\} \in E'$) implies that there exists a path $\gamma_{i,j}$ in G joining some vertex $a \in V_i$ to another vertex $b \in V_j$ without meeting any of the other V_k's ($k \neq i,j$) (the paths $\gamma_{i,j}$ are not requested to be disjoint).

We recall below the classical definition of a *minor*.

Definition 2.14. The graph $G' = (V' = \{1,\dots,p\}, E')$ is said to be a *minor* of $G = (V,E)$ if and only if there exist G-connected disjoint subsets $V_1,\dots,V_p$ of V such that ($\{i,j\} \in E'$) implies that there exists a vertex $a \in V_i$ and a vertex $b \in V_j$ such that $\{a,b\}$ is an edge of G.

The precise statement is the following:

Theorem 2.15. *Let $A_\varepsilon \in \mathrm{O}_G$. Assume that $A_\varepsilon \xrightarrow{\Gamma} (Y,B)$ with $p = \dim Y$. Then there exists a weak-minor $G' = (V',E')$ of G and an isometry J from $\mathbb{R}^{V'}$ into Y such that $J^{-1}BJ \in \mathrm{O}_{G'}$.*

The isometry J is described as follows: there exists a partition (determined by A_ε)

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

where the V_i, $i \geqslant 1$ are connected, such that if $F_i = \{\varphi \in Y$ such that $\mathrm{supp}(\varphi) \subset V_i\}$, then F_i, $i \geqslant 1$ have dimension 1 and $Y = \bigoplus_{i=1}^{p} F_i$. Each F_i is spanned by a unique function $\varphi_i \geqslant 0$ with norm 1. These φ_i are an orthonormal basis of Y. If $x = (x_i) \in \mathbb{R}^{V'}$, then $J(x) = \sum_1^p x_i \varphi_i$.

In the context of reversible Markov processes (Section 3), the partition $W_0, V_1, \dots, V_p$ has a natural interpretation in terms of recurrent and transient classes. The Γ-limit (Y,B) will be understood as a new Markov generator on a set of recurrent classes. An explicit expression for the matrix B will be given in Theorem 3.2.

Proof. The space Y. Lemma 2.6 implies that, if $\varphi \in Y$, then $|\varphi| \in Y$. Indeed,

$$Q_\varepsilon(|\varphi|) \leqslant Q_\varepsilon(\varphi).$$

Let $Z \subset V$ be the set of vertices i such that there exists $\varphi \in Y$ with $\varphi(i) \neq 0$. For $i \in Z$, let $\epsilon_i \in Y' \backslash 0$ be the linear form defined by $\epsilon_i(\varphi) = \varphi(i)$. Let us introduce the equivalence relation in Z, for which i and j are related if ϵ_i and ϵ_j are linearly dependent. Let $W_1, \ldots, W_q$ be the equivalence classes and choose $a_i \in W_i$. Then the mapping from Y into $\mathbb{R}^q$ which associates φ to $(\varphi(a_i))$ is an isomorphism. Indeed, it is obviously injective. To prove surjectivity, one can apply for instance Lemma 12.1, p. 141 of [4].

The previous argument also shows that the φ_i can be chosen to be strictly positive over W_i.

The V_i's. The W_i's are disjoint, but not necessarily connected. The sets V_i are such that $W_i \subset V_i$. The set V_i is the set of vertices $a \in V$ such that does not exist $\varphi_\varepsilon \xrightarrow{\mathrm{s}} \varphi_i$ with $\varphi_\varepsilon(a) = 0$.

Thus there exists $\varphi_\varepsilon \xrightarrow{\mathrm{s}} \varphi_i$, with $\forall\, a \notin V_i$, $\varphi_\varepsilon(a) = 0$, and V_i is minimal for that property. Indeed, it suffices to take, for each $a \notin V_i$, $\varphi_{a,\varepsilon} \xrightarrow{\mathrm{s}} \varphi$ with $\varphi_{a,\varepsilon}(a) = 0$ and $\varphi_{a,\varepsilon} \geqslant 0$; then $\varphi_\varepsilon = \inf \varphi_{a,\varepsilon}$.

The V_i's are connected. Indeed if $V_i = V_i' \cup V_i''$ with $V_i' \cap V_i'' = \emptyset$ and $W_i \cap V_i' \neq \emptyset$, one can replace φ_ε by its restriction to V_i' thus contradicting minimality.

Let $a \in V_i \cap V_j$ and assume one can find $0 \leqslant \varphi_{i,\varepsilon} \xrightarrow{\mathrm{s}} \varphi_i$, $0 \leqslant \varphi_{j,\varepsilon} \xrightarrow{\mathrm{s}} \varphi_j$ and that $\varphi_{i,\varepsilon}(a)/\varphi_{j,\varepsilon}(a) \to l$, $0 \leqslant l < \infty$.

One can then find l_ε such that $\Phi_\varepsilon(a) = \varphi_{i,\varepsilon}(a) - l_\varepsilon \varphi_{j,\varepsilon}(a) = 0$ and $\Phi_\varepsilon \xrightarrow{\mathrm{s}} \varphi_i - l\varphi_j$. Then one chooses $(\Phi_\varepsilon)_+ \xrightarrow{\mathrm{s}} \varphi_i$ and null at a.

If l_ε does not exist, it suffices to swap the roles of i and j.

The graph G'. Let us choose $a_i \in W_i$. Let $m_i = \varphi_i(a_i) > 0$ and $\Phi_{i,\varepsilon}$ be such that $\Phi_{i,\varepsilon}(a_j) = \delta_{i,j} m_i$ and $A_\varepsilon \Phi_{i,\varepsilon}(x) = 0$, $\forall\, x \neq a_1, \ldots, a_p$. This amounts to minimizing $Q_\varepsilon(\Phi)$ under constraint of fixed values at the a_j's. The $\Phi_{i,\varepsilon}$ are everywhere nonnegative, because replacing them by their absolute values conserves the constraints and diminishes the quadratic form.

Now the $\Phi_{i,\varepsilon}$ satisfy the hypotheses of Lemma 2.7. Indeed, if $\psi_{i,\varepsilon} \xrightarrow{\mathrm{s}} \varphi_i$ are such that $\psi_{i,\varepsilon}(a_j) = \delta_{i,j} m_i$, one has:

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

Hence $\Phi_{i,\varepsilon} \xrightarrow{\mathrm{s}} \varphi_i$, because the $\Phi_{i,\varepsilon}$'s have the same values at vertices a_j than the $\psi_{i,\varepsilon}$'s. Next $A_\varepsilon \Phi_{i,\varepsilon}(x) = 0$ if $x \neq a_1, \ldots, a_p$, and

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

by Cauchy-Schwarz inequality.

Thus the elements $b_{i,j}$ of the limit matrix B are given by

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

The mapping $\Phi_{i,\varepsilon}$ being nonnegative in a neighborhood of a_j, this implies that for $i \neq j$, $b_{i,j} \leqslant 0$.

Let us show that, if there does not exist any path γ from V_i to V_j not entering any of the other V_l's, then $b_{i,j} = 0$.

The first step is to show that, for $a \in V_k$, $\Phi_{i,\varepsilon}(a) = \mathrm{o}(\Phi_{k,\varepsilon}(a))$.

If this was false, after possibly extracting a subsequence, one would get $\Phi_{k,\varepsilon}(a) - c_\varepsilon \Phi_{i,\varepsilon}(a) = 0$ for some bounded c_ε. This would yield

$$|\Phi_{k,\varepsilon} - c_\varepsilon \Phi_{i,\varepsilon}| \xrightarrow{\ \mathrm{s}\ } \varphi_k + |c|\varphi_i,$$

hence

$$\min\{|\Phi_{k,\varepsilon} - c_\varepsilon \Phi_{i,\varepsilon}|, |c_\varepsilon \Phi_{i,\varepsilon}|\} \xrightarrow{\ \mathrm{s}\ } \varphi_k.$$

This proves that $a \notin V_k$.

Therefore,

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

for all $a \in \bigcup_{k\neq i} V_k$. This yields the same estimate over the set of those a's which are not joined by a path to V_i, without meeting any of the V_l's (positivity of the solutions to the Dirichlet problem, with positive boundary conditions). In particular, this is true for the neighbors of a_j. This yields

$$\langle \Phi_{i,\varepsilon} \mid A_\varepsilon \mid \Phi_{j,\varepsilon}\rangle = \mathrm{o}\Big(\sum_{k\neq i} \langle \Phi_{k,\varepsilon} \mid A_\varepsilon \mid \Phi_{j,\varepsilon}\rangle\Big),$$

and thus $b_{i,j} = 0$. ∎

In general, the graph G' is not a minor of G, as was observed in Example 2.8 which corresponds to a star-triangle transformation (see also [1]). It is interesting to give sufficient conditions on the family A_ε that ensure G' to be a minor of G. One such condition is $A_\varepsilon \mathbf{1}_V = 0$, since then $W_0 = \emptyset$. We shall meet again this particular case with quasi-decomposable processes (Section 4). Another sufficient condition is a consequence of Theorem 2.9.

Proposition 2.16. *Let A_n be an increasing sequence of operators of O_G, and $A_\infty \in O_{G'}$ its Γ-limit. Then G' is a minor of G.*

Proof. In that case, according to Theorem 2.9, it suffices to consider pointwise convergence on Y. This implies that the W_i are already connected, and that $\{i, j\}$ can be an edge of G' only if there exist vertices $a_i \in W_i$ and $a_j \in W_j$ such that $\{a_i, a_j\}$ is an edge of G. ∎

3. REVERSIBLE MARKOV PROCESSES

3.1. DEFINITIONS. Let $\Lambda = (\lambda_{ij})_{i,j\in V}$ be the infinitesimal generator of a Markov process on a finite state space V (cf. Çinclar ([5]) for a general reference). It is a square matrix whose nondiagonal elements are nonnegative, the sum of elements of each line being null.

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

Let p be a strictly positive probability measure on V. The generator Λ is said to be p-reversible, if the following *detailed balance condition* is satisfied (see for instance [22]).

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

Thus a generator is p-reversible if and only if it is self-adjoint in $L_2(V,p)$. To such a generator, one can associate its *reduced graph* $G = (V,E)$ whose set of vertices is V and set of edges is

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

The generator is irreducible if its reduced graph is connected. In that case, the eigenvalue 0 is simple.

Let D be the diagonal matrix

$$D = \operatorname{diag}\Big(\big(\sqrt{p(i)},\, i \in V\big)\Big),$$

and D^{-1} its inverse. The generator Λ is p-reversible iff the matrix $D\Lambda D^{-1}$ is symmetric. Thus $-D\Lambda D^{-1}$ is the matrix of a Schrödinger operator on the graph G, in the sense of Definition 2.12. Its only additional property, compared to Definition 2.12, is to admit 0 as an eigenvalue, associated to the eigenvector $D\mathbf{1}_V$. Here $\mathbf{1}_V$ is the constant vector with coordinates equal to 1, indexed by V. The eigenvalues of a reversible generator are all real, non positive. If Λ is p-reversible, its eigenvalues will be ranked in decreasing order, and we shall call *spectrum* of Λ the vector of its ordered eigenvalues. With that convention, the first coordinate of the spectrum is 0, and the opposite of the second one is the *spectral gap* of Λ, denoted by $\operatorname{gap}(\Lambda)$.

The speed of access to equilibrium of a Markov process with generator Λ is usually studied through its Dirichlet form (see [13], [15], [17], [31] and [33]):

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

Our point of view is to use instead the quadratic form Q, with matrix $-D\Lambda D^{-1}$:

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

We recall here the classical result on the conservation of reversibility upon truncation (Corollary 1.10, p. 26 of Kelly ([22])).

Proposition 3.1. *Let Λ be a p-reversible irreducible generator on V, and α a subset of V. Define the transition rates λ^α_{ij} by*

$$\lambda^\alpha_{ij} = \begin{cases} \lambda_{ij} & \text{if } i,j \in \alpha, \\ 0 & \text{else.} \end{cases}$$

Let Λ^α be the generator corresponding to the rates $\lambda^\alpha_{i,j}$ and assume it is irreducible on α. Then Λ^α is p_α-reversible, where p_α is the conditional distribution of p over α:

$$p_\alpha(i) = \begin{cases} \frac{p(i)}{p(\alpha)} & \text{if } i \in \alpha, \\ 0 & \text{else.} \end{cases}$$

3.2. Analytic perturbations. From now on, ε is a real positive parameter and Λ_ε is a p_ε-reversible generator, irreducible over V, such that the transition rates λ^ε_{ij} depend analytically on ε in a neighborhood of 0. It is easy to deduce from (3.1) that the measure p_ε is analytic in a neighborhood of 0 as well. The generators we want to study are almost reducible, in the sense that the limit of Λ_ε, denoted by Λ_0, is reducible. From the classical perturbation theory for symmetric operators (cf. [21]), the eigenvalues of Λ_ε as well as the associated eigenspaces converge to the eigenvalues and eigenspaces of Λ_0. In other words, the spectrum of Λ_0 gives an equivalent of the eigenvalues of order $\mathrm{O}(1)$ for Λ_ε. But Λ_0 being reducible, the eigenvalue 0 has a multiplicity strictly larger than 1. This means that some eigenvalues of Λ_ε are of order $\mathrm{O}(\varepsilon)$. We propose a general method for computing explicitly the equivalents of the eigenvalues of order $\mathrm{O}(\varepsilon^k)$ for all $k \geqslant 1$. The idea is the following. The eigenvalues of order $\mathrm{O}(\varepsilon^k)$ correspond to movements of the process on the scale of time ε^{-k}. In order to compute their equivalents, one has to determine the Γ-limit of $\varepsilon^{-k}\Lambda_\varepsilon$, the spectrum of which will give the coefficients of the desired equivalents, by Theorem 2.2. The explicit expression of this Γ-limit, and its probabilistic interpretation, are given in Theorem 3.2.

Denote by S_k the vector whose coordinates are the eigenvalues of order $\mathrm{O}(\varepsilon^k)$ of the generator Λ_ε, with their multiplicity. The Γ-limit of $\varepsilon^{-k}\Lambda_\varepsilon$ is a generator Λ_k on a state space of size $\#S_k$, coming from a partition of the state space V, as in Theorem 2.15. This partition has a concrete interpretation in the dynamical language of Markov processes.

Here is the interpretation for $k = 1$. The *irreducible classes* of Λ_0 are maximal subsets of states such that inside each of these subsets, any state is linked to any other by a path of positive transitions. An irreducible class is *transient* if there exists a positive transition coming out of it, otherwise it is called *recurrent* (see for instance [5] for these classical definitions). Denote by $\mathcal{E}_1$ the set of irreducible classes of Λ_0 and by $\mathcal{F}_1 \subset \mathcal{E}_1$ the set of recurrent classes. These classes can be observed on the dynamics of the process with generator Λ_ε. For ε small enough, starting in a transient class (element of $\mathcal{E}_1 \setminus \mathcal{F}_1$), the process comes out of it, and goes to a recurrent class in a time $\mathrm{O}(1)$ on average. In the same time scale, starting in a recurrent class (element of $\mathcal{F}_1$), the process can reach any state in that same class. But the departure from any recurrent class will not occur before a time $\mathrm{O}(1/\varepsilon)$ on average. Thus over a time scale of order $\mathrm{O}(1)$ the process reaches its equilibrium inside any recurrent class, but no communication is seen between these classes. Changing the scale of time by a factor ε, i.e. over time intervals of length $\mathrm{O}(1/\varepsilon)$, the process has reached its equilibrium inside recurrent classes almost immediately, compared to the time scale. So that one can ignore the way equilibrium has been reached, and identify all states of a given recurrent class, by aggregating them into a single state. The recurrent classes are the sets $V_1, \dots, V_p$ of Theorem 2.15. The set W_0 is the union of all transient classes.

By iterating this procedure, one can define a partition $\mathcal{E}_k$ of the state space V, and a subset $\mathcal{F}_k$ of $\mathcal{E}_k$ with the following properties. In a time interval of order $\mathrm{O}(1/\varepsilon^{k-1})$ on average, starting from a class of $\mathcal{E}_k$ the process can reach any state of the same class, and go out to a class of $\mathcal{F}_k$ if the initial one was in $\mathcal{E}_k \setminus \mathcal{F}_k$. However, it will take a time $\mathrm{O}(1/\varepsilon^k)$ on average to get out of a class in $\mathcal{F}_k$. Then the Γ-limit of $\varepsilon^{-k}\Lambda_\varepsilon$ is a Markov generator on $\mathcal{F}_k$. The equivalents of S_k are the product of the eigenvalues of this generator by ε^k.

The recursive construction of the partitions $\mathcal{E}_k$ and $\mathcal{F}_k$ has been detailed in [30] and we shall not reproduce it here. It is not new. It can be seen as a cycle decomposition in the sense of Freidlin and Wentzell ([16]). It has been obtained by different means, for discrete time Markov chains, with an analogous interpretation in terms of time scales, by Delebecque ([11]).

The partitions $\mathcal{E}_k$ are nested, and $\#\mathcal{F}_k$ is decreasing. Let K be the smallest integer such that

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

The eigenvalues of order ε^{K-1} are the smallest in the spectrum, and ε^{K-1} is the order of gap (Λ_ε).

Let

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

and denote by $\Lambda_{U_{k+1}}$ the restriction of Λ_ε over U_{k+1}. For any class α_k of $\mathcal{F}_k$, denote by p_{α_k} the reversible measure for the restriction of Λ_ε to α_k. It is the conditional distribution of p_ε over α_k, according to Proposition 3.1. For all subsets S_1, S_2 of V with $S_1 \cap S_2 = \emptyset$, denote by

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

the matrix composed of the transition rates from S_1 to S_2. Denote by $D_{U_{k+1}}$ the diagonal matrix such that

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

where $\mathbf{1}_{U_{k+1}}$ and $\mathbf{1}_{\alpha_k}$ stand for constant vectors equal to 1 over U_{k+1} and α_k respectively.

Our main result gives the explicit expression of the Γ-limit of $\varepsilon^{-k}\Lambda_\varepsilon$.

THEOREM 3.2. *Let k be an integer between 1 and $K-1$.*

(i) *For all $\alpha_k \neq \beta_k \in \mathcal{F}_k$, define the transition rate from α_k to β_k as follows:*

$$\lambda_{\alpha_k \beta_k} = \varepsilon^{-k} p_{\alpha_k} \big[G_{\alpha_k}^{\beta_k} \mathbf{1}_{\beta_k} - G_{\alpha_k}^{U_{k+1}} (\Lambda_{U_{k+1}} - D_{U_{k+1}})^{-1} G_{U_{k+1}}^{\beta_k} \mathbf{1}_{\beta_k} \big].$$

Then $\lim_{\varepsilon \searrow 0} \lambda_{\alpha_k \beta_k}$ exists. Let $l \geqslant k+1$ be the smallest integer such that α_k is included in an element of $\mathcal{F}_l$, denoted by α_l. Then if $\beta_k \not\subseteq \alpha_l$ one has

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

(ii) *For all a in U_{k+1}, denote by $p_a^{\beta_k}$ the probability for the process with generator Λ_ε, starting from a, to reach β_k for the first time without visiting any other class of $\mathcal{F}_k$. Let $P_{U_{k+1}}^{\beta_k}$ be the column vector made of the $p_a^{\beta_k}$'s for a in U_{k+1}. Then*

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

(iii) *Let Λ_k be the generator on $\mathcal{F}_k$ made of the transition rates $\lambda_{\alpha_k \beta_k}$. Then up to a permutation*

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

In (i), the invertibility of $\Lambda_{U_{k+1}} - D_{U_{k+1}}$ comes from positivity properties of the Markov generator Λ_ε. A direct proof of this theorem due to Y. Pan can be found in [30]. It can also be deduced from the results of Delebecque ([11]). In the next two sections, we shall apply Theorem 3.2 to particular cases of almost reducible Markov processes which are of practical interest.

4. QUASI-DECOMPOSABLE PROCESSES

The notion of quasi-decomposability seems to have been introduced by the economists Simon and Ando ([32]) in the framework of hierarchical analysis for complex systems. The idea, quite natural, has been applied in sociology, biology, and computer science (for a history of the subject, see the introduction of Courtois ([9])). This idea is the following. In a complex dynamical system, there usually coexist several scales of time. As an example, let us speak of seconds, hours and months. At the scale of a few seconds, the system can be decomposed into isolated sub-systems, having only internal interactions. Interactions between these components occur only at the scale of several hours. This induces an aggregation of the former components into bigger ones. Some of these big components will interact at the scale of one month, thus inducing a coarser partition of the global system. One obtains a hierarchy of nested partitions, each of them corresponding to a given time scale.

We shall now describe a formal setting in terms of continuous time Markov processes (for the discrete time analogous, see the first chapter of Courtois ([9])).

Let Λ be an irreducible generator on $V = \{i, j, \dots\}$, corresponding to transition rates $(\lambda_{ij})_{i,j \in V}$. The generator Λ is assumed to be p-reversible, where $p = (p(i))_{i \in V}$. Consider a partition of V, $\mathcal{F} = \{\alpha, \beta, \dots\}$.

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

Denote by

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

the measure of the set α. The conditional measure of p over α is still denoted by p_α:

$$p_\alpha(i) = \frac{p(i)}{q(\alpha)}, \quad \forall\, i \in \alpha.$$

For $\varepsilon > 0$, we define the generator Λ_ε multiplying by ε the transition rates between different classes of the partition $\mathcal{F}$. This amounts to slowing down by a factor ε the exchanges between different classes, or else to distinguish the time scale $1/\varepsilon$ of exchanges between classes from the time scale 1 which remains that of exchanges inside the classes:

$$\begin{aligned} \lambda^\varepsilon_{ij} &= \lambda_{ij}, \quad \forall\, \alpha \in \mathcal{F}, \forall\, i \neq j \in \alpha \\ \lambda^\varepsilon_{ij} &= \varepsilon \lambda_{ij}, \quad \forall\, \alpha \neq \beta \in \mathcal{F}, \forall\, i \in \alpha,\, j \in \beta. \end{aligned}$$

The results of the previous section are easily applied in this particular case. With the notation of Section 2, the set W_0 is empty, and the sets V_k are the classes of partition $\mathcal{F}$. As ε tends to zero, the spectrum of Λ_ε, denoted by s_ε, converges to that of Λ_0, which is a concatenation of the spectrums of the restrictions of Λ to each class of the partition.

Among the eigenvalues of Λ_ε, $\#\mathcal{F}$ converge to 0 (one of them remains null). Denote by s'_ε their vector ($\#\mathcal{F}$ first coordinates of s_ε). A direct application of Theorem 3.2 shows that $s'_\varepsilon/\varepsilon$ converges to the spectrum of a Markov generator M on $\mathcal{F}$.

Proposition 4.1. *Let M be the generator defined on $\mathcal{F}$ by the following transition rates:*

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

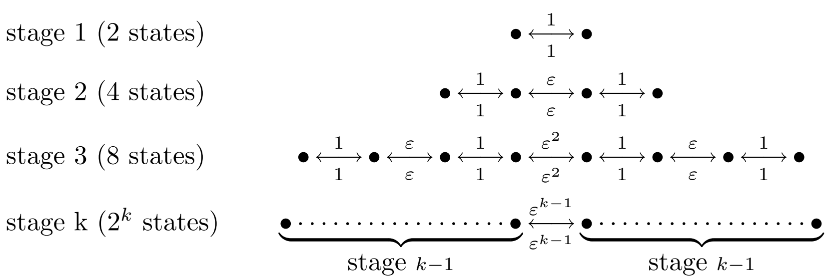
Then M is the Γ-limit of $\varepsilon^{-1}\Lambda_\varepsilon$. Let t be its spectrum; then

$$\lim_{\varepsilon\to 0} \frac{s'_\varepsilon}{\varepsilon} = t.$$

It is easy to check that the generator M is q-reversible. The rate $\mu_{\alpha\beta}$ can be interpreted as the *stationary exchange flux* from α to β for the process with generator Λ.

The above result can obviously be extended to other time scales. To do this, one has to consider coarser partitions, for which the transition rates between distinct classes are of higher order. It is not necessary to detail the full construction here, since it is a particular case of Theorem 3.2, essentially contained already in Proposition 4.1.

As an example, we describe below a family of birth and death processes (cf. [5] for a general reference) on $\{1, \ldots, 2^k\}$, with a hierarchy of time scales. The family is constructed iteratively according to the following transition diagrams.



More precisely, let Λ_2 be the symmetric generator on $\{1,2\}$,

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

For all $k \geqslant 1$, if Λ_{2^k} has been defined on $\{1,\dots,2^k\}$, define the generator $\Lambda_{2^{k+1}}$ on $\{1,\dots,2^{k+1}\}$ by

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

where $B_{2^{k+1}}$ is the generator whose transition rates between 2^k and 2^k+1 are 1, the others being null.

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

with

$$b_{ij} = \begin{cases} -1 & \text{if } i = j = 2^k \text{ or } 2^k+1; \\ 1 & \text{if } (i,j) = (2^k, 2^k+1) \text{ or } (i,j) = (2^k+1, 2^k); \\ 0 & \text{else.} \end{cases}$$

For all $k \geqslant 1$, the generator Λ_{2^k} is irreducible and symmetric. It is reversible with respect to the uniform measure on $\{1,\dots,2^k\}$. The generators Λ_{2^k} are quasi-decomposable. At time scale $1/\varepsilon^k$, for the generator $\Lambda_{2^{k+1}}$ the state space is divided into two halves, the local dynamics of each being governed by the generator Λ_{2^k}. Applying Proposition 4.1, the following result is obtained by induction.

PROPOSITION 4.2. *Let $\lambda_1 = 0 > \lambda_2 > \dots > \lambda_{2^{k+1}}$ be the eigenvalues of $\Lambda_{2^{k+1}}$, in decreasing order. One has:*

$$\lambda_1 = 0, \quad \lambda_{2^k+1} = -2$$

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

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

5. SIMULATED ANNEALING AT LOW TEMPERATURE

The Metropolis algorithm has been the subject of a large number of articles in the past ten years (see [14] and references therein). The interest for that algorithm comes from its numerous applications, from simulated annealing to Gibbs sampling ([18], [24], [19], [20], and [2]). The most important theoretical problem in that type of technique is the exact determination of the convergence speed. A review of rigorous results is given by Diaconis and Saloff-Coste ([14]). Most of them concern lower bounds of the spectral gap for the Markov chain being simulated. A lower bound of the gap can give an indication about the convergence speed, even if the full spectrum is out of reach due to the size of the state space. A complete description of the spectrum has been obtained in some particular cases by Diaconis et Hanlon ([12]) and Liu ([25]). We propose a result of this type here, for the case where some transitions are very slow. A typical case of application is the simulated annealing algorithm at low temperature, for which we shall describe the equivalents of the different eigenvalues.

Let us first recall the classical description of the simulated annealing algorithm (see [2]). Let $V = \{i, j, \ldots\}$ be a finite set, and H a function from V into $\mathbb{R}$, to be minimized. This function is interpreted as an energy and the corresponding Gibbs measure, denoted by p, and defined by

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

where T is the temperature parameter, and Z is the normalizing constant (partition function). As T tends to 0, the measure so defined converges to the uniform distribution on the set of global minima of the function H. The simulated annealing algorithm is a variant of the Metropolis algorithm that simulates a Markov chain, reversible with respect to the distribution p. The Markov chain is simulated by a rejection method, starting from a Markovian kernel K (cf. [14]). In most applications, the kernel K is that of the symmetric random walk on an unoriented graph, for which V is the set of vertices. The set of edges is still denoted by E. The algorithm will be seen as a continuous time Markov process, and described by its transition rates, which are positive only on the edges of the graph:

$$\lambda_{ij} = \begin{cases} 1 & \text{if } H(j) \leqslant H(i), \\ \mathrm{e}^{-\frac{1}{T}(H(j)-H(i))} & \text{if } H(j) > H(i); \end{cases} \qquad \forall\, \{i,j\} \in E.$$

It is immediate to check that the generator so defined is p-reversible. Our aim is to study the spectrum of that generator at low temperature. We shall set $\varepsilon = \mathrm{e}^{-1/T}$, and denote by Λ_ε the corresponding generator. In order to simplify notation, we shall assume that H can take only nonnegative integer values, 0 being the minimum, and that the difference of energy between neighboring vertices is at most 1:

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

This hypothesis is not really restrictive. Any function H can be shifted to take its values in $\mathbb{R}^+$, without changing neither the measure p nor the generator. Morrover, in order to extend our results to the general case, it would suffice to replace the integer powers of ε by exponents corresponding to the actual energy differences between neighbors (cf. Theorem 2.4). From now on the generator Λ_ε is defined on V by

$$\lambda^\varepsilon_{ij} = \begin{cases} 0 & \text{if } \{i,j\} \notin E, \\ 1 & \text{if } \{i,j\} \in E \text{ and } H(j) \leqslant H(i), \\ \varepsilon & \text{if } \{i,j\} \in E \text{ and } H(j) > H(i); \end{cases} \qquad \forall\, i,j \in V.$$

This generator is reversible with respect to the measure p_ε, now defined by:

$$p_\varepsilon(i) = \frac{1}{Z}\varepsilon^{H(i)}, \quad \forall\, i \in E.$$

This is another particular case of the general setting of Section 3. The structure of the problem allows a natural description of the hierarchical decomposition into classes. We shall be led to aggregations of states, corresponding to connected classes of the graph, with constant energy. The equivalence relation is the following:

$$i\mathcal{R}j \Leftrightarrow \begin{cases} \exists\, k_1 = i, k_2, \dots, k_\ell = j & \text{such that } \{k_h, k_{h+1}\} \in E, \\ & \text{and } H(k_1) = \cdots = H(k_\ell); \end{cases} \qquad \forall\, i,j \in V.$$

The classes of that relation are denoted by $\alpha, \beta, \dots$. We shall still denote by $H(\alpha)$ the common value of the energy function over the class α. Two classes α and β are said to *communicate* if there exists an edge of the graph linking them (the difference $|H(\alpha) - H(\beta)|$ can only be 1 in that case). A class α is said to be *minimal* for the energy function H if it communicates only with classes at a strictly larger energy level.

The following proposition describes the orders of the equivalents of the eigenvalues in terms of minimal classes of energy functions defined by induction, starting from H.

Proposition 5.1. *Set $H_0 = H$. Let $\mathcal{F}_0$ be the set of minimal classes of H_0. For any positive integer k, suppose an energy function H_k has been defined on V. Let $\mathcal{F}_k$ be the set of its minimal classes. The function H_{k+1} is defined on V by*

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

Denote by K the first integer such that $\#\mathcal{F}_k = 1$.

Then among the eigenvalues of Λ_ε, $\#V - \#\mathcal{F}_0$ converge to a strictly negative limit. For all k between 1 and K, $\#\mathcal{F}_{i-1} - \#\mathcal{F}_i$ eigenvalues are equivalent, up to a negative constant, to ε^i.

This description is not new. It is classically deduced from Freidlin et Wentzell's theory ([16]). It has been obtained by Catoni ([3]), and generalized by Trouvé ([34]) (see also in a different framework Miclo ([28] and [29])). The way it is derived here by a direct application of Theorem 2.15 is much simpler.

Instead of repeating in the particular case of simulated annealing the explicit expression of the successive generators given in Theorem 3.2, we shall illustrate it on an explicit example for a line graph with 17 vertices. The energy function H is that of Figure 1.

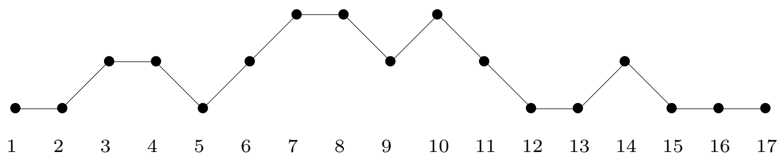


Figure 1. Energy function on a line graph with 17 vertices.

Proposition 5.1 predicts 12 eigenvalues of order 1, 3 of order ε, 1 of order ε^2 (and 1 equal to 0).

The generator Λ_0 has 3 eigenvalues equal to -3, 4 equal to -2, 5 equal to -1 and 5 equal to 0. The eigenvalues of order ε are obtained with the generator Λ_1. It is defined on the minimal classes of H which are:

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

This generator is the following:

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

Its spectrum is

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

These values are the equivalents of the eigenvalues of order ε. The equivalent of the eigenvalue of order ε^2 (spectral gap), is obtained by constructing the partition $\mathcal{E}_2$ and the generator Λ_2. The aggregation of classes α_1, α_2 on one side, and α_4, α_5 on the other side, leads to a process with two states. The transition rate from the first class to the second one is 1/15, the other one is 1/25. The reversible measure is proportional to the cumulated weights of the classes, which are 3 and 5 respectively. The eigenvalues of the corresponding generator are 0 and $-8/75$. Thus the spectral gap of Λ_ε is equivalent to $8/75\varepsilon^2$.

Numerical experiments have been performed on this example. Here are, for 4 values of ε, the maximal relative error between eigenvalues computed numerically, and their equivalents, as computed above.

$$\Delta = \max \frac{|\text{numerical} - \text{equivalent}|}{\text{numerical}}.$$

ε	10^{-1}	10^{-2}	10^{-3}	10^{-4}
Δ	0.253	0.076	0.023	0.0007

REFERENCES

1. R. BACHER, Y. COLIN DE VERDIÈRE, Multiplicités des valeurs propres et transformations étoile-triangle des graphes, *Bull. Soc. Math. France* **123**(1995), 101–117.
2. D. BERTSIMAS, J. TSITSIKLIS, Simulated annealing, *Statist. Sci.* **8**(1993), 10–15.
3. O. CATONI, Rough large deviation estimates for simulated annealing: Application to exponential schedules, *Ann. Probab.* **20**(1992), 1109–1146.
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(1996), 1–21; *J. Comb. Theory B*, to appear.
7. Y. COLIN DE VERDIÈRE, *Spectres de graphes*, Société Mathématique de France, 1998.
8. Y. COLIN DE VERDIÈRE, Spectre d'opérateurs différentiels sur les graphes, Proceedings Cortona conference, to appear.
9. P.J. COURTOIS, *Decomposability, Queuing and Computer Systems Applications*, Academic Press, New York 1977.

10. H.L. Cycon, R.G. Froese, W. Kirsch, B. Simon, *Schrödinger Operators, with Application to Quantum Mechanics and Global Geometry*, Springer Verlag, Berlin 1987.
11. F. Delebecque, A reduction process for perturbed Markov chains, *SIAM J. Appl. Math.* **43**(1983), 325–350.
12. P. Diaconis, P. Hanlon, Eigenanalysis for some examples of the Metropolis algorithm, *Contemp. Math.* **138**(1992), 99–117.
13. P. Diaconis, L. Saloff-Coste, Comparison theorems for reversible Markov chains, *Ann. Appl. Probab.* **3**(1993), 696–730.
14. P. Diaconis, L. Saloff-Coste, What do we know about the Metropolis algorithm, *J. Comput. System Sci.*, to appear.
15. P. Diaconis, D.W. Strook, Geometric bounds for eigenvalues of Markov chains, *Ann. Appl. Probab.* **1**(1991), 36–61.
16. M.I. Freidlin, A.D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer Verlag, Berlin 1984.
17. A. Frigessi, F. Martinelli, J. Sander, Computational complexity of MCMC methods, preprint 1994.
18. A. Frigessi, P. Stefano, Convergence rates of the Gibbs sampler, the Metropolis algorithm and other single-site updating dynamics. *J. Roy. Statist. Soc. Ser. B* **55**(1993), 205–219.
19. D. Geman, S. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI **6**(1984), 721–741.
20. S. Ingrassia, On the rates of convergence of Metropolis algorithm and Gibbs sampler, *Ann. Appl. Probab.* **22**(1994), 372–397.
21. T. Kato, *Perturbation Theory for Linear Operators*, Springer Verlag, New York 1978.
22. F.P. Kelly, *Reversibility and Stochastic Networks*, Wiley, London 1979.
23. R. Kindermann, J.L. Snell, *Markov Random Fields and Their Applications*, Contemp. Math., vol. 001, Amer. Math. Soc., Providence 1980.
24. S. Kirkpatrick, C.D. Gellat, M.P. Vecchi, Optimisation by simulated annealing, *Science* **220**(1983), 671–680.
25. J. Liu, Eigenanalysis for a Metropolis sampling scheme with comparisons to rejection sampling and importance sampling, to appear.
26. F. Martinelli, On the two-dimensional dynamical Ising model in phase coexistence region, *J. Statist. Phys.* **76**(1994), 1179–1246.
27. M. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, Equation of state calculation by fast computing machines, *J. Chem. Phys.* **21**(1953), 1087–1092.
28. L. Miclo, Comportement de spectres d'opérateurs de Schrödinger à basse température, *Bull. Sci. Math. (2)* **119**(1995), 529–553.
29. L. Miclo, Une étude des algorithmes de recuit simulé sous-admissibles, *Ann. Fac. Sci. Toulouse VI, Ser. Math. (4)* **4**(1995), 819–877.
30. Y. Pan, Spectres de processus de Markov, Thèse UJF, Grenoble 1997.
31. L. Saloff-Coste, Lectures on finite Markov chains, in *Ecole d'été de probabilité de Saint-Flour* XXVI, Lecture Notes in Math., vol. 1664, Springer-Verlag, New York 1997, pp. 301–413.
32. H.A. Simon, A. Ando, Aggregation of variables in dynamic systems, *Econometrica*, **29**(1961), 111–138.

33. A. Sinclair, M. Jerrum, Polynomial-time approximation algorithms for the Ising model, *SIAM J. Comput.* **22**(1993), 1087–1116.
34. A. Trouvé, Cycle decompositions and simulated annealing, *SIAM J. Control Optim.* **34**(1996), 966–986.

Y. COLIN de *VERDIÈRE*
Institut Fourier
B.P. 74
38402 Saint Martin d'Hères Cedex
FRANCE

Y. PAN
LMC/IMAG, B.P. 53
38041 Grenoble Cedex 9
FRANCE

BERNARD YCART
Bureau 39, Tour IRMA
LMC/IMAG, B.P. 53
38041 Grenoble Cedex 9
FRANCE

Received July 20, 1997; revised March 16, 1998.