ANNALES DE L'I. H. P., SECTION B

ALAIN TROUVÉ

Rough large deviation estimates for the optimal convergence speed exponent of generalized simulated annealing algorithms

Annales de l'I. H. P., section B, tome 32, nº 3 (1996), p. 299-348

<http://www.numdam.org/item?id=AIHPB_1996__32_3_299_0>

Ann. Inst. Henri Poincaré,

Vol. 32, n° 3, 1996, p. 299-348 *Probabilités et Statistiques*

Rough large deviation estimates for the optimal convergence speed exponent of generalized simulated annealing algorithms

by

Alain TROUVÉ

LMENS/DIAM, URA 762, Ecole Normale Supérieure, 45, rue d'Ulm, 75230 Paris Cedex 05, France. E-mail address: trouve@ens.ens.fr

ABSTRACT. – We study the convergence speed of generalized simulated annealing algorithms. Large deviations estimates uniform in the cooling schedule are established for the exit from the cycles in the spirit of Catoni's sequential annealing work [2]. We compute the optimal convergence speed exponent proving a conjectured of R. Azencott [1].

RÉSUMÉ. – Dans ce papier, nous étudions la vitesse de convergence des algorithmes de recuit généralisé. Nous établissons des estimées de grandes déviations uniformes en le schéma de température pour les temps et les lois de sorties des cycles dans l'esprit du travail de O. Catoni pour le recuit séquentiel [2]. Nous obtenons alors la valeur de l'exposant optimal pour la vitesse de convergence sur les schémas de température décroissants conjecturée par R. Azencott [1].

CONTENTS

1. INTRODUCTION

1.1. From sequential to generalized simulated annealing

Sequential simulated annealing is a very general and flexible method for finding good solutions to hard optimization problems. Its principle is surprisingly simple and was formulated in the early 1980s [12]. Consider a real valued function U to be minimized on a finite set E called the state space (or the configuration space). Now, let an irreducible Markov kernel q on E be given and define for all $T > 0$ the Markov kernel Q_T on E such that for any $i, j \in E$, $i \neq j$

$$Q_T(i,j) = q(i,j)\exp(-(U(j) - U(i))^+/T), \tag{1}$$

where x^+ denotes the positive part of x. For $T = 0$, denote $Q_0 = \lim_{T\to 0, T>0} Q_T$. An homogeneous Markov chain $X = (X_n)_{n\in\mathbb{N}}$ with transition kernel Q_0 satisfies $U(X_n) \geq U(X_{n+1})$ and is trapped in a finite number of steps into a local minima of U. For $T > 0$, Q_T is a perturbation of the previous mechanism allowing hill-climbing moves. Assuming that $q(i,j) = q(j,i)$, Q_T admits π_T as unique equilibrium probability measure which satisfies for all $i \in E$

$$\pi_T(i) = \exp(-U(i)/T)/Z_T. \tag{2}$$

The parameter T is physically interpreted as a temperature and one verifies that

$$\lim_{T \to 0} \pi_T(\{i \in E \mid U(i) = \min U\}) = 1. \tag{3}$$

This particular feature is the key of the simulated annealing approach. Considering a decreasing cooling schedule $\mathcal{T} = (T_n)_{n \in \mathbb{N}}$ i.e. a sequence of decreasing temperatures, we define an inhomogeneous Markov chain $X = (X_n)_{n \in \mathbb{N}}$ with transition kernel $Q_{T_{n+1}}$ at time n. The intuitive idea is that for sufficiently slowly decreasing cooling schedules, the law of X_n should be close to π_{T_n} and we should have

$$\sup_{i \in E} P(U(X_n) \neq \min U \mid X_0 = i) \underset{n \to \infty}{\longrightarrow} 0. \tag{4}$$

Many results are known on sequential simulated annealing. Since Hajek's paper [8], one knows that there exists $H \geq 0$ such that for all decreasing cooling schedules $\mathcal{T}$ vanishing to zero, (4) holds if and only if $\sum_{n \geq 0} e^{-H/T_n} = +\infty$. This problem of convergence can be extended to the more general problem of convergence on level sets $E_\lambda = \{ i \in E \mid U(i) \geq \min U + \lambda \}$ for $\lambda > 0$. This problem has been successfully solved by O. Catoni in [2] who proves that there exists $\Gamma_\lambda \geq 0$ such that if $\mathcal{T}$ is a decreasing cooling cooling schedule vanishing to zero, then

$$\sup_{i \in E} P(X_n \notin E_\lambda \mid X_0 = i) \underset{n \to \infty}{\longrightarrow} 0 \text{ iff } \sum_{n \geq 0} e^{-\Gamma_\lambda/T_n} = +\infty. \tag{5}$$

More crucial for practical use of sequential simulated annealing is the rate of convergence. More precisely, let

$$M_n = \inf_{T_0 \geq T_1 \geq \cdots \geq T_n} \sup_{i \in E} P(X_n \in E_\lambda \mid X_0 = i). \tag{6}$$

Catoni's work [2] gives that there exist $\alpha_\lambda \geq 0$, $\hat{\alpha}_\lambda \geq 0$, $K_1 > 0$ and $K_2 > 0$ such that

$$\frac{K_1}{n^{\alpha_\lambda}} \leq M_n \leq \frac{K_2}{n^{\hat{\alpha}_\lambda}}. \tag{7}$$

The lower bound says that the mass in E_λ cannot vanish faster than a power of $1/n$ and that this power is upper bounded by α_λ. Conversely, the upper bound says that for each $N \in \mathbb{N}$ there exists a decreasing cooling schedule $\mathcal{T}^N$ such that for this cooling schedule

$$\sup_{i \in E} P(X_N \in E_\lambda \mid X_0 = i) \leq \frac{2K_2}{n^{\hat{\alpha}_\lambda}}. \tag{8}$$

Note that we may have $\mathcal{T}^N \neq \mathcal{T}^{N+1}$ and in fact the proof of (8) shows that one should designed the cooling schedule according to the horizon. The value of α_λ and $\hat{\alpha}_\lambda$ are given explicitly in function of U and q (even if their numerical computation is an hard combinatorial problem). Since for sufficiently small value of λ we have $\alpha_\lambda = \hat{\alpha}_\lambda$ we deduce that if

$$\alpha_{opt} = \lim_{\lambda \to 0} \alpha_\lambda = \lim_{\lambda \to 0} \hat{\alpha}_\lambda, \tag{9}$$

then we cannot expect for the convergence to the set of global minima of U a better speed exponent with decreasing cooling schedules than α_{opt}.

In many practical situation, computer scientists have proposed modified version of the sequential simulated annealing in order to increase the speed of convergence. One of the most promising issues is the parallelization of the algorithm in order to distribute the computations on several processes. As an illustration of such an approach, let us have a look on sequential simulated annealing in image processing (see [6]). In this case, the configuration space E is a product space $E = L^S$ where S is a set of pixels and L a set of labels (usually grey levels). For each $i \in E$, $i(s)$ is the value of the grey level at pixel $s \in S$. The sequential annealing process has in this framework the following particular form. One define first a family of local updating kernels $(Q_{s,T})_{s \in S, T \geq 0}$ by

$$Q_{s,T}(i,j) = \begin{cases} \frac{\exp(-U(i^{l,s})/T)}{\sum_{l' \neq l, l' \in L} \exp(-U(i^{l',s})/T)} & \text{if } j = i^{l,s}, \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where $i^{l',s}$ is the configuration $j \in E$ such that $j(s) = l'$ and $j(s') = i(s')$ for $s' \neq s$ (only the pixel s is changed). Now, let a family $(s_p)_{1 \leq p \leq |S|}$ exhausting the elements of S be given (generally we choose a sweeping line by line of the image). We define the family $Q = (Q_T)_{T \geq 0}$ of transition kernels of the sequential simulated annealing by

$$Q_T(i,j) = (Q_{s_1,T} Q_{s_2,T} \cdots Q_{s_p,T})(i,j). \tag{11}$$

Rigorously speaking, Q_T cannot be written of the form (1) and the previous results do not apply. However, an extension of them to this case does not imply deep modifications of the simulated annealing theory since π_T is still the unique invariant probability measure of Q_T and since Q_T is reversible for $T > 0$. Now, consider the following parallelization of the sequential process: assume that one has a small processor attaches to each pixel and assume that all the pixels are updating synchronously according to the local updating rule. We define then a new transition matrix given by

$$K_T(i,j) = \prod_{s \in S} Q_{s,T}(i, i^{j(s),s}). \tag{12}$$

One can expect that the speed of convergence is increased by a factor $|S|$ since all the labels are updating in one unit of time (instead of $|S|$ units of time previously). However, some very important changes have occurred owing to the interactions between processors and we have to generalize the framework of simulated annealing. Instead of the form (1), we can only say that there exist an irreducible Markov kernel q on E, a real valued number $\kappa \in [1, +\infty[$ and a family of non negative numbers $(V(i,j))_{i,j \in E}$ such that

$$(13) \qquad \frac{1}{\kappa} q(i,j) e^{-V(i,j)/T} \leq K_T(i,j) \leq \kappa q(i,j) e^{-V(i,j)/T}.$$

For the classical sequential annealing, we have $\kappa = 1$ and $V(i,j) = (U(j) - U(i))^+$ so that $V(i,j) - V(j,i) = U(j) - U(i)$. This property which is a consequence of the reversibility of the sequential annealing transition kernel does not hold any more for the kernel K_T (in fact, there does not exist any function W such that $V(i,j) - V(j,i) = W(j) - W(i)$). As a consequence, none of the previous results can apply and we need an extended theory to handle the convergence properties of parallel version of sequential simulated annealing. This extension will be the main subject of this paper. We should emphasize the fact that many stochastic optimization algorithms fall into the scope of this extension and not only the parallel versions of the sequential annealing for image processing presented above. This extension covers in fact all the small random perturbations (with discrete time and finite state space) of dynamical systems as developed by Wentzell and Freidlin in [5]. The point is to handle here the convergence properties of such processes when the perturbation vanishes with time.

1.2. Theoretical framework

We introduce now precisely our theoretical framework. Let E be a finite set which will stay fixed throughout this work.

DEFINITION 1.1. – Let $\kappa \in [1, +\infty[$ and let q be an irreducible Markov kernel on E. We say that a continuously parametered family $Q = (Q_T)_{T \geq 0}$ of Markov kernel on E is admissible for κ and q if and only if there exists a family $V = (V(i,j))_{i,j \in E}$ such that

(1) $V(i,j) \in [0, +\infty]$ and $V(i,j) < +\infty$ iff $q(i,j) > 0$,

(2) for all $T \geq 0$, all $i, j \in E$

$$\frac{1}{\kappa} q(i,j) e^{-V(i,j)/T} \leq Q_T(i,j) \leq \kappa q(i,j) e^{-V(i,j)/T}.$$

For $T = 0$, we take the convention that $e^{-V(i,j)/0} = \mathbf{1}_{V(i,j)=0}$. The family V is called the communication cost and the set of all the admissible families Q for q and κ is denoted by $\mathcal{A}(q, \kappa)$.

Remark 1. – As a main extension of the standard sequential simulated annealing transition kernels, we do not assume that q is symmetric and and V can be completely arbitrary.

In the following definition, we define the probability measure associated with an abstract annealing process on E.

DEFINITION 1.2. – Let $X = (X_n)_{n\in\mathbb{N}}$ be the coordinate process on $E^{\mathbb{N}}$. For all $\theta = (Q, \mathcal{T}, \nu_0)$ where

• $Q = (Q_T)_{T\geq 0}$ is a continuously parametered family of Markov kernels on E,

• $\mathcal{T} = (T_n)_{n\in\mathbb{N}}$ is a sequence of non negative real valued numbers called cooling schedule,

• ν_0 is a probability measure on E called the initial distribution,

we denote P_θ the unique probability measure on $E^{\mathbb{N}}$ for the σ-algebra $\sigma(X_n, n \geq 0)$ such that X is under P_θ a Markov chain with initial distribution ν_0 and transition kernel $Q_{T_{n+1}}$ at time n.

We can now define the extended notion of generalized simulated annealing.

DEFINITION 1.3. – Let q be an irreducible Markov kernel on E, let $\kappa \in [1, +\infty[$ and let $X = (X_n)_{n\in\mathbb{N}}$ be the coordinate process on $E^{\mathbb{N}}$. We say that X is a generalized simulated annealing process (G.S.A.) with parameter (q, κ, P) if there exist $Q \in \mathcal{A}(q, \kappa)$, a cooling schedule $\mathcal{T}$ and an initial distribution ν_0 such that $P = P_\theta$ with $\theta = (Q, \mathcal{T}, \nu_0)$. Moreover, as usual, $P(\ \mid X_0 = i)$ will denote the probability measure P_{θ_i} where for all $i \in E$, $\theta_i = (Q, \mathcal{T}, \delta_i)$.

Remark 2. – An equivalent framework has been considered by Hwang and Sheu in [10]. Their results will be reported below in this section.

Considering an arbitrary communication cost V, a G.S.A. is not defined around an explicit energy function U. However, there exists an implicit function W introduced by the definitions below which plays for low temperatures the same role. This function depends on the communication cost V through a functional on A-graphs that we define now.

DEFINITION 1.4. – Let $A \subset E$. We say that a set g of arrows $i \to j$ in $A^c \times E$ is an A-graph iff:

(1) for each $i \in A^c$, there exists an unique $j \in E$ such that $i \to j \in g$;

(2) for each $i \in A^c$, there is a path in g ending on a configuration in A.

We denote by $G(A)$ the set of the A-graphs.

DEFINITION 1.5. – Let q be an irreducible Markov kernel on E, let $\kappa \in [1, +\infty[$ and let $Q \in \mathcal{A}(q, \kappa)$. Let V be the associated communication cost. For each $i \in E$ and each $g \in G(\{i\})$, we denote $V(g) = \sum_{i \to j \in g} V(i, j)$. We say that $W : E \longrightarrow \mathbb{R}_+$ is the virtual energy associated with Q if and only if for each $i \in E$

$$W(i) = \inf\{ \, V(g) \mid g \in G(\{i\}) \, \}.$$

Remark 3. – Since $V(i, j) < +\infty$ if and only if $q(i, j) > 0$ and since q is irreducible on E, one easily verifies that $W(i) < +\infty$. The above construction of W has been given by Wentzell and Freidlin in [5] as well as the following proposition.

PROPOSITION 1.6. – *Let q be an irreducible Markov kernel on E, let $\kappa \in [1, +\infty[$ and let $Q \in \mathcal{A}(q, \kappa)$. For each $T > 0$, Q_T is irreducible and we denote μ_T its unique invariant probability measure. Then if W is the virtual energy associated with Q we have*

$$T \ln(\mu_T(i)) \underset{T \to 0}{\longrightarrow} -(W(i) - \min W).$$

Remark 4. – This proposition shows that $\mu_T(\{i \in E \mid W(i) = \min W \,\}) \to 1$ when the temperature vanishes. Moreover, $\mu_T(i)$ is of the order of $\exp(-(W(i) - \min W)/T)$ for low temperature which should be compared to $\pi_T(i)$ of the order of $\exp(-(U(i) - \min U)/T)$ for the sequential simulated annealing. Hence a G.S.A. is a optimizating process for the virtual energy.

1.3. Outlines

In the general framework, one can easily show that the generalized simulated annealing algorithm converges for sufficiently slowly decreasing cooling schedule to the global minima of a virtual energy function W. For cooling schedule $T_n = c/(\ln(n + 2))$, a necessary and sufficient condition on the value of c for this convergence is given in [3], [10], [9] and [13] without giving the optimal convergence speed in particular because such a result needs large deviation estimates uniform in the cooling schedules as established by O. Catoni in [2]. Our main goal will be here to prove that the approach of O. Catoni can be successfully extended to the general framework and gives the value of the optimal convergence speed exponent conjectured by R. Azencott in [1].

We will start in Section 2 with the cycle decomposition of the state space E. Roughly speaking, a cycle is a subset of E which can be considered as a point for sufficiently low temperature. One of their important properties is that at constant temperature T, the exit time of a cycle Π is of order $e^{H_e(\Pi)/T}$ whatever the starting point in Π is. The non negative number $H_e(\Pi)$ is called the exit height of Π. The cycles are structured on a tree whose root is E and leaves are the singletons since two cycles Π_1 and Π_2 are either disjoint or included one in the other. We will recall the recursive construction of the cycles which depends only on the communication cost V of the G.S.A. and we will show the link between the virtual energy and the cycle decomposition.

In Section 3, we will introduce for each $A \subset E$, $i \in A$ and $j \in E$ a new cost $C_A(i,j)$ called the renormalized communication cost in A. Starting from $i \in A$, the probability that the process at constant temperature visits the edge (i,j) before the exit time of A (included) is of order $\exp(-C_A(i,j)/T)$. This renormalized cost in A will be expressed in function of the cycle decomposition.

Both previous sections are in fact devoted to the combinatorial definitions of the cycle decomposition and of the renormalized costs without any results on their probabilistic meaning. We will state in Section 4, large deviation estimates on exit time and exit point of subset of E which will enlighten the probabilistic role of the quantities mentioned above. We will handle large deviation estimates for arbitrary decreasing cooling schedules which are the keys for the study of the convergence of the G.S.A.

As done for the sequential simulated annealing, the problem of convergence can be written in the following way: Let $\lambda > 0$ and denotes

$$E_\lambda = \{ \ i \in E \mid W(i) \geq \min W + \lambda \ \}.$$

Our first convergence result, established in Section 5, concerns the necessary and sufficient condition on the cooling schedule $\mathcal{T}$ for which the mass in E_λ vanishes when the number of steps increases:

Let q be an irreducible Markov kernel on E and let $\kappa \in [1, +\infty[$. Let X be a G.S.A. with parameter (q, κ, P) and let $\mathcal{T}$ be the underlying cooling schedule. Assume that $\mathcal{T}$ is decreasing ($T_n \geq T_{n+1}$) and vanishes when n tends to infinity. Then for all $\lambda > 0$, there exists $\Gamma_\lambda \geq 0$ (independent of $\mathcal{T}$) such that

$$\sup_{i \in E} P(\ X_n \in E_\lambda \mid X_0 = i) \underset{n \to +\infty}{\longrightarrow} 0 \text{ iff } \sum_{n \geq 0} e^{-\Gamma_\lambda / T_n} = +\infty.$$

The value of Γ_λ is explicitly given in function of the cycle decomposition by

$$\Gamma_\lambda = \sup\{ \ H_e(\Pi) \mid \Pi \text{ cycle such that } \inf_{i \in \Pi} W(i) \geq \min W + \lambda \ \}.$$

Finally, in Section 6, we study the optimal convergence speed and we establish the following result which says that if q is an irreducible Markov kernel on E, if $\kappa \in [1, +\infty[$ and if X is a G.S.A with parameter (q, κ, P) whose underlying cooling schedule $\mathcal{T}$ is decreasing, then, assuming that the underlying family $Q \in \mathcal{A}(q, \kappa)$ satisfies an additional weak condition C_1 (which is fulfilled for instance if $Q_T(i,j) = \sum_k a_{i,j}^k e^{-b_{i,j}^k/T} / \sum_k c_{i,j}^k e^{-d_{i,j}^k/T}$), there exists $b > 0$ (which depends on Q but not on $\mathcal{T}$) such that for all level $\lambda > 0$ and all $n > 0$,

$$\sup_{i \in E} P(X_n \in E_\lambda \mid X_0 = i) \geq \frac{b}{n^{\alpha_\lambda}}$$

with

$$\alpha_\lambda = \inf \left\{ \frac{W(\Pi) - \min W}{H_e(\Pi)} \mid \Pi \text{ cycle}, W(\Pi) - \min W \geq \lambda \right\}$$

where $W(\Pi) = \inf_{i \in \Pi} W(i)$.

We finally establish a last result which say we can construct for each N a decreasing cooling schedule $\mathcal{T}^N$ such that if $P_N = P_{(Q, \mathcal{T}^N, \nu_0)}$ then there exists $K_1 > 0$ such that

$$\sup_{i \in E} P_N(X_N \in E_\lambda \mid X_0 = i) \leq \frac{K_1}{N^{\hat{\alpha}_\lambda}}$$

where

$$\hat{\alpha}_\lambda = \inf \left\{ \frac{(W(\Pi) - \min W) \wedge \lambda}{H_e(\Pi)} \mid \Pi \text{ cycle}, \ W(\Pi) > \min W \right\}.$$

This exponent is optimal for sufficiently small $\lambda > 0$ since we have then $\hat{\alpha}_\lambda = \alpha_\lambda$. As a corollary, we prove a conjecture of Azencott [1] on the optimal convergence speed exponent for the convergence towards the global minima of the virtual energy W.

2. THE CYCLE DECOMPOSITION

Let q be an irreducible Markov kernel on E and let $\kappa \geq 1$. Let $Q = (Q_T)_{T \geq 0} \in \mathcal{A}(q, \kappa)$ and denotes V the underlying communication cost. We present now the cycle decomposition of E which depends only on V. The proof of the probabilistic properties of the cycles are postponed to Section 4 where we will study exit time and exit point from a cycle.

The cycle decomposition for a given communication cost has been introduced by Wentzell and Freidlin [1] and we recall it briefly for completeness. We need first to define the usual notion of path associated with a cost function.

DEFINITION 2.1. – Let F be a finite set, $C : F \times F \to [0, +\infty]$ be a function (such a function will be called a communication cost function on F), and i and j be two distinct configurations of F.

(1) We denote $\mathrm{Pth}_F(i, j)$ the set of all paths $(g_k)_{0 \leq k \leq n}$ in F such that $g_0 = i$ and $g_n = j$. The path dependent integer n will be denoted n_g and called the length of g.

(2) Let g in $\mathrm{Pth}_F(i, j)$, we defined $C(g)$ by

$$C(g) = \sum_{k=0}^{n_g - 1} C(g_k, g_{k+1}).$$

We adopt the convention that $C(g) = +\infty$ if one of the summation terms is $+\infty$.

The decomposition in cycles is defined in a recursive way. First the set E^0 of the cycle of order 0 is defined by:

$$E^0 = \{ \{i\} \mid i \in E \}.$$

Let us consider the communication cost function V^0 on E^0 defined by

$$V^0(\{i\}, \{j\}) = V(i, j).$$

Assume that the set E^k of the cycles of order k has been constructed as well as a communication cost function V^k on E^k. The construction of the pair E^{k+1}, V^{k+1} can be split in several steps:

(1) From V^k, we define an other communication cost V_*^k on E^k by

$$V_*^k(\Pi, \Pi') = \begin{cases} 0 & \text{if } \Pi = \Pi', \\ V^k(\Pi, \Pi') - H_e^k(\Pi) & \text{otherwise,} \end{cases}$$

$$\text{where } H_e^k(\Pi) = \inf_{\Pi'' \neq \Pi} V^k(\Pi, \Pi'').$$

(2) On E^k, we define the relation $\xrightarrow{k}$ by

$$\Pi \xrightarrow{k} \Pi' \text{ if either } \Pi = \Pi' \text{ or there exists } g \in \mathrm{Pth}E^k(\Pi, \Pi') \text{ such that } V_*^k(g) = 0,$$

and the equivalence relation $\mathcal{R}_k$ by

$$\Pi\mathcal{R}_k\Pi' \text{ if either } \Pi = \Pi' \text{ or } \Pi \xrightarrow{k} \Pi' \text{ and } \Pi' \xrightarrow{k} \Pi.$$

(3) Define D^{k+1} by

$$D^{k+1} = \left\{ \bigcup_{\Pi'\mathcal{R}_k\Pi} \Pi' \mid \Pi \in E^k \right\},$$

and define on D^{k+1} the partial order $\preceq$ by $\Pi_1^{k+1} \preceq \Pi_2^{k+1}$ if there exist $\Pi_i^k \subset \Pi_i^{k+1}$ for $i = 1, 2$ such that $\Pi_2^k \xrightarrow{k} \Pi_1^k$. We note D_*^{k+1} the set of the minimal elements of D^{k+1} for the order $\preceq$.

(4) Define E^{k+1} by

$$E^{k+1} = D_*^{k+1} \cup \{ \Pi^k \in E^k \mid \exists \Pi^{k+1} \in D^{k+1} \setminus D_*^{k+1}, \ \Pi^k \subset \Pi^{k+1} \}.$$

(5) We define now a communication cost V^{k+1} on E^{k+1} by

$$V^{k+1}(\Pi_0^{k+1}, \Pi_1^{k+1}) = H_m^{k+1}(\Pi_0^{k+1}) +$$

$$\inf\{ V^k(\Pi_0^k, \Pi_1^k) - H_e^k(\Pi_0^k) \mid (\Pi_0^k, \Pi_1^k) \in E^k \times E^k, \ \Pi_0^k \subset \Pi_0^{k+1}, \ \Pi_1^k \subset \Pi_1^{k+1} \}$$

where $H_m^{k+1}(\Pi_0^{k+1}) = \sup\{ H_e^k(\Pi_0^k) \mid \Pi_0^k \subset \Pi_0^{k+1}, \ \Pi_0^k \in E^k \}$.

The construction continues until $E^k = \{E\}$. We denote n_E the integer such that $E^{n_E} \neq \{E\}$ and $E^{n_E+1} = \{E\}$.

This procedure gives an hierarchical decomposition of the states space on a tree beginning with the singletons and ending with the whole space. One could find an helful example of this recursive construction in [16]. From this tree we will introduce the following subsets.

DEFINITION 2.2. – (1) We denote $\mathcal{C}(E)$ the set of all the cycles that is

$$\mathcal{C}(E) = \cup_{k=0}^{n_E+1} E^k$$

(2) Let $A \subset E$.

- We define the maximal partition $\mathcal{M}(A)$ of A by:

$$\mathcal{M}(A) = \{ \Pi \in \mathcal{C}(E) \mid \Pi \text{ is a maximal element in } \mathcal{C}_A(E) \}$$

$$\text{where } \mathcal{C}_A(E) = \{ \Pi \in \mathcal{C}(E) \mid \Pi \subset A \}.$$

• We define the maximal proper partition $\mathcal{M}_*(A)$ of A by:

$$\mathcal{M}_*(A) = \{ \ \Pi \in \mathcal{C}(E) \mid \Pi \text{ is a maximal element in } \mathcal{C}^*_A(E) \ \}$$

$$\text{where } \mathcal{C}^*_A(E) = \{ \ \Pi \in \mathcal{C}(E) \mid \Pi \subset A, \ \Pi \neq A \ \}.$$

We will now introduce in the following definition some important parameters on the cycles.

DEFINITION 2.3. – Let $\Pi \in \mathcal{C}(E)$.

(1) We denote $H_e(\Pi)$ the real valued number

$$H_e(\Pi) = \begin{cases} \sup\{H^k_e(\Pi) \mid k \leq n_E \ \Pi \in E^k \ \} & \text{if } \Pi \neq E, \\ +\infty & \text{if } \Pi = E. \end{cases}$$

The real $H_e(\Pi)$ will be called the exit height of Π. For $\Pi = \{i\}$ we will often prefer the notation $H_e(i)$ to $H_e(\{i\})$.

(2) We denote $H_m(\Pi)$ the real number

$$H_m(\Pi) = \sup\{ \ H_e(\Pi') \mid \Pi' \in \mathcal{M}_*(\Pi) \ \} \vee 0.$$

The real $H_m(\Pi)$ will be mixing height of Π.

(3) We denote $W(\Pi)$ the real valued number

$$W(\Pi) = \inf\{ \ W(i) \mid i \in \Pi \ \}.$$

(4) We denote $F(\Pi)$ the set

$$F(\Pi) = \{ \ i \in \Pi \mid W(i) = W(\Pi) \ \}.$$

The set $F(\Pi)$ will be called the bottom of Π.

Remark 5. – The probabilistic interpretation of $H_e(\Pi)$ and $H_m(\Pi)$ can be easily given. For $T > 0$, let X be a G.S.A. with parameter (q, κ, P_T) where the underlying cooling schedule $\mathcal{T} = (T_n)_{n\in\mathbb{N}}$ is assuming to be constant and equal to T ($T_n = T$, $\forall n \in \mathbb{N}$). For each cycle Π the exit time from Π is of order $e^{H_e(\Pi)/T}$ for any starting point in Π and the probability to go from $i \in \Pi$ to $j \in \Pi$ within a time of order $e^{H_m(\Pi)/T}$ tends to 1 when the temperature T vanishes.

Some parts of the above definition can be extended to an arbitrary set.

DEFINITION 2.4. – Let $A \subset E$.

(1) We define the exit height $H_e(A)$ of A by

$$H_e(A) = \sup\{ H_e(\Pi) \mid \Pi \in \mathcal{C}(E),\ \Pi \subset A \} \vee 0.$$

(2) We denote $W(A)$ the real valued number

$$W(A) = \inf\{ W(i) \mid i \in A\}.$$

(3) We define the bottom $F(A)$ of A by

$$F(A) = \{ i \in A \mid W(i) = W(A) \}.$$

DEFINITION 2.5. – Let $i \in E$, we define by induction the increasing family of cycles $(i^k)_{0 \leq k \leq n_E}$ by $i^0 = \{i\}$ and

$$i^{k+1} \in E^{k+1},\ i^k \subset i^{k+1} \text{ for } k \leq n_E.$$

Remark 6. – A set A which is not a cycle should be considered as an inhomogeneous subset of E for the exit time from A. However, for any starting point in A, a G.S.A with parameter (q, κ, P_T) (see remark 5) can exit from A at least within a time of order $e^{H_e(A)/T}$.

We show in the next proposition proved in [15] the strong link that exists between the virtual energy W and the cycle decomposition.

PROPOSITION 2.6. – *Let $i \in E$. Then*

$$W(i) = A_E - \sum_{0 \leq k \leq n_E} (H_e^k(i^k) - H_m^k(i^k))$$

$$\textit{where } A_E = \sum_{k=0}^{n_E} \sum_{\Pi \in E^k} H_e^k(\Pi) - H_m^k(\Pi).$$

3. RENORMALIZATION OF THE COMMUNICATION COST

In this section, we define from the communication cost V and an action functional for the exit paths from a subset A of E. More precisely, starting from a point $i \in A$, we want to evaluate the probability that before the exit time from A, the process (running at constant temperature T) visits the edge (i, j). It appears that this probability is of order $e^{-C_A(i,j)/T}$. Hence $C_A(i, j)$ can be interpreted as a new cost which weights the cost of a large deviation event (the visit of the edge (i, j)) from the standard behavior of the process running in A. This cost will be called the renormalized cost in A. This section is devoted to its combinatorial definition in function of V and to preparatory lemmas for the large deviation estimates established in Section 4.

DEFINITION 3.1. – (1) Let $i, j \in E$, $i \neq j$. We denote by n_{ij} the integer uniquely defined by :

$$i^{n_{ij}} \neq j^{n_{ij}} \qquad \text{and} \qquad i^{n_{ij}+1} = j^{n_{ij}+1}.$$

(2) Let $A \subset E$, $A \neq E$. For all $i \in E$ we define $n_{i,A}$ by

$$n_{i,A} = \inf\{\ 0 \leq k \leq n_E \mid i^{k+1} \cap A^c \neq \emptyset\ \}.$$

PROPOSITION 3.2. – *Let $A \subset E$, $A \neq E$. We define on E the function C_A by :*

$$C_A(i,j) = \begin{cases} 0 & if\ i = j, \\ V(i,j) & if\ i \notin A\ and\ j \neq i, \\ V(i,j) - \sum\limits_{k=0}^{n_{ij} \wedge n_{i,A}} (H_e^k(i^k) - H_m^k(i^k)) & otherwise. \end{cases}$$

Then C_A is a communication cost called the renormalized communication cost in A.

Proof. – It is sufficient to prove that $C_A(i,j) \geq 0$. Moreover, it is sufficient to prove that for $i \neq j$ we have :

$$V(i,j) - \sum_{k=0}^{n_{ij}} (H_e^k(i^k) - H_m^k(i^k)) \geq 0.$$

This will be proved by induction on the value of n_{ij}.

Assume that $n_{ij} = 0$. Then (14) is equivalent to

$$V(i,j) - H_e(i) \geq 0. \tag{14}$$

which is obvious.

Now assume that the result has been proved for $n_{ij} \leq k-1$. Then from the recursive construction of the cycles and from the induction hypothesis we have

$$V^1(i^1, j^1) - \sum_{k=1}^{n_{ij}} (H_e^k(i^k) - H_m^k(i^k)) \geq 0. \tag{15}$$

Furthermore, from the definition of V^1 we have

$$(V(i,j) - H_e(i)) - (V^1(i^1, j^1) - H^1_m(i^1)) \geq 0. \tag{16}$$

Adding (15) and (16) we deduce (14) so that the proposition is proved. □

DEFINITION 3.3. – Let $A \subset E$, and let $i, j \in E$ be two distinct configurations. We denote by $C^*_A(i,j)$ the number

$$C^*_A(i,j) = \inf\{C_A(g) \mid (g_k)_{0 \leq k \leq n_g} \in \mathrm{Pth}_E(i,j);\ g_k \in A \text{ for } 0 < k < n_g\}.$$

Here the infimum is taken on the paths from i to j that stay within A except eventually at their extremities.

Remark 7. – Starting from the probabilistic interpretation of C_A given in the introduction of this section, we can easily deduce the interpretation of $C^*_A(i,j)$. For all $i \in A$ and $j \in E$, if $g \in \mathrm{Pth}_A(i,j)$ is a path staying in A (excepted eventually for its right hand extremity), the probability that the process starting from i visits all the edges of g before exiting from A is of order $e^{-C_A(g)/T}$. Hence, if $j \in A^c$, the probability that the process exit from A at point j is of order $e^{-C^*_A(i,j)/T}$.

We establish now four lemmas which will be useful for the next section.

LEMMA 3.4. – *Let* $\Pi \in \mathcal{C}(E)$ *and* $A \subset E$, $A \neq E$.

(1) *Assume that* $\Pi \neq E$ *and let* $\Pi' \in E^{n_\Pi}$ *be a cycle such that*

$$V^{n_\Pi}(\Pi, \Pi') = H_e^{n_\Pi}(\Pi). \tag{17}$$

Then there exist $e \in \Pi$ *and* $f \in \Pi'$ *such that* $C_\Pi(e,f) = 0$.

(2) *For all* i, $j \in \Pi$, *we have:*

$$C^*_\Pi(i,j) = 0.$$

(3) *For all* $i \in A$, *let* $C^*_A(i, A^c) = \inf\{\ C^*_A(i,j) \mid j \in A^c\ \}$. *We have*

$$C^*_A(i, A^c) = 0.$$

Remark 8. – According to the probabilistic interpretation given in the remark 7, (1) defines the arrows followed by the process when its goes out of a particular cycle. For a cycle of level 0, that is for singleton $\{e\}$, these arrows are defined by $V(e,f) = H^0_e(e)$. Moreover, (2) says that starting from $i \in \Pi$, the process visits all the configurations $j \in \Pi$ before exiting with a probability non vanishing with the temperature. The interpretation of (3) is straightforward.

Proof. – Let us begin with part 1). From the recursive definition of the cycle, there exist $e \in \Pi$ and $f \in \Pi'$ satisying:

$$(18) \qquad \begin{cases} V^r(e^r, f^r) - H_e^r(e^r) = V^{r+1}(e^{r+1}, f^{r+1}) - H_m^{r+1}(e^{r+1}) \; ; \\ \qquad\qquad 0 \leq r < n_\Pi - 1. \end{cases}$$

We verify that

$$\begin{aligned} C_\Pi(e,f) &= V^{n_\Pi}(e^{n_\Pi}, f^{n_\Pi}) - H_e^{n_\Pi}(e^{n_\Pi}) \\ &+ \sum_{l=0}^{n_\Pi - 1} \{(V^l(e_h^l, f_h^l) - H_e^l(e_h^l)) - (V^{l+1}(e_h^{l+1}, f_h^{l+1}) - H_m^{l+1}(e_h^{l+1}))\}. \end{aligned}$$

so that we get with (17) and (18) that $C_\Pi(e,f) = 0$.

We prove now the part 2) of the lemma by induction on $k = \sup_{i,j \in \Pi} n_{ij}$. Assume that $k = 0$, then $n_{ij} = 0$ and $j \in i^1$. Furthermore, for all $a, b \in \Pi$, $a \neq b$, $n_{ab} = 0$ we have

$$C_\Pi(a,b) = V(a,b) - H_e(a) = V_*^0(\{a\}, \{b\}),$$

so that the result follows easily.

Assume now that the result is true for $k-1$. If $i^k = j^k$ then the induction hypothesis gives $C_{i^k}^*(i,j) = 0$. Since we have $C_{i^k}^*(i,j) \geq C_\Pi^*(i,j)$, the result is proved. Otherwise, it is sufficient to prove that if $V^k(i^k, j^k) = H_e^k(i^k)$ then $C_\Pi(i,j) = 0$. However, using part 1) with $\Pi = i^k$, we deduce that there exist $e \in i^k$ and $f \in j^k$ such that $C_\Pi(e,f) \leq C_{i^k}(e,f) = 0$. The proof is completed if we notice that we get also from the induction hypothesis that

$$C_\Pi^*(i,e) = C_\Pi^*(f,j) = 0.$$

We are now concerned by the proof of 3). We start by proving the result for a cycle $\Pi \in \mathcal{C}(E)$ distinct of E. Since the point 2) is proved, it is sufficient to prove that there exist $e \in \Pi$ and $f \in \Pi^c$ such that $C_\Pi(e,f) = 0$. However, this has been proved in part 1). We consider now the general case of a strict subset of E. Let i be an element of A, we prove the result by induction on the value $n_{i,A}$.

We assume here that $n_{i,A} = 0$. From the definition of $n_{i,A}$, we deduce that $i^1 \cap A^c \neq \emptyset$. Let j be in $i^1 \cap A^c$. From the construction of the cycle i^1, there exists a path $g \in \mathrm{Pth}_{i^1}(i,j)$ such that:

$$V(g_k, g_{k+1}) - H_e(g_k) = 0; \; 0 \leq k < n_g.$$

Hence, if $r_g = \inf\{ 0 \leq k \leq n_g \mid g_k \in A^c \}$, the path g stopped at r_g defined by $\hat{g} = (g_k)_{0 \leq k \leq r_g}$ verifies:

$$C_A(\hat{g}_k, \hat{g}_{k+1}) = 0;\ 0 \leq k < r_g = n_{\hat{g}},$$

so that $C_A^*(\hat{g}) = 0$ and the result is proved.

Assume now that the result is proved for all i in A such that $n_{i,A} < k$. Assume that $n_{i,A} = k$. We note then $\Pi_+ = i^{n_{i,A}+1}$. From the definition of $n_{i,A}$, we have $\Pi_+ \cap A^c \neq \emptyset$. Let j be in $\Pi_+ \cap A^c$, we have from the point 1) that $C_{\Pi_+}^*(i,j) = 0$. Hence, we consider a path $g \in \mathrm{Pth}_{\Pi_+}(i,j)$ such that $C_{\Pi_+}(i,j) = 0$. As previously, we stop the path at its first exit from A, i.e. we define $\hat{g} = (g_k)_{0 \leq k \leq r_g}$ where $r_g = \inf\{ k \leq n_g \mid g_k \notin A\}$. Since for all $k \leq r_g$, $\hat{g}_k \in \Pi_+$, we have

$$\tag{19} \begin{cases} n_{\hat{g}_k,A} \leq n_{i,A} \\ \text{et} \\ n_{\hat{g}_k,\hat{g}_{k+1}} \leq n_{i,A};\ 0 \leq k < r_g \end{cases}$$

Define $s_g = \sup\{0 \leq k < r_g \mid n_{\hat{g}_k,A} = n_{i,A} \}$. For all $k \leq s_g$, we have from (19) that $n_{\hat{g}_k,\hat{g}_{k+1}} \leq n_{i,A} = n_{\hat{g}_k,A} \leq n_{\hat{g}_k,\Pi_+}$ so that

$$C_A(\hat{g}_k, \hat{g}_{k+1}) = C_{\Pi_+}(\hat{g}_k, \hat{g}_{k+1}) = 0;\ 0 \leq k \leq s_g,$$

and $C_A^*(i, \hat{g}_{s_g+1}) = 0$. Now consider the two following cases. If $s_g+1 = r_g$, then the result is proved. Otherwise, applying the induction hypothesis, we have $C_A^*(\hat{g}_{s_g+1}, A^c) = 0$ so that

$$C_A^*(i, A^c) \leq C_A^*(i, \hat{g}_{s_g+1}) + C_A^*(\hat{g}_{s_g+1}, A^c) = 0.$$

The proof of the lemma is complete. □

LEMMA 3.5. – *Let* $\Pi \in \mathcal{C}(E) \setminus \{E\}$ *and* $i \in \Pi$.
Let $C_{\Pi\setminus\{i\}}^*(i, \Pi^c) = \inf\{ C_{\Pi\setminus\{i\}}^*(i,j) \mid j \in \Pi^c \}$. *We have*

$$C_{\Pi\setminus\{i\}}^*(i, \Pi^c) = H_e(\Pi) + W(\Pi) - W(i).$$

Remark 9. – We can give an heuristic proof of the above equality with the probabilistic interpretation of the renormalized costs. Consider the process at equilibrium for a constant cooling schedule at temperature T. Now, compute the order of the mass exiting from Π at each time step. This mass is given by the probability to be in Π which is of order of $e^{-W(\Pi)/T}$ (see proposition 1.6) multiplied by the probability by unit of time to escape from Π starting from Π which is of order $e^{-H_e(\Pi)/T}$. However, choosing a point $i \in \Pi$, we can say that this mass is given by the probability to be at i which is of order $e^{-W(i)/T}$ multiplied by the probability to escape from Π without any return to i which is of order $e^{-C^*_{\Pi\setminus\{i\}}(i,\Pi^c)/T}$. Hence we get $W(\Pi) + H_e(\Pi) = C^*_{\Pi\setminus\{i\}} + W(i)$.

Proof. – From the proposition 2.6 we know that there exists a constant A_E such that:

$$W(j) = A_E - \sum_{k=0}^{n_E}(H_e^k(j^k) - H_m^k(j^k)).$$

Assume that $j \in \Pi$, then

$$W(j) = A_E - \sum_{k=n_{j,\Pi}+1}^{n_E}(H_e^k(j^k) - H_m^k(j^k)) - \sum_{k=0}^{n_{j,\Pi}}(H_e^k(j^k) - H_m^k(j^k)).$$

It is true that $n_{j,\Pi}$ does not depend on j for $j \in \Pi$ so that we will omit the index j. Furthermore, since $j^{n_\Pi} = \Pi$, only the last term depends really on j for $j \in \Pi$. Hence for $f \in F(\Pi)$, we have

$$\sum_{k=0}^{n_\Pi}(H_e^k(f^k) - H_m^k(f^k)) = \sup_{j\in\Pi}\sum_{k=0}^{n_\Pi}(H_e^k(j^k) - H_m^k(j^k)) = H_e(\Pi).$$

Hence to prove the lemma it is sufficient to prove that

$$C^*_{\Pi\setminus\{i\}}(i,\Pi^c) = \sum_{k=0}^{n_\Pi}(H_e^k(i^k) - H_m^k(i^k)).$$

Let $j \in \Pi$. Since $n_{j,\Pi\setminus\{i\}} = n_{ij}$ we have

$$C_\Pi(j,l) - C_{\Pi\setminus\{i\}}(j,l) = -\sum_{k=n_{ij}\wedge n_{jl}+1}^{n_\Pi\wedge n_{jl}}(H_e^k(j^k) - H_m^k(j^k)).$$

However, for $k > n_{ij}$, we have $j^k = i^k$ so that

$$C_\Pi(j,l) - C_{\Pi\setminus\{i\}}(j,l) = -\sum_{k=n_{ij}\wedge n_{jl}+1}^{n_\Pi\wedge n_{jl}}(H_e^k(i^k) - H_m^k(i^k)).$$

Otherwise, if $n_{jl} > n_{ij}$, then $n_{il} = n_{jl}$. We deduce then that

$$C_\Pi(j,l) - C_{\Pi\setminus\{i\}}(j,l) = -\sum_{k=n_{ij}+1}^{n_{il}}(H_e^k(i^k) - H_m^k(i^k))\mathbf{1}_{n_{il}>n_{ij}}.$$

Hence for all paths g from i to Π^c through Π we have

$$C_{\Pi\setminus\{i\}}(g) \geq C_\Pi(g) + \sum_{k=0}^{n_\Pi}(H_e^k(i^k) - H_m^k(i^k)). \tag{20}$$

Moreover, there exists a path g from i to Π^c through Π such that $C_\Pi(g) = 0$ and n_{ig_k} is increasing in k. Hence

$$C_{\Pi\setminus\{i\}}(g) = \sum_{k=0}^{n_\Pi}(H_e^k(i^k) - H_m^k(i^k)). \tag{21}$$

We deduce from (20) and (21) that

$$C_{\Pi\setminus\{i\}}(i, \Pi^c) = \sum_{k=0}^{n_\Pi}(H_e^k(i^k) - H_m^k(i^k)).$$

so that the lemma is proved. □

LEMMA 3.6. – *Let* $\Pi \in \mathcal{C}(E)$, *there exists a real valued constant* A_Π *such that for each* $\Pi' \in \mathcal{M}_*(\Pi)$, *we have*

$$W(\Pi') + H_e(\Pi') = A_\Pi.$$

Remark 10. – The probabilistic interpretation of the above equality is that at equilibrium at constant temperature T, the mass exiting from each maximal proper sub-cycle Π' of Π is of the same order.

Proof. – Let $f \in F(\Pi')$. We have

$$\begin{aligned} W(\Pi') + H_e(\Pi') &= W(f) + \sum_{k=0}^{n_{f,\Pi'}}(H_e^k(f^k) - H_m(f^k)) \\ &= A_E - \sum_{k=n_{\Pi'}+1}^{n_E}(H_e^k(f^k) - H_m(f^k)). \end{aligned}$$

The result is proved if we notice that the last summation term does not depend on Π' but on Π. □

LEMMA 3.7. – *Let* $(i,j) \in E \times E$, $i \neq j$ *and consider a cycle* $\Pi \subset E \setminus \{j\}$, $i \in \Pi$. *Then we have*

$$W(i) + V(i,j) = W(\Pi) + H_e(\Pi) + C_\Pi(i,j).$$

Remark 11. – Here again we can give a heuristic probabilistic proof of the above equality if we consider the process at equilibrium at constant temperature T. The probability to visit the edge (i,j) is given by the probability to be at i which is of order $e^{-W(i)/T}$ multiplied by the probability of the transition from i to j which is of order $e^{-V(i,j)/T}$. However, considering the cycle Π this probability can be compute in another way. The probability is given by the probability to exit from Π which is of order $e^{-(W(\Pi)+H_e(\Pi))/T}$ multiplied by the probability to visit the edge (i,j) at the exit time which is of order $e^{-C_\Pi(i,j)/T}$. Hence we get $W(\Pi) + H_e(\Pi) + C_\Pi(i,j) = W(i) + V(i,j)$.

Proof. – From the definition of C_Π we get

$$W(i) + V(i,j) = W(i) + C_\Pi(i,j) + \sum_{k=0}^{n_{ij}\wedge n_{i,\Pi}} (H_e^k(i^k) - H_m(i^k)).$$

If $j \notin \Pi$, we have $n_{ij} \geq n_{i,\Pi}$ and

$$\begin{aligned} W(i) + \sum_{k=0}^{n_\Pi}(H_e^k(i^k) - H_m(i^k)) &= A_E - \sum_{n_\Pi+1}^{n_E} (H_e^k(i^k) - H_m(i^k)) \\ &= W(\Pi) + H_e(\Pi). \end{aligned}$$

The lemma is proved. □

4. THE LOCALIZATION THEOREMS

In Section 3, we have defined the main objects needed to perform a large deviation study of a G.S.A. In this section, we turn to a rigorous setting of the heuristic probabilistic interpretations of the cycle decomposition of E and of the renormalized communication costs. One of our main tasks will be to estimate the probability that starting from a point $i \in \Pi$, the Markov chain escape from Π at a point $j \in \Pi^c$ at a given time n. We will follow the original approach of O. Catoni [2] who has handled large deviation estimates for arbitrary decreasing cooling schedules.

Throughout this section, we will consider a fixed irreducible Markel kernel q on E and a fixed $\kappa \in [1, +\infty[$. We denote X a G.S.A. with parameter (q, κ, P) , $\mathcal{T}$ the underlying cooling schedule and V the communication cost.

4.1. Survival kernels

DEFINITION 4.1. – Let $H \geq 0, a > 0$ and $b \geq 0$ be three real valued numbers. Now consider a kernel on the integers $Q : \mathbb{Z} \times \mathbb{Z} \rightarrow [0,1]$ such that

- $Q_m^n = 0$ if $m > n$
- $(1-(1+b)\prod_{l=m+1}^{n-1}(1-ae^{-H/T_l})) \leq \sum_{k=m}^{n-1} Q_m^k \leq 1$ if $m < n$.

The set of all such kernels is called the set of the right survival kernels for (H,a,b) and denoted by $\mathcal{D}^r(H,a,b)$. In the same way, we define the set $\mathcal{D}^l(H,a,b)$ of the left survival kernels for (H,a,b) by

- $Q_m^n = 0$ if $m > n$
- $(1-(1+b)\prod_{l=m+1}^{n-1}(1-ae^{-H/T_l})) \leq \sum_{k=m+1}^{n} Q_k^n \leq 1$ if $m < n$.

We will take the convention that the product $\prod_{l=m+1}^{n-1}(\quad)$ is 0 if one of the term is negative.

Remark 12. – The survival kernels has been introduced in [2] and play a central role in our large deviation estimates. The right kernel will be used to give upper and lower bound to the probability that the Markov chain starting in a subset A at time n exit from A at time n. They follows approximately an inhomogeneous exponential law where the control is done on the partial sums.

We recall now the stability lemmas given in [2].

LEMMA 4.2. (Stability under product). – *Consider two real valued numbers $a > 0$ and $b \geq 0$. Let $Q \in \mathcal{D}^r(H,a,b)$ (respectively $\mathcal{D}^l(H,a,b)$), let $R \in \mathcal{D}^r(H,a,b)$ (respectively $\mathcal{D}^l(H,a,b)$). Then, there exist $a' > 0$ and $b' \geq 0$ which depend on a and b such that the product kernel QR is a element of $\mathcal{D}^r(H,a',b')$ (respectively $\mathcal{D}^l(H,a',b')$).*

LEMMA 4.3. (Stability under convex combination). – *Let $Q \in \mathcal{D}^r(H,a,b)$ (respectively $\mathcal{D}^l(H,a,b)$), let $R \in \mathcal{D}^r(H,a,b)$ (respectively $\mathcal{D}^l(H,a,b)$) and $\lambda \in [0,1]$. Then $\lambda Q + (1-\lambda)R \in \mathcal{D}^r(H,a,b)$ (respectively $\mathcal{D}^l(H,a,b)$.*

4.2. Localization kernels

Following Catoni, we define

DEFINITION 4.4. – (1) Let $A \subset E$. We define the stopping time $\tau(A,m)$ by

$$\tau(A,m) = \inf\{\ n > m\ , X_n \notin A\ \}.$$

(2) Let $A \subset E$, $B \subset E$. We define

$$M(A,B)_{i,m}^{j,n} = \begin{cases} P(\tau(A,m) \geq n,\ X_{n-1} \notin B,\ X_n = j \mid X_m = i) & m < n,\ i \in E,\ j \in B, \\ 0 & \text{otherwise,} \end{cases}$$

$$L(A,B)^{j,n}_{i,m} = \begin{cases} P(\tau(A,m) > n, X_n = j \mid X_m = i) & m \le n,\ i \in E,\ j \in B \\ 0 & \text{otherwise.} \end{cases}$$

The kernel $M(A,B)$ gives the probability to have an entrance at time n into the set B at j, starting from i at time m and traveling in A during the intermediate time. It will be used with $B \subset A^c$ and will describe the way that the Markov chain escape from A. On the other side, $L(A,B)$ gives the probability to be in j at time n, starting from i at time m and traveling in A.

4.3. Main large deviation estimates

THEOREM 4.5. – *There exist $a > 0$, $b \ge 0$, $c > 0$, $d > 0$, $K_1 > 0$, and $K_2 > 0$ which depend only on E, q and κ (and not on V) such that for all $n \in \mathbb{N}$, the induction hypothesis $\mathcal{H}^n(a,b,c,d,K_1,K_2)$ is true :*

$\mathcal{H}^n_1(a,b,c,d,K_2)$: For all $\Pi \in \mathcal{C}(E)$, $|\Pi| \le n$, for all $A \subset \Pi$, $A \ne \emptyset$ and all $i \in \Pi$:
there exists $Q \in \mathcal{D}^r(H_e(\Pi \setminus A), a, b)$ such that

$$M(\Pi \setminus A, (\Pi \setminus A)^c)^{j,n}_{i,m} \le K_2 e^{-C^*_{\Pi\setminus A}(i,j)/T_{m+1}} Q^n_m.$$

Moreover $P(\tau(\Pi,m) > n \mid X_m = i\ ) \ge c\prod_{l=m+1}^n (1 - de^{-H_e(\Pi)/T_l})$

$\mathcal{H}^n_2(a,b,K_1)$: For all $A \subset E$, $|A| \le n$, all $i \in A$, all $j \in A^c$:
there exists $Q \in \mathcal{D}^r(H_e(A), a, b)$ such that

$$M(A, A^c)^{j,n}_{i,m} \ge K_1 e^{-C^*_A(i,j)/T_n} Q^n_m.$$

Moreover we have

$$P(\tau(A,m) > n \mid X_m = i\ ) \le (1+b) \prod_{l=m+1}^{n} (1 - ae^{-H_e(A)/T_l}).$$

$\mathcal{H}^n_3(a,b,K_2)$: For all $\Pi \in \mathcal{C}(E)$, $|\Pi| \le n$ and all $i \in \Pi$:
there exists $Q \in \mathcal{D}^r(H_e(\Pi), a, b)$ such that

$$M(\Pi, \Pi^c)^{j,n}_{i,m} \le K_2 e^{-C^*_\Pi(i,j)/T_{m+1}} Q^n_m.$$

$\mathcal{H}^n_4(a,b,K_2)$: For all $A \subset E$, $|A| \le n$, all $\Pi \in \mathcal{M}(A)$ and all $i \in A$:
there exists $Q \in D^r(H_e(A), a, b)$ such that

$$M(A, \Pi \cup A^c)^{j,n}_{i,m} \le K_2 e^{-C^*_A(i,j)/T_{m+1}} Q^n_m.$$

Remark 13. – This first theorem collects all the basic large deviation estimates needed for the study of optimal cooling schedules. These estimates will be established by induction on the size of the considered subset starting from obvious lower and upper bounds on the singletons. This is the approach proposed by O. Catoni who establishes in [2] a similar theorem for the sequential simulated annealing. We will follow the sketch of his proof. However, introducing the renormalized communication cost, we will be able to give large deviation estimates for the exit time of arbitrary subsets and for generalized simulated annealing. This theorem shows that the large deviation estimates can be obtained with uniform constants even in the time inhomogeneous setting. The uniformity of the constants will be verified step by step with the help of lemma 4.2 and lemma 4.3.

Remark 14. – Roughly speaking, the principle of the proof will be to split the trajectories of the process into pieces for which we have large deviations estimates. Combining all these estimates, we will get large deviation estimates for the whole process.

We will give a meaningful example of this approach with the upper bound of the exit flow of a cycle Π given by $\mathcal{H}^3$. Since the process spends most of its time in the configurations of lowest energies, we split the trajectory into two pieces: the part of the trajectory before its last visit to a particular point f in the bottom of Π and the remaining part of the trajectory. Hence, on one hand, we have to compute a bound for the probability to join j at a given time n without escaping from Π (this is done by the second part of $\mathcal{H}^2$ taking $A = \Pi$) and on the other hand, we have to study the probability that the process starting from $f \in F(\Pi)$ escapes from Π through a given exit point j without return to f. This last part is given by $\mathcal{H}^1$ with $A = \{f\}$.

For our induction, we will have to consider more general situations leading to the quite complicated four statements of the theorem. These statements have to be considered has the whole in the proof since they all interact during the proof. However, only statements $\mathcal{H}^1$ and $\mathcal{H}^2$ will be used in our further results.

Proof. – Consider $a = \lambda/\kappa$, $b = 0$, $K_1 = \lambda/\kappa^2$ and $K_2 = \kappa^2/\lambda$ where

$$\lambda = \inf\{\ q(i,j) \mid q(i,j) > 0;\ (i,j) \in E \times E\ \}.$$

Now consider $(c,d) = (1,\kappa)$, we will prove that $\mathcal{H}^1(a,b,c,d,K_1,K_2)$ is true.

$\mathcal{H}_1$: Consider a cycle $\Pi = \{i\}$ and a subset $A \subset \Pi$. Since we assume that A is not empty, we have $A = \Pi$ so that $\Pi \setminus A = \emptyset$. Then

$$M(\Pi \setminus A, (\Pi \setminus A)^c)_{i,m}^{j,n} = M(\emptyset, E)_{i,m}^{j,n} = 0.$$

We consider now the exit time from Π. A straightforward computation gives

$$P(\tau(\Pi, m) > n \mid X_m = i) \geq \prod_{l=m+1}^{n} (1 - \kappa e^{-H_e(i)/T_{m+1}}),$$

with the convention that the product is 0 if it contains a negative term. Hence the proof of $\mathcal{H}_1$ is complete.

$\mathcal{H}_2$: Consider $A = \{i\}$. Then we have

$$M(A, A^c)_{i,m}^{j,n} \geq \frac{\lambda}{\kappa^2} e^{-C_A^*(i,j)/T_n} Q_m^n,$$

where $Q_m^n = \begin{cases} 0 & \text{if } n \leq m, \\ \prod_{l=m+1}^{n-1}(1 - \kappa e^{-H_e(i)/T_l})(\kappa e^{-H_e(i)/T_n} \wedge 1) & \text{otherwise.} \end{cases}$

Since $H_e(i) = H_e(A)$, one easily proves that

$$Q \in \mathcal{D}^r(H_e(A), a, b).$$

Now, concerning the exit time from A, we have

$$P(\tau(A, m) > n \mid X_m = i\ ) \leq \prod_{l=m+1}^{n} (1 - (\lambda/\kappa) e^{-H_e(i)/T_l}),$$

so that the proof of $\mathcal{H}_2$ is complete.

$\mathcal{H}_3$: Consider a cycle $\Pi = \{i\}$. There exists $i_1 \in E$ such that $H_e(i) = V(i, i_1)$. Hence

$$\begin{aligned} P(X_{m+1} = i \mid X_m = i\ ) \leq 1 - P(\ & X_{m+1} \\ & = i_1 \mid X_m = i\ ) \leq 1 - (\lambda/\kappa) e^{-H_e(i)/T_{m+1}}, \end{aligned}$$

and

$$M(\Pi, \Pi^c)_{i,m}^{j,n} \leq (\kappa^2/\lambda) e^{-(V(i,j) - H_e(i))/T_{m+1}} Q_m^n,$$

where $Q_m^n = \begin{cases} \prod_{l=m+1}^{n-1}(1 - (\lambda/\kappa) e^{-H_e(i)/T_l})(\lambda/\kappa) e^{-H_e(i)/T_n} & \text{if } n > m, \\ 0 & \text{otherwise.} \end{cases}$

Since $H_e(i) = H_e(\Pi)$, one easily verifies that

$$Q \in \mathcal{D}^r(H_e(\Pi), \lambda/\kappa, 0) \subset \mathcal{D}^r(H_e(\Pi), a, b).$$

$\mathcal{H}_4$: Consider a cycle $\Pi = \{i\}$. Then $\mathcal{M}(\Pi) = \{i\}$ and $M(\{i\}, E) = 0$ so that $\mathcal{H}_4$ is proved.

The proof of $\mathcal{H}^1$ is now completed.

Assume that there exists $\gamma = (a, b, c, d, K_1, K_2)$ such that $\mathcal{H}^{n-1}(\gamma)$ is true. We will show that there exists $\gamma' = (a', b', c', d', K_1', K_2')$ such that $\mathcal{H}^n(\gamma')$ is true. Moreover, we should verify that γ' can be deduced from γ by a function which depends only on E, q and κ. This verification will be done explicitly in the proof of $\mathcal{H}_1^n$ and will be left to the reader in the remaining cases. Furthermore, we should notice that it is sufficient to prove the weaker result:

There exist $(a_i', b_i')_{1\le i\le 4}$, K_1', K_2', K_3', K_4', c' and d' such that $\mathcal{H}_1^n(a_1', b_1', c', d', K_1')$, $\mathcal{H}_2^n(a_2', b_2')$, $\mathcal{H}_3^n(a_3', b_3', K_3')$ and $\mathcal{H}_4^n(a_4', b_4', K_4')$ are true. Indeed, if we consider $a = \inf a_i'$, $b = \sup b_i'$, $K_1 = K_2'$ and $K_2 = \sup\{K_1', K_3', K_4'\}$, then $\mathcal{H}^n(a, b, c, d, K_1, K_2)$ is true.

We begin now the proof of $\mathcal{H}^n$.

$\mathcal{H}_1^n$: Let us consider the family $(\Pi_k)_{0\le k\le p}$ of the element of $\mathcal{M}_*(\Pi)$. Assume that $i \in \Pi_0$ so that

$$M(\Pi\setminus A, (\Pi\setminus A)^c)_{i,m}^{j,n} \le M_1 + M_2 + M_3,$$

where

$$\begin{aligned}
M_1 &= M(\Pi_0\setminus A, (\Pi_0\setminus A)^c)_{i,m}^{j,n},\\
M_2 &= \sum_{k=1}^{p}\{M(\Pi_0\setminus A, \Pi\setminus(A\cup\Pi_0))M(\Pi_k\setminus A, (\Pi_k\setminus A)^c)\}_{i,m}^{j,n},\\
M_3 &= \sum_{k=1}^{p}\{M(\Pi_0\setminus A, \Pi\setminus(A\cup\Pi_0))M(\Pi\setminus A, \Pi_k\setminus A)M(\Pi_k\setminus A,\\
&\quad\times(\Pi_k\setminus A)^c)\}_{i,m}^{j,n}.
\end{aligned}$$

Since $C^*_{\Pi_0\setminus A}(i,j) \ge C^*_{\Pi\setminus A}(i,j)$ we deduce from $\mathcal{H}_1^{n-1}(a, b, c, d, K_2)$ that

$$M_1 \le K_2 e^{-C^*_{\Pi\setminus A}(i,j)/T_{m+1}} Q_m^n, \tag{22}$$

with $Q \in \mathcal{D}^r(H_e(\Pi_0\setminus A), a, b)$.

Concerning M_2, we have the expansion

$$\begin{aligned}
M_2 = \sum_{k=1}^{p}\sum_{i_1\in\Pi\setminus(A\cup\Pi_0)}\sum_{m\le l\le n} & M(\Pi_0\setminus A, \Pi\setminus(A\cup\Pi_0))_{i,m}^{i_1,l}M(\Pi_k\setminus A,\\
&\times(\Pi_k\setminus A)^c)_{i_1,l}^{j,n}.
\end{aligned}$$

However, we deduce from $\mathcal{H}_1^{n-1}(a, b, c, d, K_2)$ and from the lemma 4.2 that there exists $(a_1, b_1, K_2^{(1)})$ which depends only on (a, b, K_2) and

$Q \in \mathcal{D}^r(H_e(\Pi \setminus A), a_1, b_1)$ such that

$$(23)\qquad \sum_{m \le l \le n} M(\Pi_0 \setminus A, \Pi \setminus (A \cup \Pi_0))^{i_1,l}_{i,m} M(\Pi_k \setminus A, (\Pi_k \setminus A)^c)^{j,n}_{i_1,l} \le K_2^{(1)} e^{-(C^*_{\Pi_0\setminus A}(i,i_1)+C^*_{\Pi_k\setminus A}(i_1,j))/T_{m+1}} Q^n_m.$$

Since $C^*_{\Pi_0\setminus A}(i,i_1) + C^*_{\Pi_k\setminus A}(i_1,j) \ge C^*_{\Pi\setminus A}(i,j)$, and adding the family of inequalities (23) for all the values of k and i_1, we deduce from lemma 4.3 that there exists $(a_2, b_2, K_2^{(2)})$ which depends only on the size of E and on (a, b, K_2) such that

$$(24)\qquad M_2 \le K_2^{(2)} e^{-C^*_{\Pi\setminus A}(i,j)/T_{m+1}} Q^n_m,$$

with $Q \in \mathcal{D}^r(H_e(\Pi \setminus A), a_2, b_2)$.

Concerning M_3, let us consider the family $(\Pi_{k,s})_{0\le s\le r_k}$ of the elements of $\mathcal{M}(\Pi \setminus A)$ such that $\Pi_k = \bigcup_{0\le s\le r_k} \Pi_{k,s}$. We have

$$(25)\qquad M_3 \le \sum_{k=0}^{p} \sum_{s=0}^{r_k} \sum_{i_1\in\Pi\setminus(A\cup\Pi_0)} \sum_{i_2\in\Pi_{k,s}} M_3(k,s,i_1,i_2),$$

where

$$M_3(k,s,i_1,i_2) =$$

$$\sum_{m\le l_1\le l_2\le n} M(\Pi \setminus A, (\Pi_0 \setminus (A \cup \Pi_0))^{i_1,l_1}_{i,m} M(\Pi \setminus A, \Pi_{k,s})^{i_2,l_2}_{i_1,l_1} M(\Pi_k \setminus A, \times (\Pi_k \setminus A)^c)^{j,n}_{i_2,l_2}.$$

Since $A \neq \emptyset$, we have $|\Pi \setminus A| \le n-1$ so that we deduce from $\mathcal{H}_1^{n-1}(a,b,c,d,K_2)$, $\mathcal{H}_3^{n-1}(a,b,K_2)$, $\mathcal{H}_4^{n-1}(a,b,K_2)$ and from lemma 4.3 that there exists (a_3, b_3, K_3) which depends only on (a, b, K_2) and on the size of E such that

$$(26)\qquad M_3 \le K_3 e^{-[C^*_{\Pi_0\setminus A}(i,i_1)+C^*_{\Pi\setminus A}(i_1,i_2)+C^*_{\Pi_k\setminus A}(i_2,j)]/T_{m+1}} Q^n_m,$$

with $Q \in D^r(H_e(\Pi \setminus A), a_3, b_3)$. However, we have for $i_1 \in \Pi \setminus (A \cup \Pi_0)$ and $i_2 \in \Pi_k \setminus A$

$$C^*_{\Pi_0\setminus A}(i,i_1) + C^*_{\Pi\setminus A}(i_1,i_2) + C^*_{\Pi_k\setminus A}(i_2,j) \ge C^*_{\Pi\setminus A}(i,j),$$

so that we deduce

$$M_3 \leq K_3 e^{-C^*_{\Pi\setminus A}(i,j)/T_{m+1}} Q^n_m. \tag{27}$$

Now, we deduce from (22), (24) and (27) that there exists (a_4, b_4, K_4) which depends only on (a, b, K_2) and on the size of E such that

$$M(\Pi \setminus A, (\Pi \setminus A)^c)^{j,n}_{i,m} \leq K_4 e^{-C^*_{\Pi\setminus A}(i,j)/T_{m+1}} Q^n_m,$$

with $Q \in \mathcal{D}^r(H_e(\Pi \setminus A), a_4, b_4)$. This ends the proof of the first inequality of $\mathcal{H}^n_1$.

From now, we leave to the reader, the easy verifications which prove that the parameters can always be chosen independently of V.

We are here concerned by the study of the exit time from Π. Since we can take $d \geq 1$, we will assume that $T_n > 0$ and $H_e(\Pi) > 0$ (otherwise the result is trivial). Let $f \in F(\Pi)$, considering the last visit of f by the Markov chain, we have the inequality

$$P(\tau(\Pi, m) \leq n \mid X_m = i) \leq \sum_{l=m+1}^{n} M(\Pi \setminus \{f\}, \Pi^c)^{E,l}_{i,m}$$
$$+ \sum_{k=m+1}^{n-1} \sum_{l=k+1}^{n} P(X_k = f,\ \tau(\Pi, m) > k \mid X_m = i) M(\Pi \setminus \{f\}, \Pi^c)^{E,l}_{f,k}.$$

For a fixed n, consider the new cooling schedule $(\hat{T}_l)_{l\in\mathbb{N}}$ defined by $\hat{T}_l = T_l$ for $l < n$ and $\hat{T}_l = T_n$ for $l \geq n$. For this new cooling schedule, we get from $\mathcal{H}^{n-1}_2$ (since we have assume $T_n > 0$ and $H_e(\Pi) > 0$)

$$\sum_{l=m+1}^{\infty} M(\Pi \setminus \{f\}, \{f\})^{f,l}_{i,m} \geq K_1,$$

so that

$$\sum_{l=m+1}^{\infty} M(\Pi \setminus \{f\}, \Pi^c)^{E,l}_{i,m} \leq 1 - K_1. \tag{28}$$

Then, we deduce from the proof of the first inequality of $\mathcal{H}^n_1$ that

$$P(\tau(\Pi, m) \leq n \mid X_m = 1) \leq 1 - K_1 + K_2 \sum_{k=m+1}^{n} e^{-H_e(\Pi)/T_k},$$

so that

$$(29)\qquad P(\tau(\Pi,m) > n \mid X_m = 1\ ) \geq K_1 - K_2 \sum_{k=m+1}^{n} e^{-H_e(\Pi)/T_k}.$$

The lemma 4.6 given below shows that the result follows from inequality (29). Hence the proof of $\mathcal{H}_1^n$ is completed.

LEMMA 4.6. – *Let* $\gamma \in]0,1]$ *and* $A \subset E$. *Assume that for all* $n > m$, *all* $i \in E$ *we have*

$$P(\tau(A,m) > n \mid X_m = i) \geq \gamma - K \sum_{k=m+1}^{n} e^{-V(A)/T_k},$$

then, there exist c *and* d *which depend only on* γ *and* K *such that*

$$P(\tau(A,m) > n \mid X_m = i) \geq c \prod_{k=m+1}^{n} (1 - de^{-V(A)/T_k}).$$

Proof. – This lemma has been established by O. Catoni in [2]. For completeness, we report here the proof.

Assume first that $e^{-V(A)/T_{m+1}} \leq \gamma/(4K)$, then consider the family (u_k) defined by

- $u_0 = m$,
- $u_{k+1} = \inf\{l > u_k \mid \sum_{h=u_k+1}^{l} e^{-V(A)/T_h} \geq \gamma/(4K)\ \}$.

If $u_k < \infty$, we have

$$\inf_{i\in\Pi} P(\tau(A,u_{k-1}) > u_k \mid X_{u_{k-1}} = i\ ) \geq \gamma/2.$$

However,

$$\prod_{h=u_{k-1}+1}^{u_k} (1 - ce^{-V(A)/T_h}) \leq \exp\left(-c \sum_{h=u_{k-1}+1}^{u_k} e^{-V(A)/T_h}\right).$$

Hence, if $c = \sup\{-\ln(\gamma/2)4K/\gamma, 4K/\gamma\}$, we have

$$\prod_{h=u_{k-1}+1}^{u_k} (1 - ce^{-V(A)/T_h}) \leq \gamma/2.$$

Now, for each $n \in \mathbb{N}^*$, consider the integer $k(n)$ such that $u_{k(n)} < n \leq u_{k(n)+1}$. We have

$$\begin{aligned}
&P(\tau(A, m) > n \mid X_m = i) \\
&\quad \geq (\gamma/2)^{k(n)-1} \inf_{i' \in \Pi} P(\tau(A, u_{k(n)}) > n \mid X_{u_{k(n)}} = i') \\
&\quad \geq \prod_{h=m+1}^{u_{k(n)}} (1 - ce^{-V(A)/T_h}) \left(\gamma - K \sum_{l=u_{k(n)}+1}^{n} e^{-V(A)/T_l} \right) \\
&\quad \geq (\gamma/2) \prod_{h=m+1}^{n} (1 - ce^{-V(A)/T_h}).
\end{aligned}$$

Assume now that $e^{-V(A)/T_{m+1}} > \gamma/(4K)$. since we have $c \geq (4K)\gamma$, we deduce $(1 - ce^{-V(A)/T_{m+1}}) < 0$ so that we have the result with the convention that the product is zero if one of its term has a negative value.

The proof of the lemma is complete. □

$\mathcal{H}_n(2)$: Let $g \in \mathrm{Pth}_E(i, j)$ such that $C_A(g) = C_A^*(i, j)$. As in the proof of $\mathcal{H}_1^n$, we decompose the graph g into its exit points out of the elements of $\mathcal{M}(A)$. More precisely we consider the unique family $(i_k, n_k, \Pi_k)_{0 \leq k \leq r}$ define by the following procedure :

• $i_0 = g_0 = i$, $n_0 = 0$, and Π_0 is the unique element of $\mathcal{M}(A)$ such that $i_0 \in \Pi_0$.

• Let $n_{k+1} = \inf\{ l > n_k \mid g_l \notin \Pi_k \}$ and $i_{k+1} = g_{n_{k+1}}$.

– If $i_{k+1} = j$ then the construction is completed.

– Otherwise, define Π_{k+1} in $\mathcal{M}(A)$ such that $i_{k+1} \in \Pi_{k+1}$.

Let r be the integer such that $i_r = j$.

First case: Assume that $A \notin \mathcal{C}(E)$, then for all $\Pi \in \mathcal{M}(A)$, all $i' \in \Pi$ and all $j' \in E$ we have

$$C_\Pi(i', j') = C_A(i', j'), \tag{30}$$

so that we deduce from the previous construction that

$$\begin{aligned}
C_A^*(i, j) = C_A(g) &= \sum_{k=0}^{r-1} \sum_{n_k \leq l < n_{k+1}} C_{\Pi_k}(g_l, g_{l+1}) \\
&\geq \sum_{k=0}^{r-1} C_{\Pi_k}^*(i_k, i_{k+1}).
\end{aligned} \tag{31}$$

Furthermore, we have

$$M(A, A^c)_{i,m}^{j,n} \geq \left\{ \prod_{k=0}^{r-1} M(\Pi_k, \Pi_k^c)_{i_k}^{i_{k+1}} \right\}_m^n .$$

Since the cycles Π_k verify $|\Pi_k| < n$, we deduce from $\mathcal{H}_2^{n-1}$ and from the stability lemmas that the right hand term is bounded from below by

$$K_1 e^{-\sum_{k=0}^{r-1} C^*_{\Pi_k}(i_k, i_{k+1})/T_n} Q_m^n,$$

with $Q \in \mathcal{D}^r(H_e(A), a_1, b_1)$. The result follows from inequality (31).

Second case: We assume now that A is a cycle Π. The main difference with the previous case is that the equality (30) is true only for $j' \in \Pi$. Hence the equation (31) must be changed in

$$\text{(32)} \quad C^*_\Pi(i,j) = C_\Pi(g) \geq \sum_{k=0}^{r-1} C^*_{\Pi_k}(i_k, i_{k+1}) \\ + C_\Pi(g_{n_r-1}, j) - C_{\Pi_{r-1}}(g_{n_r-1}, j).$$

Since one easily verifies that $C_\Pi(g_{n_r-1}, j) - C_{\Pi_{r-1}}(g_{n_r-1}, j) = H_m(\Pi) - H_e(\Pi)$, we deduce in the same way that above :

$$\text{(33)} \qquad M(\Pi, \Pi^c)_{i,m}^{j,n} \geq K_1 e^{-C^*_\Pi(i,j)/T_n} e^{-(H_e(\Pi) - H_m(\Pi))/T_n} Q_m^n,$$

with $Q \in \mathcal{D}^r(H_m(\Pi), a_2, b_2)$.

Define now for all $H \geq 0$, $c > 0$ and $m \in \mathbb{N}$

$$R(H, c, m) = \inf \left\{ l > m \mid \sum_{k=m+1}^{l} e^{-H/T_k} \geq c \right\}.$$

We define the sequence $(u_k)_{k \in \mathbb{N}}$ by:

- $u_0 = m$,
- $u_{k+1} = R(H_m(\Pi), \ln(2(1+b))/a, u_k)$ for $k \geq 0$.

For all $n > m$, we define $k(n) \in \mathbb{N}$ by $u_{k(n)} < n \leq u_{k(n)+1}$. We deduce then from the construction of u_k that

$$M(\Pi, \Pi^c)_{i,m}^{j,n} \geq P(\tau(\Pi, m) > u_{k(n)} \mid X_m = i) \\ \times \sum_{i_1 \in \Pi} P(X_{u_{k(n)}} = i_1 \mid \tau(\Pi, m) > u_{k(n)} \ , \ X_m = i) M(\Pi, \Pi^c)_{i_1, u_{k(n)}}^{j,n}.$$

From lemma 3.4, we deduce that $C^*_\Pi(i_1, j)$ is independent of $i_1 \in \Pi$ so that we can defined $C(\Pi, j)$ by:

$$C(\Pi, j) = C^*_\Pi(i_1, j);\ i_1 \in \Pi.$$

Hence, we deduce from the previous inequality and from (33) that:

$$\text{(34)}\quad M(\Pi, \Pi^c)^{j,n}_{i,m} \geq K_1 P(\tau(\Pi, m) > u_{k(n)} \mid X_m = i\ ) e^{-(C(\Pi,j)+H_e(\Pi)-H_m(\Pi))/T_n} Q^n_{u_{k(n)}},$$

with $Q \in \mathcal{D}^r(H_m(\Pi), a, b)$. Since we have here $|\Pi| \geq 2$, we get easily that $P(\tau(\Pi, m) > m + 1 \mid X_m = i\ ) \geq 1 - \chi$ où $\chi = 1 - \lambda/\kappa$. Hence we deduce from $\mathcal{H}^n_1$ that

$$P(\tau(\Pi, m) > u_{k(n)} \mid X_m = i\ ) \geq c \prod_{l=m+1}^{u_{k(n)}} (1 - \chi \wedge (de^{-H_e(\Pi)/T_l})),$$

and we get from (34)

$$M(\Pi, \Pi^c)^{j,n}_{i,m} \geq K_1 c \prod_{l=m+1}^{u_{k(n)}} (1 - \chi \wedge (de^{-H_e(\Pi)/T_l})) \times e^{-(C(\Pi,j)+H_e(\Pi)-H_m(\Pi))/T_n} Q^n_{u_{k(n)}}.$$

Define now A^n_m by:

$$A^n_m = \prod_{l=m+1}^{u_{k(n)}} (1 - \chi \wedge (de^{-H_e(\Pi)/T_l})) e^{-(H_e(\Pi)-H_m(\Pi))/T_n} Q^n_{u_{k(n)}}.$$

For proving $\mathcal{H}^n_2$, we have to prove that there exist a', b', K'_1 such that

$$\text{(35)}\qquad K'_1(1 - (1 + b') \prod_{l=m+1}^{n} (1 - a' e^{-H_e(\Pi)/T_l})) \leq \sum_{l=m}^{n} A^l_m \leq K'_1$$

We consider first the lower bound. We have

$$\begin{aligned}\text{(36)}\qquad \sum_{l=m}^{n} A^l_m &\geq \sum_{l=m}^{u_{k(n)}} A^l_m \\ &\geq \sum_{k=0}^{k(n)-1} \prod_{l=m+1}^{u_k} (1 - \chi \wedge (de^{-H_e(\Pi)/T_l})) \\ &\quad \times \sum_{h=u_k+1}^{u_{k+1}} e^{-(H_e(\Pi)-H_m(\Pi))/T_h} Q^h_{u_k}.\end{aligned}$$

Furthermore,

$$\sum_{h=u_k+1}^{u_{k+1}} e^{-(H_e(\Pi)-H_m(\Pi))/T_h} Q_{u_k}^h \geq e^{-(H_e(\Pi)-H_m(\Pi))/T_{u_{k+1}}} \sum_{h=u_k+1}^{u_{k+1}} Q_{u_k}^h.$$

Since

$$\sum_{h=u_k+1}^{u_{k+1}} Q_{u_k}^h \geq (1-(1+b)\prod_{u_k+1}^{u_{k+1}}(1-ae^{-H_m(\Pi)/T_h})),$$

and

$$\prod_{h=u_k+1}^{u_{k+1}} (1-ae^{-H_m(\Pi)/T_h}) \leq e^{-a\sum_{h=u_k+1}^{u_{k+1}} e^{-H_m(\Pi)/T_h}},$$

we deduce from the definition of u_n that

$$\sum_{h=u_k+1}^{u_{k+1}} Q_{u_k}^h \geq 1/2.$$

However,

$$\frac{e^{-H_e(\Pi)/T_{u_{k+1}}}}{e^{-H_m(\Pi)/T_{u_{k+1}}}} \geq \left(\sum_{h=u_{k+1}+1}^{u_{k+2}} e^{-H_e(\Pi)/T_h}\right) \Big/ \left(\sum_{h=u_{k+1}+1}^{u_{k+2}} e^{-H_m(\Pi)/T_h}\right),$$

so that we deduce from

$$\sum_{h=u_{k+1}+1}^{u_{k+2}} e^{-H_m(\Pi)/T_h} \leq \ln(2(1+b))/a+1,$$

that

$$\sum_{h=u_k+1}^{u_{k+1}} e^{-(H_e(\Pi)-H_m(\Pi))/T_h} Q_{u_k}^h \geq K \sum_{h=u_{k+1}+1}^{u_{k+2}} e^{-H_e(\Pi)/T_h},$$

with $K = (\ln(2(1+b))/a+1)/2$. Coming back to inequality (36), we get

$$\sum_{l=m}^{n} A_m^l \geq K \sum_{k=0}^{k(n)-1} \prod_{l=m+1}^{u_k} (1-\chi \wedge (de^{-H_e(\Pi)/T_l})) \sum_{h=u_{k+1}+1}^{u_{k+2}} e^{-H_e(\Pi)/T_h}.$$

Consider now $u_{k+1}+1 \leq h \leq u_{k+2}$, we have:

$$\begin{aligned}
&\prod_{l=m+1}^{u_k} (1-\chi\wedge(de^{-H_e(\Pi)/T_l}))e^{-H_e(\Pi)/T_h} \\
&\geq \left\{\prod_{l=u_1+1}^{h-1} (1-\chi\wedge(de^{-H_e(\Pi)/T_l}))e^{-H_e(\Pi)/T_h}\right\} \\
&\quad\times\left\{\prod_{l=m+1}^{u_1} (1-\chi\wedge(de^{-H_e(\Pi)/T_l}))\right\} \\
&\geq \frac{K'}{d}\prod_{l=u_1+1}^{h-1} (1-\chi\wedge(de^{-H_e(\Pi)/T_l}))\chi\wedge(de^{-H_e(\Pi)/T_h}),
\end{aligned}$$

Indeed, for all $0 \leq x \leq \chi$, we have $(1-x) \geq e^{\ln(1-\chi)x/\chi}$ so that

$$\begin{aligned}
\prod_{l=m+1}^{u_1} (1-\chi\wedge(de^{-H_e(\Pi)/T_l})) &\geq \exp\left(\frac{\ln(1-\chi)}{\chi}\sum_{l=m+1}^{u_1} de^{-H_e(\Pi)/T_l}\right) \\
&\geq \exp\left(\frac{\ln(1-\chi)}{\chi}\sum_{l=m+1}^{u_1} de^{-H_m(\Pi)/T_l}\right) \geq K',
\end{aligned}$$

where $K' = \exp\left(\frac{\ln(1-\chi)}{\chi}d(\ln(2(1+b)\right)/a+1)$.

Therefore,

$$\begin{aligned}
\sum_{l=m}^{n} A_m^l &\geq \frac{K'}{d}\left(1-\prod_{l=u_1+1}^{u_{k(n)+1}} (1-\chi\wedge(de^{-H_e(\Pi)/T_l}))\right) \\
&\geq \frac{K'}{d}\left(1-\frac{\left\{\prod_{l=m+1}^{n}(1-\chi\wedge(de^{-H_e(\Pi)/T_l}))\right\}}{\left\{\prod_{l=m+1}^{u_1}(1-\chi\wedge(de^{-H_e(\Pi)/T_l}))\right.}\right).
\end{aligned}$$

Hence, defining $b' = (1/K')-1$ and $K_1' = K'/d$, we have

$$\begin{aligned}
\sum_{l=m}^{n} A_m^l &\geq \frac{K}{d}(1-(1+b')\prod_{l=m+1}^{n} (1-\chi\wedge(de^{-H_e(\Pi)/T_l}))) \\
&\geq K_1'(1-(1+b')\prod_{l=m+1}^{n} (1-(\chi\wedge d)e^{-H_e(\Pi)/T_l}))).
\end{aligned}$$

To establish the upper bound in (35), it is sufficient to consider for all $m \in \mathbb{Z}$, the integer $n(m)$ defined by $n(m) = \inf\{ k \in \mathbb{Z} \mid \sum_{l=m}^{k} A_m^l > K_1' \}$. Then, we consider $\hat{A}_m^n$ defined by $\hat{A}_m^n = A_m^n$ for $n < n(m)$, and if $n(m) < +\infty$, we define $\hat{A}_m^n = K_1' - \sum_{l=m}^{n(m)-1} A_m^n$ and $\hat{A}_m^k = 0$ if $k > n(m)$. We can verify that

$$M(\Pi, \Pi^c)_{i,m}^{i,n} \geq \hat{A}_m^n$$

and $\hat{A}_m^n$ satisfies the inequalities (35). This ends the proof of the first inequality of $\mathcal{H}_2^n$.

We prove now the last assertion of $\mathcal{H}_2^n$. We do not any longer assume that A is a cycle. Let i be in A, from lemma 3.4 we deduce that there exists $j \in A^c$ so that $C_A^*(i,j) = 0$. Hence

$$\begin{aligned}(37)\quad P(\tau(A,m) > n \mid X_m = i) &= 1 - P(\tau(A,m) \leq n \mid X_m = i\ ) \\ &\leq 1 - \sum_{k=m+1}^{n} Q_m^k \\ &\leq 1 - K_1(1 - (1+b) \prod_{l=m+1}^{n} (1 - a e^{-H_e(A)/T_l}).\end{aligned}$$

The proof is completed as in [2] but we recall it for sake of completeness. Consider the family (u_k) defined by

- $u_0 = m$.
- $u_k = \inf\left\{ l > u_{k-1} \mid \sum_{h=u_{k-1}+1}^{l} e^{-H_e(A)/T_h} \geq \ln(2(1+b))/a \right\}$.

From (37) we deduce that for $u_{k+1} < \infty$

$$\sup_{i \in \Pi} P(\tau(A, u_k) > u_{k+1} \mid X_{u_k} = i) \leq 1 - K_1/2.$$

However, assuming that $a' \leq 1/2$, we have

$$\prod_{l=u_k+1}^{u_{k+1}} (1 - a' e^{-H_e(A)/T_l}) \geq \exp\left(2\ln(1/2) a' \sum_{l=u_k+1}^{u_{k+1}} e^{-H_e(A)/T_l} \right).$$

Now, assuming

$$a' = \inf\{1/2, \ln(1 - K_1/2)/[(\ln(2(1+b))/a + 1)(2\ln(1/2))]\}$$

we deduce

$$\prod_{l=u_k+1}^{u_{k+1}} (1 - a' e^{-H_e(A)/T_l}) \geq 1 - K_1/2.$$

Hence, using the Markov property at time u_k, we deduce

$$P(\tau(A,m) > n \mid X_m = i) \leq \frac{1}{1 - K_1/2} \prod_{l=m+1}^{n} (1 - a' e^{-H_e(A)/T_l}),$$

from which the proof of $\mathcal{H}_2^n$ is completed.

$\mathcal{H}_3^n$: Let $f \in F(\Pi)$ and consider the last visit of f by the Markov chain. We can write

$$M(\Pi, \Pi^c)_{i,m}^{j,n} \leq M(\Pi \setminus \{f\}, \Pi^c)_{i,m}^{j,n}$$

$$+ \sum_{k=m}^{n-1} P(\tau(\Pi, m) > k \ , \ X_k = f \mid X_m = i\ ) M(\Pi \setminus \{f\}, \Pi^c)_{f,k}^{j,n}.$$

From $\mathcal{H}_1^{n-1}$ we deduce that the first term of the inequality has the upper bound

$$\begin{aligned} (38) \qquad & M(\Pi \setminus \{f\}, \Pi^c)_{i,m}^{j,n} \\ & \leq K_2 e^{-C_\Pi^*(i,j)/T_{m+1}} Q_m^n \text{ with } Q \in \mathcal{D}^r(H_e(\Pi), a, b). \end{aligned}$$

Now, considering the second term, we get from $\mathcal{H}_2^n$ and $\mathcal{H}_1^{n-1}$

$$\begin{aligned} (39) \quad & \sum_{k=m}^{n-1} P(\tau(\Pi, m) > k \ , \ X_k = f \mid X_m = i\ ) M(\Pi \setminus \{f\}, \Pi^c)_{f,k}^{j,n} \\ & \leq K e^{-(C_{\Pi\setminus\{f\}}^*(f,j) - H_e(\Pi))/T_{m+1}} Q_m^n \text{ with } Q \in \mathcal{D}^r(H_e(\Pi), a', b'). \end{aligned}$$

In the same way than in the proof of lemma 3.5 we easily show that $C_{\Pi\setminus\{f\}}^*(f,j) - H_e(\Pi) = C_\Pi^*(f,j)$ and from lemma 3.4 we get $C_\Pi^*(f,j) = C_\Pi^*(i,j)$ for any $i \in \Pi$ so that the proof of $\mathcal{H}_3^n$ follows from (38) and (39).

The proof of $\mathcal{H}_3^n$ is completed.

$\mathcal{H}_4^n$: We can assume that A is not a cycle, otherwise the result is trivial. We will show at first that there exist $Q \in \mathcal{D}^r(H_e(A), a, b)$ and $K > 0$ such that

$$M(A, \Pi \cup A^c)_{i,m}^{j,n} \leq K Q_m^n.$$

Let $g \in A$ and $\Pi_g \in \mathcal{M}(A)$ such that $g \in F(\Pi_g)$ and $H_e(\Pi_g) = H_e(A)$. Considering the last visit to g by the Markov chain, we get

$$M(A, \Pi \cup A^c)_{i,m}^{j,n} \leq M(A \setminus \{g\}, \Pi \cup A^c)_{i,m}^{j,n}$$

$$+\sum_{k=m+1}^{n-1} P(\tau(A,m)>k,\ X_k=g \mid X_m=i\,)M(A\setminus\{g\},\Pi\cup A^c)^{j,n}_{g,k}.$$

Furthermore, concerning the last term, we have the decomposition for $j\in\Pi\cup A^c$

$$\begin{aligned}(40)\qquad M(A\setminus\{g\},\Pi\cup A)^{j,n}_{g,k} &\leq M(\Pi_g\setminus\{g\},\Pi_g^c)^{j,n}_{g,k}\\ &+\{M(\Pi_g\setminus\{g\},\Pi_g^c)M(A\setminus\{g\},\Pi\cup A^c)\}^{j,n}_{g,k}.\end{aligned}$$

Now, consider the family $(\Pi_p)_{0\leq p\leq r}$ of elements in $\mathcal{M}(A\setminus\{g\})$ such that

$$\Pi\cup A^c=\bigcup_{0\leq p\leq r}(\Pi_p\cup B_g)$$

where $B_g=A^c$ if $g\notin\Pi$ and $B_g=A^c\cup\{g\}$ otherwise. We have the expansion

$$M(A\setminus\{g\},\Pi\cup A^c)\leq\sum_{0\leq p\leq r}M(A\setminus\{g\},\Pi_p\cup B_g),$$

so that we obtain from $\mathcal{H}_1^{n-1}$ and $\mathcal{H}_4^{n-1}$ and lemma 3.5 that

$$\begin{aligned}(41)\quad &\{M(\Pi_g\setminus\{g\},\Pi_g^c)M(A\setminus\{g\},\Pi\cup A^c)\}^{j,n}_{g,k}\\ &\leq K_1e^{-C^*_{\Pi_g\setminus\{g\}}(g,\Pi_g^c)/T_{k+1}}Q^n_k\\ &=K_1e^{-H_e(A)/T_{k+1}}Q^n_k\qquad\text{with}\quad Q\in\mathcal{D}^r(H_e(A),a_1,b_1).\end{aligned}$$

Hence using again $\mathcal{H}_1^{n-1}$, we deduce from (40) and (41) that

$$M(A\setminus\{g\},\Pi\cup A^c)^{j,n}_{g,k}\leq K_2e^{-H_e(A)/T_{k+1}}Q^n_k \text{ with } Q\in\mathcal{D}^r(H_e(A),a_2,b_2),$$

and then from $\mathcal{H}_2^n$

$$\sum_{k=m}^{n-1}P(\tau(A,m)>k,\ X_k=g\mid X_m=i\,)M(A\setminus\{g\},\Pi\cup A^c)^{j,n}_{g,k}\leq K_3Q^n_m,$$

with $Q\in D^r(H_e(A),a_3,b_3)$. Applying $\mathcal{H}_4^{n-1}$ to $M(A\setminus\{g\},\Pi\cup A^c)^{j,n}_{i,m}$ we conclude finally that

$$M(A,\Pi\cup A^c)^{j,n}_{i,m}\leq K_4Q^n_m\qquad\text{with}\quad Q\in\mathcal{D}^r(H_e(A),a_4,b_4).$$

We can now prove $\mathcal{H}_4^n$. Let $\Pi_0 \in \mathcal{M}(A)$ such that $i \in \Pi_0$. If $j \in \Pi_0$ then $C_A^*(i,j) = 0$ and the result has been proved above. Otherwise, considering the last visit of the Markov chain in Π_0, we get

$$M(A, \Pi \cup B) \leq (I + M(A, \Pi_0))M(\Pi_0, \Pi_0^c)(I + M(A \setminus \Pi_0, \Pi \cup A^c)),$$

where $I_{i,m}^{j,m} = 1$ if $i = j$ and $m = n$, otherwise its value is zero. Then if

$$C = \inf_{i_2 \in A \setminus \Pi_0} \{C(\Pi_0, i_2) + C_{A \setminus \Pi_0}^*(i_2, j)\} \wedge C(\Pi_0, j)$$

where $C(\Pi_0, i_2)$ is the common value of $C_{\Pi_0}(i_1, i_2)$ for any $i_1 \in \Pi_0$, we deduce that

$$M(A, \Pi \cup B)_{i,m}^{j,n} \leq K_5 e^{-C/T_{m+1}} Q_m^n \text{ with } Q \in \mathcal{D}^r(H_e(A), a_5, b_5).$$

Since one easily see that $C \geq C_A^*(i,j)$, the proof of $\mathcal{H}_4^n$ is completed. □

Consider the family $(F_k)_{0 \leq k \leq r}$ of subsets of E, and the family $(H_k)_{0 \leq k \leq r}$ of positive real valued numbers defined by the following procedure:

- $F_0 = F(E)$, $H_0 = +\infty$.
- for $k \geq 0$,

– if $F_k = E$ then the procedure stops and $r = k$

– otherwise, we define $H_{k+1} = \sup\{H_e(\Pi) \mid \Pi \subset F_k^c\}$ and $F_{k+1} = \{ i \in E \mid H_e(\Pi_i^*) = H_{k+1} \}$ where Π_i^* is the largest cycle Π in E such that $i \in F(\Pi)$.

We can now establish the following theorem:

THEOREM 4.7. – *For each $0 \leq k < r$, then there exist $a > 0, b \geq 0$ and $K > 0$ which depends only on E, q, and κ (and not on V), such that for all $i \in F_k$, all $j \in F_{k+1} \setminus F_k$ we have*

$$L(F_k^c, F_k^c)_{i,m}^{j,n} \leq K e^{(W(j) - W(i))^+ / T_{m+1}} Q_m^n$$

with $Q \in \mathcal{D}^l(H_{k+1}, a, b)$.

Proof. – We first write the following expansion

$$L(F_k^c, F_{k+1}^c)_{i,m}^{j,n} \leq \sum_{j' \in \Pi_j^*} \sum_{l=m+1}^{n} M(F_k^c, \Pi_j^*)_{i,m}^{j',l} P(\ \tau(\Pi_j^*) > n \mid X_l = j'\ ).$$

Now define Π_i as the largest cycle in $E \setminus \{j\}$ which contains i. We have

$$\begin{aligned} M(F_k^c, \Pi_j^*)_{i,m}^{j',l} &\leq \{M(\Pi_i \setminus F_k, \Pi_i^c)(I + M(F_k^c, \Pi_j^*))\}_{i,m}^{l,j'} \\ &\leq K e^{-C_{\Pi_i \setminus F_k}^*(i, \Pi_i^c)/T_{m+1}} Q_m^n, \end{aligned}$$

with $Q \in \mathcal{D}^r(H_{k+1}, a_1, b_1)$. However

$$C^*_{\Pi_i \setminus F_k}(i, \Pi_i^c) \geq C^*_{\Pi_i \setminus \{i\}}(i, \Pi^c) = W(\Pi_i) + H_e(\Pi_i) - W(i),$$

so that

$$M(F_k^c, \Pi_j^*)_{i,m}^{j',l} \leq K e^{-[W(\Pi_i)+H_e(\Pi_i)-W(i)]/T_{m+1}} Q_m^n.$$

We should now study two cases.

First case: Assume that $W(j) \leq W(i)$, then $j \notin \Pi_i^*$ otherwise $j \in F(\Pi_i^*)$ and $j \in F_k$. Hence $\Pi_i^* \subset \Pi_i$ and $W(\Pi_i) + H_e(\Pi_i) - W(i) \geq W(\Pi_i^*) + H_e(\Pi_i^*) - W(i) = H_e(\Pi_i^*) \geq H_k \geq H_{k+1}$.

Second case: Assume that $W(j) > W(i)$, then let Π_i^+ be the smallest cycle Π such that $\{i, j\} \subset \Pi$. We have $\Pi_j^* \subset \Pi_i^+$ otherwise $i \in \Pi_j^*$ and $W(j) > W(i)$ leads to a contradiction. Hence $W(\Pi_i) + H_e(\Pi_i) \geq W(\Pi_j^*) + H_e(\Pi_j^*)$ so that

$$W(\Pi_i) + H_e(\Pi_i) - W(i) \geq H_{k+1} + W(j) - W(i).$$

From both previous cases, we get

$$L(F_k^c, F_{k+1} \setminus F_k)_{i,m}^{j,n} \leq \sum_{l=m+1}^{n} e^{-[H_{k+1}+(W_j-W_i)^+]/T_{m+1}} Q_m^l (1+b) \times \prod_{s=l+1}^{n} (1 - a e^{-H_{k+1}/T_s}),$$

with $Q \in D^r(H_{k+1}, a_2, b_2)$.

For proving the result, it is sufficient to show that there exist $a' > 0$ and $K' > 0$ such that

$$L_{m+1}^n \underset{\text{def}}{=} \sum_{l=m+1}^{n} Q_m^l \prod_{s=l+1}^{n} (1 - a e^{-H_{k+1}/T_s}) \leq K' \prod_{l=m+2}^{n} (1 - a' e^{-H_{k+1}/T_l}).$$

Noting $a_3 = \inf\{a, a_2\}$, we get after an integration by parts

$$L_{m+1}^n \leq \sum_{l=m+1}^{n} S_l \left(\prod_{s=l+1}^{n} (1 - a_3 e^{-H_{k+1}/T_s}) - \prod_{s=l}^{n} (1 - a_3 e^{-H_{k+1}/T_s}) \right) + S_{m+1} \prod_{s=m+1}^{n} (1 - a_3 e^{-H_{k+1}/T_s}),$$

where $S_l = \sum_{h=l}^{\infty} Q_m^h$. Since $S_l \leq (1+b)\prod_{h=m+1}^{l-1}(1 - a_3 e^{-H_{k+1}/T_h})$, we get

$$\begin{aligned} L_{m+1}^n \leq (1+b) &\sum_{l=m+1}^{n} \prod_{h=m+1}^{l-1} (1 - a_3 e^{-H_{k+1}/T_h}) a_3 e^{-H_{k+1}/T_l} \\ &\times \prod_{s=l+1}^{n} (1 - a e^{-H_{k+1}/T_s}) \\ &+ (1+b) \prod_{s=m+1}^{n} (1 - a_3 e^{-H_{k+1}/T_s}) \\ &\leq \frac{1+b}{1-a_3} \left(\sum_{l=m+1}^{n} a_3 e^{-H_{k+1}/T_l} \right) \prod_{l=m+1}^{n} (1 - a_3 e^{-H_{k+1}/T_l}) \\ &+ (1+b) \prod_{s=m+1}^{n} (1 - a_3 e^{-H_{k+1}/T_s}). \end{aligned}$$

Since $\sum_{l=m+1}^{n} a_3 e^{-H_{k+1}/T_l} \leq 2 \prod_{l=m+1}^{n} (1 + (a_3/2) e^{-H_{k+1}/T_l})$ and

$$\left(1 + \frac{a_3}{2} e^{-H_{k+1}/T_l}\right)(1 - a_3 e^{-H_{k+1}/T_l}) \leq \left(1 - \frac{a_3}{2} e^{-H_{k+1}/T_l}\right),$$

the result is proved. Hence, the proof of the theorem is completed. □

5. NECESSARY AND SUFFICIENT CONDITION FOR CONVERGENCE

With this section, we start our study of the convergence properties of the G.S.A. As mentioned in the introduction, our approach will be based on the study on the probability to be in a level set of the virtual energy W.

DEFINITION 5.1. – Let $\lambda > 0$. We denote E_λ the subset of E defined by

$$E_\lambda = \{\, i \in E \mid W(i) \geq \min W + \lambda \,\}.$$

We give now a necessary and sufficient condition on the decreasing cooling schedule for which the mass in E_λ vanishes when the number of steps increases:

THEOREM 5.2. – *Let q be an irreducible Markov kernel on E and let $\kappa \in [1, +\infty[$. Let X be a G.S.A. with parameter (q, κ, P) and let $\mathcal{T}$ be the* *underlying cooling schedule. Assume that $\mathcal{T}$ is decreasing ($T_n \geq T_{n+1}$) and vanishes when n tends to infinity. Then for all $\lambda > 0$,*

$$\sup_{i\in E} P(\ X_n \in E_\lambda \mid X_0 = i) \underset{n\to+\infty}{\longrightarrow} 0 \qquad \text{iff} \quad \sum_{n\geq 0} e^{-\Gamma_\lambda/T_n} = +\infty,$$

where

$$\Gamma_\lambda = \sup\{\ H_e(\Pi) \mid \Pi \in \mathcal{C}(E) \text{ and } \inf_{i\in\Pi} W(i) \geq \min W + \lambda\ \}.$$

Proof. – The direct implication can be got from theorem 4.5. Let Π be a cycle included in E_λ and such that $H_e(\Pi) = \Gamma_\lambda$. Then, for all $i \in \Pi$

$$\begin{aligned} P(\ X_n \in \Pi \mid X_0 = i\ ) &\geq P(\ \tau(\Pi, 0) = +\infty \mid X_0 = i\ ) \\ &\geq c \prod_{k=0}^{\infty} (1 - de^{-H_e(\Pi)/T_k}). \end{aligned}$$

For the converse implication, we have to use the theorem 4.7. From the definition of the subsets F_k, we get that there exits $k \in \mathbb{N}$ for which $E_\lambda \subset E \setminus F_k$ et $H_{k+1} = \Gamma_\lambda$. Hence, denoting $\Delta = \inf\{W(j) - W(i) \mid i \in E \setminus E_\lambda \text{and } j \in E_\lambda\ \}$, we get the inequality:

$$\begin{aligned} &P(\ X_n \in E_\lambda \mid X_0 = i\ ) \\ &\leq \sum_{m=0}^{n-1} \sum_{i'\in F_k} \sum_{j\in E_\lambda} P(\ X_m = i' \mid X_0 = i\ ) L(E \setminus F_k, E \setminus F_k)_{i',m}^{j,n} \\ &\leq K \sum_{0\leq m<n} e^{-\Delta/T_{m+1}} Q_m^n \end{aligned}$$

with $Q \in D^l(H_{k+1}, a, b)$. Since $(T_n)_{n\in\mathbb{N}}$ is decreasing and converges to 0, there exists for all $\epsilon > 0$ an non negative integer M such that for all $m \geq M$ we have $e^{-\Delta/T_{m+1}} \leq \epsilon/2$. Therefore we have

$$\sum_{m=M}^{n-1} e^{-\Delta/T_{m+1}} Q_m^n \leq \epsilon/2 \sum_{m=M}^{n-1} Q_m^n \leq \epsilon/2.$$

However

$$\sum_{m=0}^{M-1} e^{-\Delta/T_{m+1}} Q_m^n \leq \sum_{m=0}^{M-1} Q_m^n \leq \sum_{-\infty}^{M-1} Q_m^n.$$

Since $Q \in D^l(H_{k+1}, a, b)$, we get the upper bound

$$\sum_{-\infty}^{M-1} Q_m^n \leq (1 + b) \prod_{m=M}^{n-1} (1 - ae^{-\Gamma_\lambda/T_m}),$$

so that there exists $N \in \mathbb{N}$ such that for all $n \geq N$

$$(1+b) \prod_{m=M}^{n-1} (1 - ae^{-\Gamma_\lambda / T_m}) \leq \epsilon/2.$$

The proof is completed. □

6. OPTIMAL CONVERGENCE SPEED EXPONENT

We study now by the optimal convergence rate of $\sup_{i \in E} P(\ X_N \in E_\lambda \mid X_0 = i\ )$ for any $\lambda > 0$ (see definition 1.3). We will consider triangular cooling schedules e.a. we free ourself from an unique cooling schedule for all the finite horizon N and we allows us to define for each finite horizon an adapted cooling schedule $\mathcal{T}^N$. We will established successively an upper and a lower bound for this convergence speed.

6.1. Upper bound

In this part, we assume that the family $(Q_T)_{T>0}$ in $\mathcal{A}(q, \kappa)$ satisfies the following additional condition:

C_1: There exists a set $\mathcal{B}$ of continuous real valued functions on $[0, +\infty[$ such that

(1) For all f and g in $\mathcal{B}$

• $f + g \in \mathcal{B}$

• There exists $A \geq 0$ such that one of the two following inequalities is true:
- $f(x) \leq g(x)$ for all $x \geq A$.
- $f(x) \geq g(x)$ for all $x \geq A$.

(2) For all $i, j \in E$, $i \neq j$, if $q(i,j) > 0$, then there exist $A_{ij} > 0$ and $f_{ij} \in \mathcal{B}$ such that

$$Q_\beta(i,j) = A_{ij} \exp\left(\int_0^\beta f_{ij}\right) \text{ with } \beta = 1/T.$$

This extra condition is essentially a condition of monotony in a neighborhood of $+\infty$ in β which is satisfied by all the standard sequential or parallel annealing algorithms. For example, condition C_1 holds if

$$Q_\beta(i,j) = \{\sum_{k \in I} a_k e^{b_k \beta}\} / \{\sum_{j \in J} c_j e^{d_j \beta}\},$$

where $(a_k)_{k\in I}$, $(b_k)_{k\in I}$, $(c_j)_{j\in J}$ and $(d_j)_{j\in J}$ are finite family of real numbers such that $\sum_{j\in J} c_j e^{d_j\beta} \neq 0$ for all $\beta \geq 0$.

THEOREM 6.1. – *Consider a family $(Q_T)_{T>0}$ in $\mathcal{A}(q,\kappa)$ which satisfies C_1. Let X be a G.S.A. with parameter (q,κ,P) where $P = P_{(Q,\mathcal{T},\nu_0)}$ (see definition 1.3) and assume that the cooling schedule in decreasing ($T_n \geq T_{n+1}$). Then there exists $\gamma > 0$ independent of $\mathcal{T}$ and of the initial distribution ν_0 such that*

$$P(\ X_n = i\ ) \geq \left(\gamma \inf_{j\in E} \nu_0(j)\right) \times \mu_{T_n}(i),$$

where μ_T is the unique invariant probability measure of Q_T.

Proof. – Let us recall first the well known explicit expression of the invariant probability measure given by the Wentzell and Freidlin A-graphs in [5]:

$$\mu_T(i) = \{ \sum_{g\in G(\{i\})} \prod_{u\to v\in g} Q_T(u,v)\}/\{\sum_{j\in E} \sum_{g\in G(\{j\})} \prod_{u\to v\in g} Q_T(u,v)\}.$$

From now, we will use the variable $\beta = 1/T$ instead of T which seems to be more appropriate for our proof. From C_1, we get for each $u, v \in E$, $u \neq v$, a continuous function f_{uv} in $\mathcal{B}$ such that:

$$Q_\beta(u,v) = A_{uv} \exp\left(\int_0^\beta f_{uv}\right).$$

Hence, if we note $f_g = \sum_{u\to v\in g} f_{uv}$, we get

$$\mu_T(i) = \left\{\sum_{g\in G(\{i\})} A_g \exp\left(\int_0^\beta f_g\right)\right\} \Big/ \left\{\sum_{j\in E} \sum_{g\in G(\{j\})} A_g \exp\left(\int_0^\beta f_g\right)\right\}.$$

Now, considering $c^+ = \sum_{u\neq v} f_{uv}^+$ where $f_{uv}^+ = f_{uv}\mathbf{1}_{f_{uv}\geq 0}$ and defining $\widetilde{f}_g = f_g - c^+$, we have

$$\mu_\beta(i) = a_\beta(i)/Z_\beta,$$

where

$$a_\beta(i) = \sum_{g\in G(\{i\})} A_g \exp\left(\int_0^\beta \widetilde{f}_g\right) \text{ and } Z_\beta = \sum_{i\in E} a_\beta(i).$$

An obvious computation shows that $\widetilde{f}_g \leq 0$ so that $a_\beta(i)$ is decreasing in β for each $i \in E$. Finally, since $\widetilde{f}_g - \widetilde{f}_{g'} = f_g - f_{g'}$ for $g \in G(\{i\})$ and $g' \in G(\{i'\})$. We deduce from the stability under addition of the elements of $\mathcal{B}$ that there exist $i_0 \in E$, $g_0 \in G(\{i_0\})$ and $\beta_0 \geq 0$ such that for all $i \in E$ and all $g \in G(\{i\})$, the functions $\hat{f}_g(\beta) = \widetilde{f}_g(\beta) - \widetilde{f}_{g_0}(\beta)\mathbf{1}_{\beta \geq \beta_0}$ satisfy

$$\begin{cases} \hat{f}_g(\beta) \leq 0 & \text{if } \beta \geq 0, \\ \hat{f}_{g_0}(\beta) = 0 & \text{if } \beta \geq \beta_0. \end{cases}$$

Hence, $\mu_\beta(i) = \hat{a}_\beta(i)/\hat{Z}_\beta$ where

$$\hat{a}_\beta(i) = \sum_{g \in G(\{i\})} A_g \exp\left(\int_0^\beta \hat{f}_g\right) \text{ and } \hat{Z}_\beta = \sum_{i \in E} \hat{a}_\beta(i).$$

The essential fact here is that the $\hat{a}_\beta(i)$ are decreasing in β and that

$$\hat{a}_\beta(i_0) \geq A_{g_0} \exp\left(\int_0^{\beta_0} \hat{f}_{g_0}\right) > 0 \text{ for all } \beta.$$

Now, assume that we have proved

$$P(X_n = i) \geq \alpha_n \mu_{\beta_n}(i),$$

then

$$\begin{aligned} P(X_{n+1} = i) &= \sum_{j \in E} P(X_n = j) Q_{\beta_{n+1}}(j, i) \\ &\geq \alpha_n \sum_{j \in E} \mu_{\beta_n}(j) Q_{\beta_{n+1}}(j, i). \end{aligned}$$

However, since $\hat{a}_\beta(j)$ is decreasing and $\beta_{n+1} \geq \beta_n$, we get

$$\mu_{\beta_n}(j) \geq k_n \mu_{\beta_{n+1}}(j) \text{ with } k_n = \hat{Z}_{\beta_{n+1}}/\hat{Z}_{\beta_n}.$$

Hence, one easily computes that

$$P(X_{n+1} = i) \geq \alpha_n k_n \mu_{\beta_{n+1}}(i).$$

We can notice then that

$$\prod_{p=0}^{n} k_p \geq \lim_{\beta \to +\infty} \hat{Z}_\beta / \hat{Z}_0.$$

Since $\hat{Z}_\beta \geq \gamma$ with $\gamma = A_{g_0} \exp(\int_0^{\beta_0} \hat{f}_{g_0})$ and

$$\hat{Z}_0 \leq \prod_{u \in E} (\sum_{v \in E} A_{uv}) \leq \prod_{u \in E} (\sum_{v \in E} Q_0(u, v)) \leq 1,$$

we get $P(X_{n+1} = i) \geq \gamma \inf_{j \in E} \nu_0(j) \mu_{\beta_n}(i)$ so that the theorem is proved. □

The theorem above shows what can be expected from a decreasing cooling schedule. Whatever the cooling schedule is, the probability for X_n to be in a configuration i is bounded from below by the invariant probability measure at temperature T_n. We can now give an upper bound for the convergence speed of G.S.A on level sets.

THEOREM 6.2. – *Let q be an irreducible Markov kernel on E, let $\kappa \in [1, +\infty[$ and let X be a G.S.A with parameter (q, κ, P) whose underlying cooling schedule $\mathcal{T}$ is decreasing. Then, assuming that the underlying family $Q \in \mathcal{A}(q, \kappa)$ satisfies the condition C_1, there exists $b > 0$ (which depends on Q but not on $\mathcal{T}$) such that for all level $\lambda > 0$ and all $n > 0$,*

$$\sup_{i \in E} P(X_n \in E_\lambda \mid X_0 = i\ ) \geq \frac{b}{n^{\alpha_\lambda}}$$

with

$$\alpha_\lambda = \inf \left\{ \frac{W(\Pi) - \min W}{H_e(\Pi)} \mid \Pi \in \mathcal{C}(E),\ W(\Pi) - \min W \geq \lambda \right\}.$$

Proof. – A very similar theorem has been proved in [2] for sequential annealing and our proof borrows it crucial technical tricks from [2].

As noticed in [2] its is sufficient to prove that

$$P(W(X_n) - \min W \geq \lambda\ ) \geq \frac{b}{n^{\alpha\lambda}},$$

for the Markov chain with the initial distribution measure $\nu_0(j) = 1/|E|$ for all $j \in E$.

Now, consider $\Pi_\lambda \in \mathcal{C}(E)$ such that $W(\Pi) - \min W \geq \lambda$ and $\frac{W(\Pi) - \min W}{H_e(\Pi)} = \alpha_\lambda$. Then, considering the last visit to Π_λ, we have the expansion

$$\begin{aligned} (42) \qquad P(X_n \in \Pi_\lambda) \geq &\sum_{i \in \Pi_\lambda} P(X_0 = i) P(\tau(\Pi_\lambda, 0) > n \mid X_0 = i) \\ &+ \frac{1}{\kappa} \sum_{j \notin \Pi_\lambda} \sum_{i \in \Pi_\lambda} \sum_{m < n} P(X_m = j) q(j, i) e^{-V(j,i)/T_{m+1}} \\ &\times P(\tau(\Pi_\lambda, m+1) > n \mid X_{m+1} = i\ ) \end{aligned}$$

From theorem 4.7 and Proposition 1.6 we get $K > 0$ independent of $\mathcal{T}$ such that for all $j \in E$

$$P(X_m = j) \geq Ke^{-(W(j)-minW)/T_m}.$$

Hence,

$$P(X_m = j)e^{-V(j,i)/T_{m+1}} \geq Ke^{-(W(j)+V(j,i)-minW)/T_{m+1}}. \tag{43}$$

Now, consider the cycle Π_{ij} which is the smallest cycle in $\mathcal{C}(E)$ containing $\{i,j\}$ and Π_i (respectively Π_j) which is the largest cycle in $E \setminus \{j\}$ (respectively $E \setminus \{i\}$) containing i (respectively containing j). We have $\Pi_i, \Pi_j \in \mathcal{M}(\Pi_{ij})$ so that we deduce from lemma 3.6 that $W(\Pi_i) + H_e(\Pi_i) = W(\Pi_j) + H_e(\Pi_j)$ and from lemma 3.7 that $W(j) + V(j,i) \geq W(\Pi_j) + H_e(\Pi_j)$. Hence

$$\inf_{j \notin \Pi_\lambda,\ i \in \Pi_\lambda} W(j) + V(j,i) - minW = H_e(\Pi_\lambda) + W(\Pi_\lambda) - minW. \tag{44}$$

We should handle separately the case $|\Pi_\lambda| = 1$.

Assume that $|\Pi_\lambda| = 1$, then one proves easily by a direct computation that there exist $c > 0$, $0 \leq d \leq 1$ such that

$$P(\tau(\Pi_\lambda, m) > n \mid X_m = i\ ) \geq c \prod_{l=m+1}^{n} (1 - de^{-H_e(\Pi_\lambda)/T_l}) \tag{45}$$

so that we obtain from (42), (43), (44) and (45) that there exists K_1 (independent of λ and $\mathcal{T}$) such that

$$\begin{aligned} P(X_n \in \Pi_\lambda) &\geq c\prod_{l=1}^{n}(1 - de^{-H_e(\Pi_\lambda)/T_l}) \\ &+ \sum_{m=1}^{n} K_1 e^{-(H_e(\Pi_\lambda)+W(\Pi_\lambda)-minW)/T_m} \prod_{l=m+1}^{n} (1 - de^{-H_e(\Pi_\lambda)/T_l}). \end{aligned} \tag{46}$$

Assume now that $|\Pi_\lambda| \geq 2$, then denoting $\rho = \inf\{q(i,j) \mid q(i,j) > 0 \text{ and } i,j \in E\ \}$, we have $P(\tau(\Pi_\lambda, m) > m+1 \mid X_m = i\ ) \geq 1 - \chi$ where $\chi = 1 - \rho/\lambda$. Hence, we deduce from $\mathcal{H}_1$ in theorem 4.5, that there exist c, d and $K > 0$ such that

$$P(\ X_n \in \Pi_\lambda \mid X_m = i\ ) \geq c \prod_{l=m+1}^{n} (1 - (de^{-H_e(\Pi_\lambda)/T_l}) \wedge \chi),$$

which leads with (42), (43) and (44) to

$$(47)\quad P(X_n \in \Pi_\lambda) \geq c \prod_{l=m+1}^{n} (1 - (de^{-H_e(\Pi_\lambda)/T_l}) \wedge \chi)$$
$$+ \sum_{m=1}^{n} K_2 e^{-(H_e(\Pi_\lambda)+W(\Pi_\lambda)-minW)/T_m} \prod_{l=m+1}^{n} (1 - (de^{-H_e(\Pi_\lambda)/T_l}) \wedge \chi).$$

The remaining of the proof is now exactly the same that the proof of Theorem 5.2 in [2] and leads to the explicit computation of the lower bound for $P(X_n \in \Pi_\lambda)$ through inequalities (46) and (47). However, for completeness, we recall here the arguments. We consider only the case $|\Pi_\lambda| \geq 2|$. The case $|\Pi_\lambda| = 1$ can be handled similarly and is left to the reader.

Consider the sequence $(R_n)_{n\in\mathbb{N}}$ defined by

$$(48)\qquad \begin{cases} R_0 \leq c \\ R_n = (1 - (de^{-H_e(\Pi_\lambda)/T_n}) \wedge \chi) R_{n-1} \\ +K_2 e^{-(H_e(\Pi_\lambda)+\widetilde{W}(\Pi_\lambda))/T_n}, \end{cases}$$

where we have noted $\widetilde{W}(\Pi_\lambda) = W(\Pi_\lambda) - minW$.

Denoting $\delta = 1/\alpha_\lambda$, a straightforward computation gives that

$$(49)\qquad \min_{T_n>0} R_n = R_{n-1} - \frac{d}{1+\delta}\left(\frac{d}{K_2(1+\delta^{-1})}\right)^{\delta} R_{n-1}^{(1+\delta)}$$

for

$$R_{n-1} \leq \left(\frac{\chi}{d}\right)^{\delta^{-1}} \left(\frac{(1+\delta^{-1})K_2}{d}\right).$$

Hence consider

$$S_0 = c \wedge \left(\frac{\chi}{d}\right)^{\delta^{-1}} \left(\frac{(1+\delta^{-1})K_2}{d}\right)$$

and

$$(50)\qquad S_n = S_{n-1} - \frac{d}{1+\delta}\left(\frac{d}{K_2(1+\delta^{-1})}\right)^{\delta} S_{n-1}^{(1+\delta)}.$$

We will prove by induction that for all decreasing $\mathcal{T} = (T_n)_{n\in\mathbb{N}}$, the sequence $(R_n)_{n\in\mathbb{N}}$ defined by (48) and $R_0 = S_0$ verifies $R_n \geq S_n$ for all $n \in \mathbb{N}$.

Assume that $R_k \geq S_k$, then from (48) we get

$$(51) \quad R_{k+1} \geq \min_{T_{k+1}}(1 - (de^{-H_e(\Pi_\lambda)/T_{k+1}}) \wedge \chi)S_k + K_2 e^{-(H_e(\Pi_\lambda)+\widetilde{W}(\Pi_\lambda))/T_{k+1}}$$

Since $S_k \leq S_0 \leq (\chi/d)^{\delta^{-1}}(1+\delta^{-1})K_2/d$, we deduce from (49) and (51) that

$$R_{k+1} \geq S_{k+1}.$$

Hence $P(X_n \in \Pi_\lambda) \geq S_n$ for all $n \in \mathbb{N}$.

It is sufficient now to compute a lower bound for the sequence $(S_n)_{n \in \mathbb{N}}$. From (50) we deduce that

$$\begin{aligned} S_n^{-\delta} - S_{n-1}^{-\delta} &\leq -\delta(S_n - S_{n-1})S_n^{-(\delta+1)} \\ &\leq \frac{\delta}{1+\delta} d\left(\frac{d}{K_2(1+\delta^{-1})}\right)^{\delta}\left(\frac{S_{n-1}}{S_n}\right)^{(1+\delta)}. \end{aligned}$$

However,

$$\begin{aligned} \frac{S_{n-1}}{S_n} &= \left(1 - \frac{d}{1+\delta}\left(\frac{dS_{n-1}}{K_2(1+\delta^{-1})}\right)^{\delta}\right)^{-1} \\ &\leq \left(1 - \frac{d}{1+\delta}\left(\frac{dS_0}{K_2(1+\delta^{-1})}\right)^{\delta}\right)^{-1} \leq \left(1 - \frac{\chi}{1+\delta}\right) \leq (1-\chi). \end{aligned}$$

Hence

$$S_n^{-\delta} \leq S_0^{\delta} + n\frac{\delta}{1+\delta} d\left(\frac{d}{K_2(1+\delta^{-1})}\right)^{\delta}(1-\chi).$$

Denoting

$$a = S_0^{\delta}\frac{\delta}{1+\delta} d\left(\frac{d}{K_2(1+\delta^{-1})}\right)^{\delta}(1-\chi),$$

we get $S_n^{-\delta} \leq na$ so that $S_n \geq b/n^{\delta^{-1}}$ with $b = a^{-\delta^{-1}}$. The proof of theorem 6.2 is completed. □

6.2. Lower bound

Theorem 6.1 shows that there exist a constant K such that for all $i \in E$ and all decreasing cooling schedules, we have

$$\sup_{i\in E} P(\ X_n = j \mid X_0 = i) \geq K e^{-(W(j)-W_{min})/T_n}$$

To establish a lower bound, it is natural to look for cooling schedule such that

$$\sup_{i\in E} P(\ X_n = j \mid X_0 = i) \leq K' e^{-(W(j)-W_{min})/T_n} \tag{52}$$

Theorem 4.7 is the key of the inequality (52). As we have previously noticed, the statement of this theorem is exactly the same in the sequential case than in the generalized case except that we have to replace the virtual energy U by the virtual energy W. Moreover, it is the unique source of all the lower bound estimates for the convergence speed of the sequential simulated annealing algorithms. Hence , the extension the the general case of the theorem on the lower bound of the convergence speed does not demand any specific modification compared to its statement in [2] for the sequential case neither for its proof. It is sufficient to use *strictly* the same arguments and to exchange U and W.

THEOREM 6.3. – *Let q be an irreducible Markov kernel on E, let $\kappa \in [1, +\infty[$ and let $Q \in \mathcal{A}(q, \kappa)$. Let $\lambda > 0$. There exists a non negative constant K such that for all $N \in \mathbb{N}_*$, there exists a decreasing cooling schedule $\mathcal{T}^N = (T^N_{n\in\mathbb{N}})$ for which if X is a G.S.A with parameter (q, κ, P_N) where $P_N = P_(Q, \mathcal{T}^N, \nu_0)$ then*

$$\sup_{i\in E} P_N(\ X_N \in E_\lambda \mid X_0 = i) \leq \frac{K}{N^{\hat{\alpha}_\lambda}}$$

where

$$\hat{\alpha}_\lambda = \inf\left\{ \frac{(W(\Pi) - \min W) \wedge \lambda}{H_e(\Pi)} \mid \Pi \in \mathcal{C}(E),\ W(\Pi) > \min W \right\}.$$

Remark 15. – One obviously have $\alpha_\lambda = \hat{\alpha}_\lambda$ for sufficiently small λ.

Proof. – It is sufficient to consider the proof of the theorem 7.1 in [2] and to replace U by W. □

7. CONCLUSION

In this paper, we have emphasized the study of the value of the optimal convergence speed exponent for level sets. Our approach gives a semi explicit expression of the exponents in function of the communication cost V. However, in practical situation, their numeric computations are hard combinatorial problems but an implementable recursive algorithm has been proposed by the author in [15] and [16] useful to test conjecture on small state spaces. Moreover, an easier problem is the comparison of the exponents for different annealing algorithms and this approach succeeds for instance in the study of parallel algorithms based on several interacting annealing processes as done in [14]. Concerning the synchronous parallel version of the sequential annealing for image processing presented in the introduction, the problem is a bit more delicate and it appear that generally the configurations minimizing the virtual energy do not minimize the underlying cost U (one can propose more efficient parallel schemes,; see [15] for an extensive study).

One limitation of this approach is in the evaluation of the multiplicative constants appearing in the upper and lower bounds for the convergence speed. An alternative approach seems to be the computation of geometric bounds of the spectral gap of the transition kernel based on the Poincaré method. This method, used in the continuous-time setting, gives the optimal constant for the logarithmic cooling schedules ([7], [4], [11]). However, this approach does not make appear the optimal exponent α_{opt} , in particular because one needs to get estimates uniform in a large set of cooling schedules.

REFERENCES

[1] R. AZENCOTT, A common large deviation framework for sequential and parallel annealing. In R. Azencott *et al.*, editors, *Simulated annealing: Parallelization techniques*, chapter 2, Willey and Sons, 1992, pp. 11-23.

[2] O. CATONI, Rough large deviation estimates for simulated annealing. Application to exponential schedules, *Ann. Probab.*, Vol. **20**, 1992, pp. 1109-1146.

[3] T.-S. CHIANG and Y. CHOW, A limit theorem for a class of inhomogeneous markov processes, *Ann. Probab.*, 1989.

[4] J.-D. DEUSCHEL and C. MAZZA, L^2 convergence of time non-homogeneous markov processes: I. spectral estimates, *Université de Fribourg, Institut de Mathématiques*, preprint (to appear in the *Ann. of Appl. Prob.*), 1992.

[5] M. I. FREIDLIN and A. D. WENTZELL, *Random Pertubations of Dynamical Systems*, Vol. **260**, Springer-Verlag, 1984.

[6] D. GEMAN, Random fields and inverse problem in imaging. In *École d'Été de probabilités de Saint-Flour XVIII,* Springer-Verlag, 1990.

[7] F. GÖTZE, Rate of convergence of simulated annealing processes, Preprint, 1992.

[8] B. HAJEK, Cooling schedule for optimal annealing, *Math. Oper. Res.*, Vol. **13**, 1988, pp. 311-329.

[9] R. HOLLEY and D. STROOK, Annealing via sobolev inequalities, *Comm. Math. Phys.*, Vol. **115**, 1988, pp. 553-559.

[10] C.-R. HWANG and S.-J. SHEU, Singular perturbed markov chains and exact behaviors of simulated annealing process, *J. Theoret. Probab.*, Vol. 5(2), 1992, pp. 223-249.

[11] S. INGRASSIA, On the rate of convergence of the metropolis algorithm and gibbs sampler by geometric bounds, to appear in *Annals of Applied Probability*, 1993.

[12] S. KIRKPATRICK, C. GELATT and M. VECCHI, Optimization by simulated annealing, *Sciense*, Vol. **220**, 1983, pp. 671-680.

[13] L. MICLO, Recuit simulé sans potentiel sur un ensemble fini, *Séminaire de Probabilités*, Vol. 26, 1992.

[14] A. TROUVÉ, , Asymptotical behavior of several interacting annealing processes, Preprint, 1993.

[15] A. TROUVÉ, Parallélisation massive du recuit simulé, *PhD thesis*, Université d'Orsay, Jan. 1993.

[16] A. TROUVÉ, Cycle decompositions and simulated annealing, Rapport de recherche du LMENS, to appear in *SIAM J. Control Opt.*, 1996.

(Manuscript received October 4, 1993;
revised version received January 30, 1995.)