On the Convergence and the Applications of the Generalized Simulated Annealing

P. DEL MORAL, L. MICLO

Abstract

The convergence of the generalized simulated annealing with time-inhomogeneous communication cost functions is discussed. This study is based on the use of Log-Sobolev inequalities and semigroup techniques in the spirit of a previous article by one of the authors. We also propose a natural test set approach to study the global minima of the virtual energy. The second part of the paper is devoted to the application of these results. First we propose two general Markovian models of genetic algorithms and we give a simple proof of the convergence toward the global minima of the fitness function. Finally we introduce a stochastic algorithm which converges to the set of the global minima of a given mean cost optimization problem.

Introduction

Let E a finite state space and q an irreducible Markov kernel. The main purpose of this paper is to study the limiting behavior of a large class of time-inhomogeneous Markov processes controlled by two parameters $(\gamma, \beta) \in \mathbb{R}_+^2$ and associated to a family of Markov kernels $Q_{\gamma,\beta}(x, y)$ having the following property:

$$\exists k > 0 \;:\quad k^{-1}\; q(x, y)e^{-\beta V_\gamma(x,y)} \leq Q_{\gamma,\beta}(x, y) \leq k\;\; q(x, y)e^{-\beta V_\gamma(x,y)} \tag{1}$$

where $V : \mathbb{R}_+ \times E^2 \to \mathbb{R}_+ \cup \{\infty\}$, $V_\gamma(x, y) < +\infty \Longleftrightarrow q(x, y) > 0$ and, for any $x, y \in E$ $(\gamma, \beta) \to Q_{\gamma,\beta}(x, y) \in C^1$.

For a discussion on the origins of this problem the reader is referred to the introduction

---

Work partly supported by INTAS-RFBR 95-0091 and EU TMR Programme Project ERBF MRX CT 960075A .

Laboratoire de Statistique et Probabilités, CNRS-UMR C5583, Bat 1R1, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex, electronic address: delmoral@cict.fr, miclo@cict.fr.

AMS subject classifications: *60J05, 92D15*

keywords: *Simulated Annealing, Genetic Algorithms, Stochastic Optimization*

of A.Trouvé ([12]) who studies the asymptotic behavior of such chains, with time homogeneous function $V(x,y)$, using large deviation techniques. The fundamental notions here are those of the log-Sobolev constant $a(\gamma,\beta)$ of $Q_{\gamma,\beta}$ and the relative entropy of a measure with respect to another measure. Other complementary results relating to time-inhomogeneous communication cost function can be found in Frigerio/Grillo [7], Younes [13] and more recently in Löwe [9] where Sobolev inequalities are used rather than log-Sobolev inequalities for classical models where q is assumed to be reversible and V_γ is associated to an a priori potential depending on γ.

For a probability measure m on E, inverse-freezing schedule $\beta \in C^1(\mathbb{R}_+, \mathbb{R}_+)$ and $\gamma \in C^1(\mathbb{R}_+, \mathbb{R}_+)$, we denote (Ω, P, F_t, X_t) the canonical process associated to the family of generators $(L_{\gamma_t,\beta_t})_{t\geq 0} = (Q_{\gamma_t,\beta_t} - I)_{t\geq 0}$ whose initial condition is $m_0 = m$ and we denote m_t the distribution of X_t.
The aim of section 1.1 is to give several conditions on the rate of increase of $\gamma_t, \beta_t \to +\infty$ to ensure the entropy of m_t with respect to π_{γ_t,β_t} converges to 0. We shall examine as much of the theory as possible in a form applicable to general optimization problems and so in particular to mean cost optimization problems.
To illustrate our results we will restrict attention to various special classes of generalized simulated annealing. We will commence with a detailed analysis of general Markov kernels of the form

$$Q_\beta(x,y) = \sum_{u\in U} \overline{q}_\beta(x,u,y) e^{-\beta \overline{V}(x,u,y)} \tag{2}$$

where U is a given finite set, $\overline{V} : E \times U \times E \to \mathbb{R}_+$ and $\overline{q}_\beta : E \times U \times E \to \mathbb{R}_+$, $\beta \in \mathbb{R}_+$, is a family of functions satisfying some continuity and irreductibility conditions. This situation can be formulated in the general form (1). We will settle this question together with the explicit computation of the corresponding communication cost function V. In a final stage we will give several conditions on the rate of decrease of the cooling schedule to ensure the convergence in probability of the corresponding canonical process X_t, as $t \to +\infty$, to the set of global minima of the virtual energy associated to V.

Another application is the situation in which the Markov kernel $Q_{\gamma,\beta}$ has the form

$$Q_{\gamma,\beta}(x,y) = q_\beta(x,y)\; e^{\beta V_\gamma(x,y)}$$

with

$$\lim_{\gamma\to+\infty} V_\gamma(x,y) = V(x,y) \qquad \lim_{\beta\to+\infty} q_\beta(x,y) = q(x,y)$$

for some Markov kernel q and some function $V : E \times E \to \mathbb{R}^+$. In this situation, let π_β be the unique invariant probability measure of the Markov generator $L_\beta = Q_\beta - I$, where

$$Q_\beta(x,y) = q(x,y)\; e^{-\beta\, V(x,y)}$$

We will give several conditions on the rate of decrease of the cooling schedule and on the rate of convergence $\lim_{t\to+\infty} V_{\gamma_t} = V$ to ensure the entropy of m_t with respect to π_{β_t} converges to 0.

The above results imply the convergence in law of the canonical process X_t to the set of the global minima $V^\star$ of a virtual energy V. This leads us to investigate more closely the properties of such function. The purpose of section 1.2 is to introduce a natural test set approach to study $V^\star$. More precisely we will give a condition for a given subset $H \subset E$ to contain $V^\star$.

Section 2 is devoted to application of these results. An area of application of these results is the situation in which Q_β is the transition probability kernel of a genetic algorithm. Such algorithms can be formulated by a Markov process with state space $E = S^N$ ($N > 1$ and S a finite set) and whose transition probabilities Q_β includes a mutation transition $Q_\beta^{(1)}$ and a selection mechanism $Q_\beta^{(2)}$. More precisely the mutation transition is modelled by independent motion of each particles and the selection mechanism chooses randomly in the previous population according a given fitness function. The first convergence result was obtained by Cerf [2] in the case in which $Q_\beta = Q_\beta^{(1)} Q_\beta^{(2)}$ and the mutations vanish, that is $\lim_{\beta\to+\infty} Q_\beta^{(1)}(x,y) = 1_x(y)$. In section 2.1 we will use the results of section 1.1 and the test set approach introduced in section 1.2 to derive a new and simple proof of the convergence in probability of such algorithms to the set of the global minima of the fitness function in the following situations:

$$Q_\beta = Q_\beta^{(1)} Q_\beta^{(2)} \quad \text{and} \quad Q_\beta = \alpha\, Q_\beta^{(1)} + (1-\alpha)\, Q_\beta^{(2)} \quad 0 < \alpha < 1$$

Finally, in subsection 2.2 we apply the results of the first section to mean cost optimization problems. However here we touch upon a slightly different aspect of the theory. Namely, the object will be to find the global minima of a function $U : E \to \mathbb{R}_+$ defined by

$$U(x) = E(L(Z,x))$$

where Z is a random variable taking values in a finite set F and $L : F \times E \to \mathbb{R}_+$. We will solve this optimization problem by an original method based on the use of Monte Carlo simulations coupled with simulated annealing. This special case will require specific developments because the corresponding function V_γ will necessarily behave as a random process. We will present a time-inhomogeneous Markov process which converges to the global minima of U.

1 General Results

The purpose of this section is to study the limiting behavior of time-inhomogeneous Markov chains controlled by two parameters $(\gamma, \beta) \in \mathbb{R}_+^2$ and associated to a family of Markov kernels $Q_{\gamma,\beta}(x,y)$ having the following property (1) with the assumptions given in the introduction. This is in keeping with our second objective, which is to introduce some areas in which such results are useful.
The reader who is especially interested in genetic algorithms has to consult corollary 1 propositions 3 and 4. Finally, the numerical solving of mean cost optimization problems only requires the use of Theorem 2 or corollary 4.

1.1 Relative Entropy Convergence

Our analysis will be based entirely on considerations of the time continuous semigroup associated to the Markov kernels $Q_{\gamma,\beta}(x,y)$ introduced in (1). Namely, define, for $f : E \to \mathbb{R}$

$$L_{\gamma,\beta} f(x) = \sum_{y \in E} (f(y) - f(x))\, Q_{\gamma,\beta}(x,y)$$

For a probability measure m on E, an inverse-freezing schedule $\beta \in C^1(\mathbb{R}_+, \mathbb{R}_+)$ and $\gamma \in C^1(\mathbb{R}_+, \mathbb{R}_+)$, we denote (Ω, P, F_t, X_t) the canonical process associated to the family of generators $(L_{\gamma_t,\beta_t})_{t\geq 0} = (Q_{\gamma_t,\beta_t} - I)_{t\geq 0}$ whose initial condition is $m_0 = m$ and we write m_t the distribution of X_t.
Whenever X is time homogeneous (i.e. $\beta_t = \beta$ and $\gamma_t = \gamma$) it is well known that $L_{\gamma,\beta}$ has a unique invariant probability measure $\pi_{\gamma,\beta}$ so that

$$\forall f : E \to \mathbb{R} \qquad \pi_{\gamma,\beta}(L_{\gamma,\beta} f) = 0$$

and $\pi_{\gamma,\beta}$ charges all the points. It is also convenient to recall the notion of log-Sobolev constant of $Q_{\gamma,\beta}$. Namely,

$$a(\gamma,\beta) \stackrel{\text{def}}{=} \min\left\{\mathcal{E}_{\gamma,\beta}(f,f)/\mathcal{L}_{\gamma,\beta}(f) \quad , \ \mathcal{L}_{\gamma,\beta}(f) \neq 0\right\}$$

where the Dirichlet form $\mathcal{E}_{\gamma,\beta}$ and $\mathcal{L}_{\gamma,\beta}$ are defined by

$$\begin{aligned}
\mathcal{E}_{\gamma,\beta}(f,g) &= -\langle L_{\gamma,\beta} f, g\rangle_{\pi_{\gamma,\beta}} = -\sum_{x\in E} L_{\gamma,\beta} f(x)\, g(x)\, \pi_{\gamma,\beta}(x) \\
\mathcal{L}_{\gamma,\beta}(f) &= \sum_{x\in E} f(x)^2 \log\left(f(x)^2/\|f\|^2_{2,\pi_{\gamma,\beta}}\right) \pi_{\gamma,\beta}(x)
\end{aligned}$$

Let us recall the notion of relative entropy of a measure m with respect to a measure π charging all the points

$$\mathrm{Ent}_\pi(m) = \sum_{x\in E} m(x)\ \log\left(m(x)/\pi(x)\right)$$

Using these notation, whenever X is time homogeneous, one has the following basic inequality (for instance see Miclo [10])

$$\frac{d}{dt}\mathrm{Ent}_{\pi_{\gamma,\beta}}(m_t) \leq -2\, a(\gamma,\beta)\, \mathrm{Ent}_{\pi_{\gamma,\beta}}(m_t) \tag{3}$$

For an expository paper on log-Sobolev constants for general Markov chain on finite spaces the reader is referred to Diaconis/Saloff-Costes [5]. Holley and Stroock use in [8] Sobolev and log-Sobolev inequalities to study the standard simulated annealing. For another approach using only spectral gap estimates the reader should consult [4]. Using log-Sobolev inequalities one of the authors addressed the convergence of a simulated annealing associated to a Markov transition kernel of the form using the entropy distance to stationarity (see [10]). The purpose of this section is to extend these results to general Markov transition kernels of the form (1).
What follows is an exposition of some basic results regarding the description of $\pi_{\gamma,\beta}$ which is due to Bott and Mayberry [1] and also exposed in Freidlin-Wentzell [6]. For $x \in E$ we denote $G_E(x)$, or $G(x)$ when there is no possible confusions, the set of x-graphs. We shall also use the notations, for $x \in E$ and $g \in G(x)$

$$\begin{aligned} R_{\gamma,\beta}(x) &= \sum_{g\in G(x)} Q_{\gamma,\beta}(g), \qquad Q_{\gamma,\beta}(g) = \prod_{(y\to z)\in g} Q_{\gamma,\beta}(y,z) \\ V_\gamma(g) &= \sum_{(y\to z)\in g} V_\gamma(y,z), \qquad \mathbf{Q}_{\gamma,\beta}(x,y) = q(x,y)\, e^{-\beta\, V_\gamma(x,y)} \end{aligned}$$

Then, whenever X is time homogeneous, its invariant distribution $\pi_{\gamma,\beta}$ is given by

$$\pi_{\gamma,\beta}(x) = R_{\gamma,\beta}(x)/\sum_{z\in E} R_{\gamma,\beta}(z)$$

Similarly, let $\mu_{\gamma,\beta}$ be the invariant probability measure of

$$\mathbf{L}_{\gamma,\beta} f(x) = \sum_{y\in E} (f(y) - f(x))\, \mathbf{Q}_{\gamma,\beta}(x,y)$$

If $\mathbf{a}(\gamma,\beta)$ is the Log-Sobolev constant of $\mathbf{Q}_{\gamma,\beta}$ then under assumption (1) there exists some constant $k_1 > 0$ such that

$$k_1^{-1}\mu_{\gamma,\beta}(x) \leq \pi_{\gamma,\beta}(x) \leq k_1 \mu_{\gamma,\beta}(x) \tag{4}$$

Now, from the very definition of $\mu_{\gamma,\beta}$ and (4), we have the estimate

$$-\beta^{-1} \log \pi_{\gamma,\beta}(x) \underset{\beta\to+\infty}{\longrightarrow} V_\gamma(x) - \min_{z\in E} V_\gamma(z) \quad \text{with} \quad V_\gamma(x) \stackrel{\text{def}}{=} \min_{g\in G(x)} V_\gamma(g)$$

As a direct consequence of lemma 3.3, Diaconis/Saloff-Costes [5], and the inequalities (1) and (4) there exist some constant $B > 0$ such that

$$B^{-1}\, \mathbf{a}(\gamma,\beta) \leq a(\gamma,\beta) \leq B\, \mathbf{a}(\gamma,\beta)$$

Finally, by Theorem 3.23 of Holley/Stroock [8] and the inequalities stated in Miclo [10]

Proposition 1 *There exists some constant $A > 0$ such that $a(\gamma,\beta) \geq A\ \dfrac{e^{-\beta\, c(\gamma)}}{1+\beta}$, where $c(\gamma)$ is the critical height associated to the communication cost V_γ given by*

$$\begin{aligned} c(\gamma) &= \max_{x,y\in E}\left(\min_{p\in\mathcal{S}_{x,y}} e_\gamma(p) - V_\gamma(x) - V_\gamma(y)\right) + \min_{z\in E} V_\gamma(z) \\ e_\gamma(p) &= \max_{1\leq i\leq n} \min\left(V_\gamma(p_{i-1}) + V_\gamma(p_{i-1},p_i), V_\gamma(p_i) + V_\gamma(p_i,p_{i-1})\right) \end{aligned}$$

where $\mathcal{S}_{x,y}$ is the set of all the finite sequences from x to y and $e_\gamma(p)$ denotes elevation of a path. It can also be shown that

$$c(\gamma) = \max_{x,y\in E}\left(\min_{p\in C_{x,y}} \tilde{e}_\gamma(p) - V_\gamma(x) - V_\gamma(y)\right) + \min_{z\in E} V_\gamma(z)$$

where

$$\tilde{e}_\gamma(p) = \max_{1\leq i\leq n} V_\gamma(p_{i-1}) + V_\gamma(p_{i-1},p_i)$$

and $C_{x,y}$ is the set of all paths (admissible for q) from x to y.

By choosing $t \to (\gamma_t, \beta_t)$ to go to infinity and using (3) we arrive at

$$\frac{d}{dt}\mathrm{Ent}_{\pi_{\gamma_t,\beta_t}}(m_t) \leq -2\ A\ \frac{e^{-\beta_t\, c(\gamma_t)}}{1+\beta_t}\mathrm{Ent}_{\pi_{\gamma_t,\beta_t}}(m_t) - \sum_{x\in E} m_t(x)\frac{d}{dt}\log \pi_{\gamma_t,\beta_t}(x) \qquad (5)$$

Therefore it remains to estimate the derivatives $d\pi_{\gamma,\beta}/d\beta$ and $d\pi_{\gamma,\beta}/d\gamma$. For this purpose, write

$$\overline{Q}_{\gamma,\beta}(g) \;=\; Q_{\gamma,\beta}(g)/\sum_{h\in I_x} Q_{\gamma,\beta}(h) \qquad I_x = \{g\in G(x)\ :\ \prod_{(y\to z)\in g} q(y,z) > 0\}$$

By a simple analysis it is easily checked that

$$\begin{aligned} \frac{d}{d\beta}\log R_{\gamma,\beta}(x) &= \sum_{g\in I_x} \overline{Q}_{\gamma,\beta}(g)\frac{d}{d\beta}\log Q_{\gamma,\beta}(g) \\ \frac{d}{d\beta}\log \pi_{\gamma,\beta}(x) &= \sum_{z\in E}\left(\frac{d}{d\beta}\log R_{\gamma,\beta}(x) - \frac{d}{d\beta}\log R_{\gamma,\beta}(z)\right)\pi_{\gamma,\beta}(z) \end{aligned}$$

In order to derive a useful inequality we assume there exist two functions $d_1,\ d_2 : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\sup_{g\in\cup_x I_x} \frac{d}{d\beta}\log Q_{\gamma,\beta}(g) - \inf_{g\in\cup_x I_x} \frac{d}{d\beta}\log Q_{\gamma,\beta}(g) \;\leq\; d_1(\gamma) \qquad (6)$$

$$\sup_{g\in\cup_x I_x} \frac{d}{d\gamma}\log Q_{\gamma,\beta}(g) - \inf_{g\in\cup_x I_x} \frac{d}{d\gamma}\log Q_{\gamma,\beta}(g) \;\leq\; d_2(\beta) \qquad (7)$$

Remark, by the very definitions of the sets I_x, that it clearly suffices to have

$$\left|\frac{d}{d\beta}\log Q_{\gamma,\beta}(x,y)\right| \leq \tilde{d}_1(\gamma) \qquad \left|\frac{d}{d\gamma}\log Q_{\gamma,\beta}(x,y)\right| \leq \tilde{d}_2(\beta)$$

for every $q(x,y) > 0$ and two functions $\tilde{d}_1,\ \tilde{d}_2 : \mathbb{R}_+ \to \mathbb{R}_+$. Now we are ready to apply the following technical lemma

Lemma 1 (Stroock [11]) *If μ is a probability measure and $\theta \in L^1(\mu)^+$ satisfies $\int \theta d\mu = 1$, then for every $\phi \in L^\infty(\mu)$ satisfying $\int \phi d\mu = 0$, one has*

$$\left|\int \phi\ \theta\ d\mu\right| \leq \sqrt{2}\ \|\phi\|_{L^\infty(\mu)}\ \left(\int \theta\ \log\theta\ d\mu\right)^{1/2}$$

Using this lemma with $\mu = \pi_{\gamma,\beta}$, $\phi = d\log\pi_{\gamma,\beta}/d\beta$ (resp. $\phi = d\log\pi_{\gamma,\beta}/d\gamma$) and $\theta = m_t/\pi_{\gamma_t,\beta_t}$ and writing

$$I_t \stackrel{\text{def}}{=} \left(\text{Ent}_{\pi_{\gamma_t,\beta_t}}(m_t)\right)^{1/2}$$

we see from (5) that

$$\frac{dI_t}{dt} \leq -A_0\ \frac{e^{-\beta_t\, c(\gamma_t)}}{1+\beta_t} I_t + A_1\, d_1(\gamma_t)\ \left|\frac{d\beta_t}{dt}\right| + A_2\, d_2(\beta_t)\ \left|\frac{d\gamma_t}{dt}\right| \tag{8}$$

for some constants $A_0, A_1, A_2 > 0$. Hence by taking, for t sufficiently large, an inverse-temperature schedule of the form $\beta_t = K^{-1}\log t$ we obtain

$$\frac{dI_t}{dt} \leq -A_t\ I_t + B_t$$

with

$$A_t = A_0\ t^{-c(\gamma_t)/K}(1+K^{-1}\log t)^{-1} \quad \text{and} \quad B_t = A_1\frac{d_1(\gamma_t)}{Kt} + A_2\, d_2(K^{-1}\log t)\ \left|\frac{d\gamma_t}{dt}\right|$$

Now, it is well known that

$$\exists t_0 \in \mathbb{R}_+ \quad \int_{t_0}^{+\infty} A_s\ ds = +\infty \qquad \lim_{t\to+\infty}\frac{B_t}{A_t} = 0 \Longrightarrow \lim_{t\to+\infty} I_t = 0$$

We can now summarize the entire consideration in the following way.

Theorem 1 *Assume that $c = \limsup_{\gamma\to+\infty} c(\gamma) < +\infty$ and the following conditions are satisfied.*

$$\sup_{x\in E}\left|\frac{d}{d\beta}\mu_{\gamma,\beta}(x)\right| \leq d_1(\gamma) \qquad \sup_{x\in E}\left|\frac{d}{d\gamma}\mu_{\gamma,\beta}(x)\right| \leq d_2(\beta)$$

for two non-negative functions d_1, d_2. When the inverse freezing schedule has parametric form $\beta_t = K^{-1}\log t$, for t sufficiently large and $K > c$, we have

$$\lim_{t\to+\infty} Ent_{\pi_{\gamma_t,\beta_t}}(m_t) = 0 \tag{9}$$

whenever

$$\frac{d\gamma_t}{dt} = o\left(1/(d_2(\log t^{1/K})t^{c/K}\log t)\right) \qquad d_1(\gamma_t)/t = o\left(1/(t^{c/K}\log t)\right) \tag{10}$$

<u>*Remark*</u> *:*

The hypothesis above might seem difficult to check; we shall see in fact that in many cases they are indeed fulfilled. In practice the Markov kernels $Q_{\gamma,\beta}$ are known and it clearly suffices to check that

$$\begin{aligned}
\sup_{g\in\cup_x I_x} \frac{d}{d\beta}\log Q_{\gamma,\beta}(g) - \inf_{g\in\cup_x I_x} \frac{d}{d\beta}\log Q_{\gamma,\beta}(g) &\leq d_1(\gamma)\\
\sup_{g\in\cup_x I_x} \frac{d}{d\gamma}\log Q_{\gamma,\beta}(g) - \inf_{g\in\cup_x I_x} \frac{d}{d\gamma}\log Q_{\gamma,\beta}(g) &\leq d_2(\beta)
\end{aligned}$$

Because of the general orientation provided in this section we can proceed immediately to review the most important special class of generalized simulated annealing which we shall study latter. Let us consider the transition probability kernels:

$$Q_\beta(x,y) = \sum_{u\in U} \overline{q}_\beta(x,u,y)e^{-\beta\overline{V}(x,u,y)} \tag{11}$$

where U is a given finite set, $\overline{V} : U\times E^2 \to \mathbb{R}_+$ and $\overline{q}_\beta(x,u,y) \geq 0$. The proof of Theorem 1 shows the importance that the Markov kernel q is irreducible. For this purpose we will assume the existence of non negative functions $\overline{q}(x,u,y)$ so that

$$(\mathcal{C}) \qquad \overline{q}_\beta(x,u,y) > 0 \Longleftrightarrow \overline{q}(x,u,y) > 0 \qquad \lim_{\beta\to+\infty} \overline{q}_\beta(x,u,y) = \overline{q}(x,u,y)$$

and we will work with the following irreducibility condition

$$(\mathcal{I}) \quad \forall x,y\in E \quad \exists (p_k,u_k)_{1\leq k\leq r} : \ p_0 = x \quad p_k\in E\ u_k\in U\ p_r = y\ \overline{q}(p_k,u_k,p_{k+1}) > 0$$

Our immediate goal is to prove the following consequence of Theorem 1 which gives conditions assuring the convergence stated in Theorem 1 for a time-inhomogeneous Markov chain with transitions (11). This corollary is applied in section 2 to the convergence of genetic algorithms

Corollary 1 *Assume $\overline{q}_\beta$ and $\overline{q}$ satisfy the continuity and irreducibility conditions $(\mathcal{C})$ and $(\mathcal{I})$. Then, the transition probabilities*

$$Q_\beta(x,y) = \sum_{u\in U} \overline{q}_\beta(x,u,y)e^{-\beta\overline{V}(x,u,y)} \tag{12}$$

satisfies the inequalities (1) with

$$\begin{aligned} V(x,y) &= \min_{u\in U(x,y)} \overline{V}(x,u,y) \qquad U(x,y)=\{u\in U \ : \ \overline{q}(x,u,y)>0\} \\ q(x,y) &= \sum_{u\in U^\star(x,y)} \overline{q}(x,u,y) \qquad U^\star(x,y)=\{u\in U(x,y) \ : \ \overline{V}(x,u,y)=V(x,y)\} \end{aligned}$$

Let V be the virtual energy function corresponding to the above communication cost function

$$V(x)=\min_{g\in G(x)} \sum_{(y\to z)} V(y,z)$$

and let c be the corresponding critical height. In addition, suppose that for every $\overline{q}(x,u,y)>0$ and for some $\beta_0\geq 0$

$$\sup_{\beta\geq\beta_0} |\frac{d\log\overline{q}_\beta}{d\beta}(x,u,y)|<+\infty \tag{13}$$

In this case, if m is a probability measure on E and if β_t assumes the parametric form $\beta_t=K^{-1}\log t$, for sufficiently large t, with $K>c$ then

$$\lim_{t\to+\infty} Ent_{\pi_{\beta_t}}(m_t)=0 \qquad \textit{and} \qquad \lim_{t\to+\infty} P(X_t\in V^\star)=1$$

where (Ω,P,F_t,X_t) is the canonical process associated to the family of generators $(L_{\beta_t})_{t\geq 0}=(Q_{\beta_t}-I)_{t\geq 0}$ whose initial condition is $m_0=m$, m_t is the distribution of X_t, π_β is the unique invariant probability measure of L_β and

$$V^\star=\{x\in E \ : \ V(x)=\min_E V\}$$

proof:
Using the form of $\overline{q}_\beta$ we have for some suitable function $\epsilon(\beta)\to 0$, as $\beta\to+\infty$,

$$(1-\epsilon(\beta))\ K_\beta(x,y)\leq Q_\beta(x,y)\leq(1+\epsilon(\beta))\ K_\beta(x,y) \tag{14}$$

where

$$K_\beta(x,y)=\sum_{u\in U(x,y)} \overline{q}(x,u,y)e^{-\beta\overline{V}(x,u,y)}$$

But now we have that

$$\begin{aligned} K_\beta(x,y) &= \sum_{u\in U^\star(x,y)} \overline{q}(x,u,y)\ e^{-\beta V(x,y)} + e^{-\beta V(x,y)} \sum_{u\in U(x,y)-U^\star(x,y)} \overline{q}(x,u,y)\ e^{-\beta(\overline{V}(x,u,y)-V(x,y))} \\ &= q(x,y)e^{-\beta V(x,y)} + e^{-\beta V(x,y)} \sum_{u\in U(x,y)-U^\star(x,y)} \overline{q}(x,u,y)\ e^{-\beta(\overline{V}(x,u,y)-V(x,y))} \end{aligned}$$

Thus, condition $(\mathcal{I})$ implies that q is irreducible. Furthermore, if we write,

$$\begin{aligned} I &= \{(x,y) \in E^2 \ : \ U(x,y) \neq \emptyset\} \\ J &= \{(x,u,y) \in E \times U \times E \ : \ (x,y) \in I \ \ u \in U(x,y)\} \end{aligned}$$

and

$$\begin{aligned} h_1 &= \min_{(x,y)\in I} \sum_{u \in U(x,y) - U^\star(x,y)} \overline{q}(x,u,y)/q(x,y) \\ h_2 &= \min_{(x,u,y)\,:\,u \notin U^\star(x,y)} \overline{V}(x,u,y) - V(x,y) \end{aligned}$$

using (14) we get the system of inequalities

$$(1-\epsilon(\beta))\; q(x,y)\; e^{\beta V(x,y)} \leq Q_\beta(x,y) \leq (1+\epsilon(\beta))\; (1 + h_1\; e^{-\beta h_2}) q(x,y)\; e^{\beta V(x,y)}$$

To end the proof it remains to check condition (6). Choose $q(x,y) > 0$, after some computations we find

$$\left|\frac{d \log Q_\beta(x,y)}{d\beta}\right| \leq \sup_{u \in U(x,y)} \left|\frac{d \log \overline{q}_\beta}{d\beta}(x,u,y)\right| + \sup_{u\in U(x,y)} \overline{V}(x,u,y)$$

thus (13) implies (6) and using theorem 1 the proof of the first assertion is complete. Examination of the invariant distribution of L_β soon yields that for all $x \notin V^\star$ we have $\lim_{\beta\to+\infty} \pi_\beta(x) = 0$. Then, to prove the last assertion it is enough to recall the basic inequality

$$\|m_t - \pi_{\beta_t}\|_{TV}^2 \leq 2 \ \mathrm{Ent}_{\pi_{\beta_t}}(m_t)$$

where $\|\centerdot\|_{TV}$ is the distance in total variation given by

$$\|\mu - \nu\|_{TV} = 2 \sup_{A \subset E} |\mu(A) - \nu(A)|$$

∎

It is quite transparent from definition (12) that the following situations are covered

$$Q_\beta = Q_\beta^{(1)} Q_\beta^{(2)} \quad \text{and} \quad Q_\beta = \alpha\, Q_\beta^{(1)} + (1-\alpha)\, Q_\beta^{(2)} \quad 0 < \alpha < 1$$

In section 2 we will develop properties of both class of chains which we shall find includes, as a special case, the evolutionary processes studied by Cerf in [2]. For the sake of unity and to highlight issues specific to evolutionary processes, we give some examples to suggest how these results translates in this special situation.

Examples:

1. If $d_1(\gamma) = \gamma^p$ and $d_2(\beta) = \beta^q$ for some $p \geq 0$ and $q > 0$, it clearly suffices to choose $\gamma_t = \log t$.

2. Let us study a way to combine the transitions $Q_\beta^{(1)}, \ldots, Q_\beta^{(r)}$ given by

$$Q_\beta^{(k)}(x,y) = q_\beta^{(k)}(x,y)\, e^{-\beta V^{(k)}(x,y)} \qquad V^{(k)} : E^2 \to \mathbb{R}_+ \qquad q_\beta^{(k)}(x,y) \geq 0$$

It is transparent from (11) that the following situations are covered:

a) $Q_\beta(x,y) = Q_\beta^{(1)} \ldots Q_\beta^{(r)}(x,y) = \sum_{z_1,\ldots,z_{r-1} \in E} Q_\beta^{(1)}(x,z_1) \ldots Q_\beta^{(r)}(z_{r-1},y)$

This situation can be formulated in the form (11) with $U = E^{r-1}$ and

$$\overline{q}_\beta(x,u,y) = q_\beta^{(1)}(x,u_1) \ldots q_\beta^{(r)}(u_{r-1},y) \qquad \overline{V}(x,u,y) = V^{(1)}(x,u_1) + \ldots + V^{(r)}(u_{r-1},y)$$

b) $\tilde{Q}_\beta(x,y) = \sum_{k=1}^{r} \alpha_k\, Q_\beta^{(k)}(x,y) \qquad \text{with} \qquad \sum_{k=1}^{r} \alpha_k = 1$

This situation can be formulated in the form (11) with $E = \{1, \ldots, r\}$ and

$$\overline{q}_\beta(x,u,y) = \alpha_u\, q_\beta^{(u)}(x,y) \qquad \overline{V}(x,u,y) = V^{(u)}(x,y)$$

Probabilistically and in precise language 1-a) has the interpretation of being the transition of a chain obtained through overlapping $r-1$ other chains and 1-b) has the interpretation of being the transition of a chain obtained through choosing randomly at each step among r chains. Let us remark, by way of illustration, that it is also possible to consider a way of combining 1-a and 1-b which subsumes such parallel and series combinations. For instance each transition probability $Q_\beta^{(k)}$ in the expression (a) may be of type (b) and conversely. As a result one has a great freedom in the design and the physical construction of the transition probabilities $Q_{\gamma,\beta}$ and they appears ideally suited to describe a large class of processes encountered in applications.

3. Let us examine the above example when $r = 2$. In this situation we introduce the irreducibility and continuity conditions

$$(\mathcal{I})' \qquad q^{(1)} \;\text{ irreducible and }\; \forall x \in E \qquad q^{(2)}(x,x) > 0 \tag{15}$$

$$(\mathcal{C})' \qquad \lim_{\beta \to +\infty} q_\beta^{(k)}(x,y) = q^{(k)}(x,y) \quad q_\beta^{(k)}(x,y) = 0 \Longleftrightarrow q^{(k)}(x,y) = 0 \quad \forall k = 1,2 \tag{16}$$

where $q^{(k)}(x,y)$ are transition probability kernels such that

$$q^{(k)}(x,y) = 0 \Longleftrightarrow V^{(k)}(x,y) = +\infty \qquad \forall k = 1,2$$

In this situation the conditions $(\mathcal{C})$ and $(\mathcal{I})$ of corollary 1 are satisfied. In addition, if we assume that for every $k = 1,2$ and $q^{(k)}(x,y)$

$$\left|\frac{d \log q_\beta^{(2)}}{d\beta}(x,y)\right| < +\infty$$

then the last condition (13) introduced in corollary 1 is satisfied.

Corollary 2 *Suppose the Markov kernel of the chain has the form*

$$Q_{\gamma,\beta} = q_\beta(x,y)e^{-\beta V_\gamma(x,y)}$$

with

$$q_\beta(x,y) = 0 \Longleftrightarrow q(x,y) = 0 \qquad \lim_{\beta\to+\infty} q_\beta(x,y) = q(x,y)$$

and assume the following conditions are satisfied for every $\beta, \gamma \in \mathbb{R}_+$ and some constant $d > 0$

$$c = \sup_\gamma c(\gamma) < +\infty \qquad \sup_{g\in\cup_x I_x} V_\gamma(g) - \inf_{g\in\cup_x I_x} V_\gamma(g) \le d$$

$$\sup_{g\in\cup_x I_x} \frac{d}{d\beta}\log q_\beta(g) - \inf_{g\in\cup_x I_x} \frac{d}{d\beta}\log q_\beta(g) \le d \quad \sup_{g\in\cup_x I_x} \frac{d}{d\gamma}V_\gamma(g) - \inf_{g\in\cup_x I_x} \frac{d}{d\gamma}V_\gamma(g) \le d$$

When the inverse freezing schedule has parametric form $\beta_t = K^{-1}\log t$, for t sufficiently large and $K > c$, we have

$$\lim_{t\to+\infty} Ent_{\pi_{\gamma_t,\beta_t}}(m_t) = 0 \qquad \textit{whenever} \quad \frac{d\gamma_t}{dt} = o\left(1/(t^{c/K}\log^2 t)\right)$$

As a consequence we have the well known

Corollary 3 (Miclo [10]) *Suppose the Markov kernel of the chain has the form*

$$Q_{\gamma,\beta} = q(x,y)e^{-\beta V(x,y)} \stackrel{\text{def}}{=} Q_\beta(x,y)$$

When the inverse freezing schedule has parametric form $\beta_t = K^{-1}\log t$, for t sufficiently large and $K > c$, we have

$$\lim_{t\to+\infty} Ent_{\pi_{\beta_t}}(m_t) = 0$$

where π_β is the unique invariant probability measure of the Markov generator $Q_\beta - I$.

The usefulness of Theorem 1 will now be illustrated in the case where the transition probabilities $Q_{\gamma,\beta}$ converge to a transition probability kernel Q_β as $\gamma \to +\infty$.

Theorem 2 *Let $Q_\beta(x,y)$ be a Markov kernel such that $Q_\beta(x,y) = 0 \Longleftrightarrow q(x,y) = 0$. Suppose the assumptions of Theorem 1 are satisfy and for every $q(x,y) > 0$*

$$\lim_{t\to+\infty} |\log Q_{\gamma_t,\beta_t}(x,y) - \log Q_{\beta_t}(x,y)| = 0 \tag{17}$$

Then

$$\lim_{t\to+\infty} Ent_{\pi_{\beta_t}}(m_t) = 0 \tag{18}$$

where π_β is the unique invariant probability measure of the Markov generator $Q_\beta - I$.

proof:
By the same line of arguments as before π_β may be described as follows

$$\pi_\beta(x) = \frac{R_\beta(x)}{\sum_{z\in E} R_\beta(z)} \quad \text{with} \quad R_\beta(x) = \sum_{g\in I_x} Q_\beta(g) \quad \text{and} \quad Q_\beta(g) = \prod_{(y\to z)\in g} Q_\beta(y,z) \tag{19}$$

It follows that

$$\log \frac{\pi_{\gamma,\beta}}{\pi_\beta}(x) = -\log\big(\sum_{g\in I_x} \prod_{(y\to z)\in g} (Q_\beta Q^{-1}_{\gamma,\beta})(y,z)\, \overline{Q}_{\gamma,\beta}(g)\big)$$

$$-\log\big(\sum_{x'\in E, g\in I_{x'}} \prod_{(y\to z)\in g} (Q_{\gamma,\beta} Q^{-1}_{\beta})(y,z)\, \tilde{Q}_{\beta}(g)\big)$$

where

$$\overline{Q}_{\gamma,\beta}(g) = Q_{\gamma,\beta}(g) / \sum_{h\in I_x} Q_{\gamma,\beta}(h) \qquad \text{and} \qquad \tilde{Q}_\beta(g) = Q_\beta(g) / \sum_{z\in E, h\in I_z} Q_\beta(h)$$

By Jensen's inequality, we have

$$\log \frac{\pi_{\gamma,\beta}}{\pi_\beta}(x) \le 2 \sup_{g\in \cup_z I_z} |\log Q_{\gamma,\beta}(g) - \log Q_\beta(g)|$$

Finally, we obtain

$$\mathrm{Ent}_{\pi_{\beta_t}}(m_t) \le \mathrm{Ent}_{\pi_{\gamma_t,\beta_t}}(m_t) + 2 \sup_{g\in\cup_x I_x} |\log Q_{\gamma_t,\beta_t}(g) - \log Q_{\beta_t}(g)| \tag{20}$$

Using (9) the proof is complete.

∎

We now make the above observations precise by considering more specific, though general, transitions $Q_{\gamma,\beta}$.

Corollary 4 *Let V be a non negative function $V : E\times E \to \mathbb{R}_+$ and $Q_{\gamma,\beta}(x,y) = q_\beta(x,y)\, e^{\beta V_\gamma(x,y)}$.*
Suppose the assumptions of Corollary 2 are satisfied and, for every $q(x,y) > 0$,

$$|V_{\gamma_t}(x,y) - V(x,y)| = o(1/\log t) \tag{21}$$

When the inverse freezing schedule has parametric form $\beta_t = K^{-1}\log t$, for t sufficiently large and $K > c$, we have

$$\lim_{t\to+\infty} Ent_{\pi_{\beta_t}}(m_t) = 0 \qquad \textit{and} \qquad \lim_{t\to+\infty} P(X_t \in V^\star) = 1 \tag{22}$$

where π_β is the unique invariant probability measure of the Markov generator $Q_\beta - I$ with

$$Q_\beta(x,y) = q(x,y)\; e^{-\beta V(x,y)} \qquad V^\star = \{x \in E:\; V(x) = \min_E V\}$$

and

$$V(x) = \min_{g \in G(x)} \sum_{(y \to z) \in g} V(y,z)$$

If $V_{\gamma_t}(x,y) = (U_t(y) - U_t(x))^+$ with $U_t : E \to \mathbb{R}_+$ and $\lim_{t\to+\infty} U_t(x) = U(x)$ condition (21) takes the form

$$\lim_{t\to+\infty} \log t\; |U_t(x) - U(x)| = 0$$

The following examples illustrate the results and the conditions stated in the above theorems.

Examples:

1. Let us now turn our attention to the transition probability kernel

   $$Q_{\gamma,\beta}(x,y) = q(x,y)\; e^{-\beta V_\gamma(x,y)}$$

   where

   $$V_\gamma(x,y) = (U_\gamma(y) - U_\gamma(x))^+ \;\text{ if } q(x,y) > 0, \text{ and } \;\; +\infty \;\text{ otherwise}$$

   In this case the conditions (6) and (7) take the form

   $$\sup_{x\in E} U_\gamma(x) - \inf_{x\in E} U_\gamma(x) \le d_1(\gamma) \qquad \sup_{x\in E} \frac{d}{d\gamma} U_\gamma(x) - \inf_{x\in E} \frac{d}{d\gamma} U_\gamma(x) \le d_2(\beta)/\beta$$

   In addition, if

   $$\sup_{x,\gamma} U_\gamma(x) < +\infty \qquad \sup_{x,\gamma} \frac{d}{d\gamma} U_\gamma(x) < +\infty$$

   Then we can choose $d_2(\beta) = \beta\; d_2$ and $d_1(\gamma) = d_1 < +\infty$ ($d_1, d_2 > 0$) and the condition (10) takes the form

   $$\frac{d\gamma_t}{dt} = o\left(1/(t^{c/K} (\log t)^2)\right)$$

2. Let us examine the above example with time-inhomogeneous potential given by

   $$U_\gamma(x) = \theta(\gamma)^{-1} \int_0^{\theta(\gamma)} C(s,x)\; ds$$

   with $C : \mathbb{R}_+ \times E \to \mathbb{R}_+$ bounded and $\dfrac{d}{d\gamma} \log \theta(\gamma) < +\infty$.
   From the boundedness of C we can choose constants $d_1, d_2 > 0$ so that $d_1(\gamma) = d_1$

and $d_2(\beta) = \beta\ d_2$ satisfy the required conditions. Finally, if $\gamma_t = A\ \log t$ and $\theta(\gamma) = e^\gamma$ then we have

$$\frac{d\gamma_t}{dt} = A/t = o\left(1/(t^{c/K}(\log t)^2)\right) \qquad \frac{d}{d\gamma}\log\theta(\gamma) = 1$$

and

$$U_{\gamma_t}(x) = \frac{1}{t^A}\int_0^{t^A} C(s,x)\,ds$$

1.2 General results on the virtual energy

In section 1 we proved the convergence in probability of a class of stochastic algorithms to the set of the global minima $V^\star$ of a virtual energy function V. The crucial question is of course to estimate $V^\star$. In the case where q is symmetric and $V(x,y) = (U(y) - U(x))^+$ with $U : E \to \mathbb{R}_+$ it is well known that $V^\star = U^\star$. For a generalization of this see Trouvé [12]. The situation becomes considerably more involved when dispensing with the above assumptions. The purpose of this section is to introduce a natural test set approach to study $V^\star$. More precisely, we will give several conditions for a given subset $H \subset E$ to contains $V^\star$.

Let us recall some basic definitions. Let E be a finite set and q an irreducible Markov kernel. Assume a given function $V : E \times E \to \overline{\mathbb{R}}_+$ satisfies:

$$V(x,y) < +\infty \Longleftrightarrow q(x,y) > 0$$

Let us write $C_{x,y}$ the paths p in E joining x and y, that is

$$\forall k \in \{0,\ldots,|p|-1\} \quad q(p_k,p_{k+1}) > 0 \quad \text{and} \quad p_0 = x, p_{|p|} = y$$

where $|p|$ is the length of p. For $x,y \in E$, $p \in C_{x,y}$ and $g \in G(x)$ we note

$$V(p) = \sum_{k=0}^{|p|-1} V(p_k,p_{k+1}) \qquad V(g) = \sum_{(y\to z)\in g} V(y,z) \qquad V(x) = \min_{g\in G(x)} V(g)$$

For $H \subset E$ and g an x-graph over H (that is $g \in G_H(x)$) it is convenient to define a new communication cost V_H by making the set H a *taboo* set. Namely, for every $x,y \in H$

$$\begin{aligned} V_H(x,y) &= \min\{V(p) \ :\ p \in C_{x,y} \text{ with } \forall k \in \{1,\ldots,|p|-1\}\ p_k \notin H\} \\ V_H(g) &= \sum_{(y\to z)\in g} V_H(y,z) \end{aligned}$$

It is also convenient to define the virtual energy function associated to V_H

$$\forall x \in H \qquad V_H(x) = \min_{g\in G_H(x)} V_H(g) - \min_{y\in H, h\in G_H(y)} V_H(h)$$

Finally, let us write

$$V(H) = \min_{x \in H} V(x) \qquad \text{and} \qquad V_H^\star = \{x \in H \ : \ V_H(x) = \min_H V_H\}$$

Lemma 2 $\forall x \in H \qquad V_H(x) = V(x) - V(H)$

The lemma is an easy consequence of the following lemma

Lemma 3 *Let Q be an irreducible transition probability over E with invariant measure μ. Let X be a Markov chain with transition probability Q and initial measure m such that $m(x) > 0$ for all $x \in E$. Given a subset $H \subset E$ define*

$$T_1 \;=\; \inf\{n \geq 1 \ : \ X_n \in H\} \qquad \tilde{Q} = P(X_{T_1} = y / X_0 = x)$$

Then $\tilde{Q}$ is an irreducible Markov kernel over H and its invariant probability measure is given by

$$\tilde{\mu}(x) = \mu(x)/\mu(H)$$

There are several ways to prove Lemma 2 and 3. The following is may be the shortest in this context.

proof of lemma 3:

The proof is a consequence of the law of large numbers. Let us set by induction on the parameter $n \geq 1$

$$T_{n+1} = \inf\{k > T_n \ : \ X_k \in H\} \qquad n \geq 1$$

Now, under the above conditions, the random process $Y = (Y_n)_{n \geq 0}$ defined by

$$Y_0 = X_0 \qquad Y_n = X_{T_n} \qquad n \geq 1$$

is an irreducible Markov chain over H with transition probability kernel $\tilde{Q}$. Let $\tilde{\mu}$ be its invariant probability measure. First of all we note that

$$\forall x \in E \qquad \frac{1}{n} \sum_{i=1}^{n} 1_x(Y_i) \underset{n \to +\infty}{\longrightarrow} \tilde{\mu}(x) \qquad \text{P.a.e.}$$

On the other hand we have

$$\frac{1}{n} \sum_{i=1}^{n} 1_x(Y_i) = \frac{T_n}{n} \left(\frac{1}{T_n} \sum_{i=1}^{T_n} 1_x(X_i) \right) \qquad \frac{n}{T_n} = \frac{1}{T_n} \sum_{i=1}^{T_n} 1_H(X_i)$$

and

$$\frac{1}{T_n} \sum_{i=1}^{T_n} 1_x(X_i) \underset{n \to +\infty}{\longrightarrow} \mu(x) \qquad \frac{1}{T_n} \sum_{i=1}^{T_n} 1_H(X_i) \underset{n \to +\infty}{\longrightarrow} \mu(H) \qquad \text{P.a.e.}$$

The lemma follows immediately

■

We come to the proof of lemma 2.

proof of lemma 2:

We shall give a sketch of proof. Let us denote by Q_β the Markov kernel over E given by

$$Q_\beta(x,y) = \begin{cases} |E|^{-1} \exp -\beta V(x,y) & \text{if } x \neq y \\ 1 - |E|^{-1} \sum_{z \in E\, : z \neq x} \exp -\beta V(x,z) & \text{otherwise} \end{cases}$$

Let μ_β be the invariant measure of Q_β. From the description of μ_β in terms of x-graphs over E it is clear that

$$\mu_\beta(x) \underset{\beta \to +\infty}{\sim} C(x) \exp -\beta V(x)$$

for some non negative function $C : E \to \mathbb{R}_+^*$. If one now defines $\tilde{Q}_\beta$ as in lemma 3, by elementary large deviation arguments one sees the equivalence

$$\tilde{Q}_\beta(x,y) \underset{\beta \to +\infty}{\sim} \tilde{q}(x,y) \exp -\beta V_H(x,y) \qquad \forall x,y \in H$$

for some irreducible Markov kernel $\tilde{q} : E \times E \to \mathbb{R}_+$. Therefore if $\tilde{\mu}_\beta$ is the invariant measure of $\tilde{Q}_\beta$ one has for some non negative function $\tilde{C} : E \to \mathbb{R}_+^*$

$$\tilde{\mu}_\beta(x) \underset{\beta \to +\infty}{\sim} \tilde{C}(x) \exp -\beta V_H(x)$$

from which our claim follows easily.

■

We now give a definition that we will use in our formulation of our test set approach to understand the limiting behavior of a generalized simulated annealing.

Definition 1 *Let H be a subset of E. We say that a partition $\mathcal{H} = \{H_1, \ldots, H_n\}$ of H is a V-partition if for every $1 \leq i \leq n$ and $x, y \in H_i$, $x \neq y$, there exists a path $p \in C_{x,y}$ such that*

$$\forall 0 \leq k < |p| \qquad p_k \in H_i \quad \textit{and} \quad V(p_k, p_{k+1}) = 0$$

We observe that for a given subset $H \subset E$, one can always obtain a V-partition. For instance

$$\mathcal{H} = \{\{x\} \, : \, x \in H\}$$

is a V-partition. Given a V-partition $\mathcal{H} = \{H_1, \ldots, H_n\}$ of $H \subset E$, it is convenient to define a new communication cost function $V_{\mathcal{H}}$ by setting for every $x \in H_i$ and $y \in H_j$, $1 \le i, j \le n$

$$
\begin{aligned}
V_{\mathcal{H}}(x, y) \quad = \quad & \{V(p) \ : \ p \in C_{x,y} \ \ \exists 0 \le n_1 < n_2 \le |p| \\
& \forall 0 \le k \le n_1 \ \ p_k \in H_i \qquad \forall n_1 < k < n_2 \ \ p_k \notin H \qquad \forall n_2 \le k \le |p| \ \ p_k \in H_j\}
\end{aligned}
$$

As it is easily seen $V_{\mathcal{H}}(x, y)$ does not depend on the choice of $x \in H_i$ and $y \in H_j$. Moreover we note that if $\mathcal{H} = \{\{x\} \ : \ x \in H\}$ then $V_{\mathcal{H}} = V_H$.
Let $V_{\mathcal{H}}$ be the virtual energy function associated to the communication cost function $V_{\mathcal{H}}$. Namely,

$$
\forall x \in H \qquad V_{\mathcal{H}}(x) = \min_{g \in G_H(x)} V_{\mathcal{H}}(g) \qquad \text{with} \quad V_{\mathcal{H}}(g) = \sum_{(y \to z) \in g} V_{\mathcal{H}}(y, z)
$$

As usually, we put also

$$
V_{\mathcal{H}}^{\star} = \{x \in H \ : \ V_{\mathcal{H}}(x) = \min_{H} V_{\mathcal{H}}\}
$$

Proposition 2 *If $\mathcal{H}$ is a V-partition of a subset $H \subset E$ then we have $V_{\mathcal{H}} = V_H$.*

proof:
Let us prove that for every $x \in H$

$$
\min_{g \in G_H(x)} V_H(g) = \min_{g \in G_H(x)} V_{\mathcal{H}}(g)
$$

$\mathcal{H} = \{H_1, \ldots, H_n\}$ being a V-partition it is clear that $V_{\mathcal{H}} \le V_H$. So our claim will follow provided for every $g \in G_H(x)$ we can build a new $\tilde{g} \in G_H(x)$ such that $V_{\mathcal{H}}(g) \ge V_H(\tilde{g})$. For this purpose let $g \in G_H(x)$ for some $x \in H$ and let $i \in \{1, \ldots, n\}$ such that $x \in H_i$. We are going to construct an x-graph $\tilde{g} \in G_H(x)$ such that $V_{\mathcal{H}}(g) \ge V_H(\tilde{g})$. For this we introduce the set

$$
\Gamma = \{(j \to k) \ : \ \exists (y, z) \in g \text{ such that } \ y \in H_j \text{ and } z \in H_k\}
$$

Obviously Γ is not an i-graph over $\{1, \ldots, n\}$ but examination of Γ soon yields that it contains at least one i-graph G_i. Now for $j \ne i$ and $(j \to k) \in G_i$ (unique) there exists an arrow $(y_j \to z_k) \in g$ such that $y_j \in H_j$ and $z_k \in H_k$.
On the other hand, from the definition of $V_{\mathcal{H}}$ there exists a path $p \in C_{y_j, z_k}$ and $0 \le n_1 < n_2 \le |p|$ such that

$$
\forall 0 \le l \le n_1 \quad p_l \in H_j \qquad \forall n_1 < l < n_2 \quad p_l \notin H \qquad \forall n_2 \le l \le |p| \quad p_l \in H_k
$$

and $V(p) = V_{\mathcal{H}}(y_j, z_k)$. Given such a path $p \in C_{y_j, z_k}$ let us set

$$
\tilde{y}_j = p_{n_1} \qquad \tilde{z}_k = p_{n_2}
$$

Finally, $\mathcal{H}$ being a V-partition there exists for every $1 \leq i \leq n$ an $\tilde{y}_j$-graph $\tilde{g}_j \in G_{H_j}(\tilde{y}_j)$ such that $V(\tilde{g}_j) = 0$, with the convention $\tilde{y}_i = x$.
Using the above construction it is easily seen that the set of arrows

$$\tilde{g} = \bigcup_{i=1}^{n} \tilde{g}_i \quad \bigcup \bigcup_{(j \to k) \in G_i} \{(\tilde{y}_j \to \tilde{z}_k)\}$$

is an x-graph over H and, from the construction of g it follows that

$$V_{\mathcal{H}}(g) \geq V_{\mathcal{H}}(\tilde{g}) = \sum_{(j \to k) \in G_i} V_{\mathcal{H}}(\tilde{y}_j, \tilde{z}_k) = \sum_{(j \to k) \in G_i} V_H(\tilde{y}_j, \tilde{z}_k) = V_H(\tilde{g})$$

this ends the proof.

■

The following concept of λ-stability leads to a natural test set approach to study $V^\star$.

Definition 2 *Let λ be a non negative real number. A subset $H \subset E$ is called λ-stable with respect to a communication cost function V when the following conditions are satisfied:*

1. $\forall x \in H \ \ \forall y \notin H \qquad V(x,y) > \lambda$
2. $\forall x \notin H \ \ \exists y \in H \qquad V(x,y) \leq \lambda$

The importance of the notion of λ-stability resides in the following result which extends lemma 4.1 of Freidlin-Wentzell [6].

Proposition 3 *Let λ be a non negative real number and $H \subset E$. Any λ-stable subset H with respect to V contains $V^\star$ and $V_H = V_{/H}$*

proof:
Let H be a λ-stable subset of E. Let $x \notin H$ and let g be an element of $G(x)$ such that $V(g) = V(x)$. There exists almost one $y \in H$ such that $V(x,y) \leq \lambda$. Now we note p the exit path from y to x extracted from g, that is

$$p_0 = y, p_{|p|} = x \quad \text{and} \quad \forall k = 0, \ldots, |p|-1 \quad (p_k \to p_{k+1}) \in g$$

Write $k_0 = \inf\{k = 0, \ldots, |p|-1 \ : \ p_k \notin H\}$. Let $\overline{g}$ be the graph obtained from g by replacing the arrow $(p_{k_0-1} \to p_{k_0})$ by $(x \to y)$. Then we have $\overline{g} \in G(p_{k_0-1})$ and

$$V(x) = V(g) > V(\overline{g}) \geq V(p_{k_0-1}) \qquad p_{k_0-1} \in H$$

This completes the proof of the first assertion. The last assertion is a clear consequence of lemma 2.

■

Let us now reduce all of the results of this section to a proposition which we shall use for later reference.

Proposition 4 *Let A be a 0-stable subset of E with respect to V and, let $\mathcal{A}$ be a V-partition of A. If H is a λ-stable subset of A with respect to $V_{\mathcal{A}}$ for some $\lambda \geq 0$ then $V^\star \subset H$.*

proof:
Using proposition 3 it is easily seen that $V^\star \subset A$, $V_A = V_{/A} = V_{\mathcal{A}}$ and $V_{\mathcal{A}}^\star \subset H$. So one gets $V^\star = V_{\mathcal{A}}^\star = V_A^\star \subset H$ and the proof of proposition 4 is completed.

■

2 Applications

In this section we examine two applications. In section 2.1 we use the results of the first section to derive a new and simple proof of the convergence of the genetic algorithms. We shall prove this important result in a way different from the original proof of Cerf [2]. The proof splits quite naturally into two distinct parts:

1. We use the relative entropy convergence results stated in the first section to prove the convergence of the algorithm.
2. Then, we investigate the test set approach, introduced in the second part of section 1, to prove that the set of the global minima of the virtual energy is contained in the product of the set of the global minima of the fitness function.

In section 2.2 we apply corollary 4 to construct a stochastic algorithm for the numerical solving of a general mean cost optimization problem.

2.1 Genetic Algorithms

A genetic algorithm is a discrete time Markov process $\widehat{x} = (\widehat{x}_n)_n$ with state space $E = S^N$ ($N > 1$ and S a finite set) and whose transition probabilities G_n include a mutation M_n and a selection S_n mechanism. The N-tuple of elements of S, i.e. the points of the set E, are called particle systems and will be mostly denoted by the letters x, y, z. In what follows, we shall distinguish two kinds of combinations. Namely,

$$\begin{array}{llcl}
a) & P\left(\widehat{x}_n \in dx \,/\widehat{x}_{n-1} = z\right) & = & \displaystyle\int_E M_n(z, dy)\; S_n(y, dx) \\
b) & P\left(\widehat{x}_n \in dx \,/\widehat{x}_{n-1} = z\right) & = & \alpha\; M_n(z, dx) + (1-\alpha)\; S_n(z, dx) \qquad 0 < \alpha < 1
\end{array}$$

<u>**Mutations:**</u>
The mutation transition is modelled by independent motion of each particles, that is

$$M_n(z, dy) = \prod_{p=1}^{N} K_n(z^p, dy^p)$$

where K_n is a Markov kernel over S, $z = (z^p)_{1 \leq p \leq N}$ and $y = (y^p)_{1 \leq p \leq N}$.
Selection:
In the selection transition the particles are chosen randomly and independently in the previous population according a given selection function $F_n : S \to \mathbb{R}^+$. Namely,

$$S_n(y, dx) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{F_n(y^i)}{\sum_{j=1}^{N} F_n(y^j)} \, \delta_{y^i}(dx^p)$$

The study of the convergence, as $N \to +\infty$ or $n \to +\infty$, of such algorithms requires specific developments because each individual particle is no longer Markovian and it is difficult to produce mean error estimates. In [3] one of the authors applies and extended such algorithms to non linear filtering problems. One of the apparent difficulty in establishing a convergence result as $n \to +\infty$ is finding a candidate invariant measure which enable us to describe some aspects of the limiting behavior of the algorithm. To our knowledge, Cerf gives in his PhD. dissertation [2] the first convergence result $n \to +\infty$ for a genetic algorithm to converge in probability to the global minima of a given fitness function. More precisely, he studies the following situation:

1. The state space S is finite and $G_n = M_n \, S_n$
2. The mutation Markov transition kernels $K_n(x^1, y^1)$ are governed by a parameter a and a cooling schedule $\beta(n) : \mathbb{N} \to \mathbb{R}^+$. Namely,

$$M_n(x, y) = Q^{(1)}_{\beta(n)}(x, y) \stackrel{\text{def}}{=} \prod_{p=1}^{N} k_{\beta(n)}(x^p, y^p) \tag{23}$$

   with

$$k_\beta(x^1, y^1) \stackrel{\text{def}}{=} \begin{cases} k(x^1, y^1) \; e^{-a\,\beta} & \text{if} \quad x^1 \neq y^1 \\ 1 - \sum_{z^1 \neq x^1} k(x^1, z^1) \; e^{-a\,\beta} & \text{if} \quad x^1 = y^1 \end{cases} \tag{24}$$

   where k is a given irreducible Markov kernel on the space S.
3. The selection operator is build with a fitness function $f : S \to \mathbb{R}^+$ and a cooling schedule $\beta(n) : \mathbb{N} \to \mathbb{R}^+$. Namely,

$$S_n(x, y) = Q^{(2)}_{\beta(n)}(x, y) \stackrel{\text{def}}{=} \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{e^{-\,\beta(n)\, f(x^i)}}{\sum_{j=1}^{N} e^{-\,\beta(n)\; f(x^j)}} \, 1_{x^i}(y^p) \tag{25}$$

He gives several conditions on the rate of decrease of the cooling schedule $\beta(n) \to +\infty$ to ensure all the particles visit the set of global minima, as times goes on, when the number of particle is greater than a critical value. He carries out in a discrete time setting a precise study using large deviation techniques and the powerful tools developed by Trouvé [12]. Simplifying and extending techniques of Cerf and Trouvé, our results are obtained by using the relative entropy convergence result stated in corollary 1 and by investigating the test set approach introduced in section 1.2.

2.1.1 General results and notations

In this section we will consider genetic algorithms described by the transition probability kernel

$$Q_\beta = Q_\beta^{(1)} Q_\beta^{(2)} \qquad \text{or} \qquad \tilde{Q}_\beta = \alpha_1 \, Q_\beta^{(1)} + \alpha_2 \, Q_\beta^{(2)} \quad \text{with} \quad \alpha_1 + \alpha_2 = 1 \quad \text{and} \quad 0 < \alpha_1 < 1 \tag{26}$$

and non necessarily vanishing mutations. More precisely we assume that the mutation transition kernels k_β in (24) have the following property

$$\exists b > 0 \qquad b^{-1} \; k(x_1, x_2) \, e^{-\beta \, a(x_1,x_2)} \leq k_\beta(x_1, x_2) \leq b \;\; k(x_1, x_2) \, e^{-\beta \, a(x_1,x_2)} \tag{27}$$

where $a : S^2 \to \overline{\mathbb{R}}_+$, $a(x,y) < +\infty \Longleftrightarrow k(x,y) > 0$, k is an irreducible Markov kernel and the relation on S defined by

$$x_1 \sim x_2 \Longleftrightarrow a(x_1, x_2) = 0$$

is an equivalence relation. This leads us naturally to consider the partition $S_1, \ldots, S_{n(a)}$, $n(a) \geq 1$, induced by $\sim$. If x_1 is a typical element of S then the equivalence class of x_1 will be denoted by $S(x_1)$

$$S(x_1) = \{x_2 \in S \; : \; x_1 \sim x_2\}$$

We further require that

$$a(x_1, x_2) = 0 \Longrightarrow f(x_1) = f(x_2)$$

and for some $\beta_0 > 0$

$$\forall x_1, x_2 \in S \qquad \sup_{\beta \geq \beta_0} |d \log k_\beta(x_1, x_2)/d\beta| < +\infty \tag{28}$$

To our knowledge the models of evolutionary processes (26) have not been covered by the literature of genetic algorithms.

Examples:
The following mutation transition kernels have the properties (27) and (28)

1. $k_\beta(x_1, x_2) = \begin{cases} k(x_1, x_2) \, e^{-\beta \, a(x_1,x_2)} & \text{if} \quad a(x_1, x_2) > 0 \\ |S(x_1)|^{-1} \left(1 - \sum_{y_1 \notin S(x_1)} \; k(x_1, y_1) \, e^{-\beta \, a(x_1,y_1)}\right) & \text{otherwise} \end{cases}$

2. $k_\beta(x_1,x_2) = \dfrac{e^{-\beta\, a(x_1,x_2)}\; k(x_1,x_2)}{\sum_{y_1\in S}\; e^{-\beta\, a(x_1,y_1)}\; k(x_1,y_1)} \qquad \forall (x_1,x_2)\in S^2$

Finally, let us note that if a is given by

$$a(x_1,x_2) = a\;(1 - 1_{x_1}(x_2)) \qquad \forall (x_1,x_2)\in S^2\;:\; k(x_1,x_2)>0$$

then the transition probability kernel 1) is clearly the same as the mutation transition probability kernel (24) studied by Cerf.
In this special situation, the first model $Q_\beta = Q_\beta^{(1)} Q_\beta^{(2)}$ is of course identical to Cerf's model of genetic algorithm.

Let us recall some terminology introduced by Cerf in [2]. The cardinality of a set will be denoted by $|\bullet|$. If x and y belongs to $E = S^N$ and $f : S \to \mathbb{R}^+$ we write

$$\begin{aligned} [x] &= \{x_k \;:\; 1\le k\le N\} && f^\star = \{x_1\in S \;:\; f(x_1) = \min_S f\} \\ \widehat{x} &= \{k \;:\; 1\le k\le N\;,\; f(x_k) = \widehat{f}(x)\} && \widehat{f}(x) = \min_{1\le k\le N}\; f(x_k) \\ x(y_1) &= \mathrm{Card}\{k \;:\; 1\le k\le N\;,\; x_k = y_1\} \end{aligned}$$

Using these notations, an easy calculation shows that, for $k=1$ or $k=2$

$$Q_\beta^{(k)}(x,y) = q_\beta^{(k)}(x,y)\; e^{-\beta V^{(k)}(x,y)} \qquad q_\beta^{(k)}(x,y) = q^{(k)}(x,y)\;\theta_\beta^{(k)}(x,y)$$

where

$$\begin{aligned} q^{(1)}(x,y) &= \prod_{i:a(x_i,y_j)>0} k(x_i,y_i) \qquad\qquad q^{(2)}(x,y) = \prod_{i=1}^N \frac{x(y_i)}{|\widehat{x}|} \\ V^{(1)}(x,y) &= \sum_{i=1}^N a(x_i,y_i) \quad \text{if} \quad q^{(1)}(x,y)>0 \\ V^{(2)}(x,y) &= \sum_{i=1}^N \Big(f(y_i) - \widehat{f}(x)\Big) \quad \text{if} \quad q^{(2)}(x,y)>0 \\ \theta_\beta^{(1)}(x,y) &= \prod_{i:a(x_i,y_i)=0} k_\beta(x_i,y_i) \\ \theta_\beta^{(2)}(x,y) &= (1+|\widehat{x}|^{-1}\sum_{k\notin\widehat{x}} \exp -\beta(f(x_k) - \widehat{f}(x)))^{-N} \end{aligned}$$

As usual, we will use the convention:

$$\forall k\in\{1,2\} \qquad V^{(k)}(x,y) = +\infty \Longleftrightarrow q^{(k)}(x,y) = 0$$

The asymptotic mutation dynamics of the genetic algorithms is governed by the kernel k and the function a. The irreducibility condition on the kernel k and the fact that

a is an equivalence relation are sufficient conditions in order to allow the system of particles to visit all the state space E. Thus, using the above notations, it is easily checked that

- $q^{(1)}$ is irreducible and $q^{(2)}(x,x) > 0$ for every $x \in E$.
- For every $k = 1,2$ and $q^{(k)}(x,y) > 0$ we have $\sup_{\beta \geq \beta_0} |d \log \theta_\beta^k(x,y)/d\beta| < +\infty$ for some $\beta_0 \geq 0$.

Then, returning to our general model (12), the conditions introduced in corollary 1 are satisfied in both situations

1. $$Q_\beta(x,y) \stackrel{\text{def}}{=} Q_\beta^{(1)}\, Q_\beta^{(2)}(x,y) = \sum_{u \in U} \overline{q}_\beta(x,u,y) e^{-\beta \overline{V}(x,u,y)}$$

$$\begin{array}{llll} \text{with} \quad U = E & \overline{V}(x,u,y) & = V^{(1)}(x,u) + V^{(2)}(u,y) & \overline{q}(x,u,y) = q^{(1)}(x,u) q^{(2)}(u,y) \\ \text{and} & \theta_\beta(x,u,y) & = \theta_\beta^{(1)}(x,u)\, \theta_\beta^{(2)}(u,y) & \overline{q}_\beta(x,u,y) = \theta_\beta(x,u,y) \overline{q}(x,u,y) \end{array}$$

2. $$\tilde{Q}_\beta(x,y) \stackrel{\text{def}}{=} \alpha_1\, Q_\beta^{(1)}(x,y) + \alpha_2\, Q_\beta^{(2)}(x,y) = \sum_{u \in U} \overline{q}_\beta(x,u,y) e^{-\beta \overline{V}(x,u,y)}$$

$$\begin{array}{llll} \text{with} \quad U = \{1,2\} & \overline{V}(x,u,y) & = V^{(u)}(x,y) & \overline{q}(x,u,y) = \alpha_u\, q^{(u)}(x,y) \\ \text{and} & \theta_\beta(x,u,y) & = \theta_\beta^{(u)}(x,y) & \overline{q}_\beta(x,u,y) = \theta_\beta(x,u,y) \overline{q}(x,u,y) \end{array}$$

2.1.2 A convergence theorem

To clarify the notations, we will use in the remainder of section 2, the superscript $(\tilde{\cdot})$ to distinguish the communication cost functions and the virtual energy function and the critical height associated to the transition probability kernels Q_β from those associated to $\tilde{Q}_\beta$.

From the above observations and corollary 1, choosing β of the form

$$\beta_t = K^{-1} \log t \quad \text{where} \quad K > c \quad (\text{resp. } K > \tilde{c})$$

for t sufficiently large, yields that the canonical process (Ω, P, F_t, X_t) associated to the family of generators $(L_{\beta_t})_{t \geq 0} = (Q_{\beta_t} - I)_{t \geq 0}$ (resp. $(\tilde{Q}_{\beta_t} - I)_{t \geq 0}$) converges in probability to the set of the global minima $V^\star$ (resp. $\tilde{V}^\star$) of the virtual energy V (resp. $\tilde{V}$) associated to Q_β (resp. $\tilde{Q}_\beta$) and described in corollary 1. One open problem is to compare c and $\tilde{c}$. Let us remark that $\tilde{c}$ does not depends on the choice of the parameter $\alpha \in]0,1[$. In view of these observations the bulk of the proof rests on showing that $V^\star$ and $\tilde{V}^\star$ are subsets of $(f^\star)$ where $(f^\star)$ is the set in E defined by

$$(f^\star) = \{x \in E \;:\; \widehat{f}(x) = \min_E f\}$$

The main purpose of this section is to prove a more general result. We will prove that $\tilde{V}^\star$ and $\tilde{V}^\star$ are subsets of $(f^\star) \cap A$, where A is the set in E defined by

$$A = \{x \in E \ : \ x_k \sim x_l \quad \forall 1 \leq k, l \leq N\}$$

By $\mathcal{A}$ we will denote the partition of A induced by the equivalence relation $\sim$

$$\mathcal{A} = \{A_1, \ldots, A_{n(a)}\} \qquad A_i = \{x \in E \ : \ [x] \subset S_i\} \qquad \forall 1 \leq i \leq n(a)$$

As usual we associate to each typical element $x = (x_1, \ldots, x_N) \in A$ the subset

$$A(x) = \{y \in E \ : \ [y] \subset S(x_1)\}$$

Note that A is 0-stable with respect to V and $\tilde{V}$. Moreover, from our constructions, a routine proof yields that $\mathcal{A}$ is a V and $\tilde{V}$-partition of A. In view of propositions 2 and 3 it follows that $V^\star \subset A$, $\tilde{V}^\star \subset A$ and

$$\forall x \in A \qquad V(x) = \min_{g \in G_A(x)} V_{\mathcal{A}}(g) \qquad \tilde{V}(x) = \min_{g \in G_A(x)} \tilde{V}_{\mathcal{A}}(g)$$

Now, from proposition 4, to prove that $V^\star$ and $\tilde{V}^\star$ are subsets of $(f^\star)$ it clearly suffices to find a constant λ such that the subset $(f^\star) \cap A$ is λ-stable with respect to $V_{\mathcal{A}}$ and $\tilde{V}_{\mathcal{A}}$.
As we will see such results holds when the size N of the particle systems is greater that a critical value which depends on the functions a and f. We shall study this critical size now, beginning with two important lemmas.
Before to proceed we need to introduce some additional notations.
By Γ_{x_1,x_2}, $x_1, x_2 \in S$, we denote the paths q in S joining x_1 and x_2, that is

$$\forall 0 \leq l < |q| \qquad k(x_l, x_{l+1}) > 0 \qquad q_0 = x_1 \quad q_{|q|} = x_2$$

We will also note $R(a)$ the smallest integer such that for every $x_1, x_2 \in S$ in two different classes there exists a path joining x_1 and x_2 with length $|q| \leq R(a)$. More precisely

$$R(a) = \max_{1 \leq i,j \leq n(a)} \min_{(x_i,x_j) \in S_i \times S_j} \min_{q \in \Gamma_{x_i,x_j}} |q|$$

It will be also convenient to use the following definitions

$$\begin{aligned} \triangle a &= \min \{a(x_1, x_2) \ : \ a(x_1, x_2) \neq 0\} & \triangle f = \min \{|f(x_1) - f(x_2)| \ : \ f(x_1) \neq f(x_2)\} \\ \delta(a) &= \sup \{a(x_1, x_2) \ : \ x_1, x_2 \in S\} & \delta(f) = \sup \{|f(x_1) - f(x_2)| \ : \ x_1, x_2 \in S\} \end{aligned}$$

Lemma 4 *For every $x, y \in A$ such that $\widehat{f}(x) \geq \widehat{f}(y)$ we have*

$$\tilde{V}_{\mathcal{A}}(x, y) \leq \delta(a)\ R(a) \tag{29}$$

Moreover, for every $x \in A$ there exists a state $y \in (f^\star) \cap A$ such that

$$V_{\mathcal{A}}(x, y) \leq (\delta(a) + \delta(f))\ R(a) \tag{30}$$

Lemma 5 *For every $x, y \in A$ such that $\widehat{f}(x) < \widehat{f}(y)$ we have*

$$V_{\mathcal{A}}(x,y) \geq \min(\Delta a, \Delta f)\, N \quad \textit{and} \quad \tilde{V}_{\mathcal{A}}(x,y) \geq \min(\Delta a, \Delta f)\, N \tag{31}$$

Let us write

$$\begin{aligned} \lambda(a,f) &= (\delta(a) + \delta(f))\, R(a) \qquad \tilde{\lambda}(a,f) = \delta(a)\, R(a) \\ N(a,f) &= \lambda(a,f)/\min(\Delta a, \Delta f) \qquad \tilde{N}(a,f) = \tilde{\lambda}(a,f)/\min(\Delta a, \Delta f) \end{aligned}$$

Combining lemma 4 and lemma 5 one gets easily

$$\begin{aligned} N > N(a,f) &\Longrightarrow (f^\star) \cap A \text{ is } \lambda(a,f)-\text{stable with respect to } V_{\mathcal{A}} \\ N > \tilde{N}(a,f) &\Longrightarrow (f^\star) \cap A \text{ is } \tilde{\lambda}(a,f)-\text{stable with respect to } \tilde{V}_{\mathcal{A}} \end{aligned}$$

These results and proposition 4 combine to yield the following

Theorem 3 *We note (Ω, P, F_t, X_t) the canonical process associated to the family of generators $(L_{\beta_t})_{t\geq 0} = (Q_{\beta_t} - I)_{t\geq 0}$ (resp. $(\tilde{Q}_{\beta_t} - I)_{t\geq 0}$), c (resp. $\tilde{c}$) the critical height associated to the communication cost function V (resp. $\tilde{V}$) and m_t the distribution of X_t. If β_t assumes the parametric form $\beta_t = K^{-1}\log t$, for sufficiently large t, with $K > c$ and if $N > N(a,f)$ (resp. $\tilde{N}(a,f)$), then we have*

$$\lim_{t\to+\infty} Ent_{\pi_{\beta_t}}(m_t) = 0 \qquad \textit{and} \qquad \lim_{t\to+\infty} P(X_t \in (f^\star)\cap A) = 1$$

where π_β is the invariant probability measure of $L_\beta = Q_\beta - I$ (resp. $\tilde{Q}_\beta - I$).

We come to the proof of lemmas 4 and 5.
proof of lemma 4:
Let $x = (x_1, \ldots, x_N)$ and $y = (y_1, \ldots, y_N)$ be two elements of A such that $\widehat{f}(x) \geq \widehat{f}(y)$. First of all let us remark that

$$x_1 \sim y_1 \Longrightarrow \forall 1 \leq k \leq N \quad a(x_k, y_k) = 0 \Longrightarrow \forall 1 \leq k \leq N \quad f(x_k) = f(y_k)$$

In this situation, a routine proof yields

$$V_{\mathcal{A}}(x,y) = 0 \qquad \text{and} \qquad \tilde{V}_{\mathcal{A}}(x,y) = 0$$

If $a(x_1, y_1) > 0$ then the irreductibility condition implies the existence of a path $q \in \Gamma_{x_1,y_1}$ and a pair of integers $0 \leq n_1 < n_2 \leq |q|$ such that

$$\forall 0 \leq k \leq n_1 \quad q_k \in S(x_1) \qquad \forall n_1 < k < n_2 \quad q_k \notin S(x_1) \qquad \forall n_2 \leq k \leq |q| \quad q_k \in S(y_1) \tag{32}$$

Let us prove (29). For this, let $p \in \tilde{C}_{x,y}$ be the path defined by

$$\forall 0 \leq k \leq |q| \qquad p_k = (q_k, x_2, \ldots, x_N) \qquad p_{|q|+1} = (y_1) \qquad p_{|q|+2} = y$$

From the definition of q we have

$$\forall 0 \le k \le n_1 \quad p_k \in A(x) \qquad \forall n_1 < k \le |q| \quad p_k \notin A \qquad \forall k \in \{|q|+1, |q|+2\} \quad p_k \in A(y) \tag{33}$$

Moreover it follows that

$$\begin{aligned} 0 &\le \sum_{k=0}^{n_1-1} \tilde{V}(p_k, p_{k+1}) \le \sum_{k=0}^{n_1-1} V^{(1)}(p_k, p_{k+1}) = 0 \\ 0 &\le \tilde{V}(p_{|q|}, p_{|q|+1}) \le V^{(2)}(p_{|q|}, p_{|q|+1}) = 0 \quad \text{and} \quad \tilde{V}(p_{|q|+1}, p_{|q|+2}) = V^{(1)}(p_{|q|+1}, p_{|q|+2}) = 0 \end{aligned}$$

Now, it appears from the above that

$$\begin{aligned} \tilde{V}(p) &\le \sum_{k=0}^{n_1-1} \tilde{V}(p_k, p_{k+1}) + \sum_{k=n_1}^{|q|-1} \tilde{V}(p_k, p_{k+1}) + \tilde{V}(p_{|q|}, p_{|q|+1}) + \tilde{V}(p_{|q|+1}, p_{|q|+2}) \\ &= \sum_{k=n_1}^{|q|-1} \tilde{V}(p_k, p_{k+1}) \le \delta(a)|q| \end{aligned}$$

So that $\tilde{V}(x,y) \le \delta(a)\ R(a)$ and the proof of (29) is completed.

The proof of (30) is just a little more complicated.
Suppose $x \in A$ and y' is an element of A such that $\widehat{f}(x) \ge \widehat{f}(y') = \min_S f$ and $a(x_1, y_1') > 0$. Let q be the path joining x_1 and y_1' and defined as in (32). Using the above notations, let $(t_m)_m$ be the sequence of integers defined by

$$t_0 = n_1 \qquad t_m = \inf\{k > t_{m-1} \ :\ f(q_k) < f(q_{t_{m-1}})\} \qquad \forall m \ge 1$$

Using the assumption $\widehat{f}(x) \ge \widehat{f}(y') = \min_S f$, examination of q soon yields that there exists an $m_0 \ge 1$ such that $t_{m_0} \le n_2$ and $f(q_{t_{m_0}}) = \min_S f$. Consequently, we have constructed a sequence of states $(q_{t_m})_{0 \le m \le m_0}$ such that

$$\widehat{f}(x) = f(q_{t_0}) > f(q_{t_1}) > \ldots > f(q_{t_m}) > \ldots > f(q_{t_{m_0}}) = \widehat{f}(y) \tag{34}$$

where $y = (q_{t_{m_0}}, \ldots, q_{t_{m_0}}) \in A \cap (f^\star)$. To each state q_{t_m} we associate a state $p_m \in E$, $0 \le m \le m_0$, by setting

$$p_{t_0} = (q_{t_0}, x_2, \ldots, x_N) \qquad p_{t_m} = (q_{t_m}, \ldots, q_{t_m}, x_N) \quad \forall 1 \le m \le m_0$$

First of all we note that $p_{t_m} \notin A$ for all $1 \le m \le m_0$ and

$$\widehat{f}(x) = \widehat{f}(p_{t_0}) > \widehat{f}(p_{t_1}) > \ldots > \widehat{f}(p_{t_m}) > \ldots > \widehat{f}(p_{t_{m_0}}) = \min_S f = \widehat{f}(y)$$

Our next task is to construct a sequence of paths $(p^{(m)})_{0 \le m \le m_0+1}$ such that

$$p^{(0)} \in C_{x, p_{t_0}} \qquad p^{(m)} \in C_{p_{t_{m-1}}, p_{t_m}} \quad \forall 1 \le m \le m_0 \qquad p^{(m_0+1)} \in C_{p_{t_{m_0}}, y}$$

and

- The path $p^{(0)}$ has length $|p^{(0)}| = t_0$ and for every $0 \leq k \leq t_0$ the states $p_k^{(0)}$ belong to $A(x)$.
- For every $1 \leq m \leq m_0$, $p^{(m)}$ is a path joining $p_{t_{m-1}}$ and p_{t_m} in time

$$|p^{(m)}| = t_m - t_{m-1}$$

  and for every $0 \leq k \leq t_m - t_{m-1}$ the states $p_k^{(m)}$ does not belong to A except the first initial state $p_0^{(1)} = p_{t_0} \in A(x)$.
- $p^{(m_0+1)} = (p_{t_{m_0}}, y)$.

It is straightforward to see that any path $p^{(0)}$ satisfying the above conditions has null cost, that is $V(p^{(0)}) = 0$. Then, to obtain the desired upper bound it clearly suffices to have

$$\forall 1 \leq m \leq m_0 \qquad V(p^{(m)}) \leq (t_m - t_{m-1})\,(\delta(a) + \delta(f))$$

We proceed to define $(p^{(m)})_{0 \leq m \leq m_0+1}$ as follows:

1. In view of (32) and (34) it is natural to define the initial path $p^{(0)} \stackrel{\text{def}}{=} (p_0, \ldots, p_{t_0})$ by setting

$$\forall 0 \leq k \leq t_0 = n_1 \qquad p_k = (q_k, x_2, \ldots, x_N) \in A(x)$$

   As it has already been noted a simple calculation shows that

$$V(p^{(0)}) = \sum_{k=0}^{t_0-1} V^{(1)}(p_k, p_{k+1}) + V^{(2)}(p_{k+1}, p_{k+1}) = 0$$

2. Taking into account that t_1 is the first time k such that $f(q_k) < \widehat{f}(x)$ we are lead to define $p^{(1)} \stackrel{\text{def}}{=} (p_{t_0}, p_{t_0+1}, \ldots, p_{t_1})$ by setting

$$\forall t_0 \leq k < t_1 \qquad p_k = (q_k, x_2, \ldots, x_N) \qquad p_{t_1} = (q_{t_1}, \ldots, q_{t_1}, x_N)$$

   Let us write $\overline{p}_{t_1} = (q_{t_1}, x_2, \ldots, x_N)$. In this situation $p_k \notin A$ for all $t_0 < k \leq t_1$ and it is easy to verify that

$$\begin{aligned} V(p^{(1)}) &\leq \sum_{k=t_0}^{t_1-2} V^{(1)}(p_k, p_{k+1}) + V^{(2)}(p_{k+1}, p_{k+1}) + V^{(1)}(p_{t_1-1}, \overline{p}_{t_1}) + V^{(2)}(\overline{p}_{t_1}, p_{t_1}) \\ &\leq (t_1 - t_0)(\delta(a) + \delta(f)) \end{aligned}$$

3. As for 2), we define the paths $p^{(m)} \stackrel{\text{def}}{=} (p_{t_{m-1}}, p_{t_{m-1}+1}, \ldots, p_{t_m})$ for $2 \leq m \leq m_0$ by setting

$$\forall t_{m-1} \leq k < t_m \qquad p_k = (q_k, q_{t_{m-1}}, \ldots, q_{t_{m-1}}, x_N) \qquad p_{t_m} = (q_{t_m}, \ldots, q_{t_m}, x_N)$$

Here again we have $p_k \notin A$ for all $t_{m-1} \leq k \leq t_m$.
Let us introduce a new state $\overline{p}_{t_m} = (q_{t_m}, q_{t_{m-1}}, \ldots, q_{t_{m-1}}, x_N)$. It is then an elementary matter to prove the inequalities

$$\begin{aligned} V(p^{(m)}) &\leq \sum_{k=t_{m-1}}^{t_m-2} \Big(V^{(1)}(p_k, p_{k+1}) + V^{(2)}(p_{k+1}, p_{k+1}) \Big) + V^{(1)}(p_{t_m-1}, \overline{p}_{t_m}) + V^{(2)}(\overline{p}_{t_m}, p_{t_m}) \\ &\leq (t_m - t_{m-1} - 1)\ (\delta(a) + 2\,\delta(f)) + \delta(a) + \delta(f) \leq (t_m - t_{m-1})\ (\delta(a) + \delta(f)) \end{aligned}$$

4. Finally, let us note that

$$0 \leq V(p^{(m_0+1)}) \leq V^{(1)}(p_{t_{m_0}}, p_{t_{m_0}}) + V^{(2)}(p_{t_{m_0}}, y) = 0$$

Consider the path $p = (p^{(0)}, \ldots, p^{(m_0+1)}) \in C_{x,y}$ obtained by joining end to end all these paths. From the above inequalities it follows easily that $V(p) \leq |q|\ (\delta(a) + \delta(f))$. As a clear consequence one gets

$$V_{\mathcal{A}}(x, y) \leq (\delta(a) + \delta(f))\ R(a)$$

This ends the proof of lemma 4.

■

Much more is true. In view of our assumptions on the function a and the constructions given in the proof of lemma 4 we observe easily that

$$\forall x \in A \qquad \forall y \in A \cap (f^\star) \qquad V_{\mathcal{A}}(x, y) \leq (\delta(a) + \delta(f))\ R(a)$$

proof of lemma 5:
Let (x, y) a pair of point of A such that $\widehat{f}(x) < \widehat{f}(y)$. Now, let p belongs to $\tilde{C}_{x,y}$. Remark that since $\widehat{f}(x) < \widehat{f}(y)$ there exists a real number λ such that $\widehat{f}(x) < \lambda < \widehat{f}(y)$. Let us set

$$\forall 0 \leq l \leq |q| \qquad I_l = \Big\{ i \in \{1, \ldots, N\}\ :\ f(p_l^i) > \lambda \Big\} \quad \text{and} \quad n_l = |I_l|$$

It follows easily that $n_0 = 0$ and $n_{|p|} = N$.
Now, let T_k, $0 \leq k \leq N$, be the first time $l \geq 0$ such that $n_l \geq k$. More precisely,

$$T_k = \inf \{ l \in \{0, \ldots, |p|\}\ :\ n_l \geq k \} \qquad \forall 0 \leq k \leq N$$

Clearly it appears from the above that

$$T_0 = 0 \qquad T_N \leq |p| \qquad n_{T_N} = N \qquad n_{T_0} = n_0 = 0 \tag{35}$$

By definition of the communication cost function $V^{(1)}$ we can see that

$$V^{(1)}(p_{T_k-1}, p_{T_k}) \geq \sum_{i \in I_{T_k} - I_{T_{k-1}}} a(p^i_{T_k-1}, p^i_{T_k}) \geq (n_{T_k} - n_{T_k-1})\ \Delta a \qquad \forall 1 \leq k \leq N \tag{36}$$

In a more precise language p_{T_k-1} contains n_{T_k-1} individuals $p^i_{T_k-1}$ such that $f(p^i_{T_k-1}) > \lambda$ and p_{T_k} contains n_{T_k} individuals $p^j_{T_k}$ such that $f(p^j_{T_k}) > \lambda$.
Therefore, if $V^{(1)}(p_{T_k-1}, p_{T_k}) < +\infty$ then the transition $p_{T_k-1} \to p_{T_k}$ necessarily involves at least $(n_{T_k} - n_{T_k-1})$ individual mutations.
Similarly, if $V^{(2)}(p_{T_k-1}, p_{T_k}) < +\infty$ then the system p_{T_k} contains at least $(n_{T_k} - n_{T_k-1})$ new individuals $p^i_{T_k} \in [p_{T_{k-1}}]$ such that $f(p^i_{T_k}) > \lambda$. Thus a similar discussion to that above leads to

$$V^{(2)}(p_{T_k-1}, p_{T_k}) \geq (n_{T_k} - n_{T_k-1})\, \Delta f \tag{37}$$

Finally, by definition of $\tilde{V}$, we have

$$\tilde{V}(q) \geq \tilde{V}(p_{T_1-1}, p_{T_1}) + \ldots \tilde{V}(p_{T_N-1}, p_{T_N})$$

Let us remark that

$$n_{T_k-1} \leq k-1 \leq n_{T_{k-1}} \quad \forall 1 \leq k \leq N$$

Thus, combining (36) and (37) , we arrive at

$$\tilde{V}(q) \geq N \ \min(\Delta a, \Delta f)$$

Taking the minimum of all $p \in \tilde{C}_{x,y}$ and taking into account that that $V \geq \tilde{V}$ we obtain

$$V(x,y) \geq \tilde{V}(x,y) \geq N \ \min(\Delta a, \Delta f)$$

So finally we have

$$\tilde{V}_{\mathcal{A}}(x,y) \geq \tilde{V}(x,y) \geq N \ \min(\Delta a, \Delta f) \quad \text{and} \quad V_{\mathcal{A}}(x,y) \geq V(x,y) \geq N \ \min(\Delta a, \Delta f)$$

This ends the proof of lemma 5.

■

Remark : Lemma 4 and lemma 5 show that the cost of good transitions are bounded whereas the costs of the bad ones increase at least linearly with the size of the system. On the basis of the definition of V and $\tilde{V}$ and in view of the proof of these lemma it is clear that the above result is easier to establish for the cost function $\tilde{V}$. It also turns out that the estimate of the cost of bad transitions with respect to $\tilde{V}$ provide a quick and natural way to estimate their cost with respect to V.
In [2] an inductive proof of this result is presented for the genetic algorithm associated to V and without the equivalence relation considered here. The main contribution here is the extension of the results presented in [2] to any equivalence relation a and to the genetic algorithm associated to the cost function $\tilde{V}$.
On the other hand and in contrast to the inductive proof presented in [2], the approach described here is based on a precise study of the cost of bad or good paths.
The constants $\lambda(a,f), \tilde{\lambda}(a,f)$ represent the difficulty for a population to travel from

an equivalence class to a better ones. In connection with this remark it is interesting to note that $\tilde{\lambda}(a,f)$ does not depend on the fitness function f and

$$\lambda(a,f) > \tilde{\lambda}(a,f)$$

In other words, it is more difficult for the genetic algorithm associated to V to move from one configuration to a better one. The above observations also implies that the critical value $N(a,f)$ is greater than $\tilde{N}(a,f)$.

Examples:
Let see what happens when our second general model (26) specializes to the case where the state is

$$S = \{-1,+1\}^{\mathcal{S}} \qquad \mathcal{S} = [-n,n]^p \qquad p \geq 1$$

and the fitness function $f : S \to \mathbb{R}$ is given by

$$f(x) = \frac{1}{2} \sum_{s \in \mathcal{S}} \sum_{s' \in V_s} x(s)\ x(s') + \frac{1}{2} \sum_{s \in \mathcal{S}} x(s)$$

where

$$\forall s \in \mathcal{S} \qquad V_s = \{s' \in \mathcal{S}\ :\ |s_k - s'_k| \leq 1,\ \ 1 \leq k \leq p\}$$

Let k be the Markovian mutation kernel on S given by

$$k(x_1,x_2) = \frac{1}{|\mathcal{V}(x_1)|} 1_{\mathcal{V}(x_1)}(x_2) \qquad \mathcal{V}(x_1) \stackrel{\text{def}}{=} \{x_2 \in S\ :\ \text{Card}\{s \in \mathcal{S}\ :\ x_1(s) \neq x_2(s)\} \leq 1\}$$

Suppose that the function a is given by

$$a(x_1,x_2) = (1 - 1_{x_1}(x_2)) \qquad \forall (x_1,x_2) \in S^2\ :\ k(x_1,x_2) > 0$$

Then, one can check that

$$R(a) \leq \max_{x,y} \min_{q \in C_{x,y}} |q| = \text{card}(\mathcal{S}) = (2n+1)^p \qquad \text{and} \qquad \delta(a) = \Delta(a) = 1$$

Let N be an integer that $N > (2n+1)^p$. The above theorem shows that N individuals will solve the optimization problem when using the genetic algorithm associated to $\tilde{Q}_\beta$.

2.2 Mean Cost optimization

In this short section we discuss the ways in which the results of section 1 are applied in mean cost optimization problems. Namely, the object will be to find the global minima of a function $V : E \to \mathbb{R}_+$ given by

$$V(x) = E(\mathcal{U}(Z,x)) \qquad \text{or} \qquad V(x) = \min_{g \in G(x)} \sum_{(y,z) \in g} E(\mathcal{V}(Z;y,z))$$

where

- E is a finite set and $G(x)$ is the set of x-graphs over E.
- Z is a random variable taking value in a finite set F. We note μ its distribution.
- $U : F \times E \to \mathbb{R}_+$ and $\mathcal{V} : F \times E \times E \times \to \mathbb{R}_+$

We have seen how to construct a stochastic algorithm converging in probability to global minima of the virtual energy associated to a communication cost function. It is clear from the description above that the appropriate communication cost function are given by

$$V(x,y) = (E(\mathcal{U}(Z;y)) - E(\mathcal{U}(Z;x)))^+ \qquad \text{or} \qquad V(x,y) = E(\mathcal{V}(Z;y,z))$$

Unfortunately the huge size of the set F often preclude the use of such algorithm and the essential problem is to compute a mean cost function at each step. Therefore it is natural to choose, for instance, a Markovian kernel K which ensure that

$$\begin{aligned}
V_{\gamma_t}(x,y) &= \frac{1}{t^A}\int_0^{t^A} \mathcal{V}(Z_s;x,y)\,ds \underset{t\to+\infty}{\longrightarrow} V(x,y) = E(\mathcal{V}(Z;x,y)) \qquad P.a.e \\
\text{or } V_{\gamma_t}(x,y) &= \left(\frac{1}{t^A}\int_0^{t^A} \mathcal{U}(Z_s;y)\,ds - \frac{1}{t^A}\int_0^{t^A} \mathcal{U}(Z_s;x)\,ds\right)^+ \\
&\qquad \underset{t\to+\infty}{\longrightarrow} V(x,y) = (E(\mathcal{U}(Z;y)) - E(\mathcal{U}(Z;x)))^+ \qquad P.a.e
\end{aligned}$$

where

- $V_\gamma \stackrel{\text{def}}{=} \frac{1}{e^\gamma}\int_0^{e^\gamma} \mathcal{V}(Z_s;x,y)\,ds$ or $V_\gamma \stackrel{\text{def}}{=} \left(\frac{1}{e^\gamma}\int_0^{e^\gamma} \mathcal{U}(Z_s;y)\,ds - \frac{1}{e^\gamma}\int_0^{e^\gamma} \mathcal{U}(Z_s;x)\,ds\right)^+$.
- $\gamma_t = A\log t$.
- Z_s is a time-homogeneous Markov process associated to the generator $\mathcal{L} = K - I$
- μ is an invariant measure of $\mathcal{L}$.

Before starting the description of our stochastic algorithm, we just give some details about the above convergences.

Lemma 6 *Let K be a an irreducible transition kernel with unique invariant measure μ. For every $x,y \in E$ and $A > 0$ we have*

$$\lim_{t\to+\infty} \sqrt{\frac{t^A}{\log t}}\,|V_{\gamma_t}(x,y) - V(x,y)| = 0 \qquad P.a.e$$

proof:
In this situation it is well known that for every $x,y \in E$ there exists a bounded function $F(\centerdot;x,y)$ such that

$$\mathcal{V}(\centerdot;x,y) - \mu(\mathcal{V}(\centerdot;x,y)) = \mathcal{L}(\mathrm{F}(\centerdot;x,y)) \qquad \text{and} \qquad \mu(\mathrm{F}(\centerdot;x,y) = 0$$

This equation is the Poisson equation associated to $\mathcal{V}(\centerdot;x,y)$ and $\mathcal{L}$. Thus, one gets the decomposition

$$\frac{1}{t}\int_0^t \left(\mathcal{V}(Z_s;x,y)-\mu(\mathcal{V}(\centerdot;x,y))\right)ds = \frac{1}{t}(F(Z_t;x,y)-F(Z_0;x,y)-M(x,y)_t)$$

where $M(x,y)$ is a martingale with angle bracket $\langle M(x,y)\rangle$ given by

$$\langle M(x,y)\rangle_t = \sum_{z\in F}\int_0^t \left(F(z;x,y)-F(Z_s;x,y)\right)^2 K(Z_s,z)\,ds$$

Since the function F is bounded we have $\langle M(x,y)\rangle_t \le c\ t$ for some non negative constant. Finally by the standard iterated-Log law and since jumps are bounded it follows that

$$\frac{1}{t}M(x,y)_t \le \frac{\sqrt{2\,c\,t\log\log(c\,t)}}{t}$$

and thus $\lim_{t\to+\infty}\sqrt{\frac{t}{\log t}}\left|\frac{1}{t}M(x,y)_t\right| = 0$. This ends the proof.

■

Remark : The Poisson equation is a standard tool in the study of Markov processes. For instance it was also used by Younes [13] to study the convergence of a stochastic gradient algorithm to a maximum likelihood estimator. The context of Younes is more complex that those considered here, and the speed of convergence cannot be obtained by a mere application of the iterated logarithm law as before. But Younes also noticed that if the convergence is fast enough (in a negative power in time), then one can couple the estimation procedure to a simulated annealing algorithm (with the classical reversibility conditions) to get the global minima of a function depending on the parameter to be estimated. To do that, Younes uses the Dobrushin coefficients, but the entropy approach enable one to get more precise results on the admissible logarithmic schedules of temperature (the constant c given below).

Let us fix some terminology:

- Let $(\Omega^{(Z)}, P^{(Z)}, F_t^{(Z)}, Z_t)$ be the canonical process associated to the generator $\mathcal{L}$.
- For a given probability measure m on E, $\beta,\gamma\in C^1(\mathbb{R}_+,\mathbb{R}_+)$ and given the Markov process Z we note $(\Omega_{(Z)}, P_{(Z)}, F_{(Z),t}, X_t)$ the canonical process associated to the family of generators $(L_{\gamma_t,\beta_t})_{t\geq 0} = (Q_{\gamma_t,\beta_t}-I)_{t\geq 0}$ whose initial condition is $m_0 = m$ and we note m_t the distribution of X_t where

$$Q_{\gamma,\beta}(x,y) = q(x,y)\,e^{-\beta\,V_\gamma(x,y)} \qquad \text{with} \quad q \ \text{irreducible}$$

- To capture all randomness we note $\Omega = \Omega^{(Z)}\times\Omega_{(Z)}$, $F_t = F_t^{(Z)}\times F_{(Z),t}$ and we define P as follows

$$\forall A\in F_{(Z),t}\in \qquad \forall B\in F_t^{(Z)} \qquad P(A\times B) = \int_B P_{(Z)}(A)\,dP^{(Z)}$$

The above lemma and corollary 4 lead us to the following theorem

Proposition 5 *Let us set* $c = \limsup_{\gamma\to+\infty} c(\gamma) < +\infty$ *P.a.e., where* $c(\gamma)$ *is the critical height associated to the communication cost* V_γ*.*
When the inverse freezing schedule has parametric form $\beta_t = K^{-1}\log t$*, for* t *sufficiently large and* $K > c$*, we have*

$$\lim_{t\to+\infty} Ent_{\pi_{\beta_t}}(m_t) = 0 \quad P.a.e. \qquad and \qquad \lim_{t\to+\infty} P(X_t \in V^\star) = 1 \tag{38}$$

where π_β *is the unique invariant probability of* $L_\beta = Q_\beta - I$ *with*

$$Q_\beta(x,y) = q(x,y)\, e^{-\beta V(x,y)}$$

In many practical situations we also want a quantitative measure of the convergence (38). Unfortunately our method of proof is not suitable for estimating such quantitative behavior. In our settings a natural alternative approach is to look at the convergence of the mean value of the process $t \to \mathrm{Ent}_{\pi_{\beta_t}}(m_t)$ with respect to the random media (given by Z). In view of the inequality (20) we immediately observe that the speed of convergence of the mean value is related to the speed of convergence of the mean values of $Ent_{\pi_{\gamma_t,\beta_t}}(m_t)$ and $|V_{\gamma_t}(x,y) - V(x,y)|$. The first term, linked to the critical height $c(\gamma_t)$ and to the derivative of γ_t, depends in a complicated way on the constant A, but we know that it is a non decreasing function of the parameter A. On the other hand, the second term is a non increasing function of the parameter A. If we know how these quantities are linked to A a good adjustment of this parameter is then related to a classical minimization problem. We will examine this quantitative behavior in a forthcoming paper.

References

[1] R. Bott, J.P. Mayberry, *Matrices and trees*, Economics Activity Analysis, Wiley, New York, 1954.

[2] R. Cerf, *Une théorie asymptotique des algorithmes génétiques*, Thése de Doctorat, Université Montpellier II, Mars 1994.

[3] P. del Moral, *Nonlinear Filtering: Interacting Particle Resolution*, Markov Processes and Related Fields, **Vol.2**, No 4, pp 555-579, 1996.

[4] J.D. Deuschel, C. Mazza, *L^2-Convergence of Time Nonhomogeneous Markov Processes: I. Spectral Estimates*, The Annals of Applied Probability, Vol.4, No 4, 1012-1056, 1994.

[5] P. Diaconis, L. Saloff-Costes, *Logarithmic Sobolev Inequalities for Finite Markov Chains*, to appear in Ann. Prob. Comp., 1996.

[6] M.I. Freidlin, A.D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer Verlag, 1984.

[7] A. Frigerio, G. Grillo, *Simulated annealing with time-dependent energy function*, Mathematische Zeitschrift, Vol 213, pp 97-116, 1993.

[8] R. Holley, D. Stroock, *Annealing via Sobolev inequalities* , Communications in Mathematical Physics, 115; pp 553-569, 1988.

[9] M. Löwe, *Simulated annealing with time-independent energy function via Sobolev inequalities*, Stochastic Processes and their Applications 63 (1996) 221-233.

[10] L. Miclo, *Recuit simulé sans potentiel sur un ensemble fini*, in Séminaire de Probabilités, Lecture Notes in Mathematics 1526, 47-60, Springer Verlag, 1992.

[11] D. Stroock, *Logarithmic Sobolev Inequalities for Gibbs States*, Dirichlet Forms, Eds G. Dell'Antonio and U. Mosco Springer-Verlag, Lecture Notes in Mathematics 1563 pp 194-228 (1993).

[12] A. Trouvé, *Parallélisation massive du recuit simulé*, Thèse de Doctorat, Université Paris XI, Janvier 1993.

[13] L. Younes, *Estimation and annealing for Gibbsian fields*, Ann. Inst. Henri Poincaré, Vol. 24, No 2, 1988, p. 269-294.