ASYMPTOTIC CONVERGENCE OF GENETIC ALGORITHMS

Raphaël CERF

CNRS, Université d'Orsay, Paris

January 1997

Abstract. We study a markovian evolutionary process which encompasses the classical simple genetic algorithm. This process is obtained by perturbing randomly a very simple selection scheme. Using the Freidlin–Wentzell theory, we carry out a precise study of the asymptotic dynamics of the process as the perturbations disappear. We show how a delicate interaction between the perturbations and the selection pressure may force the convergence toward the global maxima of the fitness function. We put forward the existence of a critical population size, above which this kind of convergence can be achieved. We compute upper bounds of this critical population size for several examples. We derive several conditions to ensure convergence in the homogeneous case: these provide the first mathematically well-founded convergence results for genetic algorithms.

1991 *Mathematics Subject Classification.* 60F10, 60J10, 92D15.
Key words and phrases. Freidlin–Wentzell theory, genetic algorithms, stochastic optimization.
This work was done during my last year as a student at the Ecole Normale Supérieure.
Professor Jean–Arcady Meyer first spoke to me of genetic algorithms.
I am deeply grateful to Professor Gérard Ben Arous for introducing me to the Freidlin–Wentzell theory.
I thank Professor Alain Berlinet for his useful advice and for having carefully read this paper.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

I. Introduction

Stochastic optimization has been a field of intense research during the last decade: computers become more and more powerful and there is an increasing need for robust and efficient search techniques. In the jungle of optimization procedures, there exists a particular class of algorithms called genetic algorithms. Introduced by J.H. Holland in the seventies [6], these algorithms are based on the genetic mechanisms which guide natural evolution: mutation, crossover and selection. Today, they are widely used to handle a large spectrum of optimization problems ranging from the classical traveling salesman problem to the design of network architecture [5].

However the behavior of genetic algorithms relies heavily on several control parameters (for instance the population size, the probabilities of mutation and crossover). Until now, no firm result concerning the choice of these parameters and the dynamics of the genetic algorithm has been available: practical know–how and experimental simulations were the only guides for handling concrete problems.

In this paper, we study a slightly more general evolutionary scheme. We consider first a very simple selection algorithm which is extremely rapidly trapped in bad points of the search space and we perturb randomly the whole mechanism. We focus on the asymptotic dynamics of the process as the random perturbations vanish. We show how a delicate interaction between mutations, crossovers and selection pressure ensures the convergence of the algorithm toward the set of the global maxima of the fitness function. We put forward the existence of a critical population size, above which this kind of convergence can be achieved. Our results seem to be the first well–founded convergence theorems of this kind concerning genetic algorithms.

To fulfill this program, we use the powerful tools developed by Freidlin and Wentzell in a much more general framework for the study of random perturbations of dynamical systems [4].

Classical genetic algorithms are not usually described as the random perturbation of a simple process. For a fixed level of intensity of the random perturbations, our model evolves exactly as a classical genetic algorithm. We are only able to analyze rigorously the asymptotic regime when the perturbations vanish. On one hand, the asymptotic picture sheds some light on the true dynamics of classical genetic algorithms. On the other hand, this approach suggests new ways of implementing genetic algorithms, using the theory of generalized annealing processes [3,12].

To avoid lengthy and technical computations we focus here mainly on the study of the invariant measure of the process and we just quote without proof a result of convergence for the inhomogeneous case. In another work, we apply to our model the very technical tools developed by Catoni and Trouvé to control the speed of convergence of generalized annealing processes to equilibrium (see [2] and the references therein). A major difficulty with such dynamics is the lack of reversibility [3].

This paper has the following structure.
We first describe our model. Then, we study the convergence of the algorithm in the homogeneous case and derive sufficient conditions to ensure that the population settles in a global maximum of the fitness function. We show that these conditions are fulfilled when the population size is large enough. We apply our results to the classical simple genetic algorithm. In the inhomogeneous case, we give several conditions on the rate of decrease of the perturbations to ensure the convergence of the population to the set of global maxima in finite time. We finally examine the value of the critical population size for a specific problem coming from statistical mechanics, namely the search of the ground state of the Ising model in a finite box.

II. General notations and conventions

The cardinality of a set X will be noted indifferently $|X|$ or $\mathrm{card}\, X$ and its characteristic function 1_X. We adopt usual conventions concerning empty sets:

$$\min \emptyset = +\infty, \quad \max \emptyset = -\infty, \quad \prod_{\emptyset} = 1, \quad \sum_{\emptyset} = 0.$$

If s is a real number, $\lfloor s \rfloor$ denotes the unique integer such that $\lfloor s \rfloor \leq s < \lfloor s \rfloor + 1$.
We say that a real number s is positive if $s > 0$ and that it is non–negative if $s \geq 0$.
We consider a finite space of states E and a real-valued positive non-constant function f (which will be called the fitness function) defined on E. The letters i, j will denote elements of E. The set of global maxima of f is

$$f^* = \{\, i \in E : f(i) = \max_{j \in E} f(j) \,\}.$$

By $f(f^*)$ we mean the maximum of f over E i.e. $\max_{j \in E} f(j)$.
The Kronecker symbol $\delta(i, j)$ will be used to denote the identity matrix indexed by E:

$$\forall i, j \in E \qquad \delta(i, j) = 0 \quad \text{if} \quad i \neq j, \quad \delta(i, j) = 1 \quad \text{if} \quad i = j.$$

Let m be a positive integer. The state space of most of Markov chains under consideration will be E^m, the set of m–uples of elements of E: these m–uples are called populations and their components individuals. They will be denoted by the letters x, y, z.
If $x = (x_1, \cdots, x_m)$ belongs to E^m, $[x]$ is the set of individuals contained in x i.e.

$$[x] = \{\, x_k : 1 \leq k \leq m \,\}$$

and for i in E, $x(i)$ is the number of occurrences of i in the population x:

$$x(i) = \mathrm{card}\, \{\, k : 1 \leq k \leq m,\ x_k = i \,\}.$$

The group of permutations $\mathfrak{S}_m$ of the set $\{1, \cdots, m\}$ operates on E^m in the following way:

$$\forall \sigma \in \mathfrak{S}_m \qquad \sigma \cdot x = \sigma \cdot (x_1, \cdots, x_m) = (x_{\sigma(1)}, \cdots, x_{\sigma(m)}).$$

The set $E^m/\mathfrak{S}_m$ is the set of equivalence classes associated with this group operation.
We associate with f two functions defined on E^m.
The first one is still noted f and takes its values in $(\mathbb{R}_+^*)^m$:

$$f(x) = f(x_1, \cdots, x_m) = (f(x_1), \cdots, f(x_m)).$$

The second one is noted $\widehat{f}$ and is real-valued:

$$\widehat{f}(x) = \widehat{f}(x_1, \cdots, x_m) = \max_{1 \le k \le m} f(x_k).$$

For x in E^m, $\widehat{x}$ denotes the set of those elements of $[x]$ which realize the value $\widehat{f}(x)$:

$$\widehat{x} = \{\, x_k : 1 \le k \le m,\ f(x_k) = \widehat{f}(x) \,\}.$$

For a point i of E, (i) is the m–uple whose m components are equal to i and A is the set of all such m–uples (which are called the uniform populations). By S we denote the set of equi–fitness populations, that is, populations whose individuals have the same fitness:

$$S = \{\, x \in E^m : f(x_1) = \cdots = f(x_m) \,\}.$$

We sometimes identify f^* with $\{\, (i) : i \in f^* \}$ so that f^* may be seen as a subset of A.

III. Description of the model

1. The unperturbed Markov chain (X_n^∞)

In the absence of perturbations, the process under study is a Markov chain $(X_n^\infty)_{n \ge 0}$ with state space E^m. The superscript ∞ reflects the fact that this process describes the limit behavior of our model, when all perturbations have disappeared. The transition probabilities of this chain are

$$P\left(X_{n+1}^\infty = z / X_n^\infty = y\right) = \frac{1}{(\operatorname{card} \widehat{y})^m} \prod_{k=1}^m 1_{\widehat{y}}(z_k) y(z_k) = \frac{1}{(\operatorname{card} \widehat{y})^m} \prod_{i \in [z]} 1_{\widehat{y}}(i) y(i)^{z(i)}$$

that is, the individuals of the population X_{n+1}^∞ are chosen randomly (under the uniform distribution) and independently among the elements of $\widehat{X}_n^\infty$ which are the best individuals of X_n^∞ according to the fitness function f.
Suppose the chain starts up with the initial population $X_0 = x_0$. Then

$$\forall n \ge 1 \qquad [X_n^\infty] \subset \widehat{x}_0$$

and with probability one, after a finite number of steps N, the chain is absorbed in a state (i) where i belongs to $\widehat{x}_0$. In particular, if $\widehat{x}_0$ is reduced to one point i, the chain is instantaneously absorbed in (i).

2. The perturbed Markov chain (X_n^l)

The previous Markov chain (X_n^∞) is randomly perturbed by three distinct mechanisms. The first two act directly upon the population and mimic the phenomena of mutation and crossover. The third one consists in loosening the selection of the individuals.
The intensity of the perturbations is governed by an integer parameter l. As l grows toward infinity, the perturbations progressively disappear.
The perturbed Markov chain (X_n^l) is obtained through overlapping of several other chains $(U_n^l), (V_n^l)$ which represent the successive populations obtained by applying the perturbing operators. More precisely, we decompose the transition from X_n^l to X_{n+1}^l in three stages:

$$X_n^l \quad \xrightarrow{\text{mutation}} \quad U_n^l \quad \xrightarrow{\text{crossover}} \quad V_n^l \quad \xrightarrow{\text{selection}} \quad X_{n+1}^l.$$

We now proceed to a more detailed description of these three operations.

$X_n^l \longrightarrow U_n^l$**: mutation.** The mutation operator is modeled by random independent perturbations of the individuals of the population X_n^l. Such a perturbation is described by a markovian kernel p_l on the space E, that is a function defined on $E \times E$ with values in $[0,1]$ verifying

$$\forall i \in E \qquad \sum_{j \in E} p_l(i,j) = 1.$$

The transition probabilities from X_n^l to U_n^l are then given by

$$P\left(U_n^l = u / X_n^l = x\right) = p_l(x_1, u_1) \cdots p_l(x_m, u_m). \tag{1}$$

This perturbation is small whenever the matrix $(p_l(i,j))_{(i,j)\in E\times E}$ is close to the identity matrix. To ensure the vanishing of mutations when l grows toward infinity, we will impose

$$\forall i,j \in E \qquad \lim_{l\to\infty} p_l(i,j) = \delta(i,j). \tag{2}$$

$U_n^l \longrightarrow V_n^l$**: crossover.** The crossover operator is modeled by random independent perturbations of the couples formed by consecutive individuals of the population (X_n^l). As in the mutation case, such a perturbation is described by a markovian kernel q_l on the space $E \times E$, that is a function defined on $(E\times E)\times(E\times E)$ with values in $[0,1]$ verifying:

$$\forall (i_1, j_1) \in E \times E \qquad \sum_{(i_2,j_2)\in E\times E} q_l\left((i_1,j_1),(i_2,j_2)\right) = 1.$$

The transition probabilities from U_n^l to V_n^l are

$$P\left(V_n^l = v / U_n^l = u\right) = \delta_m(u_m, v_m) \prod_{1\le k\le m/2} q_l\left((u_{2k-1}, u_{2k}), (v_{2k-1}, v_{2k})\right) \tag{3}$$

where $\delta_m(i,j) = \delta(i,j)$ if m is odd (the last individual of the population remains unchanged after crossover) and $\delta_m(i,j) = 1$ if m is even.
To ensure the vanishing of crossovers when l grows toward infinity, we will impose

$$(4) \quad \forall (i_1,j_1) \in E\times E \quad \forall (i_2,j_2) \in E\times E \qquad \lim_{l\to\infty} q_l\left((i_1,j_1),(i_2,j_2)\right) = \delta(i_1,i_2)\delta(j_1,j_2).$$

$V_n^l \longrightarrow X_{n+1}^l$**: selection.** In order to build our selection operator, we will use a selection function.

Definition. A selection function of order m is a function F defined on $\{1,\cdots,m\}\times(\mathbb{R}_+^*)^m$ with values in $[0,1]$ satisfying for each $(f_1,\cdots,f_m)$ in $(\mathbb{R}_+^*)^m$:

$$a) \quad \sum_{k=1}^m F(k,f_1,\cdots,f_m) = 1,$$

$$b) \quad \forall \sigma \in \mathfrak{S}_m \quad \forall k \in \{1,\cdots,m\} \qquad F(\sigma(k),f_{\sigma(1)},\cdots,f_{\sigma(m)}) = F(k,f_1,\cdots,f_m),$$

$$c) \quad f_1 \geq f_2 \geq \cdots \geq f_m \Rightarrow F(1,f_1,\cdots,f_m) \geq F(2,f_1,\cdots,f_m) \geq \cdots \geq F(m,f_1,\cdots,f_m).$$

The value $F(k,f_1,\cdots,f_m)$ is the probability of choosing f_k among the values $f_1,\cdots,f_m$: this probability does not depend on the indices and increases with the relative value of f_k among $f_1,\cdots,f_m$. Now let F_l be a selection function. The m individuals $X_{n+1}^{l,1},\cdots,X_{n+1}^{l,m}$ who make up the population X_{n+1}^l are chosen randomly and independently in the population V_n^l from the law defined by F_l:

$$\forall\, r \in \{1,\cdots,m\} \quad \forall i \in E \qquad P\left(X_{n+1}^{l,r} = i\right) = \sum_{h:V_n^{l,h}=i} F_l\left(h, f(V_n^l)\right).$$

(recall that $f(V_n^l)$ is the m–uple $\left(f(V_n^{l,1}),\cdots,f(V_n^{l,m})\right)$) so that the transition probabilities from V_n^l to X_{n+1}^l are

$$(5) \qquad P\left(X_{n+1}^l = x / V_n^l = v\right) = \prod_{i\in[x]}\left(\sum_{k:v_k=i} F_l(k,f(v))\right)^{x(i)} = \prod_{r=1}^m \sum_{k:v_k=x_r} F_l(k,f(v)).$$

The selection pressure is maximal if the individuals of X_{n+1}^l are chosen randomly and uniformly from the fittest individuals of V_n^l. The unique selection function F_∞ which implements such a selection scheme is defined by

$$F_\infty(k,f(x)) = \frac{1_{\widehat{x}}(x_k)}{\operatorname{card}\widehat{x}}$$

i.e. we have then the uniform distribution over $\widehat{x}$.
To ensure the disappearance of the selection of individuals below peak fitness, we will impose the convergence of F_l toward F_∞ on the set $f(E)^m$:

$$(6) \qquad \forall x \in E^m \quad \forall k \in \{1, \cdots, m\} \qquad \lim_{l\to\infty} F_l(k, f(x)) = F_\infty(k, f(x)).$$

The basic example of such a sequence of selection functions is given by

$$(7) \qquad F_l(k, f_1, \cdots, f_m) = \frac{\exp(f_k \ln w_l)}{\sum_{r=1}^m \exp(f_r \ln w_l)}$$

where $(w_l)_{l\geq 0}$ is an increasing sequence of positive real numbers which tends to infinity. We will only be concerned with such sequences of selection functions in the sequel.
We can now evaluate the

Transition probabilities of the chain (X_n^l).
The transition probability $P\left(X_{n+1}^l = z/X_n^l = y\right)$ is given by the sum

$$\sum_{(u,v)\in(E^m)^2} P\left(X_{n+1}^l = z/V_n^l = v\right) P\left(V_n^l = v/U_n^l = u\right) P\left(U_n^l = u/X_n^l = y\right).$$

Conditions (2), (4) and (6) imply

$$\forall (y,z) \in E^m \times E^m \qquad \lim_{l\to\infty} P\left(X_{n+1}^l = z/X_n^l = y\right) = P\left(X_{n+1}^\infty = z/X_n^\infty = y\right)$$

so that the transition probabilities of (X_n^l) converge toward those of (X_n^∞) as l tends to infinity. Thus the Markov chain (X_n^l) appears as a perturbation of the Markov chain (X_n^∞). In order to deal with the asymptotic dynamics of the chain (X_n^l) we need more information about the way its transition probabilities converge to those of (X_n^∞). We will make the following assumptions on the sequences of kernels p_l and q_l.

Hypothesis H_p. There exist a sequence $(u_l)_{l\geq 0}$ of positive real numbers and a function α defined on $E \times E$ with values in $\mathbb{R}^+$ satisfying

a) $\lim_{l\to\infty} u_l = +\infty$

b) α is an irreducible kernel, that is

$$\forall i, j \in E \quad \exists i_0, i_1, \cdots, i_r \quad i_0 = i, \quad i_r = j, \qquad \prod_{0\leq k\leq r-1} \alpha(i_k, i_{k+1}) > 0$$

c) p_l admits the development

$$\forall i, j \in E \quad \forall s > 0 \qquad p_l(i,j) = \begin{cases} \alpha(i,j)(u_l)^{-1} + o\left((u_l)^{-s}\right) & \text{if } i \neq j \\ 1 - \alpha(i,j)(u_l)^{-1} + o\left((u_l)^{-s}\right) & \text{if } i = j \end{cases}$$

Hypothesis H_q. There exist a sequence $(v_l)_{l\geq 0}$ of positive real numbers and a function β defined on $(E\times E)\times(E\times E)$ with values in $\mathbb{R}^+$ satisfying

a) $\lim\limits_{l\to\infty} v_l = +\infty$

b) q_l admits the development

$$\forall (i_1,j_1)\in E\times E \quad \forall (i_2,j_2)\in E\times E \quad \forall s>0$$

$$q_l\left((i_1,j_1),(i_2,j_2)\right)=\begin{cases}\beta\left((i_1,j_1),(i_2,j_2)\right)(v_l)^{-1}+o\left((v_l)^{-s}\right) & \text{if } (i_1,j_1)\neq(i_2,j_2)\\ 1-\beta\left((i_1,j_1),(i_2,j_2)\right)(v_l)^{-1}+o\left((v_l)^{-s}\right) & \text{if } (i_1,j_1)=(i_2,j_2)\end{cases}$$

In fact, the asymptotic dynamics of the chain will be governed by the kernels α and β: the irreducibility condition on the kernel α (which implies the irreducibility of the kernel p_l for l large enough) is essential in order to allow the population to visit all the space E even when the perturbations are small.

IV. Convergence of the homogeneous algorithm

This section is devoted to the study of the behavior of the chain (X_n^l) when n first goes to infinity and then l does. We will be interested in the quantities

$$\lim_{l\to\infty}\quad\lim_{n\to\infty}\quad P\left(X_n^l=z/X_0^l=y\right).$$

The kernel α is irreducible so that for l large enough, the chain (X_n^l) is irreducible. Furthermore, for l sufficiently large, the diagonal coefficients $P\left(X_{n+1}^l=x/X_n^l=x\right)$ are positive so that the chain (X_n^l) is aperiodic. In addition, the state space of (X_n^l) is finite. The chain (X_n^l) admits thus a unique invariant probability measure μ^l[11, Theorems 4.1 and 4.2]. This measure charges all points of the space E^m and we have

$$\forall\, y,z\in E^m \qquad \lim_{n\to\infty} P\left(X_n^l=z/X_0^l=y\right)=\mu^l(z).$$

We are thus interested in the convergence of the stationary measures $(\mu^l)_{l\geq 0}$.
We will note P_x for the probability measure associated with the chain (X_n^l) starting at $X_0^l=x$ and E_x will be the expectation with respect to P_x.

1. The time of entrance in the set S of X_n^l

Let $\tau^l=\min\{\,n>0: X_n^l\in S\,\}$ be the first entrance time of the chain (X_n^l) into S and

$$s_l=\min_{x\in E^m}\sum_{k:x_k\in\widehat{x}} F_l(k,f(x)).$$

For each n in $\mathbb{N}$ and x in E^m we have

$$P\left(X^l_{n+1}\in S/X^l_n=x\right)\geq \min_{v\in E^m} P\left([X^l_{n+1}]\subset \widehat{v}/V^l_n=v\right)\geq (s_l)^m.$$

From this we deduce that

$$\forall q\in\mathbb{N}\quad \forall x\in E^m\qquad P_x\left(\tau^l>q\right)\leq \left(1-(s_l)^m\right)^q. \tag{8}$$

Summing up this inequality from $q=0$ to ∞ we get

$$\forall x\in E^m\qquad E_x\left(\tau^l\right)\leq (s_l)^{-m}.$$

We proceed now to a more careful study of the sequence $(s_l)_{l\geq 0}$. With our choice for the sequence F_l, it is clear that s_l is strictly positive for every l. In addition, for a fixed population x, $F_l(k,f(x))$ increases with l (recall that w_l is an increasing sequence) whenever x_k belongs to $\widehat{x}$. The sequence s_l, being the minimum over E^m of sums of such increasing sequences, is also increasing. Define

$$\delta=\min\{\,|f(i)-f(j)|:i,j\in E,\,f(i)\neq f(j)\,\}. \tag{9}$$

For x in E^m, we have

$$1-\sum_{k:x_k\in\widehat{x}}F_l(k,f(x))=\frac{\sum_{r:x_r\notin\widehat{x}}\exp((f(x_r)-\widehat{f}(x))\ln w_l)}{\operatorname{card}\widehat{x}+\sum_{r:x_r\notin\widehat{x}}\exp((f(x_r)-\widehat{f}(x))\ln w_l)}\leq (m-1)(w_l)^{-\delta}$$

whence finally

$$1-s_l\leq (m-1)(w_l)^{-\delta}. \tag{10}$$

2. Concentration of μ^l on S

We denote by (Z^l_n) the Markov chain induced by (X^l_n) on S and by ν^l its invariant probability measure μ^l_S. More precisely, (Z^l_n) is the Markov chain with state space S and with transition probabilities

$$\forall x,y\in S\qquad P(Z^l_{n+1}=y/Z^l_n=x)\;=\;P_x(X_{\tau^l}=y)$$

(we recall that τ^l is the hitting time of S). We first use the standard formula for representing the invariant measure of (X^l_n) with the help of the induced Markov chain (Z^l_n) [7, Proposition 5.3]. We obtain

$$\forall x\in E^m\qquad \mu^l(x)=\mu^l(S)\sum_{y\in S}\nu^l(y)E_y\left[\sum_{k=0}^{\tau^l-1}1_{\{X^l_k=x\}}\right]. \tag{11}$$

The study of the asymptotic behavior of μ^l will entirely rely on this representation formula. If x does not belong to S then for y in S

$$E_y\left[\sum_{k=0}^{\tau^l-1} 1_{\{X_k^l=x\}}\right] = E_y\left[\sum_{k=1}^{\tau^l-1} 1_{\{X_k^l=x\}}\right] \leq E_y[\tau^l]-1 \leq (s_l)^{-m}-1.$$

Reporting this inequality in the representation formula (11) yields

$$\mu^l(x) \leq \mu^l(S)\left((s_l)^{-m}-1\right).$$

The convergence of the selection functions $(F_l)_{l\geq 0}$ toward F_∞ implies that s_l converges to one whence

$$\lim_{l\to\infty} \mu^l(E^m\setminus S) = 0.$$

If x belongs to S then for y in S

$$E_y\left[\sum_{k=0}^{\tau^l-1} 1_{\{X_k^l=x\}}\right] = \delta(x,y).$$

Reporting this equality in the representation formula (11) yields $\mu^l(x) = \mu^l(S)\nu^l(x)$. We have just shown that $\mu^l(S)$ converges to one as l goes to infinity. Therefore it remains to study the convergence of the sequence of probability measures ν^l.

3. Asymptotics of the transition matrix of Z_n^l

To study the convergence of the measure ν^l we will use the powerful machinery developed by Freidlin and Wentzell. The first step consists in evaluating the asymptotics of the transition probabilities of the induced Markov chain (Z_n^l). Thus our next task will be to obtain estimates of the quantities $P\left(Z_{n+1}^l = y/Z_n^l = x\right)$ for x and y in S.
We will successively study the asymptotics of the transition probabilities

$$P\left(U_n^l = u/X_n^l = x\right), \quad P\left(V_n^l = v/U_n^l = u\right), \quad P\left(X_{n+1}^l = x/V_n^l = v\right).$$

Asymptotics of $P\left(U_n^l = u/X_n^l = x\right)$. Define for x and u in E^m

$$d(x,u) = \mathrm{card}\,\{\, k : 1\leq k\leq m, x_k \neq u_k \,\}$$

and extend α to $E^m\times E^m$ by putting

$$\alpha(x,u) = \prod_{k:x_k\neq u_k} \alpha(x_k,u_k).$$

Define for x in E^m the set $\mathcal{U}(x)$ as the set of populations reachable from x in one transition step through the kernel α that is $\mathcal{U}(x) = \{\, u\in E^m : \alpha(x,u)>0 \,\}$. Using the development of p_l given by hypothesis H_p in (1), we see that, as $l\to\infty$,

$$\begin{array}{lll} \text{If} \quad u\notin\mathcal{U}(x) & P\left(U_n^l = u/X_n^l = x\right) = o\left((u_l)^{-s}\right) & \forall s>0 \\ \text{If} \quad u\in\mathcal{U}(x) & P\left(U_n^l = u/X_n^l = x\right) \sim \alpha(x,u)(u_l)^{-d(x,u)}. & \end{array}$$

Asymptotics of $P\left(V_n^l = v / U_n^l = u\right)$. Define for u and v in E^m

$$\bar{d}(u,v) = \text{card}\,\{\, k : 1 \le k \le m/2, \quad (u_{2k-1}, u_{2k}) \neq (v_{2k-1}, v_{2k}) \,\}$$

and extend β to $E^m \times E^m$ by putting

$$\beta(u,v) = \delta_m(u_m, v_m) \prod_{k:(u_{2k-1},u_{2k}) \neq (v_{2k-1},v_{2k})} \beta((u_{2k-1}, u_{2k}), (v_{2k-1}, v_{2k})).$$

For u in E^m, let $\mathcal{V}(u)$ be the set of populations reachable from u in one transition step through the kernel β that is $\mathcal{V}(u) = \{\, v \in E^m : \beta(u,v) > 0 \,\}$. Using the development of q_l given by hypothesis H_q in (3), we obtain, as $l \to \infty$,

$$\begin{array}{lll} \text{If} \quad v \notin \mathcal{V}(u) & P\left(V_n^l = v / U_n^l = u\right) = o\left((v_l)^{-s}\right) & \forall s > 0 \\ \text{If} \quad v \in \mathcal{V}(u) & P\left(V_n^l = v / U_n^l = u\right) \sim \beta(u,v) (v_l)^{-\bar{d}(u,v)}. & \end{array}$$

Asymptotics of $P\left(X_{n+1}^l = z / V_n^l = y\right)$. Formulas (5) and (7) yield

$$P\left(X_{n+1}^l = z / V_n^l = y\right) = \prod_{i \in [z]} \left(\frac{y(i) \exp(f(i) \ln w_l)}{\sum_{k=1}^m \exp(f(y_k) \ln w_l)} \right)^{z(i)}$$

whence

$$P\left(X_{n+1}^l = z / V_n^l = y\right) \underset{l \to \infty}{\sim} \frac{1}{(\text{card}\,\widehat{y})^m} \prod_{i \in [z]} y(i)^{z(i)} \exp\left(\left(\sum_{i \in [z]} z(i) f(i) - m \widehat{f}(y) \right) \ln w_l \right)$$

The important quantity in this formula is the exponent of w_l which gives the rate of decreasing of the transition probabilities: it may be written

$$m \left(\frac{1}{m} \sum_{k=1}^m f(z_k) - \widehat{f}(y) \right)$$

i.e. m multiplied by the difference between the mean of f over the population z and the maximum of f over the population y.

Choice of the sequences u_l, v_l and w_l. We are interested in the dynamics of the chain (X_n^l) when the perturbations become smaller and smaller that is when l grows to infinity. In this situation, with overwhelming probability, the chain (X_n^l) behaves as the chain (X_n^∞) would do. However the occurrence of rare events (which have very low probability) allows

the chain (X_n^l) to escape from the absorbing states of the chain (X_n^∞) and to visit all the space. These rare events are caused by the three distinct mechanisms of perturbations described earlier. To ensure that these three mechanisms actually play their own role in the asymptotic dynamics of the chain, we must give the same order of intensity to the three kinds of perturbations. If one perturbation is negligible compared to another one, its asymptotic influence will be null. We will therefore focus on the situation where the three sequences u_l, v_l and w_l which express the intensity of each perturbation have asymptotics logarithmically of the same order.
In the sequel we make the choice $u_l = l^a$, $v_l = l^b$, $w_l = l^c$ where a, b, c are three positive real numbers.

Let z_1 and z_2 be two arbitrary populations. As high as l might be, the probability for the chain (X_n^l) to travel from z_1 to z_2 is non–zero. Let r be a positive integer. We will now study the

Asymptotics of $P\left(X_{n+r}^l = z_2/X_n^l = z_1\right)$. This transition probability may be written as

$$\sum P\big(X_{n+r}^l = z_2, V_{n+r-1}^l = v^{r-1}, U_{n+r-1}^l = u^{r-1}, X_{n+r-1}^l = x^{r-1}, \cdots, \\ X_{n+1}^l = x^1, V_n^l = v^0, U_n^l = u^0, X_n^l = z_1\big)$$

the sum being taken over all possible values of sequences of populations in E^m

$$u^0, \cdots, u^{r-1}, v^0, \cdots, v^{r-1} \quad \text{and} \quad x^1, \cdots, x^{r-1}.$$

Anyway the sum is finite and to estimate it we need only to take into account the terms which prevail at infinity. Since the kernel α is irreducible, for r large enough, there is always a path leading from z_1 to z_2 which involves only transition probabilities of order l^{-s} for some positive s. Therefore at least one term in the above sum is asymptotically of order l^{-t} for some positive t.
If the path $\mathcal{P}$: $z_1 = x^0 \to u^0 \to v^0 \to x^1 \to u^1 \to \cdots \to x^{r-1} \to u^{r-1} \to v^{r-1} \to x^r = z_2$ contains a transition of very low probability, the corresponding term in the sum will be negligible compared to any power of $1/l$. More precisely:
If there exists k in $\{0, \cdots, r-1\}$ such that either $u^k \notin \mathcal{U}(x^k)$ or $v^k \notin \mathcal{V}(u^k)$, the product

$$\prod_{k=0}^{r-1} P\left(X_{n+1}^l = x^{k+1}/V_n^l = v^k\right) P\left(V_n^l = v^k/U_n^l = u^k\right) P\left(U_n^l = u^k/X_n^l = x^k\right)$$

is negligible with respect to l^{-s} for any positive s.
If there exists an index k in $\{0, \cdots, r-1\}$ such that $[x^{k+1}] \not\subset [v^k]$, the above term vanishes.
Let $D_r(z_1, z_2)$ be the set of paths

$$x^0 \to u^0 \to v^0 \to x^1 \to u^1 \to \cdots \to x^{r-1} \to u^{r-1} \to v^{r-1} \to x^r$$

satisfying

$$x^0 = z_1,\ x^r = z_2 \quad \text{and} \quad \forall k \in \{0, \cdots, r-1\} \qquad u^k \in \mathcal{U}(x^k),\ v^k \in \mathcal{V}(u^k),\ [x^{k+1}] \subset [v^k]$$

(This set may be empty for small values of r).
It follows from the preceding discussion that

$$P\left(X_{n+r}^l = z_2 / X_n^l = z_1\right) = \sum_{D_r(z_1,z_2)} P\left(X_{n+r}^l = z_2 / V_{n+r-1}^l = v^{r-1}\right) \cdots P\left(U_n^l = u^0 / X_n^l = z_1\right) + o\left(l^{-s}\right)$$

for any positive s. In the above sum, the term associated with the path $\mathcal{P}$ of $D_r(z_1, z_2)$

$$z_1 = x^0 \to u^0 \to v^0 \to x^1 \to u^1 \to \cdots \to x^{r-1} \to u^{r-1} \to v^{r-1} \to x^r = z_2$$

is asymptotically of order l^{-t} where t is equal to

$$V(\mathcal{P}) = a \sum_{k=0}^{r-1} d(x^k, u^k) + b \sum_{k=0}^{r-1} \bar{d}(u^k, v^k) + c \sum_{k=0}^{r-1} \left(m\widehat{f}(v^k) - \sum_{h=1}^{m} f(x_h^{k+1}) \right).$$

This quantity, which will be called the cost of the path, reflects the difficulty for the process to move along the path under consideration. We need only to consider the terms with the lowest exponents. Define the function V_r on $E^m \times E^m$ by

$$V_r(z_1, z_2) = \min_{\mathcal{P} \in D_r(z_1,z_2)} V(\mathcal{P})$$

and let $D_r^*(z_1, z_2)$ be the set of those elements in $D_r(z_1, z_2)$ which realize the minimum $V_r(z_1, z_2)$. We define the length $l(\mathcal{P})$ of a path $\mathcal{P}$ as the number of transitions of the chain (X_n^l) it involves. With each path $\mathcal{P}$ we associate a constant $C(\mathcal{P})$ defined by

$$C(\mathcal{P}) = \prod_{0 \le k < l(\mathcal{P})} \alpha(x^k, u^k) \beta(u^k, v^k) (\operatorname{card} \widehat{v}^k)^{-m} \prod_{i \in [x^{k+1}]} (v^k(i))^{x^{k+1}(i)} \tag{12}$$

and we put

$$C_r(z_1, z_2) = \sum_{\mathcal{P} \in D_r^*(z_1,z_2)} C(\mathcal{P}).$$

In particular, $D_r(z_1, z_2)$ is exactly the set of paths $\mathcal{P}$ of length r such that $C(\mathcal{P}) > 0$.
With these notations we see that:
If $D_r(z_1, z_2)$ is empty,

$$\forall s > 0 \qquad P\left(X_{n+r}^l = z_2 / X_n^l = z_1\right) = o(l^{-s}).$$

If $D_r(z_1, z_2)$ is not empty,

$$P\left(X^l_{n+r} = z_2 / X^l_n = z_1\right) = C_r(z_1, z_2)\, l^{-V_r(z_1,z_2)} + o\left(l^{-V_r(z_1,z_2)}\right).$$

Remark that this formula is valid even if z_1 and z_2 differ only by a permutation (in this case $V_r(z_1, z_2)$ may vanish).
Our next objective is to study the way the chain (X^l_n) behaves when it escapes from the set S. We first examine how the chain may travel from z_1 to z_2 staying outside S.

Asymptotics of $P\left(X^l_{n+r} = z_2,\, \forall k \in \{1, \cdots, r-1\},\, X^l_{n+k} \notin S / X^l_n = z_1\right)$**.**
Let $\widetilde{D}_r(z_1, z_2)$ be the set of elements belonging to $D_r(z_1, z_2)$ such that

$$\forall k \in \{1, \cdots, r-1\} \qquad x^k \notin S.$$

Analogously define the function $\widetilde{V}_r$ on $E^m \times E^m$ by

$$\widetilde{V}_r(z_1, z_2) = \min_{\widetilde{D}_r(z_1,z_2)} \; a \sum_{k=0}^{r-1} d(x^k, u^k) + b \sum_{k=0}^{r-1} \bar{d}(u^k, v^k) + c \sum_{k=0}^{r-1} \left(m\widehat{f}(v^k) - \sum_{h=1}^{m} f(x^{k+1}_h) \right)$$

and let $\widetilde{D}^*_r(z_1, z_2)$ be the set of those elements in $\widetilde{D}_r(z_1, z_2)$ which realize the minimum. Put

$$\widetilde{C}_r(z_1, z_2) = \sum_{\mathcal{P} \in \widetilde{D}^*_r(z_1,z_2)} C(\mathcal{P}).$$

Using exactly the same technique as in the preceding case we get

$$P\left(X^l_{n+r} = z_2,\, \forall k \in \{1, \cdots, r-1\},\, X^l_{n+k} \notin S / X^l_n = z_1\right) = \widetilde{C}_r(z_1, z_2)\, l^{-\widetilde{V}_r(z_1,z_2)} + o\left(l^{-\widetilde{V}_r(z_1,z_2)}\right) \tag{13}$$

whenever $\widetilde{D}_r(z_1, z_2)$ is not empty.

The minimal costs V **and** $\widetilde{V}$**.** The most probable path followed by the chain (X^l_n) between two populations z_1 and z_2 will be a path of minimal cost. We consider two minimal costs, depending on whether the path is constrained outside S or not:

$$V(z_1, z_2) \;=\; \inf_{r \in \mathbb{N}} V_r(z_1, z_2), \qquad \widetilde{V}(z_1, z_2) \;=\; \inf_{r \in \mathbb{N}} \widetilde{V}_r(z_1, z_2)$$

(We make the convention that $V_0(z_1, z_2) = \widetilde{V}_0(z_1, z_2) = \infty$ if $z_1 \neq z_2$ and 0 if $z_1 = z_2$).
Since $\widetilde{D}_r(z_1, z_2) \subset D_r(z_1, z_2)$, then $V_r(z_1, z_2) \leq \widetilde{V}_r(z_1, z_2)$ and taking the minimum over r we obtain

$$V(z_1, z_2) \leq \widetilde{V}(z_1, z_2).$$

We first state some elementary properties of the functions V_r and $\widetilde{V}_r$.

Lemma 3.1. *Let r be a positive integer.*
For every integers k_1 and k_2 such that $1 \le k_1 \le k_2 \le r-1$,

$$V_r(z_1,z_2) = \min_{z_{k_1},z_{k_2}\in\ E^m} \Big\{ V_{k_1-1}(z_1,z_{k_1}) + V_{k_2-k_1}(z_{k_1},z_{k_2}) + V_{r-k_2+1}(z_{k_2},z_2) \Big\}$$

$$\widetilde{V}_r(z_1,z_2) = \min_{z_{k_1},z_{k_2}\in E^m\setminus S} \Big\{ \widetilde{V}_{k_1-1}(z_1,z_{k_1}) + \widetilde{V}_{k_2-k_1}(z_{k_1},z_{k_2}) + \widetilde{V}_{r-k_2+1}(z_{k_2},z_2) \Big\}$$

Corollary 3.2. $\forall r \in \mathbb{N}^* \quad \forall z_1,z_2 \in E^m$

$$V_r(z_1,z_2) = \min_{x^1\cdots x^{r-1}\in\ E^m} \Big\{ V_1(z_1,x^1) + V_1(x^1,x^2) + \cdots + V_1(x^{r-1},z_2) \Big\}$$

$$\widetilde{V}_r(z_1,z_2) = \min_{x^1\cdots x^{r-1}\in E^m\setminus S} \Big\{ \widetilde{V}_1(z_1,x^1) + \widetilde{V}_1(x^1,x^2) + \cdots + \widetilde{V}_1(x^{r-1},z_2) \Big\}$$

Lemma 3.3. *Let r be a positive integer. We have*

$$\begin{array}{lll}
\forall z_1 \in E^m & \forall z_2 \in E^m\setminus S & \widetilde{V}_r(z_1,z_2) \ge V_r(z_1,z_2) > 0 \\
\forall z_1 \in E^m & \forall z_2 \in E^m & V_r(z_1,z_2) = 0 \quad \Longrightarrow \quad [z_2] \subset \widehat{z}_1 \quad \text{and} \quad z_2 \in S \\
\forall z_1 \in A & \forall z_2 \in E^m & z_1 \ne z_2 \quad \Longrightarrow \quad V_r(z_1,z_2) > 0
\end{array}$$

The above results are immediate consequences of the definition of the functions V_r and $\widetilde{V}_r$. The next lemma is of direct interest for V and $\widetilde{V}$:

Lemma 3.4. *Let $r^* = card\ (E^m/\mathfrak{S}_m)$. For every r strictly greater than r^* we have*

$$\inf_{k\le r^*} V_k(z_1,z_2) \le V_r(z_1,z_2), \qquad \inf_{k\le r^*} \widetilde{V}_k(z_1,z_2) < \widetilde{V}_r(z_1,z_2)$$

Proof. Let r be an integer strictly greater than r^* and consider an element of $D_r(z_1,z_2)$:

$$u^0,\cdots,u^{r-1},v^0,\cdots,v^{r-1},x^0,\cdots,x^r.$$

Since $r >$ card $(E^m/\mathfrak{S}_m)$ then necessarily the sequence $x^1,\cdots,x^r$ contains two populations which are equivalent modulo $\mathfrak{S}_m$:

$$\exists\, k_1 \in \{1,\cdots,r\} \quad \exists\, k_2 \in \{1,\cdots,r\} \quad k_1 < k_2 \quad \exists\, \sigma \in \mathfrak{S}_m \qquad \sigma\cdot x^{k_1} = x^{k_2}.$$

In each of the three sequences u^k, v^k, x^k we remove the elements whose index lies in $\{k_1,\cdots,k_2-1\}$. We obtain an element of $D_{r-(k_2-k_1)}(z_1,z_2)$ whose cost is less than the original one, whence $V_{r-(k_2-k_1)}(z_1,z_2) \le V_r(z_1,z_2)$.
Suppose now that the original sequence was in fact in the set $\widetilde{D}_r(z_1,z_2)$: necessarily x^{k_1} and therefore x^{k_2} were in $E^m\setminus S$; by lemma 3.3, the cost of the path between x^{k_1} and x^{k_2} is strictly positive and the cost of the new path between z_1 and z_2 (i.e. the element of $\widetilde{D}_{r-(k_2-k_1)}(z_1,z_2)$ obtained by removing the cycle $x^{k_1}\to\cdots\to x^{k_2}$) is strictly smaller than the original one, whence $\widetilde{V}_{r-(k_2-k_1)}(z_1,z_2) < \widetilde{V}_r(z_1,z_2)$. An immediate descending induction on r gives the desired inequalities. □

As a consequence, we have the following

Corollary 3.5. *Let $r^* = card\ (E^m/\mathfrak{S}_m)$. Then*

$$V(z_1,z_2) = \inf_{k\le r^*} V_k(z_1,z_2), \qquad \widetilde{V}(z_1,z_2) = \inf_{k\le r^*} \widetilde{V}_k(z_1,z_2).$$

Let $I(z_1,z_2)$ (respectively $\widetilde{I}(z_1,z_2)$) be the set of integers k such that $V(z_1,z_2) = V_k(z_1,z_2)$ (respectively $\widetilde{V}(z_1,z_2) = \widetilde{V}_k(z_1,z_2)$). Let $D(z_1,z_2)$ and $\widetilde{D}(z_1,z_2)$ be the set of all paths realizing the minimum in the functions V and $\widetilde{V}$:

$$D(z_1,z_2) = \bigcup_{k\in I(z_1,z_2)} D_k^*(z_1,z_2), \qquad \widetilde{D}(z_1,z_2) = \bigcup_{k\in \widetilde{I}(z_1,z_2)} \widetilde{D}_k^*(z_1,z_2).$$

Lemma 3.4 and corollary 3.2 show that these sets are non–empty; furthermore the set $D(z_1,z_2)$ contains at least a path whose length is less than r^* and the length of each path in $\widetilde{D}(z_1,z_2)$ is less than r^* (so that $\widetilde{D}(z_1,z_2)$ is finite). Finally, we put

$$C(z_1,z_2) = \sum_{\mathcal{P}\in D(z_1,z_2)} C(\mathcal{P}), \qquad \widetilde{C}(z_1,z_2) = \sum_{\mathcal{P}\in \widetilde{D}(z_1,z_2)} C(\mathcal{P}).$$

We are now in position to estimate the transition probabilities of the chain (Z_n^l).
Let y be in E^m and z in S. A straightforward application of the Markov property to the chain (Z_n^l) shows that $P\left(Z_{n+1}^l = z/Z_n^l = y\right) = P_y\left(X_{\tau^l}^l = z\right)$. We decompose this quantity according to the possible values of τ^l:

$$P_y\left(X_{\tau^l}^l = z\right) = \sum_{k=1}^{q} P_y\left(X_{\tau^l}^l = z,\, \tau^l = k\right) + P_y\left(X_{\tau^l}^l = z,\, \tau^l > q\right). \tag{14}$$

However we have already shown in (8) and (10) that

$$P_y\left(\tau^l > q\right) \le (1-(s_l)^m)^q \le m^q(1-s_l)^q \le m^q(m-1)^q\, l^{-c\delta q}.$$

We choose the integer q large enough to ensure that $q > r^*$ and $c\delta q > \widetilde{V}(y,z)$. Since

$$\begin{aligned} P_y\left(X_{\tau^l}^l = z,\, \tau^l = k\right) &= P_y\left(X_k^l = z,\, \forall h \in \{1,\cdots,k-1\},\, X_h^l \notin S\right) \\ &= \widetilde{C}_k(y,z)\, l^{-\widetilde{V}_k(y,z)} + o\left(l^{-\widetilde{V}_k(y,z)}\right) \end{aligned}$$

then for each fixed integer q greater than r^*,

$$\sum_{k=1}^{q} P_y\left(X_{\tau^l}^l = z,\, \tau^l = k\right) \underset{l\to\infty}{\sim} \widetilde{C}(y,z)\, l^{-\widetilde{V}(y,z)}.$$

In addition, the second term in the right–hand side of (14) is dominated by $l^{-c\delta q}$ which is negligible compared to $l^{-\widetilde{V}(y,z)}$ whence finally

$$P\left(Z_{n+1}^l = z/Z_n^l = y\right) = P_y\left(X_{\tau^l}^l = z\right) \underset{l\to\infty}{\sim} \widetilde{C}(y,z)\, l^{-\widetilde{V}(y,z)}. \tag{15}$$

Now, to evaluate the stationary measure of the chain (Z_n^l) we need

4. Some results from the Freidlin–Wentzell theory

In this section we restate word for word some key results from the Freidlin–Wentzell theory. This material is extracted from [4], chapter 6, Lemmas on Markov Chains.

Let H be a finite set and let a subset W be selected in H. A graph consisting of arrows $i \to j \, (i \in H \setminus W, j \in H, i \neq j)$ is called a W–graph if it satisfies the following conditions:

(1) every point $i \in H \setminus W$ is the initial point of exactly one arrow;

(2) there are no closed cycles in the graph.

We note that condition (2) can be replaced by the following condition:

(2′) for any point $i \in H \setminus W$ there exists a sequence of arrows leading from i to some point $j \in W$.

We denote by $G(W)$ the set of W–graphs; we shall use the letter g to denote graphs. If $p_{ij} \, (i, j \in H, j \neq i)$ are numbers, then $\prod_{(i \to j) \in g} p_{ij}$ will be denoted by $\pi(g)$.

Lemma 4.1. *Let us consider a Markov chain with set of states H and transition probabilities p_{ij} and assume that every state can be reached from any other state in a finite number of steps. Then the stationary distribution of the chain is $\{(\sum_{i\in H} Q_i)^{-1} Q_i,\ i \in H\}$, where*

$$Q_i = \sum_{g \in G\{i\}} \pi(g). \tag{16}$$

Lemma 4.2. *Let us be given a Markov chain on a phase space X divided into disjoint sets X_i, where i runs over a finite set H. Suppose that there exist non–negative numbers $p_{ij} \, (j \neq i, i, j \in H)$ and a number $a > 1$ such that*

$$a^{-1} p_{ij} \leq P(x, X_j) \leq a p_{ij} \qquad (x \in X_i, i \neq j) \tag{17}$$

for the transition probabilities of our chain. Furthermore, suppose that every set X_j can be reached from any state x sooner or later (for this it is necessary and sufficient that for any j there exist a $\{j\}$–graph g such that $\pi(g) > 0$). Then

$$a^{2-2h} \left(\sum_{i \in H} Q_i \right)^{-1} Q_i \ \leq\ \nu(X_i)\ \leq\ a^{2h-2} \left(\sum_{i \in H} Q_i \right)^{-1} Q_i \tag{18}$$

for any normalized invariant measure ν of our chain, where h is the number of elements in H and the Q_i are defined by formula (16).

5. Convergence of the measure ν^l

Put for y and z in S

$$p^l_{yz} = \widetilde{C}(y, z)\, l^{-\widetilde{V}(y,z)}.$$

Since S is finite, from the estimation (15) we deduce that for any positive ϵ there exists an integer L such that

$$\forall\, l \geq L \quad \forall y, z \in S \qquad (1+\epsilon)^{-1}\, p^l_{yz} \leq P\left(Z^l_{n+1} = z / Z^l_n = y\right) \leq (1+\epsilon)\, p^l_{yz}.$$

Put for x in S

$$Q^l_x = \sum_{g \in G\{x\}} \pi^l(g)$$

where $G\{x\}$ denotes the set of x–graphs over S and for g in $G\{x\}$,

$$\pi^l(g) = \prod_{(y \to z) \in g} p^l_{yz}.$$

Application of lemma 4.2 to the chain (Z^l_n) yields for $l \geq L$

$$(1+\epsilon)^{2-2|S|} \frac{Q^l_x}{\sum_{y \in S} Q^l_y} \leq \nu^l(x) \leq (1+\epsilon)^{2|S|-2} \frac{Q^l_x}{\sum_{y \in S} Q^l_y}. \tag{19}$$

We study now the asymptotic behavior of Q^l_x. For g in $G\{x\}$,

$$\pi^l(g) = \Bigg(\prod_{(y \to z) \in g} \widetilde{C}(y,z) \Bigg) \exp\Bigg(- \sum_{(y \to z) \in g} \widetilde{V}(y,z) \ln l \Bigg)$$

so that the crucial quantity for computing the asymptotics of Q^l_x is

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

Let $G^*(x)$ be the set of x–graphs which realize the above minimum.
If g belongs to $G^*(x)$, $\pi^l(g)$ is asymptotically of order $l^{-W(x)}$.
If g does not belong to $G^*(x)$, $\pi^l(g)$ is negligible compared to $l^{-W(x)}$.
Finally

$$Q^l_x \underset{l \to \infty}{\sim} \Bigg(\sum_{g \in G^*(x)} \prod_{(y \to z) \in g} \widetilde{C}(y,z) \Bigg)\, l^{-W(x)}.$$

Put

$$\widetilde{D}(x) = \sum_{g \in G^*(x)} \prod_{(y \to z) \in g} \widetilde{C}(y,z)$$

and let

$$W^* = \big\{\, x \in S : W(x) = \min_{y \in S} W(y) \,\big\}.$$

By $W(W^*)$ we mean the value $\min_{y\in S} W(y)$. With these notations, we see that there exists an integer L' such that for $l \geq L'$,

$$(1+\epsilon)^{-1} \frac{\widetilde{D}(x)}{\sum_{y\in W^*}\widetilde{D}(y)}\, l^{W(W^*)-W(x)} \leq \frac{Q^l_x}{\sum_{y\in S} Q^l_y} \leq (1+\epsilon) \frac{\widetilde{D}(x)}{\sum_{y\in W^*}\widetilde{D}(y)}\, l^{W(W^*)-W(x)}$$

whence for $l \geq \max(L, L')$, by (19),

$$(1+\epsilon)^{1-2|S|} \frac{\widetilde{D}(x)}{\sum_{y\in W^*}\widetilde{D}(y)}\, l^{W(W^*)-W(x)} \leq \nu^l(x) \leq (1+\epsilon)^{2|S|-1} \frac{\widetilde{D}(x)}{\sum_{y\in W^*}\widetilde{D}(y)}\, l^{W(W^*)-W(x)}$$

We conclude that:
If x does not belong to W^* then $\lim_{l\to\infty} \nu^l(x) = 0$.
If x belongs to W^* then

$$\lim_{l\to\infty} \nu^l(x) = \frac{\widetilde{D}(x)}{\sum_{y\in W^*}\widetilde{D}(y)} = \frac{\sum_{g\in G^*(x)} \prod_{(y\to z)\in g} \widetilde{C}(y,z)}{\sum_{y\in W^*} \sum_{g\in G^*(y)} \prod_{(z_1\to z_2)\in g} \widetilde{C}(z_1,z_2)}$$

We denote by ν^∞ this limit measure. The complete formula for ν^∞ is

$$\nu^\infty(x) = \frac{\sum_{g\in G^*(x)} \prod_{(y\to z)\in g} \sum_{\mathcal{P}\in\widetilde{D}(y,z)} C(\mathcal{P})}{\sum_{y\in W^*} \sum_{g\in G^*(y)} \prod_{(z_1\to z_2)\in g} \sum_{\mathcal{P}\in\widetilde{D}(z_1,z_2)} C(\mathcal{P})}$$

where $C(\mathcal{P})$ is defined by (12). The measure ν^∞ is concentrated on the set W^* and charges all points of W^*. The crucial question now is: what is the set W^*?

6. The function W and the uniform populations

We derive some properties of the functional W. First, W may equivalently be defined through V instead of $\widetilde{V}$:

Lemma 6.1. *The following equality holds for all x in E^m:*

$$W(x) = \min_{g\in G(x)} \sum_{(y\to z)\in g} \widetilde{V}(y,z) = \min_{g\in G(x)} \sum_{(y\to z)\in g} V(y,z). \tag{20}$$

Proof. The proof of this lemma can be found in [4, chapter 6, lemma 4.1]. $\square$

The perturbed chain (X^l_n) is attracted by the populations of the set S: these populations play a role analogous to the compacta K_i's (the ω–limit sets of the trajectories of the solutions of the deterministic system) in the work of Freidlin and Wentzell. They distinguish two kinds of compacta according to their stability properties. In our situation, the populations in A will play the role of the stable compacta whereas the populations in $S \setminus A$ will behave as the unstable compacta.

Lemma 6.2. *If x belongs to $S \setminus A$ then for each i in $[x]$,*

$$V(x,(i)) = \widetilde{V}(x,(i)) = 0.$$

Proof. This result is straightforward and similar to lemma 4.2 of [4]. □

We restate now textually lemma 4.3 of [4].

Lemma 6.3. *a) Among the x–graphs for which the minimum in $W(x)$ is attained there is one in which from the index y, $y \neq x$ of each population in $S \setminus A$ an arrow $y \to z$ is issued with $V(y,z) = 0$ and z in A.*
b) For a population x in A, the value $W(x)$ can be calculated according to (20), considering graphs on the set of populations in A.
c) If x is a population in $S \setminus A$ then

$$W(x) = \min_{y \in A} \left(W(y) + V(y,x) \right).$$

Proof. The proof of this lemma is only a rewriting of the proof of lemma 4.3 of [4]. □

As a consequence, the set W^* is included in the set of uniform populations A (lemmas 3.3 and 3.4 imply that for y in A and $x \neq y$, $V(y,x) > 0$).

7. Some more notations

For λ in $\mathbb{R}^+$ we define

$$f_\lambda = f^{-1}(\{\lambda\}), \quad f_\lambda^+ = f^{-1}(]\lambda,\infty[), \quad f_\lambda^- = f^{-1}([0,\lambda[).$$

Since the study of W^* requires only consideration of uniform populations, we will often write i instead of (i). For instance, $V(i,j)$ stands for $V((i),(j))$.
For any graph g over A we define

$$V(g) = \sum_{(i \to j) \in g} V(i,j).$$

If X and Y are two subsets of S, we denote by $G_X(Y)$ the set of Y–graphs over $X \cup Y$. For instance, we have $G(X) = G_S(X) = G_{S \setminus X}(X)$. We put

$$W_X(Y) = \min_{g \in G_X(Y)} \sum_{(i_1 \to i_2) \in g} V(i_1,i_2) = \min_{g \in G_X(Y)} V(g).$$

We note $G_X^*(Y)$ the set of graphs in $G_X(Y)$ which realize this minimum.
We define $G_X(\emptyset)$ as the union of all x–graphs over X for all x in X:

$$G_X(\emptyset) = \bigcup_{x \in X} G_X\{x\}.$$

If g is a graph over E, its restriction $g_{|\lambda}$ to the level λ is the graph

$$g_{|\lambda} = \{\, (i_1,i_2) \in g : i_1 \in f_\lambda \}.$$

8. Sufficient conditions to ensure $W^* \subset f^*$

Theorem 8.1. *If the inequality*

$$\sum_{\lambda\in f(E)} W_{f_\lambda}(f_\lambda^+) < \sum_{\lambda\in f(E)} W_{f_\lambda}(f_\lambda^+ \cup f_\lambda^-) - \max_{i\in E\setminus f^*} \min_{j:f(j)\neq f(i)} V(i,j) \tag{21}$$

is satisfied, then the set W^* of all points which may be settled ultimately by the homogeneous algorithm with positive probability is included in the set f^* of the global maxima of the fitness function i.e.

$$\forall x \in E^m \qquad \lim_{l\to\infty} \quad \lim_{n\to\infty} \quad P\left([X_n^l] \subset f^*/X_0^l = x\right) = 1.$$

Proof. Let g be a graph over E. We may decompose the sum $V(g)$ in the following way:

$$V(g) = \sum_{\lambda\in f(E)} \sum_{\substack{i_1\in f_\lambda \\ (i_1\to i_2)\in g}} V(i_1,i_2) = \sum_{\lambda\in f(E)} \sum_{(i_1\to i_2)\in g_{|\lambda}} V(i_1,i_2)$$

i.e.

$$V(g) = \sum_{\lambda\in f(E)} V(g_{|\lambda}).$$

Suppose now that g is in $G\{i\}$ for some i in E. Put $\theta = f(i)$.
Then $g_{|\lambda}$ belongs to $G_{f_\lambda}(f_\lambda^+ \cup f_\lambda^-)$ whenever $\lambda \neq \theta$ whence

$$V(g_{|\lambda}) \geq W_{f_\lambda}(f_\lambda^+ \cup f_\lambda^-). \tag{22}$$

Let j be any point outside f_θ; then $g_{|\theta} \cup \{(i \to j)\}$ is a graph of $G_{f_\theta}(f_\theta^+ \cup f_\theta^-)$ whence

$$V(g_{|\theta}) \geq W_{f_\theta}(f_\theta^+ \cup f_\theta^-) - V(i,j)$$

and the inequality being valid for all j outside f_θ, we have

$$V(g_{|\theta}) \geq W_{f_\theta}(f_\theta^+ \cup f_\theta^-) - \min_{j:f(j)\neq f(i)} V(i,j). \tag{23}$$

Summing up inequalities (22) and (23) yields

$$V(g) \geq \sum_{\lambda\in f(E)} W_{f_\lambda}(f_\lambda^+ \cup f_\lambda^-) - \min_{j:f(j)\neq f(i)} V(i,j)$$

for any graph g in $G\{i\}$. Taking the minimum over all g in $G\{i\}$ we obtain

$$W(i) \geq \sum_{\lambda \in f(E)} W_{f_\lambda}(f_\lambda^+ \cup f_\lambda^-) - \min_{j:f(j)\neq f(i)} V(i,j)$$

so that finally

$$\min_{i\in E\setminus f^*} W(i) \geq \sum_{\lambda \in f(E)} W_{f_\lambda}(f_\lambda^+ \cup f_\lambda^-) - \max_{i\in E\setminus f^*} \min_{j:f(j)\neq f(i)} V(i,j). \tag{24}$$

We build now a near–optimal graph g.
For each λ in $f(E)$, select a graph g_λ in $G^*_{f_\lambda}(f_\lambda^+)$. If $\lambda = f(f^*)$, g_λ is an i^*–graph over f^* for some i^* in f^* (in this case, f_λ^+ is empty).
We define the graph g as the union of the graphs g_λ:

$$(i_1 \to i_2) \in g \iff \exists\, \lambda \in f(E) \quad (i_1 \to i_2) \in g_\lambda.$$

It is easy to see that g is in $G\{i^*\}$. Furthermore we have by construction

$$\forall \lambda \in f(E) \qquad g_{|\lambda} = g_\lambda$$

whence

$$V(g) = \sum_{\lambda \in f(E)} V(g_{|\lambda}) = \sum_{\lambda \in f(E)} W_{f_\lambda}(f_\lambda^+)$$

from which we deduce, for the point i^* of f^*,

$$W(i^*) \leq \sum_{\lambda \in f(E)} W_{f_\lambda}(f_\lambda^+). \tag{25}$$

Putting together the inequalities (21), (24) and (25), we obtain

$$W(i^*) < \min_{i\in E\setminus f^*} W(i)$$

where i^* is a point of f^*, which clearly implies $W^* \subset f^*$. $\square$

The condition (21) depends strongly on the structure of the optimization problem: the essential ingredients which are hidden behind the functionals W and V are the function f and the kernels α and β which determine the asymptotic dynamics of the chain (X_n^l). We derive now some stronger but more explicit conditions for the inequality (21) to hold. Clearly it is sufficient to have both

$$\forall \lambda \in f(E) \setminus f(f^*) \qquad W_{f_\lambda}(f_\lambda^+) \leq W_{f_\lambda}(f_\lambda^+ \cup f_\lambda^-) \tag{26}$$

and

$$W_{f^*}(\emptyset) < W_{f^*}(E \setminus f^*) - \max_{i\in E\setminus f^*} \min_{j:f(j)\neq f(i)} V(i,j). \tag{27}$$

We next prove two fundamental lemmas describing the behavior of the function V with respect to the size of the population m.

9. Increasing of fitness

Lemma 9.1. *Let R be the smallest integer such that*

$$\forall i,j\in E\quad \exists\, r\le R\quad \exists i_0,\cdots,i_r\in E\quad \textit{such that}$$
$$i_0=i,\ i_r=j,\quad \forall k\in\{0,\cdots,r-1\}\quad \alpha(i_k,i_{k+1})>0$$

i.e. R is the minimal number of transitions necessary to join two arbitrary points of E through the kernel α (since α is irreducible and E is finite, R is finite). Put

$$\Delta^\circ=\max_{i\in E\setminus f^*}\min\Big\{\max_{0\le k<r}(f(i)-f(i_k)):i_0=i,\,r\le R,\,f(i_r)>f(i),\,\prod_{0\le k<r}\alpha(i_k,i_{k+1})>0\Big\}$$
$$\Delta^*=\max_{i,j\in f^*}\min\Big\{\max_{0\le k<r}(f(i)-f(i_k)):i_0=i,\,r\le R,\,i_r=j,\,\prod_{0\le k<r}\alpha(i_k,i_{k+1})>0\Big\}.$$

Let i be any point in $E\setminus f^*$. With these notations we have

$$\forall m\in\mathbb{N}^*\qquad \min_{j\in E,f(j)>f(i)}V(i,j)\le aR+c(R-1)\Delta^\circ. \tag{28}$$

Let i,j be two points in f^*. We have also

$$\forall m\in\mathbb{N}^*\qquad V(i,j)\le aR+c(R-1)\Delta^*. \tag{29}$$

Corollary 9.2. *Let $\Delta^{\circledast}=\max(\Delta^\circ,\Delta^*)$. Then,*

$$\sup_{m\in\mathbb{N}^*}\Big\{\max_{i\in E\setminus f^*}\ \min_{j\in E,f(j)>f(i)}V(i,j),\quad \max_{i,j\in f^*}\ V(i,j)\Big\}\ \le\ aR+c(R-1)\Delta^{\circledast}, \tag{30}$$

$$\sup_{m\in\mathbb{N}^*}\ \max_{y\in S\setminus(f^*)^m}\ \min_{z\in S,\widehat{f}(z)>\widehat{f}(y)}\ V(y,z)\ \le\ aR+c(R-1)\Delta^\circ. \tag{31}$$

Remark. Putting
$$\Delta=\max\{\,|f(i)-f(j)|:i,j\in E\,\},$$
we have clearly $\Delta^{\circledast}=\max(\Delta^\circ,\Delta^*)\le\Delta$. All the right–hand side members of the preceding inequalities are thus smaller than $(a+c\Delta)|E|$.

Proof of lemma. Let i be a point of $E \setminus f^*$. By the very definition of Δ°, there exists a sequence $i_0, \cdots, i_r$ of points of E such that:

$$i_0 = i,\ r \leq R,\ f(i_r) > f(i), \prod_{0 \leq k < r} \alpha(i_k, i_{k+1}) > 0 \quad \text{and} \quad \max_{0 \leq k < r} (f(i) - f(i_k)) \leq \Delta^\circ.$$

Among these sequences, we choose a sequence of minimal length, so that necessarily

$$\forall k \in \{0, \cdots, r-1\} \qquad f(i_k) \leq f(i).$$

We put $j = i_r$ and we define

$$x^0 = (i_0),\ x^1 = (i, \cdots, i, i_1),\ \cdots,\ x^{r-1} = (i, \cdots, i, i_{r-1}),\ x^r = (j),$$
$$u^0 = v^0 = x^1, \cdots, u^{r-2} = v^{r-2} = x^{r-1},$$
$$u^{r-1} = v^{r-1} = (i, \cdots, i, i_r).$$

The path

$$x^0 \to u^0 \to v^0 \to x^1 \to u^1 \to \cdots \to v^{r-2} \to x^{r-1} \to u^{r-1} \to v^{r-1} \to x^r$$

clearly belongs to $D_r((i),(j))$ and its cost is $ar + c\sum_{k=0}^{r-2} (f(i) - f(i_{k+1}))$. It follows that

$$V(i,j) \leq V_r(i,j) \leq ar + c\sum_{k=0}^{r-2} \big(f(i) - f(i_{k+1})\big).$$

from which we obtain immediately

$$V(i,j) \leq aR + c(R-1)\Delta^\circ$$

and since $f(j) > f(i)$, inequality (28) is proved.
Let i, j be two points in f^*. By the very definition of Δ^*, there exists a sequence $i_0, \cdots, i_r$ leading from i to j through the kernel α such that $r \leq R$ and

$$\max_{0 \leq k < r} (f(i) - f(i_k)) \leq \Delta^*.$$

We build the sequences u^k, v^k, x^k as above and thus obtain a similar inequality for $V(i,j)$ (with Δ^* instead of Δ°), which proves inequality (29). $\square$

Proof of corollary. Inequality (30) is a direct consequence of lemma 9.1. To prove inequality (31), consider an element y of $S \setminus (f^*)^m$. Pick up a point i in $\widehat{y}$. Using inequality (28), we see that there exists a point j in E such that $f(j) > f(i)$ and $V(i,j) \leq aR + c(R-1)\Delta^\circ$. We build as above a path of length smaller than R which transforms by successive mutations the individual i in j and leaves unchanged all others individuals of y and obtain

$$V(y,(j)) \leq aR + c(R-1)\Delta^\circ$$

whence

$$\min_{j \in E, f(j) > \widehat{f}(y)} V(y,(j)) \leq aR + c(R-1)\Delta^\circ.$$

Taking the maximum over y in $S \setminus (f^*)^m$ and then over m in $\mathbb{N}^*$ yields inequality (31). $\square$

10. Decreasing of fitness

Lemma 10.1. *Let $\rho = \min(a, b/2, c\delta)$ where δ is defined by (9). Then*

$$\forall\, z^1 \in E^m \quad \forall\, z^2 \in E^m \qquad V(z^1, z^2) \geq \rho\, card\{\, k : 1 \leq k \leq m \quad f(z^1_k) > \widehat{f}(z^2)\,\}\,. \tag{32}$$

Remark. The above inequality, being true for the function V, is a fortiori true for the function $\widetilde{V}$.

Proof. We show by induction on r that V_r satisfies this inequality.
Consider first the case $r = 0$. Either $z^1 = z^2$ and the right–hand member vanishes or $z^1 \neq z^2$ and $V_0(z^1, z^2)$ is infinite.
Suppose the result is true at rank r.
Let u^k, v^k, x^k be an element of $D_{r+1}(z^1, z^2)$. We consider two situations:

- If $\widehat{f}(v^0) \leq \widehat{f}(z^2)$ then all elements of the set

$$\{\, z^1_k : 1 \leq k \leq m,\ f(z^1_k) > \widehat{f}(z^2)\,\}$$

have been destroyed during the transition between z^1 and v^0 whence

$$d(z^1, u^0) + 2\bar{d}(u^0, v^0) \geq \mathrm{card}\,\{\, k : 1 \leq k \leq m,\ f(z^1_k) > \widehat{f}(z^2)\,\}.$$

- If $\widehat{f}(v^0) > \widehat{f}(z^2)$ then putting

$$h = \mathrm{card}\,\{\, k : 1 \leq k \leq m,\ \widehat{f}(v^0) > f(x^1_k)\,\}$$

we have

$$\mathrm{card}\,\{\, k : 1 \leq k \leq m,\ f(x^1_k) > \widehat{f}(z^2)\,\} \geq m - h$$

and

$$c\Big(m\widehat{f}(v^0) - \sum_{k=1}^{m} f(x^1_k)\Big) \geq ch\delta \geq \rho h.$$

The induction hypothesis at rank r implies

$$V_r(x^1, z^2) \geq \rho(m - h).$$

Now

$$\begin{aligned} a\sum_{k=0}^{r} d(x^k, u^k) + b\sum_{k=0}^{r} \bar{d}(u^k, v^k) + c\sum_{k=0}^{r}\Big(m\widehat{f}(v^k) - \sum_{h=1}^{m} f(x^{k+1}_h)\Big) \\ \geq c\Big(m\widehat{f}(v^0) - \sum_{k=1}^{m} f(x^1_k)\Big) + V_r(x^1, z^2) \geq \rho h + \rho(m-h) = \rho m \end{aligned}$$

and the desired inequality is true in both cases for the path under consideration. Taking the minimum over all elements of $D_{r+1}(z^1, z^2)$ yields the result at rank $r + 1$ and the induction is completed. $\square$

11. Our last condition

Theorem 11.1. *Let m be an integer such that*

$$\rho m \ > \ \max_{i\in E\setminus f^*} \ \min_{j\in E, f(j)>f(i)} V(i,j)\ , \tag{33}$$

$$\rho m \ \geq \ \max_{i,j\in f^*} \ V(i,j). \tag{34}$$

For this integer m we have

$$\lim_{l\to\infty}\mu^l(f^*) \ = \ \lim_{l\to\infty}\nu^l(f^*) \ = \ 1.$$

Remark. Inequality (30) shows that such an integer m always exists: the quantities on the right–hand side are bounded by $(a+c\Delta)|E|$ which is a constant independent of m. In particular, the conditions are fulfilled as soon as $m > (aR + c(R-1)\Delta^{\circledast})/\min(a, b/2, c\delta)$.

Proof. We prove that for such an m the preceding sufficient conditions (26) and (27) are satisfied. Let λ be in $f(E)\setminus f(f^*)$ and let g be a graph in $G_{f_\lambda}(f_\lambda^+ \cup f_\lambda^-)$. This graph may contain a finite number of transitions from f_λ to f_λ^- : $i_1 \to i'_1, \cdots, i_r \to i'_r$. The first inequality (33) implies that for each i_k, $1\leq k\leq r$, there exists a state j_k in f_λ^+ such that $V(i_k,j_k) < \rho m$ and inequality (32) of lemma 10.1 implies $V(i_k,j_k) < V(i_k,i'_k)$. Let g' be the graph obtained by replacing the r arrows $i_1\to i'_1,\cdots,i_r\to i'_r$ by $i_1\to j_1,\cdots,i_r\to j_r$. The graph g' is in $G_{f_\lambda}(f_\lambda^+)$ and $V(g')\leq V(g)$. This construction being valid for any graph in $G_{f_\lambda}(f_\lambda^+\cup f_\lambda^-)$, we have

$$W_{f_\lambda}(f_\lambda^+) \leq W_{f_\lambda}(f_\lambda^+ \cup f_\lambda^-)$$

for all λ in $f(E)\setminus f(f^*)$ and the first inequality (26) is proved.

Now let g be a graph in $G_{f^*}(E\setminus f^*)$. This graph contains a finite number of transitions from f^* to $E\setminus f^*$: $i_1\to j_1,\cdots,i_r\to j_r$. Let g' be the graph obtained from g by replacing the arrows $i_2\to j_2,\cdots,i_r\to j_r$ by the $(r-1)$ arrows $i_2\to i_1,\cdots,i_r\to i_1$. Inequality (34) implies that $V(g')\leq V(g)$. In addition the graph g' is still in $G_{f^*}(E\setminus f^*)$ and may be decomposed as the union of a graph of $G_{f^*}(\emptyset)$ and the arrow $i_1\to j_1$ whence

$$V(g)\geq V(g')\geq W_{f^*}(\emptyset) + V(i_1,j_1) \geq W_{f^*}(\emptyset)+\rho m.$$

Taking the minimum over all graphs in $G_{f^*}(E\setminus f^*)$ yields

$$W_{f^*}(E\setminus f^*) \geq W_{f^*}(\emptyset)+\rho m.$$

Since by (33)

$$\rho m \ > \ \max_{i\in E\setminus f^*} \ \min_{j\in E, f(j)>f(i)} V(i,j) \quad \geq \quad \max_{i\in E\setminus f^*} \ \min_{j\in E, f(j)\neq f(i)} V(i,j)$$

then

$$W_{f^*}(E\setminus f^*) \ > \ W_{f^*}(\emptyset) \quad + \quad \max_{i\in E\setminus f^*} \ \min_{j\in E, f(j)\neq f(i)} V(i,j)$$

and the second inequality (27) is proved. □

12. Existence of a critical population size m^*

Our fundamental result may be stated as follows.

Theorem 12.1. *Fix the space E, the fitness function f, the mutation and crossover kernels α and β and the positive constants a, b, c.*
There exists a critical population size m^* depending upon all these objects such that, when m is greater than m^*, the set W^* of the minima of the virtual energy W is included in the set f^* of the global maxima of the fitness function f. Moreover we have

$$m^* \;\leq\; \frac{aR + c(R-1)\Delta^{\circledast}}{\min\big(a, b/2, c\delta\big)}.$$

(We recall that $\delta = \min\{\,|f(i) - f(j)| \,:\, i, j \in E,\, f(i) \neq f(j)\,\}$ and $\Delta^{\circledast}$ is defined in Lemma 9.1 *and Corollary* 9.2*. Moreover $\Delta^{\circledast} \leq \Delta = \max\{\,|f(i) - f(j)| : i, j \in E\,\}$.)*

Of course, we expect that in most practical situations, the critical size m^* is much smaller than the cardinality of the search space E, which is a basic requirement for a reasonable algorithm. In fact, considering the upper bound given in theorem 12.1 with $b = \infty$, we obtain that m^* is less than $R\,(a + c\Delta^{\circledast})/\min(a, c\delta)$. The quantities $\Delta^{\circledast}$ and δ depend only on the values of the fitness function. In a concrete situation, δ can be chosen as the level of precision that we require from the genetic algorithm to feel. The coefficient R can be thought of as the diameter of the search space, which is of course much less than its cardinality. In addition, the general upper bound given above is very rough and might be considerably enhanced (for instance by analyzing carefully the lefthand side of (28) and using the condition (33)). In fact, the critical size is related to the difficulty of escaping from a local maximum to go to a better local maximum, rather than to the cardinality of the search space. We will illustrate this on a concrete example on the space $\{0,1\}^{N^d}$ in the last section of the paper. Let us first examine more simple examples.

Consider the space $E = \{\,0, 1, \cdots, N\,\}$ and let α be a markovian mutation kernel on E such that

$$\forall i \in \{1, \cdots, N-1\} \qquad \alpha(i,i-1)+\alpha(i,i+1)+\alpha(i,i) = 1, \quad \alpha(i,i-1) > 0, \quad \alpha(i,i+1) > 0,$$

(the possible mutations from i are $i \to i-1$ and $i \to i+1$), and for the points 0 and N,

$$\alpha(0,1) + \alpha(0,0) = 1, \quad \alpha(0,1) > 0, \quad \alpha(N,N-1) + \alpha(N,N) = 1, \quad \alpha(N,N-1) > 0.$$

For the crossover kernel β, we choose the identity matrix over $E \times E$.
We consider the fitness function f defined by

$$\forall k \in E \qquad f(k) \;=\; 2\cos\Big(k\frac{\pi}{2}\Big) + k\,.$$

This function has local maxima at the points $4k$. If we apply directly the rough upper bound given above, we obtain $m^* \leq (aN + 2c(N-1))/\min(a, b/2, c)$. This bound is very bad, because here the diameter of the space is of the same order as the cardinality of the space itself. However, if we compute the lefthand side of (28) and use the condition (33), we get $m^* \leq 3(a+c)/\min(a, b/2, c)$ which does not even depend on N. Taking b large and $a = c$, we see that $m^* \leq 6$.

Clearly, the algorithm can't possibly work with one single individual: when $m = 1$, the selection operator does nothing at all and the fitness function f does not intervene in the transition mechanism.
Perhaps two individuals is enough to solve all the optimization problems, i.e., $m^* = 2$? The answer is no. We give now examples where the critical size m^* is arbitrarily large. We consider the space E defined in the previous example, with the same crossover and mutation kernels. Let f be a function defined on E, with values in $\mathbb{R}_+^*$ such that

$$f(1) = f(2) = \cdots = f(N-1), \quad f(0) = f(1) + \gamma, \quad f(N) = f(N-1) + \Gamma,$$

where $0 < \gamma < \Gamma$. We have

$$\forall i \in \{1, \cdots, N-1\} \qquad V(i, i-1) = V(i, i+1) = a,$$

(these transitions need only the mutation of one individual: the remainder of the population, driven by the process (X_n^∞), then comes on the new point)

$$V(0,1) = \min\big(am, a + cm\gamma\big), \quad V(N, N-1) = \min\big(am, a + cm\Gamma\big).$$

The transition from 0 to 1, for instance, takes place either trough a general mutation, of cost am, or by a general anti-selection after an individual has mutated to 1, of cost $a+cm\gamma$. Finally

$$V(0, N) = \min\big(am + a(N-1), a + cm\gamma + a(N-1), aN + c(N-1)\gamma\big).$$

There are two possible trajectories to go from 0 to N; either the whole population go from 0 to 1 (cost $V(0,1)$) and then from 1 to N (cost $a(N-1)$), or an explorer starts alone from 0 and mutates until the point N, while the remainder of the population waits in 0 (cost $aN + c(N-1)\gamma$): when the explorer reaches the point N, the whole population throw themselves on N, since $\gamma < \Gamma$.
Similarly

$$V(N, 0) = \min\big(am + a(N-1), a + cm\Gamma + a(N-1), aN + c(N-1)\Gamma + cm(\Gamma - \gamma)\big).$$

Notice yet a crucial difference for the passage from N to 0: for the second kind of trajectories, whenever the explorer reaches 0, it is necessary to perform a general anti–selection of cost $cm(\Gamma-\gamma)$ to bring everyone in 0. In particular, we check that

$$\lim_{m\to\infty} V(0,N) \;=\; aN+c(N-1)\gamma \;<\; \infty$$

and

$$\lim_{m\to\infty} \frac{V(N,0)}{m} \;=\; \min\big(a,c(\Gamma-\gamma)\big) \;>\; 0.$$

We evaluate now the virtual energy W.
For i in $\{1,\cdots,N-1\}$, the i–graph of minimal cost is

$$0\to 1\cdots\to i-1\to i\leftarrow i+1\leftarrow\cdots\leftarrow N-1\leftarrow N$$

so that

$$W(i) \;=\; a(N-2)\;+\;V(0,1)\;+\;V(N,N-1).$$

For the point 0, there are two possible 0–graphs:

$$N\to N-1\to\cdots\to 1\to 0 \quad \text{et} \quad N-1\to\cdots\to 1\to 0\leftarrow N$$

whence

$$\begin{aligned}W(0) \;&=\; a(N-1)\;+\;\min\big(V(N,N-1),V(N,0)\big)\\ &=\; a(N-1)\;+\;\min\big(am,a+cm\Gamma,aN+c(N-1)\Gamma+cm(\Gamma-\gamma)\big).\end{aligned}$$

Similarly, we have

$$W(N) \;=\; a(N-1)\;+\;\min\big(am,a+cm\gamma,aN+c(N-1)\gamma\big).$$

Since $N>1$ and $V(0,1)>a$, $V(N,N-1)>a$, for all i in $\{1,\cdots,N-1\}$, we have $W(0)<W(i)$, $W(N)<W(i)$, so that $W^*\subset\{0,N\}$. Suppose

$$am\le a+cm\gamma \quad \text{et} \quad am\le aN+c(N-1)\gamma.$$

For these inequalities to hold, it is enough that $a\le c\gamma$ and $m\le N$.
In this situation

$$W(0)=W(N)=a(N-1)+am$$

and the set $W^*=\{0,N\}$ is not included in $f^*=\{N\}$.
Thus, whenever $a\le c\gamma$, we have $m^*>N$.
Let us try to explain this phenomenon: the exits from the attraction basins of 0 and N

take place with a general mutation, of cost am, which brings the whole population in 1 and $N-1$. The heights γ and Γ do not intervene and the dynamics can't discriminate between the points 0 and N. The limiting law is then concentrated on both points $\{0, N\}$.

Whenever the mutation kernel α is not symmetric, we may in addition build examples where $f^* \cap W^* = \emptyset$ whenever $m < m^*$.
Consider the space $E = \{\, 0, 1, \cdots, N, N+1, \cdots, N+M-1 \,\}$ and let $f : E \mapsto \mathbb{R}_+^*$ be such that

$$f(1) = f(2) = \cdots = f(N-1) = f(N+1) = \cdots = f(N+M-1),$$
$$f(0) = f(1) + \gamma, \quad f(N) = f(N-1) + \Gamma,$$

where $0 < \gamma < \Gamma$. Let α be a markovian mutation kernel satisfying

$$\forall i \in \{0, \cdots, N+M-2\} \qquad \alpha(i, i+1) \,=\, 1$$

and $\alpha(N+M-1, 0) = 1$. We have

$$\begin{aligned} V(N,0) \;&=\; \min\big(am + a(M-1), a + cm\Gamma + a(M-1), aM + c(M-1)\Gamma + cm(\Gamma - \gamma)\big), \\ V(0,N) \;&=\; \min\big(am + a(N-1), a + cm\gamma + a(N-1), aN + c(N-1)\gamma\big), \end{aligned}$$

and

$$\begin{aligned} W(0) \;&=\; a(N+M-2) + \min\big(am, a + cm\Gamma, aM + c(M-1)\Gamma + cm(\Gamma - \gamma)\big), \\ W(N) \;&=\; a(N+M-2) + \min\big(am, a + cm\gamma, aN + c(N-1)\gamma\big). \end{aligned}$$

To have $W(0) < W(N)$ (and $W^* = \{0\}$), it is enough that

$$aM + c(M-1)\Gamma + cm(\Gamma - \gamma) \;<\; \min\big(am, a + cm\gamma, aN + c(N-1)\gamma\big).$$

Equivalently

$$\begin{aligned} aM + c(M-1)\Gamma \;&<\; m(a - c(\Gamma - \gamma)), \\ aM + c(M-1)\Gamma \;&<\; a + cm(2\gamma - \Gamma), \\ aM + c(M-1)\Gamma + cm(\Gamma - \gamma) \;&<\; aN + c(N-1)\gamma. \end{aligned}$$

We choose γ, Γ such that $0 < \Gamma/2 < \gamma < \Gamma$ and a, c such that $a > c(\Gamma - \gamma)$.
Fix an integer M. The first two inequalities are fulfilled for m sufficiently large. Fix such a value of m. The parameter N appears only in the third inequality. All other parameters being fixed, this inequality is satisfied for N sufficiently large.
We have thus built a space E, a mutation kernel α, a function $f : E \mapsto \mathbb{R}_+^*$ and a set of parameters (a, c, m) (where the population size m may be arbitrarily large) such that $W^* \cap f^* = \emptyset$.

V. Application to the classical simple genetic algorithm

We specialize our general model in order to apply our results to

1. The simple genetic algorithm

We take $E = \{0,1\}^N$ for some integer N. A point i of E is a word of length N over the alphabet $\{0,1\}$ and is noted $i = i_1 \cdots i_N$ where $i_k \in \{0,1\}$. The Hamming distance $H(i,j)$ between two points i,j of E is the number of letters where i and j differ:

$$H(i,j) = \mathrm{card}\,\{\, k : 1 \le k \le m,\ i_k \neq j_k \,\}.$$

The mutation kernel p_l is defined by

$$p_l(i,j) = \begin{cases} 0 & \text{if } H(i,j) > 1, \\ l^{-a} & \text{if } H(i,j) = 1, \\ 1 - N l^{-a} & \text{if } H(i,j) = 0. \end{cases}$$

The associated kernel α is

$$\alpha(i,j) = \begin{cases} 0 & \text{if } H(i,j) > 1, \\ 1 & \text{if } H(i,j) = 1, \\ N & \text{if } H(i,j) = 0. \end{cases}$$

It is irreducible: the minimal number of transitions necessary to join two arbitrary points of E through the kernel α is $R = N$.
In order to build the crossover operator, we define now a cutting operator T_k for k in $\{1, \cdots, N-1\}$; T_k maps $E \times E$ onto $E \times E$ and for i,j in E, we put $T_k(i,j) = (i',j')$ where

$$i' = i_1 \cdots i_k j_{k+1} \cdots j_N, \qquad j' = j_1 \cdots j_k i_{k+1} \cdots i_N.$$

The kernel β is then defined to be

$$\beta((i,j),(i',j')) = \mathrm{card}\,\{\, k : 1 \le k \le N-1,\ T_k(i,j) = (i',j') \,\}$$

and the kernel q_l is

$$q_l((i,j),(i',j')) = \beta((i,j),(i',j'))\, l^{-b} \qquad \text{if } (i,j) \neq (i',j')$$

and

$$q_l((i,j),(i,j)) = 1 - \sum_{(i',j') \neq (i,j)} \beta((i,j),(i',j'))\, l^{-b}.$$

Both hypothesis H_p and H_q are clearly fulfilled. We put as before

$$\Delta^\circ = \max_{i\in E\setminus f^*} \min\left\{ \max_{0\le k<r}(f(i)-f(i^k)) : i^0=i,\, r\le R,\, f(i^r)>f(i),\, \prod_{0\le k<r} H(i^k,i^{k+1})>0\right\}$$

$$\Delta^* = \max_{i,j\in f^*} \min\left\{ \max_{0\le k<r}(f(i)-f(i^k)) : i^0=i,\, r\le R,\, i^r=j,\, \prod_{0\le k<r} H(i^k,i^{k+1})>0\right\}$$

and let $\Delta^{\circledast} = \max(\Delta^\circ, \Delta^*)$.
Within this framework, all convergence results proved in part IV do apply. We restate and comment only the rough results, which take a more precise form in this context.

2. The homogeneous case

Theorem 2.1. *(Convergence of the homogeneous simple genetic algorithm)*
If

$$m \quad > \quad \frac{aN + c(N-1)\Delta^{\circledast}}{\min(a, b/2, c\delta)} \tag{35}$$

then

$$\forall x \in E^m \qquad \lim_{l\to\infty} \quad \lim_{n\to\infty} \quad P\left([X_n^l] \subset f^*/X_0^l = x\right) \;=\; 1.$$

Our condition shows that the optimization problem may be solved with a sufficiently large population size. In addition, we are completely free for the choice of the parameters a, b and c: if we take a very large c, so that $\min(a, b/2) \le c\delta$, we see that δ does not intervene any more in the above condition. The parameter $\Delta^{\circledast}$ describes the difficulty for an individual to travel from one point to a better point. If $\Delta^{\circledast}$ is fixed and N becomes large, a population size of order CN for some constant C will always suffice to handle the problem. Finally, let us remark that our condition on the population size m is very rough: neither the fine structure of the optimization problem nor the possibility of using crossovers to travel between two populations have been taken into account to obtain it. The crucial point is that the perturbation mechanism allows the process to visit all the space, even when the random perturbations are very small. The role of the crossover is thus not fundamental (the algorithm without crossover corresponds to the case $b=\infty$): however this operator creates a lot of new possible transitions for the chain (X_n) and thus certainly decreases the values of the functions V and $\widetilde{V}$ as well as the optimal population size.

3. The inhomogeneous case

In practice, one will not wait for the Markov chain to reach equilibrium before decreasing the perturbations. The idea is then to decrease the intensity of the perturbations as time goes, and one wishes do to this in the most efficient way to obtain the same limiting law. From now onwards we assume that l and n increase simultaneously. More precisely, we take l as an increasing function of n with $\lim_{n\to\infty} l(n) = \infty$.
The Markov chains of our model become thus time inhomogeneous and the transition probabilities depend on the time via the function $l(n)$. We suppress the index l in our notations; for instance X_n stands for X_n^l. Recall that the convergence exponent [10, chapter 3] of the increasing sequence $l(n)$ is defined as the unique non–negative real number λ having the following property: the series

$$l(1)^{-\theta} + l(2)^{-\theta} + \cdots + l(n)^{-\theta} + \cdots$$

converges for $\theta > \lambda$ and diverges for $\theta < \lambda$.
We denote by $T_1, \cdots, T_n, \cdots$ the instants of the successive visits of the chain (X_n) in the set S of equi–fitness populations i.e. $T_n = \inf\{\, k : k > T_{n-1},\, X_k \in S\,\}$. We are mostly interested in the behavior of the chain (X_{T_n}).

Theorem 3.1. *(Convergence of the inhomogeneous simple genetic algorithm)*
1) For the chain (X_{T_n}) to be trapped in f^* after a finite number of transitions, i.e., to have

$$\forall x \in E^m \qquad P\,(\exists N \quad \forall n \geq N \quad [X_{T_n}] \subset f^* / X_0 = x) = 1$$

the convergence exponent of the sequence $l(n)$ must be a positive real number; that is, there must exist two positive real numbers θ_1 and θ_2 such that

$$\sum_{n\geq 0} l(n)^{-\theta_1} = \infty \qquad \text{and} \qquad \sum_{n\geq 0} l(n)^{-\theta_2} < \infty.$$

2) If the convergence exponent λ of the sequence $l(n)$ and the population size m satisfy the inequalities

$$aN + c(N-1)\Delta^{\circledast} \;<\; \lambda \;<\; \min\,(a, b/2, c\delta)\, m$$

then, with probability one, the chain (X_{T_n}) is trapped in f^* after a finite number of transitions, i.e.,

$$\forall x \in E^m \qquad P\,(\exists N \quad \forall n \geq N \quad [X_{T_n}] \subset f^* / X_0 = x) = 1.$$

3) Suppose there exists a real number t strictly greater than one such that for all r in $\mathbb{N}$, the sequences $l(\lfloor tn \rfloor + r)$ and $l(n)$ are logarithmically equivalent. If the convergence exponent λ of the sequence $l(n)$ and the population size m both satisfy the inequalities

$$aN + c(N-1)\Delta^{\circledast} \;<\; \min\,(a, b/2, c\delta)\, m \;\leq\; \lambda$$

then

$$\forall x \in E^m \qquad \lim_{n\to\infty} \quad P\left([X_n] \subset f^*/X_0 = x\right) \,=\, 1.$$

The proofs of these results involve technical large deviations estimates. To avoid lengthy developments, we do not reproduce them here. To ensure the convergence toward f^* in the inhomogeneous case, we must first ensure that the homogeneous algorithm converges (condition (35)) and that the rate of increasing of the sequence $l(n)$ is carefully adapted to avoid the process to be trapped in any sub–optimal population.

The next very important issue is the speed of convergence of the algorithm. This question is addressed in [2] using the very technical tools developed by Catoni and Trouvé for analyzing generalized simulated annealing processes.

4. The ground state of the Ising model

We consider a finite box $\Lambda = \{1 \cdots N\}^d$ of side length N in the d–dimensional integer lattice $\mathbb{Z}^d$. A point of this box is called a site. Sites will be denoted by the letters x, y. We wrap this box into a torus and we define a neighbourhood relation on Λ by: $x \sim y$ if all the coordinates of x and y are equal except one which differs by 1 or $N-1$. At each site x of Λ there is a spin taking the values -1 or $+1$. The set of all possible spins configurations is $X = \{-1,+1\}^\Lambda$. Configurations of spins will be denoted by the letters η, σ. The value of the spin at site x for a configuration σ is denoted by $\sigma(x)$.

The energy of a configuration σ is

$$E(\sigma) \;=\; -\frac{1}{2}\sum_{\{x,y\}:x\sim y} \sigma(x)\sigma(y) \;-\; \frac{h}{2}\sum_{x\in\Lambda}\sigma(x) \tag{36}$$

where $h>0$ is the external magnetic field. We consider the situation where N is large and h is small, with $Nh > 2d$. A ground state of the Ising model is a configuration realizing the global minimum of the energy. Because h is small, in order to minimize the energy, we should first try to minimize the first term appearing in the energy. The effect of this term is to make neighbouring sites have the same sign (this is the ferromagnetic interaction). The effect of the second term is to make the spins choose the sign of the magnetic field. Hence the unique ground state is the configuration with all spins up. There is another very stable configuration, the one with all spins down. We use the mutation kernel of the simple genetic algorithm defined in paragraph V.1. For any configurations σ, η, we set

$$p_l(\sigma,\eta) = \begin{cases} l^{-a} & \text{if } \eta = \sigma^x \text{ for some site } x \\ 1 - N^d l^{-a} & \text{if } \eta = \sigma \\ 0 & \text{otherwise} \end{cases} \tag{37}$$

where σ^x is the configuration σ flipped at x i.e.

$$\sigma^x(y) \;=\; \begin{cases} +\sigma(y) & \text{if } y \neq x \\ -\sigma(y) & \text{if } y = x \end{cases}$$

Hence the mutation consists in making one spin-flip at any site of the configuration. The energy landscape associated with this energy and this mutation kernel is very complicated and full of local minima. For instance any configuration which is the union of parallelepipedic droplets of $+$ in a sea of $-$ such that any two droplets are at a distance larger than two is a local minimum. Some features of this energy landscape have been analyzed in dimension two and three in order to deal with metastability questions associated to the Metropolis dynamics [1,9]. Partial results are also available in arbitrary dimension [8]. We will rely on these results to give upper bounds on the critical population size m^* associated to the problem of minimizing the energy E. The critical population size of the simple genetic algorithm running on the space $X = \{-1,+1\}^{N^d}$, with the mutation kernel defined in (37), with no crossover, and with the fitness function $-E$ defined in (36) satisfies

$$m^* \;\leq\; N^d \left(a + c\,\frac{C(d)}{h^{d-1}} \right) (\min(a,ch))^{-1}$$

where $C(d)$ is a constant depending only on the dimension. Thus the critical size is less than a polynomial in N, while the size of the state space is 2^{N^d}. In fact we have used here the rough upper bound given in (28). A more careful analysis would yield

$$m^* \;\leq\; \left(\frac{2d}{h} \right)^d \left(a + c\,\frac{C(d)}{h^{d-1}} \right) (\min(a,ch))^{-1}$$

which is even independent of N! Notice however that the length N has to satisfy $Nh > 2d$. This phenomenon illustrates the fact that the critical size is related to a fitness barrier the population has to overcome to find the true global minimum, rather than to the cardinality of the search space. Here the fitness barrier corresponds to the energy necessary to escape from the metastable state (all spins down) to reach the ground state (all spins up). This energy barrier is the energy of a critical droplet (see [1,8,9]).

References

1. G. Ben Arous–R. Cerf, *Metastability of the three dimensional Ising model on a torus at very low temperatures*, Electronic Journal of Probability **1** (1996), 1–55.
2. R. Cerf, *The dynamics of mutation–selection algorithms with large population sizes*, Ann. Inst. Henri Poincaré **32 no. 4** (1996), 455–508.
3. J.D. Deuschel–C. Mazza, *L^2 convergence of time nonhomogeneous Markov processes: I. Spectral Estimates*, Annals Applied Probab. **Vol. 4 no. 4** (1994), 1012–1056.

4. M.I. Freidlin–A.D. Wentzell, *Random perturbations of dynamical systems*, Springer–Verlag, New York, 1984.
5. D. Goldberg, *Genetic algorithms in search, optimization and machine learning*, Addison–Wesley, 1989.
6. J.H. Holland, *Adaptation in natural and artificial systems*, The University of Michigan Press, Ann Arbor, 1975.
7. Y. Kifer, *Random perturbations of dynamical systems*, Birkhäuser, Boston–Basel, 1988.
8. E.J. Neves, *A discrete variational problem related to Ising droplets at low temperatures*, Jour. Stat. Phys. **80 nos 1/2** (1995), 103–123.
9. E.J. Neves and R.H. Schonmann, *Critical droplets and metastability for a Glauber dynamics at very low temperatures*, Commun. Math. Phys. **137** (1991), 209–230.
10. G. Pólya–G. Szegö, *Problems and theorems in analysis I*, Springer–Verlag, Berlin Heidelberg, 1972.
11. E. Seneta, *Non–negative matrices and Markov chains*, Springer–Verlag, New York, 1981.
12. A. Trouvé, *Asymptotical behaviour of several interacting annealing processes*, Probab. Theory Relat. Fields **Vol. 102** (1995), 123–143.

Université Paris sud, Mathématiques, bâtiment 425, 91405 Orsay Cedex–France
E-mail address: Raphael.Cerf@math.u-psud.fr