THE DYNAMICS OF MUTATION-SELECTION ALGORITHMS WITH LARGE POPULATION SIZES

LA DYNAMIQUE DES ALGORITHMES DE MUTATION–SÉLECTION AVEC DE GRANDES TAILLES DE POPULATION

Raphaël CERF

Université de Montpellier

June 1994

Abstract. We build the mutation–selection algorithm by randomly perturbing a very simple selection scheme. Our process belongs to the class of the generalized simulated annealing algorithms studied by Trouvé. When the population size m is large, the various quantities associated with the algorithm are affine functions of m and the hierarchy of cycles on the set of uniform populations stabilizes. If the mutation kernel is symmetric, the limiting distribution is the uniform distribution over the set of the global maxima of the fitness function. The optimal convergence exponent defined by Azencott, Catoni and Trouvé is an affine strictly increasing function of m.

Résumé. Nous construisons l'algorithme de mutation–sélection en perturbant aléatoirement un mécanisme de sélection très simple. Le processus obtenu entre dans la classe des algorithmes de recuit simulé généralisés étudiés par Trouvé. Lorsque la taille de la population m est grande, les différentes quantités associées à l'algorithme sont des fonctions affines de m et la hiérarchie des cycles sur l'ensemble des populations uniformes se stabilise. Si le noyau de mutation est symétrique, la distribution limite est la distribution uniforme sur l'ensemble des maxima globaux de la fonction fitness. L'exposant optimal de convergence défini par Azencott, Catoni et Trouvé est une fonction affine strictement croissante de m.

1991 *Mathematics Subject Classification.* 60F10, 60J10, 92D15.

Key words and phrases. Freidlin–Wentzell theory, evolutionary algorithms, stochastic optimization.

Professor Jean–Arcady Meyer first spoke to me of evolutionary 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.

I thank Alain Trouvé for several discussions about the generalized simulated annealing and for comments which improved this work.

I thank an anonymous Referee for his meticulous work and numerous changes which increased significantly the clarity of this paper. He gave a proof of proposition 10.1 which is considerably simpler than the original one.

Typeset by $\mathcal{AMS}$-TEX

1. Introduction

Evolutionary algorithms are optimization techniques based on the mechanics of natural selection [6]. Experimental simulations have demonstrated their efficiency: they are robust, flexible and in addition extremely easy to implement on parallel computers [5].

Unfortunately, there is a critical lack of theoretical results for the convergence of this kind of algorithms. In a previous paper [3], the author constructed a Markovian model of Holland's Genetic Algorithm. This model was built by randomly perturbing a very simple selection scheme. The study of the asymptotic dynamics of the process was carried out with the powerful tools developed by Freidlin and Wentzell for the study of random perturbations of dynamical systems, yielding what seemed to be the first mathematically well–founded convergence results for genetic algorithms.

It was proved that the convergence toward the global maxima of the fitness function becomes possible when the population size is greater than a critical value (which depends on all the characteristics of the optimization problem). Surprisingly, the crossover did not play a fundamental role in this model. The crucial point to ensure the desired convergence was the delicate asymptotic interaction between the local perturbations of the individuals (i.e. the mutations) and the selection pressure.

In this paper, we deal with the two–operators mutation–selection algorithm. We introduce a new parameter (analogous to the temperature of the simulated annealing) which controls the intensity of the random perturbations. We study the asymptotic dynamics of the process when the perturbations disappear. This new point of view in the field of genetic algorithms allows to prove several results concerning the convergence of the law of the process toward the maxima of the fitness function. We focus here on the dynamics of the algorithm when the population size m becomes very large.

Once more, we follow the road opened by Freidlin and Wentzell [4]. However we will use the concepts and tools introduced by Catoni in his precise study of the sequential simulated annealing [1,2]; these were put in a more general framework by Trouvé, who carried out a systematic study (initiated by Hwang and Sheu [7]) of a class of processes he baptized "generalized simulated annealing" [9,10,11]. Our algorithm belongs to this class of processes and our study will thus heavily rely on Catoni and Trouvé's work.

This paper has the following structure. First we describe our model of the mutation–selection algorithm. We show how it fits in the framework of the generalized simulated annealing and give a sufficient condition to ensure the concentration of the law of the algorithm on the global maxima of the fitness function. The dynamics of the algorithm is best described through the decomposition of the space into a hierarchy of cycles, which are the most attractive and stable sets we can build for the perturbed process. We prove a result valid for the generalized simulated annealing which reduces the study of the dynamics to the bottom of the cycles (the uniform populations in our situation). The key result for our study lies in the structure of the most probable trajectories of populations joining two uniform populations: a small group of individuals sacrifice themselves in order to create an ideal path which is then followed by all other individuals. As a consequence, the various quantities associated with the algorithm (such as the communication cost, the virtual energy, the communication altitude ...) are affine functions of the population size. We then prove that the hierarchy of cycles on the set of the uniform populations stabilizes when the population size is large. The structure of the limiting hierarchy of cycles depends upon the relative values of the mutation cost and the variations of the fitness function. However, the fitness function is constant on the "bad" cycles (i.e. those which do not contain the minima of the virtual energy). Furthermore, if the mutation kernel is symmetric and if the population size is greater than a critical value, the limiting distribution is the uniform distribution over the set of the global maxima of the fitness function. Finally, we investigate the critical constants describing the convergence of the algorithm. The optimal convergence speed exponent defined by Azencott, Catoni and Trouvé [1,2,9,10,11] (which gives the optimal convergence rate toward the maxima in finite time) increases linearly with m. This fact shows that our algorithm is intrinsically parallel: it involves only local independent computations.

2. Description of the model

2.1. The fitness landscape. 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 set E is endowed with an irreducible Markov kernel α, that is a function defined on $E \times E$ with values in $[0,1]$ satisfying

$$\forall i \in E \qquad \sum\nolimits_{j \in E} \alpha(i,j) = 1,$$

$$\forall i,j \in E \quad \exists e_1,\cdots,e_r \in E \quad e_1 = i,\, e_r = j,\, \forall k \in \{1 \cdots r-1\} \quad \alpha(e_k, e_{k+1}) > 0.$$

The three objects E, f, α define an abstract fitness landscape. We are searching for the set f^* of the global maxima of f i.e.

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

By $f(f^*)$ we mean the maximum value of f over E i.e. $\max_{j\in E} f(j)$. Symbols with a star $*$ in superscript will denote sets realizing the minimum or the maximum of a particular functional. The points of E will be called individuals and will be mostly denoted by the letters i, j, e.

2.2. The population space. We will consider Markov chains with state space E^m where m is the population size of our algorithm; the m–uples of elements of E, i.e. the points of the set E^m, are called populations and will be mostly denoted by the letters x, y, z. For x in E^m and i in E, $x(i)$ is the number of occurrences of i in x:

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

With f we associate a function $\widehat{f}$ defined on E^m by

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

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

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

If i belongs to E, (i) is the m–uple whose m components are equal to i and U is the set of all such m-uples (which are called the uniform populations). We sometimes identify an element i of E with the uniform population (i). Thus f^* may be seen as a subset of U.

2.3. 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

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

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. Notice that after a while, the chain (X_n^∞) starting from a population x is trapped in a uniform population selected in $\widehat{x}$:

$$P\left(\exists i \in \widehat{x} \quad \exists N \in \mathbb{N} \quad \forall n \ge N \quad X_n^\infty = (i) / X_0^\infty = x\right) = 1.$$

In fact, the absorbing states of the chain (X_n^∞) are exactly the uniform populations U.

2.4. The perturbed Markov chain (X_n^l). The previous Markov chain (X_n^∞) is randomly perturbed by two distinct mechanisms. The first one acts directly upon the population and mimics the phenomenon of mutation. The second one consists in loosening the selection of the individuals. The intensity of the perturbations is governed by a positive parameter l so that we obtain a family of Markov chains $(X_n^l)_{n \in \mathbb{N}}$ indexed by l. As l grows toward infinity, the perturbations progressively disappear. The transition from X_n^l to X_{n+1}^l includes two stages corresponding to the mutation and the selection operators:

$$X_n^l \xrightarrow{\text{mutation}} Y_n^l \xrightarrow{\text{selection}} X_{n+1}^l.$$

We now describe more precisely these two operators.

2.5. $X_n^l \longrightarrow Y_n^l$: mutation. The mutation operator is modelled by random independent perturbations of the individuals of the population X_n^l. With the Markov kernel α (which is initially given with the set E and the function f), we build a family of Markov kernels $(\alpha_l)_{l\geq 1}$ on the set E. Let a be a positive real number. Define for i, j in E and $l \geq 1$

$$\alpha_l(i,j) = \begin{cases} \alpha(i,j)\, l^{-a} & \text{if } i \neq j \\ 1 - \sum_{e\neq i} \alpha(i,e)\, l^{-a} & \text{if } i = j \end{cases}$$

The quantity $\alpha_l(i,j)$ is the probability that a mutation transforms the individual i into the individual j at the perturbation level l. The transition probabilities from X_n^l to Y_n^l are then given by

$$P\left(Y_n^l = y / X_n^l = x\right) = \alpha_l(x_1, y_1) \cdots \alpha_l(x_m, y_m).$$

Note that the mutations vanish when l goes to infinity i.e.

$$\lim_{l\to\infty} P\left(Y_n^l = y / X_n^l = x\right) = \delta(x_1, y_1) \cdots \delta(x_m, y_m) \tag{1}$$

where $\delta(i,j)$ is the Kronecker symbol (the identity matrix indexed by E):

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

2.6. $Y_n^l \longrightarrow X_{n+1}^l$: selection. The selection operator is built with a selection function [3]. We evaluate the fitness of the individuals of the population Y_n^l and we build a distribution probability over Y_n^l which is biased toward the best fit individuals and which progressively concentrates on the set $\widehat{Y}_n^l$ as l goes to infinity. We then select independently m individuals from Y_n^l to form the population X_{n+1}^l. Let c be a positive real number. The transition probabilities from Y_n^l to X_{n+1}^l are

$$P\left(X_{n+1}^l = x / Y_n^l = y\right) = \prod_{r=1}^{m} \frac{y(x_r)\exp(cf(x_r)\ln l)}{\sum_{k=1}^{m} \exp(cf(y_k)\ln l)}.$$

We have

$$\lim_{l\to\infty} P\left(X_{n+1}^l = x / Y_n^l = y\right) = \prod_{k=1}^{m} \frac{1_{\widehat{y}}(x_k) y(x_k)}{\operatorname{card} \widehat{y}} \tag{2}$$

i.e. the selections of individuals below peak fitness tend to disappear when l goes to infinity.

2.7. The vanishing perturbations. The two operators play antagonistic roles: whereas the mutation tends to disperse the population over the space E, the selection tends to concentrate the population on the current best individual. Both are built through random perturbations: random perturbations of the identity for the mutations, random perturbations of the very strong selection mechanism of the chain (X_n^∞) for the selection. With this scheme of non–overlapping generations, it is essential that the mutations vanish: otherwise, the current population could be destroyed at any time through a massive mutation event with a null perturbation cost, preventing the concentration of the law of the process on the set of the global maxima of the fitness function. It is yet questionable whether the vanishing mutations and the growing selection pressure have a biological meaning. However, the optimization problem leads to the following question: is it possible to exert a control over these very rudimentary operators to ensure the convergence toward a state of maximum fitness?
Formulas (1) and (2) yield

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

so that the process (X_n^l) is a perturbation of the process (X_n^∞). A crucial point is to give the same intensity to the two kinds of perturbations so that they interact properly when l goes to infinity. More precisely, the rate of convergence of the transition probabilities in formulas (1) and (2) should be logarithmically of the same order (this is the reason why we introduce two more parameters a and c). The parameters a and c are not independent: we will in fact compare a/c to quantities related to the fitness function f. Rescaling the function f with the parameter c has the same effect as changing the mutation cost a. To avoid breaking the symmetry, we keep both parameters a and c in the sequel. Finally, we are entirely free for the choice of the control parameter l: here l goes to infinity. We could also take $\ln l$ or the temperature $T=1/\ln l$.

3. Asymptotic expansion of $P\left(X_{n+1}^l=z/X_n^l=y\right)$

For a fixed value of l, the expression of the transition matrix of the chain (X_n^l) is quite complicated. We will focus on the asymptotic behavior of the chain, when l goes to infinity. The first step consists in estimating the transition matrix of (X_n^l).
By the very construction of the process (X_n^l), we have

$$P\left(X_{n+1}^l=z/X_n^l=x\right)=\sum_{y\in E^m} P\left(X_{n+1}^l=z/Y_n^l=y\right)P\left(Y_n^l=y/X_n^l=x\right).$$

For each y in E^m,

$$P\left(X_{n+1}^l=z,Y_n^l=y/X_n^l=x\right)\underset{l\to\infty}{\sim}\alpha(x,y)\gamma(y,z)\exp-\left(\left(ad(x,y)+c\sum_{k=1}^m\left(\widehat{f}(y)-f(z_k)\right)\right)\ln l\right)$$

where we note for x, y, z in E^m

$$\alpha(x,y) = \prod_{k: x_k \neq y_k} \alpha(x_k, y_k), \quad \gamma(y,z) = \prod_{r=1}^{m} \frac{y(z_r)}{\operatorname{card} \widehat{y}}$$

and $d(x,y)$ is the Hamming distance between the vectors x and y i.e.

$$d(x,y) = \operatorname{card} \{ \, k : 1 \leq k \leq m, \, x_k \neq y_k \, \}.$$

The above transition probability vanishes whenever $\alpha(x,y)\gamma(y,z) = 0$. There are two different terms in the exponent. The quantity $d(x,y)$ is simply the number of mutations necessary to go from x to y and $ad(x,y)$ represents the perturbation cost necessary to achieve these mutations. The second term originates from the fact that individuals of y with a fitness strictly less than $\widehat{f}(y)$ may have been selected to form z: such an event will be called an anti–selection. The term $c\sum_k(\widehat{f}(y) - f(z_k))$ represents the perturbation cost necessary to achieve all the anti–selections.
We define next the communication cost V_1 on $E^m \times E^m$ by

$$V_1(x,z) = \min \Big\{ \, ad(x,y) + c\sum_{k=1}^{m} (\widehat{f}(y) - f(z_k)) : y \in E^m, \alpha(x,y)\gamma(y,z) > 0 \, \Big\}.$$

We put finally $q_1(x,z) = \sum \alpha(x,y)\gamma(y,z)$, the sum being extended over the populations y realizing the minimum in $V_1(x,z)$. With these notations, we see that

$$P\left(X^l_{n+1} = z / X^l_n = x\right) \underset{l \to \infty}{\sim} q_1(x,z) \exp\left(-V_1(x,z) \ln l\right).$$

Remark in addition that for each x, z in E^m

$$P\left(X^l_{n+1} = z / X^l_n = x\right) = 0 \iff V_1(x,z) = \infty \iff q_1(x,z) = 0.$$

We are now in the framework of the generalized simulated annealing studied by Trouvé [8,9,10,11]. That is, the transition probabilities of the process (X^l_n) form a family of Markov kernels on the space E^m indexed by l which is admissible for the communication kernel q_1 and the cost function V_1 [9, Definition 3.1].

4. Convergence of the mutation–selection algorithm

We will use graphs over the set E^m. By a graph over E^m we mean a set of arrows $x \to y$ with endpoints in E^m. Thus our graphs are all oriented.

Notation 4.1. Let g be a graph on E^m. The V_1–cost of g is

$$V_1(g) = \sum_{(x\to y)\in g} V_1(x,y).$$

We recall that an x–graph is a graph with no arrow starting from x and such that for any $y \neq x$ there exists a unique path in g leading from y to x. The set of all x–graphs is denoted by $G(x)$. For more details and notations concerning graphs, see [4, chapter 6].

Definition 4.2. The virtual energy W_1 associated with the cost function V_1 is defined by

$$\forall x \in E^m \qquad W_1(x) = \min\{\, V_1(g) : g \in G(x)\,\}.$$

We put also

$$W_1(E^m) = \min\{\, W_1(x) : x \in E^m\,\}, \qquad W_1^* = \{\, x \in E^m : W_1(x) = W_1(E^m)\,\}.$$

Proposition 4.3. *(Freidlin and Wentzell)*

$$\forall x \in E^m \qquad \lim_{l\to\infty} \lim_{n\to\infty} P\left(X_n^l \in W_1^* / X_0^l = x\right) = 1.$$

Theorem 4.4. *(convergence of the homogeneous algorithm)*
Fix the space E, the fitness function f, the kernel α and the positive constants a, c. There exists a critical population size m^* depending upon all these objects such that

$$m \geq m^* \qquad \Longrightarrow \qquad W_1^* \subset f^*.$$

This theorem has been proved in [3] and various upper bounds for the value m^* were given there; for instance

$$m^* \leq \frac{aR + c(R-1)\Delta}{\min(a, c\delta)}$$

where

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

and R is the minimal number of transitions necessary to join two arbitrary points of E through the kernel α i.e. R is the smallest integer such that

$\forall i,j \in E \quad \exists\, r \leq R \quad \exists e_1, \cdots, e_{r+1} \in E \quad$ such that

$$e_1 = i,\ e_{r+1} = j, \quad \forall k \in \{1, \cdots, r\} \quad \alpha(e_k, e_{k+1}) > 0.$$

The preceding bound may be improved in several ways with more complicated constants: however, since we will deal only with very large population sizes, we do not care about a precise estimation of m^*. Anyway, it is obvious that m^* depends strongly on the fitness landscape (E, f, α).

We restate now Trouvé's result for the convergence in the inhomogeneous case, where the parameter l is an increasing function of n. We have then an inhomogeneous Markov chain $(X_n^{l(n)})_{n\geq 0}$ and we suppress the superscript l: X_n stands for $X_n^{l(n)}$.

Theorem 4.5. *(Trouvé, [9, Theorem 3.23])*
There exists a constant H_1, called the critical height, such that for all increasing sequences $l(n)$, we have the equivalence

$$\forall x \in E^m \qquad P\left(X_n \in W_1^* / X_0 = x\right) \underset{n\to\infty}{\longrightarrow} 1 \qquad \Longleftrightarrow \qquad \sum_{n=0}^{\infty} l(n)^{-H_1} = \infty.$$

Note that H_1 may depend on m.

Corollary 4.6. *Suppose $m \geq m^*$ and $\sum_{n=0}^{\infty} l(n)^{-H_1} = \infty$. Then*

$$\forall x \in E^m \qquad P\left(X_n \in f^* / X_0 = x\right) \underset{n\to\infty}{\longrightarrow} 1.$$

All the tools introduced by Trouvé in [9,10,11] for the study of the generalized simulated annealing may be used for analyzing the mutation–selection algorithm. The particular interest of our process lies in the presence of several control parameters among which the population size m is of paramount importance. In the sequel, we will try to understand the behaviour of the mutation–selection algorithm for very large populations. The next section contains several results valid for the generalized simulated annealing; they will allow us to reduce the study of the dynamics of the process to the set U of uniform populations (whose cardinality is independent of m).

5. General results about the virtual energy and the communication altitude

Let us introduce some notations and definitions.
If $\mathcal{S}$ is an arbitrary set, $\mathcal{S}^{(\mathbb{N})}$ denotes the set of paths in $\mathcal{S}$, that is the finite sequences of elements of $\mathcal{S}$. A path s in $\mathcal{S}$ is noted indifferently

$$s = (s_1, \cdots, s_r), \quad s = (s^1, \cdots, s^r) \quad \text{or} \quad s = s_1 \to \cdots \to s_r$$

and its length is noted $|s|$ (r in the above example). A path s in $\mathcal{S}$ is said to join two elements t_1 and t_2 if $s_1 = t_1$ and $s_{|s|} = t_2$; the set of all paths in $\mathcal{S}$ joining the points t_1 and t_2 is noted $\mathcal{S}^{(\mathbb{N})}(t_1, t_2)$.

Notation 5.1. By D^m we denote the paths in E^m which correspond to possible trajectories of the process (X_n^l) i.e. the paths p in E^m satisfying

$$\forall k \quad 1 \leq k < |p| \qquad V_1(p_k, p_{k+1}) < \infty.$$

The V_1–cost of such a path is

$$V_1(p) = \sum_{k=1}^{|p|-1} V_1(p_k, p_{k+1}).$$

This definition coincides with the cost of a graph (see notation 4.1) if we consider the path as a graph over E^m. Notice that for the empty path (which has a null length), the cost is zero. If p belongs to $E^{m(\mathbb{N})} \setminus D^m$, we put $V_1(p) = \infty$. We put also for y, z in E^m

$$D^m(y,z) = D^m \cap E^{m(\mathbb{N})}(y,z).$$

Definition 5.2. We define the minimal communication cost V for y and z in E^m by

$$V(y,z) = \inf \{ V_1(p) : p \in D^m(y,z) \}$$

and let $D^{m*}(y,z)$ be the paths of $D^m(y,z)$ realizing the above minimum.

Notice that for all x in E^m, we have $V(x,x) = 0$. With this new communication cost V, we associate a cost function for the graphs and a virtual energy.

Definition 5.3. For a graph g on E^m, we define its V–cost by

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

and we define the virtual energy for any element x of E^m by

$$W(x) = \min \{ V(g) : g \in G(x) \}.$$

We put also $W(E^m) = \min \{ W(x) : x \in E^m \}$, $W^* = \{ x \in E^m : W(x) = W(E^m) \}$.

We will now show that the minimal cost V contains all the relevant information to carry out the study of the dynamics of the generalized simulated annealing: the following results are not specific to the case of the mutation–selection algorithms and are valid in the general framework. Propositions 5.4 and 5.6 below are consequences of the more general lemma 5.6 of [10]. However we give here direct proofs which do not involve the hierarchy of cycles.

Proposition 5.4. *The virtual energy W associated with the communication cost V coincides with the virtual energy W_1 associated with the communication cost V_1 i.e.*

$$\forall x \in E^m \quad \min \left\{ \sum_{(y \to z) \in g} V(y,z) : g \in G(x) \right\} = \min \left\{ \sum_{(y \to z) \in g} V_1(y,z) : g \in G(x) \right\}.$$

As a consequence, we have $W(E^m) = W_1(E^m)$ and $W^* = W_1^*$.

Proof. Since $V \le V_1$ we have clearly $W \le W_1$. The proof of the reverse inequality $W \ge W_1$ is similar to the proof of lemma 4.1 of [4, chapter 6]. □

The fundamental quantity used to build the hierarchical decomposition of the space into cycles is the communication altitude.

Definition 5.5. (Trouvé, [9, Definition 3.15])
The communication altitude $A_1(x,y)$ between two distinct points x and y of E^m is

$$A_1(x,y) = \inf\Big\{ \max_{1\le k<|p|} W_1(p_k) + V_1(p_k,p_{k+1}) : p \in D^m(x,y) \Big\}.$$

For any x in E^m, we put $A_1(x,x) = W_1(x)$.

The communication altitude may equivalently be defined through the cost function V.

Proposition 5.6. *Let for x and y two distinct points of E^m*

$$A(x,y) = \inf\Big\{ \max_{1\le k<|p|} W(p_k) + V(p_k,p_{k+1}) : p \in D^m(x,y) \Big\}.$$

For x in E^m, put $A(x,x) = W(x)$.
Then $A = A_1$.

Proof. We have $W = W_1$ and $V \le V_1$: clearly $A \le A_1$.
Conversely, let p belong to $E^{m(\mathbb{N})}(x,y)$. For each k, $1 \le k < |p|$, let $p_1^k \to \cdots \to p_{n_k}^k$ be a path in E^m joining p_k and p_{k+1} such that

$$V(p_k,p_{k+1}) = V_1(p_1^k,p_2^k) + \cdots + V_1(p_{n_k-1}^k,p_{n_k}^k).$$

Consider the path $\bar p$ obtained by joining end to end all these paths:

$$p_1^1 \to \cdots \to p_{n_1-1}^1 \to p_1^2 \to \cdots \to p_{n_2-1}^2 \to \cdots \to p_1^{|p|-1} \to \cdots \to p_{n_{|p|-1}-1}^{|p|-1} \to p_{n_{|p|-1}}^{|p|-1}.$$

Let k and h be two integers such that $1 \le k < |p|$, $1 \le h < n_k$. We have

$$\begin{aligned} W_1(p_h^k) + V_1(p_h^k,p_{h+1}^k) &\le W(p_k) + V(p_k,p_h^k) + V_1(p_h^k,p_{h+1}^k) + V(p_{h+1}^k,p_{k+1}) \\ &= W(p_k) + V(p_k,p_{k+1}) \end{aligned}$$

whence

$$\max_{1\le k<|p|}\ \max_{1\le h<n_k} W_1(p_h^k) + V_1(p_h^k,p_{h+1}^k) \le \max_{1\le k<|p|} W(p_k) + V(p_k,p_{k+1}).$$

However, the left–hand side member of the above inequality coincides with

$$\max_{1\le k<|\bar p|} W_1(\bar p_k) + V_1(\bar p_k,\bar p_{k+1}).$$

Taking the minimum over all paths of $D^m(x,y)$, we obtain $A_1(x,y) \le A(x,y)$. □

Suppose we wish only to examine the trace of the dynamics on a subset H of E^m. We build a new communication cost V_H on H by making the set $E^m \setminus H$ a "taboo" set. Once more, the following definitions and results are not specific to mutation–selection algorithms.

Definition 5.7. Let H be a subset of E^m. We define a communication cost V_H by

$$V_H(x,y) = \inf\{\, V_1(p) : p \in D^m(x,y),\ \forall k \quad 1 < k < |p|,\ p_k \notin H\,\}.$$

(for the definition of $V_1(p)$, see notation 5.1).
We define a virtual energy W_H on H by

$$\forall x \in H \qquad W_H(x) = \min\{\, V_H(g) : g \in G_H(x)\,\}$$

where $G_H(x)$ is the set of x–graphs over H and the V_H–cost of a graph g over H is

$$V_H(g) = \sum_{(x\to y)\in g} V_H(x,y).$$

Finally, we define a communication altitude A_H: if x and y are two distinct points of H,

$$A_H(x,y) = \inf\Big\{ \max_{1\le k<|p|} W_H(p_k) + V_H(p_k,p_{k+1}) : p \in H^{(\mathbb{N})}(x,y) \Big\}.$$

For any x in H, we put $A_H(x,x) = W_H(x)$.

Theorem 5.8. *Let H be a subset of E^m such that*

$$\forall x \in E^m \quad \exists y \in H \quad V(x,y) = 0.$$

Then $W_H = W$ and $A_H = A$ on the set H.

Proof. Let g belong to $G_H(x)$, where $x \in H$. By hypothesis, for each y in $E^m \setminus H$, there exists z in H such that $V(y,z) = 0$. Let $\bar g$ be an x–graph over E^m obtained by adding one such arrow for each point of $E^m \setminus H$. We have $V(\bar g) = V(g) \le V_H(g)$ whence, by taking the minimum over all x–graphs, $W(x) \le W_H(x)$.
Conversely, let g belong to $G(x)$, where x is in H, and consider the graph $\bar g$ of $G_H(x)$ defined by: the arrow $(y \to z)$ is in $\bar g$ if and only if y and z are in H and there exists a path $x^1 = y \to x^2 \to \cdots \to x^{r-1} \to x^r = z$ in E^m such that

$$\forall k \in \{1\cdots r-1\} \quad (x^k \to x^{k+1}) \in g, \qquad \forall k \in \{2\cdots r-1\} \quad x^k \in E^m \setminus H.$$

We have $V_H(\bar g) \le V_1(g)$. It follows that $W_H(x) \le W_1(x) = W(x)$ and the first equality $W_H = W$ is proved.

Let p belong to $H^{(\mathbb{N})}(x,y)$, where x,y are elements of H. For each k, $1 \le k < |p|$, let $p^k_1 \to \cdots \to p^k_{n_k}$ be a path in E^m such that $p^k_1 = p_k$, $p^k_{n_k} = p_{k+1}$ and

$$V_H(p_k,p_{k+1}) = V_1(p^k_1,p^k_2) + \cdots + V_1(p^k_{n_k-1},p^k_{n_k}).$$

Consider the path $\bar{p}$ obtained by joining end to end all these paths:

$$p_1^1 \to \cdots \to p_{n_1-1}^1 \to p_1^2 \to \cdots \to p_{n_2-1}^2 \to \cdots \to p_1^{|p|-1} \to \cdots \to p_{n_{|p|-1}-1}^{|p|-1} \to p_{n_{|p|-1}}^{|p|-1}.$$

Let k and h be two integers such that $1 \le k < |p|$, $1 \le h < n_k$. We have

$$\begin{aligned} W_1(p_h^k) + V_1(p_h^k, p_{h+1}^k) &\le W_H(p_k) + V_1(p_k, p_2^k) + \cdots + V_1(p_{h-1}^k, p_h^k) + V_1(p_h^k, p_{h+1}^k) \\ &\le W_H(p_k) + V_H(p_k, p_{k+1}) \end{aligned}$$

whence

$$\max_{1 \le k < |p|} \max_{1 \le h < n_k} W_1(p_h^k) + V_1(p_h^k, p_{h+1}^k) \le \max_{1 \le k < |p|} W_H(p_k) + V_H(p_k, p_{k+1}).$$

However, the left–hand side member of the above inequality coincides with

$$\max_{1 \le k < |\bar{p}|} W_1(\bar{p}_k) + V_1(\bar{p}_k, \bar{p}_{k+1}).$$

Taking the minimum over all paths of $H^{(\mathbb{N})}(x,y)$, we obtain $A_1(x,y) = A(x,y) \le A_H(x,y)$. The proof of the reverse inequality will use the following little lemma.

Lemma 5.9. *Let x belong to E^m. There exists a graph g in $G(x)$ such that $V(g) = W(x)$ and for each arrow $(y \to z)$ of g we have either $\big(y \in E^m \setminus H,\ z \in H,\ V(y,z) = 0\big)$ or $\big(y \in H,\ z \in H \cup \{x\}\big)$. In particular, no arrow ends in a population of $E^m \setminus \{H \cup \{x\}\}$.*

Remark. This lemma is a slight improvement of the first part of Lemma 4.3 of [4, chapter 6].

Proof. Let g be an element of $G(x)$ such that $V(g) = W(x)$. First, we proceed as in the proof of Lemma 4.3 of [4, chapter 6] to get rid of all the arrows from $E^m \setminus H$ to $E^m \setminus H$. Let $(y \to z)$ be such an arrow. Let z' be in H such that $V(y,z') = 0$. We replace the arrow $(y \to z)$ by $(y \to z')$. If a cycle $y \to x^1 \to \cdots \to x^r \to y$ is formed, we replace the arrow $(x^r \to y)$ by $(x^r \to z)$. Since $V(x^r, z) \le V(x^r, y) + V(y,z)$ the cost of the graph does not increase. We obtain a graph g in $G(x)$ such that $V(g) = W(x)$ and all arrows starting from $E^m \setminus H$ end in H and have a null cost. We now remove the arrows from H to $E^m \setminus \{H \cup \{x\}\}$. Let $(y \to z)$ be such an arrow. Let z' be the unique element of H such that $(z \to z') \in g$. We replace the arrow $(y \to z)$ by $(y \to z')$. This operation does not increase the cost of the graph since $V(y,z') \le V(y,z) + V(z,z') = V(y,z)$. The resulting graph has the desired properties. □

Now let p belong to $E^{m(\mathbb{N})}(x,y)$.
Let k belong to $\{1 \cdots |p|\}$. If p_k is in H, we put $p_1^k = p_k$ and $r_k = 1$.
Suppose p_k is not in H and let p_1^k be an element of H such that $V(p_k, p_1^k) = 0$ and there exist

populations $p_1^\circ, \cdots, p_r^\circ$ in $E^m \setminus H$ such that $V_1(p_k, p_1^\circ) = V_1(p_1^\circ, p_2^\circ) = \cdots = V_1(p_r^\circ, p_1^k) = 0$.
Let g belong to $G(p_k)$ be as in lemma 5.9. We add the arrow $(p_k \to p_1^k)$ to g.
Let $\bar{g}$ be the graph on H defined by: the arrow $(y \to z)$ is in $\bar{g}$ if and only if y and z are in H and there exists a path $x^1 = y \to x^2 \to \cdots \to x^{r-1} \to x^r = z$ in E^m such that

$$\forall k \in \{1 \cdots r-1\} \quad (x^k \to x^{k+1}) \in g, \qquad \forall k \in \{2 \cdots r-1\} \quad x^k \in E^m \setminus H.$$

(In particular, we suppress from g all the arrows $(y \to z), y \in E^m \setminus \{H \cup \{p_k\}\}$).
We have $V_H(\bar{g}) \leq V(g) = W(p_k)$.
Moreover, the graph $\bar{g}$ contains at most one cycle $p_1^k \to p_2^k \to \cdots \to p_{r_k}^k \to p_1^k$ (where the populations p_h^k, $2 \leq h \leq r_k$, belong to H). This cycle has been created by the addition of the arrow $(p_k \to p_1^k)$ in g which causes the arrow $(p_{r_k}^k \to p_1^k)$ to appear in $\bar{g}$. Necessarily this arrow $(p_{r_k}^k \to p_1^k)$ comes from the sequence $p_{r_k}^k \to p_k \to p_1^k$ in g (recall that no arrow of g ends in $E^m \setminus \{H \cup \{p_k\}\}$) so that the arrow $(p_{r_k}^k \to p_k)$ is present in g. If we now remove this arrow from g we obtain an element of $G(p_{r_k}^k)$ whence $V(g) - V(p_{r_k}^k, p_k) \geq W(p_{r_k}^k)$ or $W(p_{r_k}^k) + V(p_{r_k}^k, p_k) \leq V(g) = W(p_k)$ and it follows that $W(p_k) = W(p_{r_k}^k) + V(p_{r_k}^k, p_k)$.
If there is no cycle, we put $r_k = 1$ and the preceding equality is still valid.
Removing the arrow $(p_h^k \to p_{h+1}^k)$ from $\bar{g}$ $(1 \leq h < r_k)$ gives an element of $G_H(p_h^k)$ so that we have the inequalities

$$(3) \qquad \forall h \in \{1 \cdots r_k - 1\} \qquad W_H(p_h^k) + V_H(p_h^k, p_{h+1}^k) \leq V_H(\bar{g}).$$

Yet, there exists a sequence of populations $p_{r_k+1}^k, \cdots, p_{n_k}^k$ in H such that

$$V(p_k, p_{k+1}) = V(p_k, p_{r_k+1}^k) + V(p_{r_k+1}^k, p_{r_{k+2}}^k) + \cdots + V(p_{n_k-1}^k, p_{n_k}^k) + V(p_{n_k}^k, p_{k+1})$$

and each cost appearing in the right–hand side is realized by a path in $E^m \setminus H$. That is

$$\forall h \quad r_k + 1 \leq h \leq n_k - 1 \qquad V(p_h^k, p_{h+1}^k) = V_H(p_h^k, p_{h+1}^k)$$

and there exist populations $p_1^-, \cdots, p_s^-, p_1^+, \cdots, p_t^+$ of $E^m \setminus H$ such that

$$\begin{aligned} V(p_k, p_{r_k+1}^k) &= V_1(p_k, p_1^-) + V_1(p_1^-, p_2^-) + \cdots + V_1(p_s^-, p_{r_k+1}^k), \\ V(p_{n_k}^k, p_{k+1}) &= V_1(p_{n_k}^k, p_1^+) + V_1(p_1^+, p_2^+) + \cdots + V_1(p_t^+, p_{k+1}). \end{aligned}$$

To obtain such a sequence $p_{r_k+1}^k, \cdots, p_{n_k}^k$, we just take the successive populations of H which appear in a path of the set $D^{m*}(p_k, p_{k+1})$ (see definition 5.2). Let $\bar{p}$ be the path obtained by joining end to end all these paths in the following way:

$$p_1^1, \cdots, p_{r_1}^1, \cdots, p_{n_1}^1, p_1^2, \cdots, p_{r_2}^2, \cdots, p_{n_2}^2, \cdots, p_1^{|p|-1}, \cdots, p_{r_{|p|-1}}^{|p|-1}, \cdots, p_{n_{|p|-1}}^{|p|-1}, p_1^{|p|}.$$

For each k in $\{1\cdots|p|-1\}$, we have

$$\forall h \qquad 1\le h<r_k \qquad W_H(p_h^k)+V_H(p_h^k,p_{h+1}^k)\le W(p_k), \qquad \text{(by inequality (3))}$$

$$\forall h \qquad r_k\le h\le n_k \qquad W_H(p_h^k)+V_H(p_h^k,p_{h+1}^k)\le$$

$$W(p_{r_k}^k)+V(p_{r_k}^k,p_k)+V(p_k,p_{r_k+1}^k)+V(p_{r_k+1}^k,p_{r_k+2}^k)+\cdots+V(p_{n_k-1}^k,p_{n_k}^k)+V(p_{n_k}^k,p_{k+1})$$

$$=W(p_k)+V(p_k,p_{k+1}).$$

Thus

$$\max_{1\le k<|p|}\ \max_{1\le h\le n_k}\ W_H(p_h^k)+V_H(p_h^k,p_{h+1}^k)\le \max_{1\le k<|p|}\ W(p_k)+V(p_k,p_{k+1})$$

(we put $p_{n_k+1}^k=p_1^{k+1}$ for k in $\{1\cdots|p|-1\}$).
However the left–hand side member of the above inequality is exactly

$$\max_{1\le k<|\bar p|}\ W_H(\bar p_k)+V_H(\bar p_k,\bar p_{k+1}).$$

Taking the minimum over all paths of $E^{m(\mathbb{N})}(x,y)$ yields $A_H(x,y)\le A(x,y)$. $\square$

Coming back to the mutation–selection algorithm, we see that the set U of the uniform populations verify

$$\forall x\in E^m \quad \exists y\in U \qquad V(x,y)=0$$

(more precisely: $\forall x\in E^m \quad \forall i\in\widehat{x} \quad V(x,(i))=0$) so that the preceding results may be applied to the set U. Our next task is to study the cost function V_U, or equivalently (by propositions 5.4 and 5.6 and theorem 5.8), the restriction of V to U, as a function of m. We will often omit the parenthesis when speaking of uniform populations: for instance $V(i,j)$ will stand for $V((i),(j))$. The crucial result is that, for m sufficiently large,

6. $V(i,j)$ is an affine function of m

Before proceeding to the proof of the main theorem 6.12, we give several notations and definitions. We will consider paths in the sets E,E^m and $\mathcal{P}(E)$. Paths of E^m will mostly be denoted by the letter p and paths of $\mathcal{P}(E)$ by the letter q.

Definition 6.1. We define a bracket operator [ ] from $E^{(\mathbb{N})}=\bigcup_{m\in\mathbb{N}}E^m$, the set of all finite sequences of elements of E, onto $\mathcal{P}(E)$, the set of all subsets of E, by

$$x=(x_1,\cdots,x_m)\in E^m \ \longmapsto\ [x]=\{\,x_k:1\le k\le m\,\}$$

i.e. $[x]$ is the set of all individuals present in the population x. The bracket operator [ ] provides a natural projection from $\cup_{m\in\mathbb{N}^*}E^{m(\mathbb{N})}$ (the set of all finite sequences of elements of E^m) onto $\mathcal{P}(E)^{(\mathbb{N})}$: with each path $p=(p_1,\cdots,p_r)$ in E^m we associate the path $[p]=([p_1],\cdots,[p_r])$ in $\mathcal{P}(E)$.

From now on, the population size m will vary and we will sometimes consider simultaneously several paths, possibly with different population sizes. Thus, if p is a path in E^m, we put $m(p)=m$.

Notation 6.2. By $\overline{D}^m$ we denote the paths in E^m which correspond to possible trajectories for the whole process

$$X_n^l \to Y_n^l \to X_{n+1}^l \to Y_{n+1}^l \to \cdots \to Y_{n+t-1}^l \to X_{n+t}^l$$

i.e. such a path p includes the intermediate populations Y_n^l, has an odd length and satisfies

$$\forall k \quad 1 \leq 2k < |p| \qquad \alpha(p^{2k-1}, p^{2k})\,\gamma(p^{2k}, p^{2k+1}) > 0.$$

The corresponding cost function $\overline{V}$ is defined by

$$\overline{V}(p) = \sum_{1 \leq 2k < |p|} \Big(a\, d(p^{2k-1}, p^{2k}) + c \sum_{h=1}^{m} \big(\widehat{f}(p^{2k}) - f(p_h^{2k+1}) \big) \Big)$$

if the path p belongs to $\overline{D}^m$ (here p_h^{2k+1} is the h–th component of the vector p^{2k+1}) and $\overline{V}(p) = \infty$ otherwise (we recall that $d(x, y)$ is the Hamming distance between x and y). We put also for y, z in E^m

$$\overline{D}^m(y, z) = \overline{D}^m \cap E^{m(\mathbb{N})}(y, z).$$

We denote by $\overline{D}^{m*}(y, z)$ the elements p of $\overline{D}^m(y, z)$ such that $\overline{V}(p) = V(y, z)$.

Definition 6.3. We define two cost functions $[V]$ and $[\overline{V}]$ on $\mathcal{P}(E)^{(\mathbb{N})}$; for $q \in \mathcal{P}(E)^{(\mathbb{N})}$,

$$[V](q) = \inf \Big\{ V_1(p) : p \in \bigcup_{m \in \mathbb{N}^*} E^{m(\mathbb{N})},\ [p] = q \Big\}$$
$$[\overline{V}](q) = \inf \Big\{ \overline{V}(p) : p \in \bigcup_{m \in \mathbb{N}^*} E^{m(\mathbb{N})},\ [p] = q \Big\}.$$

(for the definitions of V_1 and $\overline{V}$, see notations 5.1 and 6.2).

Notation 6.4. We put for i, j in E

$$\overline{D}(i, j) = \bigcup_{m \in \mathbb{N}^*} \overline{D}^m((i), (j)), \quad \mathcal{P}(i, j) = \{ [p] : p \in \overline{D}(i, j) \}$$
$$\overline{D} = \bigcup_{m \in \mathbb{N}^*} \overline{D}^m, \quad [\overline{D}] = \{ [p] : p \in \overline{D} \}.$$

Definition 6.5. Let q be a path in $\mathcal{P}(E)$. We say that the path $e_1 \to \cdots \to e_r$ in E is admissible for q if $r = |q|$, $\forall k \in \{1 \cdots |q|\} \quad e_k \in q_k$ and

$$\forall k \quad 1 \leq 2k < |q| \qquad \text{either} \quad e_{2k-1} = e_{2k} \quad \text{or} \quad \alpha(e_{2k-1}, e_{2k}) > 0.$$

The set of all admissible paths for q is denoted by $\mathcal{A}(q)$. For instance, if p belongs to $\overline{D}$, for each h in $\{1 \cdots m(p)\}$, the path $p_h^1 \to \cdots \to p_h^{|p|}$ is admissible for $[p]$ (here p_h^k denotes the h–th component of the vector p^k).

Notation 6.6. We define a function Ω on the set $\mathcal{P}(E)^{(\mathbb{N})}$ by: if $q \in \mathcal{P}(E)^{(\mathbb{N})} \setminus [\overline{D}]$, then $\Omega(q) = \infty$; if $q \in [\overline{D}]$, then we put

$$\Omega(q) = \min \Big\{ \sum_{1 \leq 2k < |q|} a(1 - \delta(e_{2k-1}, e_{2k})) + c\big(\widehat{f}(q_{2k}) - f(e_{2k+1})\big) : (e_k)_{1 \leq k \leq |q|} \in \mathcal{A}(q) \Big\}$$

(we recall that δ is the Kronecker symbol).
We use the function Ω defined on $\mathcal{P}(E)^{(\mathbb{N})}$ to build a function Ω on $E \times E$:

$$\forall i, j \in E \qquad \Omega(i,j) = \inf \big\{ \Omega(q) : q \in \mathcal{P}(i,j) \big\}.$$

Since the set $\mathcal{P}(i,j)$ is never empty and is included in $[\overline{D}]$ (see notation 6.4), the quantity $\Omega(i,j)$ is finite. We denote by $\mathcal{P}^*(i,j)$ the elements of $\mathcal{P}(i,j)$ realizing the above minimum.

Let p be a path in E^m and $\tilde{p}$ a path in $E^{\tilde{m}}$.
We say that $\tilde{p}$ is included in p (noted $\tilde{p} \subset p$) if $|p| = |\tilde{p}|$ and

$$\forall k \in \{1 \cdots |p|\} \quad \forall i \in E \qquad \tilde{p}^k(i) \leq p^k(i)$$

$$\forall k, 1 < 2k < |p| \qquad \{ (p_h^{2k-1}, p_h^{2k}) : 1 \leq h \leq m(p) \} \;=\; \{ (\tilde{p}_h^{2k-1}, \tilde{p}_h^{2k}) : 1 \leq h \leq m(\tilde{p}) \}$$

(we recall that $p^k(i)$ is the number of occurrences of i in the population p^k).
A path $\tilde{p}$ included in p may be obtained from the path p by destroying some individuals in each population of the path in such a way that the set of mutations occurring at odd times is preserved.
Let q be a path in $\mathcal{P}(E)$. Among the paths p such that $[p] = q$, we single out the paths which are minimal with respect to the above inclusion relation:

Definition 6.7. The path p in $\overline{D}^m$ is a minimal path realizing the path q in $\mathcal{P}(E)$ if $[p] = q$ and for each $\tilde{m}$ in $\mathbb{N}^*$ and each path $\tilde{p}$ in $\overline{D}^{\tilde{m}}$, we have the implication

$$[\tilde{p}] = q \quad \text{and} \quad \tilde{p} \subset p \qquad \Longrightarrow \qquad m = \tilde{m}$$

(whence also $\forall k \in \{1 \cdots |p|\} \quad \forall i \in E \quad \tilde{p}^k(i) = p^k(i)$).
The path p in $\overline{D}^m$ is minimal if it is a minimal path realizing the path $[p]$.

This definition has the following interpretation: the path p is minimal if we can't destroy a fixed number of individuals in each population of p and still have a path belonging to $\overline{D}$ without altering $[p]$ and the set of mutations occurring in p.
We have the following upper bound for the population sizes of minimal paths:

Lemma 6.8. *Let p be a minimal path realizing the path q in $\mathcal{P}(E)$. Then*

$$m(p) \quad \leq \quad \max_{1\leq 2k<|q|} (\mathrm{card}\, q_{2k-1})(\mathrm{card}\, q_{2k}) \quad \leq \quad (\mathrm{card}\, E)^2.$$

Proof. Suppose

$$\forall k \quad 1 \leq 2k < |q| \qquad m(p) \;>\; (\mathrm{card}\, q_{2k-1})(\mathrm{card}\, q_{2k}).$$

For each k, $1 \leq 2k < |q|$, consider the pairs $(p_h^{2k-1}, p_h^{2k})_{1\leq h\leq m(p)}$: they belong to the set $q_{2k-1} \times q_{2k}$ whose cardinality is strictly less than $m(p)$; necessarily at least two pairs are identical. We choose an index $h(k)$ such that the pair $(p_{h(k)}^{2k-1}, p_{h(k)}^{2k})$ is present twice. Let $\tilde{p}$ be a path obtained in the following way:

- for each k, $1 \leq 2k < |p|$, we remove the individuals $p_{h(k)}^{2k-1}$ and $p_{h(k)}^{2k}$ from p^{2k-1} and p^{2k} to obtain the populations $\tilde{p}^{2k-1}$ and $\tilde{p}^{2k}$;
- to build $\tilde{p}_{|q|}$, we remove from the population $p_{|q|}$ an individual which is present twice (such an individual necessarily exists since $[p_{|q|}] \subset [p_{|q|-1}]$ and the individuals of $p_{|q|-1}$ are not all distinct).

Finally we have for this path $\tilde{p}$

$$\tilde{p} \in \overline{D}, \quad m(\tilde{p}) = m(p) - 1, \quad [\tilde{p}] = q \quad \text{and} \quad \tilde{p} \subset p,$$

thus contradicting the minimality of p. $\square$

Definition 6.9. We define a function θ on $\mathcal{P}(E)^{(\mathbb{N})}$ which associates with each path q the value

$$\theta(q) = \min \big\{ \overline{V}(p) - m(p)\,\Omega([p]) : p \text{ minimal path realizing } q \big\}.$$

By the very definition of Ω, all paths p satisfy the inequality $\overline{V}(p) \geq m(p)\,\Omega([p])$ so that the quantity $\theta(q)$ is non–negative. More precisely, we have the following

Lemma 6.10. *For all paths p belonging to $\overline{D}$,*

$$\overline{V}(p) \quad \geq \quad \theta([p]) + m(p)\,\Omega([p]).$$

Proof. Let p be a path of $\overline{D}$. Necessarily, there exists a minimal path $\tilde{p}$ included in p which realizes the path $[p]$. We may rearrange the individuals of the populations $p^1, \cdots, p^{|p|}$ so that $\tilde{p}$ appears as the trajectories of the first $m(\tilde{p})$ components of the path p, that is $\tilde{p}^k = (p_1^k, \cdots, p_{m(\tilde{p})}^k)$ for each k in $\{1 \cdots |p|\}$. The condition that the set of mutations occurring at odd times is preserved in the minimal path is essential for this operation to be possible. Although this reordering might be quite complex, it does not affect the cost of the path p. Yet

$$\begin{aligned}\overline{V}(p) &= \sum_{h=1}^{m(p)} \sum_{1\le 2k<|p|} a\big(1-\delta(p_h^{2k-1},p_h^{2k})\big)+c\big(\widehat{f}(p^{2k})-f(p_h^{2k+1})\big)\\ &= \overline{V}(\tilde{p}) \;+\; \sum_{h=m(\tilde{p})+1}^{m(p)} \sum_{1\le 2k<|p|} a\big(1-\delta(p_h^{2k-1},p_h^{2k})\big)+c\big(\widehat{f}(p^{2k})-f(p_h^{2k+1})\big).\end{aligned}$$

By the very definition of Ω, we have for all h in $\{1\cdots m(p)\}$

$$\Omega([p]) \;\le\; \sum_{1\le 2k<|p|} a\big(1-\delta(p_h^{2k-1},p_h^{2k})\big)+c\big(\widehat{f}(p^{2k})-f(p_h^{2k+1})\big)$$

(remark that since p belongs to $\overline{D}$, the path $p_h^1\to\cdots\to p_h^{|p|}$ is an admissible path for $[p]$). It follows that

$$\overline{V}(p) \;\ge\; \overline{V}(\tilde{p})+(m(p)-m(\tilde{p}))\,\Omega([p]) \;=\; \overline{V}(\tilde{p})-m(\tilde{p})\,\Omega([\tilde{p}])+m(p)\Omega([p]).$$

Now $\tilde{p}$ is a minimal path realizing $[p]$, whence $\overline{V}(\tilde{p})-m(\tilde{p})\,\Omega([\tilde{p}])\ge\theta([p])$ and finally

$$\overline{V}(p)\ge\theta([p])+m(p)\Omega([p]). \quad \square$$

Definition 6.11. We define for i,j in E

$$\theta(i,j)=\inf\big\{\,\theta(q) : q\in\mathcal{P}^*(i,j)\,\big\}$$

and

$$m(i,j)=\min\big\{\,m(p) : p\in\overline{D},\ [p]\in\mathcal{P}^*(i,j),\ p\text{ minimal},\ \overline{V}(p)-m(p)\Omega([p])=\theta(i,j)\,\big\}.$$

(for the definition of $\mathcal{P}^*(i,j)$, see notation 6.6).
Notice that $m(i,j)$, being the population size of a minimal path, is smaller than $|E|^2$.

We are now in position to state our fundamental

Theorem 6.12. *For all pair (i,j) of points of E, there exists an integer M such that*

$$\forall m\ge M \qquad V(i,j) \;=\; \theta(i,j) \;+\; m\,\Omega(i,j).$$

Proof. Let i, j be two points of E. Let m be an integer greater than $m(i,j)$ and let p be an element of $\overline{D}^{m(i,j)}$ satisfying: $[p] \in \mathcal{P}^*(i,j)$, p is minimal, $\overline{V}(p) - m(p)\,\Omega([p]) = \theta(i,j)$ (see notations 6.2, 6.6, 6.11). Let $e_1^* \to \cdots \to e_{|p|}^*$ be an admissible path for $[p]$ such that

$$\Omega([p]) \quad = \sum_{1 \le 2k < |q|} a(1 - \delta(e_{2k-1}^*, e_{2k}^*)) + c\big(\widehat{f}(p_{2k}) - f(e_{2k+1}^*)\big)$$

i.e. this path realizes the minimum defining $\Omega([p])$.
Consider the path $\bar{p}$ in E^m defined by

$$\forall k \in \{1 \cdots |p|\} \qquad \bar{p}^k = (\ \underbrace{p^k}_{E^{m(i,j)}},\ \underbrace{e_k^*,\ \cdots,\ e_k^*}_{E^{m-m(i,j)}}\ )$$

obtained by adding $m - m(i,j)$ copies of the path $(e_k^*)_{1 \le k \le |p|}$ to the path p. We have

$$\begin{aligned}
\overline{V}(\bar{p}) &= \overline{V}(p) + (m - m(i,j))\,\Omega(i,j) \\
&= \overline{V}(p) - m(p)\,\Omega([p]) + m\,\Omega(i,j) \\
&= \theta(i,j) + m\,\Omega(i,j)
\end{aligned}$$

whence clearly

$$\forall m \ge m(i,j) \qquad V(i,j) \ \le\ \theta(i,j) + m\,\Omega(i,j).$$

Conversely, put

$$M(i,j) \quad = \quad \sup \left\{ \frac{\theta(i,j) - \theta(q)}{\Omega(q) - \Omega(i,j)} : q \in \mathcal{P}(i,j),\ \Omega(q) \neq \Omega(i,j) \right\}.$$

By the very definition of Ω, which is a sum of terms of the form a and $c\,|f(j_1) - f(j_2)|$, we have the implication

$$q \in \mathcal{P}(i,j),\ \Omega(q) \neq \Omega(i,j) \quad \Longrightarrow \quad \Omega(q) \ge \Omega(i,j) + \min(a, c\delta)$$

(where $\delta = \min\{\,|f(i) - f(j)| : i, j \in E, f(i) \neq f(j)\,\}$).
Since in addition $\theta(q)$ is non–negative, we see that

$$M(i,j) \ \le\ \frac{\theta(i,j)}{\min(a, c\delta)} \ <\ \infty$$

so that $M(i,j)$ is finite (and independent of m: neither $\mathcal{P}(i,j)$, nor θ, nor Ω depend on m).

Let m be an integer such that $m \geq m(i,j)$ and $m > M(i,j)$. Let p belong to $\overline{D}^m(i,j)$. Suppose $\Omega([p]) \neq \Omega(i,j)$. Then, by definition of $M(i,j)$, we have

$$\theta(i,j) + m\,\Omega(i,j) \;<\; \theta([p]) + m\,\Omega([p]).$$

However we know already (by lemma 6.10) that $\overline{V}(p) \geq \theta([p]) + m\,\Omega([p])$ and (by the first part of the proof) $V(i,j) \leq \theta(i,j) + m\,\Omega(i,j)$, whence $\overline{V}(p) > V(i,j)$ and p does not belong to $\overline{D}^{m*}(i,j)$ (see notation 6.2).
Henceforth each path p of $\overline{D}^{m*}(i,j)$ satisfies $\Omega([p]) = \Omega(i,j)$, or equivalently $[p] \in \mathcal{P}^*(i,j)$. For such a path p, we have also $\theta([p]) \geq \theta(i,j)$ (by the very definition of $\theta(i,j)$, see definition 6.11). Application of lemma 6.10 yields the desired inequality $\overline{V}(p) \geq \theta(i,j) + m\,\Omega(i,j)$ from which we deduce

$$V(i,j) \;\geq\; \theta(i,j) + m\,\Omega(i,j). \quad \square$$

Remark 1. This theorem may be interpreted in the following way: for $m \geq M$, each new individual added to the population follows the ideal path $e_1^* \to \cdots \to e_{|p|}^*$ which realizes the minimal cost $\Omega(i,j)$. However, it is necessary to have at least $m(i,j)$ individuals for such a path to become possible i.e. the first $m(i,j)$ individuals are used to build the path p having an admissible path realizing the minimal cost $\Omega(i,j)$.

Remark 2. Another way to understand this result is the following: to travel between the uniform populations (i) and (j), some events of positive cost must necessarily take place i.e. specific sequences of mutations and selections of individuals below peak fitness (these events are called anti–selections). We distinguish informally two kinds of such events:

- the local events which affect only a limited number of individuals,
- the collective events which affect almost all the individuals.

As the reader would have guessed, $\theta(i,j)$ corresponds to the cost of the local events whereas $\Omega(i,j)$ corresponds to the cost of the collective events.

Remark 3. We have no useful practical information about the values $m(i,j)$ and $M(i,j)$: they depend strongly on the structure of the fitness landscape (E, f, α). Although they may be very large in theory, it is likely that in most cases they will be "reasonably" small: for instance if (i) and (j) are very close (i.e. the points i and j may be joined through the kernel α with a short path) or if $\Omega(i,j)$ is null.

Remark 4. We have also $V(i,j) \geq m\,\Omega(i,j)$ for each m in $\mathbb{N}$. This inequality is an improvement of lemma 12.1 of [3] which was the key to prove the existence of the critical population size m^*. In particular, it could be used to obtain better bounds on m^*. However, in this paper, we deal only with large values of m.

Remark 5. Of course, the coefficients $\Omega(i,j)$ and $\theta(i,j)$ are of crucial interest. We will derive numerous properties of these coefficients in the sequel.

We know that for m sufficiently large, the cost function $V(i,j)$ is affine. All the important quantities concerning our algorithm are defined through V as the minimum or the maximum of a set of finite sums involving V. The following elementary result shows immediately that this procedure yields also functions which become affine when m is large.

Lemma 6.13. *Let $(y_i)_{i\in I}$ be a finite family of affine functions. Put*

$$y(t) \;=\; \max\big\{\, y_i(t) : i \in I \,\big\} \quad \textit{for } t \textit{ in } \mathbb{R}.$$

Then for t sufficiently large, y is affine and coincides with one of the functions y_i i.e.

$$\exists T \in \mathbb{R} \quad \exists i \in I \quad \forall t \geq T \qquad y(t) = y_i(t).$$

Corollary 6.14. *The virtual energy restricted to U is an affine function of m for m sufficiently large:*

$$\forall i \in E \quad \exists M(i) \quad \forall m \geq M(i) \qquad W(i) = \theta(i) + m\Omega(i)$$

where

$$\Omega(i) = \inf\Big\{ \sum_{(j_1 \to j_2)\in g} \Omega(j_1,j_2) : g \in G_U(i) \Big\}$$

and

$$\theta(i) = \inf\Big\{ \sum_{(j_1 \to j_2)\in g} \theta(j_1,j_2) : g \in G_U(i)\,, \sum_{(j_1 \to j_2)\in g} \Omega(j_1,j_2) = \Omega(i) \Big\}.$$

Corollary 6.15. *The communication altitude restricted to U is an affine function of m for m sufficiently large:*

$$\forall i,j \in E \quad \exists \rho(i,j) \quad \exists \psi(i,j) \quad \exists M \quad \forall m \geq M \qquad A(i,j) = \rho(i,j) + m\psi(i,j).$$

7. The limiting decomposition in cycles

When the random perturbations are small, the dynamics of the process is well described by the decomposition of the space into cycles (this notion was originally introduced by Freidlin and Wentzell). Let us try to give a brief survey of these objects. Suppose the process starts from the uniform population (e_1). It leaves (e_1) after a finite amount of time. Among all the exit trajectories, there exists one trajectory which is the most probable one, which leads to another uniform population, for instance (e_2). Again, from (e_2), the process goes to (e_3). The set U of uniform populations being finite, the process one day visits a uniform population twice. For instance, from (e_3), it returns to (e_1). We obtain then a cycle, $(e_1) \to (e_2) \to (e_3) \to (e_1)$, and the process "cycles" over it a very long time. Put now these three populations in a box. Again, the perturbations will force the process to leave this box, and once more, there exists a canonical path of exit which leads the process to another uniform population, or more generally, to another cycle. Since there is a finite number of such cycles, the process visits in the end a cycle twice: we obtain then a "cycle of cycles" into which the process remains trapped for a very very long time. Going on this way, it is possible to build a whole hierarchy of cycles which exhausts the set of uniform populations and yields a very accurate picture of the asymptotic dynamics of the process.

The good tool to perform this hierarchical decomposition of the space E^m into cycles is the communication altitude. For the construction of cycles and the various related quantities, we refer the reader to [9,10]. Nevertheless, our notion of cycles differs slightly from Catoni and Trouvé's one. Whereas they consider as a cycle any set from which the process can't escape but without having forgotten its entrance point (when the perturbations are small), we merely consider the cycles as the most attractive and stable sets we can build for the perturbed process. We thus eliminate some cycles appearing in Trouvé's work: not all singletons are cycles (only those which are local minima of the virtual energy) and our cycles are all strict in the sense of Trouvé [9, Definition 3.20]. As a consequence, the unperturbed process will never leave a cycle. We believe that this point of view is closer to the initial notion introduced by Freidlin and Wentzell.

Definition 7.1. Let $\lambda \in \mathbb{R}$. On the set

$$W_\lambda = \{\, x \in E^m : W(x) \leq \lambda \,\}$$

we define an equivalence relation $\mathcal{R}_\lambda$ by

$$\forall y, z \in W_\lambda, \qquad y \,\mathcal{R}_\lambda\, z \quad \Longleftrightarrow \quad A(y,z) \leq \lambda.$$

Proposition 7.2. *(Trouvé,* [9, Proposition 3.21]*)*
The set of cycles in E^m associated with the cost function V_1 and the kernel q_1 is

$$\mathcal{C}(E^m) \;=\; \bigcup_{\lambda \in \mathbb{R}^+} W_\lambda \,/\, \mathcal{R}_\lambda$$

where $W_\lambda \,/\, \mathcal{R}_\lambda$ is the quotient set of the equivalence classes of W_λ for the relation $\mathcal{R}_\lambda$.

Unfortunately, the set of cycles $\mathcal{C}(E^m)$ increases dramatically with m. Since we are mostly interested in the way the mutation–selection algorithm visits the set U of uniform populations (which are the attractors of the unperturbed process: the bottom of the cycles contain only uniform populations), we will study the projection of the cycles on U.

We define a projection T_U from the set $\mathcal{P}(E^m)$ onto $\mathcal{P}(E)$ (or equivalently $\mathcal{P}(U)$, since we may identify U and E) by

$$\forall F \in \mathcal{P}(E^m) \qquad T_U(F) = \{\, i \in E : (i) \in F\,\}$$

and we put

$$\mathcal{C}_U(E^m) = \{\, T_U(\pi) : \pi \in \mathcal{C}(E^m)\,\}\,.$$

Since the set U of the uniform populations verify

$$\forall x \in E^m \quad \exists y \in U \qquad V(x,y) = 0$$

we are in position to apply theorem 5.8: the communication altitude and the virtual energy on the set U may be evaluated by considering only paths in U, with either the cost function V_U or V. As a consequence, we have the following

Theorem 7.3. *The set of cycles in U associated with the cost function V_U i.e.*

$$\mathcal{C}(U) = \bigcup_{\lambda\in\mathbb{R}+} W_\lambda \cap U \,/\, \mathcal{R}_\lambda$$

coincides with the set $\mathcal{C}_U(E^m)$, which is the projection of the cycles in E^m on U.

Notice that the cardinality of the set U is equal to $|E|$ and does not depend on m. The hierarchy of cycles over U is thus built by taking equivalence classes of comparison relations on U induced by the communication altitude which is an affine function of m (by corollary 6.15). Yet, the relative order of a finite family of affine functions does not change any more when the variable is sufficiently large.

Corollary 7.4. *There exists an integer M such that for all i_1, j_1, i_2, j_2 in E, we have*

$$\begin{array}{lll} \textit{either} & \forall m \geq M & A(i_1,j_1) = A(i_2,j_2) \\ \textit{or} & \forall m \geq M & A(i_1,j_1) < A(i_2,j_2) \\ \textit{or} & \forall m \geq M & A(i_1,j_1) > A(i_2,j_2). \end{array}$$

Let M^* be the smallest integer such that the limit behavior described by the preceding results is achieved for $m \geq M^*$; that is V, W, A are affine functions of m and the relative positions of these affine functions does not change any more on $[M^*, \infty[$. The preceding results yield the

Theorem 7.5. *(stabilization of the cycles of U)*
When m is greater than M^*, the set $\mathcal{C}(U)$ of cycles in U does not depend any more on m.

Proof. Let $\mathcal{A} = \{\, A(i,j) : i \in E,\, j \in E \,\}$. The cardinality of $\mathcal{A}$ is less than $|E|^2$.
Put $\mathcal{A} = \{\, a_1, \cdots a_s \,\}$ where s is a function of m and $0 \le a_1 < a_2 < \cdots < a_{s-1} < a_s$.
Let $N : E \times E \longmapsto \mathbb{N}$ be the function defined by

$$\forall i, j \in E \qquad a_{N(i,j)} = A(i,j).$$

Consider the equivalence relation $\mathcal{T}_k$ defined on the set $\{\, i \in E : N(i,i) \le k \,\}$ by $i\ \mathcal{T}_k\ j \iff A(i,j) \le a_k \iff N(i,j) \le k$. We have

$$\begin{aligned}\mathcal{C}(U) &= \bigcup_{\lambda \in \mathcal{A}} \{\, i : W(i) \le \lambda \,\} \,/\, \mathcal{R}_\lambda \\ &= \bigcup_{k=1}^{s} \{\, i : W(i) \le a_k \,\} \,/\, \mathcal{R}_{a_k} \\ &= \bigcup_{k=1}^{s} \{\, i : N(i,i) \le k \,\} \,/\, \mathcal{T}_k.\end{aligned}$$

so that in fact $\mathcal{C}(U)$ depends only on the functions $N(i,j)$, the integer s and the equivalence relations $(\mathcal{T}_k)$: corollary 2 shows that these objects do not vary when $m \ge M^*$. $\square$

To analyze the hierarchy of cycles, we need some information about the communication altitude, and first about the coefficients $\Omega(i,j), \Omega(i), \psi(i,j)$. This is the purpose of the next two lemmas.

Lemma 7.6. *Let i, j be two points of E. The coefficient $\Omega(i,j)$ vanishes if and only if there exists a path $e_1 \to \cdots \to e_r$ in E joining i and j such that*

$$\forall k \in \{1 \cdots r-1\} \quad \alpha(e_k, e_{k+1}) > 0 \quad \textit{and} \quad f(e_k) \le f(j).$$

Proof. Let q be an element of $\mathcal{P}(i,j)$ (see notation 6.4) such that $\Omega(q) = 0$ and let $(e^*_k)_{1 \le k \le |q|}$ be a path of $\mathcal{A}(q)$ such that

$$\sum_{1 \le 2k < |q|} a\left(1 - \delta(e^*_{2k-1}, e^*_{2k})\right) + c\left(\widehat{f}(q_{2k}) - f(e^*_{2k+1})\right) = 0.$$

Necessarily

$$\forall k \quad 1 \le 2k < |q| \qquad e^*_{2k-1} = e^*_{2k}, \quad \widehat{f}(q_{2k}) = f(e^*_{2k+1}),$$

whence $\widehat{f}(q_{2k}) \leq \widehat{f}(q_{2k+1})$. Moreover, since $[q_{2k+1}] \subset [q_{2k}]$, we have also $\widehat{f}(q_{2k+1}) \leq \widehat{f}(q_{2k})$. It follows that $\widehat{f}(q_{2k}) = \widehat{f}(q_{2k+1})$, $1 \leq 2k < |q|$ and the individual e^*_{2k+1} belongs to $\widehat{q}_{2k+1}$ (also $e^*_1 \in \widehat{q}_1$). Since for each k, $1 \leq 2k < |q|$, $e^*_{2k-1} = e^*_{2k}$, we have $\widehat{f}(q_{2k-1}) \leq \widehat{f}(q_{2k})$ and the sequence $\big(\widehat{f}(q_k)\big)_{1\leq k\leq |q|}$ is increasing. In particular

$$\forall k \in \{1 \cdots |q|\} \qquad \widehat{f}(q_k) \leq f(j).$$

Now, there exists a path $e_1 \to \cdots \to e_{|q|}$ in E which is admissible for q and such that

$$\forall k \quad 1 \leq 2k < |q| \qquad e_{2k} = e_{2k+1}$$

(this path is the path which leads to the creation of the individual j).
This path necessarily satisfies

$$\forall k \in \{1 \cdots |q|\} \qquad f(e_k) \leq f(j).$$

We suppress the elements e_k of this path such that $e_k = e_{k-1}$ to obtain a path with the desired properties.
The reverse implication is easy: we build a path p which contains an individual following the path $e_1 \to \cdots \to e_r$ given by the hypothesis of the lemma and we just let the remaining individuals evolve "naturally" i.e. there are no mutation and no selection of individuals below peak fitness apart from those of the path $e_1 \to \cdots \to e_r$ (this is explicitly done in the proof of lemma 9.1 below). $\square$

The next lemma will be used in the proofs of corollary 7.8 and proposition 8.2.

Lemma 7.7. $\forall i \in f^* \quad \forall j \in E \qquad \Omega(i) = 0$ *and* $\Omega(j) = \Omega(i,j)$.

Proof. Let i belong to f^*. For each j_1 in $E\backslash f^*$, we choose a point j_2 such that $f(j_1) < f(j_2)$ and there exists a path $e_1 \to \cdots \to e_r$ in E joining j_1 and j_2 satisfying

$$\forall k \in \{1 \cdots r-1\} \qquad \alpha(e_k, e_{k+1}) > 0, \quad f(e_k) \leq f(j_2).$$

Let g be the i–graph built with all these arrows $j_1 \to j_2$ and the arrows $j \to i$, $j \in f^* \setminus \{i\}$.
Lemma 7.6 shows that

$$\sum_{(j_1 \to j_2)\in g} \Omega(j_1, j_2) = 0$$

whence $\Omega(i) = 0$ and the first part of the lemma is proved.
Let j belong to E. Since $\Omega(i) + \Omega(i,j) \geq \Omega(j)$, we have $\Omega(j) \leq \Omega(i,j)$.
To prove the reverse inequality, first notice that Ω satisfies the triangular inequality

$$\forall j_1, j_2, e \in E \qquad \Omega(j_1, j_2) \leq \Omega(j_1, e) + \Omega(e, j_2)$$

(to prove this inequality, we just put end to end paths joining (j_1) to (e) and paths joining (e) to (j_2) to obtain paths joining (j_1) to (j_2) and then take the infimum defining Ω).
Yet each j–graph contains a path $e_1 \to \cdots \to e_r$ joining i to j whence

$$\Omega(j) \geq \Omega(e_1, e_2) + \cdots + \Omega(e_{r-1}, e_r) \geq \Omega(i,j)$$

and finally $\Omega(j) = \Omega(i,j)$. $\square$

Corollary 7.8. *The rate of increase of the communication altitude is*

$$\forall i,j \in E \quad \psi(i,j) = \lim_{m\to\infty} \frac{A(i,j)}{m} = \max\big(\Omega(i),\Omega(j)\big).$$

Proof. We already know that $A(i,j)$ is affine for m sufficiently large (by lemma 6.15). It remains only to prove that

$$\lim_{m\to\infty} \frac{A(i,j)}{m} = \max\big(\Omega(i),\Omega(j)\big).$$

Since $A(i,j) \geq \max\big(W(i),W(j)\big)$, we have

$$\lim_{m\to\infty} \frac{A(i,j)}{m} \geq \max\big(\lim_{m\to\infty} \frac{W(i)}{m}, \lim_{m\to\infty} \frac{W(j)}{m}\big) = \max\big(\Omega(i),\Omega(j)\big).$$

Now pick a point i^* in f^* and consider the path $i \to i^* \to j$. Lemma 7.7 yields

$$\lim_{m\to\infty} \frac{A(i,j)}{m} \leq \max\big(\Omega(i),\Omega(j)\big)$$

and the result of the corollary is proved. □

We focus on the limiting decomposition. From now onwards, the population size m is assumed to be greater than M^*.

Theorem 7.9. *(structure of the cycles)*
Let π be a cycle of $\mathcal{C}(U)$. Suppose π is not included in W^*.
Then either Ω is constant on π or the set

$$\{\, i : \Omega(i) < \max_{j\in\pi} \Omega(j)\,\}$$

(which contains f^*) is included in the cycle π.

Proof. Let π be a cycle of $\mathcal{C}(U)$ not included in W^*. There exists a real number $\lambda(m)$ such that π is an equivalence class of $W_{\lambda(m)} \cap U$ for the relation $\mathcal{R}_{\lambda(m)}$ (for all $m \geq M^*$). Suppose Ω is not constant on π, so that there exist j and e in π such that

$$\Omega(j) = \max_{i\in\pi} \Omega(i), \quad \Omega(e) < \Omega(j).$$

Since $A(e,j) \leq \lambda(m)$, necessarily

$$m \max\big(\Omega(e),\Omega(j)\big) = m\,\Omega(j) \leq \lambda(m).$$

Let i be a point of E such that $\Omega(i) < \Omega(j)$. We have, as $m \to \infty$

$$\frac{A(i,e)}{m} \longrightarrow \max\big(\Omega(i),\Omega(e)\big) < \Omega(j)$$

so that for m sufficiently large $A(i,e) < m\,\Omega(j) \leq \lambda(m)$ and i belongs to the cycle π. □

Finally, we prove a general fact concerning the "bad" cycles.

Theorem 7.10. *(cycles disjoint from W^*)*
The function f is constant over the cycles of $\mathcal{C}(U)$ not intersecting W^*.

Remark. The cycles not intersecting W^* are the "bad" cycles which slow down the convergence of the process toward W^* (the critical height H_1 is the maximal height of exit of these cycles [9,10,11]). The above result shows that these cycles are in a way "transverse" to the "good" cycles which intersect W^*.

Proof. Let π belong to $\mathcal{C}(U)$ be such that $\pi \cap W^* = \emptyset$. Let i, j be two points of π and let $e_1 \to \cdots \to e_r$ be a path in E joining i and j such that

$$A(i,j) = \max_{1 \le k < r} W(e_k) + V(e_k, e_{k+1}). \tag{4}$$

Once more, for $m \ge M^*$, the quantities involved in the above formula are affine functions of m so that the path $e_1 \to \cdots \to e_r$ actually realizes the value $A(i,j)$ for all values of m in $[M^*, \infty[$ (the minimizing path should a priori depends on m). Necessarily, all the points of this path are in the cycle π, so that by theorem 7.9,

$$\Omega(i) = \Omega(j) = \Omega(e_1) = \cdots = \Omega(e_r).$$

Identity (4) implies also

$$\max\big(\Omega(i), \Omega(j)\big) = \max_{1 \le k < r} \Omega(e_k) + \Omega(e_k, e_{k+1})$$

whence

$$\forall k \in \{1 \cdots r-1\} \qquad \Omega(e_k, e_{k+1}) = 0$$

and by lemma 7.6

$$\forall k \in \{1 \cdots r-1\} \qquad f(e_k) \le f(e_{k+1}).$$

Finally, we have $f(i) \le f(j)$ and by symmetry it follows that $f(i) = f(j)$. Thus f is constant over the cycle π. □

Translating this result on E^m with the projection T_U, we obtain

Corollary 7.11. *The function $\widehat{f}$ is constant over the cycles of $\mathcal{C}(E^m)$ not intersecting W^*.*

Remark. When the process starts from a uniform population (i), it explores very intensively the neighbourhood of i until a point j of greater fitness is found. The cost to find such a point j corresponds to the mutations and anti–selections necessary to lead an individual from i to j while the remainder of the population waits in i. Then the process moves to the uniform population (j) and the mechanism starts again from scratch at j, until a global maxima of f is finally reached. With high probability, the process will thus visit f^* before coming back to (i). Therefore, as soon as a cycle contains two populations with distinct maximal fitness, it contains also points of f^*.

8. Influence of the mutation cost

We consider first the case where the mutation cost (more precisely the ratio a/c) is very high compared to the variations of the fitness function f.

Proposition 8.1. *(upper bound of Ω with selection)*

Define

$$\omega(i,j)=\inf\big\{\max_{1\le k\le r} f(e_k)-f(j): e_1=i,\, e_r=j,\, \forall k\in\{1\cdots r-1\}\quad \alpha(e_k,e_{k+1})>0\big\}.$$

We have

$$\Omega(i,j)\ \le\ c\,\omega(i,j)$$

and

$$a\ >\ c\,\omega(i,j)\quad\Longrightarrow\quad \Omega(i,j)\ =\ c\,\omega(i,j).$$

Remark. Notice here something very queer: the quantity $\omega(i,j)$ plays also an essential role in the theory of the sequential simulated annealing, although this algorithm is always studied as a minimization procedure.

Proof. Let $e_1=i\to e_2\to\cdots\to e_r=j$ be a path in E such that $\alpha(e_k,e_{k+1})>0$ for all k in $\{1\cdots r-1\}$. Put $q_1=\{e_1\}$,

$$q_{2k}=q_{2k+1}=\{e_1,\cdots,e_k,e_{k+1}\}\quad\text{for } k \text{ in } \{1\cdots r-2\},$$

and $q_{2r-2}=\{e_1,\cdots,e_r\}$, $q_{2r-1}=\{j\}$.
Clearly, the path $q: q_1\to\cdots\to q_{2r-1}$ belongs to $\mathcal{P}(i,j)$ (see notation 6.4).
Let $(e^*_k)_{1\le k\le 2r-1}$ be a path in E such that $e^*_1=e_1$, $e^*_2=e^*_1$,

$$e^*_{2k-1}\in\widehat{q}_{2k-1},\, e^*_{2k}=e^*_{2k-1}\qquad\text{for } k \text{ in } \{2\cdots r-1\},$$

and $e^*_{2r-1}=j$. For this path, which belongs to $\mathcal{A}(q)$ (see definition 6.5) and which involves no mutations, we have

$$\sum_{1\le 2k<|q|} a\left(1-\delta(e^*_{2k-1},e^*_{2k})\right)+c\left(\widehat{f}(q_{2k})-f(e^*_{2k+1})\right)=$$

$$\sum_{1\le 2k<|q|} c\left(\widehat{f}(q_{2k})-f(e^*_{2k+1})\right)=c\left(\widehat{f}(q_{2r-2})-f(j)\right)=c\left(\max_{1\le k\le r} f(e_k)-f(j)\right)$$

from which we deduce $\Omega(i,j)\ \le\ c\,\omega(i,j)$.
Suppose now $a>c\,\omega(i,j)$. Let the pair $(q,(e^*_k)_{1\le k\le|q|})$ realize the minimum in $\Omega(i,j)$

(where q belongs to $[\overline{D}]$ (see notation 6.4) and $(e^*_k)_{1\leq k\leq |q|}$ is an admissible path for q). Since $\Omega(i,j)<a$, necessarily

$$\forall k\quad 1\leq 2k<|q|,\qquad e^*_{2k}=e^*_{2k-1}$$

and

$$\Omega(i,j)=c\sum_{1\leq 2k<|q|}\big(\widehat{f}(q_{2k})-f(e^*_{2k+1})\big).$$

Since the path e^* realizes the minimum of the above quantity among all the paths admissible for q, we have in addition

$$\forall k\quad 1\leq 2k<|q|,\qquad e^*_{2k+1}\in\widehat{q}_{2k+1}$$

so that

$$\Omega(i,j)=c\sum_{1\leq 2k<|q|}\big(\widehat{f}(q_{2k})-\widehat{f}(q_{2k+1})\big).$$

Yet $e^*_{2k-1}=e^*_{2k}$ whence $\widehat{f}(q_{2k})\geq\widehat{f}(q_{2k-1})$ and $\widehat{f}(q_{2k})-\widehat{f}(q_{2k+1})\geq\big(\widehat{f}(q_{2k-1})-\widehat{f}(q_{2k+1})\big)^+$ (if s is a real number, s^+ denotes the maximum $\max(s,0)$).
Let r be an index such that $\widehat{f}(q_{2r-1})=\max_{1\leq 2k-1\leq|q|}\widehat{f}(q_{2k-1})$.
We have (even if $2r-1=|q|$)

$$\Omega(i,j)\ \geq\ c\sum_{2r-1\leq 2k<|q|}\big(\widehat{f}(q_{2k-1})-\widehat{f}(q_{2k+1})\big)^+\ \geq\ c\,\big(\widehat{f}(q_{2r-1})-f(j)\big)$$

whence

$$\Omega(i,j)\ \geq\ c\,\big(\max_{1\leq 2k-1\leq|q|}\widehat{f}(q_{2r-1})-f(j)\big). \tag{5}$$

Necessarily, there exists a path $\tilde{e}$ admissible for q such that

$$\forall k\quad 1\leq 2k<|q|,\qquad \tilde{e}_{2k}=\tilde{e}_{2k+1}$$

(this path corresponds for instance to the creation of the individual j). Clearly

$$\max_{1\leq k\leq|q|}f(\tilde{e}_k)-f(j)\ \geq\ \omega(i,j) \tag{6}$$

(just delete all the elements $\tilde{e}_k$ of the path $(\tilde{e}_k)_{1\leq k\leq|q|}$ such that $\tilde{e}_k=\tilde{e}_{k-1}$ to obtain a path admissible for evaluating $\omega(i,j)$). However

$$\max_{1\leq k\leq|q|}f(\tilde{e}_k)\ =\ \max_{1\leq 2k-1\leq|q|}f(\tilde{e}_{2k-1})\ \leq\ \max_{1\leq 2k-1\leq|q|}\widehat{f}(q_{2k-1})$$

which, together with inequalities (5) and (6), imply $\Omega(i,j)\ \geq\ c\,\omega(i,j)$. $\square$

Proposition 8.2. *Suppose* $a \geq c\big(f(f^*) - f(i)\big)$*, where* i *is a point of* E*. Then*

$$\Omega(i) = c\big(f(f^*) - f(i)\big).$$

In particular, if $a \geq c\Delta$*, then (we recall that* $\Delta = \max\{\,|f(i) - f(j)| : i, j \in E\,\}$*)*

$$\forall i \in E \qquad \Omega(i) = c\big(f(f^*) - f(i)\big).$$

Remark. When the cost of the mutations is much higher than the cost of the anti–selections, the coefficient Ω is obtained by applying an affine transformation on the fitness function f.

Proof. Let i belong to E and e to f^*. We have $\omega(e,i) = c\big(f(f^*) - f(i)\big)$ and the inequality $a \geq c\big(f(f^*) - f(i)\big)$ implies by proposition 8.1 that $\Omega(e,i) = c\big(f(f^*) - f(i)\big)$. Finally, it follows from lemma 7.7 that $\Omega(i) = c\,\big(f(f^*) - f(i)\big)$. $\square$

Corollary 8.3. *Let* π *be a cycle of* $\mathcal{C}(U)$ *included in the set*

$$\big\{\, i : f(f^*) - f(i) \leq \frac{a}{c} \,\big\}.$$

Either f *is constant on* π *or* π *contains the set*

$$\big\{\, i : \min_{j \in \pi} f(j) < f(i) \,\big\}$$

(whence f^* *is included in* π*).*
In particular, if $a \geq c\Delta$*, then the function* f *is constant on all the cycles not containing* f^**. The other cycles verify*

$$\big\{\, i : \min_{j \in \pi} f(j) < f(i) \,\big\} \;\subset\; \pi.$$

Proof. This corollary is an easy consequence of theorem 7.9 and proposition 8.2.

Corollary 8.4. *(cycles of* E^m*)*
Suppose $a \geq c\Delta$*. Let* π *belong to* $\mathcal{C}(E^m)$*. Either* $\widehat{f}$ *is constant on* π *or the set*

$$\big\{\, (i) \in U : f(i) > \min_{j \in \pi} \widehat{f}(j) \,\big\}$$

is included in π *(whence* $f^* \subset \pi$*).*

Proof. It is enough to remark that

$$\forall \pi \in \mathcal{C}(E^m) \quad \forall x \in \pi \quad \forall i \in \widehat{x} \qquad V(x,(i)) = 0 \quad \text{and} \quad (i) \in \pi$$

and to translate the preceding corollary on E^m with the projection T_U. $\square$

We consider now the case where the mutation cost (more precisely the ratio a/c) is very low compared to the variations of the fitness function f.

Proposition 8.5. *(upper bound of Ω with mutation)*

Define

$$\varpi(i,j) = \inf\,\big\{\,\mathrm{card}\{\,k : 1\le k\le r,\ e_{2k-1}\neq e_{2k}\,\}\,\big\}$$

where the infimum is taken over the set of paths $e_1\to\cdots\to e_{2r+1}$ in E joining i and j (the length $2r+1$ is also variable) and satisfying

- $\forall k\in\{1\cdots r\}$ *either* $e_{2k}=e_{2k-1}$ *or* $\alpha(e_{2k-1},e_{2k})>0,\quad f(e_{2k-1})>\max\limits_{2k-1<h\le 2r+1} f(e_h)$;
- *for each k in $\{1\cdots r\}$, $f(e_{2k+1})\ge f(e_{2k})$ and there exists a path $j^k : j_1^k\to\cdots\to j^k_{n^k_k}$ of length n^k_k in E joining i and e_{2k+1} (i.e. $j_1^k=i$, $j^k_{n^k_k}=e_{2k+1}$) and $k-1$ integers $n_1^k,\cdots,n^k_{k-1}$ such that $(n_0^k=)1<n_1^k<\cdots<n^k_{k-1}<n^k_k$ and*

$$\forall h\in\{0\cdots k-1\}\quad \forall t\in\{n_h^k\cdots n^k_{h+1}-1\}\qquad f(j_t^k)\le f(e_{2h+1}).$$

Finally, for each h in $\{1\cdots n^k_k-1\}$, $\quad\alpha(j_h^k,j^k_{h+1})>0$.

We have

$$\Omega(i,j)\ \le\ a\,\varpi(i,j)$$

and

$$c\delta\ >\ a\,\varpi(i,j)\quad\Longrightarrow\quad \Omega(i,j)\ =\ a\,\varpi(i,j).$$

Remark. To obtain an upper bound for $\Omega(i,j)$ we will build a path in E^m joining the populations (i) and (j), containing only $\varpi(i,j)$ collective events, which are mutations. The path $e_1\to\cdots\to e_{2r+1}$ is the collective path followed by almost all individuals. The indexes k such that $e_{2k-1}\neq e_{2k}$ correspond to the collective mutations (that is almost all individuals mutate from e_{2k-1} to e_{2k}): such an event is justified if and only if $f(e_{2k-1})>\max_{2k-1<h\le 2r+1} f(e_h)$. Since our path should not contain collective anti–selections events we impose also $f(e_{2k+1})\ge f(e_{2k})$. Finally, the path j^k corresponds to the sequence of mutations leading to the appearance of e_{2k+1}: this creation should not interfere with the collective behavior of the other individuals (i.e. it should involve only local events) and the condition required on j^k (i.e. the existence of the subdivision n^k and the linked inequalities $f(j_t^k)\le f(e_{2h+1})$) exactly tells that e_{2k+1} may be created from i by a path of individuals whose fitness is always less than the maximal fitness of the current population. With the elements in hand, the only delicate problem to build the path is to stay the right amount of time in each state e_{2k+1} before leaving in order to let enough time to the path j^k to be completed without disturbing the collective evolution of the population. This is the object of the (quite intricate) construction done in the first part of

the proof. Finally, this result, which describes the situation dual to proposition 8.1, will not be used in the sequel, and the following (painful) proof may well be skipped. Its main interest is to give some insight into the structure of the trajectories of the populations.

Proof. Let $e_1 = i \to \cdots \to e_{2r+1} = j$ be a path in E verifying the conditions imposed on the paths used for defining ϖ and let $(j^k)_{1\leq k\leq r}$ and $(n_h^k)_{1\leq k\leq h\leq r}$ be the associated paths and subdivisions. Put $n_0 = 1$, $n_1 = \max_{1\leq k\leq r} n_1^k$, define by induction for h in $\{2\cdots r-1\}$

$$n_h = n_{h-1} + \max_{h\leq k\leq r} n_h^k - n_{h-1}^k$$

and finally $n_r = n_{r-1} + n_r^r - n_{r-1}^r$. We define now a path p of length $2n_r - 1$ in E^{r+1}. First, we let $p^1 = (i)$. For h and s such that $0 \leq s < h \leq r$, we put

$$p_h^{2(n_s+t-1)} = \begin{cases} j^h_{n_s^h+t} & \text{if } \; 1 \leq t < n_{s+1}^h - n_s^h \\ j^h_{n_{s+1}^h-1} & \text{if } \; n_{s+1}^h - n_s^h \leq t < n_{s+1} - n_s \\ j^h_{n_{s+1}^h} & \text{if } \; t = n_{s+1} - n_s \end{cases}$$

and

$$p_h^{2t+1} = p_h^{2t} \qquad 3 \leq 2t+1 \leq 2n_h - 1.$$

For h and s such that $1 \leq h \leq s < r$, we put $p_h^{2n_{s+1}-2} = e_{2s+2}$ and

$$p_h^t = e_{2s+1} \qquad 2n_s - 1 \leq t < 2n_{s+1} - 2.$$

For the $(r+1)$–th individual, for each s in $\{0\cdots r-1\}$,

$$p_{r+1}^t = \begin{cases} e_{2s+1} & \text{if } \; 2n_s - 1 \leq t < 2n_{s+1} - 2 \\ e_{2s+2} & \text{if } \; t = 2n_{s+1} - 2 \end{cases}$$

Finally, for each h in $\{1\cdots r+1\}$, $p_h^{2n_r-1} = e_{2r+1} = j$. We then consider the path $(e_k^*)_{1\leq k\leq 2n_r-1}$ followed by the last individual: $e_k^* = p_{r+1}^k$, $1 \leq k \leq 2n_r - 1$. We have

$$\begin{aligned} \sum_{1\leq 2k<|p|} a\left(1-\delta(e_{2k-1}^*, e_{2k}^*)\right) + c\left(\widehat{f}(p_{2k}) - f(e_{2k+1}^*)\right) &= \sum_{s=0}^{r-1} \sum_{2n_s-1\leq 2k<2n_{s+1}-1} \cdots \\ &= \sum_{s=0}^{r-1} a\left(1-\delta(e_{2s+1}, e_{2s+2})\right) \end{aligned}$$

from which we deduce the inequality $\Omega(i,j) \;\leq\; a\,\varpi(i,j)$.

Suppose now $c\delta > a\,\varpi(i,j)$. Let $(q,(e^*_k)_{1\leq k\leq |q|})$ be a pair realizing the minimum in $\Omega(i,j)$ (where q belongs to $[\overline{D}(i,j)]$ and e^* is a path in E admissible for q). Thus

$$\Omega(i,j) = \sum_{1\leq 2k<|q|} a(1-\delta(e^*_{2k-1},e^*_{2k})) + c\big(\widehat{f}(q_{2k}) - f(e^*_{2k+1})\big)$$

and since $c\delta > a\,\varpi(i,j) \geq \Omega(i,j)$, necessarily for each k, $1\leq 2k<|q|$, we have $e^*_{2k+1}\in \widehat{q}_{2k}$ (whence in particular $f(e^*_{2k+1})\geq f(e^*_{2k})$) and $\widehat{f}(q_{2k+1})=\widehat{f}(q_{2k})$.
Let us show that the path $(e^*_k)_{1\leq k\leq |q|}$ satisfies the conditions imposed on the paths defining the infimum in $\varpi(i,j)$. Let k be an integer such that $1\leq 2k<|q|$. Suppose $e^*_{2k}\neq e^*_{2k-1}$. Then $\alpha(e^*_{2k-1},e^*_{2k})>0$. Suppose by absurd there exists an index h, $2k-1<2h+1\leq |q|$, such that $f(e^*_{2k-1})\leq f(e^*_{2h+1})$. We define now a new pair $(\tilde{q},\tilde{e}^*)$; we put

$$\tilde{q}_t = \begin{cases} q_t & \text{if } \ 1\leq t<2k, \quad 2h<t\leq |q| \\ q_t\cup\{e^*_{2k-1}\} & \text{if } \ 2k\leq t\leq 2h \end{cases}$$

and we choose a sequence $(\tilde{e}^*)$ such that

$$\begin{array}{lll} \tilde{e}^*_t & = e^*_t & 1\leq t<2k, \quad 2h<t\leq |q| \\ \tilde{e}^*_{2t+1} & \in \widehat{\tilde{q}}_{2t+1} & 2k\leq 2t+1\leq 2h \\ \tilde{e}^*_{2t} & = e^*_{2t-1} & 2k\leq 2t<2h \end{array}$$

The pair $(\tilde{q},\tilde{e}^*)$ satisfies the conditions related to the definition of the infimum in $\Omega(i,j)$ and

$$\sum_{1\leq 2k<|\tilde{q}|} a(1-\delta(\tilde{e}^*_{2k-1},\tilde{e}^*_{2k})) + c\big(\widehat{f}(\tilde{q}_{2k}) - f(\tilde{e}^*_{2k+1})\big)$$

$$\leq\ a\ + \sum_{1\leq 2k<|q|} a(1-\delta(e^*_{2k-1},e^*_{2k})) + c\big(\widehat{f}(q_{2k}) - f(e^*_{2k+1})\big)$$

which contradicts the fact that (q,e^*) realizes the value $\Omega(i,j)$.
Thus

$$f(e^*_{2k-1}) < \max_{2k-1<h\leq |q|} f(e^*_h).$$

The remaining condition, i.e. the existence of the paths (j^k) and the subdivision (n^h_k), is a consequence of the following facts:

- the path q belongs to $[\overline{D}]$
- for each k, $1\leq 2k<|q|$, the individual e^*_{2k+1} belongs to $\widehat{q}_{2k}$.

To find the path j^k, we just look at the sequence of mutations which leads to the creation

of the individual e^*_{2k+1} in the sequence of populations $q^1 \to \cdots \to q^{2k}$.
Since e^* satisfies the conditions imposed on the paths defining the infimum in $\varpi(i,j)$, we obtain $\Omega(i,j) \geq a\,\varpi(i,j)$ whence finally $\Omega(i,j) = a\,\varpi(i,j)$. $\square$

The remarkable fact is that for $a \geq c\Delta$, the structure of the cycles is essentially determined by the level sets of the function f (the kernel α just plays a role within the level sets). We have the opposite phenomenon when $Ra \leq c\delta$ (where R is the minimal number of transitions necessary to join two arbitrary points of E through the kernel α, see section 4). The structure of the cycles is then essentially determined by the kernel α and the fitness landscape (E, f, α). However, since there is no easy way of describing this dynamics ($\Omega(i,j)$ is then equal to $\varpi(i,j)$), we do not state the results dual to corollaries 8.3 and 8.4.

Example. Consider the fitness landscape of figure 1.

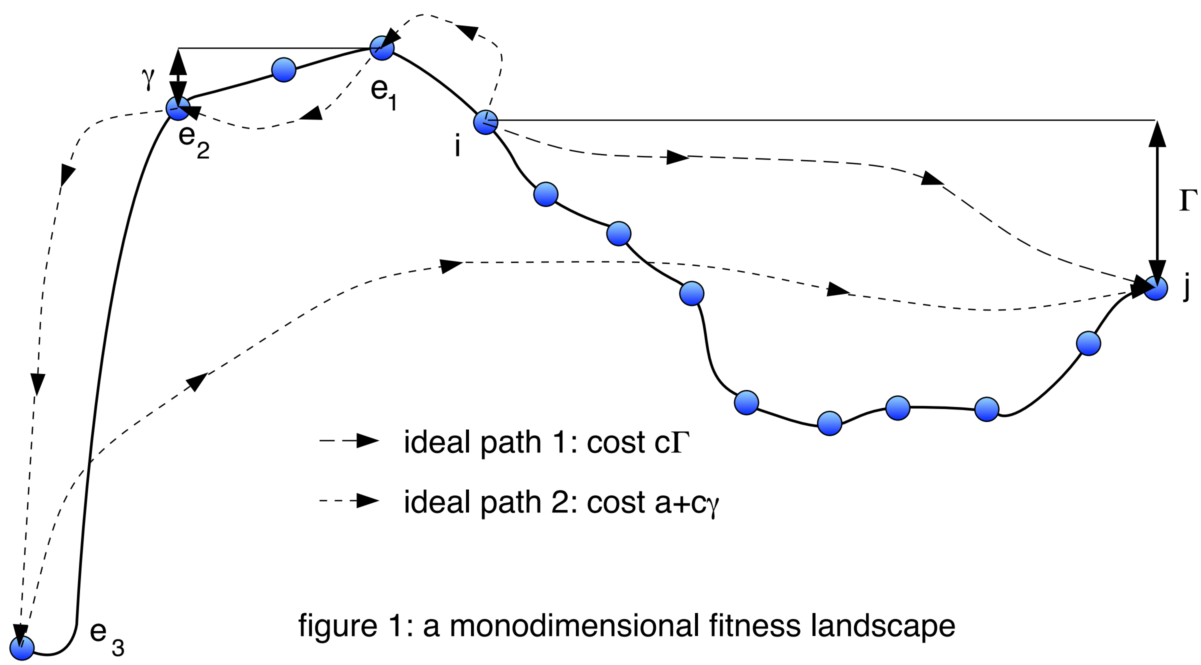

figure 1: a monodimensional fitness landscape

There are essentially two candidates for the ideal path between the points i and j; either the whole population stays in i while an explorer goes alone to reach j and then a massive anti–selection brings everybody in j (anti–selection cost of $c\Gamma$) or the whole population goes successively to e_1 (null cost), to e_2 (anti–selection cost of $c\gamma$), to e_3 (mutation cost a) and finally to j (null cost). Notice that the structure of the second path is much more intricate than the first one. It requires that two explorers leave i: one will go to j while the other will successively go to e_1 and e_2. The moves of the explorers and of the population should be carefully synchronized, so as to minimize the global cost. The best path depends upon the value of the mutation cost (if $a > c(\Gamma - \gamma)$) the path 1 is less expensive). When the mutation cost is low, the way the algorithms wanders in the fitness landscape depends strongly upon the mutation kernel α.

9. The limiting distribution

The aim of this section is to study precisely the limiting distribution when l is infinite and the mutation kernel α is symmetric. We first give several results necessary to establish the main theorem 9.7. We start by studying the coefficient $\theta(i,j)$.

Lemma 9.1. *Suppose* $\Omega(i,j) = 0$. *Then*

$$\theta(i,j) = \inf\Big\{\, a(r-1) \,+\, c \sum_{k=1}^{r} \big(\max_{1\le l\le k} f(e_l) - f(e_k)\big) \Big\}$$

where the infimum is taken over the paths $e_1 \to \cdots \to e_r$ *in* E *(of variable length* r*) verifying* $e_1 = i$, $e_r = j$ *and for each* k *in* $\{1 \cdots r-1\}$,

$$e_k \neq e_{k+1}, \quad \alpha(e_k, e_{k+1}) > 0, \quad f(e_k) \le f(j).$$

Remark. The infimum does not change if we suppress the condition $e_k \neq e_{k+1}$: the infimum is attained with a path satisfying this additional condition.

Proof. Let $e_1 \to \cdots \to e_r$ be a path in E joining i and j satisfying the above conditions. Let $(e_k^*)_{1\le k<r}$ be a sequence such that

$$\forall k \in \{1 \cdots r-1\} \qquad e_k^* \in \{e_1 \cdots e_k\} \quad \text{and} \quad f(e_k^*) = \max_{1\le l\le k} f(e_l)$$

and let p be the element of $\overline{D}^2(i,j)$ (see notation 6.2) defined by

$$p^{2k-1} = \{e_k^*, e_k\}, \quad p^{2k} = \{e_k^*, e_{k+1}\} \qquad \text{for } k \text{ in } \{1 \cdots r-1\}$$

and $p^{2r-1} = \{e_r, e_r\}$. We have $\Omega([p]) = 0$ so that

$$\theta(i,j) \;\le\; \overline{V}(p) \;=\; a(r-1) + c\sum_{k=1}^{r} \big(\max_{1\le l\le k} f(e_l) - f(e_k)\big). \tag{7}$$

Conversely, let q belong to $\mathcal{P}^*(i,j)$: thus $\Omega(q)=0$ and the sequence $\big(\widehat{f}(q_k)\big), 1\leq k\leq |q|$, is increasing (see the proof of lemma 7.6). Let p be an element of $\overline{D}(i,j)$ (see notation 6.4) such that $[p]=q$. Necessarily, there exists a path $(e_k), 1\leq k\leq |p|$ in $\mathcal{A}([p])$ (see definition 6.5) such that $e_{2k}=e_{2k+1}$ and $(e_{2k-1},e_{2k})\in\big\{(p_h^{2k-1},p_h^{2k}) : 1\leq h\leq m(p)\big\}$ for each k, $1\leq 2k<|p|$. Yet

$$\overline{V}(p) \;\geq\; \sum_{1\leq 2k<|p|} a\,\big(1-\delta(e_{2k-1},e_{2k})\big)+c\,\big(\widehat{f}(p_{2k})-f(e_{2k+1})\big).$$

Since the sequence $\big(\widehat{f}(q_k)\big)_{1\leq k\leq |q|}$ is increasing, we have $\widehat{f}(p_{2k})\geq \max_{1\leq l\leq 2k+1} f(e_l)$ so that

$$\overline{V}(p) \;\geq\; \sum_{1\leq 2k<|p|} a\,\big(1-\delta(e_{2k-1},e_{2k})\big)+c\,\big(\max_{1\leq l\leq 2k+1} f(e_l)-f(e_{2k+1})\big).$$

Let $\tilde{e}_1\to\cdots\to\tilde{e}_r$ be the path obtained by deleting from the path $e_1\to\cdots\to e_{|p|}$ all the individuals e_k such that $e_k=e_{k-1}$. We have

$$\overline{V}(p) \;\geq\; a(r-1)\,+\,c\,\sum_{k=1}^{r}\big(\max_{1\leq l\leq k} f(\tilde{e}_l)-f(\tilde{e}_k)\big)$$

which together with inequality (7) yields the desired result. $\square$

Corollary 9.2. *Suppose the kernel α is symmetric i.e.*

$$\forall i,j\in E \qquad \alpha(i,j)=\alpha(j,i)$$

and let i,j be two points of E such that $f(i)=f(j)$ and $\Omega(i,j)=0$.
Then $\Omega(j,i)=0$ and $\theta(i,j)=\theta(j,i)$.

Proof. In this situation, the symmetry of α together with lemma 7.6 yield $\Omega(j,i)=0$. In addition, the quantity

$$\theta(i,j) \;=\; \inf\Big\{\,a(r-1)\,+\,c\sum_{k=1}^{r}\big(f(i)-f(e_k)\big) : e_1=i,\ e_r=j,$$
$$\forall k\in\{1\cdots r-1\}\quad e_k\neq e_{k+1},\ \alpha(e_k,e_{k+1})>0,\ f(e_k)\leq f(j)\Big\}$$

becomes symmetric with respect to i and j. Thus $\theta(i,j)=\theta(j,i)$. $\square$

We now state a general lemma about graphs.

Lemma 9.3. *Let i belong to E and let g be a graph on E such that for each j in $E \setminus \{i\}$ there exists a path $j = e_1 \to e_2 \to \cdots \to e_r = i$ in g leading from j to i. Then there exists an i–graph $\tilde{g}$ which is contained in g i.e.*

$$\forall j_1, j_2 \in E \qquad (j_1 \to j_2) \in \tilde{g} \quad \Longrightarrow \quad (j_1 \to j_2) \in g.$$

Proof. We build the graph $\tilde{g}$ by removing arrows from g. First, we suppress all arrows starting at i. We then consider successively each point of $E \setminus \{i\}$. Let j be such a point. By hypothesis, there exists at least a path $j = e_1 \to e_2 \to \cdots \to e_r = i$ in g leading from j to i. We remove all arrows starting at j distinct from $(e_1 \to e_2)$. We continue this procedure until each point of $E \setminus \{i\}$ is the starting point of exactly one arrow. □

We now prove two lemmas which describe very accurately the paths joining the points of f^* in minimal graphs (these are the paths which will appear in the formula expressing the limiting distribution).

Notation 9.4. For a point i of E, we denote by $G_U^*(i)$ the elements of the set $G_U(i)$ (i.e. the i–graphs over E) whose V–cost is minimal. Equivalently, an element g of $G_U(i)$ belongs to $G_U^*(i)$ if and only if $V(g) = W(i)$ (see definition 5.3).

Remark. Let g belong to $G_U^*(i)$. We have clearly $V_U(g) = V(g) = W_U(i) = W(i)$ and for all arrows $(j_1 \to j_2)$ of g, $V_U(j_1, j_2) = V(j_1, j_2)$.

Lemma 9.5. *Let i belong to f^* and g to $G_U^*(i)$. Suppose the arrow $(j_1 \to j_2)$ belong to g, where j_1 and j_2 are elements of f^*. Then any path $e_1 = j_1 \to \cdots \to e_r = j_2$ in E realizing the value*

$$\theta(j_1, j_2) \;=\; \inf\Big\{ a(r-1) + c\sum_{k=1}^{r}\big(f(f^*) - f(e_k)\big) : e_1 = j_1,\, e_r = j_2,$$
$$\forall k \in \{1 \cdots r-1\} \quad e_k \neq e_{k+1},\ \alpha(e_k, e_{k+1}) > 0 \Big\}$$

is such that

$$\forall k \in \{2 \cdots r-1\} \qquad e_k \notin f^*.$$

Proof. Let i and g be as in the hypothesis and suppose the result is false: there exists an arrow $(j_1 \to j_2)$ in g, where j_1, j_2 belong to f^*, and a path $e_1 \to \cdots \to e_r$ in E joining j_1 and j_2 such that

$$\forall k \in \{1 \cdots r-1\} \quad e_k \neq e_{k+1},\ \alpha(e_k, e_{k+1}) > 0, \qquad \exists h \in \{2 \cdots r-1\} \qquad e_h \in f^*$$

$$\text{and} \qquad \theta(j_1, j_2) \;=\; a(r-1) + c\sum_{k=1}^{r}\big(f(f^*) - f(e_k)\big).$$

We have then

$$\theta(j_1,e_h) \;=\; a(h-1) + c\sum_{k=1}^{h}\big(f(f^*)-f(e_k)\big),$$

$$\theta(e_h,j_2) \;=\; a(r-h) + c\sum_{k=h}^{r}\big(f(f^*)-f(e_k)\big).$$

We build from g a graph $\tilde{g}$ in the following way:
- if $e_h = i$, we replace the arrow $(j_1 \to j_2)$ by $(j_1 \to i)$;
- if $e_h \neq i$, we replace the arrow $(j_1 \to j_2)$ by the arrows $j_1 \to e_h \to j_2$ and we remove the arrow starting at e_h.

Since $V(e_h,j)$ is strictly positive for any j distinct from e_h, we obtain in both cases a graph $\tilde{g}$ such that $V(\tilde{g}) < V(g)$ and for each point j in $E\setminus\{i\}$, there exists a path $e_1 = j \to \cdots \to e_r = i$ in $\tilde{g}$ leading from j to i. Lemma 9.3 shows that $\tilde{g}$ contains an i–graph $\tilde{\tilde{g}}$, and necessarily $V(\tilde{\tilde{g}}) < V(g)$, which is absurd since g belongs to $G^*_U(i)$. □

Lemma 9.6. *Let i belong to f^* and g to $G^*_U(i)$ (see notation 9.4). Suppose the arrow $(j_1 \to j_2)$ belong to g, where j_1 and j_2 are elements of f^*.*
A path p in E^m belongs to $D^{m*}(j_1,j_2)$ (see definition 5.2) if and only if:
there exists a path $e_1 = j_1 \to \cdots \to e_r = j_2$ in E realizing the value $\theta(j_1,j_2)$ verifying

$$\forall k\in\{1\cdots r-1\}\qquad e_k\neq e_{k+1},\ \alpha(e_k,e_{k+1})>0$$

and two integers $0\le t_1 < t_2 < |p|$ such that

$$\begin{array}{ll} p^s = (j_1) & 1\le s\le t_1+1\\ p^s(j_1)=m-1,\ p^s(e_{s-t_1})=1 & t_1+2\le s\le t_1+r-1\\ {[p^s]}=\{j_1,j_2\} & t_1+r\le s\le t_2\\ p^s=(j_2) & t_2+1\le s\le |p| \end{array}$$

Proof. Any path p satisfying the above conditions belongs to $D^m(j_1,j_2)$, has a cost $\theta(j_1,j_2)$ and thus belongs to $D^{m*}(j_1,j_2)$ (see definition 5.2).
Conversely, let p be an element of $D^{m*}(j_1,j_2)$. Put

$$t_1=\min\{\,s: p^s\neq(j_1)\,\}-2,\qquad t_2=\max\{\,s: p^s\neq(j_2)\,\}+1.$$

Necessarily, there exists a path $e_1 = j_1 \to \cdots \to e_r = j_2$ in E satisfying

$$\forall k\in\{1\cdots r-1\}\qquad e_k=e_{k+1}\quad\text{or}\quad \alpha(e_k,e_{k+1})>0,$$

$$\forall k\in\{2\cdots r-1\}\qquad e_k\in[p_{t_1+k}].$$

To obtain this path, we just look at the sequence of mutations which leads to the creation of j_2 in the path p. We suppress the elements e_k of this path such that $e_k = e_{k-1}$ to obtain a path with $\tilde{e}_1 = j_1 \to \cdots \to \tilde{e}_{\tilde{r}} = j_2$ in E satisfying the additional condition $\tilde{e}_k \neq \tilde{e}_{k+1}$. Thus the cost of the path p satisfies

$$(8) \qquad V(p) \geq a(\tilde{r}-1) + c\sum_{k=1}^{r}\big(f(f^*) - f(e_k)\big) \geq a(\tilde{r}-1) + c\sum_{k=1}^{\tilde{r}}\big(f(f^*) - f(\tilde{e}_k)\big)$$

(necessarily, $\Omega([p]) = \Omega(j_1, j_2) = 0$ so that $\forall k \in \{2 \cdots r-1\}$ $\widehat{f}(p_{t_1+k}) = \widehat{f}(f^*)$, see for instance the proof of lemma 7.6). However, the cost of p is precisely $\theta(j_1, j_2)$: thus the path $\tilde{e}_1 \to \cdots \to \tilde{e}_r$ realizes the value $\theta(j_1, j_2)$. The preceding lemma 9.5 implies that $\tilde{e}_k \notin f^*$ for k in $\{2 \cdots \tilde{r}-1\}$, so that $e_k \notin f^*$ for k in $\{2 \cdots r-1\}$. It follows from inequalities (8) that $r = \tilde{r}$ and the paths e and $\tilde{e}$ coincide (whence in particular $e_k \neq e_{k+1}$ for k in $\{1 \cdots r-1\}$). Furthermore the only events of positive cost in the path p are those concerning this individual path: that is there are no mutations nor selection of bad individuals apart from the individual which follows the path $e_1 \to \cdots \to e_r$. Thus

$$\begin{array}{lll} \forall s & t_1 + 2 \leq s \leq t_1 + r - 1 & p^s(j_1) = m-1,\ p^s(e_{s-t_1}) = 1 \\ \forall s & t_1 + r \leq s \leq t_2 & [p^s] = \{j_1, j_2\} \end{array}$$

and the path p satisfies the required conditions. □

We always suppose that m is greater than M^*. We are now in position to prove the

Theorem 9.7. *(limiting distribution)*
We have $W^* \subset f^*$. Suppose that the kernel α is symmetric. Then $W^* = f^*$ and the limit distribution ν^∞ is the uniform distribution on f^*:

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

Proof. The inclusion $W^* \subset f^*$ is given by theorem 4.4 (which was proved in [3]). Suppose that the kernel α is symmetric. Let i and j be two points of f^*. Let g be a graph belonging to $G_U^*(j) : V(g) = W(j)$. There exists a unique path $e_1 = i \to e_2 \to \cdots \to e_r = j$ in g leading from i to j. Since $\Omega(j) = 0$, necessarily

$$\forall k \in \{1 \cdots r-1\} \qquad \Omega(e_k, e_{k+1}) = 0 \quad \Longrightarrow \quad f(e_k) \leq f(e_{k+1}).$$

Since i is in f^*, we have in fact

$$\forall k \in \{1 \cdots r\} \qquad e_k \in f^*.$$

Let $\Phi_{ij}(g)$ be the i–graph obtained from g by reversing the arrows of this path in the following way: $e_r = j \to e_{r-1} \to \cdots \to e_1 = i$. By corollary 9.2, V is symmetric on f^*, so that $V(g) = V(\Phi_{ij}(g))$. It follows that $W(i) \le W(j)$ and by symmetry we conclude that $W(i) = W(j)$ i.e. the virtual energy is constant on f^* and $W^* = f^*$. Thus Φ_{ij} is a one-to-one map between $G^*_U(j)$ and $G^*_U(i)$. Moreover, $\Phi_{ij} \circ \Phi_{ji} = \mathrm{Id}_{G^*_U(i)}$. In addition, the operator Φ_{ij} reverses only arrows between points of f^*. Put

$$q(i,j) = \sum_{p \in D^{m*}(i,j)} \prod_{k=1}^{|p|-1} q_1(p_k, p_{k+1})$$

(for the definition of q_1, see section 3). It turns out that the limiting distribution is a rational fraction of the numbers $q(i,j)$.

Lemma 9.8. *The limiting distribution ν^∞ of the Markov chain X^l_n is concentrated on the uniform populations corresponding to the points of W^* and for i in W^* we have*

$$\nu^\infty(i) = \frac{\sum_{g \in G^*_U(i)} \prod_{(j_1 \to j_2) \in g} q(j_1, j_2)}{\sum_{j \in f^*} \sum_{g \in G^*_U(j)} \prod_{(j_1 \to j_2) \in g} q(j_1, j_2)}.$$

Proof. Let μ^l be the stationary measure of the Markov chain X^l_n i.e.

$$\forall x \in E^m \qquad \mu^l(x) \;=\; \lim_{n \to \infty} P\left(X^l_n = x / X^l_0 = y\right).$$

Let $({}^U X^l_n)$ be the Markov chain of successive visits of (X^l_n) to the set U (${}^U X^l_n$ is the chain induced by X^l_n on U). By identifying the sets U and E we may consider that the chain $({}^U X^l_n)$ takes its values in E. Let τ^l be the first entrance time of the chain (X^l_n) into U i.e. $\tau^l = \min\{\, n > 0 : X^l_n \in U \,\}$. We have

$$\forall i, j \in E \qquad P\left({}^U X^l_{n+1} = j / {}^U X^l_n = i\right) \;=\; P\left(X^l_{\tau^l} = (j) / X^l_0 = (i)\right).$$

Let ν^l be the invariant probability measure of the Markov chain $({}^U X^l_n)$. We have the representation formula

$$\forall x \in E^m \qquad \mu^l(x) \;=\; \mu^l(U) \sum_{j \in E} \nu^l(j)\, E_{(j)} \left[\sum_{k=0}^{\tau^l - 1} 1_{\{X^l_k = x\}} \right]$$

(where $E_{(j)}$ denotes the expectation for the chain starting at (j)).
For $x = (i)$ the above formula reduces to $\mu^l((i)) = \mu^l(U)\nu^l(i)$. Since $\mu^l(U)$ tends to one as l tends to infinity, we see that

$$\forall i \in E \qquad \lim_{l \to \infty} \mu^l((i)) \;=\; \lim_{l \to \infty} \nu^l(i) \;=\; \nu^\infty(i).$$

The measure ν^l may be expressed through Freidlin–Wentzell graphs [4, chapter 6]. For i in E, we have

$$\nu^l(i) = \frac{\sum_{g\in G_U(i)} \prod_{(j_1\to j_2)\in g} P\left({}^U X^l_{n+1} = j_2 / {}^U X^l_n = j_1\right)}{\sum_{j\in E}\sum_{g\in G_U(j)} \prod_{(j_1\to j_2)\in g} P\left({}^U X^l_{n+1} = j_2 / {}^U X^l_n = j_1\right)}.$$

Yet we have the expansion

$$P\left({}^U X^l_{n+1} = j / {}^U X^l_n = i\right) \underset{l\to\infty}{\sim} q_U(i,j)\exp\left(-V_U(i,j)\ln l\right)$$

where V_U is the cost defined on U by making the set U taboo (see definition 5.7) and

$$q_U(i,j) = \sum_{p\in D^{m*}_U(i,j)} \prod_{k=1}^{|p|-1} q_1(p_k,p_{k+1}),$$

the sum being carried over the set

$$D^{m*}_U(i,j) = \{\, p\in D^m(i,j) : \forall k, 1<k<|p|,\, p_k\notin U,\, V_1(p) = V_U(i,j)\,\}.$$

By passing through the limit as $l\to\infty$, we obtain that for i in W^*

$$\nu^\infty(i) = \frac{\sum_{g\in G^*_U(i)} \prod_{(j_1\to j_2)\in g} q_U(j_1,j_2)}{\sum_{j\in W^*}\sum_{g\in G^*_U(j)} \prod_{(j_1\to j_2)\in g} q_U(j_1,j_2)}.$$

Let g belong to $G^*_U(i)$ and let $(j_1\to j_2)$ be an arrow of g. Since $V_U(j_1,j_2) = V(j_1,j_2)$ (see the remark after notation 9.4) we have $D^{m*}_U(j_1,j_2)\subset D^{m*}(j_1,j_2)$. Conversely let p be a path belonging to $D^{m*}(j_1,j_2)$ and suppose $p\notin D^{m*}_U(j_1,j_2)$. Then there exists an index k, $1<k<|p|$, such that $p_k = (e)$, where $e\in E$. Thus $V(p)\geq V(j_1,e)+V(e,j_2)$. We replace the arrow $(j_1\to j_2)$ by the two arrows $(j_1\to e)$ and $(e\to j_2)$ in the graph g and we obtain a new graph with the same cost which is not any more an i–graph. By lemma 9.3, this graph contains an i–graph $\tilde{g}$. Since each transition between two distinct uniform populations has a positive cost (such a transition requires at least a mutation) and we have to delete some arrows to build $\tilde{g}$, we have $V(\tilde{g}) < V(g)$ which is absurd. Thus for each arrow $(j_1\to j_2)$ of a graph g of $G^*_U(i)$, we have $D^{m*}_U(j_1,j_2) = D^{m*}(j_1,j_2)$ and also $q_U(j_1,j_2) = q(j_1,j_2)$ (which is of course false for an arbitrary pair (j_1,j_2)). This fact yields the desired formula for the limiting distribution ν^∞. $\square$

The end of the proof of theorem 9.7 rests on the

Lemma 9.9. *Let i belong to f^* and g to $G^*_U(i)$. Suppose the arrow $(j_1 \to j_2)$ belong to g, where j_1 and j_2 are elements of f^*. Then $q(j_1, j_2) = q(j_2, j_1)$.*

Assume the lemma is true. It follows that

$$\forall i, j \in f^* \quad \forall g \in G^*_U(j) \qquad \prod_{(j_1 \to j_2) \in g} q(j_1, j_2) = \prod_{(j_1 \to j_2) \in \Phi_{ij}(g)} q(j_1, j_2)$$

whence

$$\sum_{g \in G^*_U(j)} \prod_{(j_1 \to j_2) \in g} q(j_1, j_2) = \sum_{g \in \Phi_{ij}\left(G^*_U(j)\right)} \prod_{(j_1 \to j_2) \in g} q(j_1, j_2) = \sum_{g \in G^*_U(i)} \prod_{(j_1 \to j_2) \in g} q(j_1, j_2)$$

and the above quantity is the same for all points of f^*: the limit distribution is thus uniform on f^*. □

Proof of lemma 9.9. Let i, j_1, j_2 be as in the hypothesis of the lemma. Let p be an element of $D^{m*}(j_1, j_2)$ (see definition 5.2). By lemma 9.6, there exists a path $e_1 = j_1 \to \cdots \to e_r = j_2$ realizing the value $\theta(j_1, j_2)$ and satisfying

$$\forall k \in \{1 \cdots r-1\} \qquad e_k \neq e_{k+1}, \quad \alpha(e_k, e_{k+1}) > 0$$

and two integers $0 \leq t_1 < t_2 < |p|$ such that

$$\begin{array}{ll} p^s = (j_1) & 1 \leq s \leq t_1 + 1 \\ p^s(j_1) = m-1,\ p^s(e_{s-t_1}) = 1 & t_1 + 2 \leq s \leq t_1 + r - 1 \\ [p^s] = \{j_1, j_2\} & t_1 + r \leq s \leq t_2 \\ p^s = (j_2) & t_2 + 1 \leq s \leq |p| \end{array}$$

With the path p we associate the path $\Phi(p) = \tilde{p}$ defined by

- $\forall s \quad 1 \leq s \leq |p| - t_2 \qquad \tilde{p}^s = (j_2)$

For each s such that $t_1 + 2 \leq s \leq t_1 + r - 1$, there exists a unique index $\sigma(s)$ such that $p^s_{\sigma(s)} = e_{s-t_1} \neq j_1$. We define

- $\forall s \quad |p| - t_2 + 1 \leq s \leq |p| - t_2 + r - 2$

$$\tilde{p}^s_h = \begin{cases} e_{r-|p|+t_2-s} & \text{if} \quad h = \sigma(s) \\ j_2 & \text{if} \quad h \neq \sigma(s) \end{cases}$$

Let $\tau_{j_1 j_2}$ be the transposition of E which exchanges the points j_1 and j_2.
For $x = (x_1, \cdots, x_m)$ in E^m, we put $\tau_{j_1 j_2} \cdot x = (\tau_{j_1 j_2}(x_1), \cdots, \tau_{j_1 j_2}(x_m))$.
We define

- $\forall s \quad |p| - t_2 + r - 1 \leq s \leq |p| - t_1 - 1 \qquad \tilde{p}^s = \tau_{j_1 j_2} \cdot p_{s-|p|+t_1+t_2+1}$

and finally

- $\forall s \quad |p|-t_1 \le s \le |p| \qquad \tilde{p}_s = (j_1)$.

The path $\tilde{p}$ is built by reversing the path $e_1 = j_1 \to \cdots \to e_r = j_2$ and by reproducing (with j_1 and j_2 exchanged) the portion of p which contains only the individuals j_1 and j_2. Since the kernel α is symmetric, the path $\Phi(p)$ belongs to $D^m(j_2, j_1)$ and its cost is $\theta(j_1, j_2)$; corollary 9.2 shows that θ is symmetric on f^*, so that $\Phi(p)$ is actually an element of $D^{m*}(j_2, j_1)$. Finally, the very definition of Φ yields the following facts:

- $\forall p \in D^{m*}(j_1, j_2) \qquad \Phi(\Phi(p)) = p$.
- Φ is one-to-one between $D^{m*}(j_1, j_2)$ and $D^{m*}(j_2, j_1)$.
- $\forall p \in D^{m*}(j_1, j_2)$

$$\prod_{k=1}^{|p|-1} q_1(p^k, p^{k+1}) = \prod_{k=1}^{|p|-1} q_1(\Phi(p)^k, \Phi(p)^{k+1})$$

(where $\Phi(p)^k$ denotes the k–th population of the path $\Phi(p)$).
As a consequence, we have

$$\begin{aligned} q(j_1, j_2) = & \sum_{p \in D^{m*}(j_1, j_2)} \prod_{k=1}^{|p|-1} q_1(p^k, p^{k+1}) = \sum_{p \in D^{m*}(j_1, j_2)} \prod_{k=1}^{|p|-1} q_1(\Phi(p)^k, \Phi(p)^{k+1}) \\ = & \sum_{p \in \Phi(D^{m*}(j_1, j_2))} \prod_{k=1}^{|p|-1} q_1(p^k, p^{k+1}) = \sum_{p \in D^{m*}(j_2, j_1)} \prod_{k=1}^{|p|-1} q_1(p^k, p^{k+1}) = q(j_2, j_1). \end{aligned}$$

and the lemma is proved. □

10. The critical height H_1

For the definitions and the properties of the quantities $H_e(\pi)$ and $H_m(\pi)$ (the height of exit and height of mixing of a cycle π), we refer the reader to Trouvé's work [9,10,11].

Proposition 10.1. *For each cycle π of $\mathcal{C}(E^m)$, we have*

$$H_e(\pi) \ \le \ \max_{i \in \pi} \ \min_{j \notin \pi} V(i, j)$$

Suppose α is symmetric. Then the critical height H_1 is bounded as a function of m and for m large we have

$$H_1 \ \le \ \max_{i \notin f^*} \ \min_{j \in f^*} V(i, j).$$

Proof. Let π be a cycle and let $F(\pi)$ be the points of π whose virtual energy is minimal. For each x belonging to $F(\pi)$, we have

$$H_e(\pi) = \min_{y \notin \pi} A(x, y) - W(x) \le \min_{(j) \notin \pi} A(x, (j)) - W(x) \le \min_{(j) \notin \pi} V(x, (j)).$$

Since in addition $F(\pi)$ is included in $\pi \cap U$, we obtain the first inequality.
The second inequality is an immediate consequence of the first one, the definition of H_1, $H_1 = \sup \{ H_e(\pi) : \pi \in \mathcal{C}(E^m), \pi \cap W^* = \emptyset \}$ ([9, Definition 3.22]), and the fact that $\sup_{m\in\mathbb{N}^*} \max_{i\notin f^*} \min_{j\in f^*} V(i,j) < \infty$ (see [3, Corollary 11.1]). □

We now restate Trouvé's convergence result, which is an extension of a result by Hajek for the simulated annealing.

Theorem 10.2. *(Trouvé [9, Theorem 3.23])*
Suppose α is symmetric and m is large enough to have $W^* = f^*$.
For all increasing sequences $l(n)$ going to infinity, we have the equivalence

$$\sup_{x\in E^m} \quad P\left(X_n \notin f^*/X_0 = x\right) \underset{n\to\infty}{\longrightarrow} 0 \qquad \Longleftrightarrow \qquad \sum_{n=0}^{\infty} l(n)^{-H_1} = \infty.$$

A remarkable fact is that we may adapt the sequence $l(n)$ in order to be trapped in f^*.

Theorem 10.3. *Define $H_e^* = H_e(\{ x \in E^m : [x] \cap f^* \neq \emptyset \})$ (for [ ], see definition 6.1).*
Suppose α is symmetric and m is large enough to have $W^* = f^*$ and $H_1 < H_e^*$.
For all increasing sequences $l(n)$, we have the equivalence

$$\begin{aligned} &\forall x \in E^m \qquad P\left(\exists N \quad \forall n \geq N \quad [X_n] \cap f^* \neq \emptyset / X_0 = x\right) = 1 \\ &\Longleftrightarrow \qquad \sum_{n=0}^{\infty} l(n)^{-H_1} = \infty, \qquad \sum_{n=0}^{\infty} l(n)^{-H_e^*} < \infty. \end{aligned}$$

Proof. The two above conditions on the sequence $l(n)$ exactly express that

- the chain (X_n) has a null probability of being trapped in a cycle disjoint from f^* i.e.

$$\forall x \in E^m \quad \forall I \subset E \setminus f^* \qquad P\left(\exists N \quad \forall n \geq N \qquad [X_n] \subset I / X_0 = x\right) = 0,$$

- the chain has a positive probability of being trapped in the set of populations containing an individual of f^* i.e.

$$\forall x \in E^m \qquad P\left(\exists N \quad \forall n \geq N \qquad [X_n] \cap f^* \neq \emptyset / X_0 = x\right) > 0. \quad \square$$

Remark. Since $H_e^* \geq m \min(a, c\delta^*)$, where $\delta^* = \min \{ f(f^*) - f(i) : i \notin f^* \}$, the hypothesis of the theorem is fulfilled when m is large, and any increasing sequence $l(n)$ satisfying

$$\sum_{n=0}^{\infty} l(n)^{-H_1} = \infty, \qquad \sum_{n=0}^{\infty} l(n)^{-m \min(a,c\delta^*)} < \infty$$

will achieve the desired behavior.

Example. Figure 2 shows graphically the critical height for both the simulated annealing and the genetic algorithm.

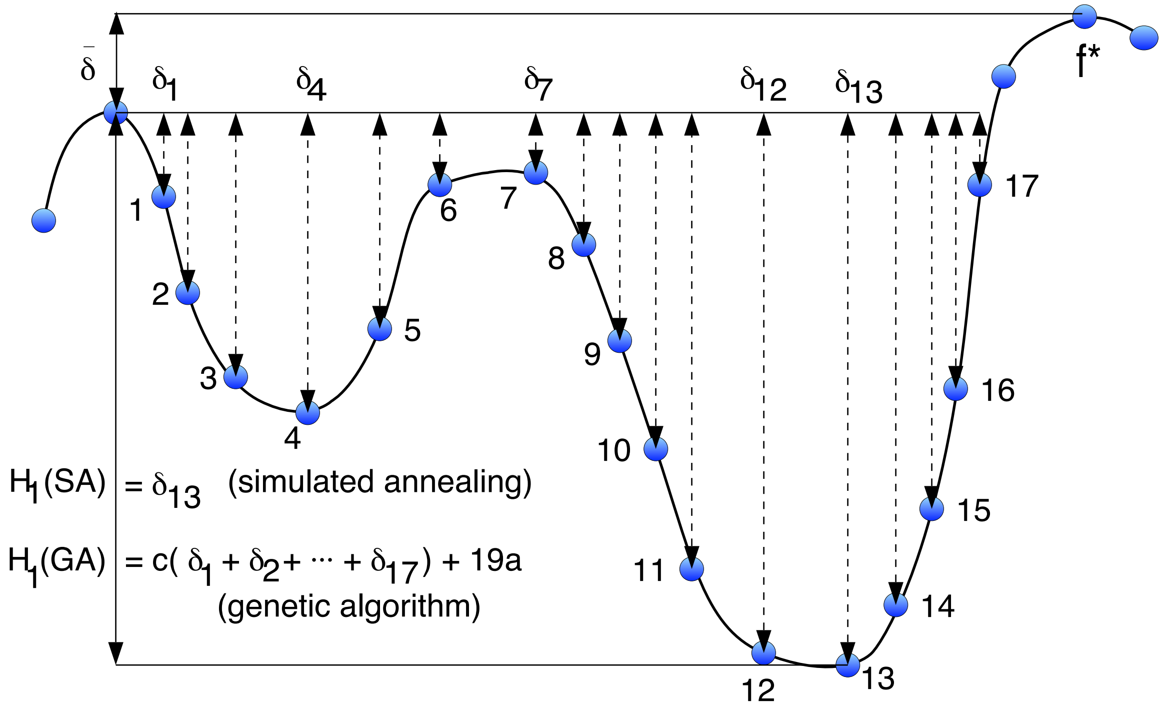

figure 2: the critical height H_1

The above values of H_1 concern one simulated annealing algorithm and a genetic algorithm with a large population (so that H_1 is equal to its limiting value and do not depend any more upon m).

11. The optimal rate of convergence

For the meaning and the properties of the optimal convergence exponent and the logarithmic convergence exponent we refer the reader to [1,2,9,10,11].

Proposition 11.1. *Suppose α is symmetric. The optimal convergence exponent α_{opt} is an affine strictly increasing function of m for m large. The rate of increase of α_{opt} is*

$$\lim_{m\to\infty}\frac{\alpha_{\text{opt}}}{m} \;=\; \min\left\{\frac{\Omega(\pi)}{H_e(\pi)} : \pi\in\mathcal{C}(E^m),\,\pi\cap f^*=\emptyset\right\} \;\geq\; \frac{\min(a,c\delta^*)}{H_1}$$

where $\delta^* = \min\{\, f(f^*) - f(i) : i \notin f^* \,\}$.

Proof. The definition of α_{opt} is [9, Definition 3.22]

$$\alpha_{\text{opt}} = \min\left\{ \frac{W(\pi) - W(W^*)}{H_e(\pi)} : \pi \in \mathcal{C}(E^m),\ \pi \cap f^* = \emptyset \right\}.$$

For π in $\mathcal{C}(E^m)$ with $\pi \cap f^* = \emptyset$, we have

$$\Omega(\pi) = \min\{\, \Omega(i) : i \in \pi \,\} \geq \min(a, c\delta^*) > 0$$

so that $W(\pi)$ is an affine strictly increasing function of m. The result of the proposition follows easily. □

We restate now Trouvé's result for the optimal convergence rate, which generalizes Catoni's work.

Theorem 11.2. *There exist two strictly positive constants* R_1 *and* R_2 *such that for all* n

$$\frac{R_1}{n^{\alpha_{\text{opt}}}} \quad \leq \quad \inf_{0 \leq l(1) \leq \cdots \leq l(n)} \quad \max_{x \in E^m} \quad P\left(X_n \notin f^* / X_0 = x\right) \quad \leq \quad \frac{R_2}{n^{\alpha_{\text{opt}}}}.$$

Proof. In order to apply Trouvé's result to the mutation–selection algorithm, we need only to check that condition C_1 (see [10,11]) concerning the transition probabilities is fulfilled; but these are precisely obtained as sums of fractions involving the powers of l. □

Remark. The fact that the optimal convergence exponent α_{opt} increases linearly with m shows that the mutation–selection algorithm is intrinsically parallel: it involves only local independent computations. We have here a quantitative measurement of this parallelism.

Example. Let us come back to the fitness landscape of figure 2. In this situation, the optimal convergence exponent of the sequential simulated annealing is $\alpha_{\text{opt}} = \bar{\delta}/\delta_{13}$. The optimal convergence exponent of the genetic algorithm satisfies

$$\lim_{m \to \infty} \frac{\alpha_{\text{opt}}}{m} \quad = \quad \frac{\min(a, c\bar{\delta})}{c(\delta_1 + \cdots + \delta_{17}) + 19a} \quad < \quad \frac{\bar{\delta}}{\delta_{13}}.$$

Consider now m independent simulated annealing algorithms running over this fitness landscape. We keep track of the best point found by the m algorithms. The optimal convergence exponent of this process is $m\,\bar{\delta}/\delta_{13}$, which is better than the exponent α_{opt} of the genetic algorithm with population size m (m being large). We suspect this result is true in the general case.

12. The logarithmic convergence exponent

We consider here the best rate of convergence which can be achieved with sequences of the form $l(n) = n^\kappa$, $\kappa \in \mathbb{R}_+^*$. The exponent $\alpha_{\log}$, when computed on the space E^m, is bounded. We have

$$\alpha_{\log} = \min\left\{ \frac{W(x) - W(W^*)}{H_1} : x \notin W^* \right\}.$$

Consider a state x of the form $(i, \cdots, i, j)$ where $i \in W^*$ and $j \notin W^*$, $\alpha(i,j) > 0$.
Clearly $W(x) \leq W(W^*) + a + c\big(f(i) - f(j)\big)$.
However, it is not fair to compare α_{opt} with this $\alpha_{\log}$ in this situation.
In fact, we are much more interested in the rate of convergence of

$$P\left(\widehat{f}(X_n) = f(f^*)/X_0 = x\right) = P\left([X_n] \cap f^* \neq \emptyset/X_0 = x\right)$$

than of $P\left([X_n] \subset f^*/X_0 = x\right)$. For $l(n) = n^{\frac{1}{H_1}}$, we obtain [10, Théorème 1.49]

$$P\left(\widehat{f}(X_n) < f(f^*)/X_0 = x\right) \quad \geq \quad \frac{K}{n^{\,\bar{\alpha}_{\log}}}$$

where K is a positive constant and $\bar{\alpha}_{\log}$ is defined by

$$\bar{\alpha}_{\log} = \min\left\{ \frac{W(x) - W(W^*)}{H_1} : x \in E^m, \widehat{f}(x) < f(f^*) \right\}.$$

Proposition 12.1. *The exponent $\bar{\alpha}_{\log}$ is an affine increasing function of m with*

$$\lim_{m\to\infty} \frac{\bar{\alpha}_{\log}}{m} \; = \; \frac{\min(a, c\delta^*)}{H_1}.$$

Proof. Let x in E^m be such that $\widehat{f}(x) < f(f^*)$. Let i_x be an element of $\widehat{x}$.
We have $V(x, (i_x)) = 0$ whence $W((i_x)) \leq W(x)$. Thus

$$\begin{aligned} \min\{\, W(x) : x \in E^m, \widehat{f}(x) < f(f^*)\,\} &\geq \min\{\, W((i_x)) : x \in E^m, \widehat{f}(x) < f(f^*)\,\} \\ &= \min\{\, W(i) : i \in E, f(i) < f(f^*)\,\}. \end{aligned}$$

The reverse inequality is straightforward, so that in fact

$$\min\{\, W(x) : x \in E^m, \widehat{f}(x) < f(f^*)\,\} = \min\{\, W(i) : i \in E, f(i) < f(f^*)\,\}.$$

For any point j in f^*, we have $\Omega(j) = 0$ (by lemma 7.7) whence $\lim_{m\to\infty} W(W^*)/m = 0$ and

$$\lim_{m\to\infty} \frac{\bar{\alpha}_{\log}}{m} = \min\left\{ \frac{\Omega(i)}{H_1} : i \in E \setminus f^* \right\}.$$

For i not belonging to f^*, the coefficient $\Omega(i)$ is positive and its minimal value over $E \setminus f^*$ is precisely $\min(a, c\delta^*)$. □

The exponent $\bar{\alpha}_{\log}$ is the right object to compare with α_{opt}. Nevertheless, it may very well happen that

$$\min\left\{ \frac{\Omega(\pi)}{H_e(\pi)} : \pi \in \mathcal{C}(E^m),\, \pi \cap f^* = \emptyset \right\} \;>\; \frac{\min(a, c\delta^*)}{H_1}$$

and thus

$$\lim_{m\to\infty} \frac{\alpha_{\text{opt}}}{\bar{\alpha}_{\log}} \;>\; 1.$$

It is actually the case for the fitness landscape of figure 2 whenever $a > c\delta^*$.

References

1. O. Catoni, *Large deviations for annealing*, PhD Thesis, University Paris XI (1990).
2. O. Catoni, *Rough large deviations estimates for simulated annealing. Application to exponential schedules*, The Annals of Probability **20** (1992), 1109–1146.
3. R. Cerf, *Asymptotic convergence of genetic algorithms*, Preprint, submitted (1993).
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. C.R. Hwang–S.J. Sheu, *Singular perturbed Markov chains and exact behaviours of simulated annealing process*, Journal of Theoretical Probability **Vol. 5 No. 2** (1992), 223–249.
8. A. Trouvé, *Convergence optimale pour les algorithmes de recuits généralisés*, C. R. Acad. Sci. Paris **t.315 Série I** (1992), 1197–1202.
9. A. Trouvé, *Cycle decompositions and simulated annealing*, Preprint (1993).
10. A. Trouvé, *Parallélisation massive du recuit simulé*, PhD Thesis, University Paris XI (1993).
11. A. Trouvé, *Rough large deviation estimates for the optimal convergence speed exponent of generalized simulated annealing algorithms*, Preprint (1993).

Raphaël Cerf, Université Montpellier II, Département des Sciences Mathématiques, case 051, Place Eugène Bataillon, 34095 Montpellier cedex 05–France

E-mail address: Raphael.Cerf@ens.fr or cerf@helios.montpellier.inra.fr