# Perturbation theory for Evolutionary Algorithms: towards an estimation of convergence speed.

Yann LANDRIN-SCHWEITZER and Evelyne LUTTON

*Projet Fractales — INRIA Rocquencourt, B.P. 105, 78153 Le Chesnay Cedex, France*
*Yann.Landrin-schweitzer@inria.fr, Evelyne.Lutton@inria.fr*
*Tel : +33 (0)1 39.63.55.52 — Fax : +33 (0)1 39.63.89.95*
`http://www-rocq.inria.fr/fractales`

**Abstract.** When considering continuous spaces EA, a convenient tool to model these algorithms is perturbation theory. In this paper we present preliminary results, derived from Freidlin-Wentzell theory, related to the convergence of a simple EA model. The main result of this paper yields a bound on sojourn times of the Markov process in subsets centered around the maxima of the fitness function. Exploitation of this result opens the way to convergence speed bounds with respect to some statistical measures on the fitness function (likely related to irregularity).

## 1 Introduction

While strong results have been obtained in a recent past concerning the convergence behaviour of particular applications of Evolutionary Algorithms (EAs), especially in the discrete case, no generic theory has been proposed to deal at once with a wider variety of frameworks.

In this paper, we will show how classical stochastic analysis tools can be used in order to build a model of EA, first step toward a theoretical toolkit that could apply to a wide range of evolutionary problems.

The main mathematical results we will rely on are inspired by Freidlin and Wentzell's [1] fundamental work about stochastic perturbations of dynamic systems, a necessary tool to obtain global time-related equivalents on Markov processes. And EA modeling principles are based on Raphaël Cerf's [8] and Olivier François [10] works on GA convergence. The principles of Olivier François' MOSES model, using only selection or mutation in a discrete context, have been partly reused in this paper to build a continuous-space model of EA.

Considering a very simple model of EA, presented in section 2, with no crossover and a very basic selection scheme, enables us to focus on EAs' global behaviour along time, with affordable computational complexity.

With stronger hypotheses on exponential moments in the main theorem of Freidlin and Wentzell's perturbation theory, we prove a global bound on measures (theorem 2, section 3), in place of an asymptotic bound.

In section 4 this stronger version of the theorem is used to get an exponential bound on probabilities for populations to lie in well-chosen sets, yielding an existence theorem (theorem 3) on time-exponential bounds for sojourn times.

## 2 A Markov Model for Evolutionary Algorithms

### 2.1 Optimization model

In the following, we will consider stochastic processes, or random functions, which are time-indexed families of random variables on a probabilistic space $(\Omega, \mathbb{P})$, and more precisely, Markov processes. We aim at finding the maximum of a function $h : \mathbb{R}^n \mapsto \mathbb{R}^+$. We suppose that the problem is consistent:

- $h$ does not reach its upper bound at infinity, i.e. there does not exist a sequence $x_i$ of points in $\mathbb{R}^n$ such that $\lim_{i\to\infty} |x_i| = \infty$ and $\lim_{i\to\infty} h(x_i) = \sup_{\mathbb{R}^n} h$
- the set of maxima $\text{Argmax}\,(h) = \{x \in \mathbb{R}^n | h(x) = \sup_{\mathbb{R}^n} h\}$ is finite

An evolutionary algorithm can be defined as an operator acting on populations, i.e. subsets of $X = \mathbb{R}^n$. Sequences of populations $\{\xi_t\}$, indexed by the set $T$ of simulation times, a subset of $\mathbb{N}$, are produced by iterating this operator. Considering only populations of a fixed size $d$, an EA will be described as an operator $R$ from $X^d$ into itself. A new population $\xi_{t+1}$ is computed from the current one $\xi_t$ at time $t$ by applying $R$, so that $\xi_{t+1} = R(\xi_t)$. This algorithm will converge, from an initial population $\xi_o \in X^d$, if $\xi_t$ converges towards the maximal set of $h$ when $t$ goes to infinity, that is: $\lim_{t\to\infty} \max\{d(x, \text{Argmax}\,(h)), x \in \xi_t\} = 0$.

We can even suppose that the operator itself depends on $t$, we thus write $\xi_{t+1} = R_t(\xi_t)$. $R_t$ is usually a random operator.

### 2.2 Markov Model

The collection of populations over time is a discrete-time stochastic process $\xi_t \in X, t \in T$. At this point, assumptions on the random operator $R_t$ must be made. We will define a reproduction operator $R_t$ by an elementary transition probability $\Phi(t, x, \Gamma) = \mathbb{P}(\xi_{t+1} \in \Gamma | \xi_t = x)$, the probability for the population at time $t+1$ to be in a set of possible populations $\Gamma$ if the population at time $t$ was $x \in X$.
The existence of such a transition probability makes $\{\xi_t\}$ a Markov process.
We deduce from $\Phi$ a global transition probability $P(s, x, t, \Gamma) = \mathbb{P}(\xi_t \in \Gamma | \xi_s = x)$ to get from $x \in X$ at time $s \in T$ into $\Gamma$ at time $t \in T$:

$$P(s, x, s, \Gamma) = \mathbb{1}_\Gamma(x) \text{ ( that is 1 if } x \in \Gamma \text{ , 0 otherwise )}$$

$$\forall t \in T, x \in X, P(s, x, t+1, \Gamma) = \int_X \Phi(t, y, \Gamma) P(s, x, t, dy)$$

We will not actually work on these transition probabilities, but rather on the associated densities $\phi(t, x, \Gamma) = \int_\Gamma \phi(t, x, u)\, du$, and $P(s, x, t, \Gamma) = \int_\Gamma p(s, x, t, u)\, du$. Here are some basic properties ($\delta$ is the Dirac distribution centered at 0):

$$p(s, x, s, y) = \delta(y - x) \ , \ p(s, x, s+1, y) = \phi(s, x, y)$$

$$\text{and } p(s, x, t+1, y) = \int_X p(s, x, t, u) \phi(t, u, y)\, du$$

A operator $D_t$ can be introduced, defined for a measurable function $f$ from a vector space $E$ into $X$ by, $\forall t \in T, x \in X$:

$$D_t f(x) = \int_X f(u)\phi(t,u,x)\,du \tag{1}$$

If $f$ represents some criterion on the population at time $t$, then $D_t f$ is an estimation of this criterion at time $t+1$. We look forward in time, using the known values of $f$ on the current population together with the information on possible offsprings to compute this estimation. We can rewrite the above equations using $D_t$:

$$p(s,x,s,y) = \delta(y-x) \text{ and } p(s,x,t+1,y) = D\left(p(s,x,t,\bullet)\right)(y)$$

Moreover, since the populations are meant to be only sets, the order of coordinates on $X$ is irrelevant: any permutation must leave $\phi$ unchanged. If we define, $\forall \sigma \in \mathbb{S}_d$ (the set of all permutations on $d$ elements), $x_\sigma = \left(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(d)}\right)$, we can write this condition as: $\forall (x,u) \in X^2, \forall i \in [\![1,d]\!]$,

$$\phi(t,x,u) = \phi(t,x,u_\sigma) \text{ and } \phi(t,x,u) = \phi(t,x_\sigma,u)$$

### 2.3 Evolutionary Model

To shorten the proofs, we have restricted the model to individuals taken in $\mathbb{R}$, so that $X = \mathbb{R}^d$. This does not affect the model's generality. In the first approach described here, we will only use mutation to generate offsprings, so that one individual has one, and only one, offspring. This condition is very restrictive, and in doing this we are stepping back from the general definition of evolutionary algorithms, involving selection (individuals can have none or many offsprings) and crossover (an individual is born from at least two parents).

The initial population is generated with respect to a random isotropic law $Q$, that is in terms of distribution: $Q(\Gamma) = \int_\Gamma q(u)du$ where $q$ is isotrope. Defining $p(t,x), t \in T, x \in X$, by $\mathbb{P}(\xi_t \in \Gamma) = \int_\Gamma p(t,x)\,dx$, we get:

$$p(0,x) = q(x) \text{ and } p(t+1,x) = \int_X p(t,u)\phi(t,u,x)\,du$$

The mutation process can be split into two phases: the strength of the mutation of a particular individual, and the "shape" of the mutation itself. If we do not discriminate between individuals, we should fix once for all the shape at a given time. Keeping in mind we should use isotropic functions since populations are unordered sets, we can mutate individuals following a density $g_t$, with $g_t(-u) = g_t(u)$. Let $m(x)$ be a "mutation vector" in ${\mathbb{R}^+}^d$, where $m_i(x)$ is our mutation decision on the i-th individual in the population. $m_i(x) = 0$ stands for "no mutation at all", and the probability of mutations grows with $m_i(x)$. We can rewrite $\phi$ as:

$$\phi(t,x,y) = \prod_{m_i(x)=0} \delta(y_i - x_i) \prod_{m_i(x)\neq 0} g_t(y_i - x_i)$$

$$\text{or} \quad \phi(t,x,y) = \prod_{i=1}^{d} \frac{1}{m_i(x)} g_t\left(\frac{y_i - x_i}{m_i(x)}\right)$$

A very simple example of decision function can be defined as follows: let $\alpha$, $0 \leq \alpha \leq 1$ be a threshold value,

$$m_i(x) = \begin{cases} 0 \text{ if } h(x_i) - a(x) \geq \alpha(b(x) - a(x)) \\ 1 \text{ otherwise} \end{cases}$$

where $a(x)$ is the lowest value of the fitness function $h$ on the population $x$, and $b(x)$ its highest value. That means the individuals with fitness lower than the threshold $\alpha$ mutate.

We could also use a variation of elitist selection, by letting $m_i(x) = 0$ if $x_i$ is among the $\alpha.d$ better individuals in population $x$, and 1 otherwise.

But there is no need to restrict $m_i$ to binary values. For instance, $m_i(x) = \frac{h(x_i)}{\sum_{j=1}^{d} h(x_j)}$ could be used, which represents still another adaptation of the selection concept to our model.

We will need in the following a generic definition of this model:

▹**Definition 1**: **EA process**

Using the conventions stated above, we will call "EA process" with parameters $[q, \phi]$ a Markov process $\{X_t\}$, taking its values in $X = \mathbb{R}^d$ and its times in $\mathbb{R}_+$, with the properties:

$$\mathbb{P}\left(X_t \in \Gamma\right) = \int_{\Gamma} p(t,x)\ dx \tag{2}$$

$$p(0,x) = q(x) \text{ and } p(t+1,x) = \int_{X} p(t,u).\phi(t,u,x)\ du \tag{3}$$

We will call $q$ the initialization function of the EA, $\phi$ its transition function, and $p$ its density.

## 3 Perturbation theory

### 3.1 Exponential moments

Freidlin and Wentzell's perturbation theory makes an extensive use of exponential moments of the measures of interest. The measures $\mu^t$, where $\mu^t(\Gamma) = \mathbb{P}(\xi_t \in \Gamma) = \int_{\Gamma} p(t,x)\ dy$, are examined here. Their exponential moments are:

$$H^t(\alpha) = \ln \int_{X} e^{\langle \alpha, x \rangle} p(t,x) dx \tag{4}$$

As in section 2.2, we can introduce an operator $G_t$, defined for a measurable function $f$ from a vector space $E$ into $X$ by, $\forall t \in T, x \in X$:

$$G_t f(x) = \int_X f(u)\phi(t,x,u)\ du \tag{5}$$

(the variables are switched in comparison to the ones of the $D_t$ definition). $G_t$ does the opposite of $D_t$: instead of looking forward in time, trying to estimate the future value of some criterion $f$, it looks backward, estimating the value of this criterion on the previous population. We then write:

$$H^{t+1}(\alpha) = \ln \int_X p(t,u) G_t \left(e^{\langle \alpha, \bullet \rangle}\right)(u)\ du$$

Iterating this result, and letting $\Gamma_t f(x) = \ln G_t \left(e^f\right)(x)$, we get:

$$H^t(\alpha) = \ln \int_X q(x) e^{\left(\prod_{s=1}^{t} \Gamma_s\right)\left(e^{\langle \alpha, \bullet \rangle}\right)}(x) dx$$

### 3.2 Moments properties

The properties we are interested in apply to the iterates of the $\Gamma$ operator on linear functions. Let us define a function $\gamma_t$ from $\mathbb{R}$ into $\mathbb{R}_+$ by:

$$\gamma_t(s) = \ln \int_{\mathbb{R}} e^{s\ u} g_t(u)\ du \tag{6}$$

$\gamma_t$ is actually the exponential moment associated with $g_t$, the mutation law, and summarizes all the useful information on it from the point of view of exponential moments. We get:

$$H^{t+1}(\alpha) = \ln \int_X p(t,u) e^{\langle \alpha, u \rangle} e^{\sum_{i=1}^{d} \gamma_t(\alpha_i\ m_i(u))}\ du$$

where $d$ is the population's size, $p$ its density at time $t$, and $m$ the mutation decision.

Freidlin and Wentzell's perturbation theorem gives only asymptotic informations on a probability distribution. However, under somewhat more restrictive conditions and minor changes in the original proof, we have established an upper bound on the distributions, that applies globally instead of asymptotically (see appendix B for proofs).

Let us recall the Legendre's transform definition:

▷**Definition 2**: **Legendre's transform**

Given $H : X \to \mathbb{R}$ convex, the Legendre's transform of $H$ is, for $x \in X$:

$$\mathcal{L}H(x) = \sup_{u \in X} \left(\langle u, x \rangle - H(u)\right) \tag{7}$$

▷**Theorem 1**: **Uniform bound**

*Let* $\left\{\mu^h\right\}, h \in \mathbb{R}^+$ *be a family of probability measures on* $X$. *For* $\alpha \in X$, *we define* $H^h(\alpha) = \ln \int_X e^{\langle \alpha, x \rangle}\ d\mu^h(x)$. $H^h$ *is convex. Let us make the following hypotheses:*

– *for any fixed* $\alpha$, $h \mapsto H^h(\alpha)$ *is a non-increasing function*
– *there exists some non-increasing function* $\lambda : \mathbb{R}_+^* \to \mathbb{R}_+^*$ *such that:*
  • $\lim_{h\to 0} \lambda(h) = \infty$
  • $\forall \alpha \in X, H(\alpha) = \lim_{h\to 0} \frac{1}{\lambda(h)} H^h(\lambda(h)\alpha)$ *exists and the limit* $H$ *is* $\mathcal{C}^1$

*Then for* $r \in \mathbb{R}, v \in X$, *we have:* $\mu^h(\{x \in X, \langle v, x\rangle \geq r\}) \leq e^{-\lambda(h)\ \inf_{\langle v,x\rangle \geq r} \mathcal{L}H(x)}$

**Note 1**: To preserve Freidlin and Wentzell's notations, we have used here a parameter $h$ to index the family of measures, and examined their behavior as $h \to 0$. However, to be consistent with the indexing of populations by $t$ in EAs, results will be stated using $t \to \infty$ in the following.
**Note 2**: In the following we will respectively speak, for $\mathcal{L}H$ and $\lambda$, of an "action functional" and a normalization factor for the familly of measures $\{\mu^h\}$.

We are specially interested in the case $v = \pm e_i$, where $e_i$ is the base vector with null components except for the coordinate $i$: we then have $x_i \leq r$ or $x_i \geq r$. Let us suppose the conditions of the theorem are fulfilled. We will build hypercubic "balls", by excluding $2.d$ half-spaces. These half-spaces being defined by some relation like $x_i > r$, a permutation on the coordinates doesn't change anything, and we will get an exponential upper bound on the probability to be outside this ball. Taking into account sections 3.1 and 3.2's results on the $G$ operator (equation 5), we get:

▷**Theorem 2**: **Probability asymptotics**
*Let us examine an EA process with isotropic initialization function* $q$, *and with transition function* $\phi$, *with:*

$$\phi(t,x,y) = \prod_{i=1}^{d} \frac{1}{m_i(x)} g_t\left(\frac{y_i - x_i}{m_i(x)}\right)$$

*If there exists:*

– *an non-decreasing function* $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$, *with* $\lim_{h\to 0} \lambda(h) = \infty$,
– *a function* $\alpha \mapsto \overline{H}(\alpha)$ *such that* $\forall \alpha, \frac{1}{\lambda(t)} H^t(\lambda(t)\alpha) \leq \overline{H}(\alpha)$,

*then* $\forall \alpha \in X, H(\alpha) = \lim_{h\to 0} \frac{1}{\lambda(h)} H^h(\lambda(h)\alpha)$ *exists. And if we let*
$A = \{x \in X | \exists i, |x_i - x_o| \geq r\}$, *we get:* $P(t, A) \leq 2.d.e^{-\lambda(t)\ \inf_A \mathcal{L}H(x)}$

# 4 Convergence speed

## 4.1 Particular values of $g_t$

In order to use theorem 2, let us assume that our mutation kernel $g_t$ has the following form: $g_t(u) = k_t g(k_t.u)$ where $g = \frac{1}{2}\mathbb{1}_{[-1,1]}$. This yields: $\gamma_t = \gamma\left(\frac{s}{k_t}\right)$, where: $\gamma(s) = \ln \frac{1}{2}\int_{-1}^{1} e^{su}\, du = \ln \sinh s - \ln s$.

Let us recall that $m(u)$ is the mutation decision vector, where $m_i(u)$ stands for the mutation rate of the $i$-th individual in population $u$. As $0 \leq \gamma(s) \leq s$, we obtain:

$$H^{t+1}(\alpha) = \ln \int_X p(t,u) e^{\langle \alpha, u \rangle} e^{\sum_{i=1}^{d} \gamma\left(\frac{\alpha_i \ m_i(u)}{k_t}\right)} \ du$$

$$H^t(\alpha) \leq H^{t+1}(\alpha) \leq \ln \int_X p(t,u) e^{\left\langle \alpha, u + \frac{m(u)}{k_t} \right\rangle} \ du$$

### 4.2 How to get an usable action functional $\mathcal{L}H$?

The previous relation shows which criterion on $k_t$ and $m$ implies the existence of some $\lambda$ for which the limit exists, and for which theorem 2 is not trivially verified:

- If $H^0(\alpha) > C|\alpha|^{1+\epsilon}$, the limit $H(\alpha)$ will not exist.
- If $\sum_{s=0}^{\infty} \frac{1}{k_t}$ converges, $H^0(\alpha) \leq H^t(\alpha) \leq H^0(\alpha) + |\alpha| \sum_{s=0}^{\infty} \frac{1}{k_t}$, and $\mathcal{L}H$ is infinite almost everywhere.

The limit will exist if $\sum_{t=0}^{\infty} \frac{1}{k_t} \left\langle \alpha, \int_X p(t,u) m(u) \ du \right\rangle$ is finite.
When $\frac{1}{\lambda(t)} H^t(\lambda(t)\alpha)$ converges towards $H(\alpha)$, the upper bound theorems apply. To get a non-trivial result, we need that $\infty > \inf_{A_{r,x_o}} \mathcal{L}H(x) > 0$, which will be true, if $H$ is convex, for large enough radius $r$. Finally, we get an exponential bound on probabilities of being outside a disk $D$ of radius $r$ around an optimum.

### 4.3 Sojourn times

The sojourn time of an EA process in a region $D$ of the populations space $X$, for a given simulation (that is, a run of the algorithm), can be defined as:

$$\theta(D) = \sum_{t=0}^{\infty} \mathbb{1}_D(\xi_t)$$

This random variable can be understood as the number of elementary time units during which the population is inside $D$. If theorem 2 applies, we get the immediate upper bound:

$$\mathrm{E}\left(\theta(D)\right) = \sum_{t=0}^{\infty} P\left(t, X \backslash D\right) \leq \sum_{t=0}^{\infty} 2.d.e^{-\lambda(t) \ \inf_{X \backslash D} \mathcal{L}H(x)}$$

that, together with Chebychev's inequality, provides an estimate of the time after which the EA has "converged" into $D$:

▷**Theorem 3**: **Convergence speed**
*If theorem 2 (probability asymptotics) applies, then:*

$$\mathbb{P}\left(\sum_{0}^{\infty} \mathbb{1}_{X \backslash D}(\xi_t) dt > l\right) \leq \frac{1}{l} \sum_{t=0}^{\infty} 2.d.e^{-\lambda(t) \ \inf_{X \backslash D} \mathcal{L}H(x)}$$

In other words, the probability to be outside $D$ during more than $l$ time units is bounded by $\frac{1}{l}\mathrm{E}\left(\theta(D)\right)$.

## 5 Limits and prospects

The use of the convergence theorems presented above is tricky: without an explicit form of $\lambda$ in the general case, we need to treat this problem for every particular case, taking into account the exact mutation operator (that is, the properties of $m$) and the characteristics of the objective function.

The critical role of the values $\frac{1}{k_t}\left\langle \alpha, \int_X p(t,u)m(u)\ du \right\rangle$ emphasizes the importance of the regularity of the $m$ decision factor, strongly linked to the regularity of the fitness function $h$. An analysis of the impact of some regularity measure on the values of $\lambda$, $H$ and $\mathcal{L}H$ is first to be done. There are evidences that "fractal" quantities such as Legendre multifractal spectra have common features with quantities involved in this study. Future works will also concern the analysis of irregularities for some constrained classes of fitness functions, as in [11].

Finally, we worked on a voluntarily simplified framework respectively to the traditional models of EAs. Further efforts will also concern a more realistic EA model involving crossover and advanced selection schemes.

## References


1. M. I. Freidlin, A. D. Wentzell, "*Random perturbations of dynamical systems.*" Trad. de : "Fluktuatsii u Dinamicheskikh Sistemakh Pod Deistviem Malykh Sluchainykh Vozmushchenii". Springer, 1984.
2. M. Jerrum, "*Mathematical Foundations of the Markov Chain Monte Carlo Method*". In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, B. Reed (Eds.), Probabilistic Methods for Algorithmic Discrete Mathematics, Springer Verlag, 1998.
3. E. Aarts and P. Van Laarhoven, "Simulated annealing : a pedestrian review of the theory and some applications". NATO, *ASI Series*, Vol. F30, Springer Verlag.
4. E. Lutton and J. Lévy Véhel, Some remarks on the optimization of Hölder functions with Genetic Algorithms, IEEE trans on Evolutionary computing, 1998, see also Inria Research Report No 2627.
5. B. Leblanc, and E. Lutton. Bitwise Regularity Coefficients as a Tool for Deception Analysis of a Genetic Algorithm, 1997, INRIA research report, RR-3274.
6. A. Agapie. Genetic algorithms : Minimal conditions for convergence. In *Artificial Evolution, European Conference, AE 97, Nimes, France, October 1997,*. Springer Verlag, 1997.
7. Lee Altenberg. The Schema Theorem and Price's Theorem . In *Foundation of Genetic Algorithms 3*, 1995. ed. Darrell Whitley and Michael Vose. pp. 23-49. Morgan Kaufmann, San Francisco.
8. R. Cerf. *Artificial Evolution, European Conference, AE 95, Brest, France, September 1995, Selected papers*, volume Lecture Notes in Computer Science 1063, chapter Asymptotic convergence of genetic algorithms, pages 37–54. Springer Verlag, 1995.
9. Thomas E. Davis and Jose C. Principe. A Simulated Annealing Like Convergence Theory for the Simple Genetic Algorithm . In *Proceedings of the Fourth International Conference on Genetic Algorithm*, pages 174–182, 1991. 13-16 July.
10. O. François "*An Evolutionary Strategy for Global Minimization and its Markov Chain Analysis* IEEE transactions on evolutionary computation, septembre 1998
11. Evelyne Lutton. *Genetic Algorithms and Fractals* . In *Evolutionary Algorithms in Engineering and Computer Science*, K. Mietttinen, M. M. Mäkelä, P. Neittaanmaki, J. Périaux, Eds, John Wiley & Sons, 1999.
12. G. Rudolph. Asymptotical convergence rates of simple evolutionary algorithms under factorizing mutation distributions. In *Artificial Evolution, European Conference, AE 97, Nimes, France, October 1997,*. Springer Verlag, 1997.

13. A. Trouvé, *Rough large deviation estimates for the optimal convergence speed exponent of generalized simulated annealing algorithms.* 1993
14. M.D. Vose. Formalizing genetic algorithms. In *Genetic Algorithms, Neural Networks and Simulated Annealing Applied to Problems in Signal and Image processing.* The institute of Electrical and Electronics Engineers Inc., 8th-9th May 1990. Kelvin Conference Centre, University of Glasgow.
15. Y. Landrin-Schweitzer, E. Lutton. Perturbation theory for Evolutionary Algorithms. Inria Research Report, 2000.

## Appendix A: Notations

| | |
|---|---|
| $[\![m, m']\!]$ | set of the integers $i \in \mathbb{N}, m \leq i \leq m'$ |
| $[x, x']$ | set of the reals $u \in \mathbb{R}, x \leq u \leq x'$ |
| $Y^X$ | set of the functions $f : X \mapsto Y$ |
| $\mathbb{S}_d$ | permutations group on $[\![m, m']\!]$ |
| $\mathbb{1}_X$ | characteristic function of a set $X$ |
| $\mathbb{P}(E)$ | probability of an event $E$ |
| $d$ | population size |
| $X$ | set of all individuals |
| $p(t, x)$ | probability density to have a population $x$ at time $t$ |
| $q(x)$ | probability density for the initial population |
| $\phi(t, u, x)$ | probability density to produce a population $x$ at time $t+1$ if the population at time $t$ is $u$ |

## Appendix B: Sketch of the proofs

### Uniform bound theorem

We begin by stating here some lemma usefull in the proof of our main theorem, and whose proofs, rather straightforward, can be found in [15].

▷**Lemma 1**: **Hyperplane separation, real case**
*Let $H$ be a convex function on $\mathbb{R}$, such that $H(0) = 0$ and $\mathcal{L}H$ be strictly convex. Let $A$ be a range in $\mathbb{R}$, with $s(A) = \inf_{x \in A} \mathcal{L}H(x)$. With these assumptions, $\forall \gamma \in \mathbb{R}_+^*$, there exists some $w \in \mathbb{R}$ verifying $\forall x \in A, w.x - H(w) \geq s(A) - \gamma$*

▷**Lemma 2**: **Hyperplane separation**
*Let $H \in U(X)$, $r \in \mathbb{R}$ and $v \in X$ such that $\forall t \in \mathbb{R}, \forall q \in X \mid \langle v, q \rangle = 0$, we have $H(tv + q) = H(tv - q)$ (that means $v$ is a symmetry axis for $H$). If $A = \{x \in X, \langle v, x \rangle \geq r\}$, $s(A) = \inf_{x \in A} \mathcal{L}H(x)$, $\forall \gamma \in \mathbb{R}_+^*$ there exists $w \in X$ verifying: $\forall x \in A, \langle w, x \rangle - H(w) \geq s(A) - \gamma$*

▷**Lemma 3**: **Limit from below**
*Let $H^h : X \to \mathbb{R}$ convex such that $\forall \alpha \in X, \forall (h, h') \in {\mathbb{R}_+^*}^2$ with $h' \leq h$, we have $H^{h'}(\alpha) \geq H^h(\alpha)$. Let $\lambda \in (\mathbb{R}_+^*)^{\mathbb{R}_+^*}$, strictly decreasing, verifying:*
- $\lim_{h \to 0} \lambda(h) = \infty$
- *$\forall \alpha \in X, H(\alpha) = \lim_{h \to 0} \frac{1}{\lambda(h)} H^h(\lambda(h)\alpha)$ exists*

*Then $\forall \alpha \in X, H(\alpha) = \lim_{h \to 0} \frac{1}{\lambda(h)} H^h(\lambda(h)\alpha)$ is reached from below.*

**We are now able to prove theorem 1:**

**Proof** Let $s = \inf_{\langle v, x \rangle > r} \mathcal{L}H(x)$. If $s = \infty$, $\mu^h(\{x \in X, \langle v, x \rangle > r\}) = 0$
In the other case, let $w \in X$ such that $\forall x \mid \langle v, x \rangle > r, \langle w, x \rangle - H(w) \geq s$.

Using the exponential Chebychev's inequality, we get:

$$\mu^h\left(\{x \in X, \langle v, x\rangle > r\}\right) \leq \mu^h\left(\{\langle w, x\rangle - H(w) \geq s\}\right)$$
$$\leq \int_X e^{\lambda(h)(\langle w,x\rangle - H(w) - s)} \leq e^{\lambda(h)\left(\frac{1}{\lambda(h)} H^h(\lambda(h)w) - H(w)\right)} e^{-\lambda(h)s} \leq e^{-\lambda(h)s}$$

since the $H$ limit is reached from below. □

## Application to EAs

▷**Lemma 4**: **Exponential moments' properties I**
*Let us consider an EA process with density $p$ and transition function $\phi$. For $\alpha \in X$, we define $H^t(\alpha) = \ln \int_X e^{\langle \alpha, x\rangle} p(t, y) dy$*
*Let $E$ be a vector space, and $f \in E^X$ measurable, $G_t f(x) = \int_X f(u)\phi(t, x, u)\, du$, and $\Gamma_t f(x) = \ln \int_X e^{f(u)} \phi(t, x, u)\, du = \ln G_t\left(e^f\right)(x)$*
*Then $H^{t+1}(\alpha) = \ln \int_X p(t, u) G_t\left(e^{\langle \alpha, \bullet\rangle}\right)(u)\, du$, and $H^t(\alpha) = \ln \int_X q(x) e^{\left(\prod_{s=1}^t \Gamma_s\right)\left(e^{\langle \alpha, \bullet\rangle}\right)}(x) dx$.*

▷**Lemma 5**: **Exponential moments' properties II**
*Let us consider an EA process with density $p$ and transition function $\phi$ given by the expression:*

$$\phi(t, x.y) = \prod_{i=1}^d \frac{1}{m_i(x)} g_t\left(\frac{y_i - x_i}{m_i(x)}\right)$$

*Then $H^{t+1}(\alpha) = \ln \int_X p(t, u) e^{\langle \alpha, u\rangle} e^{\sum_{i=1}^d \gamma_t(\alpha_i\; m_i(u))}\, du$, where $\gamma_t(s) = \ln \int_{\mathbb{R}} e^{s\; u} g_t(u)\, du$.*

▷**Lemma 6**: **Exponential moments' properties III**
*Let us consider an EA process with isotropic initialization function $q$, and with transition function $\phi = \prod_{i=1}^d \frac{1}{m_i(x)} g_t\left(\frac{y_i - x_i}{m_i(x)}\right)$, with a mutation decision $m$ verifying: $\forall \sigma \in \mathbb{S}_d$, $\forall x \in X, \forall i \in [\![1, d]\!], m_i(x_\sigma) = m_{\sigma(i)}(x)$ (with $x_\sigma = \left(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(d)}\right)$).*
*Then $\forall \sigma \in \mathbb{S}_d$, $H^t(\alpha) = H^t(\alpha_\sigma)$.*

**Using lemma 4, 5 and 6, we can now prove theorem 2:**

**Proof** Lemma 5, taking into account that $\gamma_t(s) \geq 0$, yields $H^{t+1}(\alpha) \geq H^t(\alpha)$.
We then use lemma 3 to prove that $\frac{1}{\lambda(t)} H^t(\lambda(t)\alpha)$ is increasing, so it converges if it remains bounded, and theorem 1 applies. Since $H$ is unchanged when coordinates are permutated, $\inf_{\langle \pm e_i, x\rangle \geq r} \mathcal{L}H(x)$ does not depend on $i$.
Let $A_i^+(r) = \{x \in X | x_i \geq r\}$, and $A_i^-(r) = \{x \in X | x_i \leq r\}$. $\inf_{A_i^+(r)} \mathcal{L}H(x)$ does not depend on $i$, and $\inf_{A_i^-(r)} \mathcal{L}H(x)$ neither.
If $A = \{x \in X | \exists i, |x_i - x_o| \geq r\}$, then $A = \bigcup_i A_i^+(x_o + r) \bigcup_i A_i^-(x_o - r)$, and:

$$\inf_A \mathcal{L}H(x) = \min\left(\inf_{A_i^+(x_o+r)} \mathcal{L}H(x), \inf_{A_i^-(x_o-r)} \mathcal{L}H(x)\right)$$

$$\text{We thus get} \quad P(t, A) \leq 2.d.e^{-\lambda(t)\; \inf_A \mathcal{L}H(x)}$$

□