Statistical distribution of the convergence time for longpath problems

Josselin Garnier and Leila Kallel

CMAP – UMR CNRS 7641 , Ecole Polytechnique,
Palaiseau 91128, France.
E-mail: *garnier@cmapx.polytechnique.fr*
E-mail: *kallel@cmapx.polytechnique.fr*

July 7, 1998

Abstract

The asymptotical behavior of a (1+1)-ES process on Rudolph's long k-paths is extensively investigated in this paper. First, in the case of $k = l^{\alpha}$, we prove that the long k-path is a longpath for the (1+1)-ES, in the sense that the entire path has to be followed before convergence occurs. For $\alpha < 1/2$, expected convergence time is still exponential but some shortcuts will occur meanwhile which speeds up the process.

Second, in the case of constant k, the statistical distribution of convergence time is calculated, and the influence of population size is investigated for different $(\mu+\lambda)-ES$. Besides, the histogram of the first hitting time of the solution shows an anomalous peak close to zero, which corresponds to an exceptional set of events that speed up the expected convergence time with a factor of l^2. A direct consequence of this exceptional set is that performing independent (1+1)-ES processes proves to be more advantageous than any population based $(\mu + \lambda) - ES$.

Introduction

Long paths are unimodal problems with only one path (in hamming space) to the optimum. The length of this path is *exponential*[1]. A random mutation Hill Climber *can* solve it. But it takes an exponential time in string length to reach the solution, then we cannot wait for it.

Long path problems [6] have been introduced in GA literature to precise the notion of problem difficulty for optimization algorithms. Unimodality have so been added to the list of well known difficulties: Isolation (*needle in a haystack*), deception, and multimodality -studied in [4] [3] [2]- are no longer the only responsible for search difficulty.

Horn's longpath [6] for the single-bit-flip hill-climber is an example of such a fitness function. It is a sequence of strings with the property that two successive strings are at Hamming distance 1 from each other. The k-th string in the sequence has a fitness value of k. All strings s which are not in the sequence get a fitness value of $(l - d)/l$, where d is the Hamming distance between the string s and the first string of the sequence.

[1] Unless the growth rate is explicitly stated, any asymptotic behavior beyond a polynomial growth in string length will be called *exponential*

In this way, a single-bit-flip hill-climber is first lead towards the first string of the path, and then through the whole path. Because of its length, Horn's longpath is an intractable search problem for the single-bit-flip hill-climber.
Rudolph [10] presents a generalization of Horn's longpath to a class of fitness functions called *long k-paths*.
The long k-path $P(l,k)$ is defined by a recursion on the length l of the strings, with $P(1,1) = \{0,1\}$ as the base path. Given the long k-path $P(l,k)$, $P(l+k,k)$ is constructed as follows: A sub-path S_0 of $P(l+k,k)$ is created by taking the path $P(l,k)$ and by prepending k zeros to each point on $P(l,k)$. A sub-path S_1 of $P(l+k,k)$ is created by taking $P(l,k)$ *in reverse order* and prepending k ones to each point on the reversed path. A third path B consists of $k-1$ points created by prepending the following substrings of length k to the last point in the path $P(l,k)$: $0\ldots01$, $0\ldots011$, $\ldots$, $001\ldots11$ and $01\ldots11$. The resulting long k-path $P(l+k,k)$ on strings of length $l+k$ is then defined by the concatenation of S_0, the "bridge path" B and S_1. Horn's longpath is the long k-path for $k=2$.
As an example, consider the construction of the path $P(7,3)$. Starting from $P(1,3) = \{0,1\}$, we obtain the sequence

$$0000_1 \quad 0001_2 \quad 0011_3 \quad 0111_4 \quad 1111_5 \quad 1110_6 \tag{1}$$

from which the path $P(7,3)$ is constructed:

$$\begin{array}{lll} 0000000_1 & 0001110_6 & 1110111_{11} \\ 0000001_2 & 0011110_7 & 1110011_{12} \\ 0000011_3 & 0111110_8 & 1110001_{13} \\ 0000111_4 & 1111110_9 & 1110000_{14} \\ 0001111_5 & 1111111_{10} & \end{array} \tag{2}$$

- The Hamming distance between two consecutive points of the path is 1.
- A mutation of $i < k$ bits can only lead to a point which is i positions away from the original point. To jump over more than k consecutive positions in the path, all the bits of at least one of the blocks $2j+1,\ldots,2(j+1)$, for $j \in \{0,\ldots,(l-1)/k\}$, have to be flipped. Such a jump will be called a *shortcut*.
- The length of the long k-path is given by $|P(l,k)| = (k+1)2^{(l-1)/k} - k + 1$. Note that for a fixed string length, the path becomes shorter with increasing k.

As to the behavior of the GA with only one-point crossover and no mutation, experiments carried out in [5] demonstrate that a GA with complementary crossover (a one-point crossover between a string and its complement) exploits a kind of Royal Road structure (of Mitchell and al. [8] [9]) present in the long k-path. This results in the occurrence of many shortcuts, and a growth rate of l^2 for the time to reach the optimum of the long k-path. The authors expect that the complementary crossover approximates the one-point crossover if the latter is used in combination with a big enough population size. However, no conclusion can be drawn about the asymptotic behavior of the GA.

Rudolph [10] gives $O(l^{k+1}/k)$ as an asymptotic upper bound on the expected time spent by a (1+1)-ES with $1/l$ mutation rate on a long k-path. For a fixed value of k, this bound is a polynomial of the string length. It shows that for large values of l the (1+1)-ES is a lot quicker than the single-bit-flip hill-climber in optimizing the path. The difference between

the two algorithms is that the (1+1)-ES can do more than one bit-flip at the same time. Note that in the case of k being a function of l, as in $k = \sqrt{l}$, the upper bound changes to an uninteresting exponential. So the problem of convergence time with ($k = l^\alpha$, $\alpha > 0$) remained unsolved.

In this paper, the exact expected time for convergence is calculated for k constant and l dependent. In the case of constant k, the statistical distribution of convergence time and its variance are given. This allows to study the influence of population size (as a function of l) on the expected convergence time, with $(\mu + \lambda) - ES$ strategies.

The first section illustrates the construction of the long k-path, and defines some notations. Sections 3 is devoted to the study of expected convergence time for $k = l^\alpha$. Section 4 considers the case of constant k: expected convergence time and its fluctuations are calculated. In the light of these results, different $(\mu+\lambda) - ES$ are compared. The influence of mutation choice is also discussed.

1 Construction of the (k,l)-path.

We first construct the (1,l)-paths recursively with respect to l. Let $P_{1,l}$ be the (1,l)-path. Create sub-path S_0 by prepending 0 to each string in path $P_{1,l}$ and sub-path S_1 by prepending 1 to each string in the reverse of path $P_{1,l}$. Finally concatenate sub-paths S_0 and S_1 to obtain the (1,l+1)-path.
We can now build the (k,l)-path for any k as follows.

1) Let us start with the path $P_{1,(l-1)/k}$, which contains $2^{(l-1)/k}$ strings with length $(l-1)/k$.

$$\text{For } \frac{l-1}{k} = 3: \qquad \begin{array}{ccccc} & & 0 & & 000 \\ & 0 < & & & 001 \\ & & 1 & & 011 \\ 0 < & & & \Longleftrightarrow & \\ & 1 < & 0 & & 010 \\ & & 0 & & 110 \\ & 1 < & & & \\ & & 1 & & 111 \\ 1 < & & & & \\ & & 1 & & 101 \\ & 0 < & & & \\ & & 0 & & 100 \end{array}$$

2) Let us substitute for each 0 (resp. each 1) a substring with length k consisting of 0 (resp. 1). You then get a sequence of $2^{(l-1)/k}$ strings with length $l-1$.

$$\text{For } \frac{l-1}{k} = 3: \qquad \begin{array}{ccccc} & & 0\ldots0 & & 0\ldots0 \\ & 0\ldots0 < & & & 1\ldots1 \\ & & 1\ldots1 & & 1\ldots1 \\ 0\ldots0 < & & & & \\ & 1\ldots1 < & 0\ldots0 & & \\ & & 1\ldots1 & & 0\ldots0 \\ 1\ldots1 < & & & & \\ & & 0\ldots0 & & 1\ldots1 \\ & 0\ldots0 < & & & \\ & & 1\ldots1 & & 0\ldots0 \end{array}$$

3) You add to each leave an ultimate branching consisting alternatively of $< \begin{matrix} 0 \\ 1 \end{matrix}$ and $< \begin{matrix} 1 \\ 0 \end{matrix}$.
You then get the so-called skeleton of the (k,l)-path, which consists of $2^{(l+k-1)/k}$ strings with length l. This is actually a sub-sequence extracted from the (k,l)-path as built in [10].

$$
\text{For } \frac{l-1}{k} = 3: \quad
\begin{array}{l}
0\ldots0 < \\
\qquad 0\ldots0 < \\
\qquad\qquad 0\ldots0 < \begin{matrix} 0 \\ 1 \end{matrix} \\
\qquad\qquad 1\ldots1 < \begin{matrix} 1 \\ 0 \end{matrix} \\
\qquad 1\ldots1 < \\
\qquad\qquad 0\ldots0 < \begin{matrix} 0 \\ 1 \end{matrix} \\
\qquad\qquad 1\ldots1 < \begin{matrix} 1 \\ 0 \end{matrix} \\
1\ldots1 < \\
\qquad 1\ldots1 < \\
\qquad\qquad 0\ldots0 < \begin{matrix} 0 \\ 1 \end{matrix} \\
\qquad\qquad 0\ldots0 < \begin{matrix} 1 \\ 0 \end{matrix}
\end{array}
$$

4) To get the whole path from the skeleton, you simply need to add bridges between each two strings of the skeleton. A bridge consists of $k-1$ strings which contain either the sequence of substrings $*0....01*$, $*0...011*$, $*001..11*$, and $*01...11*$ or the inverse one $*01...11*$, $*001..11*$, $*0...011*$, and $*0....01*$. You then build the whole (k,l)-path with $(k+1)2^{(l-1)/k} - k + 1$ strings.

2 The process $X(t)$

The definition of the (1+1)-ES considered in this paper is given below. Note that the whole study supposes the search starts at the first string of the path. Let us denote by $\mathbf{e}_i := 0.....0$ the first string of the path and by $\mathbf{e}_f := 1..10....0$ the last one.
The search follows a process $X(t)$ starting from $X(0) = \mathbf{e}_i$. At each iteration time $t \in \mathbb{N}$, you build some mutated string $X'(t)$ from $X(t)$ by inverting each bit of $X(t)$ with probability l^{-1} independently. If $X'(t)$ lies in the path farther than $X(t)$, then $X(t+1) = X'(t)$ otherwise $X(t+1) = X(t)$. The last string $\mathbf{e}_f$ of the path is therefore the unique absorbing state of the Markov chain X.

In the following T stands for the first hitting time of the last string $\mathbf{e}_f$ of the (k,l)-path starting from the first string $\mathbf{e}_i$.

3 Asymptotic study for large l and $k = l^\alpha$

3.1 Main results

This section considers the case of $k = l^\alpha$ and large l values. Recall that the case of $k = O(l)$ is not interesting because it yields a path length of $O(l)$ and an expected number of trials of no more than $O(l^2)$. The main result of the section is given in proposition 3.1. It states that expected convergence time is exponential for $k = l^\alpha$.
In subsection 3.1, some simple estimates show that shortcuts speed up the convergence only in the case of $\alpha < 1/2$. In case $\alpha \geq 1/2$, the path is entirely visited before any shortcut is likely to happen.

PROPOSITION **3.1**
Let us assume that $k = [l^\alpha]$ *for* $\alpha \in (0, 1)$.[2]
1. If $\alpha < 1/2$, *then*

$$\lim_{l\to\infty} \frac{1}{l^\alpha \ln(l)} \ln \mathbb{E}[T] = 1. \tag{3}$$

2. If $\alpha \geq 1/2$, *then*

$$\lim_{l\to\infty} \frac{1}{l^{1-\alpha} \ln(2)} \ln \mathbb{E}[T] = 1. \tag{4}$$

Before turning to the proof we give some results about the waiting times of some shortcuts.

3.2 Waiting times of some particular events

Let us introduce the family $\mathcal{S}$ which consists of $\frac{l-1}{k}$ shortcuts denoted by $\mathcal{S}_j$ for $j = 1, ..., \frac{l-1}{k}$. The shortcut $\mathcal{S}_j$ is the operator which inverts the k bits whose labels lie within the segment $[(j-1)k + 1, jk + 1]$.
We also introduce the family $\mathcal{S}'$ which consists of all the shortcuts which invert at least the k successive bits whose labels lie within one of the intervals $[(j-1)k + 1, jk + 1]$, $j = 1, ..., (l-1)/k$. Note that some of the shortcuts of the family $\mathcal{S}'$ do not give rise to a jump for the process $X(t)$ but we don't care. Indeed we clearly have $\mathcal{S} \subset \mathcal{S}'$ and $\mathcal{S}'$ actually contains much more shortcuts than $\mathcal{S}$. Nevertheless the waiting time for one particular event of $\mathcal{S}$ or for any event of $\mathcal{S}'$ are almost the same when $k \sim l^\alpha$, of the order of l^{l^α}.

LEMMA **3.2**
1. If $l \gg 1$, *then for any fixed* j, *the waiting time* $T_{\mathcal{S}_j}$ *for the shortcut* $\mathcal{S}_j$ *obeys an exponential distribution with mean* el^k.
2. The waiting time T_{seq} *for the sequence* $\mathcal{S}_1$,..., $\mathcal{S}_{\frac{l-1}{k}}$ *(which means that we wait for* $\mathcal{S}_1$, *and then we wait for* $\mathcal{S}_2$, *and so on until* $\mathcal{S}_{\frac{l-1}{k}}$*) obeys a* $\chi^2(2\frac{l-1}{k})$ *distribution with mean* $el^k(l-1)/k$ *and density:*

$$p_{T_{seq}}(t) = \frac{1}{\Gamma(\frac{l-1}{k}) el^k} \left(\frac{t}{el^k}\right)^{\frac{l-1}{k}-1} \exp -\frac{t}{el^k}.$$

3. The waiting time $T_{\mathcal{S}'}$ *for some shortcut of the family* $\mathcal{S}'$ *obeys an exponential distribution with mean* $l^k k/(l-1)$.

[2] $[x]$ is the integer part of a real number x.

Although $T_{\mathcal{S}_j}$, T_{seq} and $T_{\mathcal{S}'}$ are integer-valued, we describe their respective distributions by using the notion of "densities", which is usually reserved for real-valued random variables. As an example the statement "$T_{\mathcal{S}_j}$ has an exponential density with mean el^k" should be understood as follows:

$$\mathbb{P}(T_{\mathcal{S}_j} = t) = \frac{1}{el^k} \exp - \frac{t}{el^k} \text{ for any integer } t$$

Proof. At each time $t \in \mathbb{N}$, the probability that the shortcut $\mathcal{S}_j$ occurs is:

$$P_{\mathcal{S}_j} = l^{-k}(1 - l^{-1})^{l-k}.$$

Since the events at different times are independent, the statistical distribution of the first time $T_{\mathcal{S}_j}$ when $\mathcal{S}_j$ occurs is given by:

$$\mathbb{P}(T_{\mathcal{S}_j} > t) = (1 - P_{\mathcal{S}_j})^t$$

For $1 \ll k \ll l$, we have $P_{\mathcal{S}_j} = l^{-k}e^{-1}$ and consequently:

$$\mathbb{P}(T_{\mathcal{S}_j} \geq l^k t) = \exp(-e^{-1}t).$$

which proves the first statement of the Lemma. Since the exponential variables have no memory (see Appendix A), we have:

$$\mathbb{P}(T_{seq} > t) = \mathbb{P}(T_{\mathcal{S}_1} + ... + T_{\mathcal{S}_{\frac{l-1}{k}}} > t)$$

Furthermore the sum of $\frac{l-1}{k}$ independent exponential variables with mean m obeys the so-called $\chi^2(2\frac{l-1}{k})$ distribution with mean $\frac{l-1}{k}m$ [1, Vol 2, ch. I-3], which proves the second statement.
At each time $t \in \mathbb{N}$ and for any fixed j, the probability that some shortcut which inverts at least the k successive bits whose labels lie within one of the intervals $[(j-1)k+1, jk+1]$ occurs is:

$$P_{\mathcal{S}'_j} = l^{-k}.$$

As above this proves that the statistical distribution of the first time $T_{\mathcal{S}'_j}$ when one of these shortcuts occurs is given by:

$$\mathbb{P}(T_{\mathcal{S}'_j} \geq l^k t) = \exp(-t).$$

The third statement of the proposition then follows from the fact that the infimum of $\frac{l-1}{k}$ independent exponential variables with mean m obeys an exponential distribution with mean $m\frac{k}{l-1}$:

$$\mathbb{P}\left(\inf_{j=1,...,(l-1)/k} T_{\mathcal{S}'_j} \geq l^k t\right) = \mathbb{P}(T\mathcal{S}'_1 \geq l^k t)^{(l-1)/k} = \exp\left(-t\frac{l-1}{k}\right)$$

□

LEMMA **3.3**
The waiting times between two successive jumps of the process $X(t)$ are independent from each other and have means bounded below by l and above by el.

Proof. The independence is insured by the mutation process, which is completely blind. Furthermore, whatever $X(t)$ is (except $\mathbf{e}_f$ of course), there is a string just after it in the longpath which differs only from $X(t)$ in one bit, so that at each time the probability p that a jump occurs is bounded below by:

$$p \geq l^{-1}(1-l^{-1})^{l-1} \simeq e^{-1}l^{-1}$$

Furthermore, for $j = 1, ..., k-1$, there is at most one string in the longpath which is after $X(t)$ and which differs from $X(t)$ in exactly j bits. As a consequence, the next jump of $X(t)$ must be either a jump to one of these strings, or a shortcut of the family $\mathcal{S}'$. As a consequence at each time the probability p that a jump occurs is bounded above by:

$$p \leq \sum_{j=1}^{k-1} l^{-j}(1-l^{-1})^{l-j} + P_{\mathcal{S}'} \simeq e^{-1}l^{-1} + l^{-k+1}/k \underset{\text{for } k\geq 2}{\leq} (e^{-1}+2^{-1})l^{-1} \leq l^{-1}$$

The bounds on p then imply the bounds on the mean waiting time T, since $\mathbb{P}(T > t) = (1-p)^t$ and $\mathbb{E}[T] = \sum_{t=1}^{\infty} \mathbb{P}(T \geq t) = 1/p$. □

The above proof actually shows that, if $k \geq 3$, the mean waiting time between two jumps of $X(t)$ is $el + O(1)$.

3.3 Proof of Proposition 3.1

We now turn to the proof of Proposition 3.1. We can split the all possible ways to go from $\mathbf{e}_i$ to $\mathbf{e}_f$ into two complementary groups:
- G1: the process does not contain any shortcut of the family $\mathcal{S}'$,
- G2: the process contains at least one shortcut of the family $\mathcal{S}'$,

so that the expectation of T writes:

$$\mathbb{E}[T] = \mathbb{E}[T/G_1]\mathbb{P}(G_1) + \mathbb{E}[T/G_2]\mathbb{P}(G_2) \tag{5}$$

We shall estimate the four terms of the right-hand member of this equality in the following.

Given G1 The process visits at least one over two strings of the skeleton, or a string of the neighboring bridge, that is to say at least $2^{(l-1)/k}$ strings. Since each visit has mean bounded below by l, we get:

$$\mathbb{E}\left[T/G_1\right] \geq l2^{(l-1)/k} \tag{6}$$

On the other hand, in the worst case the process visits all strings of the path, and it spends a mean time el for each of them. Thus:

$$\mathbb{E}\left[T/G_1\right] \leq e(k+1)l2^{(l-1)/k} \tag{7}$$

Combining (6) and (7), if $k = [l^\alpha]$ then:

$$\lim_{l\to\infty} \frac{1}{l^{1-\alpha}\ln(2)} \ln \mathbb{E}\left[T/G1\right] = 1. \tag{8}$$

Given G2 On the one hand, in the best case we need only one shortcut of the family $\mathcal{S}'$ to arrive at $\mathbf{e}_f$. By Lemma 3.2 we have:

$$\mathbb{E}\left[T/G_2\right] \geq \frac{k}{l-1} l^k \tag{9}$$

On the other hand, in the worst case the following sequence of events insures that the process reaches $\mathbf{e}_f$, whatever the intermediate events: shortcut $\mathcal{S}_1$, then $\mathcal{S}_2$, etc until $\mathcal{S}_{(l-1)/k}$. By Lemma 3.2 we get the estimate:

$$\mathbb{E}\left[T/G_2\right] \leq e \frac{l-1}{k} l^k \tag{10}$$

Combining (9) and (10), if $k = [l^\alpha]$ then:

$$\lim_{l\to\infty} \frac{1}{l^\alpha \ln(l)} \ln \mathbb{E}\left[T/G2\right] = 1. \tag{11}$$

We now introduce some new notations. The process $\bar{X}(t)$ is equal to the process $X(t)$ where we remove all the steps corresponding to the shortcuts of the family $\mathcal{S}'$. We denote by T_{ws} the first time when $\bar{X}$ hits the string $\mathbf{e}_f$.
The process $\bar{X}$ visits at most every string of the whole path (that is to say at most $N_l := (k+1)2^{(l-1)/k} - k + 1$ strings), and the waiting times between two jumps have means bounded above by el. Thus:

$$\mathbb{E}[T_{ws}] \leq elN_l \tag{12}$$

Furthermore, since no shortcut of $\mathcal{S}'$ happens for $\bar{X}$, the process must visit at least one over two strings of the skeleton, or a string of the neighboring bridge, that is to say at least $N_s^l := 2^{(l-1)/k}$ strings. Furthermore the waiting times between two jumps are statistically independent and have means bounded below by l. By the usual large deviations principle for the sum of independent random variables [11, theorem 3-8], we get that, for any $\delta > 0$, there exists some $c_\delta > 0$ such that:

$$\mathbb{P}\left(T_{ws} < (1-\delta) l N_l^s\right) \leq \exp(-c_\delta l N_l^s), \tag{13}$$

Estimate of $\mathbb{P}(G_1)$ in the case $\alpha < 1/2$ Under G_1 we have $T = T_{ws}$ so that $\mathbb{P}(G_1) = \mathbb{P}(T_{ws} < T_{\mathcal{S}'})$. The third statement of Lemma 3.2 establishes that $\mathbb{P}(T_{\mathcal{S}'} > t) = \exp -\frac{t}{l^k k/(l-1)}$ for any t. Since T_{ws} and $T_{\mathcal{S}'}$ are independent, we then have:

$$\mathbb{P}(G_1) = \mathbb{P}(T_{\mathcal{S}'} > T_{ws}) = \mathbb{E}\left[\exp -\frac{T_{ws}}{l^k k/(l-1)}\right] \tag{14}$$

We decompose the expectation with respect to the event: $T_{ws} \geq \frac{1}{2} l N_l^s$:

$$\begin{aligned}
\mathbb{E}\left[\exp -\frac{T_{ws}}{l^k k/(l-1)}\right] &= \mathbb{E}\left[\exp -\frac{T_{ws}}{l^k k/(l-1)} / T_{ws} < \frac{1}{2} l N_l^s\right] \mathbb{P}\left(T_{ws} < \frac{1}{2} l N_l^s\right) \\
&\quad + \mathbb{E}\left[\exp -\frac{T_{ws}}{l^k k/(l-1)} / T_{ws} \geq \frac{1}{2} l N_l^s\right] \mathbb{P}\left(T_{ws} \geq \frac{1}{2} l N_l^s\right) \\
&\leq \mathbb{P}\left(T_{ws} < \frac{1}{2} l N_l^s\right) + \exp -\frac{l N_l^s}{2 l^k k/(l-1)}
\end{aligned}$$

If $k = [l^\alpha]$, $\alpha < 1/2$, then (13) and the fact that $l N_l^s \sim l 2^{l^{1-\alpha}} \gg 2 l^{l^\alpha} l^{\alpha-1}$ implies:

$$\mathbb{P}(G_1) \simeq 0, \text{ when } k = [l^\alpha],\ \alpha < 1/2 \text{ and } l \gg 1 \tag{15}$$

Substituting (8), (11) and (15) into (5) proves the first point of the Proposition.

Estimate of $\mathbb{P}(G_1)$ in the case $\alpha \geq 1/2$ Applying Jensen's inequality to Eq. (14):

$$\mathbb{P}(G_1) \geq \exp - \frac{\mathbb{E}[T_{ws}]}{l^k k/(l-1)}$$

From Eq. (12), if $k = [l^\alpha]$, $\alpha \geq 1/2$, then the fact that $elN_l \sim el2^{l^{1-\alpha}} \ll l^{l^\alpha} l^{\alpha-1}$ implies:

$$\mathbb{P}(G_1) \simeq 1, \text{ when } k = [l^\alpha], \ \alpha \geq 1/2 \text{ and } l \gg 1 \tag{16}$$

Substituting (8), (11) and (16) into (5) then proves the second point of the Proposition.

3.4 Remarks

1. Note that for $\alpha \geq 1/2$, the waiting time for one shortcut occurrence (equation 9 under G2) is much longer than –the upper bound of– expected convergence time when following the entire path (equation 7 under G1). Though, this is not enough to describe the effective behavior of the process as described below.

2. A direct consequence of equations (15) and (16) is that shortcuts speed up the convergence in the case of $\alpha < 1/2$ only. Otherwise, for $\alpha \geq 1/2$, the process will reach $\mathbf{e}_f$ by following the whole path before any shortcut occurs.

3. Note also that the critical value $\alpha_c = 1/2$ holds true in the limit $l \to \infty$, but is slightly different for large but finite l. Indeed comparing the different expressions, we get that the transition between the two regimes occurs when $l^{\alpha_c} \ln(l) = l^{1-\alpha_c} \ln(2)$, i.e. for:

$$\alpha_c = \frac{1}{2} - \frac{\ln\ln l - \ln\ln 2}{2\ln l}$$

which converges to the value $1/2$ at rate $\ln\ln l / \ln l$.

4 Asymptotic behavior for large l and fixed k

This section is devoted to the case of fixed value of k in the asymptotic framework $l \gg 1$. The main result is stated in Subsection 4.1 and claims that the expected first hitting time is $el^{k+1}/k \ln(l/k)$. This result is not an estimate but the dominant term of the expansion of $\mathbb{E}[T]$ with respect to l. We then give at the same order of precision the complete probability density of T. It shows that, although the normalized variance of T goes to 0 as $l \to \infty$, the decay rate is so slow that the variance actually remains of order 1 for a large band of values of l. We are also able to exhibit a class of exceptional realizations of probability kl^{-1} where the process hits the final string much quicker than the expected value $\mathbb{E}[T]$. After studying this set of realizations, we describe how we can take advantage of it in Subsection 4.3. Finally subsection 4.5 compares different population based evolution strategies behavior.

4.1 Main results

Let us consider the first hitting time T of the process $X(t)$ (defined in section 2).

PROPOSITION **4.1**
The expectation of the first hitting time T is for large l:

$$\mathbb{E}[T] \underset{l\to\infty}{\simeq} \frac{el^{k+1}}{k\ln(l/k)}. \tag{17}$$

which means that $\mathbb{E}[T] \times \left(\frac{el^{k+1}}{k\ln(l/k)}\right)^{-1}$ *converges to* 1 *as* l *goes to infinity.*

Proof. See Subsection 4.2. □

The following Proposition gives more detail about the statistical distribution of the first hitting time T.

PROPOSITION **4.2**
1. The normalized variance of T *goes to* 0 *as* $l \to \infty$*:*

$$V(T) := \frac{\sqrt{\mathbb{E}[T^2] - \mathbb{E}[T]^2}}{\mathbb{E}[T]} \underset{l\to\infty}{\longrightarrow} 0$$

2. As $l \to \infty$*, the statistical distribution of* T *can be represented at the first order as* $el^k\bar{T}$*, where* $\bar{T}$ *is a random variable with density :*

$$p_{\bar{T}}(t) = \sum_{m=1}^{\infty} \sum_{\substack{n_1,\ldots,n_m=1 \\ n_1+\ldots+n_{m-1}<L \\ n_1+\ldots+n_m\geq L}}^{\infty} \int_0^\infty \ldots \int_0^\infty d\tau_1 \ldots d\tau_m \prod_{j=1}^{m} e^{-n_j\tau_j}(1-e^{-\tau_j})\delta_{\tau_1+\ldots+\tau_m=t},$$

and $L = (l-1)/k$*.*

Proof. See Subsection 4.2. □

In the above expression of the density $p_{\bar{T}}(t)$ the integer m stands for the number of shortcuts which are necessary to reach the final string, τ_j, $j = 1, ..., m$ are the times between two shortcuts, and n_j, $j = 1, ..., m$ are the ranges of the corresponding shortcuts, as explained in Subsection 4.2.
By integrating $tp_{\bar{T}}(t)$ and $t^2p_{\bar{T}}(t)$ with respect to t we can compute the normalized variance of T. It then appears that the relative fluctuations with respect to the mean stay of order 1 for a large band of values of l, although it decays to 0, as shown by Figure 1, where the normalized variance is plotted.

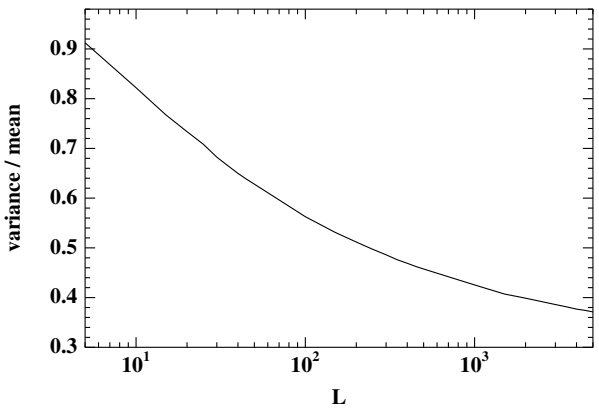

FIG 1 : *Normalized variance of the variable* T*, which depends at the first order only on the ratio* $L := (l-1)/k$*.*

In Figure 2 are plotted the probability densities of T for typical and relevant values of the parameters k and l. The theoretical histograms of T present an anomalous peak close

to zero. This corresponds to outliers which are not at all artifacts of the theory. This leftmost peak in the histogram of T, corresponds to the event Ω_1: shortcut $\mathcal{S}_1$ is the first to occur, i.e. before $\mathcal{S}_2$, ..., $\mathcal{S}_{(l-1)/k}$. The peak disappears when the first term ($m = 1$) in the expression of $p(\bar{t})$ is omitted. The event Ω_1 results in a smaller convergence time than average. This exceptional set of realizations is studied carefully in Subsection 4.3.

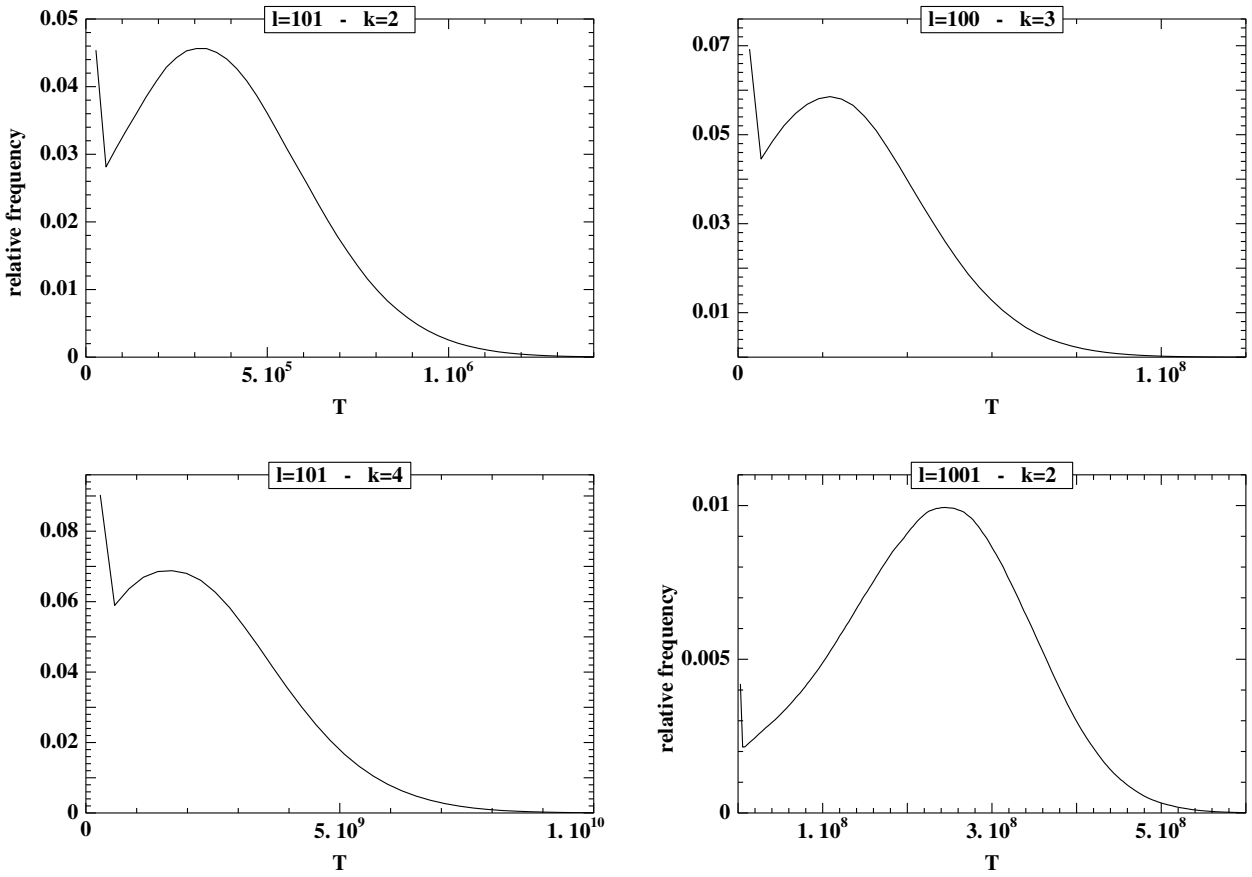

FIG 2 : *Theoretical histograms of the random variable T for different values of the parameters k and l.*

4.2 Proofs of Propositions 4.1 and 4.2

The (k,l)-path can be decomposed in the following manner:

$$
\begin{array}{rcllll}
\mathcal{M}_1 & = & \{0..0**\ & ... & *\} & \bigcup \{bridges\} \\
\mathcal{M}_2 & = & \{1..11..1**\ & ... & *\} & \bigcup \{bridges\} \\
\mathcal{M}_3 & = & \{1..10..01..1**\ & ... & *\} & \bigcup \{bridges\} \\
\mathcal{M}_4 & = & \{1..10..00..01..1**\ & ... & *\} & \bigcup \{bridges\} \\
 & \vdots & & & \vdots & \\
\mathcal{M}_{(l-k-1)/k} & = & \{1..10..0 & ... & 0..01..1*\} & \bigcup \{bridges\} \\
\mathcal{M}_{(l-1)/k} & = & \{1..10..0 & ... & 0..00..0*\} &
\end{array}
$$

Note that $\mathcal{M}_1$ contains about one half of all the strings of the path, and more generally $\mathcal{M}_j$ contains about a proportion 2^{-j} of all the strings of the path. The longpath also presents

some remarkable symmetries:
If you consider the string n^o j, with $j < N_l/2$, and you apply the shortcut $\mathcal{S}_1$, then you get the string n^o $N_l - j + 1$.
If you consider the string n^o j, with $j < N_l/4$, and you apply the shortcut $\mathcal{S}_2$, then you get the string n^o $N_l/2 - j + 1$.
Furthermore since the waiting times $T_{\mathcal{S}_j}$ are independent and identically distributed, we have for any fixed j_0:

$$\mathbb{P}(T_{\mathcal{S}_{j_0}} < T_{\mathcal{S}_1}, .., T_{\mathcal{S}_{j_0-1}}, T_{\mathcal{S}_{j_0+1}}, ..., T_{\mathcal{S}_{(l-1)/k}}) = \frac{k}{l-1}.$$

Let us discuss *when* the process enters $\mathcal{M}_1^c$.
• If no shortcut of $\mathcal{S}'$ occurs, then it will take an exponentially long time.
• If neither $\mathcal{S}_1$ nor $\mathcal{S}_2$ occurs, then the process must still visit an exponentially large number of strings, which takes an exponentially long time.
• If $\mathcal{S}_2$ occurs, but not $\mathcal{S}_1$. If one of the $\mathcal{S}_j$, $j \geq 3$, occurs before $\mathcal{S}_2$, then the process must then visit an exponentially large number of strings, which takes an exponentially long time. The only way that could produce a not-exponentially large time is that $\mathcal{S}_2$ be the first shortcut of the family $\mathcal{S}'$ to occur. Then the process jumps at the end part of $\mathcal{M}_1$ and will soon enter $\mathcal{M}_2$. This event that we denote by $\mathcal{E}$ therefore provides a probability $k/(l-1)$ to enter $\mathcal{M}_2$ (this is the probability that $\mathcal{S}_2$ be smaller than the other $\mathcal{S}_j$). We shall see that the contribution of $\mathcal{E}$ is negligible because its probability is too small to be taken into account.
The only probable way to enter $\mathcal{M}_1^c$ is indeed that $\mathcal{S}_1$ occurs. The access time to $\mathcal{M}_1^c$ obeys then an exponential distribution with mean el^k. It will be denoted in the following by $el^k\tau_1$:

$$\mathbb{P}\left(\tau_1 > t\right) = \exp(-t)$$

Let us now study the point *where* the process enters $\mathcal{M}_1^c$.
From the above discussion, this point depends on the shortcuts $\mathcal{S}_j$, $j \geq 2$, which may have occurred before $\mathcal{S}_1$.
If $\mathcal{S}_2$ occurs before $\mathcal{S}_1$, then the process enters $\mathcal{M}_2$ (Thus the process has probability 1/2 to enter $\mathcal{M}_2$ by this way. That is why the contribution of the event $\mathcal{E}$ can be neglected).
If $\mathcal{S}_2$ does not occur before $\mathcal{S}_1$, but $\mathcal{S}_3$ does, then the process enters $\mathcal{M}_3$.
More generally, if $\mathcal{S}_i$, $i \leq j-1$, do not occur before $\mathcal{S}_1$, but $\mathcal{S}_j$ does, then the process enters $\mathcal{M}_j$. Let us denote by N_1 this particular j. Since the waiting times for the shortcuts $\mathcal{S}_i$ obey independent exponential distributions with mean el^k, we have:

$$\mathbb{P}\left(N_1 = j/\tau_1 = t\right) = \mathbb{P}(\tau_2', ..., \tau_{j-1}' > t, \tau_j' < t/\tau_1 = t\Big) = \left(1 - e^{-t}\right) e^{-(j-1)t}$$

We can then iterate the above arguments. We then find that we need to do it M times, where M is the first time when $\sum_{i=1}^m N_j$ reaches $(l-1)/k$, which means that the process enters $\mathcal{M}_{(l-1)/k}$.
As a consequence, denoting

$$L = \frac{l-1}{k}, \tag{18}$$

we can describe the first hitting time T as follows.
For large l, the first hitting time T obeys the distribution of:

$$T = el^k \bar{T}, \tag{19}$$

where $\bar{T}$ is defined by the following way. Consider a sequence of independent and identically distributed real-valued random variables τ_j, $j = 1, 2, ...$ with exponential density: $p(\tau) = \mathbb{1}_{\tau \geq 0} e^{-\tau}$. Consider also a sequence of independent integer-valued random variables N_j, $j = 1, 2, ...$, whose distributions depend on τ_j. More exactly, $\mathbb{P}(N_j = n/\tau_j = \tau) = p_\tau(n)$, where

$$p_\tau(n) = e^{-(n-1)\tau}(1 - e^{-\tau}), \qquad n = 1, 2, ... \tag{20}$$

If we denote by M_L the first exit time:

$$M_L = \inf \left\{ m \geq 1, \sum_{j=1}^{m} N_j \geq L \right\}, \tag{21}$$

then the random variable $\bar{T}$ is defined by:

$$\bar{T} = \sum_{j=1}^{M_L} \tau_j. \tag{22}$$

The closed-form expression of the probability density of $\bar{T}$ given in Proposition 4.2 can then be deduced directly from the definition of $\bar{T}$.
Roughly speaking, we then expect from (19) that $\mathbb{E}[T] = el^k \mathbb{E}[\bar{T}]$, from (22) that $\mathbb{E}[\bar{T}] \simeq \mathbb{E}[M_L]\mathbb{E}[\tau_1] = \mathbb{E}[M_L]$ and from (21) that $\mathbb{E}[M_L] \simeq L/\mathbb{E}[N_1]$. However the distributions of M_L and τ_j are joint, since M_L depends on $(N_j)_j$ whose distribution depends on $(\tau_j)_j$. Therefore the assertion "$\mathbb{E}[\bar{T}] \simeq \mathbb{E}[M_L]\mathbb{E}[\tau_1]$" is not so obvious, and actually it is false in some sense. Furthermore, the statistical distribution of N_j is:

$$\mathbb{P}(N_j = n) = \int_0^\infty p(\tau) p_\tau(n) d\tau = \frac{1}{n(n+1)}. \tag{23}$$

A noticeable feature which will appear crucial in the following is that $\mathbb{E}[N_j] = +\infty$. It will prevent us from applying the standard theorems of probability theory (strong law of large numbers and central limit theorem) and it will give rise to anomalous behaviors for M_L and the other relevant quantities. The asymptotic behaviors of the variables M_L and $\bar{T}$ are given by the following Propositions.

PROPOSITION **4.3**
The sequence of normalized random variables $\frac{\ln L}{L} M$ converges in mean, in L^2 and in probability as $L \to \infty$ to 1, which means that, for any $\delta > 0$:

$$\frac{\ln L}{L} \mathbb{E}[M_L] \overset{L \to \infty}{\longrightarrow} 1 \ , \ \left(\frac{\ln L}{L} \right)^2 \mathbb{E}[M_L^2] \overset{L \to \infty}{\longrightarrow} 1 \text{ and } \mathbb{P}\left(\left| \frac{\ln L}{L} M_L - 1 \right| \geq \delta \right) \overset{L \to \infty}{\longrightarrow} 0. \tag{24}$$

PROPOSITION **4.4**
For any $q = 1, 2$, we have

$$\frac{\mathbb{E}[\bar{T}^q]}{\mathbb{E}[M_L^q]} \underset{L \to \infty}{\longrightarrow} 1 \tag{25}$$

Combining these Propositions yield the statement of Proposition 4.1 ($q = 1$) and the first point of Proposition 4.2 ($q = 2$). In the Appendix B, we shall first study the statistical distribution of the variable M_L and prove Proposition 4.3, and then we study the distribution of $\bar{T}$ and prove Proposition 4.4.

4.3 A class of exceptional events

We now go back to a very interesting point which consists in the anomalous peak close to 0 in the histogram of the first hitting time T. We can describe the corresponding set of realizations (*shortcut $\mathcal{S}_1$ happens first*) which gives rise to a very quick convergence of the process.

PROPOSITION **4.5**
There exists a set of realizations Ω_1 with probability

$$\mathbb{P}(\Omega_1) \underset{l\to\infty}{\simeq} kl^{-1} \tag{26}$$

such that the conditional distribution of the first hitting time given Ω_1 is an exponential distribution with mean $2el^{k-1}k$:

$$\mathbb{P}\left(T \geq a/\Omega_1\right) \underset{l\to\infty}{\simeq} \exp(-\frac{a}{2el^{k-1}k}). \tag{27}$$

Proposition 4.5 shows that the expected convergence time under Ω_1 ($2el^{k-1}k$) is roughly l^2 times shorter than the global expected time T ($\frac{el^{k+1}}{k\ln(l/k)}$ in Proposition 4.1). A noticeable fact is then that the convergence time with $N = O(l)$ processes will be l^2 times shorter than the expected time for a single process with very high probability, because it is then highly probable that at least one of the N processes will realize the favorable event Ω_1: shortcut $\mathcal{S}_1$ occurs first. We quantify this statement in the following corollary.

COROLLARY **4.6**
Assume that we deal with a set of N processes which evolve independently. Then if N is large, denoting $c = \dfrac{N}{l/k}$, there exists a set of realizations Ω_N with probability

$$\mathbb{P}(\Omega_N) \underset{l\to\infty}{\simeq} 1 - e^{-c} \tag{28}$$

such that the conditional expectation of the first hitting time is:

$$\mathbb{E}\left[T/\Omega_N\right] \underset{l\to\infty}{\simeq} 2ef(c)l^{k-1}k \tag{29}$$

where $f(c) = \frac{\mathrm{Ei}(c)-\gamma-\ln c}{e^c-1}$ and γ is the Euler constant $\gamma \simeq 0.57...$ (see Figure 3).
More exactly, the conditional distribution of T given Ω_N is:

$$T \underset{l\to\infty}{\simeq} -2el^{k-1}k\ln\left(\frac{1-Z}{c}\right), \tag{30}$$

where Z is a $[0,1]$-valued random process with density $p(z) = \mathbb{1}_{z\in[0,1]}e^{-z}(1-e^{-1})^{-1}$.

Proof of Proposition 4.5 We still represent the statistical distribution of the first hitting time T as (19)-(22). In this framework we define the set Ω_1 as the set of the realizations which fulfill the condition $M_L = 1$:

$$\Omega_1 = \{M_L = 1\} = \{N_1 \geq L\}.$$

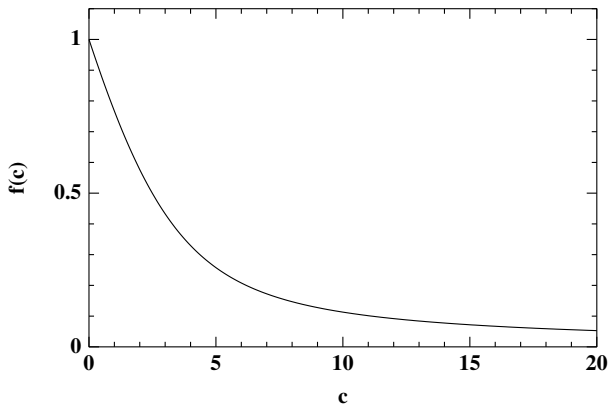

FIG 3 : *Function* $c \mapsto f(c)$.

Since the distribution of N_1 is given by (23), the probability of this set is simply $\mathbb{P}(\Omega_1) = L^{-1}$. Given Ω_1, the process $X(t)$ obeys the following evolution: before $el^k\tau_1$, it evolves without shortcut and arrives at the string $\mathrm{n}^o j$; at time $el^k\tau_1$, by shortcut $\mathcal{S}_1$ the process jumps at the string $\mathrm{n}^o(k+1)2^{(l-1)k} - k + 1 - j$; finally the process goes to the final string. Furthermore the statistical distribution of τ_1 given Ω_1 is the statistical distribution of τ_1 given $\tau_2, ..., \tau_L > \tau_1$, where the τ_j are independent random variables with exponential distribution and mean 1:

$$\begin{aligned}
\mathbb{P}(\tau_1 > t/\Omega_1) &= \frac{\mathbb{P}(\tau_2, ..., \tau_L > \tau_1 > t)}{\mathbb{P}(\tau_2, ..., \tau_L > \tau_1)} \\
&= \frac{\int_t^\infty d\tau_1 e^{-\tau_1} \left(\int_{\tau_1}^\infty d\tau_2 e^{-\tau_2}\right)^{L-1}}{\int_0^\infty d\tau_1 e^{-\tau_1} \left(\int_{\tau_1}^\infty d\tau_2 e^{-\tau_2}\right)^{L-1}} \\
&= e^{-Lt}
\end{aligned}$$

The result of the Proposition then follows readily.

Proof of Corollary 4.6 We adopt the same notations as (19)-(22). We add a subscript $i = 1, ..., N$ to each quantity which stands for the labels of the N independent processes. The set Ω_N is then defined as:

$$\Omega_N = \left\{ \inf_{i=1,...,N} M_{L,i} = 1 \right\}.$$

The probability of the complementary set is:

$$\mathbb{P}(\Omega_N^c) = (1 - \mathbb{P}(M_{L,1} = 1))^N = (1 - L^{-1})^N \overset{l\to\infty}{\longrightarrow} e^{-c}.$$

where $c = \dfrac{N}{l/k}$. The global first hitting time is the infimum of the first hitting times corresponding to each process:

$$T = \inf_{i=1,...,N} T_i,$$

where the T_i are independent and identically distributed random variables defined as in (19)-(22). Furthermore, given Ω_N, the statistical distribution of T is:

$$T = \inf_{i\in I} T_i,$$

where $I = \{i = 1, ..., N \text{ s.t. } M_{L,i} = 1\}$ and T_i are independent and identically distributed random variables defined as (27). We have

$$\mathbb{P}(|I| = j/\Omega_N) = \mathbb{P}(|I| = j/|I| \geq 1) = \frac{\mathbb{P}(|I| = j)}{\sum_{j'=1}^{N} \mathbb{P}(|I| = j')}.$$

For any j', the probability that $|I|$ be equal to j' is the probability that j' of the N processes satisfy $M_{L,i} = 1$, and that the other ones satisfy $M_{L,i} > 1$. Since $\mathbb{P}(M_{L,i} = 1) = L^{-1}$, this implies:

$$\mathbb{P}(|I| = j') = C_N^{j'} L^{-j'} (1 - L^{-1})^{N-j'}$$

For $N = cL$ and $L \gg 1$, the binomial distribution becomes equivalent to a Poisson distribution:

$$\mathbb{P}(|I| = j') = e^{-c} \frac{c^{j'}}{j'!}$$

and consequently:

$$\mathbb{P}(|I| = j/\Omega_N) = \frac{e^{-c}}{1 - e^{-c}} \frac{c^j}{j!}, \text{ for } j \geq 1.$$

Thus

$$\begin{aligned} \mathbb{P}(T \geq t) &= \sum_{j=1}^{\infty} \mathbb{P}\left(|I| = j, T_1 \geq t, ..., T_j \geq t\right) \\ &= \sum_{j=1}^{\infty} \frac{e^{-c}}{1 - e^{-c}} \frac{c^j}{j!} e^{-jt/(2el^{k-1}k)} \\ &= \frac{\exp\left(ce^{-t/(2el^{k-1}k)}\right) - 1}{e - 1}. \end{aligned}$$

The interested reader can then check that this is exactly the statistical distribution of the random variable defined by (30).

4.4 Optimization of the mutation probability

PROPOSITION **4.7**
Let us assume that the mutation probability is dl^{-1}.
1. For large l, the first hitting time T_d obeys the statistical distribution of $e^{d-1}d^{-k}T_1$, where T_1 is as described in Section 4. In particular its mean is:

$$\mathbb{E}[T_d] \underset{l \to \infty}{\simeq} \frac{e^d}{d^k} \frac{l^{k+1}}{k \ln(l/k)}. \tag{31}$$

2. Knowing k, the best choice for d is $d = k$.

Proof. The same arguments as for the case $d = 1$ yield the result. The only difference comes from the fact that the waiting times $T_{\mathcal{S}_j}$ for the shortcuts $\mathcal{S}_j$ obey different distributions. Indeed, at each time the probability that the shortcut $\mathcal{S}_j$ occurs is:

$$P_{\mathcal{S}_j} = \left(\frac{d}{l}\right)^k \left(1 - \frac{d}{l}\right)^{l-k} \simeq d^k e^{-d} l^{-k}.$$

Since the events at different times are independent, the statistical distribution of $T_{\mathcal{S}_j}$ is given by:

$$\mathbb{P}(T_{\mathcal{S}_j} > t) = (1 - P_{\mathcal{S}_j})^t \simeq \exp - \frac{t}{e^d d^{-k} l^k},$$

which proves that $T_{\mathcal{S}_j}$ obeys an exponential distribution with mean $e^d d^{-k} l^k$. The remainder of the proof is then exactly the same as for the case $d = 1$. □

4.5 Discussion and comparison of different algorithms

In this subsection we disccuss the asymptotic behavior for large l and fixed k of different evolution strategies. The main result is that all population based evolution strategies are less efficient (in number of evaluations) than a (1+1)-ES. Moreover, the best strategy is to run simultaneously several independent (1+1)-ES processes.
This latter result has been already stated in subsection 4.3 : Corollary (4.6) shows that with cl/k independent processes ($c > 1$) evolving simultaneously, the first hitting time (of the fastest process) is about l^{k-1} with very high probability ($1 - e^{-c}$). Recall that with one process $X(t)$, the first hitting time is about $l^{k+1}/\ln(l)$ (proposition 4.1). The first strategy of independent processes is therefore very advantageous and much more rapid than the second one in number of generations, and also in number of evaluations.

Consider now a number of processes that communicate together by exchanging the best so-far string, as in the case of a $(1 + \lambda) - ES$. Expected first hitting time is necessarily longer than that of N processes evolving independently. This is due to the fact that the order of shortcut occurrence determines the convergence time (as explained in 4.2). More precisely, the only way to get a different order of convergence time is that the event Ω_1 happens for some parent: the shortcut $\mathcal{S}_1$ is then the first shortcut to occur. Exchanging the best so-far string forces the N processes to follow the same order of shortcuts occurrence. In fact, the string resulting from the first shortcut contaminates –more or less quickly– the other processes, and so on.

In the case of a $(1 + \lambda) - ES$, the string resulting from the first shortcut immediately colonizes the population. But expected time for a shortcut and for any improvement gets N times smaller, hence expected first hitting time is N times smaller than that of a single process. Hence, the number of evaluations remains the same, and there is no advantage to increase λ.

Consider now a $(\mu+\lambda)-ES$ with $\lambda = \alpha\mu$, starting at string $\mathbf{e}_i$. Due to elitism, the μ parents are always on the path. If a shortcut occurs, it takes a time of order $\sum_{k=1}^{\mu-1} l/(k\alpha)$ (so less than $l\alpha^{-1}\ln(\mu)$) for the string resulting from the shortcut to dominate the population. It is then very unlikely that another shortcut (exponential distribution) happens within this time, and happens first. An upper bound of this probability is given in the following:
Suppose a shortcut different from $\mathcal{S}_1$ happens for one of the μ parents. The probability that the shortcut $\mathcal{S}_1$ happens for one of the remaining $(\mu - 1)$ parents, within a time $l\alpha^{-1}\ln(\mu)$ (even if it happens after another contaminating shortcut) is:

$$\mathbb{P}\left(\inf_{i=1..\mu-1}(T_i) \leq l\alpha^{-1}\ln(\mu)\right) = 1 - \mathbb{P}\left(T_i > l\alpha^{-1}\ln(\mu)\right)^{\mu-1} \simeq 1 - \exp - \frac{l\mu\alpha^{-1}\ln(\mu))}{l^k}$$

If $k = 2$ and $\mu = O(l)$, this probability is 1 as l goes to infinity, hence the event Ω_1 happens before the first shortcut colonizes the population. Though, this is not the probability of the event Ω_1. And, as explained above, μ independent processes are obviously more advantageous even in this case.

If $k \geq 3$, then the probability tends to zero as l goes to infinity (even if $\mu = O(l)$). Hence all μ individuals follow the first unlucky shortcut, and looses any chance to realize the event Ω_1. $(\mu + \lambda) - ES$ only speeds the convergence with a factor of λ in number of generations.

Summary and conclusion

This paper studies the behavior of a (1+1)-ES process on Rudolph's long k-paths, in the asymptotic framework $l \gg 1$: First, for $k = l^\alpha$, we prove that expected convergence time is exponential. Besides, shortcuts speed up the convergence only if $\alpha < 1/2$. Otherwise (if $\alpha \geq 1/2$), the process reaches the solution by following the whole path before any shortcut occurs.
Second, in the case of fixed value of k, expected first hitting time T is equal to $el^{k+1}/k \ln(l/k)$ (at the first order with respect to l). The normalized variance of T goes to 0 as $l \to \infty$ at very slow rate. Further, the study of the statistical distribution of convergence time shows an anomalous peak close to zero, corresponding to the event *shortcut $\mathcal{S}_1$ happens first*. This event happens with probability kl^{-1}, and yields a convergence time –a factor of l^2– smaller than the expected value $\mathbb{E}[T]$. Therefore, the best strategy to take advantage of this distribution, is to perform $O(l)$ independent (1+1)-ES processes, so that one of these processes is very likely to realize the exceptional event. On the other hand, population based processes limit diversity in shortcuts occurrence, drifting all individuals to blindly follow the first occurring shortcut. This implies the same number of evaluations as with a (1+1)-ES.
Finally, expected convergence time distribution is given for any mutation rate d/l, $d > 0$. As expected, the smallest convergence time is obtained with a k/l mutation probability.

This answers the question: is there a longpath for a (1+1)-ES ? Indeed, in the case $k = l^\alpha$ and $\alpha \geq 1/2$, the process follows the path and convergence time is exponential. As for other algorithms, one can still wonder whether such a longpath exists. Yet the notion of "following a path" has to be defined for population based algorithms. Partial answers to these questions are given in [7], where a fitness function is constructed in the binary framework so that the GA empirically follows a *path* arbitrarily chosen in the search space.

Appendix

A - The lack of memory of exponential random variables

LEMMA **A.1**
Let τ be an exponential random variable with mean m. Then for any stopping time τ' (for instance an independent exponential random variable), the distribution of τ given $\tau > \tau'$ is still an exponential distribution with mean m.

Proof. This is proved in [1, Vol 1, ch. III-9]. □

B - Technical results valid for large l and fixed k

We aim at studying the statistical distribution of the stopping time M_L as defined by (21) in Subsection 4.2. We shall use the well known Tauberian theorem (Theorem 4 [1, Vol. 2, ch. XIII-5]):

LEMMA **B.1**
If $u(x) \geq 0$ is monotone and

$$\int_0^\infty e^{-\lambda x} u(x) dx \underset{\lambda \to 0^+}{\sim} \lambda^{-\rho} L(\lambda^{-1})$$

where L is slowly varying (which means that $L(sx)/L(s) \to 1$ as $s \to \infty$ for any $x > 0$), then:

$$u(x) \underset{x \to +\infty}{\sim} \frac{1}{\Gamma(\rho)} x^{\rho - 1} L(x).$$

The main result is the Proposition 4.3.

Proof of Proposition 4.3. Let us denote by S_m the partial sums:

$$S_m = \sum_{j=1}^{m} N_j.$$

Then we can express the statistical distribution of the variable M_L in terms of S_m:

$$\mathbb{P}(M_L = m) = \mathbb{P}(S_{m-1} < L) - \mathbb{P}(S_m < L).$$

Furthermore we introduce F the repartition function of the random variable N_1:

$$F(n) = \mathbb{P}(N_1 \leq n) = 1 - \frac{1}{n+1},$$

whose Laplace transform is:

$$\begin{aligned} \hat{F}(\lambda) &:= \mathbb{E}[e^{-\lambda N_1}] \\ &= 1 + (e^\lambda - 1) \ln(1 - e^{-\lambda}). \end{aligned}$$

Since $\mathbb{P}(S_m \leq L) = F^{*m}(L)$, where the star stands for the convolution, we can express the mean of the variable M_L as follows:

$$\begin{aligned} \mathbb{E}[M_L] &= \sum_{m=1}^{\infty} m \left(\mathbb{P}(S_{m-1} < L) - \mathbb{P}(S_m < L) \right) = \sum_{m=0}^{\infty} \mathbb{P}(S_m < L) \\ &= \sum_{m=0}^{\infty} F^{*m}(L-1) \\ &= I(L-1) + F * \sum_{m=0}^{\infty} F^{*m}(L-1), \end{aligned}$$

where $F^{*0}(L) = I(L) = \mathbb{1}_{L \geq 0}$. Denoting by $\hat{H}_1(\lambda)$ the Laplace transform of the series $\mathbb{E}[M_L]$:

$$\hat{H}_1(\lambda) := \sum_{L=0}^{\infty} \mathbb{E}[M_L] e^{-\lambda L},$$

we then get:

$$\hat{H}_1(\lambda) = \frac{e^{-\lambda}}{\lambda} + e^{-\lambda} \hat{F}(\lambda) \hat{H}_1(\lambda)$$

from which we deduce that:

$$\hat{H}_1(\lambda) = \frac{e^{-\lambda}}{\lambda(1-e^{-\lambda}\hat{F}(\lambda))} \underset{\lambda\to 0^+}{\simeq} \frac{1}{\lambda^2|\ln\lambda|}.$$

Applying Lemma B.1 we then get:

$$\mathbb{E}[M_L] \underset{L\to+\infty}{\simeq} \frac{L}{\ln L}. \tag{B.1}$$

Let us now deal with the fluctuations.

$$\begin{aligned}\mathbb{E}[M_L^2] &= \sum_{m=1}^{\infty} m^2 \left(\mathbb{P}(S_{m-1} < L) - \mathbb{P}(S_m < L)\right) \\ &= \sum_{m=0}^{\infty} (2m+1) F^{*m}(L-1).\end{aligned}$$

Denoting by $H_2(L) := \sum_{m=0}^{\infty} mF^{*m}(L)$ and by $\hat{H}_2(\lambda)$ its Laplace transform, we have:

$$H_2(L-1) = F * H_2(L-1) + \mathbb{E}[M_L]$$

which implies:

$$\hat{H}_2(\lambda) = e^{-\lambda}\hat{F}(\lambda)\hat{H}_2(\lambda) + \frac{e^{-\lambda}}{\lambda(1-e^{-\lambda}\hat{F}(\lambda))}$$

and consequently:

$$\hat{H}_2(\lambda) = \frac{e^{-\lambda}}{\lambda(1-e^{-\lambda}\hat{F}(\lambda))^2} \underset{\lambda\to 0^+}{\simeq} \frac{1}{\lambda^3(\ln\lambda)^2}.$$

Still applying Lemma B.1 we get:

$$\mathbb{E}[M_L^2] \underset{L\to+\infty}{\simeq} \frac{L^2}{(\ln L)^2}. \tag{B.2}$$

Combining (B.1) with (B.2), we obtain the convergence in probability of the variable $\frac{\ln L}{L} M_L$ to 1:

$$\begin{aligned}\mathbb{P}\left(\left|M_L\frac{\ln L}{L} - 1\right| \le \delta\right) &\le \frac{\mathbb{E}\left[\left(M_L\frac{\ln L}{L} - 1\right)^2\right]}{\delta^2} \\ &= \frac{\mathbb{E}\left[M_L^2\right]\left(\frac{\ln L}{L}\right)^2 - 2\mathbb{E}[M_L]\frac{\ln L}{L} + 1}{\delta^2} \underset{L\to\infty}{\longrightarrow} 0\end{aligned}$$

□

This result is asymptotic, that is to say holds true for large L. However, analyzing precisely the convergence of M_L, it appears that the convergence rate is very slow (logarithmic). If we deal with value of L which are of order 100 or 1000, it is then necessary to be very careful when using Proposition 4.3. The mean $\mathbb{E}[M_L]$ actually behaves as $L/\ln L$ even for -relatively- small values of L, as shown by Figure 4. However the relative fluctuations with respect to the mean stay of order 1 for a large band of values of L, and vanish only for $L \gg 10^9$ (Figure 4). In Figures 5 we have plotted the computed histograms of the variable M_L for different values of L.

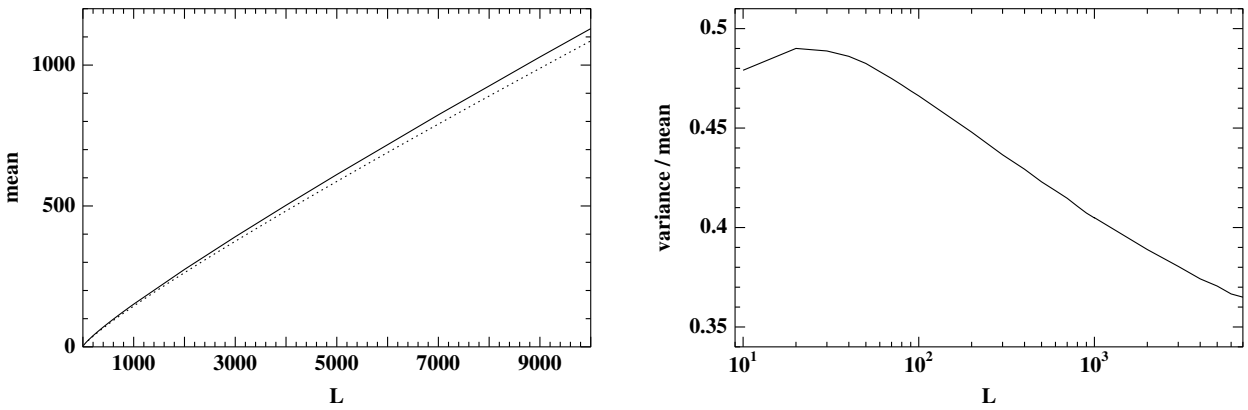

FIG 4 : *Left picture: Mean of the random variable M_L as a function of L (solid line). In dashed line is plotted the curve $L \mapsto L/\ln L$. Right picture: Normalized variance $\mathbb{E}\left[M_L^2 - \mathbb{E}[M_L]^2\right]^{1/2} / \mathbb{E}\left[M_L\right]$ of the random variable M_L as a function of L.*

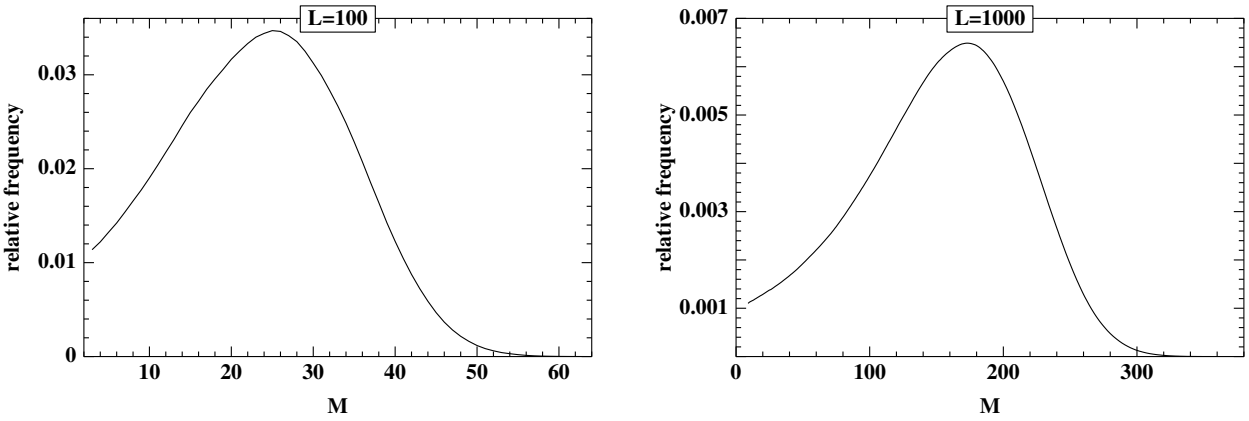

FIG 5 : *Theoretical histograms of the random variables M_L for $L = 100$ (left picture) and $L = 1000$ (right picture).*

LEMMA **B.2**
Let $p \in \{1, 2\}$, $q \in \mathbb{N}^*$ and ψ_1,...,ψ_q be continuous functions from $\mathbb{N}^*$ into $\mathbb{R}^+$ satisfying $0 < a \leq \psi_j(.) \leq b < \infty$. Then:

$$\left| \frac{\mathbb{E}[M_L^p \psi_1(N_1)...\psi_q(N_q)] - \mathbb{E}[M_L^p]\mathbb{E}[\psi_1(N_1)]...\mathbb{E}[\psi_q(N_q)]]}{\mathbb{E}[M_L^p]\mathbb{E}[\psi_1(N_1)]...\mathbb{E}[\psi_q(N_q)]} \right| \underset{L\to\infty}{\longrightarrow} 0.$$

This result is not surprising. Indeed for -relatively- small L, M_L and N_1,...,N_q are correlated. But if L is large, then M_L gets large too, so that the statistical distribution of M_L involves many N_j and becomes independent of the realization of the first particular N_1, ..., N_q. However the convergence rate is very slow (logarithmic) and all the slower as q is large.

Proof. For simplicity we give the result in the case $q = 1$. The extension to arbitrary q is straightforward. By the definition of M_L:

$$\mathbb{P}(M_L = m/N_1 = n) = \mathbb{P}\left(\sum_{j=2}^{m-1} N_j < l - n \leq \sum_{j=2}^{m} N_j\right) = \mathbb{P}(M_{L-n} = m - 1/N_1 = n),$$

which establishes that:

$$\mathbb{E}[M_L^p \psi_1(N_1)] = \mathbb{E}[(1 + M_{L-\tilde{N}_1})^p \psi_1(\tilde{N}_1)]$$

where $\tilde{N}_1$ obeys the same distribution of N_1, but is independent from $M_.$. Thus:

$$\mathbb{E}[M_L^p \psi_1(N_1)] = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \psi_1(n) \mathbb{E}[(1 + M_{L-n})^p]$$

We finally apply Lebesgue's theorem:

$$\lim_{L\to\infty} \frac{\ln L}{L} \mathbb{E}[M_L^p \psi_1(N_1)] = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \psi_1(n) = \mathbb{E}[\psi_1(N_1)]$$

□

Proof of Proposition 4.4. By Proposition 4.2, we can compute $\mathbb{E}[\bar{T}]$ by integrating the density $p_{\bar{T}}(t)$ multiplied by t. Since

$$\int_0^\infty e^{-n\tau}(1 - e^{-\tau})d\tau = \frac{1}{n(n+1)} \quad \text{and} \quad \int_0^\infty \tau e^{-n\tau}(1 - e^{-\tau})d\tau = \frac{2n+1}{n^2(n+1)^2}$$

we get:

$$\mathbb{E}[\bar{T}] = \sum_{m=1}^{\infty} \sum_{n_1,\ldots,n_m=1}^{\infty} \mathbb{1}_{n_1+\ldots+n_{m-1}<L\leq n_1+\ldots+n_m} \sum_{i=1}^{m} \frac{2n_i+1}{n_i^2(n_i+1)^2} \prod_{j\neq i}^{m} \frac{1}{n_j(n_j+1)}.$$

The sum over i contains m terms. The first $m-1$ terms are equal by simple permutation of the index. Only the last one ($i = m$) differs. We can therefore rewrite the long expression as:

$$\mathbb{E}[\bar{T}] = \mathbb{E}\left[(M_L - 1)\frac{2N_1+1}{N_1(N_1+1)}\right] + \mathbb{E}\left[\frac{2N_{M_L}+1}{N_{M_L}(N_{M_L}+1)}\right].$$

The second term of the right-hand side is uniformly bounded, while the first term converges to $\mathbb{E}[M_L - 1]\mathbb{E}\left[\frac{2N_1+1}{N_1(N_1+1)}\right]$ by Lemma B.2. Since the second expectation is equal to 1, we get the result in the case $q = 1$. The case $q = 2$ can be dealt with the same way, since the expectation $\mathbb{E}[\bar{T}^2]$ can be expressed in terms of the processes N_j and M_L:

$$\begin{aligned}\mathbb{E}[\bar{T}^2] = \quad & \mathbb{E}\left[(M_L - 1)(M_L - 2)\frac{2N_1+1}{N_1(N_1+1)}\frac{2N_2+1}{N_2(N_2+1)}\right] + \mathbb{E}\left[(M_L - 1)\frac{3N_1^2+3N_1+1}{N_1^2(N_1+1)^2}\right] \\ & +2\mathbb{E}\left[(M_L - 1)\frac{2N_1+1}{N_1(N_1+1)}\frac{2N_{M_L}+1}{N_{M_L}(N_{M_L}+1)}\right] + \mathbb{E}\left[\frac{3N_{M_L}^2+3N_{M_L}+1}{N_{M_L}^2(N_{M_L}+1)^2}\right].\end{aligned}$$

The result can be interpreted as follows. The statistical distribution of the first hitting time $\bar{T}$ depends on the fluctuations of M_L and τ_j. Both are correlated, but for large L, we deal with a large M_L, and consequently with the sum of a large number of τ_j. By the strong law of large numbers, the fluctuations of this sum goes to 0, so that the fluctuations of T become only dependent on those of M_L.

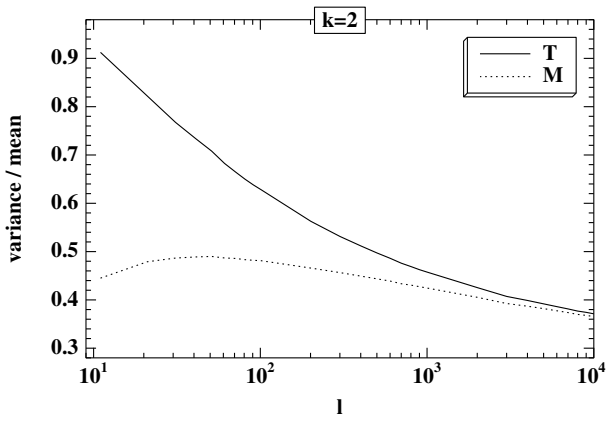

FIG 6 : *Normalized variances of the variables T and M_L ($L = (l-1)/k$).*

References

[1] W. Feller. *An introduction to probability theory and its applications.* Wiley, New York, 1971.

[2] S. Forrest and M. Mitchell. Relative building-block fitness and the building-block hypothesis. *Foundations of Genetic Algorithms 2*, pages 109–126, 1993.

[3] D. E. Goldberg. *Genetic algorithms in search, optimization and machine learning.* Addison Wesley, 1989.

[4] D. E. Goldberg. Making genetic algorithms fly:a lesson from the wright brothers. *Advanced Technology for Developpers*, 2:1–8, February 1993.

[5] C. Höhn and C. Reeves. Are long path problems hard for genetic algorithms? In H.-M. Voigt, W. Ebeling, I. Rechenberg, and H.-P. Schwefel, editors, *Proceedings of the 4^{th} Conference on Parallel Problems Solving from Nature*, volume 1141 of *LNCS*, pages 134–143. Springer Verlag, 1996.

[6] J. Horn, D.E. Goldberg, and K. Deb. Long path problems. In Y. Davidor, H.-P. Schwefel, and R. Manner, editors, *Proceedings of the 3^{rd} Conference on Parallel Problems Solving from Nature*, pages 149–158. Springer Verlag, 1994.

[7] L. Kallel and B. Naudts. Longpaths for the genetic algorithm. In *Foundations of Genetic Algorithms 5.* Morgan Kaufmann, 1998. to appear.

[8] M. Mitchell, S. Forrest, and J.H. Holland. The Royal Road for genetic algorithms : Fitness landscapes and GA performance. In F. J. Varela and P. Bourgine, editors, *Proceedings of the First European Conference on Artificial Life-93*, pages 245–254. MIT Press/Bradford Books, 1993.

[9] M. Mitchell and J.H. Holland. When will a genetic algorithm outperform hill-climbing ? In S. Forrest, editor, *Proceedings of the 5^{th} International Conference on Genetic Algorithms*, page 647, 1993.

[10] G. Rudolph. Convergence of non-elitist strategies. In Z. Michalewicz, J. D. Schaffer, H.-P. Schwefel, D. B. Fogel, and H. Kitano, editors, *Proceedings of the First IEEE International Conference on Evolutionary Computation*, pages 63–66. IEEE Press, 1994.

[11] D. W. Stroock. *An introduction to the theory of large deviations.* Springer, New York, 1984.