Statistical distribution of the convergence time of evolutionary algorithms for longpath problems

Josselin Garnier and Leila Kallel

Centre de Mathématiques Appliquées, Unité Mixte de Recherche 7641 du Centre National de la Recherche Scientifique, Ecole Polytechnique, Palaiseau 91128, France.
E-mail: *garnier@cmapx.polytechnique.fr*
E-mail: *kallel@cmapx.polytechnique.fr*

Abstract— **The behavior of a (1+1)-ES process on Rudolph's binary long k-paths is investigated extensively in the asymptotic framework with respect to string length l.**

First, the case of $k = l^\alpha$ is addressed. For $\alpha \geq 1/2$, we prove that the long k-path is a longpath for the (1+1)-ES, in the sense that the process follows the entire path with no shortcuts resulting in an exponential expected convergence time. For $\alpha < 1/2$, the expected convergence time is also exponential but some shortcuts occur meanwhile that speed 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 $(\mu+\lambda)-ES$. 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 $(\mu+\lambda)-ES$.

Keywords— **Longpath problems, statistical analysis, hitting times.**

Introduction

Longpaths 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. The intention of paper is a further study of the behavior of evolutionary algorithms (EAs) on longpath problems, in order to get a better insight about possible EA dynamics within the class of unimodal fitness functions in Hamming space.

Longpath problems [10] have been introduced to deal with the notion of problem difficulty for optimization algorithms. Unimodality can involve difficulties for GAs. Isolation (*needle in a haystack*), deception, and multimodality [7], [6], [4] are no longer the only properties that make a search difficult.

Horn's longpath [10] 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 j-th string in the sequence has a fitness value of j. 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 and l denotes the string length.

In this way, a single-bit-flip hill-climber is first led 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 [15] presents a generalization of Horn's longpath to a class of fitness functions called *long k-paths*. The long k-path $P(k,l)$ (where $l-1$ is a multiple of k) is defined by a recursion with respect to the length l of the strings, with $P(k,1) = \{0,1\}$ as the base path. Given the long k-path $P(k,l)$, $P(k,l+k)$ is constructed as follows: A subpath S_0 of $P(k,l+k)$ is created by taking the path $P(k,l)$ and by prepending k zeros to each point on $P(k,l)$. A subpath S_1 of $P(k,l+k)$ is created by taking $P(k,l)$ *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 $P(k,l)$: $0\ldots01$, $0\ldots011$, $\ldots$, $001\ldots11$ and $01\ldots11$. The resulting long k-path $P(k,l+k)$ 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 $P(3,7)$. Starting from $P(3,1) = \{0,1\}$, we obtain the sequence $P(3,4)$:

$$\underbrace{0000_1 \quad 0001_2}_{S_0} \quad \underbrace{0011_3 \quad 0111_4}_{Bridge} \quad \underbrace{1111_5 \quad 1110_6}_{S_1}$$

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

0000000_1	0011110_7	1111110_9
0000001_2	0111110_8	1111111_{10}
0000011_3		1110111_{11}
0000111_4		1110011_{12}
0001111_5		1110001_{13}
0001110_6		1110000_{14}
S_0	*Bridge*	S_1

Long k-paths share the following properties:

- 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

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

the bits of at least one of the blocks $kj+1, \ldots, k(j+1)$, for $j \in \{0, \ldots, (l-1)/k\}$, must be flipped. Such a jump will be called a *shortcut*.

- The length of the long k-path is given by $|P(k,l)| = (k+1)2^{(l-1)/k} - k + 1$. Note that for a fixed string length the path becomes shorter with increasing k.

Experiments demonstrate that a GA with complementary crossover (a one-point crossover between a string and its complement) exploits a kind of Royal Road structure [13], [14] present in the long k-path [9]. 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. Authors of [9] expect that complementary crossover approximates one-point crossover if the latter is used in combination with a sufficiently large population. However, no conclusion can be drawn about the asymptotic behavior of the GA.

Rudolph [15] gives $O(l^{k+1}/k)$ as an asymptotic upper bound of the expected time spent by a (1+1)-ES (evolution strategy) 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 much 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 a 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, we restrict to the case of $1/l$ mutation, and the exact expected time for convergence is calculated for both fixed and l-dependent k. The first section illustrates the construction of the long k-path, and introduces some notations. Section II is devoted to the study of expected convergence time for $k = l^\alpha$. Section III 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 the mutation rate choice is also discussed.

I. The (k,l)-path and search algorithm

A. *Construction of the $(1,l)$-path*

This is done recursively with respect to l. Let $P(1,l)$ be the $(1,l)$-path. A subpath S_0 is created by prepending 0 to each string in path $P(1,l)$ and subpath S_1 by prepending 1 to each string in the reverse of path $P(1,l)$. The $(1,l+1)$-path is obtained by concatenating subpaths S_0 and S_1. The (k,l)-path can now be built for any k as follows.

B. *Construction of the (k,l)-path*

We denote by L the ratio:

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

Notice that L should be an integer so that the corresponding (k,l)-path could exist.

1) Start with the path $P(1,L)$, which contains 2^L strings with length L. As an example, for $L=3$:

$$\begin{array}{l} 0 < \begin{array}{l} 0 < \begin{array}{c}0\\1\end{array}\\ 1 < \begin{array}{c}1\\0\end{array}\end{array}\\ 1 < \begin{array}{l} 1 < \begin{array}{c}0\\1\end{array}\\ 0 < \begin{array}{c}1\\0\end{array}\end{array}\end{array} \iff \begin{array}{c}000\\001\\011\\010\\110\\111\\101\\100\end{array}$$

2) Substitute for each 0 (resp. each 1) a substring with length k consisting of 0 (resp. 1). This yields a sequence of 2^L strings with length $l-1$. For $L=3$:

$$\begin{array}{l} 0\ldots0 < \begin{array}{l} 0\ldots0 < \begin{array}{c}0\ldots0\\1\ldots1\end{array}\\ 1\ldots1 < \begin{array}{c}1\ldots1\\0\ldots0\end{array}\end{array}\\ 1\ldots1 < \begin{array}{l} 1\ldots1 < \begin{array}{c}0\ldots0\\1\ldots1\end{array}\\ 0\ldots0 < \begin{array}{c}1\ldots1\\0\ldots0\end{array}\end{array}\end{array}$$

3) Add to each leaf an ultimate branching that consists alternatively of $< \begin{array}{c}0\\1\end{array}$ and $< \begin{array}{c}1\\0\end{array}$. This yields the so-called skeleton of the (k,l)-path, which consists of 2^{L+1} strings with length l. This is actually a subsequence extracted from the (k,l)-path as built in [15]. Notice that the strings labeled $\{2j+1, 2j+2\}$ are at distance larger than k for all $j = 0, ..., 2^L - 1$. For $L=3$:

$$\begin{array}{l} 0\ldots0 < \begin{array}{l} 0\ldots0 < \begin{array}{l} 0\ldots0 < \begin{array}{c}0\\1\end{array}\\ 1\ldots1 < \begin{array}{c}1\\0\end{array}\end{array}\\ 1\ldots1 < \begin{array}{l} 1\ldots1 < \begin{array}{c}0\\1\end{array}\\ 0\ldots0 < \begin{array}{c}1\\0\end{array}\end{array}\end{array}\\ 1\ldots1 < \begin{array}{l} 1\ldots1 < \begin{array}{l} 0\ldots0 < \begin{array}{c}0\\1\end{array}\\ 1\ldots1 < \begin{array}{c}1\\0\end{array}\end{array}\\ 0\ldots0 < \begin{array}{l} 1\ldots1 < \begin{array}{c}0\\1\end{array}\\ 0\ldots0 < \begin{array}{c}1\\0\end{array}\end{array}\end{array}\end{array}$$

4) To get the whole path from the skeleton, simply add bridges between successive pairs of 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 $*0....01*$. This completes the construction of the (k,l)-path with $(k+1)2^L - k + 1$ strings.

Remarks

1. Although the path takes only a small fraction of the search space for large l, a process starting at a random point can get through some strings which are less than k-bits distant from optimum (or from other strings of the path). This situation is not considered in the following study: We assume that the process starts on the first string of the path, and hence focus on studying the likelihood of k-bit-flip shortcuts.

2. In the following, only elitist evolution strategies are studied, but some experiments show that the lack of elitism of (μ, λ)-ES can improve the convergence speed when starting from a random point. A (1,5)-ES (with $k = 4$, $l = 45$) takes a *shortcut* from the beginning of the path towards the global optimum. In fact, once it reaches the beginning of the path, the $(1,5)$-ES process does not find at once fitter strings on the path and falls down more or less immediately. This situation is repeated a number of times, which maximizes the chances of hitting the global optimum (at successive visits of the first point of the path). On the other hand, when λ/μ increases (e.g., $(1,30)$-ES), the process follows the path.

C. *The process $X(t)$*

In the following, $\mathbf{e}_i := 0.....0$ denotes the first string of the path, $\mathbf{e}_f := 1..10....0$ the last one. T stands for the first hitting time of string $\mathbf{e}_f$ by the process $X(t)$. The search follows a process $X(t)$ starting from $X(0) = \mathbf{e}_i$, and evolves as a (1+1)-ES: At each generation $t \in \mathbb{N}$, build some mutated string $X'(t)$ from $X(t)$ by independently inverting each bit of $X(t)$ with probability l^{-1}. If $X'(t)$ lies farther in the path 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 order to characterize the distribution of the mutations used in the process $X(t)$, we first introduce some new notation:

- $\mathcal{MU}$ denotes the set of all possible mutations $(\{\text{ invert, no-invert}\}^l)$

- $\mathcal{MU}_j$ denotes the set of mutations that invert exactly j bits (we have $\mathcal{MU} = \cup_{j=0}^{l} \mathcal{MU}_j$).

- J is the random variable representing the number of bits flipped per bitstring with the $1/l$ mutation:

$$p_j^J = \mathbb{P}(J = j) = C_j^l l^{-j}(1 - l^{-1})^{l-j}. \tag{2}$$

To select a mutation from $\mathcal{MU}$, first choose j with probability p_j^J, then choose μ uniformly in $\mathcal{MU}_j$. The sequence of mutations of the process $X(t)$ are independent and identically distributed with distribution:

$$p_\mu^{\mathcal{MU}} = \sum_{j=0}^{l} p_j^J \frac{\mathbb{I}_{\mu \in \mathcal{MU}_j}}{|\mathcal{MU}_j|}. \tag{3}$$

D. *Waiting times of some particular events*

Let $\mathcal{S}$ be the family that consists of $L = (l-1)/k$ shortcuts denoted by $\mathcal{S}_j$ for $j = 1, ..., L$. The shortcut $\mathcal{S}_j$ is the operator which inverts the k bits whose labels lie within the segment $[(j-1)k+1, jk+1]$.

Similarly let $\mathcal{S}'$ be the family which consists of all the shortcuts which invert at least the k successive bits whose labels lie within one of the intervals $[(j-1)k+1, jk+1]$, $j = 1, ..., L$. Yet, it is clear that $\mathcal{S} \subset \mathcal{S}'$, and $\mathcal{S}'$ actually contains many more shortcuts than $\mathcal{S}$.

Lemma *I.1: 1.* For any fixed j, the waiting time $T_{\mathcal{S}_j}$ for the shortcut $\mathcal{S}_j$ obeys a geometric distribution with mean $1/p_{\mathcal{S}_1}^{\mathcal{MU}}$:

$$\mathbb{P}\left(T_{\mathcal{S}_j} = t\right) = p_{\mathcal{S}_1}^{\mathcal{MU}} \left(1 - p_{\mathcal{S}_1}^{\mathcal{MU}}\right)^{t-1} \text{ for any } t \geq 1.$$

2. The waiting time $T_{\mathcal{S}'}$ for some shortcut of the family $\mathcal{S}'$ obeys a geometric distribution with mean $1/\sum_{m \in \mathcal{S}'} p_m^{\mathcal{MU}}$.

Proof: At each time $t \in \mathbb{N}$, the probability that a subclass $\mathcal{S}'' \subset \mathcal{MU}$ of mutations occurs is: $\sum_{m \in \mathcal{S}''} p_m^{\mathcal{MU}}$ Since the mutations are independent and identically distributed, the statistical distribution of the first time $T_{\mathcal{S}''}$ when $\mathcal{S}''$ occurs is given by:

$$\mathbb{P}(T_{\mathcal{S}''} > t) = \left(1 - \sum_{m \in \mathcal{S}''} p_m^{\mathcal{MU}}\right)^t.$$

Computing $\mathbb{P}(T_{\mathcal{S}''} = t) = \mathbb{P}(T_{\mathcal{S}''} > t-1) - \mathbb{P}(T_{\mathcal{S}''} > t)$ establishes the result. ■

Lemma *I.2:* If $l \gg 1$, then the means of the waiting times $T_{\mathcal{S}_j}$ and $T_{\mathcal{S}'}$ are respectively $1/p_{\mathcal{S}_1}^{\mathcal{MU}} \simeq el^k$ and $1/\sum_{m \in \mathcal{S}'} p_m^{\mathcal{MU}} \simeq l^k/L$.

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

$$p_{\mathcal{S}_1}^{\mathcal{MU}} = l^{-k}(1 - l^{-1})^{l-k}.$$

For $k \ll l$, we have $p_{\mathcal{S}_1}^{\mathcal{MU}} = l^{-k}e^{-1}$ since $(1 - 1/n)^n \to e^{-1}$ as $n \to \infty$ which proves the first statement of the lemma. The probability that some shortcut of $\mathcal{S}'$ occurs is:

$$P_{\mathcal{S}'}^{(l)} = \sum\nolimits_{m \in \mathcal{S}'} p_m^{\mathcal{MU}}.$$

If we denote by $\mathcal{S}_j'$ the mutations which invert at least the k successive bits whose labels lie within the interval $[(j-1)k+1, jk+1]$, then we clearly have:

$$P_{\mathcal{S}'}^{(l)} \leq \sum_{j=1}^{L} P_{\mathcal{S}_j'}^{(l)} = Ll^{-k}. \tag{4}$$

Further we introduce the set $\hat{S}_j'$ of all mutations that invert at least the bits at positions $[(j-1)k+1, jk+1]$, and that do not invert the successive k bits at positions $[(n-1)k+1, nk+1]$ for $n \neq j$. We have the recurrence:

$$P_{\hat{S}_1'}^{(l)} = l^{-k}(1 - P_{S'}^{(l-k)}).$$

Since $P_{S'}^{(l-k)} \to 0$ for $l \gg 1$, we have that

$$\lim_{l \to \infty} l^k P_{\hat{S}_1'}^{(l)} = 1.$$

Back to the the family $\mathcal{S}'$, we have

$$\liminf_{l \to \infty} l^k L^{-1} P_{S'}^{(l)} \geq \liminf_{l \to \infty} l^k L^{-1} P_{\cup_j \hat{S}_j'}^{(l)} = \lim_{l \to \infty} l^k P_{\hat{S}_1'}^{(l)} = 1.$$

Combining the latter inequality with Eq. (4) yields:

$$l^k L^{-1} P_{\mathcal{S}'}^{(l)} \to 1 \text{ for } l \gg 1,$$

which completes the proof of the lemma. ∎

LEMMA *I.3:* The waiting times between two successive jumps of the process $X(t)$ are independent from each other and have means bounded by l below and by el above if $l \gg 1$.

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

$$p_{jump} \geq l^{-1}(1-l^{-1})^{l-1} \simeq e^{-1} l^{-1}.$$

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, the probability p_{jump} that a jump occurs is always bounded above by:

$$\begin{aligned} p_{jump} &\leq \sum_{j=1}^{k-1} l^{-j}(1-l^{-1})^{l-j} + P_{\mathcal{S}'}^{(l)} \\ &\simeq e^{-1} l^{-1} + l^{-k+1}/k \\ &\underset{\text{for } k\geq 2}{\leq} (e^{-1} + 2^{-1}) l^{-1} \leq l^{-1}. \end{aligned}$$

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

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

II. ASYMPTOTIC STUDY FOR LARGE l AND $k = l^\alpha$

This subsection considers the case of $k = [l^\alpha]$ and large values of l. 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 $O(l^2)$.

A. Main results

The main result of the section is given in Proposition II.1. It states that the expected convergence time is exponential for $k = [l^\alpha]$.

PROPOSITION *II.1:* Assume that $k = [l^\alpha]$ for $\alpha \in (0,1)$.[2]

1. If $\alpha < 1/2$, then

$$\lim_{l\to\infty} \frac{1}{l^\alpha \ln(l)} \ln \mathbb{E}[T] = 1. \tag{5}$$

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

$$\lim_{l\to\infty} \frac{1}{l^{1-\alpha} \ln(2)} \ln \mathbb{E}[T] = 1. \tag{6}$$

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

The estimates which are used in the proof are basically Lemmas I.1-I.3. We shall also show that shortcuts speed up the convergence only in the case $\alpha < 1/2$. If $\alpha \geq 1/2$, the path is entirely visited before any shortcut is likely to happen.

B. Proof of Proposition II.1

We now give an equivalent construction of the process $X(t)$. Recall that for the process $X(t)$ (defined in Section I-C), the mutations are chosen in $\mathcal{MU}$ according to the distribution $p_\mu^{\mathcal{MU}}$ defined by (3). We introduce:

- **A sequence of independent mutations** $(\mu_t)_{t\geq 1}$ chosen in $\mathcal{MU} \setminus \mathcal{S}'$, each with probability
$$\frac{p_\mu^{\mathcal{MU}}}{\sum_{m\notin\mathcal{S}'} p_m^{\mathcal{MU}}} \mathbb{I}_{\mu\notin\mathcal{S}'}.$$
- **A process** $X_{ws}(t)$ without shortcuts starting from $X_{ws}(0) = \mathbf{e}_i$, the first point of the path, defined as follows: mutation μ_{t+1} introduced here above is applied to string $X_{ws}(t)$, and the mutated string is accepted if it lies farther than $X_{ws}(t)$ in the (k,l)-path, otherwise $X_{ws}(t+1) = X_{ws}(t)$. In the following T_{ws} stands for the first hitting time of the last string $\mathbf{e}_f$ by the process X_{ws}.
- **A sequence of random variables** $(T_j)_{j\geq 0}$ such that $T_0 = 0$ and $T_j = T_{j-1} + \sigma_j$, where (σ_j) are independent geometric variables with parameter $p^\sigma := \sum_{m\in\mathcal{S}'} p_m^{\mathcal{MU}}$:
$$\forall n \in \mathbb{N}^*,\ \mathbb{P}(\sigma_j = n) = p^\sigma \left(1-p^\sigma\right)^{n-1}.$$
Notice that σ_j obeys the same distribution as $T_{\mathcal{S}'}$.
- **A sequence of independent mutations** $(\mu'_j)_{j\geq 1}$ chosen in $\mathcal{S}'$, each with probability
$$\frac{p_\mu^{\mathcal{MU}}}{\sum_{m\in\mathcal{S}'} p_m^{\mathcal{MU}}} \mathbb{I}_{\mu\in\mathcal{S}'}.$$

We then **construct a process** $\bar{X}(t)$ by setting $j = 1$, then iterating:

1. While $T_{j-1} \leq t < T_j - 1$, apply mutation μ_{t+1} (from the above sequence) to string $\bar{X}(t)$ and accept the mutated if it lies farther than $\bar{X}(t)$ in the path, otherwise $\bar{X}(t+1) = \bar{X}(t)$.
2. If $t = T_j - 1$, apply mutation μ'_{j+1} to string $\bar{X}(t)$ and accept the mutated if it lies farther than $\bar{X}(t)$ in the path, otherwise $\bar{X}(t+1) = \bar{X}(t)$.
3. Indent $j \to j+1$ and go to step 1.

Let us prove that the law of the process $\bar{X}(t)$ is the same as that of $X(t)$. Since the selection rule is the same for both processes (following a (1+1)-ES scheme), we only need to prove that the mutations of the process $\bar{X}(t)$ obey the same distribution as the mutations of $X(t)$. First, we prove the following Lemma.

LEMMA *II.2:* If $(T_i)_{i\geq 0}$ is the sequence of random variables as described here above and

$$G(n) = \mathbb{P}\left(n \in \{T_j\}_{j\geq 0}\right) = \sum_{i\geq 0} \mathbb{P}(T_i = n),$$

then, for all $n \in \mathbb{N}^*$, we have $G(n) = p^\sigma$.

Proof: Let us denote by F the generating function of σ_i and by Φ the generating function of the sequence $\{G(n), n = 0, 1, \ldots\}$:

$$F(z) = \sum_{n\geq 0} z^n \mathbb{P}(\sigma_1 = n), \;\; \Phi(z) = \sum_{n\geq 0} z^n G(n).$$

Applying [8, Theorem 2, ch III-4] establishes that Φ is characterized by the equation :

$$\Phi(z) = 1 + F(z)\Phi(z). \tag{7}$$

One can then check that the choice:

$$G(n) = p^\sigma \text{ for } n \geq 1, \text{ and } G(0) = 1$$

involves $\Phi(z) = [1-(1-p^\sigma)z]/(1-z)$ which fulfills Eq. (7) since $F(z) = p^\sigma z/[1-(1-p^\sigma)z]$, which characterizes the function $G(n)$. ■

Let $(\bar{\mu}_n)$ be the sequence of mutations of $\bar{X}$. We now prove that the distribution of $(\bar{\mu}_n)$ is the same as that of (μ_n).

$$\begin{aligned} \mathbb{P}(\bar{\mu}_n = \mu) &= \mathbb{P}(n \in \{T_j\}_{j\geq 1}, \bar{\mu}_n = \mu) \\ &\quad + \mathbb{P}(n \notin \{T_j\}_{j\geq 1}, \bar{\mu}_n = \mu) \\ &= \mathbb{P}(n \in \{T_j\}_{j\geq 1})\mathbb{P}(\mu'_1 = \mu) \\ &\quad + \mathbb{P}(n \notin \{T_j\}_{j\geq 1})\mathbb{P}(\mu_1 = \mu) \\ &= p^\sigma \frac{p_\mu^{\mathcal{MU}}}{\sum_{m\in\mathcal{S}'} p_m^{\mathcal{MU}}} \mathbb{I}_{\mu\in\mathcal{S}'} \\ &\quad + (1-p^\sigma)\frac{p_\mu^{\mathcal{MU}}}{\sum_{m\notin\mathcal{S}'} p_m^{\mathcal{MU}}} \mathbb{I}_{\mu\notin\mathcal{S}'}. \end{aligned}$$

Using the fact that $p^\sigma = \sum_{m\in\mathcal{S}'} p_m^{\mathcal{MU}}$, we obtain the desired result:

$$\mathbb{P}(\bar{\mu}_n = \mu) = p_\mu^{\mathcal{MU}}.$$

The mutation distributions of $X(t)$ and $\bar{X}(t)$ are identical, hence the processes $X(t)$ and $\bar{X}(t)$ obey the same law. In the following we study the convergence time T of the process $\bar{X}(t)$. Within this framework, $T_{\mathcal{S}'} = T_1$ which shows that T_{ws} and $T_{\mathcal{S}'}$ are independent. We can split all the possible ways to go from $\mathbf{e}_i$ to $\mathbf{e}_f$ into two complementary groups:

1. **G1**: $T_{ws} < T_{\mathcal{S}'}$,
2. **G2**: $T_{ws} \geq T_{\mathcal{S}'}$.

In the following, to improve readability of the text, we often consider the asymptotic exponential laws of the waiting times studied above, rather than the exact geometrical laws. Indeed, if τ_0 is a random variable with geometric distribution and mean $t_0 \gg 1$, then for any positive real τ we have:

$$\mathbb{P}(\tau_0 \geq \tau t_0) = (1 - 1/t_0)^{\tau t_0} \simeq \exp(-\tau)$$

since $(1-x/n)^n \to \exp(-x)$ as $n \to \infty$. Note however that the proof and conclusions would follow in the same manner using the exact geometrical law.

B.1 Study of the process X_{ws}

On the one hand, the process X_{ws} visits at most every string of the whole path (that is to say at most $N_l := (k+1)2^L - k + 1$ strings, where $L = (l-1)/k$), and the waiting times between two jumps have means bounded above by el. Thus:

$$\mathbb{E}[T_{ws}] \leq elN_l \simeq elk2^L. \tag{8}$$

On the other hand, since no shortcut from the family $\mathcal{S}'$ happens for X_{ws}, the process must visit at least one over two strings of the skeleton (see Subsection I-B), that is to say at least $N_s^l := 2^L$ strings. strings. Furthermore the waiting times between two jumps are statistically independent and have means bounded below by l. Hence

$$\mathbb{E}\left[T_{ws}\right] \geq l2^L. \tag{9}$$

Let us denote by τ_i the waiting times for the successive jumps of the process $X(t)$. We have:

$$\mathbb{E}[T_{ws}^2] \leq \mathbb{E}[(\sum_1^{N_l} \tau_i)^2] = \sum_1^{N_l} \mathbb{E}[\tau_i^2] + \sum_{i\neq j}^{N_l} \mathbb{E}[\tau_i]\mathbb{E}[\tau_j].$$

Since the variables τ_i are exponential with mean bounded from above by el:

$$\mathbb{E}[T_{ws}^2] \leq 2(el)^2 N_l^2. \tag{10}$$

By applying a standard large deviations principle for the sum of independent random variables we get the following proposition:

PROPOSITION *II.3:* For any $\delta > 0$, there exists some $c_\delta > 0$ such that:

$$\begin{aligned} \mathbb{P}\left(T_{ws} < (1-\delta)lN_l^s\right) &\leq \exp(-c_\delta lN_l^s), && (11) \\ \mathbb{P}\left(T_{ws} \geq (1+\delta)elN_l\right) &\leq \exp(-c_\delta elN_l). && (12) \end{aligned}$$

Proof: Apply [16, Theorem 3-8]. ■

If we consider the convergence time T of the process $\bar{X}(t)$, the expectation of T writes:

$$\begin{aligned} \mathbb{E}[T] &= \mathbb{E}_{ws}\mathbb{E}_{\mathcal{S}'}[T/T_{ws} < T_{\mathcal{S}'}]\mathbb{P}(G_1) \\ &\quad + \mathbb{E}_{ws}\mathbb{E}_{\mathcal{S}'}[T/T_{\mathcal{S}'} \leq T_{ws}]\mathbb{P}(G_2), \end{aligned} \tag{13}$$

where $\mathbb{E}_{ws}$ is the expectation with respect the distribution of the sequence $(\mu_n)_{n\geq 1}$, and $\mathbb{E}_{\mathcal{S}'}$ is the expectation with respect the distribution of the sequence $(\mu'_j, T_j)_{j\geq 1}$. In the following, the notation $\mathbb{E}$ stands for $\mathbb{E}_{ws}\mathbb{E}_{\mathcal{S}'}$. We shall now estimate the four terms of the right-hand side of this equality.

B.2 Estimate of $\mathbb{P}(G_1) = 1 - \mathbb{P}(G_2)$

$\mathbb{P}(G_1) = \mathbb{P}(T_{ws} < T_{\mathcal{S}'})$. Lemmas I.1-I.2 establish that $\mathbb{P}(T_{\mathcal{S}'} > t) = \exp(-Ll^{-k}t)$ for any t. Since T_{ws} and $T_{\mathcal{S}'}$ are independent, we then have:

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

• Case $k=[l^\alpha]$, $\alpha<1/2$
We decompose the expectation with respect to the event: $T_{ws}\geq lN_l^s/2$:

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

Eq. (11) then implies that the first term of the right-hand side is smaller than the second one, so that:

$$\mathbb{P}(G_1)\leq 2\exp\left(-\frac{lN_l^sL}{2l^k}\right)\simeq 2\exp\left(-\frac{2^{l^{1-\alpha}-1}}{l^{\alpha-2+\alpha}}\right)\underset{l\to\infty}{\longrightarrow}0. \tag{15}$$

• Case $k=[l^\alpha]$, $\alpha\geq 1/2$
Applying Jensen's inequality to Eq. (14):

$$\mathbb{P}(G_1)\geq\exp\left(-\frac{L\mathbb{E}[T_{ws}]}{l^k}\right).$$

From Eq. (8) we get:

$$\mathbb{P}(G_1)\geq\exp\left(-\frac{Lelk2^L}{l^k}\right).$$

The fact that $Lelk2^L/l^k\simeq el^2 2^{l^{1-\alpha}}/l^{l^\alpha}\ll 1$ then implies:

$$\mathbb{P}(G_2)\leq 1-\exp\left(-\frac{el^2 2^{l^{1-\alpha}}}{l^{l^\alpha}}\right)\simeq\frac{el^2 2^{l^{1-\alpha}}}{l^{l^\alpha}}\underset{l\to\infty}{\longrightarrow}0. \tag{16}$$

B.3 Expectation of T under $G1$

Under G1, $T_{S'}>T_{ws}$, thus $T=T_{ws}$ and

$$\begin{aligned}\mathbb{E}[T/T_{S'}>T_{ws}]&=\mathbb{E}[T_{ws}/T_{S'}>T_{ws}]\\&\leq\mathbb{E}[T_{ws}]\leq elN_l.\end{aligned}\tag{17}$$

• We now restrict ourselves to the case $k=[l^\alpha]$, $\alpha\geq 1/2$

$$\begin{aligned}\mathbb{E}[T/T_{S'}>T_{ws}]&=\frac{\mathbb{E}[T_{ws}\mathbb{1}_{T_{S'}>T_{ws}}]}{\mathbb{P}(T_{S'}>T_{ws})}\\&\geq\mathbb{E}_{ws}[T_{ws}\mathbb{P}_{S'}(T_{S'}>T_{ws})].\end{aligned}$$

Since $T_{S'}$ obeys an exponential distribution, we have $\mathbb{P}_{S'}(T_{S'}>T_{ws})=\exp(-T_{ws}/\mathbb{E}[T_{S'}])$, so that:

$$\mathbb{E}[T/T_{S'}>T_{ws}]\geq\mathbb{E}\left[T_{ws}\exp(-\frac{T_{ws}}{\mathbb{E}[T_{S'}]})\right]. \tag{18}$$

Further, using the fact that $1-e^{-\rho}\leq\rho\wedge 1$, and the bounds on $\mathbb{E}[T_{ws}]$ and $\mathbb{E}[T_{ws}^2]$ of Eqs. (8) and (10), we have:

$$\begin{aligned}&\mathbb{E}\left[T_{ws}\left(1-\exp(-\frac{T_{ws}}{\mathbb{E}[T_{S'}]})\right)\right]\\&\leq\rho\mathbb{E}[T_{ws}\mathbb{1}_{T_{ws}<\rho\mathbb{E}[T_{S'}]}]+\mathbb{E}[T_{ws}\mathbb{1}_{T_{ws}\geq\rho\mathbb{E}[T_{S'}]}]\\&\leq\rho\mathbb{E}[T_{ws}]+\mathbb{E}[T_{ws}^2]^{1/2}\mathbb{P}(T_{ws}>\rho\mathbb{E}[T_{S'}])^{1/2}\\&\leq\mathbb{E}[T_{ws}](\rho+2\mathbb{P}(T_{ws}>\rho\mathbb{E}[T_{S'}])^{1/2}).\end{aligned}\tag{19}$$

By Eq. (12) and the fact that $lN_l\ll E[T_{ws}]/l\sim l^{\alpha-2}l^{l^\alpha}$, we have $\mathbb{P}(T_{ws}>\rho\mathbb{E}[T_{S'}]/l)\to 0$ at exponential rate as $l\to\infty$. Hence, if we choose $\rho=1/l$ in Eq. (19), then we get that

$$\mathbb{E}\left[T_{ws}\left(1-\exp(-\frac{T_{ws}}{\mathbb{E}[T_{S'}]})\right)\right]\leq\frac{1}{l}\mathbb{E}[T_{ws}].$$

Substituting into Eq. (18) establishes that:

$$\mathbb{E}[T/T_{S'}>T_{ws}]\geq\mathbb{E}[T_{ws}](1-\frac{1}{l}). \tag{20}$$

Combining (17) and (20) we obtain:

$$\lim_{l\to\infty}\frac{1}{l^{1-\alpha}\ln(2)}\ln\mathbb{E}\left[T/T_{S'}>T_{ws}\right]=1. \tag{21}$$

B.4 Expectation of T under G2

In the worst case, the following sequence of events (happening after $T_{S'}$) insures that the process reaches $\mathbf{e}_f$, whatever the intermediate event: shortcut $\mathcal{S}_1$, then $\mathcal{S}_2$, etc, until $\mathcal{S}_L$. Hence

$$\mathbb{E}[T/T_{S'}<T_{ws}]\leq\mathbb{E}[T_{S'}/T_{S'}<T_{ws}]+\sum_{j=1}^{L}\mathbb{E}[T_{S_j}]. \tag{22}$$

Note that the expectations in the second term of the right-hand side of the above inequality are no longer conditional. Indeed, after the first shortcut happens (first term), the subsequent events can be described as independent exponential variables, and in particular independent from $T_{S'}$.

LEMMA *II.4:* Let τ_0 be a random variable with exponential distribution and mean t_0. Then

$$\mathbb{E}[\tau_0/\tau_0<\tau]=t_0 f(\frac{\tau}{t_0}),\text{ where } f(r)=\frac{1-e^{-r}(r+1)}{1-e^{-r}}.$$

Proof: Write

$$\mathbb{E}[\tau_0/\tau_0<\tau]=\frac{\mathbb{E}[\tau_0\mathbb{1}_{\tau_0<\tau}]}{\mathbb{P}(\tau_0<\tau)}=\frac{\int_0^\tau\frac{s}{t_0}e^{-s/t_0}ds}{\int_0^\tau\frac{1}{t_0}e^{-s/t_0}ds}$$

and compute the integrals. ■

Combining the fact that f is uniformly bounded by 1 and Lemma I.2, Eq. (22) now reads:

$$\mathbb{E}[T/T_{S'}<T_{ws}]\leq\mathbb{E}[T_{S'}]+eLl^k\leq e(1/L+L)l^k. \tag{23}$$

In the best case we need only one shortcut of the family $\mathcal{S}'$ to reach $\mathbf{e}_f$.

$$\mathbb{E}[T/T_{S'}<T_{ws}]\geq\mathbb{E}[T_{S'}/T_{S'}<T_{ws}]. \tag{24}$$

First, applying Lemma II.4 to the right term of the latter inequality, and then using the fact that $f(t)\geq(t/4)\wedge(1/2)$ (with $t=T_{ws}/\mathbb{E}[T_{S'}]$) we get:

$$\begin{aligned}\mathbb{E}[T_{S'}/T_{S'}<T_{ws}]&\geq\mathbb{E}[T_{S'}]\,\mathbb{E}[f(\frac{T_{ws}}{\mathbb{E}[T_{S'}]})]\\&\geq\mathbb{E}[T_{S'}]\mathbb{E}[\frac{T_{ws}}{4\mathbb{E}[T_{S'}]}\wedge\frac{1}{2}]\\&=\mathbb{E}[\frac{T_{ws}}{4}\wedge\frac{\mathbb{E}[T_{S'}]}{2}].\end{aligned}$$

- We now restrict ourselves to the case $k=[l^\alpha]$, $\alpha<1/2$

$$\begin{aligned}\mathbb{E}[\frac{T_{ws}}{4}\wedge\frac{\mathbb{E}[T_{S'}]}{2}] &= \mathbb{E}[\frac{T_{ws}}{4}\mathbb{I}_{T_{ws}<2\mathbb{E}[T_{S'}]}] \\ &\quad+\frac{\mathbb{E}[T_{S'}]}{2}\mathbb{P}(T_{ws}\geq 2\mathbb{E}[T_{S'}]) \\ &\geq \frac{\mathbb{E}[T_{S'}]}{2}(1-\mathbb{P}(T_{ws}<2\mathbb{E}[T_{S'}])).\end{aligned}$$

Since $\alpha<1/2$ we have $\mathbb{E}[T_{S'}]\ll N_l^s$, then by Eq. (11):

$$\mathbb{P}(T_{ws}<2\mathbb{E}[T_{S'}])\leq\mathbb{P}(T_{ws}<\frac{l}{2}N_l^s)\leq e^{-c_{1/2}lN_l^s}.$$

Hence, in the asymptotic framework $l\to\infty$), we have

$$\mathbb{E}[T/T_{S'}<T_{ws}]\geq\mathbb{E}[T_{S'}](1-\frac{1}{l}). \qquad (25)$$

Combining (23) and (25), if $k=[l^\alpha]$ then for $\alpha<1/2$:

$$\lim_{l\to\infty}\frac{1}{l^\alpha\ln(l)}\ln\mathbb{E}\left[T/T_{S'}<T_{ws}\right]=1. \qquad (26)$$

B.5 Global expectation of T

If $\alpha<1/2$, then we substitute (15), (17) and (26) into Eq. (13):

$$\begin{aligned}\mathbb{E}[T] &= \underbrace{\mathbb{E}_{ws}\mathbb{E}_{S'}[T/T_{ws}<T_{S'}]}_{\leq el^{2-\alpha}2^{l^{1-\alpha}}}\quad\underbrace{\mathbb{P}(G_1)}_{\leq 2\exp-\frac{l^{2-\alpha}2^{l^{1-\alpha}}}{2l^{l^\alpha}}} \\ &\quad+\underbrace{\mathbb{E}_{ws}\mathbb{E}_{S'}[T/T_{S'}<T_{ws}]}_{\sim l^{l^\alpha}}\underbrace{\mathbb{P}(G_2)}_{\simeq 1}.\end{aligned}$$

Here $A_l\sim B_l$ means $\ln A_l/\ln B_l\to 1$ as $l\to$. The first term of the right-hand side goes to 0 as $l\to\infty$, which proves the first point of the proposition.

If $\alpha\geq 1/2$, then we substitute (16), (21) and (23) into Eq. (13):

$$\begin{aligned}\mathbb{E}[T] &= \underbrace{\mathbb{E}_{ws}\mathbb{E}_{S'}[T/T_{ws}<T_{S'}]}_{\sim 2^{l^{1-\alpha}}}\underbrace{\mathbb{P}(G_1)}_{\simeq 1} \\ &\quad+\underbrace{\mathbb{E}_{ws}\mathbb{E}_{S'}[T/T_{S'}<T_{ws}]}_{\leq el^{1-\alpha}l^{l^\alpha}}\quad\underbrace{\mathbb{P}(G_2)}_{\leq el^2 2^{l^{1-\alpha}}/l^{l^\alpha}}.\end{aligned}$$

The first term of the right-hand side is equivalent to $2^{l^{1-\alpha}}$:

$$\liminf_{l\to\infty}\frac{\ln\mathbb{E}[T]}{l^{1-\alpha}\ln(2)}\geq 1.$$

More exactly the first term is bounded by $el^{2-\alpha}2^{l^{1-\alpha}}$ (by Eq. (17)), while the second term is bounded by $e^2l^{3-\alpha}2^{l^{1-\alpha}}$:

$$\limsup_{l\to\infty}\frac{\ln\mathbb{E}[T]}{l^{1-\alpha}\ln(2)}\leq 1+\limsup_{l\to\infty}\frac{\ln\left(el^{2-\alpha}+e^2l^{3-\alpha}\right)}{l^{1-\alpha}\ln(2)}=1.$$

Combining these results proves the second point of the Proposition.

Remarks

1. Note that for $\alpha\geq 1/2$, the waiting time for one shortcut occurrence (Eq. (23) under G2) is much longer than the upper bound of the expected convergence time when following the entire path. This is not enough, however, to describe the effective behavior of the process as indicated 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 $\ln\ln l/\ln l$.

III. Asymptotic behavior for large l and fixed k

This section is devoted to the case of a fixed value of k in the asymptotic framework $l\gg 1$. The main result is stated in Subsection III-A 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, with 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 also 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 III-E. The influence of mutation probability is investigated in Section III-F. Finally Subsection III-G compares different population-based evolution strategies behavior.

A. Main results

Let us consider the first hitting time T of the process $X(t)$ (defined in Section I-C).

Proposition *III.1:* The expectation of the first hitting time T for large l is

$$\mathbb{E}[T]\underset{l\to\infty}{\simeq}\frac{el^{k+1}}{k\ln(l/k)}, \qquad (27)$$

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 III-B. ■

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

Proposition *III.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:

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

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

Proof: See Subsection III-B. ■

In the above expression of density $p_{\bar{T}}(\tau)$, the integer m stands for the number of shortcuts which are necessary to reach the final string, τ_j, $j = 1, \ldots, m$ are the times between two shortcuts, and n_j, $j = 1, \ldots, m$ are the ranges of the corresponding shortcuts, as explained in Subsection III-B.

By integrating $\tau p_{\bar{T}}(\tau)$ and $\tau^2 p_{\bar{T}}(\tau)$ with respect to τ, 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.

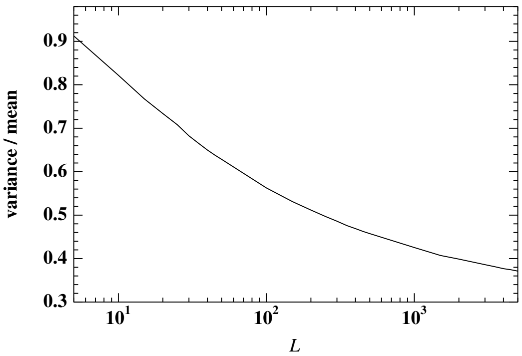

Fig. 1. Normalized variance of the variable T, which, at the first order only, depends on the ratio $L = (l-1)/k$.

The probability densities of T for typical and relevant values of the parameters k and l, are plotted in Figures 2-5. 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 $\mathcal{S}_2$, ..., $\mathcal{S}_L$). 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 III-E.

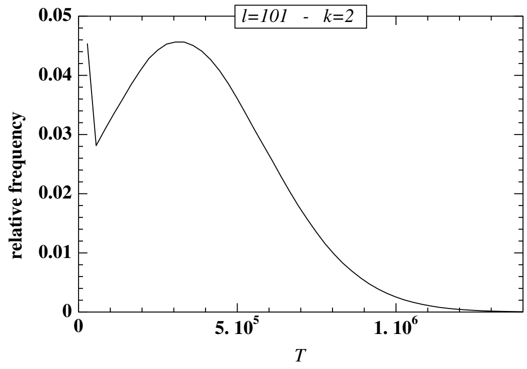

Fig. 2. Theoretical histogram of the random variable T for $k = 2$ and $l = 101$.

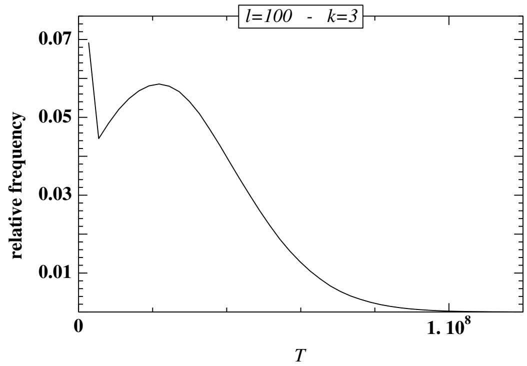

Fig. 3. Theoretical histogram of the random variable T for $k = 3$ and $l = 100$.

B. Proofs of Propositions III.1 and III.2

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

$$\begin{array}{rcl}
\mathcal{M}_1 & = & \{0..0**\ \ldots\ldots\ldots\ldots..*\} \bigcup \{bridges\} \\
\mathcal{M}_2 & = & \{1..11..1**\ \ldots\ldots..*\} \bigcup \{bridges\} \\
\mathcal{M}_3 & = & \{1..10..01..1**\ \ldots.*\} \bigcup \{bridges\} \\
\vdots & & \vdots \\
\mathcal{M}_{L-1} & = & \{1..10..0\ \ldots\ 0..01..1*\} \bigcup \{bridges\} \\
\mathcal{M}_L & = & \{1..10..0\ \ldots\ 0..00..0*\}
\end{array}$$

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 propor-

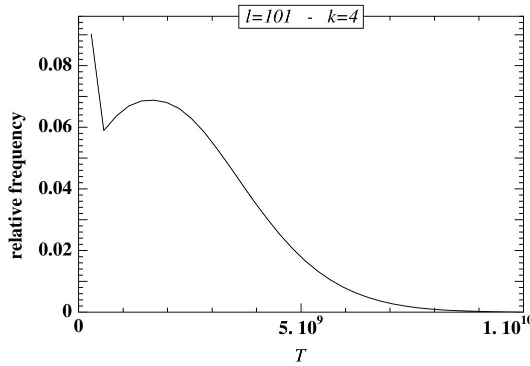

Fig. 4. Theoretical histogram of the random variable T for $k = 4$ and $l = 101$.

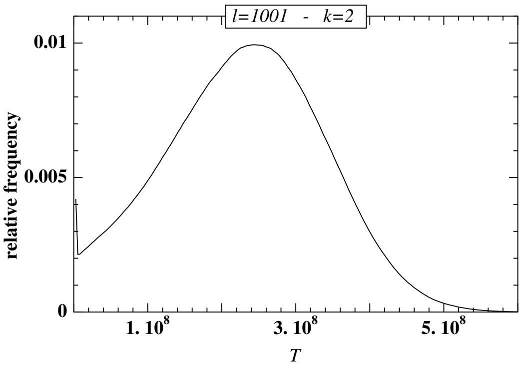

Fig. 5. Theoretical histogram of the random variable T for $k = 2$ and $l = 1001$.

tion 2^{-j} of all the strings of the path. The longpath also presents some remarkable symmetries:

- Applying $\mathcal{S}_1$ to string n^o j, with $j < N_l/2$, yields the string n^o $N_l - j + 1$.
- Applying $\mathcal{S}_2$ to string n^o j, with $j < N_l/4$, yields 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:

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

Let us discuss *when* the process enters $\mathcal{M}_1^c$ (the complementary of $\mathcal{M}_1$).

- 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 $\mathcal{S}_1$, then we may distinguish two situations:
  1. If one of the $\mathcal{S}_j$, $j \geq 3$, occurs before $\mathcal{S}_2$, then the process must still visit an exponentially large number of strings, which takes an exponentially long time.
  2. The only way that could produce a non-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 latter event, that we denote by $\mathcal{E}$, therefore provides a probability $1/L$ to enter $\mathcal{M}_2$ (this is the probability that $\mathcal{S}_2$ occurs before 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 $\mathcal{M}_1^c$, hence, obeys an exponential distribution with mean el^k (Lemmas I.1-I.2). It will be denoted in the following by $el^k\tau_1$:

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

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 occur 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 $\mathcal{M}_2$ in 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 this particular j by N_1. Since the waiting times for the shortcuts $\mathcal{S}_i$ obey independent exponential distributions with mean el^k, we have:

$$\begin{aligned}\mathbb{P}\left(N_1 = j/\tau_1 = \tau\right) &= \mathbb{P}\left(\tau_2', ..., \tau_{j-1}' > \tau, \tau_j' < \tau/\tau_1 = \tau\right)\\ &= \left(1 - e^{-\tau}\right)e^{-(j-1)\tau}.\end{aligned}$$

We can then iterate the above arguments. We then find that we need to do it M times, where M is the first time that $\sum_{i=1}^m N_j$ reaches L, which means that the process enters $\mathcal{M}_L$.

As a consequence, still denoting $L = (l-1)/k$, we can describe the first hitting time T as follows.

For large l, the first hitting time T obeys the distribution of:

$$T = el^k\bar{T}, \tag{28}$$

where $\bar{T}$ is defined in the following way. Consider a sequence of independent and identically distributed real-valued random variables τ_j, $j = 1, 2, ...$ with exponential density: $p(\tau) = \mathbb{I}_{\tau \geq 0}e^{-\tau}$. 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

$$p_\tau(n) = e^{-(n-1)\tau}(1 - e^{-\tau}), \qquad n = 1, 2, ... \tag{29}$$

If we denote by M_L the first exit time:

$$M_L = \inf\left\{m \geq 1, \sum_{j=1}^{m} N_j \geq L\right\}, \tag{30}$$

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

$$\bar{T} = \sum_{j=1}^{M_L} \tau_j. \tag{31}$$

The closed-form expression of the probability density of $\bar{T}$ given in Proposition III.2 can then be deduced directly from the definition of $\bar{T}$.

Roughly speaking, from (28), we expect that $\mathbb{E}[T] = el^k\mathbb{E}[\bar{T}]$, from (31) that $\mathbb{E}[\bar{T}] \simeq \mathbb{E}[M_L]\mathbb{E}[\tau_1] = \mathbb{E}[M_L]$ and from (30) 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:

$$\mathbb{P}(N_j = n) = \int_0^\infty p(\tau)p_\tau(n)d\tau = \frac{1}{n(n+1)}. \tag{32}$$

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 the 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 *III.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 $\delta > 0$:

$$\begin{aligned}
\frac{\ln L}{L}\mathbb{E}[M_L] &\overset{L\to\infty}{\longrightarrow} 1,\\
\left(\frac{\ln L}{L}\right)^2 \mathbb{E}[M_L^2] &\overset{L\to\infty}{\longrightarrow} 1,\\
\mathbb{P}\left(\left|\frac{\ln L}{L}M_L - 1\right| \geq \delta\right) &\overset{L\to\infty}{\longrightarrow} 0.
\end{aligned}$$

PROPOSITION *III.4:* For any $q = 1, 2$, we have

$$\frac{\mathbb{E}[\bar{T}^q]}{\mathbb{E}[M_L^q]} \underset{L\to\infty}{\longrightarrow} 1. \tag{33}$$

Combining these propositions yields the statement of Proposition III.1 ($q = 1$) and the first point of Proposition III.2 ($q = 2$). In the following, we shall first study the statistical distribution of the variable M_L and prove Proposition III.3. The study of the distribution of $\bar{T}$ then proves Proposition III.4.

C. Proof of Proposition III.3.

We aim at studying the statistical distribution of the stopping time M_L as defined by (30). We shall use a Tauberian theorem:

LEMMA *III.5:* 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:

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

Proof: This is precisely the Theorem 4 in [3, Vol. 2, ch. XIII-5]. ■

Let us denote by S_m the partial sums:

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

We can then 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 III.5 we then get:

$$\mathbb{E}[M_L] \underset{L\to+\infty}{\simeq} \frac{L}{\ln L}. \tag{34}$$

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 III.5 we get:

$$\mathbb{E}[M_L^2] \underset{L\to+\infty}{\simeq} \frac{L^2}{(\ln L)^2}. \tag{35}$$

Combining (34) with (35), we obtain the convergence in probability of the variable $\frac{\ln L}{L} M_L$ to 1:

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

This result is asymptotic, that is to say, it holds true for large L. However, precisely analyzing the convergence of M_L, it appears that the convergence rate is very slow (logarithmic). If we deal with values of L which are of order 100 or 1000, it is then necessary to be very careful when using Proposition III.3. The mean $\mathbb{E}[M_L]$ actually behaves as $L/\ln L$ even for -relatively- small values of L, as shown by Figures 6-7. 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 $L \gg 10^9$ (Figures 6-7). In Figures 8-9 we have plotted the computed histograms of the variable M_L for different values of L.

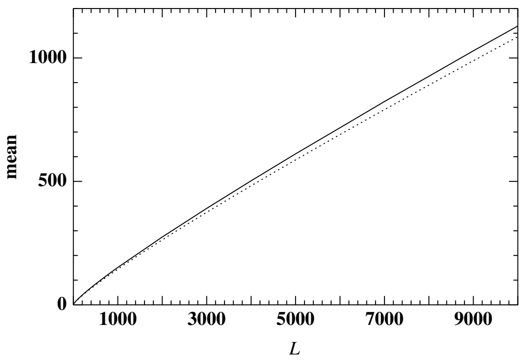

Fig. 6. Mean of the random variable M_L as a function of L (solid line). In dashed line is plotted the curve $L \mapsto L/\ln L$.

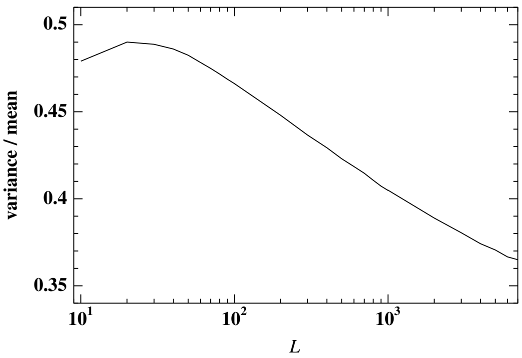

Fig. 7. Normalized variance $\mathbb{E}\left[M_L^2 - \mathbb{E}[M_L]^2\right]^{1/2}/\mathbb{E}\left[M_L\right]$ of the random variable M_L as a function of L.

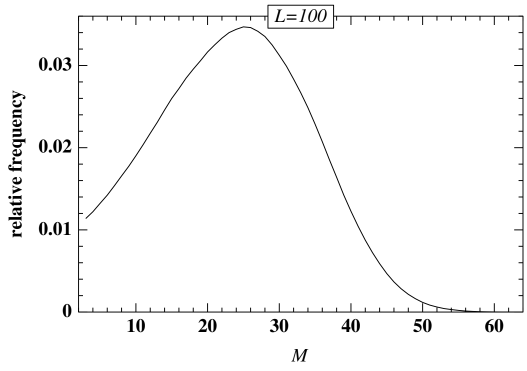

Fig. 8. Theoretical histogram of the random variables M_L for $L = 100$.

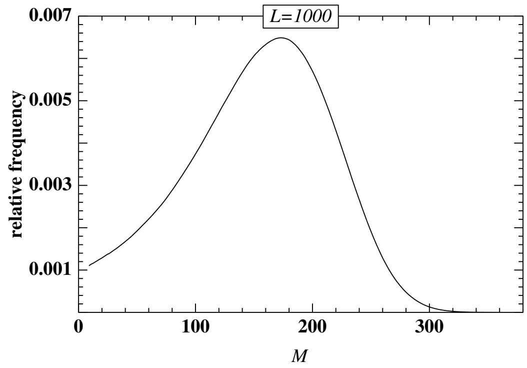

Fig. 9. Theoretical histogram of the random variables M_L for $L = 1000$.

LEMMA *III.6:* Let $p \in \{1,2\}$, $q \in \mathbb{N}^*$ and $\psi_1,\ldots,\psi_q$ be continuous functions from $\mathbb{N}^*$ into $\mathbb{R}^+$ satisfying $0 < a \leq \psi_j(.) \leq b < \infty$. Then:

$$\left|\frac{\mathbb{E}[M_L^p \psi_1(N_1)\ldots\psi_q(N_q)]}{\mathbb{E}[M_L^p]\mathbb{E}[\psi_1(N_1)]\ldots\mathbb{E}[\psi_q(N_q)]} - 1\right| \underset{L\to\infty}{\longrightarrow} 0.$$

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 becomes large.

Proof: For simplicity we prove the result for $q = 1$. The extension to an arbitrary q is straightforward. By the definition of M_L:

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

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:

$$\begin{aligned}\lim_{L\to\infty}\left(\frac{\ln L}{L}\right)^p \mathbb{E}[M_L^p \psi_1(N_1)] &= \sum_{n=1}^{\infty}\frac{1}{n(n+1)}\psi_1(n)\\ &= \mathbb{E}[\psi_1(N_1)].\end{aligned}$$

■

D. Proof of Proposition III.4.

By Proposition III.2, we can compute $\mathbb{E}[\bar{T}]$ by integrating the density $p_{\bar{T}}(t)$ multiplied by t. Since

$$\int_0^\infty e^{-n\tau}(1-e^{-\tau})d\tau = \frac{1}{n(n+1)}, \quad (36)$$

$$\int_0^\infty \tau e^{-n\tau}(1-e^{-\tau})d\tau = \frac{2n+1}{n^2(n+1)^2}, \quad (37)$$

we get:

$$\begin{aligned}\mathbb{E}[\bar{T}] &= \sum_{m=1}^{\infty}\sum_{n_1,\dots,n_m=1}^{\infty} \mathbb{I}_{n_1+\dots+n_{m-1}<L\le n_1+\dots+n_m}\\ &\quad\times\sum_{i=1}^{m}\frac{2n_i+1}{n_i^2(n_i+1)^2}\prod_{j\neq i}^{m}\frac{1}{n_j(n_j+1)}.\end{aligned}$$

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:

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

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 III.6. 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 as:

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

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.

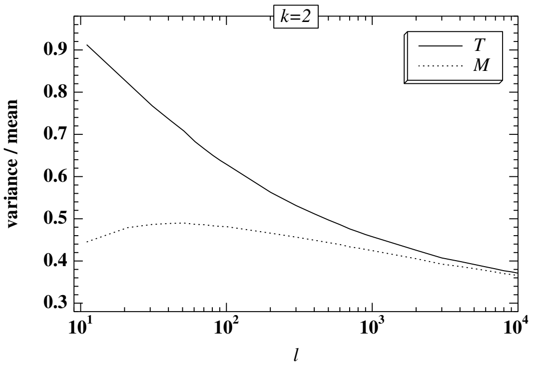

Fig. 10. Normalized variances of the variables T and M_L ($L = (l-1)/k$).

E. 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 *III.7:* There exists a set of realizations Ω_1 with probability

$$\mathbb{P}(\Omega_1) \underset{l\to\infty}{\simeq} kl^{-1}, \quad (38)$$

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

$$\mathbb{P}\left(T \ge 2el^{k-1}k\tau/\Omega_1\right) \underset{l\to\infty}{\simeq} \exp(-\tau) \text{ for any real } \tau. \quad (39)$$

Proposition III.7 shows that the expected convergence time under Ω_1 is equal to $2el^{k-1}k$, which is roughly l^2 times shorter than the global expected time $\frac{el^{k+1}}{k\ln(l/k)}$ by Proposition III.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 *III.8:* Assume that we deal with a set of N processes which evolve independently. Then if N is large, denoting $c = Nk/l$, there exists a set of realizations Ω_N with probability

$$\mathbb{P}(\Omega_N) \underset{l\to\infty}{\simeq} 1-e^{-c}, \quad (40)$$

such that the conditional expectation of the infimum T of the first hitting times of the N processes is:

$$\mathbb{E}\left[T/\Omega_N\right] \underset{l\to\infty}{\simeq} 2ef(c)l^{k-1}k, \quad (41)$$

where $f(c) = \frac{\mathrm{Ei}(c)-\gamma-\ln c}{e^c-1}$, γ is the Euler constant $\gamma \simeq 0.57$ and Ei is the exponential integral function [1, p. 228] (see

Figure 11). More exactly, the conditional distribution of T given Ω_N is:

$$T \underset{l\to\infty}{\simeq} -2el^{k-1}k\ln\left(\frac{1-Z}{c}\right), \tag{42}$$

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

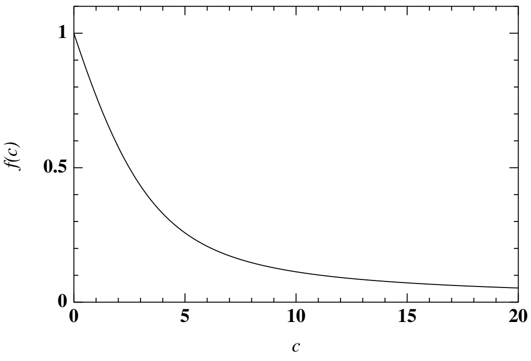

Fig. 11. Function $c \mapsto f(c)$.

E.1 Proof of Proposition III.7.

We still represent the statistical distribution of the first hitting time T as (28)-(31). In this framework, we define the set Ω_1 as the set of the realizations which fulfill the condition $M_L = 1$:

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

Since the distribution of N_1 is given by (32), 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 n$^o j$; at time $el^k\tau_1$, by shortcut $\mathcal{S}_1$, the process jumps at string n$^o(k+1)2^L - k + 1 - j$; 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 distributions and means 1:

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

The result of the Proposition then follows readily.

E.2 Proof of Corollary III.8.

We adopt the same notations as (28)-(31). 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:

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

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 $c = Nk/l$. The global first hitting time is the minimum of the first hitting times corresponding to each process:

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

where the T_i are independent and identically distributed random variables defined as in (28)-(31). Furthermore, given Ω_N, the statistical distribution of T is:

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

where $I = \{i = 1, ..., N \text{ such that } M_{L,i} = 1\}$ and T_i are independent and identically distributed random variables defined as (39). We have

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

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

$$\begin{aligned}
\mathbb{P}(T \geq t/\Omega_N) &= \sum_{j\geq 1} \mathbb{P}\left(|I| = j, T_i \geq t \ \ \forall i \in I/\Omega_N\right) \\
&= \sum_{j\geq 1} \frac{e^{-c}}{1-e^{-c}}\frac{c^j}{j!}\mathbb{P}(T_1 \geq t/\Omega_1)^j.
\end{aligned}$$

Since $\mathbb{P}(T_1 \geq 2el^{k-1}k\tau/\Omega_1) = \exp(-\tau)$, we get by summing :

$$\mathbb{P}(T \geq 2el^{k-1}k\tau/\Omega_N) = \frac{\exp\left(ce^{-\tau}\right) - 1}{e^c - 1}.$$

The interested reader can then check that this is exactly the statistical distribution of the random variable defined by (42).

F. Optimization of the mutation probability

In the following we investigate the influence of c (for stochastic c/l-mutation) on the time to convergence of the (1+1)-ES process on long k-path problems. This shows that the best mutation rate is k/l.

PROPOSITION *III.9:* Let us assume that the mutation probability is c/l.

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 III. In particular its mean is:

$$\mathbb{E}[T_d] \underset{l\to\infty}{\simeq} \frac{e^d}{d^k}\frac{l^{k+1}}{k\ln(l/k)}. \tag{43}$$

2. 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, for each instant, the probability that the shortcut $\mathcal{S}_j$ occurs is:

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

so that the statistical distribution of $T_{\mathcal{S}_j}$ is given by:

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

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 $d = 1$. ■

G. Discussion and comparison of different algorithms

In this subsection we discuss 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 than a (1+1)-ES. Moreover, the best strategy to minimize the time to convergence is to run several independent (1+1)-ES processes simultaneously.

This latter result has already been stated in Subsection III-E : Corollary (III.8) 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 III.1). Independent processes are therefore very advantageous and much more rapid than one process in terms of number of generations, as well as in terms of number of fitness 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 III-B). 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 achieve shortcuts at the same order among the families $\mathcal{S}_j$. In fact, the string resulting from the first shortcut contaminates, more or less quickly, the other processes because of the selection pressure.

In the case of a $(1+\lambda)-ES$, the string resulting from the first shortcut immediately colonizes the population. But the expected time for a shortcut and for any improvement gets N times smaller, hence the 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 increasing 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:

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

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. However, this is not the probability of the event Ω_1, but an upper bound ($\mathcal{S}_1$ does not necessarily happens first). 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 lose any chance of realizing the event Ω_1. $(\mu+\lambda)-ES$ only speeds the convergence with a factor of λ in terms of number of generations and implies no gain in number of fitness evaluations.

IV. Summary and conclusion

This paper investigates the behavior of a (1+1)-ES process using the $1/l$ bit flipping mutation on Rudolph's long k-paths, in the asymptotic framework $l \gg 1$. Both cases of variable and fixed k values are addressed.

First, for $k = l^\alpha$, we prove that expected convergence time is exponential. 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 a fixed value of k, the 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 a very slow rate. Further, the study of the statistical distribution of convergence time T 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 for taking 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 shortcut 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.

A number of interesting issues are raised by the results of this paper on the *long k-path* problem:

- For fixed k, independent parallel (1+1)-ES processes perform a lot better with no migration than they perform with migration of fit individuals between the processes. Equivalently, increasing the population size decreases convergence speed in number of evaluations and generations compared to independent (1+1)-ES processes.
- EA dynamic can be very sensitive to rare events related to the exploration properties of the $1/l$ mutation (for fixed k, shortcuts yield a very big variance in convergence time). This phenomena is confirmed by experimental work [12] which demonstrates that exceptional properties of operators sometimes reflect EA behavior more accurately than average typical properties do. However, EA dynamic is no longer sensitive to shortcuts for $k \geq \sqrt{l}$, resulting in a convergence behavior that is predictable.
- This study provides upper bounds on the complexity of simple search procedures in l-dimensional spaces, for big l values, as detailed below.

If we combine the complexity results of this paper with some previous results, we conclude that the convergence time of mutation only hill-climbers, ranges from a linear (Onemax) to an exponential (Long k-path) rate in terms of l. In the OneMax case, convergence in number of fitness evaluations is quicker with the 1-bit-flip mutation than it is with the $1/l$ mutation [5], whereas the opposite happens for the long k-path problem. On the other hand, random walks directed by both mutations require the same numbers of fitness evaluations to hit a target point of the search space [5].

Note finally that another important result follows from this study, as it 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 evolutionary 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 [11], where a *controlled path* fitness function is constructed so that the crossover and mutation evolutionary algorithms empirically follow a *path* that has been arbitrarily chosen in the search space.

To sum up, longpaths (and more generally *controlled paths*) provide an example of smooth and unimodal landscapes on which EA convergence time can be very long and predictable. An interesting but currently unanswered question is whether real world problems can present such a longpath structure? Note that there exist physical systems (spin glass [2]) which evolve very slowly towards an equilibrium state, through successive unstable states (the glass of medieval cathedrals, for example, is now reaching its final equilibrium state, more crystalline and fragile, after 400 years of slow decay).

References

[1] M. Abramowitz and I. Stegun. *Handbook of mathematical functions.* Dover publications, New York, 1965.

[2] D. Chowdhury. *Spin glasses and other frustrated systems.* World Scientific, Singapore, 1986.

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

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

[5] J. Garnier, L. Kallel, and M. Schoenauer. Rigourous results of the first hitting times of some Markov chains. Technical Report 389, CMAP, Ecole Polytechnique, Juin 1998. to appear in Journal of Evolutionary Computation.

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

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

[8] I. Guikhman and A. Skorokhod. Introduction à la théorie des processus aléatoires. Mir, Moscou, 1986.

[9] 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.

[10] 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.

[11] L. Kallel and B. Naudts. Candidate longpaths for the simple genetic algorithm. In W. Banzhaf and C. Reeves, editors, *Foundations of Genetic Algorithms 5.* Morgan Kaufmann, 1998. to appear.

[12] L. Kallel and M. Schoenauer. A priori predictions of operator efficiency. In J.-K. Hao, E. Lutton, E. Ronald, M. Schoenauer, and D. Snyers, editors, *Artificial Evolution'97*, LNCS, pages 287–299. Springer Verlag, October 1997.

[13] 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.

[14] 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. Morgan Kaufmann, 1993.

[15] 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.

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