Brazilian Journal of Probability and Statistics
2011, Vol. 25, No. 2, 121–144
DOI: 10.1214/09-BJPS113
© Brazilian Statistical Association, 2011

On the estimation of the potential of Sinai's RWRE

Pierre Andreoletti

Université d'Orléans

Abstract. We consider a one-dimensional random walk in random environment. We prove that the logarithm of the local time can be used as an estimator of the random environment. We give a constructive method allowing us to locally built, up to a translation, the random potential associated to the environment from a single trajectory of the random walk.

1 Introduction and results

In this paper we are interested in Sinai's walk, denoted $(X_l, l \in \mathbb{N})$, that is, a one-dimensional random walk in random environment (RWRE) with three conditions on the random environment: two necessary hypotheses to get a recurrent process (see Solomon (1975)) which is not a simple random walk and the uniform ellipticity hypothesis which allows us to have a good control on the fluctuations of the random environment.

The asymptotic behavior of such walk has been understood by Sinai (1982): this walk is subdiffusive $X_n \approx (\log n)^2$, and at given instant n is localized in the neighborhood of a well-defined point of the lattice. It is well known—see, for example, Zeitouni (2001) for a survey—that this behavior is strongly dependent of the random environment or, equivalently, to the associated random potential defined Section 2.1.

The question we solve here is the following: given a single trajectory of the random walk $(X_l, 1 \leq l \leq n)$ where the time n is fixed, can we estimate the trajectory of the random potential where the walk lives? Let us remark that the law of this potential is unknown as well.

In their paper, Adelman and Enriquez (2004) are interested in the question of the distribution of the random environment that could be deduced from a single trajectory of the walk; on the other hand, our purpose is to get an approximation of the trajectory of the random potential.

That kind of result is of great interest to biophysicists; indeed Baldazzi et al. (2006) are interested in a method to predict the sequence of DNA molecules. They model the unzipping of the molecule as a one-dimensional biased random walk for the fork position (number of open base pair) k in this landscape. The elementary opening ($k \to k+1$) and closing ($k \to k-1$) transitions happen with a probability

Key words and phrases. Random potential, random walk, Sinai's regime, nonparametric statistics.
Received December 2008; accepted December 2009.

that depends on the unknown sequence. This probability of transition follows an Arrhénius law which is close to the one we discuss here. The question they answer is: given an unzipping signal can we predict the unzipping sequence? Their approach is based on a Bayesian inference method which gives very good probabilities of prediction for a large amount of data. This means, in term of the walk, several trajectory on the same environment.

Our approach is purely probabilistic; it is based on good properties of the *local time* of the random walk which is the amount of time the walk spends on the different points of the lattice. We treat a general case with very little information on the random environment. We are able to reconstruct the difference of the random potential in a significant interval where the walk spends most of its time. Our proof is based on improvements of the results of Andreoletti (2006), in particular, in a weak law of large numbers for the local time on the neighborhood of the point of localization of the walk.

The largest part of this paper is devoted to the proof of a theoretical result (Theorem 1.6). We also present, at the end of the document, numerical simulations to illustrate our result. We give the main steps of the algorithm we use to rebuilt the random potential only by considering a trajectory of the walk. As an introduction we would like to comment on one of these simulations (see Figure 1).

In black we have represented the logarithm of the local time and in grey the potential associated to the random environment. First, note that we get a good approximation on a large neighborhood of the bottom of the valley around the coordinate -80. Outside this neighborhood and especially after the coordinate -20, the approximation is not precise at all. We will explain this phenomena by the fact that after the walk has reached the bottom of the valley, it will not return frequently to the points with coordinate larger than -20, so we lose information for this part of the latice.

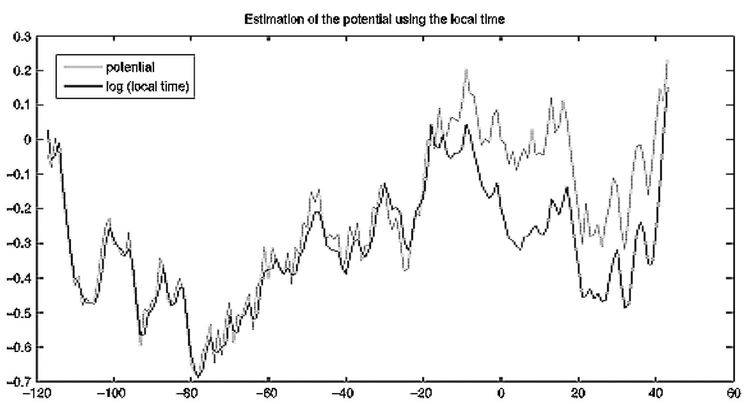

Figure 1 *The logarithm of the local time (in black) and the random potential (in grey).*

Our method of estimation gives us two crucial pieces of information: a confidence interval for the differences of potential in sup-norm, on an observable set of sites "sufficiently" visited by the walk, and a localization result for the bottom of the valley linked with the coordinate where the local time achieves its maximum. First, let us define the process.

1.1 Definition of Sinai's walk

Let $\alpha = (\alpha_i, i \in \mathbb{Z})$ be a sequence of i.i.d. random variables taking values in $(0, 1)$ defined on the probability space $(\Omega_1, \mathcal{F}_1, Q)$; this sequence will be called random environment. A random walk in random environment $(X_n, n \in \mathbb{N})$ is a sequence of random variable taking values in $\mathbb{Z}$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that:

• for every fixed environment α, $(X_n, n \in \mathbb{N})$ is a Markov chain with the following transition probabilities, for all $n \geq 1$ and $i \in \mathbb{Z}$

$$\mathbb{P}^{\alpha}[X_n = i + 1 | X_{n-1} = i] = \alpha_i,$$
$$\mathbb{P}^{\alpha}[X_n = i - 1 | X_{n-1} = i] = 1 - \alpha_i.$$

We denote $(\Omega_2, \mathcal{F}_2, \mathbb{P}^{\alpha})$ the probability space associated to this Markov chain.

• $\Omega = \Omega_1 \times \Omega_2$, $\forall A_1 \in \mathcal{F}_1$ and $\forall A_2 \in \mathcal{F}_2$, $\mathbb{P}[A_1 \times A_2] = \int_{A_1} Q(dw_1) \times \int_{A_2} \mathbb{P}^{\alpha(w_1)}(dw_2)$.

The probability measure $\mathbb{P}^{\alpha}[\cdot | X_0 = a]$ will be denoted $\mathbb{P}^{\alpha}_a[\cdot]$, the expectation associated to $\mathbb{P}^{\alpha}_a$: $\mathbb{E}^{\alpha}_a$, and the expectation associated to Q: $\mathbb{E}_Q$.

Now we introduce the hypothesis we will use in all this work. The first two are needed to get a nontrivial RWRE

$$\mathbb{E}_Q\left[\log \frac{1 - \alpha_0}{\alpha_0}\right] = 0, \tag{1.1}$$

$$\mathrm{Var}_Q\left[\log \frac{1 - \alpha_0}{\alpha_0}\right] = \sigma^2 > 0. \tag{1.2}$$

Solomon (1975) shows that under (1.1) the process $(X_n, n \in \mathbb{N})$ is $\mathbb{P}$ almost surely recurrent and (1.2) implies that the model is not reduced to the simple random walk. In addition to (1.1) and (1.2) we will consider the uniform ellipticity hypothesis: there exists $0 < \eta_0 < 1/2$ such that

$$\sup\{x, Q[\alpha_0 \geq x] = 1\} = \sup\{x, Q[\alpha_0 \leq 1 - x] = 1\} \geq \eta_0. \tag{1.3}$$

We call *Sinai's random walk* the random walk in random environment previously defined with the three hypothesis (1.1), (1.2) and (1.3).

Let us define the local time $\mathcal{L}$, at k $(k \in \mathbb{Z})$ within the interval of time $[1, T]$ $(T \in \mathbb{N}^*)$ of $(X_n, n \in \mathbb{N})$

$$\mathcal{L}(k, T) = \sum_{i=1}^{T} \mathbb{1}_{\{X_i = k\}},$$

$\mathbb{1}$ is the indicator function. Let $V \subset \mathbb{Z}$, we denote

$$\mathcal{L}(V,T) = \sum_{j \in V} \mathcal{L}(j,T) = \sum_{i=1}^{T} \sum_{j \in V} \mathbb{1}_{\{X_i = j\}}.$$

To finish, we define the following, associated to $\mathcal{L}$, random variables: $\mathcal{L}^*(n)$ the maximum of the local times (for a given instant n), $\mathbb{F}_n$ the set of all the favorite sites, and k^* the smallest favorite site,

$$\begin{aligned} \mathcal{L}^*(n) &= \max_{k \in \mathbb{Z}} (\mathcal{L}(k,n)), \\ \mathbb{F}_n &= \{k \in \mathbb{Z}, \mathcal{L}(k,n) = \mathcal{L}^*(n)\}, \end{aligned} \tag{1.4}$$

$$k^* = \inf\{|k|, k \in \mathbb{F}_n\}. \tag{1.5}$$

1.2 The random potential and the valleys

From the random environment we define what we will call random potential, let

$$\varepsilon_i = \log \frac{1-\alpha_i}{\alpha_i}, \qquad i \in \mathbb{Z},$$

define:

Definition 1.1. The random potential $(S_m, m \in \mathbb{Z})$ associated to the random environment α is defined in the following way:

$$S_k = \begin{cases} \sum_{1 \le i \le k} \varepsilon_i, & \text{if } k > 0, \\ -\sum_{k+1 \le i \le 0} \varepsilon_i, & \text{if } k < 0, \end{cases}$$

$$S_0 = 0 \qquad \text{(see Figure 2)}.$$

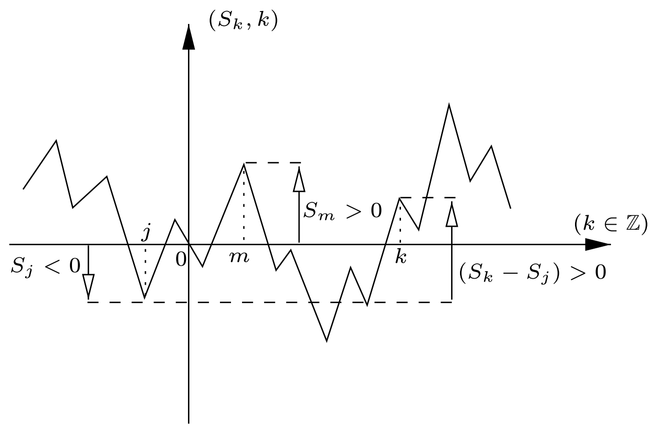

Figure 2 *Trajectory of the random potential.*

Definition 1.2. We will say that the triplet $\{M', m, M''\}$ is a *valley* if

$$S_{M'} = \max_{M' \leq t \leq m} S_t,$$

$$S_{M''} = \max_{m \leq t \leq \tilde{M}''} S_t,$$

$$S_m = \min_{M' \leq t \leq M''} S_t.$$

If m is not unique we choose the one with the smallest absolute value.

Definition 1.3. We will call *depth of the valley* $\{M', m, M''\}$ and we will denote it $d([M', M''])$ the quantity

$$\min(S_{M'} - S_m, S_{M''} - S_m).$$

Now we define the operation of *refinement*.

Definition 1.4. Let $\{M', m, M''\}$ be a valley and let M_1 and m_1 be such that $m \leq M_1 < m_1 \leq M''$ and

$$S_{M_1} - S_{m_1} = \max_{m \leq t' \leq t'' \leq M''} (S_{t'} - S_{t''}).$$

We say that the couple (m_1, M_1) is obtained by a *right refinement* of $\{M', m, M''\}$. If the couple (m_1, M_1) is not unique, we will take the one such that m_1 and M_1 have the smallest absolute value (see Figure 3). In a similar way we define the *left refinement* operation.

We denote $\log_2 = \log\log$, in all this section we will suppose that n is large enough such that $\log_2 n$ is positive.

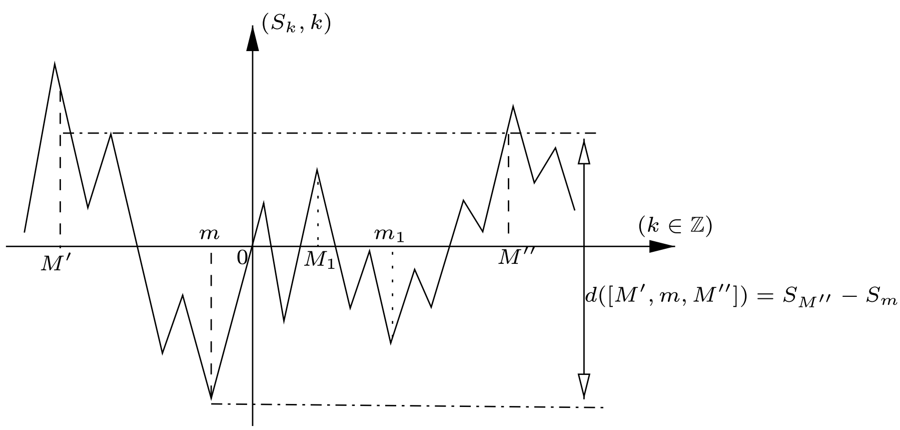

Figure 3 *Depth of a valley and refinement operation.*

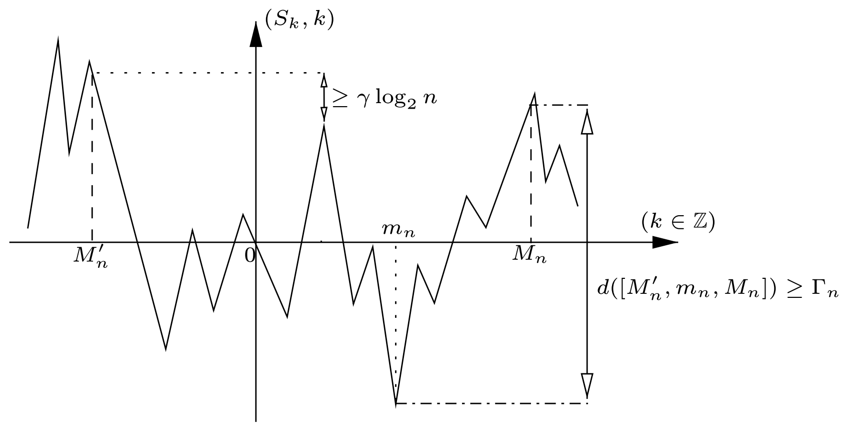

Figure 4 *Basic valley, case $m_n > 0$.*

Definition 1.5. Let $n > 3$, $\gamma > 0$, and $\Gamma_n = \log n + \gamma \log_2 n$, we say that a valley $\{M', m, M''\}$ contains 0 and is of depth larger than Γ_n if and only if

1. $0 \in [M', M'']$,
2. $d([M', M'']) \geq \Gamma_n$,
3. if $m < 0$, $S_{M''} - \max_{m \leq t \leq 0}(S_t) \geq \gamma \log_2 n$, if $m > 0$, $S_{M'} - \max_{0 \leq t \leq m}(S_t) \geq \gamma \log_2 n$.

The basic valley $\{M'_n, m_n, M_n\}$ (see Figure 4). We recall the notion of *basic valley* introduced by Sinai and denoted here $\{M'_n, m_n, M_n\}$. The definition we give is inspired by the work of Kesten (1986). First let $\{M', m_n, M''\}$ be the smallest *valley that contains* 0 *and of depth larger than* Γ_n. Here smallest means that if we construct, with the operation of refinement, other valleys in $\{M', m_n, M''\}$ such valleys will not satisfy one of the properties of Definition 1.5. M'_n and M_n are defined from m_n in the following way: if $m_n > 0$

$$M'_n = \sup\left\{l \in \mathbb{Z}_-, l < m_n, S_l - S_{m_n} \geq \Gamma_n, S_l - \max_{0 \leq k \leq m_n} S_k \geq \gamma \log_2 n\right\},$$

$$M_n = \inf\{l \in \mathbb{Z}_+, l > m_n, S_l - S_{m_n} \geq \Gamma_n\}.$$

If $m_n < 0$

$$M'_n = \sup\{l \in \mathbb{Z}_-, l < m_n, S_l - S_{m_n} \geq \Gamma_n\},$$

$$M_n = \inf\left\{l \in \mathbb{Z}_+, l > m_n, S_l - S_{m_n} \geq \Gamma_n, S_l - \max_{m_n \leq k \leq 0} S_k \geq \gamma \log_2 n\right\}.$$

If $m_n = 0$

$$M'_n = \sup\{l \in \mathbb{Z}_-, l < 0, S_l - S_{m_n} \geq \Gamma_n\},$$

$$M_n = \inf\{l \in \mathbb{Z}_+, l > 0, S_l - S_{m_n} \geq \Gamma_n\}.$$

1.3 Main results

We start with some definitions that will be used all along this work.

Let $x \in \mathbb{Z}$, define

$$T_x = \begin{cases} \inf\{k \in \mathbb{N}^*, X_k = x\}, \\ +\infty, \qquad \text{if such } k \text{ does not exist.} \end{cases} \tag{1.6}$$

Let $n > 1$, $(k, l) \in \mathbb{Z}^2$ and $c_0 > 0$, define:

$$S_{k,l}^n = 1 - \frac{1}{\log n}(S_k - S_l),$$

$$\hat{S}_k^n = \frac{\log(\mathcal{L}(k, n))}{\log n},$$

$$u_n = \frac{c_0 \log_3 n}{\log n}.$$

The random variable $S_{k,l}^n$ is the function of the potential we want to estimate, $\hat{S}_k^n$ is the estimator and u_n is an error function.

Now let us define the following random subset of $\mathbb{Z}$:

$$\mathbb{L}_n^\gamma = \left\{ l \in \mathbb{Z}, \sum_{j=T_{k^*}}^{n} \mathbb{1}_{X_j = l} \geq (\log n)^\gamma \right\},$$

recall that $\gamma > 0$. This set $\mathbb{L}_n^\gamma$ is fundamental for our result, we notice that it depends only on the trajectory of the walk and more especially on its local time: $\mathbb{L}_n^\gamma$ is the set of points for which we are able to give an estimator of the random potential. We will see that this set is large and contains a great amount of the points visited by the walk (see Proposition 1.8). We recall that T_{k^*} is the first time the walk hit the smallest favorite site. In words, $l \in \mathbb{L}_n^\gamma$, if and only if the local time of the random walker in l after the instant T_{k^*} is large enough (larger than $(\log n)^\gamma$). Our main result is the following:

Theorem 1.6. *Assume* (1.1), (1.2) *and* (1.3) *hold, there exist four constants* c_0, c_1, c_2 *and* c_2' *such that for all* $\gamma > 6$, *there exists* n_0 *such that for all* $n > n_0$ *there exists* $G_n \subset \Omega_1$ *with* $Q[G_n] \geq 1 - \phi_1(n)$ *and*

$$\inf_{\alpha \in G_n} \mathbb{P}^\alpha \left[\bigcap_{k \in \mathbb{L}_n^\gamma} \{ |\hat{S}_k^n - S_{k,m_n}^n| < u_n \} \right] \geq 1 - \phi_2(n), \tag{1.7}$$

where

$$\phi_1(n) = \frac{c_1 \gamma \log_2 n}{\log n}, \tag{1.8}$$

$$\phi_2(n) = \frac{c_2 (\log_2 n)^2}{(\log n)^{\gamma/2-2}} + \frac{c_2' (\log_2 n)^8}{(\log n)^{\gamma-6}}. \tag{1.9}$$

The fact that our result depends on m_n seems to be restrictive, we would like to know where is the bottom of the valley only by considering the local time of the walk, so we prove the following:

Proposition 1.7. *Assume* (1.1), (1.2) *and* (1.3) *hold*, *there exists a constant* $c_3 > 0$ *such that for all* $\gamma > 6$, *there exists* n_0 *such that for all* $n > n_0$ *there exists* $G_n \subset \Omega_1$ *with* $Q[G_n] \geq 1 - \phi_1(n)$ *and*

$$\inf_{\alpha \in G_n} \mathbb{P}_0^\alpha\Big[\max_{x \in \mathbb{F}_n} |m_n - x| \leq (\log_2 n)^2\Big] \geq 1 - \phi_3(n), \tag{1.10}$$

$$\inf_{\alpha \in G_n} \mathbb{P}_0^\alpha[|T_{m_n} - T_{k^*}| \leq (\log n)^3] \geq 1 - \phi_3(n), \tag{1.11}$$

where $\mathbb{F}_n$ *is defined in* (1.4), *and* $\phi_3(n) = c_3(\log_2 n)^8/(\log n)^{\gamma-6}$.

Notice that the distance between m_n (coordinate of the minimum of the potential) and a favorite site is negligible comparing to a typical fluctuation of the walk (of order $(\log n)^2$). Thanks to Proposition 1.7 we can replace (1.7) in Theorem 1.6 by

$$\inf_{\alpha \in G_n} \mathbb{P}^\alpha\Big[\bigcap_{k \in \mathbb{L}_n^\gamma} \{|\hat{S}_k^n - S_{k,k^*}^n| < u_n\}\Big] \geq 1 - \phi_2(n).$$

Now let us present a result giving some properties of $\mathbb{L}_n^\gamma$:

Proposition 1.8. *Assume* (1.1), (1.2) *and* (1.3) *hold*, *for all* $\gamma > 6$, *there exists* n_0 *such that for all* $n > n_0$ *there exists* $G_n \subset \Omega_1$ *with* $Q[G_n] \geq 1 - \phi_1(n)$, *and*

$$\inf_{\alpha \in G_n} \mathbb{P}_0^\alpha\big[\mathcal{L}(\mathbb{L}_n^\gamma, n) = n\big(1 - o(1)\big)\big] \geq 1 - \phi_2(n), \tag{1.12}$$

$$\inf_{\alpha \in G_n} \mathbb{P}_0^\alpha[\text{const}(\log n)^2/(\log_2 n)^2 \leq |\mathbb{L}_n^\gamma| \leq \text{const}(\log n \log_2 n)^2] \geq 1 - \phi_2(n). \tag{1.13}$$

The inequality (1.12) shows that the time spent by the walk in $\mathbb{L}_n^\gamma$ is almost all the lifetime of the walk: n, (1.13) shows that the size of $\mathbb{L}_n^\gamma$ is comparable to a typical fluctuation of the walk X.

Remark 1.9. The result we get gives information on the difference of potential $S_k - S_{m_n}$ when $k \in \mathbb{L}_n^\gamma$. So thanks to the definition of S ($S_0 = 0$), if $0 \in \mathbb{L}_n^\gamma$, then we also know the value S_{m_n} and so the full potential in the interval $\mathbb{L}_n^\gamma$. In the other case ($0 \notin \mathbb{L}_n^\gamma$), which also appears with a strictly positive probability, we cannot say anything precise about the value of S_{m_n}. Figure 5 corresponds to a "nice" random environment where S_{m_n} can be deduced from the local time. Conversely, Figure 6 shows an environment for which the walk does not get back to 0 once

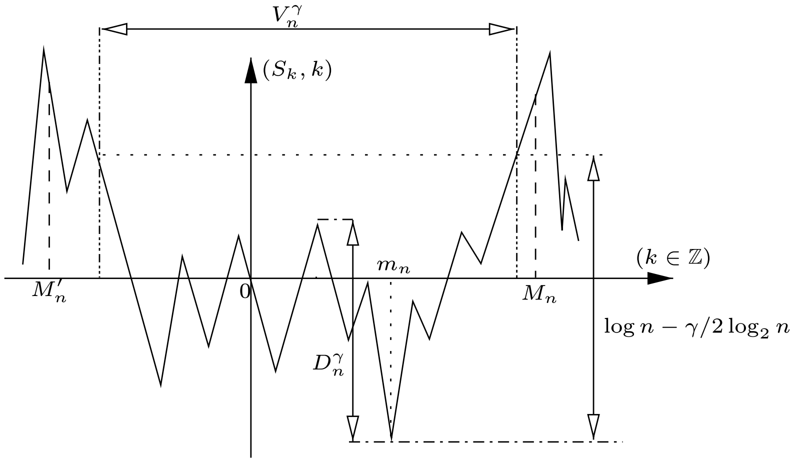

Figure 5 V_n^γ *with* $m_n > 0$, *with* $D_n^\gamma = \max_{0 \le k \le m_n}(S_j - S_{m_n}) \le \log n - \gamma \log_2 n$.

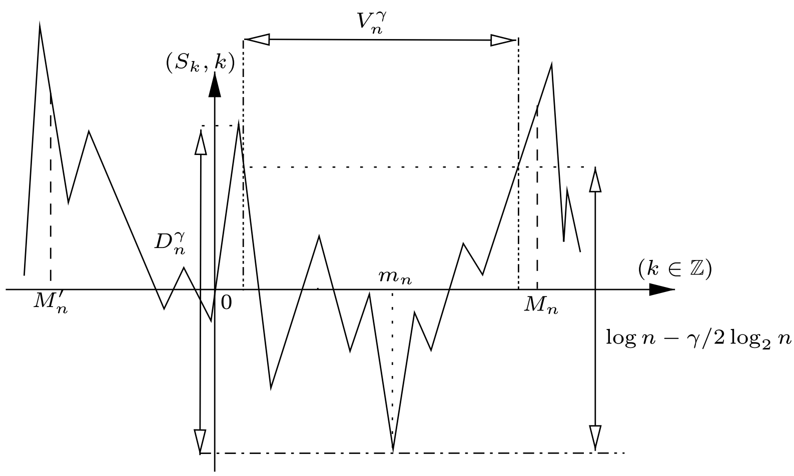

Figure 6 V_n^γ *with* $m_n > 0$, *with* $D_n^\gamma = \max_{0 \le k \le m_n}(S_k - S_{m_n}) > \log n - \gamma \log_2 n$.

it has reached m_n, in this case the value of S_{m_n} cannot be deduced from the local time.

However, notice that to predict the close future of the walk (after the instant n) the knowledge of the differences of potential together with the value of m_n are enough. Indeed the probability of transition of X (for a fixed environment) can be given as an explicit formula of the increments of the potential: for all $i > 0$

$$\alpha_i = \frac{\exp(-(S_i - S_{i-1}))}{\exp(-(S_i - S_{i-1})) + 1}, \tag{1.14}$$

and we have a similar expression when $i \leq 0$.

Theorem 1.6 is known to be the quenched result, which means we work under $\mathbb{P}^{\alpha}$ at a fixed environment α. To get the annealed result, that is to say, that we work under $\mathbb{P}$ instead under $\mathbb{P}^{\alpha}$ we need the following elementary remark:

Remark 1.10. Let $\mathcal{C}_n \in \sigma(X_i, i \leq n)$ and $G_n \subset \Omega_1$, we have:

$$\mathbb{P}[\mathcal{C}_n] = \int_{\Omega_1} Q(d\omega) \int_{\mathcal{C}_n} d\mathbb{P}^{\alpha(\omega)} \tag{1.15}$$

$$\geq \int_{G_n} Q(d\omega) \int_{\mathcal{C}_n} d\mathbb{P}^{\alpha(\omega)}. \tag{1.16}$$

So assume that $Q[G_n] = e_1(n) \geq 1 - \phi_1(n)$ and that for all $\omega \in G_n$, $\int_{\mathcal{C}_n} d\mathbb{P}^{\alpha(\omega)} \equiv e_2(\omega, n) \geq 1 - \phi_2(n)$ we get that

$$\mathbb{P}[\mathcal{C}_n] \geq e_1(n) \times \min_{w \in G_n} (e_2(w, n)) \geq 1 - \phi_1(n) - \phi_2(n). \tag{1.17}$$

Then a simple consequence of Theorem 1.6 is the following:

Corollary 1.11. *Assume* (1.1), (1.2) *and* (1.3) *hold*, *there exist four constants* c_0, c_1, c_2 *and* c_2' *such that for all* $\gamma > 6$, *there exists* n_0 *such that for all* $n > n_0$

$$\mathbb{P}\left[\bigcap_{k \in \mathbb{L}_n^{\gamma}} \{|\hat{S}_k^k - S_{k,k^*}^n| < u_n\}\right] \geq 1 - \phi(n), \tag{1.18}$$

where $\phi(n) = \phi_1(n) + \phi_2(n)$.

We just notice that, for our purpose, the above result is not useful because the aim is to reconstruct one trajectory of the random environment, that is to say, one α. In the above result we get a mean over all the possible random environments.

This paper is organized as follows. In Section 2 we give the proof of Theorems 1.6 (we easily get Corollary 1.11 from Remark 1.10), we have split this proof into two parts, the first one (Section 2.1) deals with the random environment and the other one (Section 2.2) with the random walk itself. Then in Section 2.3 we give the proofs of Propositions 1.7 and 1.8. In Section 3, as an application of our result, we present an algorithm and some numerical simulations. For completeness, we recall in the Appendix some basic facts on birth and death processes.

2 Proof of Theorem 1.6

The proof of a result with a random environment involves both arguments and properties for the random environment and arguments for the random walk itself. We start with the properties we need for the random environment, then we will use it to get the result for the walk.

2.1 Properties needed for the random environment

2.1.1 *Construction of* $(G_n, n \in \mathbb{N})$. Let k and l be in $\mathbb{Z}$, define

$$E_k^\alpha(l) = \mathbb{E}_k^\alpha[\mathcal{L}(l, T_k)] \tag{2.1}$$

in the same way, let $A \subset \mathbb{Z}$, define

$$E_k^\alpha(A) = \sum_{l \in A} \mathbb{E}_k^\alpha[\mathcal{L}(l, T_k)]. \tag{2.2}$$

Definition 2.1. Let $d_0 > 0$, $d_1 > 0$, and $\omega \in \Omega_1$, we will say that $\alpha = \alpha(\omega)$ is a *good environment* if there exists n_0 such that for all $n \geq n_0$ the sequence $(\alpha_i, i \in \mathbb{Z}) = (\alpha_i(\omega), i \in \mathbb{Z})$ satisfies the properties (2.3)–(2.6)

- $\{M_n', m_n, M_n\} \neq \varnothing$, (2.3)
- $M_n' \geq -d_0(\sigma^{-1} \log_2 n \log n)^2, \qquad M_n \leq d_0(\sigma^{-1} \log_2 n \log n)^2.$ (2.4)

Define M_1' and m_1', respectively, the maximizer and minimizer obtained by the first *left refinement* of the valley $\{M_n', m_n, M_n\}$ and in the same way M_1 and m_1, respectively, the maximizer and minimizer obtained by the first *right refinement* of the valley $\{M_n', m_n, M_n\}$.

- $S_{M_1'} - S_{m_1'} \leq \log n - \gamma \log_2 n, \qquad S_{M_1} - S_{m_1} \leq \log n - \gamma \log_2 n,$ (2.5)
- $1 \leq E_{m_n}^\alpha(W_n) \leq d_1(\log_2 n)^2,$ (2.6)

where $W_n = \{M_n', M_n' + 1, \dots, m_n, \dots, M_n\}$.

Define the *set of good environments*

$$G_n = G_n(d_0, d_1) = \{\omega \in \Omega_1, \alpha(\omega) \text{ is a } \textit{good environment}\}. \tag{2.7}$$

Notice that G_n depends on d_0, d_1 and n, however we only make explicit the n dependence.

Proposition 2.2. *There exist three constants* $d_0 > 0$, $d_1 > 0$ *and* $c_1 > 0$ *such that if* (1.1), (1.2) *and* (1.3) *hold, there exists* n_0 *such that for* $n > n_0$

$$Q[G_n] \geq 1 - \phi_1(n), \tag{2.8}$$

where ϕ_1 *is given by* (1.8).

Proof. We can find the proof of this proposition in Andreoletti (2006); see Definition 4.1 and Proposition 4.2. □

2.2 Arguments for the walk

Let $(\rho_1(n), n \in \mathbb{N})$ a strictly positive and strictly decreasing sequence such that $\lim_{n\to\infty} \rho_1(n) = 0$ and for all n large enough, $\rho_1(n) > 1/\log_2 n$. First let us show that Theorem 1.6 is a simple consequence of the following:

Proposition 2.3. *Assume (1.1), (1.2) and (1.3) hold, there exists n_0 such that for all $n > n_0$ there exists $G_n \subset \Omega_1$ with $Q[G_n] \geq 1 - \phi_1(n)$ and*

$$\sup_{\alpha\in G_n}\left\{\mathbb{P}_0^\alpha\left[\bigcup_{k\in\mathbb{L}_n^\gamma}\left\{\left|\frac{\mathcal{L}(k,n)}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)}\right| \leq w_{k,n}\right\}\right]\right\} \geq 1 - \phi_2(n), \tag{2.9}$$

where $w_{k,n} = \rho_1(n)\frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)}$, $\phi_1(n)$ and $\phi_2(n)$ are given just after (1.7).

Taking the logarithm and for n large enough, using Taylor series expansion, we remark that

$$\frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)}\big(1-\rho_1(n)\big) \leq \frac{\mathcal{L}(k,n)}{n} \leq \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)}\big(1+\rho_1(n)\big) \tag{2.10}$$

implies

$$\begin{aligned} &-2\rho_1(n) - \log(E_{m_n}^\alpha(W_n)) \\ &\quad \leq \log\mathcal{L}(k,n) - \log n - \log(E_{m_n}^\alpha(k)) \\ &\quad \leq -\log(E_{m_n}^\alpha(W_n)) + \rho_1(n). \end{aligned}$$

Rearranging the terms and using (A.1) (see the Appendix) we get

$$\frac{1}{\log n}\big(R_n^\alpha(k) - 2\rho_1(n)\big) \leq \hat{S}_k^n - S_{k,m_n}^n \leq \frac{1}{\log n}\big(R_n^\alpha(k) - \rho_1(n)\big) \tag{2.11}$$

where $R_n^\alpha(k) = \log(\frac{\alpha_{m_n}}{\beta_k}a_{k,m_n}) - \log(E_{m_n}^\alpha(W_n))$ and a_{k,m_n} is given by (A.1). Now using (A.3) and Property (2.6) we get the theorem. The proof of Proposition 2.3 is based on four lemmata presented in the following subsections.

2.2.1 *Known facts and local time at m_n.* For all $n \in \mathbb{N}$ large enough, let $\rho(n) = 1/\log_2 n$, we define

$$\mathcal{A}_1 = \left\{\left|\frac{\mathcal{L}(m_n,n)}{n} - \frac{1}{E_{m_n}^\alpha(W_n)}\right| > \frac{\rho(n)}{E_{m_n}^\alpha(W_n)}\right\}, \tag{2.12}$$

$$\mathcal{A}_2 = \{T_{m_n} \leq n/(\log n)^4, \mathcal{L}(W_n,n) = n\}. \tag{2.13}$$

First we recall an elementary result originaly due to Sinai (1982):

Lemma 2.4. *Assume* (1.1), (1.2) *and* (1.3) *hold, there exists a constant* $b_1 > 0$ *such that for all* $\gamma > 2$, *there exists* n_0 *such that for all* $n > n_0$ *there exists* $G_n \subset \Omega_1$ *with* $Q[G_n] \geq 1 - \phi_1(n)$ *and*

$$\sup_{\alpha \in G_n} \{\mathbb{P}_0^\alpha[\mathcal{A}_2]\} \leq r_1(n), \tag{2.14}$$

where $r_1(n) = b_1(\log_2 n)^2/(\log n)^{\gamma-2}$.

Proof. One can find the proof of this result in Sinai (1982); see also Andreoletti (2006): Proposition 4.7 and Lemma 4.8. □

Lemma 2.5. *Assume* (1.1), (1.2) *and* (1.3) *hold, there exists a constant* $b_2 > 0$ *such that for all* $\gamma > 6$, *there exists* n_0 *such that for all* $n > n_0$ *there exists* $G_n \subset \Omega_1$ *with* $Q[G_n] \geq 1 - \phi_1(n)$ *and*

$$\sup_{\alpha \in G_n} \{\mathbb{P}_0^\alpha[\mathcal{A}_1]\} \leq r_2(n), \tag{2.15}$$

where $r_2(n) = b_2(\log_2 n)^6/((\rho(n))^2(\log n)^{\gamma-6})$.

Proof. A weaker version of this result is already present in Andreoletti (2006, Theorem 3.8), here we get a better rate of convergence for the probability and, for completeness, we also give a shorter proof. Let us denote

$$\mathcal{A}_1^+ = \left\{\frac{\mathcal{L}(m_n, n)}{n} - \frac{1}{E_{m_n}^\alpha(W_n)} > \frac{\rho(n)}{E_{m_n}^\alpha(W_n)}\right\},$$

$$\mathcal{A}_1^- = \left\{\frac{\mathcal{L}(m_n, n^-)}{n} - \frac{1}{E_{m_n}^\alpha(W_n)} < -\frac{\rho(n)}{E_{m_n}^\alpha(W_n)}\right\},$$

$$\mathcal{A}_2^+ = \{T_{m_n} \leq n/(\log n)^4\}, \qquad \mathcal{A}_2^- = \{\mathcal{L}(W_n, n) = n\},$$

where $n^- = n - n/(\log n)^4$. Thanks to Lemma 2.4 we have

$$\begin{aligned} \mathbb{P}_0^\alpha[\mathcal{A}_1] &\leq \mathbb{P}_0^\alpha[\mathcal{A}_1, \mathcal{A}_2] + r_1(n) \\ &\leq \mathbb{P}_{m_n}^\alpha[\mathcal{A}_1^+, \mathcal{A}_2^-] + \mathbb{P}_{m_n}^\alpha[\mathcal{A}_1^-, \mathcal{A}_2^-] + r_1(n). \end{aligned} \tag{2.16}$$

Let

$$T_{m_n,j} = \begin{cases} \inf\{k > T_{m_n,j-1}, X_k = m_n\}, & j \geq 2, \\ +\infty, & \text{if such } k \text{ does not exist,} \end{cases}$$

$$T_{m_n,1} = T_{m_n} \qquad \text{(see (1.6))},$$

we can check that

$$\mathbb{P}_{m_n}^\alpha[\mathcal{A}_1^+, \mathcal{A}_2^-] \leq \mathbb{P}_{m_n}^\alpha[\mathcal{L}(W_n, T_{m_n,n_1}) \leq n], \tag{2.17}$$

where $n_1 = \frac{n}{E^{\alpha}_{m_n}(W_n)}(1+\rho(n))$ (notice that n_1 is not necessarily an integer but for simplicity we disregard that). The strong Markov property implies that $\mathcal{L}(k, T_{m_n,n_1})$ is a sum of n_1 i.i.d. random variables, so by Chebyshev's inequality we get

$$\mathbb{P}^{\alpha}_{m_n}[\mathcal{L}(W_n, T_{m_n,n_1}) \le n] \le \frac{n_1 \operatorname{Var}(\mathcal{L}(W_n, T_{m_n}))}{n^2(\rho(n))^2} \le \frac{n_1 |W_n| \sum_{k\in W_n} \operatorname{Var}(\mathcal{L}(k, T_{m_n}))}{n^2(\rho(n))^2}.$$

Using (A.2) together with the fact that $\alpha \in G_n$ [Properties (2.4) and (2.5)] we finally obtain

$$\mathbb{P}^{\alpha}_{m_n}[\mathcal{L}(W_n, T_{m_n,n_1}) \le n] \le \frac{n_1 |W_n|^3 n}{n^2(\rho(n))^2(\log n)^{\gamma}} \le \frac{\operatorname{const}(\log_2 n)^6}{(\rho(n))^2(\log n)^{\gamma-6}}. \tag{2.18}$$

We get the same estimate for $\mathbb{P}^{\alpha}_{m_n}[\mathcal{A}_1, \mathcal{A}_2^{+}]$, then collecting (2.16), (2.17) and (2.18) we get the lemma. □

Lemma 2.5 is a key point to prove Proposition 2.3; note that we will also need the result (1.11) of Proposition 1.7 which proof is postponed Section 2.3.

2.2.2 *Proof of Proposition* 2.3. We split the proof into two steps, the first step makes the link between $\mathbb{L}^{\gamma}_n$ and particular points of the random environment contained in the set V^{γ}_n. The second step is to prove a law of large numbers for the local time.

Step 1. Let us define the following subsets of W_n:

$$\bar{\mathcal{V}}_n = \left\{ M'_n \le k \le m_n - 1, \left(\max_{k\le j\le m_n} S_j - S_{m_n}\right) < \log n - \frac{\gamma}{2}\log_2 n \right\},$$

$$\bar{\mathcal{V}}'_n = \left\{ m_n + 1 \le k \le M_n, \left(\max_{m_n\le j\le k} S_j - S_{m_n}\right) < \log n - \frac{\gamma}{2}\log_2 n \right\},$$

and

$$V^{\gamma}_n = \bar{\mathcal{V}}_n \cap \bar{\mathcal{V}}'_n. \tag{2.19}$$

In words V^{γ}_n is a subset of W_n, such that for all $k \in V^{\gamma}_n$ the largest difference of potential between m_n and k is smaller than $\log n - \gamma/2 \log_2 n$ (see also Figures 5 and 6). For the walk, we will see (lemma below) that if $k \in V^{\gamma}_n$ then the walk will hit k after it has reached m_n and it will hit this point k a large number of times.

First let us prove the following lemma:

Lemma 2.6. *Assume* (1.1), (1.2) *and* (1.3) *hold for all* $\gamma > 6$, *there exists* n_0 *such that for all* $n > n_0$ *there exists* $G_n \subset \Omega_1$ *with* $Q[G_n] \geq 1 - \phi_1(n)$ *and*

$$\sup_{\alpha \in G_n} \{\mathbb{P}_0^\alpha[\mathbb{L}_n^\gamma \subseteq V_n^\gamma]\} \geq 1 - \text{const} \cdot r_3(n) - \text{const} \cdot r_2(n), \tag{2.20}$$

where $r_3(n) = (\log_2 n)^2/(\log n)^{\gamma/2-2}$.

Notice that $\mathbb{L}_n^\gamma$ is a $\mathbb{P}$ random set (with two levels of randomness) whereas V_n^γ is only a Q random set (with one level of randomness), this lemma makes the link between a trajectory of the walk and the random environment.

Proof. First notice that

$$\mathbb{P}_0^\alpha[\mathbb{L}_n^\gamma \subseteq V_n^\gamma] = 1 - \mathbb{P}_0^\alpha\Bigg[\bigcup_{k \in (\mathcal{V}_n \cup \mathcal{V}_n')} \{k \in \mathbb{L}_n^\gamma\}\Bigg], \tag{2.21}$$

where

$$\mathcal{V}_n = \left\{M_n' \leq k \leq m_n - 1, \left(\max_{k \leq j \leq m_n} S_j - S_{m_n}\right) \geq \log n - \frac{\gamma}{2}\log_2 n\right\},$$
$$\mathcal{V}_n' = \left\{m_n + 1 \leq k \leq M_n, \left(\max_{m_n \leq j \leq k} S_j - S_{m_n}\right) \geq \log n - \frac{\gamma}{2}\log_2 n\right\}.$$

Let $k \in \mathcal{V}_n$ we get that

$$\begin{aligned}
&\mathbb{P}_0^\alpha[k \in \mathbb{L}_n^\gamma, |T_{k^*} - T_{m_n}| \leq (\log n)^3] \\
&\quad \leq \mathbb{P}_0^\alpha\Bigg[\sum_{j=T_{k^*}}^{n} \mathbb{1}_{X_j=k} \geq (\log n)^\gamma, |T_{k^*} - T_{m_n}| \leq (\log n)^3\Bigg] \\
&\quad \leq \mathbb{P}_0^\alpha\Bigg[\sum_{j=T_{m_n}}^{n} \mathbb{1}_{X_j=k} \geq (\log n)^\gamma - (\log n)^3\Bigg].
\end{aligned}$$

The strong Markov property together with the fact that $T_{m_n,n} \geq n$ yields

$$\begin{aligned}
&\mathbb{P}_0^\alpha\Bigg[\sum_{j=T_{m_n}}^{n} \mathbb{1}_{X_j=k} \geq (\log n)^\gamma - (\log n)^3\Bigg] \\
&\quad \leq \mathbb{P}_{m_n}^\alpha\Bigg[\sum_{j=1}^{T_{m_n,n}} \mathbb{1}_{X_j=k} \geq (\log n)^\gamma - (\log n)^3\Bigg].
\end{aligned}$$

By using the Markov inequality and (A.1) we finally get

$$\begin{aligned}\mathbb{P}_0^\alpha[k \in \mathbb{L}_n^\gamma, |T_{k^*} - T_{m_n}| \le (\log n)^3] &\le \frac{n\mathbb{E}_{m_n}^\alpha[\mathcal{L}(k, T_{m_n})]}{(\log n)^\gamma - (\log n)^3} \\ &\le \frac{n}{\eta_0 \exp(S_k - S_{m_n})((\log n)^\gamma - (\log n)^3)} \\ &\le \frac{1}{\eta_0 (\log n)^{\gamma/2}(1 - (\log n)^3/(\log n)^\gamma)},\end{aligned}$$

notice that in the last inequality we have used the fact that $k \in \mathcal{V}_n$. A similar computation gives the same inequality when $k \in \mathcal{V}_n'$. We can conclude as follow, thanks to (1.11)

$$\begin{aligned}&\mathbb{P}_0^\alpha\Big[\bigcup_{k\in(\mathcal{V}_n\cup\mathcal{V}_n')}\{k \in \mathbb{L}_n^\gamma\}\Big] \\ &\quad\le |\mathcal{V}_n \cup \mathcal{V}_n'| \max_{k\in\mathcal{V}_n\cup\mathcal{V}_n'} \mathbb{P}_0^\alpha[k \in \mathbb{L}_n^\gamma, |T_{k^*} - T_{m_n}| \le (\log n)^3] \\ &\qquad + \text{const}\frac{(\log_2 n)^8}{(\log n)^{\gamma-6}} \\ &\quad\le \frac{|\mathcal{V}_n \cup \mathcal{V}_n'|}{\eta_0 (\log n)^{\gamma/2}(1 - (\log n)^3/(\log n)^\gamma)} + \text{const}\frac{(\log_2 n)^8}{(\log n)^{\gamma-6}}.\end{aligned} \tag{2.22}$$

Collecting (2.21), (2.22) and Property (2.4) yields the lemma. □

Step 2. This second step is devoted to the proof of the following lemma. Recall that $\mathcal{A}_1$ (resp. $\mathcal{A}_2$) is defined in (2.12) (resp. (2.13)).

Lemma 2.7. *For all α and n we have*

$$\mathbb{P}_0^\alpha\left[\left|\frac{\mathcal{L}(k, n)}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)}\right| > w_{k,n}, \mathcal{A}_1, \mathcal{A}_2\right] \le 2\exp\big(-n/4\psi_2^\alpha(n)\big) \tag{2.23}$$

recall that $w_{k,n} = \rho_1(n)E_{m_n}^\alpha(k)/E_{m_n}^\alpha(W_n)$ and $\psi_2^\alpha(n) = 2\frac{(\rho_1(n)-\rho(n))^2}{1+\rho(n)}\frac{(\alpha_{m_n}\wedge\beta_{m_n})}{|k-m_n|} \times \frac{\exp(-(S_{M_k}-S_{m_n}))}{E_{m_n}^\alpha(W_n)}$. M_k is such that $S_{M_k} = \max_{m_n+1\le j\le k} S_j$ if $k > m_n$ and conversely if $k < m_n$ $S_{M_k} = \max_{k\le j\le m_n-1} S_j$. Also $\rho_1(n)$ is defined just above Proposition 2.3.

Proof. We essentially use a concentration inequality for sum of i.i.d. random variables, for simplicity we only give the proof for $k > m_n$, the other case ($k \le m_n$) is very similar. Using the strong Markov property and the fact that $\mathcal{L}(k, T_{m_n}) = 0$, we get

$$\mathbb{P}_0^\alpha\Big[\Big|\frac{\mathcal{L}(k,n)}{n}-\frac{E^\alpha_{m_n}(k)}{E^\alpha_{m_n}(W_n)}\Big|>w_{k,n},\mathcal{A}_1,\mathcal{A}_2\Big]$$
$$\le \mathbb{P}^\alpha_{m_n}\Big[\frac{\mathcal{L}(k,n)}{n}-\frac{E^\alpha_{m_n}(k)}{E^\alpha_{m_n}(W_n)}>w_{k,n},\mathcal{A}_1\Big]$$
$$+\mathbb{P}^\alpha_{m_n}\Big[\frac{\mathcal{L}(k,n^-)}{n}-\frac{E^\alpha_{m_n}(k)}{E^\alpha_{m_n}(W_n)}<-w_{k,n},\mathcal{A}_1\Big]$$

and n^- is defined just above (2.16). We have

$$\mathbb{P}^\alpha_{m_n}\Big[\frac{\mathcal{L}(k,n)}{n}-\frac{E^\alpha_{m_n}(k)}{E^\alpha_{m_n}(W_n)}>w_{k,n},\mathcal{A}_1\Big]$$
$$\le \mathbb{P}^\alpha_{m_n}\Big[\frac{\mathcal{L}(k,n)}{n}-\frac{E^\alpha_{m_n}(k)}{E^\alpha_{m_n}(W_n)}>w_{k,n},\frac{\mathcal{L}(m_n,n)}{n}-\frac{1}{E^\alpha_{m_n}(W_n)}\le\frac{\rho(n)}{E^\alpha_{m_n}(W_n)}\Big]$$
$$\le \mathbb{P}^\alpha_{m_n}\Big[\frac{\mathcal{L}(k,T_{m_n,n_1})}{n}-\frac{E^\alpha_{m_n}(k)}{E^\alpha_{m_n}(W_n)}>w_{k,n}\Big]$$
$$= \mathbb{P}^\alpha_{m_n}\Big[\frac{\mathcal{L}(k,T_{m_n,n_1})}{n}-\frac{E^\alpha_{m_n}(k)}{E^\alpha_{m_n}(W_n)}(1+\rho(n))>w'_{k,n}\Big]$$

where n_1 is defined just below (2.17), and $w'_{k,n}=\frac{E^\alpha_{m_n}(k)}{E^\alpha_{m_n}(W_n)}(\rho_1(n)-\rho(n))$. The concentration inequality (see equation 6.12, page 164 of Ledoux and Talagrand (1991)) gives for n large enough

$$\mathbb{P}^\alpha_{m_n}\Big[\frac{\mathcal{L}(k,T_{m_n,n_1})}{n}-\frac{E^\alpha_{m_n}(k)}{E^\alpha_{m_n}(W_n)}>w'_{k,n},\mathcal{A}_1\Big]$$
$$\le \exp\Big[-\frac{n}{4}\frac{E^\alpha_{m_n}(W_n)}{\mathrm{Var}_{m_n}(\mathcal{L}(k,T_{m_n}))}\frac{(w'_{k,n})^2}{1+\rho(n)}\Big].$$

With the same method we get the same estimation for $\mathbb{P}^\alpha_{m_n}[\frac{\mathcal{L}(k,n^-)}{n}-\frac{E^\alpha_{m_n}(k)}{E^\alpha_{m_n}(W_n)}<-w_{k,n},\mathcal{A}_1]$. Using (A.2) we get Lemma 2.7. □

End of the proof of the proposition.
Using Lemmata 2.4, 2.5 and 2.6 we have

$$\mathbb{P}_0^\alpha\Big[\bigcup_{k\in\mathbb{L}_n^a}\Big\{\Big|\frac{\mathcal{L}(k,n)}{n}-\frac{E^\alpha_{m_n}(k)}{E^\alpha_{m_n}(W_n)}\Big|>w_{k,n}\Big\}\Big]$$
$$\le |V_n^\gamma|\sup_{k\in V_n^\gamma}\mathbb{P}_0^\alpha\Big[\Big|\frac{\mathcal{L}(k,n)}{n}-\frac{E^\alpha_{m_n}(k)}{E^\alpha_{m_n}(W_n)}\Big|>w_{k,n},\mathcal{A}_1,\mathcal{A}_2\Big]$$
$$+\mathrm{const}\cdot r_1(n)+\mathrm{const}\cdot r_2(n).$$

Then using Lemma 2.7 we get

$$\sup_{k\in V_n^\gamma} \mathbb{P}_0^\alpha\left[\left|\frac{\mathcal{L}(k,n)}{n}-\frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)}\right|>w_{k,n},\mathcal{A}_1,\mathcal{A}_2\right]$$
$$\le 2\sup_{k\in V_n^\gamma}\exp\bigl(-n/2\psi_2^\alpha(k,n)\bigr)\le 2\exp\bigl(-(\log n)^{\gamma/2-2}/(\rho_1(n)\log_2 n)^2\bigr),$$

where the last inequality comes from the definition of V_n^γ (see (2.19)), properties (2.4) and (2.6) of the random environment. To finish we use again the property (2.4) together with the fact that $V_n^\gamma\subset W_n$. Notice that, for n large enough, the contribution of the above probability is negligible comparing to $r_2(n)$ and $r_1(n)$ so we get our result.

2.3 Proof of Propositions 1.7 and 1.8

Proof of Proposition 1.7. Notice that (1.10) is a slight improvement of Corollary 3.17 of Andreoletti (2006), for completeness we give a short proof based on Lemma 2.5. Let $K(n)=[m_n-(\log_2 n)^2, m_n+(\log_2 n)^2]$ and $\bar{K}(n)$ the complementary of $K(n)$ in W_n. We prove that the local time on all the points belonging to $\bar{K}(n)$ is smaller than $\mathcal{L}(m_n,n)$. By using Lemma 2.4 and 2.5 we get

$$\mathbb{P}_0^\alpha\Bigl[\max_{x\in\mathbb{F}_n}|m_n-x|\le(\log_2 n)^2\Bigr]$$
$$\le \mathbb{P}_0^\alpha\Bigl[\bigcup_{l\in\bar{K}(n)}\{\mathcal{L}(l,n)\ge\mathcal{L}(m_n,n)\},\mathcal{A}_1\Bigr]+r_1(n)$$
$$\le \mathbb{P}_0^\alpha\Bigl[\bigcup_{l\in\bar{K}(n)}\{\mathcal{L}(l,T_{m_n})\ge n_2\},\mathcal{A}_1\Bigr]$$
$$+\mathbb{P}_{m_n}^\alpha\Bigl[\bigcup_{l\in\bar{K}(n)}\{\mathcal{L}(l,T_{m_n,n_1})\ge n_2\}\Bigr]+r_1(n)+r_2(n),$$

where $n_2=\frac{n^-}{E_{m_n}^\alpha(W_n)}(1-\rho(n))$, (recall that n^- is defined above (2.16)). Notice that $\mathbb{P}_0^\alpha[\bigcup_{l\in\bar{K}(n)}\{\mathcal{L}(l,T_{m_n})\ge n_2\},\mathcal{A}_1]\le\mathbb{P}_0^\alpha[T_{m_n}\ge n_2]\le r_1(n)$ thanks to property (2.6) and Lemma 2.4. Also we have

$$\mathbb{P}_{m_n}^\alpha\Bigl[\bigcup_{l\in\bar{K}(n)}\{\mathcal{L}(l,T_{m_n,n_1})\ge n_2\}\Bigr]\le\sum_{l\in\bar{K}(n)}\mathbb{P}_{m_n}^\alpha[\mathcal{L}(l,T_{m_n,n_1})\ge n_2]$$
$$\le\sum_{l\in\bar{K}(n)}\frac{n_1\operatorname{Var}(\mathcal{L}(l,T_{m_n}))}{(n_2-n_1\mathbb{E}_{m_n}^\alpha(\mathcal{L}(l,T_{m_n})))^2} \tag{2.24}$$
$$\le\frac{\text{const}(\log_2 n)^2(\log n)}{n},$$

so we get (1.10).

To get (1.11) we can use (1.10). Indeed, further on, we show that within an interval of time of length $(\log n)^3$ all the points in $K(n)$ are at least visited once, so as m_n and k^* (thanks to (1.10)) belongs to $K(n)$ we get (1.11). We have

$$\begin{aligned}
&\mathbb{P}_0^\alpha[|T_{m_n} - T_{k^*}| > (\log n)^3] \\
&\quad \le \mathbb{P}_{m_n}^\alpha\Big[\bigcup_{l \in K(n)} \{\mathcal{L}(l, (\log n)^3) < 1\}\Big] + \text{const} \cdot r_2(n) \\
&\quad \le \sum_{l \in K(n)} \mathbb{P}_{m_n}^\alpha[\{\mathcal{L}(l, (\log n)^3) < 1\}] + \text{const} \cdot r_2(n) \\
&\quad \le \exp(-\text{const} \log n) + \text{const} \cdot r_2(n),
\end{aligned}$$

and for the last inequality we have used Lemma 2.7. □

Proof of Proposition 1.8. The two properties can be deduced from the following inequality, let $\varepsilon > 1$, for all n large enough and all $\alpha \in G_n$:

$$\mathbb{P}_0^\alpha\big[V_n^{2(\gamma+\varepsilon)} \subseteq \mathbb{L}_n^\gamma\big] \ge 1 - \text{const} \cdot \big(r_2(n) + r_3(n)\big), \tag{2.25}$$

where the definition of $V_n^{2(\gamma+\varepsilon)}$ is the same as V_n^γ, replacing γ by $2(\gamma+\varepsilon)$. Indeed, thanks to (2.25) we have

$$\begin{aligned}
&\mathbb{P}_0^\alpha\big[\mathcal{L}(\mathbb{L}_n^\gamma, n) \ge n\big(1 - o(1)\big)\big] \\
&\quad \ge \mathbb{P}_0^\alpha\big[\mathcal{L}\big(V_n^{2(\gamma+\varepsilon)}, n\big) \ge n\big(1 - o(1)\big)\big] - \text{const} \cdot \big(r_2(n) + r_3(n)\big) \\
&\quad \ge \mathbb{P}_0^\alpha\big[\mathcal{L}\big(V_n^{2(\gamma+\varepsilon)}, n\big) \ge n\big(1 - o(1)\big), \mathcal{A}_1, \mathcal{A}_2\big] \\
&\qquad - \text{const} \cdot \big(r_2(n) + r_3(n)\big) \\
&\quad \ge 1 - \mathbb{P}_{m_n}^\alpha\big[\mathcal{L}\big(\bar{V}_n^{2(\gamma+\varepsilon)}, T_{m_n,n_1}\big) \ge n/(\log n)\big] - \mathbb{P}_0^\alpha[\mathcal{A}_1, \mathcal{A}_2] \\
&\qquad - \text{const} \cdot \big(r_2(n) + r_3(n)\big),
\end{aligned} \tag{2.26}$$

where $\bar{V}_n^{2(\gamma+\varepsilon)}$ is the complementary of $V_n^{2(\gamma+\varepsilon)}$ in W_n. By Markov inequality

$$\begin{aligned}
&\mathbb{P}_{m_n}^\alpha\big[\mathcal{L}\big(\bar{V}_n^{2(\gamma+\varepsilon)}, T_{m_n,n_1}\big) \ge n/(\log n)\big] \\
&\quad \le \frac{\log n}{n} n_1 \big|\bar{V}_n^{2(\gamma+\varepsilon)}\big| \max_{k \in \bar{V}_n^{2(\gamma+\varepsilon)}} \mathbb{E}_{m_n}^\alpha[\mathcal{L}(k, T_{m_n})],
\end{aligned}$$

then using (A.1) and the definition of $\bar{V}_n^{2(\gamma+\varepsilon)}$ we get

$$\begin{aligned}
&\mathbb{P}_{m_n}^\alpha\big[\mathcal{L}\big(\bar{V}_n^{2(\gamma+\varepsilon)}, T_{m_n,n_1}\big) \ge n/(\log n)\big] \\
&\quad \le \frac{\text{const} \cdot n_1 |\bar{V}_n^{2(\gamma+\varepsilon)}| (\log n)^{1+(\gamma+\varepsilon)}}{n^2}.
\end{aligned} \tag{2.27}$$

Collecting (2.26), (2.27) and using property (2.4) we get (1.12).

To get (1.13), first we notice that thanks to (2.15) and (2.25) we have

$$\mathbb{P}_0^{\alpha}\left[V_n^{2(\gamma+\varepsilon)} \subset \mathbb{L}_n^{\gamma} \subset W_n\right] \geq 1 - \text{const} \cdot \big(r_2(n) + r_3(n)\big), \tag{2.28}$$

so we only need to show that $|V_n^{\gamma+\varepsilon}| \geq (\log n)^2/(\log_2 n)^2$ and $|W_n| \leq (\log n)^2 \times (\log_2 n)^2$ with a high probability but this fact is already included in property (2.4).

We are left to prove (2.25), by using the same method to that of the proof of Theorem 1.6 we can get

$$\mathbb{P}_0^{\alpha}\left[V_n^{2(\gamma+\varepsilon)} \not\subseteq \mathbb{L}_n^{\gamma}\right] \leq |V_n^{\gamma+\varepsilon}| \max_{k \in V_n^{2(\gamma+\varepsilon)}} \mathbb{P}_{m_n}^{\alpha}\left[\sum_{j=1}^{n_1} \eta_j^k < (\log n)^{\gamma}\right] + \text{const} \cdot r_2(n) + \text{const} \cdot r_3(n)$$

where (η_j^k, j) is a i.i.d. sequence with the law of $\mathcal{L}(k, T_{m_n})$. Then using again a concentration inequality we get (2.25). □

3 Algorithm and numerical simulations

3.1 Main steps of the algorithm

First notice that we have no criteria to determine whether or not we can apply this method to an unknown series of data. All we know is that it works for Sinai's walk, however we can apply the following algorithm to every processes. Let us recall the basic random variables that will be used for our simulations, let $x \in \mathbb{Z}$, $n \in \mathbb{N}$,

$$T_x = \begin{cases} \inf\{k \in \mathbb{N}^*, X_k = x\}, \\ +\infty, \qquad \text{if such } k \text{ does not exist,} \end{cases}$$

$$\mathcal{L}(x, n) \equiv \sum_{i=1}^{n} \mathbb{1}_{\{X_i = x\}};$$

$$\mathcal{L}^*(n) = \max_{k \in \mathbb{Z}}(\mathcal{L}(k, n)), \qquad \mathbb{F}_n = \{k \in \mathbb{Z}, \mathcal{L}(k, n) = \mathcal{L}^*(n)\},$$

$$k^* = \inf\{|k|, k \in \mathbb{F}_n\};$$

$$\mathbb{L}_n^{\gamma} = \left\{k \in \mathbb{Z}, \sum_{j=T_{k^*}}^{n} \mathbb{1}_{X_j = k} \geq (\log n)^{\gamma}\right\}, \qquad \gamma > 6,$$

$$S_{k,m_n}^n = 1 - \frac{1}{\log n}(S_k - S_{m_n}), \qquad k \in \mathbb{L}_n^{\gamma},$$

$$\hat{S}_k^n = \frac{\log(\mathcal{L}(k, n))}{\log n}.$$

Notice also that, thanks to Proposition 1.7, in probability we have $|m_n - k^*| \leq \text{const}(\log_2 n)^2$. The algorithm is the following:

Step 1: We have to determine $\mathbb{L}_n^\gamma$ and to get it we have to compute T_{k^*} and therefore the local time of the process. First we compute $\mathcal{L}(k,n)$ for every k, notice that $\mathcal{L}(k,n)$ is not equal to zero only if k has been visited by the walk within the interval of time $[1,n]$. Then we can compute $\mathcal{L}^*(n)$ and determine k^* and T_{k^*}. Notice that T_{k^*} is not a stopping time, therefore we need to run the algorithm two times to compute what we need. We are now able to determine $\mathbb{L}_n^\gamma$ computing $\sum_{j=T_{k^*}}^n \mathbb{1}_{X_j=k}$.

Step 2: We can check that $\mathbb{L}_n^\gamma$ is connected, contains k^* and that its size is of the order of a typical fluctuation of the walk. Now, keeping only the k that belongs to $\mathbb{L}_n^\gamma$ we compute for those k: $\hat{S}_k^n = \frac{\log(\mathcal{L}(k,n))}{\log n}$ the estimator of the potential. We localize the bottom of the valley m_n using k^*.

3.2 Simulations

For the first simulation (Figure 7) we show a case where $\mathbb{L}_n^\gamma$ is large, that is, $\mathbb{L}_n^\gamma$ contains most of the points visited by the walk. The trajectory of the random potential is in light grey, the interval of confidence in black and grey. We took $n = 500{,}000$ and $\gamma = 7$, notice that the larger is γ, the smaller is $\mathbb{L}_n^\gamma$ but better is the rate of convergence of the probability. We get that $\mathbb{L}_n^\gamma = [10, 94]$. In Figure 8 we plot the difference $S_{x,m_n}^n - \hat{S}_x^n$ and its linear regression. We notice that the slope of the linear regression is of order 10^{-5}.

Now let us choose another example where $\mathbb{L}_n^\gamma$ is smaller. For the following simulation (Figure 9) we have only changed the sequence of random numbers. We get that $\mathbb{L}_n^\gamma = [-150, -85]$. We notice that for the coordinates larger than -85 and especially after -40, our estimator is not good at all. In fact, once the walk has reached the minimum of the valley (coordinate -111) it will never reach again one of the points of coordinate larger than -40 before $n = 500{,}000$, so our estimator cannot say anything precise about the difference $(S_{x,m_n}^n - \hat{S}_x^n, k \geq -40)$. However,

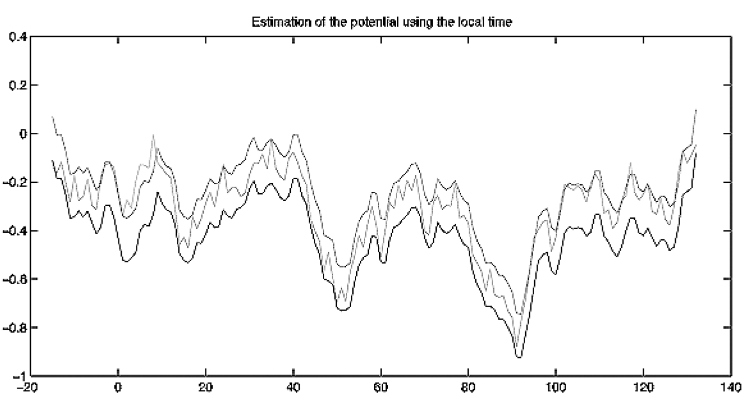

Figure 7 *In light grey S_{x,m_n}^n, in black $\hat{S}_x^n - u_n$, in grey $\hat{S}_x^n + u_n$.*

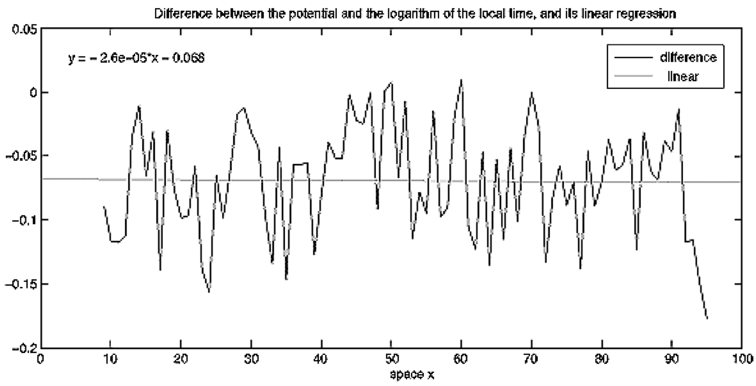

Figure 8 *In black* $S^n_{x,m_n} - \hat{S}^n_x$, *in grey the linear regression.*

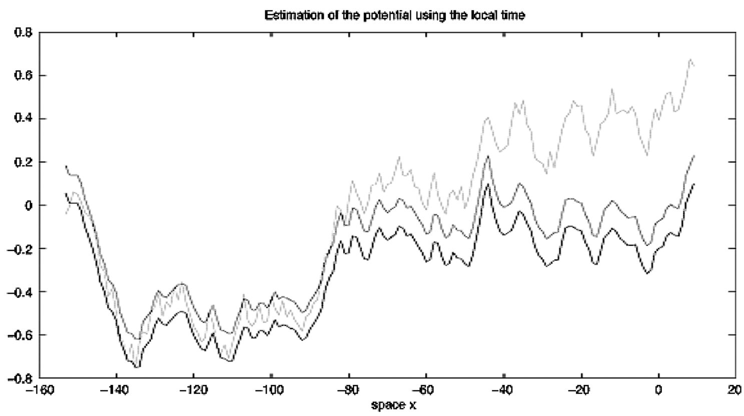

Figure 9 *In light grey* S^n_{x,m_n}, *in black* $\hat{S}^n_x - u_n$, *in grey* $\hat{S}^n_x + u_n$.

if we look in the past of the walk and especially before time T_{k^*}, then we may be able to get information about $(S^n_{x,m_n} - \hat{S}^n_x, k \geq -40)$. We can expect that the favorite point at that time is localized around the point -2, so a good estimator between the coordinate -40 and 10 may be given by $(\frac{\log(\mathcal{L}(k,T^*))}{\log T^*}, k)$.

The difference $S^n_{x,m_n} - \hat{S}^n_x$ and the linear regression in the interval $\mathbb{L}^\gamma_n = [-150, -85]$ is presented on Figure 10.

Appendix: Basic results for birth and death processes

For completeness we recall an explicit expression for the mean and an upper bound for the variance of the local times at a certain stopping time, we can find a proof of these elementary facts in Révész (1990, page 279).

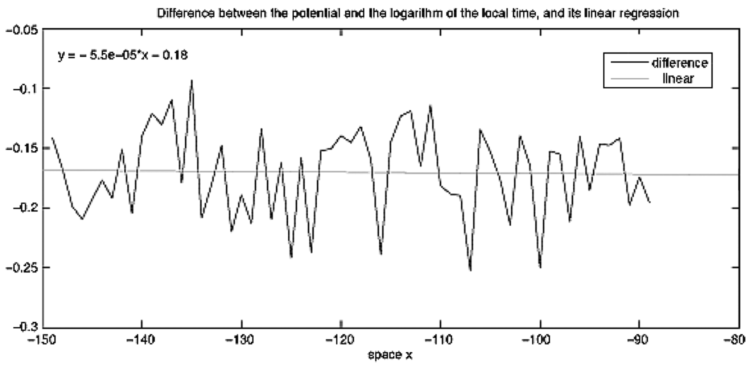

Figure 10 *In black* $S^n_{x,m_n} - \hat{S}^n_x$, *in grey the linear regression.*

Lemma A.1. *For all* α, *let* $k > m_n$

$$\mathbb{E}^{\alpha}_{m_n}[\mathcal{L}(k, T_{m_n})] = \frac{\alpha_{m_n}}{\beta_k} \frac{1}{e^{S_k - S_{m_n}}} a_{k,m_n}, \tag{A.1}$$

where

$$a_{k,m_n} = \frac{\sum_{i=m_n+1}^{k-1} e^{S_i} + e^{S_k}}{\sum_{i=m_n+1}^{k-1} e^{S_i} + e^{S_{m_n}}},$$

and

$$\mathrm{Var}_{m_n}[\mathcal{L}(k, T_{m_n})] \le 2(\mathbb{E}^{\alpha}_{m_n}[\mathcal{L}(k, T_{m_n})])^2 \frac{e^{S_{M_k} - S_{m_n}}}{\beta_k} |k - m_n|, \tag{A.2}$$

where M_k *is such that* $S_{M_k} = \max_{m_n+1 \le j \le k-1} S_j$. *For Q-a.a. environment* α

$$\frac{\eta_0}{1-\eta_0} \le \frac{\alpha_{m_n}}{\beta_k} a_{k,m_n} \le \frac{1}{\eta_0}. \tag{A.3}$$

A similar result is true for $k < m_n$ *and* $\mathbb{E}^{\alpha}_{m_n}[\mathcal{L}(m_n, T_{m_n})] = 1$.

Acknowledgments

I would like to thank Nicolas Klutchnikoff for stimulating discussions, and the anonymous referee for helpful comments on the paper.

References

Adelman, O. and Enriquez, N. (2004). Random walks in random environment: What a single trajectory tells. *Israel J. Math.* **142**, 205–220. MR2085716

Andreoletti, P. (2006). On the concentration of Sinai's walk. *Stochastic Process. Appl.* **116**, 1377–1408. MR2260740

Baldazzi, V., Cocco, S., Marinari, E. and Monasson, R. (2006). Infering DNA sequences from mechanical unzipping: An ideal-case study. *Physical Review Letters E* **96**, 128102-1-4.

Kesten, H. (1986). The limit distribution of Sinai's random walk in random environment. *Physica* **138A**, 299–309. MR0865247

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces*. Springer-Verlag, Berlin. MR1102015

Révész, P. (1990). *Random Walk in Random and Non-random Environments*. World Scientific, Singapore.

Sinai, Y. G. (1982). The limit behaviour of a one-dimensional random walk in a random medium. *Theory Probab. Appl.* **27**, 256–268. MR0657919

Solomon, F. (1975). Random walks in random environment. *Ann. Probab.* **3**, 1–31. MR0362503

Zeitouni, O. (2001). Lectures notes on random walks in random environment. St. Flour Summer School. Available at http://www.ee.technion.ac.il/~zeitouni/ps/notes1.ps.

Laboratoire MAPMO – C.N.R.S. UMR 6628 – Fédération Denis-Poisson
Université d'Orléans, UFR Sciences
Bâtiment de mathématiques – Route de Chartres
B.P. 6759 – 45067 Orléans cedex 2
France
E-mail: pierre.andreoletti@univ-orleans.fr