Random Walks in Random Environments

Erwin Bolthausen IHP, Paris, 2008

Contents

1 Introduction

We consider Markov chains on the lattice $\mathbb{Z}^d$ or on subset, which have site dependent transition probabilities. We denote by $\mathcal{P}$ a set of probability distributions on $\mathbb{Z}^d$. $\mathcal{P}$ will always contain only probabilities on a fixed finite subset of $\mathbb{Z}^d$, for instance $\left\{e \in \mathbb{Z}^d : |e| = 1\right\}$, but we will also consider other situations. A *field* of transition probabilities is described by an element $\omega \in \Omega \stackrel{\text{def}}{=} \mathcal{P}^{\mathbb{Z}^d}$, $\omega = (\omega_x)_{x \in \mathbb{Z}^d}$. $\mathcal{F}$ is the appropriate product σ-field on Ω. The transition probabilities of our "random walk in random environment" (RWRE for short) are then given by

$$p_\omega(x, x+y) \stackrel{\text{def}}{=} \omega_x(y). \tag{1.1}$$

We write $P_{x,\omega}$ for the law of a Markov chain with these transition probabilities, starting in x. The chain itself is denoted by $X_0 = x, X_1, X_2, \ldots$. We write Γ for the set of paths in $\mathbb{Z}^d$, equipped with the appropriate σ-field $\mathcal{G}$. It is evident that P defines a Markovian kernel from $\mathbb{Z}^d \times \Omega$ to Γ, i.e. for any $G \in \mathcal{G}$, the mapping $(x, \omega) \rightarrowtail P_{x,\omega}(G)$ is a measurable mapping.

We will choose ω randomly. This means that we fix a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$. A special case is when we take the ω_x as i.i.d. random variables, i.e. when $\mathbb{P}$ is a product measure $\mu^{\mathbb{Z}^d}$, μ being a probability distribution on $\mathcal{P}$.

The semi-direct product on $\Omega \times \Gamma$ is denoted by $\hat{P}_x$, i.e.

$$\hat{P}_x (A \times B) \stackrel{\text{def}}{=} \int_A P_{x,\omega} (B) \, \mathbb{P} (d\omega) .$$

It has become common to call this the *annealed law*, although it is kind of a misnomer. In contrast, $P_{x,\omega}$ is called the *quenched law*. If $x = 0$, we usually leave out x in the notation. One should note that the sequence (X_n) is not a Markov chain under the law $\hat{P}$.

An annealed property of the (X_n) is a property this sequence has under the law $\hat{P}$. In contrast, one speaks of a quenched property if for $\mathbb{P}$-a.a. ω, the property holds under $P_{x,\omega}$. For a law of large numbers, there is no difference: If $X_n/n \to v \in \mathbb{R}^d$ holds $\hat{P}$-a.s. then this holds $P_{0,\omega}$-a.s. for $\mathbb{P}$-a.a. ω. However, for convergence in law, e.g. for a CLT, there is a big difference.

Example 1.1
In the one-dimensional case, of course $\omega_x (-1) = 1 - \omega_x (1)$, $x \in \mathbb{Z}$, *and therefore, the random environment is a sequence of* $(0,1)$*-valued random variables. We will often write* p_x *instead of* $\omega_x (1)$. p_x *is the probability to move from* x *to the right, and* $q_x \stackrel{\text{def}}{=} 1 - p_x$ *is the probability to move to the left.*

We also write $\mathbf{p} = (p_x)_{x \in \mathbb{Z}}$ *for the whole sequence. In this case we simply take* Ω *to be* $(0,1)^{\mathbb{Z}}$. *A minimal requirement will always be that the sequence is stationary and ergodic under* $\mathbb{P}$: *If* θ *is the usual shift operator on* Ω, $\theta (\mathbf{p})_x = p_{x+1}$, *then* θ *is assumed to be measure preserving and ergodic. As remarked above, we always assume that* $0 < p_x < 1$.

Example 1.2
Although I essentially concentrate on the above model where the transitions are given by (1.1), one should mention that there is another one, which is simpler in many respects due to irreversibility. This is the model of random currents. One attaches the randomness to the bonds, in the form of a random strengths. Denote by $\mathbb{B}^d$, *the set of nearest neighbor bonds (undirected) in* $\mathbb{Z}^d$. *Consider furthermore a law* μ *on the positive real line, and then the product measure on* $(\mathbb{R}^+)^{\mathbb{B}^d}$. *This measure is again denoted by* $\mathbb{P}$. *For* $\omega \in (\mathbb{R}^+)^{\mathbb{B}^d}$, $b \in \mathbb{B}^d$, *the weights are* $\xi_b (\omega) \stackrel{\text{def}}{=} \omega_b$. *The transition probabilities of the RWRE are defined in the following way. If* $x \in \mathbb{Z}^d$, *denote by* n_x *the set of bonds one of the endpoints equal to* x. *If* y *is a nearest neighbor of* x, *then*

$$p_\omega (x,y) \stackrel{\text{def}}{=} \frac{\xi_{\{x,y\}} (\omega)}{\sum_{b \in n_x} \xi_b (\omega)}.$$

In this case, $p_\omega (x, \cdot)$ *and* $p_\omega (x', \cdot)$ *are not independent if* $|x - x'| = 1$. *The main advantage is that the above transition probabilities satisfy the detailed balance equation*

$$\nu_\omega (x) \, p_\omega (x,y) = \nu_\omega (y) \, p_\omega (y,x) , \ \forall x, y$$

where

$$\nu_\omega(x) \stackrel{\text{def}}{=} \sum_{b \in n_x} \xi_b(\omega).$$

The Example 1.1 does satisfy the detailed balance equation, too, but this is due to the special nearest neighbor one-dimensional situation. (Every nearest neighbor Markov chain on a graph without loops satisfies the detailed balance equation). Our models (1.1) don't satisfy a detailed balance equation in higher dimensions, and also not in one dimension if one steps away from the nearest neighbor situation. We will come to these issues later.

Natural questions to address are laws of large numbers, i.e. does S_n/n converge $\hat{P}$-a.s., to a constant $v \in \mathbb{R}^d$, and is there a central limit theorem, i.e. is

$$\frac{X_n - nv}{\sqrt{n}}$$

asymptotically Gaussian distributed. As remarked above, one has clearly to distinguish between an annealed and a quenched CLT. The former means that the above expression is asymptotically normal under $\hat{P}$, but a quenched CLT would mean that for $\mathbb{P}$-a.a. ω, there is a CLT under the measure $P_{0,\omega}$. The latter is rather more difficult to establish. There had been a lot of progress on this question recently, see [1], [12], and the older paper [4].

For the LLN, a naive guess might be that

$$\lim_{n\to\infty} \frac{X_n}{n} = \sum_e e\mathbb{E}(\omega_e),$$

but a moments reflection reveals that things cannot be as simple. The point is that the environment acts on the path in a "non-linear" way, so it is clear that

$$\hat{E}(X_n) \neq n \sum_e e\mathbb{E}(\omega_e),$$

in general, except for $n = 1, 2$.

2 The quasi one-dimensional case

2.1 Transience and recurrence

In the strictly one-dimensional nearest neighbor case of Example 1.1, a classical result by Solomon is the following:

Theorem 2.1
Assume that the sequence $\{p_x\}_{x\in\mathbb{Z}}$ is i.i.d. with $\mathbb{E}(-\log p_x)$, $\mathbb{E}(-\log q_x) < \infty$, and write

$$\lambda^+ \stackrel{\text{def}}{=} \mathbb{E}\log\frac{q_x}{p_x}.$$

Then

a) If $\lambda^+ > 0$ *then* $\lim_{t\to\infty} X_t = -\infty$, $\hat{P}_0$ *almost surely.*

b) If $\lambda^+ < 0$ *then* $\lim_{t\to\infty} X_t = \infty$, $\hat{P}_0$ *almost surely.*

c) If $\lambda^+ = 0$ *then* $\limsup_{t\to\infty} X_t = \infty$, *and* $\liminf_{t\to\infty} X_t = -\infty$, $\hat{P}_0$ *almost surely.*

As a consequence, one sees that $\lambda^+ = 0$ holds if and only if the RWRE is recurrent.

The RWRE on $\mathbb{Z}$ with finite range jumps is already considerably more delicate. In the independent case, the ω_x are i.i.d. random variables taking values in the set of probability measures on $\{-R, \ldots, R\}$, R some natural number. $\omega_x(y)$ is then the probability with which the RWRE (under the quenched law) jumps from x to $x+y$.

A quantity of crucial importance is the (random) function $h_{a,b}(x)$, $a < b$, $x \in \mathbb{Z}$,

$$h_{a,b}(x) \stackrel{\text{def}}{=} P_{x,\omega}\left(T_{[b,\infty)} < T_{(-\infty,a]}\right), \tag{2.1}$$

where T_A denotes the first entrance time into the set A. Clearly this function satisfies for $a < x < b$

$$h_{a,b}(x) = \sum_{|y|\le R} \omega_x(y)\, h_{a,b}(x+y),$$

and boundary conditions $h = 1$ on $[b, \infty)$, $h = 0$ on $(-\infty, a]$.

In the nearest neighbor case, one can explicitly solve this equation which is not possible in the non-nearest-neighbor case. A natural approach is to represent a vector of $h_{a,b}$-values of length $2R$ by a shifted one through

$$\begin{pmatrix} h(x+R) \\ \vdots \\ \vdots \\ \vdots \\ h(x-R+1) \end{pmatrix} = \begin{pmatrix} -\frac{\omega_x(R-1)}{\omega_x(R)} & \ldots, & \frac{1-\omega_x(0)}{\omega_x(R)} & \ldots, & -\frac{\omega_x(-R)}{\omega_x(R)} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ 0 & & & 1 & 0 \end{pmatrix} \begin{pmatrix} h(x+R-1) \\ \vdots \\ \vdots \\ \vdots \\ h(x-R) \end{pmatrix},$$

and match the boundary conditions. This is the approach of Key [**?**] who generalized the Solomon result to the non-nearest neighbor case. The recurrence and transience are then expressed in terms of the middle Lyapunov-exponent of the products of the above random matrices. The approach has recently been taken up by Julien Brémont [**?**].

In two papers with Ilya Goldsheid [2][3], we developed a somewhat different approach in a slightly more general setup.

We consider a RWRE in a strip $\mathbb{Z}\times\{1, \ldots, m\}$. We call the subset $\{k\}\times\{1, \ldots, m\}$ the k-th layer, and write it sometimes as $\mathrm{Lay}(k)$. Transitions are possible from $\mathrm{Lay}(k)$ to $\mathrm{Lay}(k+1)$ and to $\mathrm{Lay}(k-1)$, but also inside the same layer. The transitions probabilities are therefore described by a sequence of triplets (P_k, Q_k, R_k) of positive matrices, where P_k describes the transition probabilities from layer k to layer $k+1$, Q_k from layer k to layer $k-1$, and R_k the transitions inside layer k. For instance

$$P_k(i,j) = P\left(X_{t+1} = (k+1, j) \middle| X_t = (k, i)\right),$$

and similarly for the other matrices. If the triplets are chosen randomly, we have a RWRE on the strip with transitions possible only to the neighbor layers.

It is clear that a RWRE on $\mathbb{Z}$ with possible jumps of maximal size R can be described in the above setup, simply by chopping $\mathbb{Z}$ into pieces

$$\ldots, \{-R+1, \ldots, 0\} \{1, \ldots, R\}, \{R+1, \ldots, 2R\}, \ldots,$$

and declaring these piece to be the layers. Due to the restrictions of the jumps, there are then only jumps to the neighbor layers possible.

We will write the RWRE on the strip by $X_t = (\xi_t, \eta_t)$, where ξ is the component in $\mathbb{Z}$, and η the component in $\{1, \ldots, m\}$.

Our basic assumptions for the first result are

- The sequence $\{(P_k, Q_k, R_k)\}_{k\in\mathbb{Z}}$ is stationary and ergodic.
- An irreducibility assumption I don't spell out explicitly. For instance, the requirement that all matrix elements are positive, almost surely, is more than enough.
- Two moment assumption:

$$\begin{aligned} \mathbb{E} \log (1 - \|R_n + Q_n\|)^{-1} &< \infty, \\ \mathbb{E} \log (1 - \|R_n + P_n\|)^{-1} &< \infty. \end{aligned}$$

Here the matrix norm is

$$\|A\| \stackrel{\text{def}}{=} \max_i \sum_j |A(i,j)|.$$

We will also use the maximum norm in $\mathbb{R}^m$: $\|x\| \stackrel{\text{def}}{=} \max_i |x_i|$.

Crucial in our approach are three sequences of $m \times m$-matrices φ_n, ψ_n, and A_n, $n \in \mathbb{Z}$.

If $a < n$, we define

$$\varphi_{a,n}(i,j) = P_{(n-1,i)} \left(\eta_{T_n} = j, T_n < T_a\right),$$

where T_n is the first hitting time of layer n. Clearly, these matrices are substochastic, and pointwise non-decreasing or $a \downarrow$. So we can define

$$\varphi_n(i,j) \stackrel{\text{def}}{=} \lim_{a\to-\infty} \varphi_{a,n}(i,j).$$

These matrices are also substochastic. By the Markov property, they satisfy the equations

$$\begin{aligned} \varphi_{n+1} &= P_n + R_n\varphi_{n+1} + Q_n\varphi_n\varphi_{n+1}, \\ \varphi_{n+1} &= [I - R_n - Q_n\varphi_n]^{-1} P_n. \end{aligned}$$

We next define a sequence of *stochastic* matrices ψ_n which dominate φ_n. $\psi_n = \varphi_n$ provided φ_n is already stochastic. We define first the vector

$$y_{n+1}(j) \stackrel{\text{def}}{=} \lim_{a \to -\infty} P_{(a,i)}\left(\eta_{T_{n+1}} = j \middle| T_{n+1} < S_a\right).$$

Here S_a is the first *return* time to layer a, i.e. the first hitting time after the first step. If $T_{n+1} < S_a$, and the chain starts in $\text{Lay}(a)$, then the chain moves to the right of a. The limit exists, and is independent of i. This of course needs an argument which I don't give here in detail. Conditioned on $\{T_{n+1} < S_a\}$, the chain is an inhomogeneous Markov chain, and if we start far to the left, the chain "forgets" the starting point. This comes from the irreducibility assumption. We also define

$$\rho_n(i) \stackrel{\text{def}}{=} \sum_j \varphi_{n+1}(i,j),$$

which is the probability that when starting in point i in $\text{Lay}(n)$, the chain escapes to $-\infty$ before reaching $\text{Lay}(n+1)$. Then we set

$$\psi_{n+1}(i,j) \stackrel{\text{def}}{=} \varphi_{n+1}(i,j) + (1 - \rho_n(i))\, y_{n+1}(j),$$

or in short

$$\psi_n = \varphi_n + (1 - \rho_n) \otimes y_{n+1}.$$

$(1 - \rho_n(i))\, y_{n+1}(j)$ is the probability, starting in i in $\text{Lay}(n)$ that the chain first escapes to $-\infty$ before reaching $\text{Lay}(n+1)$, and "after reaching $-\infty$", it hits $\text{Lay}(n+1)$ at j, when hitting it the first time. Of course, in a proper formulation, one has to replace $-\infty$ by a and let $a \to -\infty$.

Lemma 2.2

a) The matrices ψ_n are stochastic matrices

b) They satisfy the equation

$$\psi_{n+1} = P_n + R_n\psi_{n+1} + Q_n\psi_n\psi_{n+1}.$$

c) $\{\psi_n\}$ is the only sequence of stochastic matrices which satisfies this equation.

d) The sequence $(P_n, Q_n, R_n, \varphi_n, \psi_n)$ is stationary and ergodic.

Proof. a) is evident. We give a sketch of the proof of b), assuming for simplicity, that $R_n = 0$.

We already know that $\{\varphi_n\}$ satisfies this equation. Writing $\chi_{n+1} \stackrel{\text{def}}{=} (1 - \rho_n) \otimes y_{n+1}$, we have

$$\begin{aligned}
\psi_{n+1} &= P_n + Q_n\varphi_n\varphi_{n+1} + \chi_{n+1} \\
&= P_n + Q_n(\psi_n - \chi_n)(\psi_{n+1} - \chi_{n+1}) + \chi_{n+1} \\
&= P_n + Q_n\psi_n\psi_{n+1} + \chi_{n+1} \\
&\quad -Q_n\chi_n(\psi_{n+1} - \chi_{n+1}) - Q_n(\psi_n - \chi_n)\chi_{n+1} - Q_n\chi_n\chi_{n+1}.
\end{aligned}$$

We therefore have to show that

$$\begin{aligned}\chi_{n+1} &= Q_n\chi_n\left(\psi_{n+1}-\chi_{n+1}\right)+Q_n\left(\psi_n-\chi_n\right)\chi_{n+1}+Q_n\chi_n\chi_{n+1}\\ &= Q_n\chi_n\varphi_{n+1}+Q_n\varphi_n\chi_{n+1}+Q_n\chi_n\chi_{n+1}.\end{aligned}$$

Now, χ_{n+1} describes the transition for a path "escaping to $-\infty$" before reaching $n+1$. Of course, strictly speaking, one considers the event that it reaches a before $n+1$, and then conditions that it goes back, and one let $a\to-\infty$. Anyway, if it reaches $-\infty$ before $n+1$, it certainly moves to the left when starting in layer n. These transitions are described by Q_n. $\chi_n\varphi_{n+1}$ then describes the transitions (starting in layer $n-1$) under the event $\{T_{-\infty}<T_n\}\cap\{T_{n+1}\circ\theta_{T_n}<T_{-\infty}\circ\theta_{T_n}\}$, where θ_{T_n} means the shift to the time layer n is reached, $\varphi_n\chi_{n+1}$ describes the transitions under the event $\{T_n<T_{-\infty}\}$, and $\chi_n\chi_{n+1}$ under the event $\{T_{-\infty}<T_n\}\cap\{T_{n+1}\circ\theta_{T_n}>T_{-\infty}\circ\theta_{T_n}\}$ (everything starting after the transition to layer $n-1$). This proves (slightly sketchily) the above relations.

For the proof of c), and d) we refer to [2]. ■

The last sequence of matrices we need is

$$A_n\overset{\text{def}}{=}\left[I-R_n-Q_n\psi_n\right]^{-1}Q_n.$$

Remark that in the nearest neighbor one-dimensional case one of course has just $\psi_n=1$, and $A_n=q_n/p_n$.By Kingmans subadditive ergodic theorem

$$\lambda^+=\lim_{n\to\infty}\frac{1}{n}\log\|A_nA_{n-1}\ldots A_1\| \tag{2.2}$$

exists almost surely, and is non-random. The following theorem describes the recurrence-transience behavior of the RWRE in terms of λ^+. It is an extension of the Solomon's Theorem 2.1

Theorem 2.3

a) If $\lambda^+>0$, then $\lim_{t\to\infty}\xi(t)=-\infty$ almost surely

b) If $\lambda^+<0$, then $\lim_{t\to\infty}\xi(t)=\infty$ almost surely

c) If $\lambda^+=0$, then $\limsup_{t\to\infty}\xi(t)=\infty$, and $\liminf_{t\to\infty}S(t)=\infty$, almost surely.

The theorem also states that the RWRE is almost surely recurrent, if and only if $\lambda^+=0$.

It is important that there is a symmetry between "movement to the right" and "movement to the left". Precisely, given the stationary sequence $\{(P_n,Q_n,R_n)\}_{n\in\mathbb{Z}}$, we define the sequence $\left\{\left(\hat{P}_n,\hat{Q}_n,\hat{R}_n\right)\right\}_{n\in\mathbb{Z}}$ by $\hat{P}_n\overset{\text{def}}{=}Q_{-n}$, $\hat{Q}_n\overset{\text{def}}{=}P_{-n}$, $\hat{R}_n\overset{\text{def}}{=}R_{-n}$. This defines a Lyapunov exponent $\lambda^-\overset{\text{def}}{=}\lambda^+\left(\left\{\left(\hat{P}_n,\hat{Q}_n,\hat{R}_n\right)\right\}\right)$.

Lemma 2.4

$\lambda^++\lambda^-=0$.

The proof is not entirely trivial, due to the fact that λ^+ has been introduced in an asymmetric way. I am not giving the details of the proof, and refer to [2] for details. The proof is done by relating our matrices to the stationary distribution of the chain. This stationary distribution is invariant under reflection of the movement.

Corollary 2.5
If (P_n, Q_n, R_n) *is i.i.d. and identical in law to* (Q_n, P_n, R_n), *then* $\lambda^+ = 0$, *and therefore the walk is recurrent.*

This result has also been proved by Mike Keane, and Silke Rolles [13], using Key's approach.

There is an easily proved **dichotomy**. Either one has

A) φ_n is stochastic for all n, almost surely, in which case it follows

$$\limsup_{t\to\infty} \xi_t = \infty, \text{ a.s.}$$

or

B)

$$\sum_j \varphi_n(i,j) < 1$$

holds for all $i \in \{1, \ldots, m\}$, and all n, almost surely, in which case one has

$$\lim_{t\to\infty} \xi_t = -\infty.$$

The dichotomy follows easily from ergodicity, and irreducibility.

Proof of Theorem 2.3.

By Lemma 2.4, a) is equivalent to b). Furthermore, c) follows easily from a), b) and the above dichotomy.

We introduce for $n > a$:

$$\Delta_{a,n} \stackrel{\text{def}}{=} \psi_n - \varphi_{a,n}.$$

Then

$$\Delta_{a,n+1} = A_n \Delta_{a,n} \varphi_{a,n+1}.$$

We can iterate that. Because $\varphi_{a,a+1} = 0$, we get on the one hand

$$\Delta_{a,n+1} = A_n A_{n-1} \cdot \cdots \cdot A_{a+1} \psi_{a+1} \varphi_{a,a+2} \cdot \cdots \cdot \varphi_{n,a}, \tag{2.3}$$

and on the other, by letting $a \to -\infty$ first, putting $\Delta_n \stackrel{\text{def}}{=} \psi_n - \varphi_n$, we get for $n > m$

$$\Delta_{n+1} = A_n A_{n-1} \cdot \cdots \cdot A_m \Delta_m \varphi_{m+1} \cdot \cdots \cdot \varphi_n. \tag{2.4}$$

We prove first "$\Longrightarrow$" in b). We therefore assume that $\lambda^+ < 0$. We use (2.3) with a fixed, e.g. $a = 0$. This then implies that for n large, $\varphi_{0,n}$ is exponentially close to the

stochastic matrix ψ_n. However, this implies that ξ_t converges to ∞, at least with positive probability. But this implies by our dichotomy, that it converges to ∞ with probability 1.

We next prove "$\Longleftarrow$" in a), and here we apply (2.4). Assuming that $\lim_{t\to\infty} \xi_t = -\infty$ implies that we are in situation B). We apply (2.4) to $m = 1$. As Δ_1 is strictly positive, we see that

$$\lambda^+ = -\lim_{n\to\infty} \log \|\varphi_2 \cdot \cdots \cdot \varphi_n\| > 0,$$

because the φ_i form a stationary sequence of strictly substochastic matrices. ■

Remark 2.6
One dimensional RWREs can have the interesting property that $X_t \to \infty$ holds almost surely, but $X_t/t \to 0$. In fact, if $\mathbb{E}(q_x/p_x) \geq 1$ and $\mathbb{E}(p_x/q_x) \geq 1$ then

$$\lim_{t\to\infty} \frac{X_t}{t} = 0, \ \hat{P}-\text{a.s.}$$

even if $\lambda^+ \neq 0$ (see e.g. [18]). Similar phenomena appear in the strip case, too, see the recent paper by Goldsheid [7]. I will not discuss this here.

2.2 The Sinai-behavior: Small displacement in the recurrent case

In the strictly one-dimensional i.i.d. nearest-neighbor case with $\lambda^+ = \mathbb{E}\log(q/p) = 0$, Sinai.[14] proved the remarkable result that the displacement of the RWRE after time t is only of order $(\log t)^2$. This is based on the fact that the functions $h_{a,b}$ defined in (2.1) can be expressed in terms of the random walk

$$\phi_k \stackrel{\text{def}}{=} \sum_{j=1}^{k} \log \frac{q_j}{p_j}$$

(extended properly to a two-sided walk with $\phi_0 = 0$). For instance, if $a < k < b$, then

$$P_{k,\omega}(T_a < T_b) \leq \text{const} \times \sum_{j=k}^{b} \exp\left[\phi_j - \phi_a\right].$$

The random walk $\{\phi_k\}$ has on k-intervals of size m fluctuations of size $\sqrt{m}$. From standard fluctuation properties of a random walk, one therefore has "valleys" of depth const $\times\sqrt{m}$ on k-intervals of size m. Using the above type of estimates, one can derive that it takes $\exp[\text{const} \times \sqrt{m}]$ time steps for the RWRE to escape from such a valley. From that one sees that the RWRE typically moves away from starting point of order m, after a time of order $\exp[\text{const} \times \sqrt{m}]$. In other words, after time t, one has moved away of order $(\log t)^2$. There is a more precise formulation given below for our strip RWRE which states that the RWRE in fact stays most of the time close to the "bottom" of properly chosen valleys.

As this has been discussed in many survey articles (see e.g. [18]), I don't spell out the details here, and focus on the methods which are needed in the strip case (or non-nearest neighbor case). For the non-nearest neighbor one-dimensional case, there is also a paper by Letchikov [11] on the topic.

We will always assume here that the sequence $\{(P_n, Q_n, R_n)\}_{n\in\mathbb{Z}}$ is i.i.d., and we write μ for its law. The rôle of the above random walk in the strip case is played by

$$\phi_k \stackrel{\text{def}}{=} \begin{cases} \log \|A_k \cdot \cdots \cdot A_1\|, & k \geq 1 \\ 0 & k = 0 \\ -\log \|A_0 \cdot \cdots \cdot A_{k+1}\| & k \leq -1 \end{cases}.$$

In contrast to the strictly neighbor one-dimensional situation, the $\{\phi_k\}$ is not a sum of independent random variables, but it can be represented as a functional of an ergodic Markov chain.

Let $\mathcal{J} \stackrel{\text{def}}{=} \{(P,Q,R) : P,Q,R \text{ non} - \text{negative}, P+Q+R \text{ stochastic}\}$, and

$$\mathcal{J}_\varepsilon \stackrel{\text{def}}{=} \left\{(P,Q,R) \in \mathcal{J} : (I-R)^{-1} P, (I-R)^{-1} Q \geq \varepsilon, \text{ pointwise}, \exists l \in \mathbb{N} : \left\|R^l\right\| \leq 1-\varepsilon\right\}.$$

We will assume that our transition probabilities are in $\mathcal{J}_\varepsilon$ for some $\varepsilon > 0$. This is much stronger than really needed, but keeps somewhat uninteresting technical issues out. It is evident that if $(P,Q,R) \in \mathcal{J}_\varepsilon$, then there is a unique probability $\pi = \pi_{(P,Q,R)}$ on $\{1,\ldots,m\}$ satisfying $\pi(P+Q+R) = \pi$. We set

$$\mathcal{J}_{\text{al}} \stackrel{\text{def}}{=} \left\{(P,Q,R) \in \mathcal{J}_\varepsilon : \pi_{(P,Q,R)}(P-Q)\,\mathbf{1} = 0\right\}.$$

Remark 2.7
If $(P,Q,R) \in \mathcal{J}_\varepsilon$, we can consider the Markov chain on the strip which has these as fixed transition probabilities. Evidently, this Markov chain is recurrent if and only if $(P,Q,R) \in \mathcal{J}_{\text{al}}$. On the other hand, we can characterize recurrence also in terms of $\lambda^+_{(P,Q,R)}$ which is defined as (2.2), but with this fixed "environment". Recurrence is then also characterized by $\lambda^+_{(P,Q,R)} = 0$. Therefore one has

$$(P,Q,R) \in \mathcal{J}_{\text{al}} \Longleftrightarrow \lambda^+_{(P,Q,R)} = 0.$$

Theorem 2.8
Assume that the (P_n, Q_n, R_n) are i.i.d. with law μ, and

(a) $\lambda^+ = 0$,

(b) $\text{supp}(\mu) \subset \mathcal{J}_\varepsilon$ *for some* $\varepsilon > 0$, *and*

(c) $\text{supp}(\mu) \not\subseteq \mathcal{J}_{\text{al}}$.

Then there exists a sequence $\{b_t\}_{t\in\mathbb{N}}$ of random variables, converging in distribution, such that for any $\varepsilon' > 0$, one has

$$\lim_{t\to\infty} \mathbb{P}\left(\left\{\omega : P_{0,\omega}\left(\left|\frac{\xi_t}{(\log t)^2} - b_t\right| \leq \varepsilon'\right) \geq 1-\varepsilon'\right\}\right) = 1. \tag{2.5}$$

The sequence $\{b_t\}$ is defined in the same way as in the standard one-dimensional RWRE (see [18]), but in terms of our random walk $\{\phi_k\}$. b_t is described as the "bottom place" of properly chosen valleys of depth 1 in the rescaled random walk $\left\{ \phi_{[k \log^2 t]} \Big/ \log t \right\}$. For details of the construction, see [18]. The limit distribution of b_t has been derived by Kesten [9]. The proof of the theorem runs somewhat parallel to the proof in the classical case, provided one proves a functional CLT for $\{\phi_k\}$. There are some complications as the estimates are typically not quite as sharp as in the classical case, but the changes are mainly technical. I will not go into this here, and refer the reader to [3] for details. I will however give some explanations how to derive the CLT, and how the condition (c) appears. This condition is not necessary for the Sinai-type behavior, but curiously it is so for the non-nearest neighbor $\mathbb{Z}$-case which can be stated as follows.

Consider a sequence $\{p_k\}_{k \in \mathbb{Z}}$ of i.i.d. random probability distributions on $\{-R, \ldots, R\}$, describing the transitions of a RWRE $\{X_t\}$ on $\mathbb{Z}$, $p_k(i)$ being the probability of changing from k to $k+i$. Set

$$\widetilde{\mathcal{J}}_{\text{al}} \stackrel{\text{def}}{=} \left\{ p : \sum\nolimits_k kp(k) = 0 \right\}.$$

Theorem 2.9
Assume that the p_k are i.i.d., that there exists $\varepsilon > 0$ such that $p_k(1), p_k(-1), p_k(R)$, $p_k(-R) \geq \varepsilon$, almost surely, and that the support of the law of the p_k is not contained in $\widetilde{\mathcal{J}}_{\text{al}}$. Then one has Sinai-type behavior, i.e. $\{X_t\}$ satisfies (2.5).

In that case, the condition on the support is actually necessary, as one has the following

Theorem 2.10
Assume the same as above, except that the support of the p_k is contained in $\widetilde{\mathcal{J}}_{\text{al}}$. Then there exists $\sigma^2 > 0$ such that for almost all ω, one has

$$\lim_{t \to \infty} P_{0,\omega} \left(t^{-1/2} X_t \leq x \right) = \Phi\left(x; 0, \sigma^2\right), \ \forall x \in \mathbb{R},$$

where $\Phi\left(\cdot; a, \sigma^2\right)$ denotes the distribution function of the Gaussian distribution with mean a and variance σ^2.

This result is somewhat similar to a result by Lawler [10], and can be proved by an adaptation of his methods. In fact, of the support of the law of the p_k is in $\widetilde{\mathcal{J}}_{\text{al}}$, then the Markov chain is a martingale, and one can apply the martingale central limit theorem, but there remains the difficulty to see that the variance σ^2 is positive. I will not discuss this here, and refer to [3].

The Theorem 2.9 is an immediate consequence of Theorem 2.8.

I discuss now, how to derive the necessary CLT for $\{\phi_k\}$ which leads to Theorem 2.8. One uses the machinery of random transformations. Consider the space $M \stackrel{\text{def}}{=} \Psi \times \mathbf{X}$, where Ψ is the space of stochastic matrices, and $\mathbf{X}$ is the space of non-negative vectors x with $\|x\| = 1$. ($\|\cdot\|$ was the maximum norm). Then $g = (P, Q, R) \in \mathcal{J}_\varepsilon$ operates on M

by

$$g.\,(\psi,x)\stackrel{\text{def}}{=}\left(\left(I-R-Q\psi\right)^{-1}P,Ax/\left\|Ax\right\|\right),$$

where $A\stackrel{\text{def}}{=}\left(I-R-Q\psi\right)^{-1}Q$.

Starting with (ψ_a,x_a), and an i.i.d. sequence $\{g_n\}$, one defines recursively for $n\geq a$: $\left(\psi_{n+1},x_{n+1}\right)\stackrel{\text{def}}{=}g_{n+1}.\,(\psi_n,x_n)$. Letting $a\to-\infty$, one then obtains a stationary sequence $\{(g_n,\psi_n,x_n)\}_{n\in\mathbb{Z}}$. The distribution of (g_n,ψ_n,x_n) (for fixed n) is $\mu\otimes\nu$, where μ is the distribution of the g_n, and ν is the stationary distribution on M. The above sequence is a stationary ergodic Markov chain with a transition operator $\mathcal{A}$, operating on (continuous) functions F on $\mathcal{J}_\varepsilon\times M$ by

$$\mathcal{A}F\,(g,(\psi,x))\stackrel{\text{def}}{=}\int F\left(g',g.\,(\psi,x)\right)\mu\left(dg'\right).$$

The random variables of interest, namely ϕ_k can then be written (also in the case of $\lambda^+\neq0$) as

$$\begin{aligned}\phi_n&=\log\left\|A_n\cdot\cdots\cdot A_1\right\|-n\lambda^+\\&=\sum_{k=1}^{n}\left(\log\left\|A_kx_k\right\|-\lambda^+\right)+O\left(1\right),\end{aligned}$$

for $n\geq1$, with a natural definition also for $n\leq0$. From the standard CLT for functionals of ergodic Markov chains one then gets the desired statement that

$$\left\{\phi_{[nt]}/\sigma\sqrt{n}\right\}_{t\in\mathbb{R}}$$

converges weakly as $n\to\infty$ to a (two-sided) Brownian motion, provided the variance $\sigma^2>0$. This variance is given by

$$\sigma^2=\int_{\mathcal{J}_\varepsilon\times M}\left[\mathcal{A}F^2-\left(\mathcal{A}F\right)^2\right]d\left(\mu\otimes\nu\right),$$

where $F:\mathcal{J}_\varepsilon\times M\to\mathbb{R}$ is the unique (continuous) solution of the equation

$$F\,(g,(\psi,x))-\mathcal{A}F\,(g,(\psi,x))=\log\left\|\left(I-R-Q\psi\right)^{-1}Qx\right\|-\lambda^+.\tag{2.6}$$

If $\sigma^2>0$, then we get the functional CLT for the sequence $\{\phi_k\}$ which leads to the Sinai-type behavior. We prove now that our assumptions imply that $\sigma^2>0$.

Lemma 2.11
If $\lambda^+=0$, *and* $\sigma^2=0$ *then* $\mathrm{supp}\,(\mu)\subset\mathcal{J}_{\mathrm{al}}$.

Proof. We first observe that if $g=(P,Q,R)\in\mathcal{J}_\varepsilon$ is fixed, then there exists a unique $z_g=\left(\psi_g,x_g\right)\in M$ with $g.z_g=z_g$. This corresponds to just taking *fixed* transition

probabilities given by (P,Q,R) (not depending on n). $\left(\psi_g, x_g\right)$ is obtained by applying the constant operation N times, and let $N \to \infty$. There exists $\lambda_g^+ \in \mathbb{R}$, such that

$$\left(I - R - Q\psi_g\right)^{-1} Q x_g = \mathrm{e}^{\lambda_g^+} x_g,$$

where λ_g^+ is the number introduced in Remark 2.7. Therefore, $\lambda_g^+ = 0$ holds if and only if $g \in \mathcal{J}_{\mathrm{al}}$. The statement of the lemma therefore says that if $\lambda^+ = 0$ and $\sigma^2 = 0$, then $\lambda_g^+ = 0$ is true for all $g \in \mathrm{supp}\,(\mu)$.

Now,

$$\sigma^2 = \int \left[F\left(g', g.z\right) - \int F\left(\widetilde{g}, g.z\right) \mu\left(d\widetilde{g}\right) \right]^2 \mu\left(dg'\right),$$

$z = (\psi, x)$, where F is the solution of (2.6). It is readily checked that the solution can be chosen to be continuous. Therefore, if $\sigma^2 = 0$, then $F(g,z)$ does not depend on the first coordinate, i.e. $F(g,z) = \overline{F}(z)$ for some continuous function $\overline{F}$, for all (g,z) in the support of $\mu \otimes \nu$. From this it follows that

$$\log \left\| (I - R - Q\psi)^{-1} Qx \right\| = \overline{F}(z) - \overline{F}(g.z),$$

$g = (P,Q,R)$, $z = (\psi, x)$. In particular,

$$\lambda_g^+ = \log \left\| \left(I - R - Q\psi_g\right)^{-1} Q x_g \right\| = 0,$$

which is the required conclusion. ■

3 Exit distributions in dimensions ≥ 3

We switch now to the nearest-neighbor RWRE in $\mathbb{Z}^d$. For $x \in \mathbb{Z}^d$, ω_x is concentrated on $\left\{e \in \mathbb{Z}^d : |e| = 1\right\}$. We use e exclusively for a nearest neighbor point to 0 in $\mathbb{Z}^d$. We also assume a smallness assumption of the disorder, i.e. we require that $\mathbb{P}$ concentrates on transition probabilities $\omega_x(e)$, satisfying

$$\left| \omega_x(e) - \frac{1}{2d} \right| \leq \varepsilon.$$

Furthermore, of course, the ω_x are chosen i.i.d. We denote by μ the distribution of ω_x :

Condition 3.1

1. *μ is invariant under lattice isometries, i.e. under orthogonal transformations which leave the lattice $\mathbb{Z}^d$ invariant.*
2. *$d \geq 3$*
3. *$\varepsilon > 0$ is small.*

If $U \subset\subset \mathbb{Z}^d$, $x \in U$, we denote the exit distribution of the ordinary random walk by $\pi_U(x,z)$, $z \in \partial U$, the outer boundary of U. For the RWRE, we write $\Pi_{U,\omega}(x,z)$. The aim is to show that the RWRE exit distribution is close π_U.

The basis is a perturbation argument. We represent the RWRE through the ordinary one by a standard perturbation expansion for the Green's function. We write g_U for the Green's function of the ordinary RW with Dirichlet boundary, i.e.

$$g_U \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} (1_U p)^n,$$

where $p(x, x+e) = (2d)^{-1}$, and $(1_U p)(x,y) = 1_U(x)\, p(x,y)$. For two function $F, G : \mathbb{Z}^d \times \mathbb{Z}^d \to \mathbb{R}$, we write FG for the kernel

$$(FG)(x,y) \stackrel{\text{def}}{=} \sum_z F(x,z)\, G(z,y),$$

if this is defined. F^n is just the n-fold product, defined in this way, and $F^0(x,y) \stackrel{\text{def}}{=} \delta_{x,y}$. For $x \in U, z \in \partial U$, we have $g_U(x,z) = \pi_U(x,z)$. We also write $G_{U,\omega}$ for the corresponding kernel for the RWRE. Put

$$\Delta_{U,\omega}(x, x+e) \stackrel{\text{def}}{=} 1_U(x) \left(\omega_x(e) - \frac{1}{2d}\right).$$

Then

$$G_{U,\omega} = g_U + g_U \Delta_{U,\omega} G_{U,\omega}. \tag{3.1}$$

We can iterate this, and obtain

$$G_U = \sum_{n=0}^{\infty} (g_U \Delta)^n g_U. \tag{3.2}$$

(Convergence of such expansions will never be a problem in our setup). In particular, if $x \in U, z \in \partial U$

$$\Pi_U(x,z) = \left(\sum_{n=0}^{\infty} (g_U \Delta)^n \pi_U\right)(x,z).$$

I present here an outline of the approach developed together with Ofer Zeitouni [5]. We have not been the first to investigate this problem. A classical paper is the one by Bricmont and Kupiainen [6]. A CLT for diffusions in random environments is given in [17].

3.1 Why is the disorder contracting in $d \geq 3$?

We apply the perturbation expansion to the situation where ε (i.e. the disorder) is small and U is a large centered box of side length L, call it U_L, but we will have to see in which relation with ε they can be.

The crucial point is the behavior of the linear term with $n = 1$. For $z \in \partial U_L$, and starting point 0, we have

$$\Pi(0,z) = \pi(0,z) + \sum_{y \in U, e} g(0,y)\,\Delta(y,y+e)\,\pi(y+e,z) + \sum_{n=2}^{\infty} \mathrm{NL}_n(0,z),$$

where NL_n denotes the non-linear terms. The first important fact one observes is that in the linear summand, one has $\sum_e \Delta(y,y+e) = 0$, and therefore, one can rewrite it as

$$\sum_{y \in U, e} g(0,y)\,\Delta(y,y+e)\,[\pi(y+e,z) - \pi(y,z)].$$

The summation close to the boundary needs some care, but as this issue becomes much more delicate later anyway, I consider here only the summation over y in the bulk, say in $U_{L/2}$. Then $\pi(y,z)$ is of order L^{-d+1}, and $\pi(y+e,z) - \pi(y,z)$ is of order L^{-d}. (These estimates can be found in Lawler's green random walk book). The only place where randomness enters into this summand is in Δ and evidently from the symmetry assumption, one has $\mathbb{E}\gtrless = 0$, $\mathbb{E}\gtrless^2 \leq \varepsilon^2$. Therefore

$$\mathrm{var}\left(\sum_{y \in U_{L/2}, e} g(0,y)\,\Delta(y,y+e)\,[\pi(y+e,z) - \pi(y,z)]\right) \leq \mathrm{const} \times L^{-2d}\varepsilon^2 \sum_{y \in U_{L/2}} g(0,y)^2.$$

In $d \geq 3$, we can estimate the Green's function on U_L by the Green's function on the infinite lattice which is bounded by $\mathrm{const} \times (1+|y|)^{-d+2}$. Therefore

$$\sum_{y \in U_{L/2}} g(0,y)^2 \leq \mathrm{const} \times \begin{cases} 1 & d \geq 5 \\ \log L & d = 4 \\ L & d = 3 \end{cases}.$$

For the contribution of this part to the exit distribution, we therefore get in the "worst case" $d = 3$

$$\mathbb{E} \sum_{z \in \partial U_L} \left| \sum\nolimits_{y \in U_{L/2}, e} g(0,y)\,\Delta(y,y+e)\,[\pi(y+e,z) - \pi(y,z)] \right| \leq \mathrm{const} \times \varepsilon L^{-1/2}.$$

By using an exponential inequality, e.g. Hoeffding's inequality, then one easily gets good exponential bounds. We will come to this later, but one sees that the disorder is contracting for $d \geq 3$, when one looks only at the linear term, and if one neglects boundary effects. If one does the calculation carefully, one gets that the disorder is also borderline contracting in $d = 2$[1].

However, the nonlinear terms start to make trouble for fixed ε and large L. First of all, there is a complicated dependency structure emerging. One can try to estimate the non-linear terms very crudely, by estimating uniformly in ω. Roughly, estimating

[1] A paper treating the two-dimensional case is in preparation (with Ofer Zeitouni).

$(g_U \Delta)^n \pi_U$, taking only the contribution from the bulk, one gets (for $d = 3$), an L^2 from every g_U, then an ε from the Δ, and finally an L^{-1} from π_U (after summing over the boundary, but taking into account the gain by one derivative). Therefore a crude bound for the total variation is given by $\left(L^2\varepsilon\right)^n L^{-1}$. In order to sum that, one has the requirement that $L^2\varepsilon < 1$. Of course, there is no possibility to argue in this crude way, if ε is fixed and L become large.

The basic idea is to try to safe this kind of arguments by doing a multiscale argument. This means that one does not jump from scale 1 immediately to scale L, but one jumps so in a number of intermediate steps, for which one can argue somewhat crudely when jumping from one scale to the next.

3.2 The exact setting

The main scheme is to describe an induction scheme how to transfer information about exit distributions one has on scales $\leq l$ to information on a bigger scale L (or better on scales $\leq L$). One essentially takes

$$L \stackrel{\text{def}}{=} l \left(\log l\right)^3.$$

I will be more precise about this later. The reason one is increasing the scale faster than exponential is that this enables one to treat "bad" regions very crudely. The "badness" is washed out quickly just by the fast increase of the scale, as we will see.

Instead of boxes, we take (lattice) Euclidean balls

$$V_L \stackrel{\text{def}}{=} \left\{x \in \mathbb{Z}^d : |x| \leq L\right\},$$

where $|x|$ is the Euclidean norm. (L here is not necessarily an integer). We also write $V_L(x) \stackrel{\text{def}}{=} x + V_L$. (The reason that we didn't work with square boxes is that there are some "corner problems" which create troubles in the induction step. For the final result, this does not play any rôle, but for the induction procedure, we found that square boxes are cumbersome to handle).

For $x \in V_L, z \in \partial V_L$, we write $\pi_L(x, z)$ for the probability that *ordinary* random walk starting in x exits V_L at z, and $\Pi_{L,\omega}(x, z)$ for the corresponding quantity of the RWRE.

We now want to represent the exit distribution for the large scale L through exit distributions on the smaller scale, and then use an averaging argument for the linear term in the perturbation expansion, and crude estimates in the non-linear terms. There are a number of difficulties.

1. We found that we should always work with exits from centered balls. The advantage is that the distribution (under $\mathbb{P}$) of the exit distribution inherits trivially the symmetry properties of the original RWRE transition probabilities.

2. Close to the boundary of the big L-box, one has to refine the scale of the smaller boxes. This creates a number of technical difficulties.

3. One of the main difficulties: It is clear that the disorder is *not* contracting when measured in total variation. This is (hopefully) coming from random effects close to the boundary. The main result is therefore a result on the contracting of the disorder when looking at
$$(\Pi_{L,\omega} - \pi_L)\,\Sigma_L$$
where Σ_L is a certain smoothing kernel.

4. With $\mathbb{P}$-probability 1 there are arbitrary large "bad" regions in $\mathbb{Z}^d$, and one has to show that they don't have a big influence (which they do for $d = 1$).

The smoothing kernel is defined in a somewhat complicated way. The motivation for this choice is that it is the "right one" for the induction procedure, as we will see. Here is the definition:
$$\Sigma_L(0,z) \overset{\text{def}}{=} \int_1^2 \psi(t)\,\pi_{tL}(0,z)\,dt,$$
where $\psi : \mathbb{R} \to \mathbb{R}^+$ is a smooth probability density with support on $[1,2]$. $\Sigma_L(x,z) \overset{\text{def}}{=} \Sigma_L(0,z-x)$. The reason for making this averaging over the radius is simply to have a kernel which is smooth enough. For instance, clearly
$$\sup_{x,z} \Sigma_L(x,z) \leq \text{const} \times L^{-d},$$
and with a bit of work
$$\sup_{x,z,e} |\Sigma_L(x+e,z) - \Sigma_L(x+e,z)| \leq \text{const} \times L^{-d-1}.$$

The main result we have is the following

Theorem 3.2
There exists $\varepsilon_0 > 0$ such that for $\varepsilon \leq \varepsilon_0$, the following is true: For any $\eta > 0$, there exists a smoothing radius $\zeta(\eta)$ such that
$$\lim_{L\to\infty} \mathbb{P}\left(\left\| (\Pi_L - \pi_L)\,\Sigma_{\zeta(\eta)}(0,\cdot) \right\|_{\text{var}} \geq \eta \right) = 0.$$

The theorem states that if we want to have an estimate η for the deviation in total variation, then one has to apply a certain fixed smoothing on a scale $\zeta(\eta)$. The form of the smoothing kernel is actually not important in the final result. The special form above is only useful in the induction procedure.

Remark 3.3
If one is ready to increase the smoothing scale, then the difference goes to 0 : If $\zeta_L \to \infty$, then there is a sequence $\eta_L \to 0$ such that
$$\lim_{L\to\infty} \mathbb{P}\left(\left\| (\Pi_L - \pi_L)\,\Sigma_{\zeta_L}(0,\cdot) \right\|_{\text{var}} \geq \eta_L \right) = 0.$$

The main induction procedure, we apply, makes only statement about "globally smoothed" exit distributions, and "non-smoothed" ones. I describe this now precisely:

We call the ball $V_L(x)$ good provided

$$\|(\Pi_L - \pi_L)\Sigma_L(x,\cdot)\|_{\text{var}} \leq (\log L)^{-9}, \tag{3.3}$$

and

$$\|(\Pi_L - \pi_L)(x,\cdot)\|_{\text{var}} \leq \delta. \tag{3.4}$$

(This depends also on a parameter $\delta > 0$).

It however turns out that the appearance of "bad" regions makes it necessary to distinguish between different degrees of badness, but fortunately, only four levels are needed. We call a box $V_L(x)$ "bad on level i", $i \leq 3$, provided

$$(\log L)^{-9+9(i-1)/4} < \|(\Pi_L - \pi_L)\Sigma_L(x,\cdot)\|_{\text{var}} \leq (\log L)^{--9+9i/4},$$

and still (3.4) is satisfied. Balls which are neither good, nor bad on any of the levels $1,2,3$, are called "really bad", or bad on level 4.

Our main inductive result is the following: For $\delta > 0$, $K \in \mathbb{N}$, we denote by $\text{Cond}(\delta, K)$ the condition that for all $L \leq K$

$$\mathbb{P}(V_L(0) \text{ is bad on level } i) \leq \exp\left[-(1-(4-i)/13)(\log L)^2\right],\ i = 1,2,3,4.$$

Proposition 3.4
There exists $\delta_0 > 0$ such that for $0 < \delta \leq \delta_0$, there exist $\varepsilon_0(\delta)$, and $L_0 \in \mathbb{N}$, such that if $\varepsilon \leq \varepsilon_0$, $L_1 \geq L_0$, then

$$\text{Cond}(\delta, L_1) \Longrightarrow \text{Cond}\left(\delta, L_1(\log L_1)^2\right).$$

We take here only $L_1(\log L_1)^2$ on the right hand side, the point being that if $L \leq L_1(\log L_1)^2$, then $L/(\log L)^3 \leq L_1$. In the above formulation, I am actually cheating slightly. One has in fact to include still a slightly more complicated version regarding the smoothing scheme. This is due to some additional complications close to the boundary, but it is a very minor issue, and I leave a discussion of it out.

It is trivial to start the recursion scheme by choosing ε small enough. The above proposition at first sight gives only results about globally smoothed exit distributions (and non-smoothed ones, but the differences between RWRE and ordinary RW does not go to 0). The proof of the above proposition however derives from $\text{Cond}(\delta, L_1)$ properties about exits on the larger scale with smoothings on intermediate scale, too, but for the advancement of the induction, only the globally smoothed, and the non-smoothed information is needed.

We have to fix transition kernels for the movement inside a "big" ball V_L. The movement is by exit distributions on smaller centered balls, but we also perform an additional randomization of the radius. We also have to refine the jumping radius close to the boundary. We call this procedure the "coarse graining scheme". The coarse

graining scheme is described by fixing for every $x \in V_L$ the relevant radius $\rho_L(x)$. This is always chosen such that in such a way that $V_{2\rho(x)}(x) \subset V_L$, except possibly for the "very last layer". Let $d_L(x) \stackrel{\text{def}}{=} \operatorname{dist}(x, \partial V_L)$. If $d_L(x) \geq 2L/(\log L)^3$, then $\rho_L(x) \stackrel{\text{def}}{=} \gamma L/(\log L)^3$, where $\gamma > 0$ is still a small parameter, we did not specify. $1/10$ is probably fine. For $d_L(x) < 2L/(\log L)^3$ the coarse graining radius starts to shrink. It shrinks linearly in the region $d_L(x) < L/(\log L)^3$, where we put $\rho_L(x) \stackrel{\text{def}}{=} \gamma d_L(x)$. In the region $L/(\log L)^3 \leq d_L(x) < 2L/(\log L)^3$, we make a smooth transition between these behaviors. Our basic transition kernel for the ordinary RW is $\hat{p}_L(x, \cdot) \stackrel{\text{def}}{=} \Sigma_{\rho_L(x)}(x, \cdot)$. (Remark that if $\rho_L(x) < 1/2$, then this is just the ordinary nearest neighbor transition). It is also evident that the exit distribution from V_L for a Markov chain with these transition probabilities is just the ordinary RW exit distribution π_L. We do the same coarse graining for the RWRE, and write $\hat{P}_{L,\omega}(x, \cdot)$ for these kernels.

This is the coarse-graining scheme we use for the "non-smoothed" estimate. For the other one, we make a small modification: We stop the refinement of the coarse-graining for points $d_L(x) \leq L/(\log L)^{10}$. For such points, we just take the exit distribution from $V_{10L/(\log L)^{10}}(x) \cap V_L$, also without averaging over the radius. I don't distinguish the two coarse-graining schemes in the notation, but will tell always which one to take. In this case we write $\rho_L(x) \stackrel{\text{def}}{=} 10L/(\log L)^{10}$.

We write $\hat{g}_L$ for the Green's function of $\hat{p}_L$;

$$\hat{g}_L \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} \left(1_{V_L} \hat{p}_L\right)^n,$$

and $\hat{\Delta}_{L,\omega}$ for the difference $\hat{P}_{L,\omega} - \hat{p}_L$. Then the perturbation expansion gives

$$\Pi_{L.\omega} - \pi_L = \hat{g}_L \hat{\Delta}_{L,\omega} \pi_L + \hat{g}_L \hat{\Delta}_{L,\omega} \hat{g}_L \hat{\Delta}_{L,\omega} \pi_L + \dots \tag{3.5}$$

The reason for choosing the smoothing kernel in the particular way we did is that

$$\hat{g}_L = I + 1_{V_L} \hat{p}_L \hat{g}_L.$$

Therefore, we split

$$\hat{\Delta}_{L,\omega} \hat{g}_L = \hat{\Delta}_{L,\omega} + \left(\hat{\Delta}_{L,\omega} 1_{V_L} \hat{p}_L\right) \hat{g}_L,$$

and then we use our induction assumption on the smoothed version in the second summand, and the non-smoothed one in the first summand. The above expansion is then

$$\Pi_{L.\omega} - \pi_L = \sum_{k=1}^{\infty} \hat{g}_L \hat{\Delta}_{L,\omega}^k \pi_L + \sum_{k_1,k_2=1}^{\infty} \hat{g}_L \hat{\Delta}_{L,\omega}^{k_1} 1_{V_L} \hat{p}_L \hat{g}_L \hat{\Delta}_{L,\omega}^{k_2} \pi_L + \dots$$

and the philosophy is to estimate $\left\| \hat{\Delta}_{L,\omega}^k 1_{V_L} \hat{p}_L \right\|_{\text{var}} \leq \delta^{k-1} (\log \rho)^{-9}$ in good regions.

One of the main issues are to have good estimates on $\hat{g}_L$. Essentially it should be like the Green's function of the ordinary nearest neighbor RW with a scaling due to

the spreading. The ORW Green's function should be like the Brownian motion Green's function, which is explicitly known in a ball of radius L

$$g_L^{\mathrm{BM}}(x,y) = \frac{1}{d(2-d)\omega_d}\left(|x-y|^{2-d} - \left|\frac{L}{|x|}x - \frac{|x|}{L}y\right|^{2-d}\right),$$

where ω_d is the volume of the unit ball. The ordinary RW Green's function g_L^{ORW} is then essentially the same except for a cutoff of the singularities. For discussing $\hat{g}_L$, one has to use the appropriate "coarse-graining" scale $\rho_L(x)$.

Pseudotheorem
For all what is necessary, $\hat{g}_L(x,y)$ behaves for $y \neq x$ like $\rho_L(y)^{-d} g_L^{\mathrm{ORW}}(x,y)$, and also the (discrete) derivatives behave like the derivatives of $\rho_L(y)^{-d} g_L^{\mathrm{ORW}}(x,y)$.

We didn't prove such a theorem, which is certainly not true in this strong form, but whenever we needed something, we proved an appropriate statement ad hoc. For instance, an easy (and evident) property is that the expected total time spent by our coarse grained walk in a region $\{x \in V_L : t \leq d_L(x) \leq 2t\}$ for $t \leq L/(\log L)^3$ is of order 1, uniformly in t, and uniform in the starting point. Therefore, the expected time, the walk spends in $\left\{d_L(x) \leq L/(\log L)^3\right\}$ is of order $\log L$. These things are easy. More delicate are estimates on derivatives.

3.3 Advancement of the smoothed estimate

We use the coarse graining scheme with the stopped boundary refinement in the "last layer" $\left\{x : d_L(x) \leq L/(\log L)^{10}\right\}$.

An essential issue is how to treat bad regions. The main reason for stopping the refinement of the coarse-graining scheme (for the smoothed estimates) at $L/(\log L)^{10}$ is that we don't have to cope with complicated bad regions. The reason is the following elementary estimate. We call a point $x \in V_L$ bad if $V_l(x)$ is bad (on any of the four levels), for any l between $\rho_L(x) \leq l \leq 2\rho_L(x)$, of $d_L(x) > L/(\log L)^{10}$, or in case $d_L(x) \leq L/(\log L)^{10}$, if $V_{\rho_L(x)}(x)$ is bad. We denote by $B_{L,\omega}$ the set of bad points.

Lemma 3.5
If $\mathrm{Cond}(\delta, L_1)$ *is satisfied, and* $L \leq L_1(\log L_1)^2$, *then*

$$\mathbb{P}\left(\bigcup_{x\in V_L}\left\{B_{L,\omega} \subset V_{5\rho_L(x)}(x)\right\}\right) \geq 1 - \frac{1}{100}\exp\left[-(\log L)^2\right].$$

Proof. If $B_{L,\omega}$ is not contained in some $V_{5\rho_L(x)}(x)$, then there are points $x, y \in B_{L,\omega}$ sufficiently apart such that $\{x \in B_L\}$ and $\{y \in B_L\}$ are independent events. Therefore, under $\mathrm{Cond}(\delta, L_1)$, one estimates this probability by

$$\mathrm{const} \times L^{2d}\left[\exp\left[-\left(1-\frac{3}{13}\right)(\log L)^2\right]\right]^2 \leq \frac{1}{100}\exp\left[-(\log L)^2\right].$$

■

If the bad region B_L cannot be confined in the way above, we simply trash the box V_L declare it to be "really bad". (Of course, we do the same with any of the boxes $V_L(x)$, when used in the next induction step). From the above estimate, this is well inside the desired bound. Therefore, we essentially have to deal only with at most region of the form $V_{5\rho_L(x)}(x) \subset V_L$ which contains any bad points. The first thing to do is to get estimates when all points are "good". We want to derive under this condition estimates which then can be used also to treat the possible one bad region. In order to do this (formally), we "goodify" possible bad points by exchange there the RWRE coarse grained transition probabilities $\hat{P}_L(x,\cdot)$, simply by $\hat{p}_L(x,\cdot)$. Then we want to do two things

- Prove that after this "goodifying" manipulation, we have

$$\mathbb{P}^{\text{goodified}}\left(\left\|\left(\Pi_L - \pi_L\right)\Sigma_L\right\|_{\text{var}} \geq (\log L)^{-9}\right) \leq \frac{1}{10}\exp\left[-(\log L)^2\right]. \tag{3.6}$$

- If a constant $C > 0$ is chosen properly, then

$$\begin{aligned} &\mathbb{P}^{\text{goodified}}\left(\hat{G}_L\left(x, V_{\rho_L(y)}(y)\right) \geq C\hat{g}_L\left(x, V_{\rho_L(y)}(y)\right), \text{ some } x, y \in V_L\right) \\ \leq\ &\frac{1}{10}\exp\left[-(\log L)^2\right]. \end{aligned} \tag{3.7}$$

The last point will help to control the bad region.

I will sketch the argument leading to the first estimate. In the perturbation expansion (3.5) we first consider the linear term, but in fact we have also take the ones coming from the non-linear ones, without any smoothing possibility within the Δ's. Therefore, we consider we have to estimate

$$\sum_{k\geq 1} \hat{g}_L \hat{\Delta}_L^k \pi_L \Sigma_L. \tag{3.8}$$

The summation over k is no problem, we in fact estimate $\hat{\Delta}_L^{k-1}$ in total variation simply by δ^{k-1}. There are some minor complications from these terms, but essentially they can be handled like the $k = 1$ term with an additional exponentially decreasing factor, so we consider only the $k = 1$ term. I would like to emphasize that the main reason for propagating the non-smoothed estimate is to be able to handle this sum over k. We write

$$\left(\hat{g}_L \hat{\Delta}_L \pi_L \Sigma_L\right)(0,\cdot) = \sum_{y\in V_L} \hat{g}_L(0,y)\left(\hat{\Delta}_L \pi_L \Sigma_L\right)(y,\cdot).$$

First, what happens with the summation over the last layer $d_L(y) \leq L/(\log L)^{10}$? There, the coarse graining is also of order $L/(\log L)^{10}$, and $\pi_L\Sigma_L$ is a smooth kernel, spread out on scale L, and it is easy to see that

$$\left\|\pi_L\Sigma_L(x,\cdot) - \pi_L\Sigma_L(x+e,\cdot)\right\|_{\text{var}} \leq \text{const} \times L^{-1}. \tag{3.9}$$

Therefore, as the sum over the $\hat{g}_L(0,y)$ over this last layer is of order 1, one gets deterministically that this part gives only a contribution of order $(\log L)^{-10}$, which is below what we are shooting for. So this is harmless. Actually the whole summation over the boundary region $d_L(y) \leq 2L/(\log L)^3$ is harmless: A bit deeper inside, the coarse graining my be worse than $L/(\log L)^9$, so that we can not argue as crudely, but there we $\pi_L\Sigma_L = \hat{p}_L\pi_L\Sigma_L$ and use that $\hat{p}_L\hat{\Delta}_L$ is only of order $(\log L)^{-9}$, by the induction assumption, and we gain an additional $(\log L)^{-3}$ from (3.9). The summation of the Green's function in $\left\{d_L(x) \leq L/(\log L)^3\right\}$ is only a factor $\log L$, so we see the whole boundary region is harmless. Therefore, there remains the y-summation in the bulk $d_L(y) \geq 2L/(\log L)^3$. There we chop this bulk into subboxes of side-length $L/(\log L)^3$. There are $(\log L)^{3d}$ such subboxes. Then we split up things into summation over y in these subboxes. These parts are not completely independent through some overlapping problems, but they are nearly so, and essentially one can handle them as if they were independent, and then apply some exponential inequality, like Hoeffding's inequality, which together with the estimates on $\hat{g}_L$, and the smoothness of $\pi_L\Sigma_L = \hat{p}_L\pi_L\Sigma_L$, using again that $\hat{\Delta}_L\hat{p}_L$ is of order $(\log L)^{-9}$ does the job of proving that

$$\mathbb{P}\left(\left\|\sum\nolimits_{y\in\text{Bulk}}\hat{g}_L(0,y)\left(\hat{\Delta}_L\pi_L\Sigma_L\right)(y,\cdot)\right\|_{\text{var}} \geq \frac{1}{10(\log L)^9}\right) \leq \frac{1}{20}\exp\left[-(\log L)^2\right].$$

There is however a crucial issue here, namely that for applying the exponential estimates, we have to *center* the random part, namely, we have to subtract

$$\sum_{y\in V_L}\hat{g}_L(0,y)\left(\mathbb{E}\hat{\Delta}_L\pi_L\Sigma_L\right)(y,\cdot).$$

Now, it is crucial that the symmetry assumption we have transfers to the symmetry of $\mathbb{E}\hat{\Delta}_L$, at least in the bulk, and as $\pi_L(x,y)$ is harmonic in the first variable, this leads to a cancellation below the level we are shooting for.

The point is that if one gives up the symmetry assumption on the distribution of the random environment, then one has to take care precisely of expressions of the type $\mathbb{E}\hat{\Delta}_L$, i.e. of the annealed exit distributions, simultaneously with the quenched exit distributions. This looks being quite delicate, and has not been done.

The next thing to do is to estimate the other part in (3.5). Any of the other parts contain a factor $\hat{\Delta}\hat{p}$ twice. Look for instance at

$$\hat{g}\hat{\Delta}\hat{p}\hat{g}\hat{\Delta}\hat{p}\pi_L\Sigma_L.$$

We consider $\hat{g}\hat{\Delta}\hat{p}\pi_L\Sigma_L(x,\cdot)$, which for each x is by the above estimate $\leq \frac{1}{10(\log L)^9}$ up to a probability of order $\frac{1}{20}\exp\left[-(\log L)^2\right]$, so that there is any $x \in V_L$ where this is violated has probability at most of order

$$\frac{L^d}{20}\exp\left[-(\log L)^2\right] \leq \frac{1}{15}\exp\left[-(\log L)^2\right].$$

$\hat{g}\hat{\Delta}\hat{p}$ can be estimated deterministically:

$$\sum_{y} \hat{g}\left(0,y\right) \leq \text{const} \times \left(\log L\right)^{6}, \tag{3.10}$$

and as we have that $\hat{\Delta}\hat{p}$ is in total variation of order $\left(\log L\right)^{-9}$, we are clearly on the good side, and in this way we can handle easily the rest of the perturbation expansion, and in this way get (3.6), and (3.7) is somewhat similar.

Now we have to discuss how to handle the possible one bad region. This region is of size at maximum $L/\left(\log L\right)^{3}$. We write down the perturbation expansion, by splitting it into the various possibilities of summations where in the $\hat{\Delta}\left(x,\cdot\right)$ parts, the x are in the bad regions or are not. If there is no summation in over the bad part, then the contribution is the same as in the "goodified" environment, and we know how it behaves. We argue now, that we don't have to consider "multientries" into the bad region. This is a bid tricky, but essentially (3.7) gives the means to prove that any additional reentry to the last one, gives a factor $\leq 1/2$, so we can do with one entry into the bad region. (This needs some additional small manipulation). Therefore, we have to look at typical contribution of the type

$$\sum_{y\in B_L} \hat{g}_L\left(0,y\right)\left(\hat{\Delta}_L \pi_L \Sigma_L\right)\left(y,\cdot\right).$$

The y summation over B_L of the Green's function is harmless: This is at most of order 1, (if the bad region is close to the center). Now, the bad region can be as bad as it is, the badness is bounded by 2 in total variation. Therefore, a really bad region is getting improved to a bad region of order degree 3, unless something is going wrong in the good part, which is happening only with our less than $\exp\left[-\left(\log L\right)^{2}\right]$ probabilities. In the same way, something bad of level 3 is upgraded to level 2, level 2 is upgraded to level 1, and level 1 is upgraded to "good" on the next scale. This is the reason for having these 4 levels of badness.

(Of course one might ask why we don't increase the scale in such a way that "bad" becomes "good" in one shot. This could be done by increasing scales in steps $l \to l\left(\log l\right)^{10}$, but the problem would be that in (3.10), we would catch something like $\left(\log L\right)^{20}$ on the right hand side, which would kill us there, and then one would have to do the nonlinear part of the perturbation expansion in a *much* more sophisticated way. Therefore, I think that to distinguish several levels of badness is unavoidable).

3.4 Advancing the non-smoothed estimate

For getting information about $\left\|\Pi_L - \pi_L\right\|_{\text{var}}$, we cannot stop to refine the coarse graining close to the boundary, and therefore, we take the original coarse graining scheme. The drawback is that we have now to take into account multiple bad regions close to the boundary. For instance, if the scale is L^a, $a<1$, (meaning that we are about at the same distance from the boundary), then the probability for $V_{L^a}\left(x\right)$ to be bad is estimated

only by $\exp\left[-(1-3/13)(\log L^a)^2\right]$, so it is clear that we cannot exclude multiple bad boxes if we shoot for an estimate of order $\exp\left[-(\log L)^2\right]$.

We split the boundary region of V_L up into layers $\Lambda_j \stackrel{\text{def}}{=} \left\{x \in V_L : 2^{j-1} \leq d_L(x) < 2^j\right\}$, and we chop this layer again into subboxes of about square size, and we distinguish between "good" and "bad" of these subboxes. The probability that any of these subboxes in layer j is bad is about $\exp\left[-\log^2\left(2^j\right)\right] = \exp\left[-\text{const} \times j^2\right]$, and therefore, the further inside we go with the layer, the less frequent are the bad regions inside the layer. Due to essential independence of these bad regions, we can do a large deviation estimate which lead to the fact that if X_j is the number of bad subboxes in layer Λ_j, and N_j is the total number of boxes, then

$$\mathbb{P}\left(X_j > j^{-3/2} N_j, \text{ some } j \text{ with } 2^j \leq \frac{2L}{(\log L)^3}\right) \leq \frac{1}{10} \exp\left[-(\log L)^2\right].$$

We write B for the union of these bad subboxes, neglecting for the moment the possibility that there might still be a bad region in the bulk (which can be incorporated easily). The part of the perturbation expansion where there are *no* summands over $\hat{\Delta}(x, \cdot)$ with $x \in B$ is simply the exit distribution in a goodified environment, and this can be estimated easily. So there remains the part which has summands, perhaps multiple ones, in the bad region. In principle, this might become quite complicated as there may be clusters of bad regions, but we want to avoid any discussion of this issue, which we can, because we are shooting only for *non-smoothed* exit distributions. The point is that we stop the expansion after the first appearance of a summation in B. Stopping means, that we leave afterwards the original RWRE object. To see this, we remark that when iterating (3.1), we don't have to iterate it infinitely many times to arrive at the expansion (3.2), but we can stop the expansion whenever we like, for instance after entering a bad region in the boundary for the first time. The price we have to pay for this is that we then have to use the RWRE kernels G_U or Π_U as the last factor, but Π_U is bounded in total variation by 2. Then we resume the part before the first summation inside B, which again just gives the RWRE Green's function, evaluated only in the goodified environment. So the "bad" part of the expansion for $\|(\Pi_L - \pi_L)(0, \cdot)\|_{\text{var}}$ gives

$$\sum_{y \in B} \left\| \hat{G}_L^{\text{goodified}}(0, y)\, \hat{\Delta}\Pi_L(y, \cdot) \right\|_{\text{var}}.$$

Now, we estimate by brute force $\left\|\hat{\Delta}\Pi_L(y, \cdot)\right\|_{\text{var}} \leq 2$, and get

$$\sum_{y \in B} \left\| \hat{G}_L^{\text{goodified}}(0, y)\, \hat{\Delta}\Pi_L(y, \cdot) \right\|_{\text{var}} \leq 2 \sum_{y \in B} \hat{G}_L^{\text{goodified}}(0, y),$$

but the goodified RWRE Green's function, we can estimate $\hat{G}_L^{\text{goodified}}(0, y)$ using (3.7) by $\text{const} \times \hat{g}_L(0, y)$, and we are done. (There is some fooling around with the layers very close to the boundary, but for any fixed $J \in \mathbb{N}$, we can choose ε small enough that there is absolutely no bad region in these layers).

In this way one can advance the estimates of the smoothed estimate and the non-smoothed estimate. In the induction procedure, the non-smoothed estimate is used also to advance the estimate of the smoothed estimate, namely to handle the expression (3.8) for $k \geq 2$. The advancement of the non-smoothed estimate is relying on the smoothed estimate at many places. However, besides helping to advance the smoothed estimate, the proof of the advancement of the non-smoothed estimate can be slightly modified to really prove the statement of the theorem, namely that depending on the deviation we are shooting for, we can choose an appropriate smoothing, and furthermore, or we increase the smoothing with L in an arbitrary way, the total variation deviation goes to 0. (Theorem 3.2 and Remark 3.3). This just need a slight modification of the argument: If we have a fixed smoothing scale ζ, then one can handle some of the layers close to the boundary in the way we did in the previous section, because there we use the smoothing by the scale ζ, and therefore the bigger ζ is the more of the layers we can leave out, so that in the summation above over $y \in B$, we can restrict to layers deeper inside which give then less of a contribution. So in fact the whole proof gives that $\mathrm{Cond}\,(\delta, L_1)$ implies that for $L \leq L_1 \left(\log L_1\right)^2$

$$\mathbb{P}\left(\left\|\left(\Pi_L - \pi_L\right)\Sigma_{\zeta(\eta)}\left(0,\cdot\right)\right\|_{\mathrm{var}} \geq \eta\right) \leq \exp\left[-\left(\log L\right)^2\right]$$

provided $\zeta\left(\eta\right)$ is chosen properly.

References

[1] Berger, N., and Zeitouni, O., *A quenched invariance principle for certain ballistic random walks in i.i.d. environments.* Preprint

[2] Bolthausen, E., and Goldsheid, I., *Recurrence and transience of random walks in random environments on a strip.* Comm. Math. Phys. **214**, 429-447 (2000)

[3] Bolthausen, E., and Goldsheid, I., *Lingering random walks in random environment on a strip.* Comm. Math. Phys. **278**, 253-288 (2008)

[4] Bolthausen, E., and Sznitman. A.-S., *On the static and dynamic points of view for certain random walks in random environment.* Methods Appl. Anal., **9**, 345–375, 2002. Special issue dedicated to Daniel W. Stroock and Srinivasa S. R. Varadhan on the occasion of their 60th birthday.

[5] Bolthausen, E., and Zeitouni, O., *Multiscale analysis for exit distributions of random walks in random environments.* Prob. Theory and Rel. Fields. **138**, 581-645 (2007)

[6] Bricmont, J., and Kupiainen, A., *Random walks in asymmetric random environments.* Comm. Math. Phys. **142**, 345-420 (1991)

[7] Goldsheid, I.: *Linear and sub-linear growth and the CLT for hitting times of a random walk in random environment on a strip.* Prob. Theory and Rel. Fields **141**, 471-511 (2007)

[8] Kalikov, S.A., *Generalized random walk in random environment.* Ann. Prob. **9**, 753-768 (1981).

[9] Kesten, H. *The limiting distribution of Sinai's random walk in random environment.* Physics A **138**, 299-309 (1986).

[10] Lawler, G., *Weak convergence of a random walk in a random environment.* Comm. Math. Phys. **87**. 81-87 (1982)

[11] Letchikov, A.V., *Localization of one-dimensional random walk in random environment.* Soviet Scientific Reviews Section C. Harwood Academic Publishers 173-220 (1989).

[12] Rassoul-Agha, F., and Seppäläinen, T., *Almost sure functional limit theorem for ballistic random walk in random environment.* Preprint

[13] Keane, M.S., and Rolles, S., *Tubular recurrence.* Acta Mathematica Hungarica **97**, 207-221 (2002)

[14] Sinai, Ya. G., *The limiting behavior of a one-dimensional random walk in a random environment.* Theory Prob. Appl. **27**, 247-258 (1982).

[15] Solomon, F., *Random walks in random environments.* Ann. Prob. **3**, 1-31 (1975).

[16] Sznitman, A.-S., and Zerner, M., *A law of large numbers for random walks in random environments.* Ann. Prob. **27**, 1851-1869 (1999).

[17] Sznitman, A.-S., and Zeitouni, O., *An invariance principle for isotropic diffusions in random environment.* Invent. Math. **164**, 455-567 (2006)

[18] Zeitouni, O., *Random walks in random environments.* Lecture Notes in Mathematics, vol. 1837 pp. 190-312. Springer, Heidelberg (2004)